



GI-Edition



**Lecture Notes
in Informatics**

Heinrich C. Mayr, Martin Pinzger (Hrsg.)

INFORMATIK 2016

**26.–30. September 2016
Klagenfurt**

Proceedings



Heinrich C. Mayr, Martin Pinzger (Hrsg.)

INFORMATIK 2016

26. – 30. September 2016
Klagenfurt, Österreich

Gesellschaft für Informatik e.V. (GI)

Lecture Notes in Informatics (LNI) - Proceedings

Series of the Gesellschaft für Informatik (GI)

Volume P-259

ISBN 978-3-88579-653-4

ISSN 1617-5468

Volume Editors

Univ.-Prof. Dr. Dr. h.c. Heinrich C. Mayr
Alpen-Adria-Universität Klagenfurt
Universitätsstraße 65-67, Österreich
heinrich.mayr@aau.at

Univ.-Prof. Dipl.-Ing. Dr. Martin Pinzger
Alpen-Adria-Universität Klagenfurt
Universitätsstraße 65-67, Österreich
martin.pinzger@aau.at

Series Editorial Board

Heinrich C. Mayr, Alpen-Adria-Universität Klagenfurt, Austria
(Chairman, heinrich.mayr@aau.at)
Dieter Fellner, Technische Universität Darmstadt, Germany
Ulrich Flegel, Infineon, Germany
Ulrich Frank, Universität Duisburg-Essen, Germany
Johann-Christoph Freytag, Humboldt-Universität zu Berlin, Germany
Michael Goedicke, Universität Duisburg-Essen, Germany
Ralf Hofestädt, Universität Bielefeld, Germany
Michael Koch, Universität der Bundeswehr München, Germany
Axel Lehmann, Universität der Bundeswehr München, Germany
Thomas Roth-Berghofer, University of West London, Great Britain
Peter Sanders, Karlsruher Institut für Technologie (KIT), Germany
Sigrid Schubert, Universität Siegen, Germany
Ingo Timm, Universität Trier, Germany
Karin Vosseberg, Hochschule Bremerhaven, Germany
Maria Wimmer, Universität Koblenz-Landau, Germany

Dissertations

Steffen Hölldobler, Technische Universität Dresden, Germany

Seminars

Reinhard Wilhelm, Universität des Saarlandes, Germany

Thematics

Andreas Oberweis, Karlsruher Institut für Technologie (KIT), Germany

© Gesellschaft für Informatik, Bonn 2016

printed by Köllen Druck+Verlag GmbH, Bonn

Vorwort

Sehr geehrte Leserin, sehr geehrter Leser,

in den Anfängen der Informatik ging es im Wesentlichen darum, aus heutiger Sicht teure, aber in ihrer Leistung sehr bescheidene Hardware für Anwendungen nutzbar zu machen. Dabei galt es, die grundlegenden Methoden und Techniken für die Konstruktion datenintensiver, sicherer und effizienter Betriebs- und Softwaresysteme zu erarbeiten. Im Tempo des Moore'schen Gesetzes hat sich diese ehemals überschaubare Disziplin zu einem Gebiet entwickelt, das aus keinem Wissenschaftsbereich und kaum einem Lebensbereich mehr wegzudenken ist. Dementsprechend heterogen und facettenreich ist heute die „Familie“ derjenigen, die sich mit Fragen der Informatik befassen, genauso wie es diese Fragen selbst sind.

Seit 1971 treffen sich einmal im Jahr zahlreiche Mitglieder vorwiegend des deutschsprachigen Zweigs dieser Familie, um sich über ihre aktuellen Ergebnisse, Forschungs- und Entwicklungsprojekte, aber auch Themen der Gestaltung von Studium und Lehre auszutauschen: auf der Jahrestagung der Gesellschaft für Informatik, die seit 1996 auch den Beinamen INFORMATIK trägt. Anders als bei Fachtagungen zu bestimmten Teilgebieten steht hier der Blick über den Tellerrand und die Kommunikation zwischen den unterschiedlichsten Interessensgruppen im Vordergrund – all dies entlang eines aktuellen Leitthemas.

Informatiksysteme werden von Menschen gemacht und sie haben den Menschen in den Mittelpunkt zu stellen: mit seinen Erwartungen, seinen Möglichkeiten und Bedürfnissen in einer Welt des beschleunigten Privat- und Arbeitslebens und der zunehmend komplexen Aufgaben und Abläufe. Funktionalität, Wirtschaftlichkeit und Gebrauchstauglichkeit stehen dabei gleichermaßen im Fokus wie Sicherheit, Effizienz und Effektivität. Vor allem müssen Informatiksysteme beherrschbar, unaufdringlich und verstehbar bleiben. Technik hat sich dem Menschen anzupassen und nicht umgekehrt. All dies spiegelt sich im Leitthema des diesjährigen Familientreffens wieder: *„Informatik: von Menschen und für Menschen“*.

Vor Ihnen öffnet sich nun die Zusammenschau der in insgesamt 205 Einzelbeiträgen behandelten Fragestellungen. Deren Spektrum ist so breit, dass der Rahmen eines Vorworts gesprengt würde, wollte man sie alle nennen. Daher begnügen wir uns hier mit einer groben Einteilung in folgende, generische Themenbereiche:

- Methoden, Techniken und Standardisierung,
- Software-, Data- und Knowledge Engineering,
- Schutz und Sicherheit,
- Informations- und Geschäftsprozessmanagement,
- Informatik und Hochschule, Lehre, und Gesellschaft,
- Informatik und Nachhaltigkeit,

- und besonders aktuell: Industrie 4.0 und Anwendungen.

Die Diskussion selbst erfolgte einerseits im Rahmen von Plenarvorträgen herausragender Expertinnen und Experten aus Wirtschaft und Wissenschaft am zentralen „*Tag der Informatik*“ und am „*Tag der Wirtschaft*“; und andererseits in einer Vielzahl von Workshops, jeweils mit eigenem Programmkomitee, Tutorials, speziellen Studierenden- und Doktorandenprogrammen, Podiumsdiskussionen und anderen Formaten. All dies ist im vorliegenden Tagungsband dokumentiert.

Damit aber nicht genug, denn zeitgleich fanden auch die folgenden weiteren Fachkonferenzen unter dem Dach der INFORMATIK 2016 statt:

- die D-A-CH Security 2016,
- die 39. German Conference on Artificial Intelligence KI 2016,
- die 14. German Conference on Multiagent System Technologies und
- die D-A-CH Energieinformatik 2016.

Ein Ereignis dieser Größenordnung kann in einer gemeinnützigen Gesellschaft nur durch ein außergewöhnliches ehrenamtliches Engagement aller Akteure erfolgreich geplant und durchgeführt werden. Zählt man die Mitglieder der einzelnen Programmkomitees, die Hauptvortragenden, die AutorInnen der Beiträge und die Organisationskomitee-Mitglieder zusammen, dann haben über 500 Menschen zum Erfolg der Informatik 2016 beigetragen. Ihnen sei an dieser Stelle herzlichst auch im Namen des Vorstands der Gesellschaft für Informatik gedankt.

Klagenfurt im September 2016

Heinrich C. Mayr und Martin Pinzger

Veröffentlicht mit Unterstützung des Forschungrates der Alpen-Adria-Universität
Klagenfurt

Sponsoren

Wir danken den nachfolgenden Unternehmen und Institutionen für die Unterstützung der
Konferenz.

**Atos IT Solutions and Services GmbH
(Platin Sponsor)**



SIC Software Internet Cluster Kärnten



**UBIT – Unternehmensberater,
Buchhaltung und IT-Dienstleister**



SER Solutions Österreich GmbH



Google International GmbH Austria



Pearson Deutschland GmbH



Walter De Gruyter GmbH



**Springer Vieweg | Fachmedien Wiesbaden
GmbH**



BOC Products & Services AG



Stadtwerke Klagenfurt AG



Amazon Web Services, Inc.



Kapsch AG



Ilogs- Mobile Software GmbH



M/O/T Alpen-Adria School of Management, Organizational Development and Technology



Infineon Technologies Austria AG



Academy Smart



build! Gründerzentrum Kärnten GmbH



Kelag- Kärntner Elektrizitäts-Aktiengesellschaft



Weingut Burg Taggenbrunn GmbH



Förderverein der Technischen Fakultät



Industriellenvereinigung Kärnten



Kärntner Universitätsbund



Alpen-Adria-Universität Klagenfurt



Tagungsteam

H. C. Mayr, M. Pinzger (**Tagungsleitung, Tag der Informatik, Workshops, Tutorials**)

D. Jandl, A. Loisel, M. Zandonella, M. Zanker (**Tag der Wirtschaft**)

P. Schartner (**D-A-CH Security**)

G. Friedrich (**KI2016**)

M. Klusch (**MATES 2016**)

F. Kupzog, W. Elmenreich (**D-A-CH Energieinformatik 2016**)

M. Lux, St. Wagner, E. Denert (**Programm für Doktorand/inn/en**)

J. Michael und Studierende (**SKILL, Studierendenprogramm**)

J. Michael (**Ausstellungen**)

Ch. Kop (**Publikationen**)

F. Al Machot, G. Morak, St. Beyer, R. S. P. Ranasinghe (**Lokale Organisation AAU**)

C. Steinberger, C. Winter (**Pressearbeit**)

D. Werdnig, E. Kerkmann, K. Unterköfler (**Sekretariat, Finanzen**)

A. Lippitsch (**Organisation Hotels, Sekretariat**)

St. Leitner, M. Maier, A. Schorn (**Technik**)

M. Wolfger (**Web-Redaktion**)

Inhalt

Hauptvorträge

Michael Wooldridge <i>From model checking to equilibrium checking</i>	35
Thomas Eiter <i>Artificial Intelligence at the Gates of Dawn?</i>	37
Hanns-Thomas Kopf <i>Security demands and attempts in the new digital world</i>	39
Andreas Zeller <i>Software verstehen, zerstören, schützen mit automatischen Software-Modellen</i>	41
Andreas Reuter <i>Forschung ohne Informatik – geht das?</i>	43
Peter Liggesmeyer <i>IoT, Industrie 4.0 und Big Data: Herausforderungen für die Informatik und die GI</i>	45
Sabine Herlitschka <i>Industrie 4.0 und zukünftige Arbeitswelten: Praxis und Vision von Infineon Technologies Austria</i>	47
Michael May <i>Towards Industrial Machine Intelligence</i>	49
Friedrich Georg Hoepfner <i>Fit für den härtesten Job der Marktwirtschaft - Wie unsere Gründer spitze werden könnten</i>	51

Workshops

Software-Entwicklungsprozesse und –Produkte im Licht kultureller Unterschiede

Gerhard Chroust, Renate Motschnig <i>Vorwort: Software-Entwicklungsprozesse und -Produkte im Licht kultureller Unterschiede</i>	59
Gerhard Chroust <i>Kulturelle Unterschiede und ihr Einfluss auf Software-Entwicklungsprozesse und –Produkte</i>	63

Eckhart Hanser <i>Erweiterung agiler Prozesse durch sozietäre Vorgehensbausteine</i>	77
Christina Böhm, Renate Motschnig <i>Ein Diversity Framework für Informatik-nahe Projekte</i>	87
Walter Renner, Shiva Khalili <i>Independent vs. interdependent identities: Eine theoretische Basis zu kulturellen Unterschieden in der Entwicklung und Nutzung von Software</i>	95
Walter Khom, MykolaTkachuk, Volodymyr Sokol, Oleksiy Kosmachov, Vladyslav Sokol <i>Communication, Management and Teambuilding Issues in Austrian-Ukrainian Outsourcing Project: 10 Years of Experience and Future Challenges</i>	103
Vladimir A. Shekhovtsov, Heinrich C. Mayr <i>View Harmonization in Software Processes: from the Idea to QuASE</i>	111
 Software Qualität ist wie Schönheit	
Sabine Wieland, Andreas Hartmann <i>Zusammenfassung: Software Qualität ist wie Schönheit</i>	127
Sabine Wieland, Andreas Hartmann <i>Software Qualität ist wie Schönheit</i>	129
Mykola Tkachuk, Iryna Martinkus, Rustam Gamzayev, Andrii Tkachuk <i>An Integrated Approach to Evaluation of Domain Modeling Methods and Tools for Improvement of Code Reusability in Software Development</i>	143
Stefan Strell, Vladimir A. Shekhovtsov, Heinrich C. Mayr <i>Beurteilung von Software Qualität stärken durch Ontologien und Wiederverwendung von Wissen</i>	157
Christoph Deppisch <i>Citrus – Integrationsframework mit Demonstration</i>	171
 Software-Engineering für Multi-Plattform-Umgebungen	
Ruth Breu, Wolfgang Glock, Gabriele Haller, Thomas Matzner, Gerhard Müller, Friederike Nickl, Oliver Wiegert <i>Software-Engineering für Multi-Plattform-Umgebungen</i>	177

Suchmaschinen: Zwischen technischer Entwicklung und gesellschaftlicher Bedeutung

Olaf Resch <i>Suchmaschinen: Zwischen technischer Entwicklung und gesellschaftlicher Bedeutung</i>	181
Dirk Lewandowski <i>Die Suchergebnisseite als Dauerwerbesendung?</i>	183
Franz Kummer, Blaise Dévaud <i>Juristische Suchmaschine Lawsearch Enterprise – drei Herausforderungen</i>	195
Timo Schulz, Björn Zessack, Bengt Rabe, Sebastian Kielmann <i>Möglichkeiten und Herausforderungen bei der Verwendung von Bildern in Suchmaschinen</i>	205
Sebastian Sünkler, Friederike Kerkmann, Sonja Quirnbach, Sebastian Schultheiß <i>Sprachgesteuerte Websuche im Test – Ein Erfahrungsbericht aus der Perspektive von Auftraggeber, Auftragnehmer und Durchführendem</i>	211
Irene Teich, Peter Schnupp <i>Sinnsuche noch günstiger: wie Google an Inbegriffen scheitert</i>	227
Olaf Resch <i>Visualisierung von Suche im Dialog</i>	241
Value Sensitive Design of Internet-Based Services: Towards an Integration of Technology and Values	
Hannes Hartenstein, Alexander Maedche, Martina Zitterbart <i>Value Sensitive Design of Internet-Based Services: Towards an Integration of Technology and Values</i>	253
Markus Ast, Martin Gaedke <i>Value-Sensitive Design in Hyper-Connected Societies</i>	255
Teemu Kärkkäinen, Jörg Ott <i>The Case for Liberouters</i>	261
Stefanie Betz, Andreas Fritsch <i>A Comparison of Value Sensitive Design and Sustainability Design</i>	267
Till Neudecker, Arsen Hayrapetyan, Alexander Degitz, Philipp Andelfinger <i>Consideration of Values in the Design of Access Control Systems</i>	275

Verena Dorner, Stefan Morana, Alexander Maedche, Christof Weinhardt <i>Towards designing individual value-oriented decision support for selecting internet-based services</i>	289
Valentin Burger, Matthias Hirth, Tobias Hoßfeld, Phuoc Tran-Gia <i>Principles of Information Neutrality and Counter Measures Against Biased Information</i>	295
 2nd International Workshop on Big Data, Smart Data and Semantic Technologies – BDS DST 2016	
Nico Roedder, Stefan Zander, Natalja Kleiner, Suad Sejdovic, Stefan Jähnichen, Rudi Studer <i>2nd International Workshop on Big Data, Smart Data and Semantic Technologies – BDS DST 2016</i>	311
Florian Weber, Andreas Bihlmaier, Heinz Wörn <i>Semantic Object-Oriented Programming (SOOP)</i>	313
Lars Bollweg, Richard Lackes, Markus Siepermann, Peter Weber <i>In-Store Customer Analytics – Metriken & Reifegradszenarien zur Erfassung physischer Kundenkontakte im stationären Einzelhandel</i>	327
Timo Schindler, Christoph Skornia <i>Secure Parallel Processing of Big Data Using Order-Preserving Encryption on Google BigQuery</i>	343
Marco F. Huber, Martin Voigt, Axel-Cyrille Ngonga Ngomo <i>Big Data Architecture for the Semantic Analysis of Complex Events in Manufacturing</i>	353
Christoph Pflügler, Thomas Köhn, Maximilian Schreieck, Manuel Wiesche, Helmut Krcmar <i>Predicting the Availability of Parking Spaces with Publicly Available Data</i>	361
Maximilian Schreieck, Christoph Pflügler, Christoph Dehner, Samit Vaidya, Simon Bönisch, Manuel Wiesche, Helmut Krcmar <i>A Concept of Crowdsourced Delivery for Small Local Shops</i>	375
Kevin Laubis, Viliam Simko, Christof Weinhardt <i>Weighted Aggregation in the Domain of Crowd-Based Road Condition Monitoring</i>	385
Daniel Breucker, Benedikt Kämpgen, Andreas Harth <i>SEO4OLAP – Search Engine Optimized Presentation of Statistical Linked Data</i>	395

Recht und Technik – Datenschutz im Diskurs

Rüdiger Grimm, Christoph Sorge, Indra Spiecker genannt Döhmann <i>Recht und Technik: Datenschutz im Diskurs</i>	411
Sebastian Volkmann, Linus Feiten, Christian Zimmermann, Sebastian Sester, Laura Wehle, Bernd Becker <i>Digitale Tarnkappe: Anonymisierung in Videoaufnahmen</i>	413
Paul Georg Wagner, Pascal Birnstill, Erik Krempel, Sebastian Bretthauer, Jürgen Beyerer <i>Privacy-Dashcam – Datenschutzfreundliche Dashcams durch Erzwingen externer Anonymisierung</i>	427
Meiko Jensen, Nils Gruschka, Jens Lüssem <i>Datenschutz im Fahrzeug der Zukunft: Vernetzt, Autonom, Elektrisch</i>	441
Christoph Bier <i>Datenschutzziele im Konflikt: Eine Metrik für Unverkettbarkeit als Hilfestellung für den Betroffenen</i>	455
Maren Pollmann, Dennis-Kenji Kipker <i>Eingeschränkte Selbstbestimmung im Onlineverkehr: Stärkung der Ein- willigungserklärung durch Einführung vorformulierter Datenschutz- bestimmungen</i>	469
 Nutzerzentrierte Sicherheit – NzS 2016	
Andreas Heinemann, Steffen Lange, Marian Margraf, Jörg Willomitzer <i>Workshop: Nutzerzentrierte Sicherheit – NzS 2016</i>	485
Peter Weierich, Tobias Schmidt, Sebastian Abeck <i>Usability der Identity- und Accessmanagementkomponenten von Endkundenportalen</i>	487
Helmar Hutschenreuter, Dieter Hutter <i>Beherrschbarkeit komplexer Firmeninfrastrukturen durch eigenschafts- basierte Sicherheit</i>	497
Alexandra Kunz, Melanie Volkamer, Simon Stockhardt, Sven Palberg, Tessa Lottermann, Eric Piegert <i>NoPhish: Evaluation of a web application that teaches people being aware of phishing attacks</i>	509
Sandra Kostic, Andreas Heinemann, Marian Margraf <i>Usability-Untersuchung eines Papierprototypen für eine mobile Online- Ausweisfunktion des Personalausweises</i>	519

Workshop on New Security Standards for IACS/SCADA Industrial Systems

Jan deMeer, Karl Waedt <i>Workshop on New Security Standards for IACS/SCADA Industrial Systems</i>	531
Jan deMeer, Karl Waedt <i>New Security Standards for Industrial Automation and Control Systems, based on IEC 62443-4-2 (IACS/SCADA) - Extended Abstract</i>	537
Scott Cadzow, Alexander Cadzow <i>Humans - the weakest and strongest link in securing systems</i>	543
Mithil Parekh, Yuan Gao, Deeksha Gupta, Christian Luschmann <i>OPANSec – Security Integrity Monitoring for Controllers</i>	547
Yuan Gao, Xin Xie, Mithil Parekh, Edita Bajramovic <i>SIEM: Policy-based Monitoring of SCADA Systems</i>	559
Simon Seibt, Karl Waedt, Hans Delfs, Simon Odorfer <i>3D Modeling of Selected Assets, Security Zones and Conduits</i>	571
Jianghai Li, Edita Bajramovic, Yuan Gao, Mithil Parekh <i>Graded Security Forensics Readiness of SCADA Systems</i>	581

Risikomodelle von komplexen Systemen

Erich Schweighofer, Felix Breitenecker <i>Risikomodelle von komplexen Systemen</i>	595
Erich Schweighofer, Thomas Preiß <i>Risikoanalyse im Recht – eine neue juristische Methodik?</i>	597
Ingrid Schäumüller-Bichl, Andrea Kolberger <i>Information Security Risk Analysis in komplexen Systemen - neue Herausforderungen und Lösungsansätze</i>	609
Manfred Holzbach <i>Effiziente Risikoanalyse anhand praktischer Erfahrungsbeispiele</i>	619
Martin Setnicka <i>Predictive Analytics in der österreichischen Finanzverwaltung</i>	629
Maik Günther <i>Einsatz eines Gasmarktmodells zur Bewertung von Risiken</i>	635
Thomas Preiß <i>Konzept einer Risikoanalyse im Verwaltungsverfahren durch agentenbasierte Simulation</i>	649

Business Process Management und Strategisches Informationsmanagement im Öffentlichen Sektor

Dagmar Lück-Schneider, Konrad Walser <i>Business Process Management und Strategisches Informationsmanagement im Öffentlichen Sektor</i>	657
Dagmar Lück-Schneider <i>Process-Mining in der öffentlichen Verwaltung</i>	659
Konrad Walser, Anna Barbara Brechbühl, Reto Zraggen <i>IT-Sourcing in Schweizer Kommunen – Fallstudien aus dem Kanton Bern</i>	671
Christian Kücherer, Jan David Liebe, Michael Schaaf, Johannes Thye, Barbara Paech, Alfred Winter, Franziska Jahn <i>The Status Quo of Information Management in Hospitals - Results of an Online Survey</i>	685
André Nursinski-Stolberg, Kiritharan Gangatharan, Christian Czarnecki <i>Development of a Subject-Oriented Reference Process Model for the Telecommunications Industry</i>	699
Workshop zum Stand, den Herausforderungen und Impulsen des Geschäftsprozessmanagements	
Michael Fellmann, Ralf Laue, Agnes Koschmider, Andreas Schoknecht <i>Workshop zum Stand, den Herausforderungen und Impulsen des Geschäftsprozessmanagements</i>	715
Martin Bergner, Hans-Georg, Fill Florian Johannsen <i>Supporting Business Process Improvement with Natural Language Processing: A Model-based Approach</i>	717
Michaela Baumann, Michael Heinrich Baumann, Stefan Jablonski <i>On Behavioral Process Model Similarity Matching: A Centroid-based Approach</i>	731
Michaela Baumann, Michael Heinrich Baumann, Lars Ackermann, Stefan Schöning, Stefan Jablonski <i>Ansätze zum Ähnlichkeitsabgleich von deklarativen Geschäftsprozessmodellen</i>	733
Jana-Rebecca Rehse, Philip Hake, Peter Fettke, Peter Loos <i>Inductive Reference Model Development: Recent Results and Current Challenges</i>	739

Christina Niemöller, Dirk Metzger, Michael Fellmann, Deniz Özcan, Oliver Thomas <i>Shaping the Future of Mobile Service Support Systems – Ex-Ante Evaluation of Smart Glasses in Technical Customer Service Processes</i>	753
Enterprise Architecture Management	
Carsten Brockmann, Eldar Sultanow <i>Enterprise Architecture Management – Aktueller Stand und Ausblick</i>	771
Benjamin Weinert, Axel Hahn, Oliver Norkus <i>A domain-specific architecture framework for the maritime domain</i>	773
André Ullrich, Edzard Weber <i>Einsatz stilisierter Fakten zur Bewertung wandlungsfähiger Unternehmensarchitekturen</i>	785
Lars Santesson, Tim Piontek, Marvin Zielaskowski <i>Leichtgewichtige IT-Modernisierung in der öffentlichen Verwaltung mit der Hilfe von EAM</i>	799
Christian Czarnecki <i>Design und Nutzung einer industriespezifischen Referenzarchitektur für die Telekommunikationsindustrie</i>	807
Benedict Bender, Marcus Grum <i>Entwicklung eines Architekturkonzepts zum flexiblen Einsatz von Analytics</i>	815
Eldar Sultanow, Carsten Brockmann, Kai Schroeder, Carsten Breithaupt <i>Lufthansa Aviation Standard: Developing an Open Group Reference Architecture for the Aviation Industry</i>	825
IT-Governance und Strategisches Informationsmanagement (ITG-SIM)	
Daniel F. Abawi, Matthias Goeken, André Miede <i>IT-Governance und Strategisches Informationsmanagement (ITG-SIM)</i>	839
Nicola Bigler, Konrad Walser <i>Entwicklung eines Werkzeugs zur Standortbestimmung von IT-Abteilungen basierend auf COBIT 5</i>	841
Matthias Goeken, Stefanie Looso <i>Anwendung von Referenzmodellen der IT-Governance: Konzeptionelle Grundlage und Anwendungsszenarien</i>	855

Tomasz Kusber, Steffen Schwalm <i>Elektronische Langzeitspeicherung als SOA-Dienst – Kernelement eines vertrauenswürdigen Informationsmanagements</i>	869
Michael Adelmeyer, Jan Heinrich Beinke, Marc Walterbusch, Ricardo Ramos Gameiro, Peter König, Frank Teuteberg <i>Eye-Tracking zur Untersuchung von Vertrauenssignalen auf Webseiten von Cloud Computing-Anbietern</i>	883
Hochschule 2026	
Ulrike Lucke, Markus von der Heyde, Wolfgang Meyer <i>Hochschule 2026</i>	899
Sebastian Herwig, Stefan Schlattmann <i>Eine wirtschaftsinformatische Standortbestimmung von Forschungsinformationssystemen</i>	901
Gunnar Auth, Steffen Künstler <i>Erfolgsfaktoren für die Einführung integrierter Campus-Management-Systeme – eine vergleichende Literaturanalyse mit praxisbezogener Evaluation</i>	915
Ingo Hengstebeck, Sarah Grzemski <i>Neue Wege des Kundensupports im IT-ServiceDesk des IT Centers der RWTH Aachen University</i>	933
Jannica Budde, Gudrun Oevel <i>Innovationsmanagement an Hochschulen: Maßnahmen zur Unterstützung der Digitalisierung von Studium und Lehre</i>	947
Christoph Igel, Ansgar Baums, Marc Göcks, Martin Haag, Bernd Krämer, Stephan Pfisterer, Siegfried Stiehl, Guido Wirtz, Volker Zimmermann, Jens Mühlner <i>Intelligente Bildungs- und Hochschulnetze in Deutschland - Wo stehen wir in 2016? Was ist zu tun bis 2020?</i>	961
Kai Dornseiff, Yvonne Groening, Stephan Kassarke <i>Gelebte Kooperationsformen zwischen Hochschulen – Möglichkeiten der Ausgestaltung kritischer Erfolgsfaktoren anhand von Praxisbeispielen</i>	967
Fritz Rose, Dirk Bernds, Michael Gassert, Folke-Gert Stümpel, Ulrike Lucke <i>Integration von Lösungen für das Bewerbungsmanagement mit Job-Portalen und beruflichen Netzwerken</i>	981

Digitalisierung in Lehre und Lernen: des Kaisers neue Kleider?!

Sarah Sahl, Alke Martens <i>Digitalisierung in Lehre und Lernen: des Kaisers neue Kleider?!</i>	999
Sarah Freytag, Sarah Sahl, Birger Lantow, Alke Martens <i>Vorgehensmodelle für die Konzeption digitaler Lehr-Lern-Angebote: Theoretisches Gerüst oder praktische Hilfe?</i>	1001
Robin Nicolay, Bastian Schwennigke, Sarah Sahl, Alke Martens <i>Visualisierung konzeptuellen Lernens durch semantische Vernetzung sequenzieller Lehrinhalte</i>	1009
Agnes Koschmider, Detlef Buschfeld <i>Shifting the Process of Exam Preparation Towards Active Learning: A Crowdsourcing Based Approach</i>	1017
 Vorlesung – Quo Vadis? Lehre an der Schnittstelle zwischen Mensch und Technik	
Clemens Cap, Robin Nicolay <i>Vorlesung – Quo Vadis? Lehre an der Schnittstelle zwischen Mensch und Technik</i>	1025
Christopher Könitz, Jakob Diel, Jürgen Cleve <i>Lab 360. Interaktive Lernräume zur Verbesserung der Qualität der Lehre an der Hochschule Wismar</i>	1027
Tariqul Islam, Philipp Jäcks, Clemens H. Cap <i>Multiscript – an online student-teacher collaboration platform for classroom lectures</i>	1033
Iris Braun, Felix Kapp, Tenshi Hara, Frank Siebert <i>SMART Lecture - die technologiegestützte Neuausrichtung der Vorlesung hin zu einer interaktiven Lernumgebung für selbstreguliertes Lernen</i>	1041
 SAG WAS Studentische Ausbildung und berufliche Weiter- bildung in Allgemeinen Schlüsselqualifikationen	
Karin Vosseberg <i>SAG WAS – Studentische Ausbildung und berufliche Weiterbildung in Allgemeinen Schlüsselqualifikationen</i>	1053
Yvonne Sedelmaier, Dieter Landes <i>Schlüsselqualifikationen in der Ausbildung des Software Engineering – ein Dilemma und ein möglicher Ausweg</i>	1055

Veronika Thurner, Axel Böttcher, Kathrin Schlierkamp, Daniela Zehetmeier <i>Nerd++ – Vermitteln von Schlüsselkompetenzen an angehende Softwerker/innen</i>	1059
Teilhabe an der allgegenwärtigen Kommunikation	
Henning Lübbecke <i>Teilhabe an der allgemeinen Kommunikation</i>	1067
Taisiia Makarova, Andreas Mladenow, Christine Strauss <i>Barrierefreiheit im Internet und Suchmaschinenranking – eine empirische Untersuchung</i>	1071
Nicole Najemnik, Isabel Zorn <i>Digitale Teilhabe statt Doing Disability: Assistive Technologien für inklusive Medienbildung im Kindergarten</i>	1087
Gottfried Zimmermann, Mareen Gutsche, Jonas Häfele, Lena Krächan, Damaris Rothfuß, Hina Marquart, Hermann Frank <i>Barrierefreie Kommunikation beim Paritätischen Wohlfahrtsverband</i>	1097
Dieter Meiller, Karlheinz Müller, Hildegard Legat, Bernd Gerlang <i>Barrierefreiheit im Web für Bildungseinrichtungen</i>	1105
Petra Gröber <i>Zufriedenheit blinder Nutzer mit Zugangsmethoden zu Sozialen Netzwerken: eine theoretische Modellentwicklung</i>	1109
Svetlana Matiouk <i>Re-imagining Deafness as a Resource for Design</i>	1119
Arne Berger, Andreas Bischof, Kevin Lefeuvre, Albrecht Kurze, Michael Storz, Sören Totzauer <i>Loaded Dice: Ein Co-Design Werkzeug zum Entwerfen mit blinden und sehbehinderten Partizipanten im IoT</i>	1131
Mathias Haimerl <i>Interferenzmessung auf Bildschirmen zur Revision überholter Grenzwerte für photosensitive Epilepsie</i>	1137
Workshop Informatik-Begeisterung 2016	
Judith Michael, Ingo Scholtes <i>Workshop: Informatik-Begeisterung 2016</i>	1147
Wolfgang Pohl <i>BWINF: Informatik mit Begeisterung entdecken</i>	1149

Lena Winter, Sebastian Schäfer, Michael Gauß <i>Das Science Camp Informatik – hier können Mädchen Informatik neu erleben</i>	1153
Claudia Steinberger <i>INFORMATIK – wie geil ist denn das!</i>	1157
Alexander Best, Uwe Thierschmann <i>Erste Erfahrungen beim Einsatz von Unterrichtsbausteinen zur kooperativen Entwicklung und Erprobung von Informatikstunden an Grundschulen</i>	1161
Johann Stockinger, Gerald Futschek <i>Informatisches Denken in der Primarstufe: die Wiener Zauberschule der Informatik (WIZIK)</i>	1165
Peter Micheuz <i>Anmerkungen zu einem begeisternden Informatikunterricht</i>	1169
Niklas Kühl, Jonas Lehner <i>Programming for Refugees – An Active Learning Approach for Teaching Java to Heterogeneous Groups</i>	1175
Stefan Pasterk, Heike Demarle-Meusel, Barbara Sabitzer, Andreas Bollin <i>Informatik-Werkstatt: Entwicklungen und Erfahrungen einer Lern- und Lehrwerkstatt für Informatik</i>	1179
Bernhard Löwenstein <i>Initiative IFIT – Den Nachwuchs für IT und Technik begeistern!</i>	1185
Thorsten Schoormann, Dennis Behrens, Ralf Knackstedt <i>Das Hildesheimer Denkwerk-Projekt – Schülerinnen und Schüler modellieren und analysieren Geschäftsprozesse</i>	1189
Daniel Fett, Guido Schmitz <i>Pi and More – Eine Veranstaltungsreihe rund um “kleine Computer”</i>	1195
Alois Bachinger, Anton J. Knierzinger <i>FROM SKETCH TO SCRATCH - schrittweise zu „computational thinking“ geführt werden</i>	1197
Tanja Tomitsch, Thomas Aschinger <i>Begeisterung für Robotik – ein ganzheitlicher Ansatz der (Hoch-)Begabten- förderung mittels Blended-Learning am Beispiel des Talentehauses NÖ</i>	1203
Umwelthinformatik zwischen Nachhaltigkeit und Wandel (UINW 2016)	
Stefan Naumann, Kristina Voigt <i>4. Workshop: Umwelthinformatik zwischen Nachhaltigkeit und Wandel (UINW 2016)</i>	1211

Maik Günther, Michael Schimpf <i>Auswirkungen von Energieeinsparungen auf den Gasbedarf und den Gaspreis in Deutschland bis 2040</i>	1213
Dominik Schöner, Richard Pump, Christian Schmicke, Jan P. Minnrich, Henrik Rüscher, Volker Ahlers, Arne Koschel <i>IT-Unterstützung von BHKW-Prüfständen in der angewandten Forschung</i>	1227
Stefan Naumann, Andrea Christian, Christoph Göttert, Klaus-Uwe Gollmer, Rainer Michels, Stefan Rüffler <i>Energieeinsparungen im Gebäudebetrieb durch visualisiertes Feedback an Nutzer: Datenerfassung und Datenvisualisierung in Nicht-Wohngebäuden</i>	1239
Andrej Matusевич, Volker Wohlgemuth <i>Konzeptionierung einer Webanwendung zur Beurteilung von Stoff- und Energieströmen am Beispiel des Industrie- und Gewerbestandortes Berlin-Schönevide</i>	1251
Maximilian Schneider, Volker Wohlgemuth <i>MOPS - Mobile Unterstützung eines prozessorientierten Stoffstrommanagements in KMU</i>	1263
Hans-Knud Arndt <i>Der Eingabestift – Mode oder ein nachhaltiges Bedienkonzept?</i>	1273
Eva Kern <i>The Development of an Eco-Label for Software Products – a Transdisciplinary Process?</i>	1285
 IT und Dienstleistungen für die Energiewende und Elektromobilität (IDEE)	
Gerhard Satzger, Daniel Beverungen, Martin Matzner, Carola Stryja <i>IT und Dienstleistungen für die Energiewende und die Elektromobilität (IDEE 2016)</i>	1299
Dennis Behrens, Thorsten Schoormann, Ralf Knackstedt <i>Datensets für Demand-Side-Management – Literatur-Review-Basierte Analyse und Forschungsagenda</i>	1301
Erik Kolek, Britta Reinecke, Olexander Filevych <i>Erfolg innovativer E-Carsharing-Geschäftsmodelle – E-Carsharing-Geschäftsmodell-, Elektromobilitäts- und Carsharing-Akzeptanz im öffentlichen Hochschulsektor</i>	1317
Thorsten Schoormann, Dennis Behrens, Ralf Knackstedt <i>Softwaregestützte Modellierung von Geschäftsmodellen – Vergleich und Weiterentwicklungsperspektiven am Beispiel der Business Model Canvas</i>	1333

Franz Lambrecht, Carsten Sommer <i>Konzepte für Mietertickets zur Verknüpfung von Wohnen und Verkehr</i>	1349
Niklas Kühl, Marc Goutier <i>"Need Tweets": New Insights about Customer Needs from Micro Blog Data in the Field of E-mobility</i>	1363
ICT-based Interventions for Water and Heat Energy Conservation	
Thorsten Staake, Anna Kupfer <i>ICT-based Interventions for Water and Heat Energy Conservation</i>	1379
Ilya Kozlovskiy, Samuel Schöb, Mariya Sodenkamp <i>Non-intrusive disaggregation of water consumption data in a residential household</i>	1381
Liliane Ableitner, Samuel Schöb, Verena Tiefenbeck <i>Digitalization of Consumer Behavior – A Descriptive Analysis of Energy Use in the Shower</i>	1389
Liliane Ableitner, Samuel Schöb, Verena Tiefenbeck <i>Quantifying the Potential of Voluntary Energy Efficiency Measures: The Case of Flow Restrictors</i>	1399
Anja Peters, Daniel Hanss, Christian Sartorius <i>Determinants of adopting technological feedback devices and water conservation behaviors</i>	1407
Industrie 4.0 – Wünschenswertes, Machbares und Grenzen für Technik, Wirtschaft und Mensch	
Detlev Buchholz <i>Industrie 4.0 – Wünschenswertes, Machbares und Grenzen für Technik, Wirtschaft und Mensch</i>	1413
Detlev Buchholz <i>Industrie 4.0 und der Mensch</i>	1415
Thuy Duong Oesterreich, Frank Teuteberg <i>Chancen und Risiken der Digitalisierung in der Bauindustrie im Kontext von Industrie 4.0 – Situationsanalyse und Zieldefinition im Zuge einer Technikfolgenabschätzung</i>	1429
Kathrin Schwinghammer, Ricardo Buettner <i>Industrie 4.0-Readiness: Eine empirische Fallstudie in der Automobilproduktion</i>	1445

Jürgen Ensthaler, Martin S. Haase <i>Industrie 4.0 – Datenhoheit und Datenschutz</i>	1459
Herbert Jodlbauer, Michael Schagerl <i>Reifegradmodell Industrie 4.0 - Ein Vorgehensmodell zur Identifikation von Industrie 4.0 Potentialen</i>	1473
 14. Workshop Automotive Software Engineering	
Marcel Baunach, Leandro B. Ribeiro <i>14. Workshop Automotive Software Engineering</i>	1491
Marcus Mikulcak, Thomas Göthel, Paula Herber, Sabine Glesner <i>Towards Identifying Spurious Paths in Combined Simulink/Stateflow Models</i>	1495
Jürgen Wurzinger, Peter Priller, Aleš Kolar, Markus Nager <i>Real World Evaluation of a Novel Security Testing Environment for Vehicular Control Units via CAN Networks</i>	1509
Martin Beckmann, Aaron Schlutter <i>Automatische Duplikateliminierung in Aktivitätsdiagrammen von Fahrzeugfunktionen</i>	1523
Jonas Winkler <i>Automatische Klassifikation von Anforderungen zur Unterstützung von Qualitätssicherungsprozessen</i>	1537
Natalia Ogulenko, Sören Frey, Jens Nahm, Manfred Rössle <i>Cloud-Basierte Optimierung von Fahrzeugbetriebsstrategien durch Clustering mit Genetischen Algorithmen</i>	1551
Peter Brungs, Marcel Baunach <i>Einsatz von dynamisch rekonfigurierbaren FPGAs in Fahrzeugen</i>	1565
Andreas Richter, Michael Scholz <i>Systematische und effiziente Erhebung hochgenauer Straßengeodaten für Geoinformationssysteme und Fahrsimulationen</i>	1579
Sabrina Heppner, Marius Dransfeld, Gitta Domik <i>Adding Atmospheric Scattering and Transparency to a Deferred Rendering Pipeline for Camera Based ADAS Tests</i>	1591
 Additive Fertigung/3D Druck – Technologie, Auswirkungen und Chancen	
Felix Baumann, Dieter Roller <i>Additive Fertigung/3D Druck – Technologie, Auswirkungen und Chancen</i>	1607

Felix Baumann, Oliver Kopp, Dieter Roller <i>Universal API for 3D Printers</i>	1611
Christoph Hochreiner, Philipp Waibel, Michael Borkowski <i>Bridging Gaps in Cloud Manufacturing with 3D Printing</i>	1623
Nils Darwin Abele, Sven Hoffmann, Christopher Kuhnhen, Thomas Ludwig, Walter Schäfer, Marcus Schweitzer, Volker Wulf <i>Supporting the Set-up Processes by Cyber Elements based on the Example of Tube Bending</i>	1627
Jessica Mihalyi, Anne-Kathrin Müller <i>The custom-made body – Legal aspects of bioprinted tissue and organs</i>	1637
Oliver Stickel, Volkmar Pipek <i>Infrastructuring & Digital Fabrication</i>	1649
 Der Arbeitsplatz der Zukunft	
Jan Haase, Gerhard Leitner, Daniel Versick, Raphael Zender <i>Perspektiven des Arbeitsplatzes der Zukunft</i>	1655
Anton Josef Fercher, Gerhard Leitner <i>Das eigene Zuhause als Arbeitsumgebung - Augmented Reality zur Konfiguration und Steuerung im Smart Home</i>	1663
Jan Haase, Dominik Meyer <i>New Trends in Building Automation for Offices of the Future</i>	1677
Hauke Coltzau <i>Integration von User Privacy Mechanismen in clientseitige Standard- bibliotheken</i>	1691
Lilian Schröder, Aljoscha Pörtner, Matthias König, Martin Hoffmann <i>Überlegungen zu Multi-Sensor-Aktor-Systemen und Kontrolle in intelligenten Arbeitsumgebungen</i>	1699
Sarah Brauns, Tobias Käfer, Dirk Koriath, Andreas Harth <i>Individualisiertes Gruppentraining mit Datenbrillen für die Produktion</i>	1713
Benedikt Zobel, Lisa Berkemeier, Sebastian Werning, Oliver Thomas <i>Augmented Reality am Arbeitsplatz der Zukunft: Ein Usability-Framework für Smart Glasses</i>	1727
 IT-Unterstützung in Emergency Management & Response (IT- Rettung)	
Jens Pottebaum, Christian Erfurth <i>IT-Unterstützung in Emergency Management & Response (IT-Rettung)</i>	1743

Michael Klafft, Maurice Said, Susan Anson, Hayley Watson, Amanda Hughes, Eridy Lukau <i>Challenges in Designing and Distributing a Not For Profit First Aid App Worldwide</i>	1745
Christian Reuter, Gordian Geilen, Robin Gellert <i>Sicherheit vs. Privatsphäre: Zur Akzeptanz von Überwachung in sozialen Medien im Kontext von Terrorkrisen</i>	1759
B. Barth, M.A. Marchitti, J. Mulero Chaves, U. Raape, C. Strobl, M. Borràs, O. Vilalta, H. Ballart, J. Vendrell, N. Prat, M. Mendes, T. Ladoire, G. Gardikis, S. Pantazis, S. Costicoglou, K. Jäckel, W. van Setten <i>An Open Service Platform for Multi-Hazard in Action – the PHAROS Pilot Demonstration</i>	1775
Christian Raffelsberger, Martina Umlauf, Andreas Kercek, Alexander Almer, Thomas Schnabel, Patrick Luley <i>Flexible Kommunikations- und Informationslösungen für eine optimierte Einsatzführung von Interventionskräften</i>	1791
Christian Reuter, Christoph Amelunxen, Matthias Moi <i>Semi-Automatic Alerts and Notifications for Emergency Services based on Cross-Platform Social Media Data –Evaluation of a Prototype</i>	1805
 Softwarearchitekturen und –entwicklung für Smartwatch und Wearable Apps	
Klemens Waldhör, Rainer Lutze <i>Vorwort: Softwarearchitekturen und –entwicklung für Smartwatch und Wearable Apps</i>	1821
Diethelm Bienhaus <i>Smartwatch und Wearables im Gesundheitsbereich: Grundlagen und Anwendungen</i>	1825
Roman Englert, Jörg Muschiol <i>An Extendible Communication as a Service Platform for Wearables and Future-Oriented Devices</i>	1837
Manuel Guttenberger, Klemens Waldhör <i>xHealth: Eine MQTT und REST basierte Architektur zum Zugriff auf Sensordaten smarterer Objekte</i>	1851
Rainer Lutze, Klemens Waldhör <i>The Application Architecture of Smartwatch Apps – Analysis, Principles of Design and Organization</i>	1865

Christian Märtin, Christian Herdin, Sanim Rashid <i>Situationsbewusste, patternbasierte Adaption interaktiver Anwendungen durch Auswertung von Emotions- und Biosignal-Daten</i>	1879
---	------

Seminare, Symposia, Gespräche

Jens-Martin Loebel, Agata Królikowski <i>Symposium des Fachbereichs Informatik und Gesellschaft</i>	1887
--	------

Georg Heidenreich <i>Software as a Medical Device - ein Seminar mit Fallbeispielen zum Mit- diskutieren und Mitarbeiten</i>	1891
--	------

Günter Koch <i>Von der Messbarkeit des Wissens - Versuch und Irrtum auf dem Weg in die Wissensgesellschaft</i>	1897
---	------

Tutorien

Hans-Georg Fill <i>Semantic-based Modeling for Information Systems using the SeMFIS Platform</i>	1901
---	------

Michael Fothe <i>Informatik hat Geschichte!</i>	1909
--	------

Volker Skwarek, Thorsten Kistler, Mark Rawer, Stefan Schauer <i>IoT und Sensornetzwerke: Entwurf und Programmierung von Niedrigst- energiesystemen anhand einer Metaarchitektur</i>	1917
--	------

Markus Endres, Andreas Pfandler <i>Unsere Empfehlung für Sie: Präferenzen und Personalisierung in der Informatik</i>	1927
---	------

Markus Lepper, Baltasar Trancón y Widemann <i>D2d —Kreatives Schreiben von XML-codierten Texten</i>	1935
--	------

Harry M. Sneed <i>Planung von Migrationsprojekten</i>	1941
--	------

Doktoranden-Symposium

Ernst Denert, Mathias Lux, Stefan Wagner <i>Doktoranden-Symposium der Informatik 2016</i>	1949
--	------

Sebastian Bayerl <i>Managing Distributed Statistical Data in the Semantic Web</i>	1951
--	------

Timm Caporale <i>Geschäftsprozessmodellierung mit kontrollierter natürlicher Sprache</i>	1957
Michael Grunwald, Matthias O. Franz <i>Wahrnehmungsorientierte optische Inspektion von texturierten Oberflächen ...</i>	1963
Daniel Hallmann, Ute Schmid, Rüdiger von der Weth <i>Gemeinsame mentale Modelle in der agilen Softwareentwicklung: Ein Ansatz zur Erstellung von Gestaltungsempfehlungen für „gute“ erfahrungsspezifische User Stories</i>	1969
Oliver Knodel <i>Virtualisierung rekonfigurierbarer Hardware zur Steigerung der Rechenleistung und Sicherheit in einer flexiblen Cloud-Architektur</i>	1975
Jonas Lehner <i>Personal BPM – Anwenderorientierte Prozessmodellierung durch persönliches Workflow Management</i>	1981
Franziska Marx <i>Adaptive Informationsvisualisierung: Der Einsatz von Machine-Learning zur Entscheidungsunterstützung</i>	1987
Benjamin Saul <i>Domänenspezifische Vorabanalyse und geschickte Generierung von Optimierungsproblemen</i>	1993
Nicolai Spicher <i>Cardiac activity measurement from video signals of the human skin in ultra-high-field magnetic resonance imaging</i>	1999
Maciej Sumalvico <i>Statistisches Lernen wortbasierter Morphologie</i>	2005
Malte Wattenberg <i>Bestimmung von Erfolgsfaktoren der Unternehmenskommunikation von KMU auf Facebook: Design und Zwischenergebnisse eines Multimethoden-Ansatzes</i>	2011
Matthäus Zloch <i>Methods for Automatic Selection of Database Systems for Optimized Query Performance</i>	2019
Studierendenkonferenz Informatik 2016 (SKILL 2016)	
Michael Becker, Agnes Koschmider, Judith Michael, Thomas Riechert <i>Studierendenkonferenz Informatik 2016 (SKILL 2016)</i>	2027
Tom Standfuß <i>Das Ende der Miniaturisierung</i>	2029

Sigurd Sippel <i>Domain-specific recommendation based on deep understanding of text</i>	2041
Philipp Samfaß, Raphael Schaller <i>A Parallel Non-Hydrostatic Shallow Water Model on Adaptive Triangular Meshes in sam(oa)²</i>	2053
Felix Knispel <i>Überführung von arithmetischen Ausdrücken in ein normalisiertes Polynom mittels Baumtransformation</i>	2065
Rémi Mabon <i>Sigreturn Oriented Programming is a real Threat</i>	2077
Isabel Hagen <i>Nutzerzentrierte Konzeption, Entwicklung und Evaluation: Informatik beGreifen - Schülerinnen und Schüler mit Lego Mindstorms und Raspberry Pi für die Informatik begeistern</i>	2089
Patrick Kuhn, Sven Mischkewitz, Nico Ring, Fabian Windheuser <i>Type Inference on Wikipedia List Pages</i>	2101
Franziska Lehmann <i>new/s/leak - Anforderungsanalyse einer interaktiven Visualisierung für Data-Driven Journalism</i>	2113
Christoph Dobler <i>μ-Force Control - A Device for Controlling Power Wheelchairs for Severely Mobility Impaired Persons</i>	2127
Harry Schilling <i>Twistor – Simulation des Twitterstroms für Evaluationszwecke</i>	2133
Felix Konstantin Maurer <i>A survey on approaches to anonymity in Bitcoin and other cryptocurrencies</i> ...	2145
Dennis Schmidt <i>Ein Eclipse-Plugin zur kontrollierten Schema-Evolution im NoSQL Datenbanksystem MongoDB</i>	2151
Le Hoang Thanh <i>A Learning Algorithm Based on λ-Policy Iteration and Its Application to the Video Game "Tetris Attack"</i>	2157
Paul Jähne <i>Overview of the current state of research on parallelisation of evolutionary algorithms on graphic cards</i>	2163

Carsten Rudolph <i>Recent Developments in Example-based Texture Synthesis for Graphics Rendering</i>	2175
Jan-Patrick Lehr <i>Counting Performance: Hardware Performance Counter and Compiler Instrumentation</i>	2187
Christoph Klemenjak, Peter Goldsborough <i>Non-Intrusive Load Monitoring: A Review and Outlook</i>	2199
Sebastian Niehaus <i>Ein Ansatz zur Einführung von Complex Event Processing zum workflow- orientierten Software-Monitoring</i>	2211

Hauptvorträge

From model checking to equilibrium checking

Michael Wooldridge¹

Abstract

Equilibrium checking is concerned with establishing whether a given temporal logic formula φ is satisfied in some or all equilibrium computations of a multi-agent system – that is, whether the system will exhibit the behaviour φ under the assumption that agents within the system act rationally in pursuit of their preferences.

After motivating and introducing the framework of equilibrium checking, we present formal models through which rational verification can be studied, and survey the complexity of key decision problems. We give an overview of a prototype software tool for rational verification, and conclude with a discussion and related work.

Michael Wooldridge is a Professor in the Department of Computer Science at the University of Oxford. He has been active in multi-agent systems research since 1989, and has published over three hundred articles in the area. His main interests are in the use of formal methods for reasoning about autonomous agents and multi-agent systems.

Wooldridge is an ACM Fellow, a AAAI Fellow, an ECCAI Fellow, and a member of Academia Europaea. He was the recipient of the ACM Autonomous Agents Research Award in 2006.

¹ University of Oxford, Department of Computer Science, mjw@cs.ox.ac.uk

Artificial Intelligence at the Gates of Dawn?

Thomas Eiter¹

Abstract

In recent years, there have been significant achievements in different areas of Artificial Intelligence (AI) which made it to the headlines. After Deep Blue was the first computer program winning against a ruling world champion of chess in 1997, which was a milestone in the history of AI, the Mars Rover missions started in the early 2000s for which intelligent autonomous planning and control is instrumental. The DARPA challenges in 2004 and 2005 have stimulated intensive research on autonomous vehicles, and due to efforts by global players like Apple and Google in cooperation with car industry, encountering such vehicles on the road appears to be realistic in the near future. On another end, IBM's Watson victory over Jeopardy champions in 2011, and the most recent defeat of top Go players in 2015 and 2016 by Alpha-Go, a computer program by Google's DeepMind company, are further milestones in the history of AI, which are based on novel use of search and deep learning techniques.

Fueled by this and many success stories in other application areas, e.g. in speech recognition and robotics, there is a hype about Artificial Intelligence these days. A lot is written in the press, and concerns about emerging super-intelligence have been uttered by prominent people, in fear of the "technological singularity" that the intellectual power of machines will surpass the one of humans, which they see as a threat to humanity. Less apocalyptic yet still serious are worries about the effects of AI on employment. Whether AI will end jobs has been raised in a number of articles and has been discussed by leading economists and AI researchers at the World Economic Forum.

This talk reviews some of the issues above and emphasizes that despite the recent achievements and advances in the state of the art, there is still a number of issues that, from a cognitive perspective, are not resolved yet, among them understanding of the mind, deeper real world reasoning, natural language understanding, philosophical aspects, and ethical or emotional aspects, to mention a few.

Specifically, we will look at recent results, trends and issues in knowledge representation and reasoning, with a focus on declarative problem solving. Arguably deep and complex reasoning is a key feature of intelligence, but current capabilities, let alone for super-intelligence, are still limited. Nonetheless, advances in computational logic have led to technology such as description logics, constraint programming, SAT/SMT and Answer Set Programming that offers tools for solving important problems in a range of

¹ Technische Universität Wien, Institute of Information Systems, eiter@kr.tuwien.ac.at

applications, where logic-based reasoning is at the core, with more to come.

Thomas Eiter is Professor of Knowledge-Based Systems in the Faculty of Informatics at Technische Universität Wien, Austria, since 1998. He worked in different fields of Computer Science and Artificial Intelligence, but his main area is knowledge representation and reasoning, where he has been active since the 1990s and published extensively.

His current interests are declarative problem solving and computational reasoning methods.

Eiter has been serving on many editorial boards, steering bodies, and program committees (e.g. chairing KI 2001 and KR 2014). He is an EurAI (formerly ECCAI) Fellow (2006) and a Corresponding Member of the Austrian Academy of Sciences (2007).

Security demands and attempts in the new digital world

Hanns-Thomas Kopf¹

Zusammenfassung

Cyber-Kriminalität bzw. Cyber-Attacken auf Personen, Unternehmen, kritische Infrastrukturen bis hin zu Staaten werden zu neuen Waffen unserer Zivilisation. Terror ist eine allgegenwärtige Bedrohung unserer Zeit. Kein Tag vergeht, an dem man nicht in der Berichterstattung darauf stößt und damit konfrontiert ist. Mobile Kommunikationstechnologien und deren flächendeckende Verbreitung führen darüber hinaus zu geänderten Bedrohungen und Herausforderungen für die Sicherheit von Personen, Unternehmen und Infrastrukturen.

Andererseits haben sich jedoch auch die technischen Möglichkeiten, sich gegen diese neue Bedrohungen zu stemmen, mittlerweile mannigfaltig und vielversprechend entwickelt. Die zur Verfügung stehende Technologie - etwa im Umfeld der Communication Intelligence - ist in vielen Bereichen einsetzbar, um Risiken zu minimieren bzw. im Ernstfall Angriffe abzuwehren.

Aber wie können neue Technologien tatsächlich unterstützen? Wie genau tragen sie zur Sicherheit der Zivilbevölkerung bei? Welche Einsatzmöglichkeiten haben einzelne Länder? Wie steht Europa im internationalen Vergleich da? Welchen Platz nimmt die europäische Forschung zum Thema Sicherheit im globalen Vergleich ein? Sind wir Vorreiter oder Nachzügler?

***Hanns-Thomas Kopf** studierte an der Universität Wien, an der Friedrich Alexander Universität in Erlangen und an der Babson University Boston - Wellesley (Massachusetts). Internationales Recht hat er mit dem Master of Law in Innsbruck abgeschlossen. Weitere Studien des internationalen Rechts an der Paneuropa Universität.*

In den neunziger Jahren leitete er 10 Jahre lang Marketing von Siemens Nixdorf Informationssysteme in Österreich. Mitte der neunziger Jahre übernahm er die Verantwortung für das Südosteuropageschäft der damaligen Siemens Nixdorf. In der Siemens Business Services Ges.m.b.H. & Co wurde Hanns-Thomas Kopf im Jahr 2004 in die Geschäftsführung berufen. Als Country Manager für Österreich und die

¹ Generaldirektor Atos IT Solutions and Services GmbH und CEO Atos Central and Eastern Europe, austria.at@atos.net

südosteuropäischen Länder war er bis 2010 als Chief Operation Officer für Siemens IT Solutions and Services CEE verantwortlich.

Die Berufung zum Chief Executive Officer für Siemens IT Solutions and Services CEE erfolgte zum 1. Juni 2010. Hanns-Thomas Kopf wurde im Juli 2011 zum Generaldirektor der Atos Gruppe Österreich bestellt und auch in der CEO-Funktion der Atos IT Solutions and Services in Central Eastern Europe bestätigt. Seit 2011 ist Kopf Mitglied im Executive Committee der Atos SE in Paris. Im Jahr 2013 wurde auch die Verantwortung für die Länder Italien und Schweiz in die österreichische Verantwortung übertragen. Derzeit hat Kopf zwölf Organschaften bzw. Aufsichtsmandate.

Hanns-Thomas Kopf ist neben seiner Managementtätigkeit auch laufend Vortragender an österreichischen Universitäten und hat zwei Publikationen zu den Themen Vergaberecht und der rechtlichen Rahmenbedingungen der Energiemarktliberalisierung veröffentlicht.

Im Jahr 2011 wurde Hanns-Thomas Kopf in den Vorstand der Industriellenvereinigung Wien gewählt. Im Jahr 2014 wurde er mit dem Goldenen Ehrenzeichen der Republik Österreich ausgezeichnet.

Software verstehen, zerstören, schützen mit automatischen Software-Modellen

Andreas Zeller¹

Zusammenfassung

Modelle und Spezifikationen sind die Grundlage jeder vernünftigen Software-Entwicklung. Was aber, wenn man Modelle und Spezifikationen aus bestehenden Programmen ableiten könnte? Unsere aktuellen Arbeiten nehmen ein Programm und einige Beispieleingaben und leiten automatisch *Grammatiken* ab, die *Eingabe- und Ausgabeformat* beschreiben. So kann unser AUTOGRAM-System etwa für die Java-URL-Klasse aus den folgenden Beispiel-URLs

```
http://user:password@www.google.com:80/command?  
foo=bar&lorem=ipsum#fragment  
http://www.guardian.co.uk/sports/worldcup#results  
ftp://bob:12345@ftp.example.com/oss/debian7.iso
```

diese *Eingabegrammatik* ableiten, die recht anschaulich die Struktur der URL erfasst:

```
URL ::= PROTOCOL '://' AUTHORITY PATH ['?' QUERY] ['#' REF]  
AUTHORITY ::= [USERINFO '@'] HOST [':' PORT]  
PROTOCOL ::= 'http' | 'ftp'  
USERINFO ::= /[a-z.]+/ ':' /[a-z.]+/  
HOST ::= /[a-z.]+/  
PORT ::= '80'  
PATH ::= /\[/[a-z0-9.]*/  
QUERY ::= 'foo=bar&lorem=ipsum'  
REF ::= /[a-z.]+/
```

Wir erzeugen diese Grammatiken, indem wir den dynamischen Datenfluss eines jeden Zeichens verfolgen, und Zeichen, die in die gleiche Funktion oder Variable fließen, zu syntaktischen Einheiten zusammenfassen – die wir auch nach dem Funktions- oder Variablennamen benennen können. Da etwa „`http`“ und „`ftp`“ in eine Variable namens „`protocol`“ fließen, können wir direkt die entsprechende Regel ableiten. Dies sorgt für sehr lesbare Grammatiken, wie die obige unbearbeitete Rohausgabe unseres AUTOGRAM-Werkzeugs demonstriert.

Grammatiken wie diese helfen unmittelbar, Datenformate zu verstehen. Sie ermöglichen vollautomatisches massives Sicherheitstesten auf einer Vielzahl von Systemen, indem sie die Produktion syntaktisch korrekter Eingaben unterstützen. Diese können so tief ins Programm vordringen, um dort Fehler zu finden – insbesondere Sicherheitslücken. Da

¹ Universität des Saarlandes, Saarbrücken, Lehrstuhl für Softwaretechnik, zeller@cs.uni-saarland.de

sowohl das Lernen der Grammatik wie auch ihre Anwendung in der Testgenerierung vollautomatisch sind, ergeben sich hier viele neue Möglichkeiten – für Entwickler wie für Angreifer. Auf der anderen Seite ermöglichen erlernte Grammatiken aber auch Schutz vor illegalen Eingaben, indem nicht konforme Ein- oder Ausgaben abgeblockt werden – Techniken, die ich anhand konkreter Beispiele demonstriere.

Besuchen Sie unsere Webseite, um mehr über unsere aktuellen Projekte zu erfahren:

<http://www.st.cs.uni-saarland.de/>

*Andreas Zeller ist seit 2001 Professor für Softwaretechnik an der Universität des Saarlandes in Saarbrücken. Seine **Forschung** beschäftigt sich mit der Analyse großer Software-Systeme und ihre Entwicklungsgeschichte. In 2010 wurde Zeller zum **Fellow der ACM** ernannt für seine Beiträge zur automatischen Fehlersuche und der Analyse von Software-Archiven, für die er auch jeweils mit einem 10-Jahres-Impact Award der ACM SIGSOFT und der ICSE ausgezeichnet wurde.*

Bereits 2012 hatten die Saarbrücker Holler, Herzig und Zeller das LANGFUZZ-System entwickelt, das Grammatik-basiert syntaktisch gültige Javascript-Eingaben für Webbrowser erzeugt. Bis heute hat LANGFUZZ als Teil der Firefox-Entwicklung mehr als 4000 Sicherheitslücken in Firefox aufgedeckt. Die aktuellen Arbeiten sind mit Matthias Hörschele entstanden und werden aus einem ERC Advanced Grant für Arbeiten über Specification Mining und Testerzeugung gefördert.

Forschung ohne Informatik – geht das?

Andreas Reuter¹

Zusammenfassung

Man könnte im klassischen Radio-Eriwan-Stil antworten: „Im Prinzip ja, aber ...“, und dann käme eine ziemlich lange Liste von Ausnahmen. Es ist ja unstreitig, dass von der Informatik entwickelte Werkzeuge in fast allen Wissenschaftsgebieten eingesetzt werden, sei es in den Natur- und Ingenieurwissenschaften, in der Medizin, oder in zunehmendem Maße auch in den Geistes- und Sozialwissenschaften. Nach einigen Studien entfallen in wissenschaftlichen Großprojekten mittlerweile 25 % - 50 % der Gesamtkosten über die Projektlaufzeit auf die Entwicklung der (projektspezifischen) Software-Werkzeuge. Kleinere Projekte sind teilweise so abhängig von Software-Tools, dass eine Erhöhung der Rechenzeitkosten die Projekt-Durchführung verhindern kann.

Aber das sind nur die ökonomischen Aspekte. Aus methodischer Sicht ist festzustellen, dass Software-Werkzeuge „Experimente“ ermöglichen, die in der Realität nicht möglich wären; und dass sie die Analyse von Datenmengen erlauben, die anders nicht verarbeitet werden könnten. In diesen Aufzählungen ist aber immer von „Werkzeugen“ die Rede, wenn es um den Beitrag der Informatik geht.

Heißt das, dass die Informatik „nur“ eine Hilfswissenschaft ist, die anderen Disziplinen ermöglicht, ihre Arbeiten schneller, besser, effektiver zu erledigen? Auch wenn das nicht grundsätzlich schlecht ist, muss man doch fragen, ob es über die Bereitstellung von Werkzeugen hinaus genuin „informatische“ Beiträge zum Methodenvorrat der „Computational Sciences“ gibt. Diese Frage diskutiert der Vortrag aus verschiedenen Perspektiven.

Andreas Reuter ist wissenschaftlicher Leiter des „Heidelberg Laureate Forum“, Vorstand der HITS-Stiftung, Geschäftsführer der EML European Media Laboratory GmbH und Seniorprofessor der Universität Heidelberg. Sein Forschungsschwerpunkt liegt auf Datenbanken, Transaktionssystemen sowie parallelen und verteilten Systemen. Gemeinsam mit Jim Gray veröffentlichte er 1992 das Buch „Transaction Processing: Concepts and Techniques“, ein Standardwerk für Forscher und Entwickler auf der ganzen Welt.

Andreas Reuter promovierte 1981 an der TU Darmstadt und lehrte im Anschluss an der TU Kaiserslautern. Von 1983 bis 1984 war er Gastwissenschaftler am IBM Research

¹ HITS gGmbH, EML und Universität Heidelberg, andreas.reuter@h-its.org

Lab in San Jose. Von 1985 bis 1997 war er C4-Professor an der Universität Stuttgart und Leiter der Forschungsgruppe für die Anwendung paralleler und verteilter Systeme. Er hatte eine Vielzahl akademischer Funktionen inne, u.a. war er Prorektor der Universität Stuttgart, Mit-Gründer, Dekan (ab 1998) und Vizepräsident (ab 2001) der International University in Germany mit Sitz in Bruchsal. Von 2006 bis 2011 hatte Andreas Reuter eine Stiftungsprofessur für „Verlässliche Systeme“ an der TU Kaiserslautern. 2010 gründete er zusammen mit Klaus Tschira das Heidelberger Institut für Theoretische Studien (HITS).

IoT, Industrie 4.0 und Big Data: Herausforderungen für die Informatik und die GI

Peter Liggesmeyer¹

Zusammenfassung

Deutschland hat eine „Digitale Agenda“ und einen „IT-Gipfel“. Wichtige aktuelle informatikbasierte Themen, die dringend mit öffentlicher Förderung bearbeitet werden müssen, sind identifiziert worden – z.B. Industrie 4.0. Das Bundesministerium, das sich in der Vergangenheit um den „traditionellen“ Verkehr gekümmert hat, verantwortet nun auch die Infrastruktur für den „digitalen“ Verkehr.

Dessen ungeachtet ist eine kontinuierliche Zunahme an Cyber-Kriminalität zu beobachten, die „scheibchenweise“ an Dramatik gewinnt. Viele Menschen machen sich Gedanken darüber, wo ihre Daten für welche Zwecke ausgewertet werden und ob daraus Nachteile für die eigene Person entstehen können.

Aus Sicht der Gesellschaft für Informatik ist es natürlich zu begrüßen, dass die Bedeutung der Informatik als Schlüsseltechnologie zur Beherrschung dieser Probleme enorm ansteigt, und dies nun auch in der Bevölkerung und der Politik wahrgenommen wird. Aber es gibt eben auch viele Fragestellungen im Zusammenhang mit neuen Entwicklungen, zu denen zur Zeit noch keine zufriedenstellenden Antworten existieren. An dieser Stelle sieht sich die GI in der Pflicht, die Chancen und Risiken konstruktiv und kritisch, fachkompetent und neutral zu kommentieren und Empfehlungen auszusprechen. Wir gestalten unsere Zukunft aktiv mit.

Der Vortrag analysiert wichtige Themen und hinterfragt insbesondere Standpunkte, die in der öffentlichen Diskussion vielleicht manchmal ein wenig zu oberflächlich behandelt wurden.

Peter Liggesmeyer ist seit 2004 Professor für Software Engineering am Fachbereich Informatik der Technischen Universität Kaiserslautern und seit 2015 geschäftsführender Leiter des Fraunhofer-Instituts für Experimentelles Software Engineering IESE in Kaiserslautern.

Seine Forschungsinteressen sind Sicherheits- und Zuverlässigkeitsanalysetechniken für Cyber-Physical Systems und Verfahren zur umfassenden Sicherheitsanalyse (Security

¹ Technische Universität Kaiserslautern, Arbeitsgruppe Software Engineering: Dependability; Fraunhofer Institut für Experimentelles Software Engineering, peter.liggesmeyer@iese.fraunhofer.de

und Safety) für Smart Ecosystems, insbesondere in den Anwendungsgebieten Nutzfahrzeugtechnologie, Industrie 4.0 und „Smart Rural Areas“. Er ist an mehreren Patenten beteiligt.

Seit 2014 ist Peter Liggesmeyer Präsident der GI, zuvor war er u.a. Vizepräsident und langjähriger Sprecher der Fachgruppe „Softwaretechnik“ sowie des Fachbereichs “Softwaretechnologie”.

Prof. Liggesmeyer ist Autor zahlreicher Fachartikel und verbreiteter Fachbücher, insbesondere des Standardwerks „Software-Qualität“ (2002, 2. Aufl. 2009). Darüber hinaus wirkt er regelmäßig in nationalen und internationalen Programmkomitees mit. Er war bzw. ist Mitherausgeber mehrerer Fachzeitschriften und Mitglied einer Vielzahl von Beiräten und Steering Committees.

Industrie 4.0 und zukünftige Arbeitswelten: Praxis und Vision von Infineon Technologies Austria

Sabine Herlitschka¹

Zusammenfassung

Der Schlüsselfaktor für wettbewerbsfähige Produktion in Europa ist das intelligente Nutzen des Wissens und daraus rasch zu lernen.

Industrie 4.0 führt die reale und virtuelle Welt zusammen und ermöglicht bessere, schnellere und effizientere Produktion bei größerer Flexibilität. Dies wird insbesondere die Arbeitsplätze verändern und neue Herausforderungen an Teams stellen.

Infineon Technologies Austria erarbeitet proaktiv Lösungen für die Arbeitswelt der Zukunft.

Sabine Herlitschka is Chief Executive Officer and Chief Technology Officer of Infineon Technologies Austria. Her professional career includes industrial biotechnology research, international cooperation and financing in research, technology and innovation, Internships at leading organizations in the USA, Fulbright Scholar at George Washington University and Johns Hopkins University/School of Advanced International Studies, as well as founding Vice-Rector for Research Management and International Cooperation at the Medical University of Graz/Austria.

Before joining Infineon Technologies Austria, Herlitschka was Director of the Division European and International Programmes in the Austrian Research Promotion Agency with responsibility for implementation of European & international Research and Technology Programmes, particularly the 7th EU Framework Programme in Austria.

Herlitschka holds a Ph.D. in Food- and Biotechnology with Postdoc specialization in molecular biology and genetic engineering and an Master of Business Administration (MBA) in General Management.

¹ Vorstandsvorsitzende & CTO, Infineon Technologies Austria

Towards Industrial Machine Intelligence

Michael May¹

Abstract

The next decade will see a deep transformation of industrial applications by big data analytics, machine learning and the internet of things. Industrial applications have a number of unique features, setting them apart from other domains. Central for many industrial applications in the internet of things is time series data generated by often hundreds or thousands of sensors at a high rate, e.g. by a turbine or a smart grid. In a first wave of applications this data is centrally collected and analyzed in Map-Reduce or streaming systems for condition monitoring, root cause analysis, or predictive maintenance. The next step is to shift from centralized analysis to distributed in-field or in situ analytics, e.g., in smart cities or smart grids. The final step will be a distributed, partially autonomous decision making and learning in massively distributed environments.

In this talk, I give an overview on Siemens' journey through this transformation, highlight early successes, products and prototypes and point out future challenges on the way towards machine intelligence. I also discuss architectural challenges for such systems from a Big Data point of view.

Michael May is Head of the Technology Field Business Analytics & Monitoring at Siemens Corporate Technology, Munich, and responsible for eleven research groups in Europe, US, and Asia. Michael is driving research at Siemens in data analytics, machine learning and big data architectures. In the last two years he was responsible for creating the SAnalytics platform for Big Data applications across Siemens' business.

Before joining Siemens in 2013, Michael was Head of the Knowledge Discovery Department at the Fraunhofer Institute for Intelligent Analysis and Information Systems in Bonn, Germany. In cooperation with industry he developed Big Data Analytics applications in sectors ranging from telecommunication, automotive, and retail to finance and advertising.

Between 2002 and 2009 Michael coordinated two Europe-wide Data Mining Research Networks (KDNet, KDubiQ). He was local chair of ICML 2005, ILP 2005 and program chair of the ECML/PKDD Industrial Track 2015. Michael did his PhD on machine

¹ Leiter Business Analytics & Monitoring, Siemens Corporate Technology, München,
may.michael@siemens.com

discovery of causal relationships at the Graduate Programme for Cognitive Science at the University of Hamburg

Fit für den härtesten Job der Marktwirtschaft - Wie unsere Gründer spitze werden könnten

Friedrich Georg Hoepfner¹

Zusammenfassung

Der Zug ist abgefahren und Europa hat die erste Runde verschlafen. Die Firmen, die das Internet beherrschen, sind nicht von hier. Sind wir wirklich zu dumm, oder haben wir eine Chance in der zweiten Runde? Wie könnten Gründer europäischer Hi-Tech-Firmen ihre Chancen verbessern, zur Spitze aufzusteigen?

Berücksichtigt man die folgenden Leitlinien, so steigen die Chancen einer erfolgreichen Unternehmensgründung. Als Erfolgsfaktoren können unter anderem genannt werden:

- Erfolg fängt im Kopf ab
- Von der Geschäftsidee zum Geschäftsmodell
- Teamwork einüben
- Klein anfangen und dennoch GROSS Denken
- Verkaufen ist nicht alles, aber ohne Verkaufen ist alles nichts
- Sinnvolle Gründerförderung an den Hochschulen

Erfolg fängt im Kopf an

Ausbildung, Einstellungen und Lifestyle sind die Stichworte für diesen Abschnitt. Unternehmer ist ein selbstbestimmter Beruf, der fachübergreifende Kenntnisse erfordert. Leider arbeiten wir an vielen Universitäten dagegen:

- Die Verschulung der Studiengänge,
- das Einpauken von immer mehr Fachwissen,
- das Heranzüchten von Experten auf immer engeren Fachgebieten

widersprechen dem prinzipiell.

Wenn wir mehr erfolgreiche Unternehmer wollen, müssen wir auch in Naturwissenschaft und Technik Individualität und Kreativität zulassen!

Die Grundeinstellung erfolgreicher Unternehmer ist: Ich kann etwas bewirken, ich nehme mein Schicksal selbst in die Hand und ich investiere in die Zukunft. Dem widersprechen

¹ Geschäftsführer Hoepfner Bräu, Karlsruhe, braeu@hoepfner.de

die Idee des Versorgungs-Staats, der Lebensversicherung, des bedingungslosen Grundeinkommens und der Wunsch der Politik, ständig neue gesetzlichen Rahmenbedingungen zu schaffen.

Arbeit und Freizeit mögen für manche ein Gegensatz sein: Erfolgreiche Unternehmer wissen die beiden Elemente zu verbinden, weil ihnen ihre Arbeit Freude bereitet. Wer lieber Fußball spielen will, soll Sportlehrer werden!

Von der Geschäftsidee zum Geschäftsmodell

Der Einfluss von Ideen wird überschätzt. Ideen gibt es wie Sand am Meer, aber erfolgreich werden Sie erst durch das richtige Geschäftsmodell.

Der beste Ausgangspunkt ist nicht die neue Technologie, die nach Anwendungen sucht, sondern ein Problem der Kunden, für dessen Lösung es ein Budget gibt. Und zwar ein Problem, das wirklich wehtut (dem Kunden wohlgemerkt).

Ein Geschäftsmodell wird daraus, in dem man sich entscheidet, mit welchen Argumenten die Lösung vermarktet wird, wer dafür wie viel bezahlen soll, wie die Lösung ausgeführt werden muss, welche Ressourcen dafür eingesetzt werden und so weiter. Der Weg zum Geschäftsmodell ist oft steinig, wenn wir eine Reihe von Annahmen über die Bedürfnisse und Verhaltensweisen unserer Kunden und Konkurrenten testen müssen, um die erfolgreiche Vorgehensweise zu finden. Da hilft die Theorie des „Lean Start-Up“.

Teamwork kann man lernen

90 % unserer High-Tech-Gründer treten heute als Team an. Ob das Team 120 % seiner Leistung erbringt oder die Hälfte der Zeit mit internen Querelen beschäftigt ist, macht oft den entscheidenden Unterschied.

Häufige Probleme sind nicht die mangelnde Motivation, sondern die Unfähigkeit,

- produktiv zusammen zu arbeiten,
- aus Konflikten Chancen zu machen und
- die Qualität des Teamworks durch darauf bezogene Maßnahmen zu erhöhen.

Das muss man lernen und üben!

Allerdings hat sich auch die zu große Homogenität von Teams als Hemmschuh erwiesen, wenn zum Beispiel vier Informatiker zusammen eine Firma gründen, statt noch einen Betriebswirt und eine Juristin dazu zu nehmen.

Klein anfangen und dennoch GROSS Denken

Erfolgreiche Gründer stehen mit beiden Beinen in der Realität und fangen im Normalfall mit kleinen Budgets an. Das ist auch richtig, weil zu hohe Anlaufverluste sich später als Bürde erweisen können.

Aber schon zu einem frühen Zeitpunkt muss man an Dinge denken, die erst in der Wachstumsphase relevant werden: Software-Architektur, Internationalisierung, Team-Building, Vertriebssystem und - leistungsfähige Business Angels.

Verkaufen ist nicht alles, aber ohne Verkaufen ist alles nichts

Es ist auch ein Versäumnis der Universitäten, dass junge Gründer oft gar keine Ausbildung im Verkauf genossen haben. Im Gegenteil, die Tätigkeit des Verkaufes erscheint vielen Akademikern suspekt. Manche hoffen sogar, dass sich ihr Produkt beziehungsweise ihre Leistung aufgrund der großen technischen Überlegenheit später von selbst im Markt durchsetzen wird.

Mit dieser Grundeinstellung ist der Weg zum späteren Scheitern vorbestimmt.

Erfolg hat, wer frühzeitig mit Kunden in Kontakt kommt, und einerseits auf sie hört und andererseits doch auch überzeugend auf Sie einwirken kann. Erfolgreiches Verkaufen ist eine qualifizierte Dienstleistung, die hohe Anforderungen an die kommunikative und oft auch die analytische Fähigkeit des Unternehmers stellt.

Gründer an der Hochschule

In den letzten Jahren hat sich in den europäischen Hochschulen viel getan.

Man hat erkannt, dass wir mehr Gründer brauchen und dass wir erfolgreiche Gründer brauchen. Man hat Lehrstühle für Entrepreneurship geschaffen, Gründerzentren gebaut und Stipendien ausgelobt. Auch wenn erste Erfolge deutlich werden, gibt es an anderer Stelle gute Beispiele dafür, wie man es nicht machen sollte.

*Als Unternehmensgründer und Business-Angel war Dr. **Friedrich Georg Hoepfner** am Aufbau von weit über 50 jungen Firmen in verschiedenen Branchen beteiligt. Sein Interesse gilt der Strategie und sein Augenmerk dem Tempo, das das junge Unternehmen aufnimmt. Hoepfner\Strategie arbeitet dementsprechend auch daran, als Berater den Zielerreichungsgrad von Firmen in Umbruchsituationen zu verbessern und Führungskräfte bei ihren Kernaufgaben zu unterstützen.*

Hoepfner ist Business Angel des Jahres, Lead Investor als Partner der KfW und zukünftig des Europäischen Investitionsfonds, Initiator des CyberForum e.V. und heute Mitglied des

Kuratoriums, Vorstand der Start Angels Zürich und der Hoepfner Stiftung sowie Mitglied verschiedener Unternehmensbeiräte.

Zu seinen Veröffentlichungen gehören das eBuch „Chancen für Unternehmer – Neue Ideen am Markt durchsetzen“ und Aufsätze über Business-Pläne, strategische Due Diligence, Firmenkultur, nachhaltige Geschäftspolitik und strategische Optionen für Hi-Tech Firmen.

Workshops

Software-Entwicklungs-
prozesse und -Produkte
im Licht kultureller
Unterschiede

Vorwort: Software-Entwicklungsprozesse und -Produkte im Licht kultureller Unterschiede

Gerhard Chroust¹, Renate Motschnig²

1 Kulturelle Unterschiede und Software-Engineering

In der Software-Branche tritt der Mensch sowohl als Entwickler als auch als Benutzer von Software-Produkten auf (Abb. 1). Menschen sind weitgehend durch ihren kulturellen Background geprägt. Diese kulturellen Einflüsse fließen in den *Software-Entwicklungsprozess* ein und müssen auch bei der Gestaltung des *Softwareproduktes* berücksichtigt werden müssen, um den kulturellen Erwartungen der Benutzer zu entsprechen.

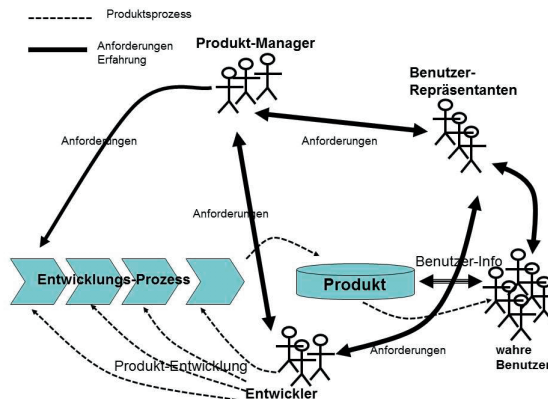


Abb. 1: Stakeholder-Einfluss

Durch die zunehmende globale Vernetzung und Arbeitsteilung wird der Entwicklungsprozess immer mehr in global verteilten, arbeitsteiligen und multinationalen Teams durchgeführt (Outsourcing!). Gute, unmissverständliche und klare Kommunikation und gegenseitiges Verständnis sind nötig [MN09]. Neben explizit übermittelter Information und dem zugehörigen impliziten Wissen sind hier nationale und kulturelle Unterschiede wesentlich: zwangsläufig wird jede Kommunikation durch die 'kulturelle Brille' interpretiert - und oft auch "umgedeutet" oder entgegen der ursprünglichen Absicht gedeutet ("verfälscht").

¹ Johannes Kepler Universität Linz, Donaustrasse 101, 2344 Maria Enzersdorf, gerhard.chroust@jku.at ,

² Universität Wien, Währinger Straße 29, 1090 Wien, renate.motschnig@univie.ac.at

Software-Produkte werden heute weltweit angeboten und erreichen zunehmend internationale, kulturell unterschiedliche Benutzergruppen. Benutzer von Software-Produkten erwarten aber nicht nur, dass ein Software-Produkt 'ihre Sprache spricht', sondern dass sein Verhalten sich nach den in dem speziellen Kulturkreis gültigen Normen und Konventionen richtet. Man wünscht sich von einem Softwareprodukt das Verhalten eines "einfühlsamen, hflischen Butlers" [Ch07]. Die Bereitstellung eines internationale Produktes ist damit mehr als nur eine Frage der Sprachübersetzung. Die Mensch-Maschine-Schnittstelle muss die kulturellen Anforderungen und mögliche (Miss-)Interpretationen berücksichtigen.

Erfolg hängt daher immer mehr vom Verständnis von national und geographisch bedingten kulturellen Unterschieden ab. Kulturelle Interaktion erfolgt auf mehreren hierarchischen Ebenen: auf der untersten Stufe geht es um die korrekte Zeichenerkennung und Interpretation und um technische Kommunikations-Standards, auf der höchsten Ebene fließen die komplexen kulturellen Traditionen, soziale Erwartungen, Tabus und Weltanschauung ein. Wesentliche Beiträge zu Fragen der kulturellen Differenz finden sich in den Arbeiten von Hofstede&Hofstede [HHM10], Hampden&Trompenaars [HTT00], Esselink [Es00] und anderen.

2 Ablauf des Workshops

Für dieses Workshop wurden insgesamt 6 Beiträge angenommen. Diese wurden annähernd in eine 'logische' Reihenfolge gebracht: vom Allgemeinen/Abstrakten zum Speziellen/-Konkreten. Im Sinne eines Workshops sind kürzere Diskussionen während den Präsentationen erwünscht, Außerdem wird nach drei und 6 Vorträgen je eine halbstündige Diskussionsrunde eingeschaltet.

Gerhard Chroust : **"Kulturelle Unterschiede und ihr Einfluss auf Software- Entwicklungsprozesse und -Produkte"** *Der Beitrag gibt eine allgemeine Einführung in die Konzepte der kulturellen Differenzen basierend auf [HHM10, HTT00]. Es werden die kulturelle Dimensionen und die Niveaus der kulturellen Differenzierung beschrieben. Konkret werden die Unterschiede für die drei in den anderen Beiträgen unmittelbar angesprochenen Länder (Österreich, Indien, Ukraine) herausgearbeitet und Folgerungen gezogen [Ch16].*

Eckhart Hanser : **"Erweiterung agiler Prozesse durch sozietaäre Vorgehensbausteine"** *Das neuartige agile Meta-Modell MAP (Meta Agile Process Model) wird vorgestellt, welches u.a. eine Evaluationsmethode zur Bestimmung der menschlichen Verhaltensweisen (Typen) in agilen Teams beinhaltet um Projektrollen optimal zu besetzen, und die Projekt- und Produktqualität zu erhöhen. Die zwei Rekursionsebenen von MAP implementieren das neurokybernetischen Viable System Models (VSM) und sind somit eine systemische Beschreibung eines lebensfähigen ('viable') Systems mit organisatorischer Intelligenz zur Weiterentwicklung des Teams. Die Praxis zeigt, dass MAP den Teamerfolg vorhersagen kann. [Ha16]*

Christina Böhm and Renate Motschnig : **"Ein Diversity Framework für Informatik-nahe Projekte"** *Soziale Kompetenzen und die Fähigkeit, effektiv in interdisziplinären*

Projektteams zu arbeiten, scheinen in der Informatik-Ausbildung wie auch in der praktischen Anwendung unterrepräsentiert. Das "Diversity Framework" bündelt Methoden, Techniken und Prozesse, um implizite, soziale Faktoren und Verhaltensweisen explizit zu machen und eine systematische und praxisnahe Integration solcher Aspekte zu unterstützen.

Der Beitrag basiert auf der eben erschienenen Dissertation der Erst-Autorin [Bo16] und geht im Prinzip von bereits bestehenden Teams aus. [BM16]

Walter Renner und Shiva Khalili : **"Independent vs. interdependent identities: Eine theoretische Basis zu kulturellen Unterschieden in der Entwicklung und Nutzung von Software"** *Wenn benutzer-orientierte Softwareprodukte die Rolle eines "einfühlsamen, höflichen Butlers" einnehmen sollen [Ch07], müssen sie den kulturellen Erwartungen und Erfordernissen der Benutzer angepasst sein. Wenn Benutzer und "Butler" zufriedenstellend interagieren sollen, stellt sich somit die Frage der kulturellen Identität auf beiden Seiten. Der Beitrag stellt die kulturpsychologische Basis zu diesen Postulaten und weiterführenden Entwicklungen bereit. In einer Weiterführung des Konzeptes individualistischer vs. kollektivistischer Gesellschaften präsentierte Markus und Kitayama die Unterscheidung einer "unabhängigen" (independent) Identität in den USA und der EU gegenüber einer "wechselseitig abhängigen" (interdependent) Identität in Asien. Diese Überlegungen werden am Beispiel Iran hinsichtlich des politischen Kontextes und bezüglich der religiös-ideologischen, soziokulturellen und genderbezogenen Konsequenzen und Implikationen veranschaulicht [RK16].*

Walter Khom, Mykola Tkachuk, Volodymyr Sokol, Oleksiy Kosmachov und Vladyslav Sokol : **"Communication, Management and Teambuilding Issues in Austrian-Ukrainian Outsourcing Project: 10 Years of Experience and Future Challenges"** *Outsourcing von Softwaredienstleistungen gehört heute zu den wichtigsten und stetig wachsenden Wirtschaftsfaktoren. Outsourcing bietet nicht nur Kosteneinsparungspotenziale, sondern vielmehr die Möglichkeit des ad-hoc Zugriffs auf Personalressourcen. Jedoch erfordert Outsourcing eine geordnete Organisation und ein effizientes Management, um auch die gewünschten positiven Ergebnisse zu erhalten und Erwartungen zu erfüllen. Eine der häufigsten Problemfelder bei Software-Outsourcing sind jedoch kulturelle Unterschiede und verschiedene Denkweisen zwischen Auftraggeber und Outsourcing-Dienstleister. Das Bewusstsein darüber gehört zu den kritischen Erfolgsfaktoren einer effizienten Kommunikation und eines effizienten Managements, gleichermaßen wie die passende Organisations- und Teamstruktur auf beiden Seiten. Dieser Artikel stellt einen Erfahrungsbericht aus 10 Jahren Software-Outsourcing zwischen Unternehmen in Österreich und der Ukraine dar und behandelt insbesondere aktuelle und künftige Herausforderungen einer derartigen Kooperation [Kh16] (Beitrag in Englisch).*

Shekhovtsov, V.A. and Mayr, H.C. : **"View Harmonization in Software Processes: from the Idea to QuASE"** *Qualitätssicherung in Software Prozessen setzt voraus, dass die Beteiligten ein gleiches Verständnis der zu erfüllenden Anforderungen haben. Ein zentraler Aspekt des "Quality Aware Software Engineering" (QuASE) ist daher die Harmonisierung von Stakeholder-Sichten bei der Systementwicklung. Der*

Vortrag zeigt auf, wie von der ersten Idee zur Unterstützung des wechselseitigen Verstehens von Qualitätsanforderungen das System QuASE entstand: dieses unterstützt nicht nur ein umfassendes "Understandability Management" sondern integriert auch Wissen aus unterschiedlichen Ticketing-Systemen, welches damit zugänglich und wiederverwertbar gemacht wird. [SM16] (Beitrag in Englisch).

Literaturverzeichnis

- [BM16] Boehm, Ch.; Motschnig, R.: Ein Diversity Framework für Informatik-nahe Projekte. In (Mayr, H.C.; Pinzger, M., Hrsg.): INFORMATIK 2016, Lecture Notes in Informatics (LNI). Gesellschaft für Informatik, Bonn, S. in this volume, 2016.
- [Bo16] Boehm, C.: A Framework for Managing Diversity in ICT Projects. Processes and Techniques for Explicating Soft Facts and Dealing with Behavioral Differences. Dissertation, Universität Wien, Wien, 2016, 2016.
- [Ch07] Chroust, G.: Software like a courteous butler - Issues of Localization under Cultural Diversity. In: ISSS 2007. The 51th Annual meeting and Conference for the System Sciences. Tokyo, Japan, August 5-10, 2007. Curran Associates, Inc. (Mar 2008), 2007.
- [Ch16] Chroust, G.: Kulturelle Unterschiede und ihr Einfluss auf Software-Entwicklungsprozesse und -Produkte. In (Mayr, H.C.; Pinzger, M., Hrsg.): INFORMATIK 2016, Lecture Notes in Informatics (LNI). Gesellschaft für Informatik, Bonn 2016, S. in this volume, 2016.
- [Es00] Esselink, B.: A Practical Guide to Localization. John Benjamins Publishing Comp., Amsterdam / Philadelphia 2000, 2000.
- [Ha16] Hanser, E.: Erweiterung agiler Prozesse durch sozietäre Vorgehensbausteine. In (Mayr, H.C.; Pinzger, M., Hrsg.): INFORMATIK 2016, Lecture Notes in Informatics (LNI). Gesellschaft für Informatik, Bonn, S. in this volume, 2016.
- [HHM10] Hofstede, G.; Hofstede, G. J.; Minkov, M.: Cultures and Organizations - Software of the Mind, 3rd edition. McGraw-Hill, NY 2010, 2010.
- [HTT00] Hampden-Turner, C.; Trompenaars, F.: Building Cross-Cultural Competence - How to Create Wealth from Conflicting Values. Yale Univ. Press 2000, 2000.
- [Kh16] Khom, W.; Tkachuk, M.; Sokol, V.; Kosmachov, O.; Sokol, V.: Communication, Management and Teambuilding Issues in Austrian-Ukrainian Outsourcing Project: 10 Years of Experience and Future Challenges. In (Mayr, H.C.; Pinzger, M., Hrsg.): INFORMATIK 2016, Lecture Notes in Informatics (LNI). Gesellschaft für Informatik, Bonn, S. in this volume, 2016.
- [MN09] Motschnig, R.; Nykl, L.: Konstruktive Kommunikation Sich und andere verstehen durch personenzentrierte Interaktion. Klett-Cotta, Stuttgart 2009, 2009.
- [RK16] Renner, W.; Khalili, S.: Independent vs. interdependent identities: kulturelle Unterschiede in der Entwicklung und Nutzung von Software. In (Mayr, H.C.; Pinzger, M., Hrsg.): INFORMATIK 2016, Lecture Notes in Informatics (LNI). Gesellschaft für Informatik, Bonn, S. in this volume, 2016.
- [SM16] Shekhovtsov, V.A.; Mayr, H.C.: View Harmonization in Software Processes: from the Idea to QuASE. In (Mayr, H.C.; Pinzger, M., Hrsg.): INFORMATIK 2016, Lecture Notes in Informatics (LNI). Gesellschaft für Informatik, Bonn, S. in this volume, 2016.

Kulturelle Unterschiede und ihr Einfluss auf Software-Entwicklungsprozesse und -Produkte

Gerhard Chroust¹

Abstract: In der heutigen globalen, arbeitsteiligen Welt - und ganz besonders in der Software-Industrie - ist gute, unmissverständliche und klare Kommunikation und gegenseitiges Verständnis ein wesentlicher Erfolgsfaktor der Kooperation. Neben explizit übermittelten Information und implizitem Wissen spielen nationale kulturelle Unterschiede eine große Rolle: alles wird durch die 'kulturelle Brille' interpretiert - und oft auch verfälscht. Das gilt besonders für die Zusammenarbeit in multinationalen Entwicklungsteams wie sie u.a. durch Outsourcing entstehen. Durch die Globalisierung des Handels (e-commerce) werden Software-Produkte weltweit angeboten. Deshalb ist auch die Mensch-Maschine-Schnittstelle den kulturellen Anforderungen und (Miss-)Interpretationen unterworfen und muss den kulturellen Unterschieden und Erwartungen angepasst werden; man spricht von 'Lokalisierung'.

In diesem Beitrag werden die hierarchisch geschichteten Niveaus der kulturellen Unterschiede klassifiziert. Es wird die Theorie der kulturellen Unterschiede von Geert und Gert Hofstede (Vater und Sohn) und nach Hampden-Trompenaars qualitativ präsentiert, die auch weltweit die Unterschiede numerisch ermittelt haben. Beispielhaft werden für drei Regionen (Österreich, Indien, und Ukraine) detailliert die entsprechenden Werte dargestellt und daraus Schlussfolgerungen für das Arbeiten in internationalen Teams gezogen. Einige Konsequenzen werden auch für die 'Produktseite', d.h. der Gestaltung der Benutzerschnittstellen gezogen.

Keywords: Globalisierung, kulturell Unterschiede, Lokalisierung, Hofstede, Hampden-Turner, Trompenaars,

1 Werkzeug-Entwicklungsprozesse

1.1 Werkzeugentwicklung in der Vorzeit

Werkzeugentwicklung und -verwendung ist einer der Schlüssel zur menschlichen Zivilisation. Ursprünglich waren Erfindung, Herstellung, Verwendung und Wartung in einer Hand (und einem Kopf!), siehe Abb. 1(a). Eine explizite Kommunikation und Weitergabe von Wissen zwischen diesen einzelnen menschlichen Funktionen war nicht notwendig. Mit der zunehmenden Zivilisation und besonders mit der Verstädterung (vgl. Mesopotamien [Be00]) entwickelte sich eine arbeitsteilige Gesellschaft, in der diese Funktionen verschiedenen Personen zufielen, Abb. 1(b). Diese Arbeitsteiligkeit war ein erster Schritt zur 'Massenfertigung' (z.B. Wachsmodele für den Guss von Götterfiguren). Dies bedeutete u.a. dass ein effektiver und effizienter Produktionsprozess an Bedeutung gewann. Ebenso musste auch auf Standardisierung Rücksicht genommen werden, besonders auffällig bei den Waffen für Heere.

¹ Johannes Kepler Universität Linz, Donaustrasse 101, 2344 Maria Enzersdorf, gerhard.chroust@jku.at

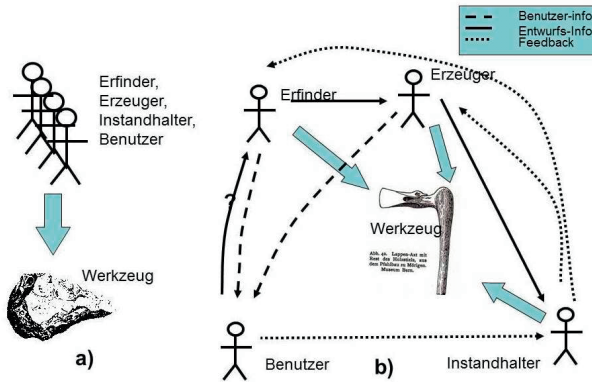


Abb. 1: Individuelle und arbeitsteilige Werkzeug-Entwicklung

1.2 Kommunikation und Kooperation im Erzeugungsprozess

Die allmähliche mit der Verstärkerung einsetzende Arbeitsteilung [Be00] erfordert auch eine entsprechende Kooperation zwischen den einzelnen Beteiligten, und damit Kommunikation, um Wissen und Erfahrung zu übertragen. Abb. 1(b) und Abb. 2 zeigen einige der wesentlichen Kommunikationswege. Mehrere wesentliche Probleme tauchen bei dieser arbeitsteiligen Produktion auf, wobei diese Probleme bei der Software-Entwicklung ebenso deutlich sichtbar werden

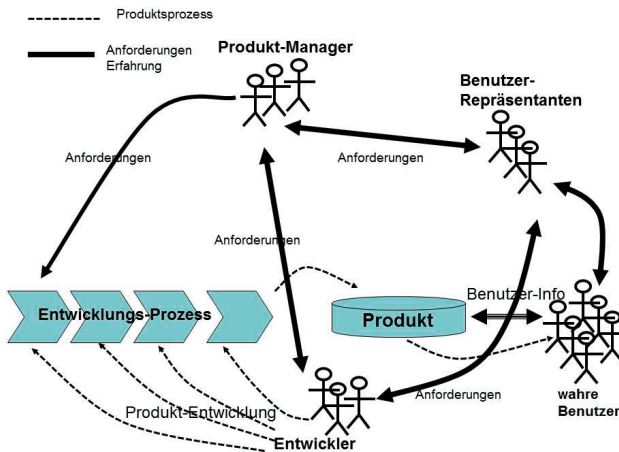


Abb. 2: Stakeholder-Einfluss

- Die Externalisierung des Informationsflusses [NT95] erfordert die Umwandlung von internalisiertem Wissen ("tacit knowledge"), das oft nur intuitiv und operationell vorhanden war, in explizites, vermittelbares Wissen ("explicit knowledge"), das von einer Person zu einer anderen übermittelt werden kann. Zu bemerken ist, dass

bei höherwertigen Produktionsprozessen auch noch eine Zweiteilung des expliziten Wissens zu unterscheiden ist: Übermittlung durch Beobachtung und Nachahmung (Meister - Schüler) und die Abstraktion des Wissens in Form von abstrakt vermittelbaren Wissen, das unabhängig von einer wissenden Person weitergegeben werden kann, z.B. in Form von Anleitungen, Lehrbüchern etc.. Letztendlich findet das Wissen auch Eingang in akademische Studien.

- Die Arbeitsteilung implizierte, dass viele der Beteiligten *mehr* als für den eigenen Bedarf produzierten. Somit musste die Verwendung der geschaffenen Werkzeuge auch den verschiedenen potentiellen Benutzern passend mitgeteilt werden.
- Bei Artefakten, deren Verwendung nicht selbstverständlich war (oder die bisher noch nicht existierten) musste auch das Bedürfnis und damit der Markt erst geschaffen werden (z.B. für Prunkgefäße).
- Bei der Kommunikation wirken meist unverzichtbar mehrere Informationskanäle (Sprache, Intonation, Körpersprache, Gesten, ...) zusammen: *"man kann nicht nicht kommunizieren"* [Wa00].
- Die Trennung der einzelnen Entwicklungsaufgaben erforderte auch expliziten *Feedback* zwischen den beteiligten Akteuren (Abb. 2). Auch bei vielen heutigen Produktionsprozessen sind nicht alle Kommunikationswege in ausreichendem Maße vorhanden, besonders Feedback von der Instandhaltung (Wartung) zurück zur Produktion ist oft mangelhaft.
- Wegen der Vielzahl der Benutzer mit ihren verschiedene Wünsche können oft nicht der eigentliche Benutzer sondern nur dessen Repräsentanten Einfluss auf das Produkt nehmen. Ebenso werden verschiedene Entwicklungsingenieur nicht unbedingt einer Meinung sein (Abb. 2).

2 Menschliche Kommunikation und kulturelle Unterschiede

2.1 Offensichtliche kulturelle Differenzen

Aus unserer Erfahrung aus Fernreisen, Fernsehen und Filmen kennen wir auffällige kulturelle Unterschiede wie Aussehen von Personen, Kleidung, Speisen, Speisevorschriften, Getränke, Sprache, Schrift, Schreibung von Datum, Zeit, Geld, und so weiter. Ein auffälliger (und auch Verwirrung stiftender) Unterschied ist, z.B., wie man auf den Fingern zählt.

2.2 Notwendigkeit der kulturellen Anpassung

Generell kann man feststellen: *Software systems do not exist in isolation, they are used in social and organizational contexts. Experience and many studies show that the major cause of most software failures is the people rather than technical issues [CKN88].*

... *It is the people and culture of the organization that determines how any system is used. ... human, social and organizational considerations affect software processes ...* [KB99].
Zwei Gründe stehen dabei im Vordergrund:

- Outsourcing von Software-Produktion, besonders in asiatische Länder, macht die internationale Kooperation und Kommunikation über nationale/kulturelle Grenzen notwendig. Probleme können sowohl bei der Gestaltung der Produkte als auch bei der Abwicklung des Entwicklungsprozesses auftreten [KH05b, KH05a, KSW04, WDH07].
- Jede Aussage wird auch im Rahmen eines kulturellen Kontext gemacht. Man unterscheidet dabei *high-context cultures* und *low-context cultures* [La06] [Sc01]. Stark kontext-abhängige Kulturen vermitteln viele Informationen nur über den Kontext des Gesagten (wer sagte was, wann, mit welchem Gleichnis, usw.) während Kulturen mit geringer Kontext-Abhängigkeit die meisten Informationen explizit verbal vermitteln [Ha76, Sc01].

Die kulturelle Anpassung erfolgt im Prinzip auf zwei eng miteinander verflochtenen Gebieten: Kulturelle Anpassung im Software-Entwicklungsteam und kulturelle Anpassung der Schnittstelle zwischen Benutzer und Computer ("Software localization", siehe Abschn. 6). Die Bedeutung der kulturellen Anpassung steigt notwendigerweise immer mehr mit dem technologischem Fortschritt, da die Computersysteme immer mehr mit pseudo-persönliche Schnittstellen (Köpfe oder ein ganze Person) ausgestattet werden und auch die Services mittels Artificial Intelligence näher an menschliches Verhalten rücken. Es ist oft schwer zu unterscheiden, ob am anderen Ende der Kommunikation ein Mensch oder Automat (Avatar) gezeigt wird, mit dem Benutzer des Software-System interagiert (Abb. 3). Kulturelle Inkompatibilitäten fallen dabei sehr unangenehm auf.

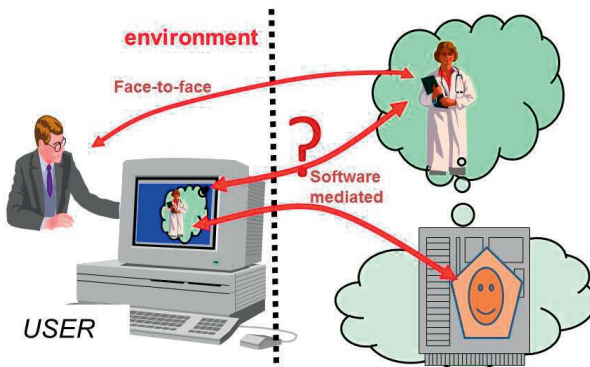


Abb. 3: Mensch oder Avatar?

2.3 Niveaus der kulturellen Differenzen

Unsere Erfahrung lehrt uns auch, dass die Anpassung an lokale/kulturelle Vorgaben auf mehreren Ebenen erfolgt:

Technologische Infrastruktur : Bei der direkten Face-to-Face-Kommunikation bedarf es kaum einer Infrastruktur. Sobald aber technische Hilfsmittel (Telefon, Video, ...) die Basis der Kommunikation sind, bedarf es technischer und organisatorischer Vorbereitungen und Anpassungen, wie z.B. Probleme bei Zeichensätzen (nationale und diakritische Zeichen, Sortierfolge), bei der Behandlung von Zwei-Byte Sprachen (Asien!) [Ad93], der richtigen Leserichtung (links-rechts, rechts-links, usw.) [Es00, HBL02] und alle Vereinbarungen, die als 'Locale' [Es00] zusammengefasst werden, wie Währungssymbole, Datumsschreibweise, ... [IM05].

Grammatikalisches Niveau : Grammatik spielt eine wesentliche Rolle bei einer über das allernotwendigste hinausgehenden Kommunikation und ist wesentlich für das Verständnis, aber auch für die barrierefreie Kommunikation.

Semantisches Niveau : Dieses Niveau befasst sich mit der *Bedeutung* von Text, seiner Interpretation. Dazu gehören Doppelbedeutungen, Verwendung von technischer Sprache versus Umgangssprache, Ausdrucksfähigkeit verschiedener Sprachen, Abkürzungen, Metapher, u.a.

Graphisches und symbolisches Niveau : Die steigende Verwendung von Symbolen und Graphiken muss auch kulturelle Unterschiede berücksichtigen. Das schließt die korrekte Verwendung von Farben, Tiersymbolen, Gesten [Mo06] etc. ein. Da viel davon optisch unbewusst aufgenommen wird, können unangenehme Differenzen zwischen gesprochenen Wort und Darstellung Unbehagen hervorrufen.

Geschäftskonventionen und Praktiken : Praktiken, Verhaltensweisen und Konventionen spiegeln deutlich die darunterliegende Kultur und Wertvorstellung wider, sie sind oft über Jahrtausende gewachsen und sind in einem gewissen Sinn optimiert. Punktuelle Änderungen und Anpassungen können ein delikates Gleichgewicht zerstören und äußerst negative Folgen haben, die in Frustration, Reibungsverlust und Entfremdung resultieren können [KSW04]. Es zählen dazu: Führungsverhalten, Organisationsstruktur, Organisation von Datensammlungen und Web-Sites, Termine und Pünktlichkeit, Leistungsbeurteilung, Überzeit-Arbeit, etc.

Gesellschaftliches und kommunikatives Niveau : Die menschliche Tendenz, menschliche Züge und Eigenschaften auch komplexen Computer-Schnittstellen zuzugestehen prägt auch das Verhalten gegenüber diesen Schnittstellen: man erwartet gute 'Umgangsformen', Höflichkeit, Verständnis und Hilfsbereitschaft und die Einhaltung gesellschaftlicher Konventionen [Mi04].

Es gehören dazu: Kommunikationsstil (Kontextabhängigkeit, Vergleiche und Analogien), Arten des Humors, Begrüßung und Anrede, Zugehörigkeit zu sozialen Klassen, gesellschaftliche Stellung und Alter, geschlechtsspezifische Stellung, Wertschätzung von 'anderen' Gruppen, Ausdrücken von Ablehnung [Da02].

Weltanschauliches Niveau : Es ist dies ein Komplex von Religions- und Werte-Vorstellungen, Moralauffassung, Tabus, Recht, Bräuchen, Kunst, und allen anderen Gewohnheiten und Sitten, die der Mensch als Mitglied einer Gesellschaft erworben hat.,

Kulturelle Diskrepanzen zeigen sich auf diesem Niveau oft am stärksten, auch bei den doch unpersönlichen Computer-Schnittstellen können kulturelle Problem besonders unangenehm auffallen.

Betrachtet man die kulturellen Niveaus (Abb. 4) so erkennt man, dass nur die obersten drei Niveaus durch stark emotional gefärbte und tradierte kulturelle Unterschiede geprägt werden. Die anderen Niveaus (vgl. auch [Es00]) enthalten kulturelle Unterschiede, die störend sein können. Diese Unterschiede sind jedoch nicht dramatisch und oft auch ohne allzu viel Emotion umlernbar. Die wesentlichen kulturellen Unterschiede (siehe Abschn. 3) prägen jedoch die Mitglieder der jeweiligen Kultur, sind oft nur unterschwellig erkennbar und beeinflussen emotionale Entscheidungen und Verhalten in starkem Ausmaß. Vieles davon ist Teil der Weltanschauung und der moralischen Werte.

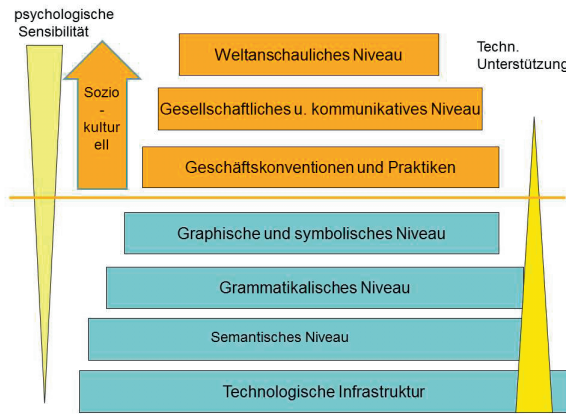


Abb. 4: Niveaus der kulturellen Differenz

3 Identifizierung kultureller Unterschiede

Bei der Identifizierung von kulturellen Unterschieden sind 2 Stufen zu unterscheiden.

Identifizierung grundlegender Dimensionen der kulturellen Ausrichtung : Das Problem dabei ist, welche Charakteristiken identifiziert man, die grundlegend und mehr oder weniger unabhängig voneinander sind.

Messung der Charakteristiken für einzelne Personen oder Personengruppen : Nach Festlegung der einzelnen Charakteristiken versucht man diese Charakteristika einer Personen zu messen, wobei die Absicht ist, darauf auf größere Aggregate (Nationen?) zu schließen. Natürlich besteht hier die Gefahr einer unreflektierten Klischee-Bildung.

In den 70-er Jahren veranlasste die Fa. IBM weltweite, detaillierte Untersuchungen über kulturellere Unterschiede, wobei die Welt in etwa 74 Regionen (meist Nationen) eingeteilt wurde. Für jede dieser Regionen wurden mit Hilfe eines Fragebogen grundlegende Einstellungen zu verschiedenen Fragen erhoben.

Basierend auf diesem weltweiten Datenmaterial und auf weiteren Analysen leiteten sowohl Geert und Gert Hofstede (Vater und Sohn), später zusammen mit Michael Minkov [HHM10], und parallel dazu Charles Hampden-Turner mit Fons Trompenaars [HTT00, Ch06] mehrere grundlegende Dimensionen kultureller Unterschiede ab. Obwohl die Analysen auf weitgehend demselben Datenmaterial beruhen, kommen die beiden Forschungsteams zu zwar ähnlichen, aber doch verschiedenen Dimensionen. Während Hofstede & Hofstede besonders auf die Orthogonalität der Dimensionen Wert legen, legen Hampden & Trompenaars mehr Wert auf die anschauliche Darstellungen und die Auswirkungen von Unterschieden in den verschiedenen Dimensionen ('im besten Fall' und 'im schlechtesten Fall'). Für Details muss auf die angegebene Literatur und auf [Ch07, Ch08] verwiesen werden.

4 Kulturelle Unterschiede nach Hofstede&Hofstede

Hofstede&Hofstede identifizieren in [HHM10] 6 Dimensionen (mit deren Gegenteil), siehe Abb. 5.

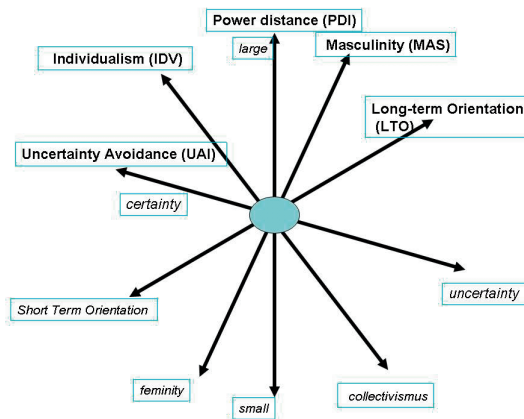


Abb. 5: Kulturelle Unterschiede nach [HHM10]

Power Difference Index (PDI) : Dieses Maß zeigt inwieweit die weniger mächtigen Mitglieder einer Organisation, Institution, oder Gesellschaft akzeptieren und erwarten, dass die Macht (und die damit zusammenhängenden Privilegien) ungleich verteilt sind. Die Ungleichheit wird sowohl von den Führern als auch den Geführten akzeptiert und gutgeheißen.

Individualism Index (IDV) : IDV ist ein Maß für Unabhängigkeit der Einzelnen im Vergleich zur Bindung an eine (kollektive) Gesellschaft, in die der Einzelnen (meist

von Geburt an) stark integriert ist. Diese Gesellschaften sind zusammengeschweißte 'In-Groups', oft eine (um enge Freunde erweiterte) Familie, die permanenten Schutz bietet im Austausch für unreflektierte Loyalität.

Masculinity (MAS) : Dieser Wert bestimmt wie groß die Kluft zwischen den Wertevorstellungen von Männern und Frauen ist. Es zeigt, inwieweit die Zuordnung der 'typisch' männlichen Werte Entschlossenheit und Konkurrenzdenken im Gegensatz zu den 'typisch' weiblichen Werten Bescheidenheit, Fürsorglichkeit und Trost, bei den beiden Geschlechtern unterschiedlich ist.

Uncertainty Avoidance Index (UAI) : UAI misst die Toleranz einer Gesellschaft für Ungewissheit und Mehrdeutigkeit, wie unbehaglich man sich in unstrukturierten Situationen fühlt, die neuartig, unbekannt, überraschend oder verschieden von dem Üblichen sind. Kulturen mit hohem UAI versuchen die Wahrscheinlichkeit derartiger Situationen durch strenge Gesetze, Regeln und Sicherheits-Maßnahmen zu minimieren. Auf der philosophischen und religiösen Ebene drückt sich ein hoher UAI durch den Glauben an eine Absolute Wahrheit aus.

Long-Time Orientation Index (LTO) : Einstellungen, die mit einem hohen LTO gekoppelt sind, sind Sparsamkeit, Beharrlichkeit und Ausdauer. Mit geringem LTO werden Werte wie Respekt für Tradition, Erfüllung soziale Verpflichtungen und das 'Bewahren des eigenen Gesichtes' geschätzt.

Indulgence versus Restraint (IVR) : "Nachgiebigkeit und Beherrschung" beschreibt das Erreichen von Glück durch die Wahrnehmung von Kontrolle über das eigene Leben und die Wichtigkeit von Freizeit und Muße. Die Dimension wurde zuerst durch Minkov formuliert und später von Hofstede als eine der Kulturdimensionen übernommen.

4.1 Kulturelle Unterschiede nach Hampden&Trompenaars

Hampden&Trompenaars kommen zu ähnlichen Überlegungen, verwenden aber teilweise andere Basisdimensionen [HTT00, Ch06]:

universalism-particularism Universalismus hat als Ziel generelle, endgültige Regeln zu finden. Wenn keine Regel passt, wird die bestmögliche Regel angewendet. Partikularismus betrachtet die Ausnahmen. Wenn keine Regel passt, wird ein Fall eher auf Grund seines inhärenten Wertes beurteilt als der Versuch gemacht, ihn in eine existierende Regel zu pressen.

individualism-communitarianism Individualismus beachtet die Rechte des Einzelnen. Man versucht jede Person eigenständig Erfolg oder Misserfolg haben zu lassen und betrachtet Kommunitarismus als einen Versuch die Einzelnen ihrer unveräußerlichen Rechte zu berauben. Kommunitarismus handelt von den Rechten der Gruppe oder Gesellschaft. Er betrachtet Individualismus als egoistisch und kurzsichtig.

specificity-diffusion 'Specificity' konzentriert sich auf die spezifische Rolle einer Person, einer Situation oder eines Dinges im Gegensatz zur ganzheitlichen Schau. Ein 'Diffusionist' würde zum Beispiel in Geschäftsverhandlungen auch die Rolle des Partners als 'Elternteil' im Privatleben berücksichtigen. Kulturen mit hoher Kontext-Abhängigkeit sind als 'diffusionistisch' einzustufen, im Gegensatz zu Kulturen mit niedriger Kontextabhängigkeit, in denen mehr Information explizit und verbal ausgedrückt wird [Ha76, Sc01].

achieved status - ascribed status In Kulturen mit 'Achieved status' erreicht man Bedeutung durch Leistung. Die Annahme ist, dass Einzelpersonen und Organisationen ihren Status täglich verdienen oder verlieren können. Andere Annahmen seien zum Scheitern verurteilt. 'Ascribed status' wird über andere Wege erreicht, wie z.B. Geburt oder Alter. Status wird hier als Recht erworben und nicht durch tägliche Leistung, die durch Glück plus Beurteilung gemessen wird. Ordnung und Sicherheit resultieren aus der Permanenz des Status.

inner direction-outer direction 'Inner-directed' beruht auf Vorstellungen und persönliches Urteil 'in unseren Köpfen'. Denken wird als das mächtigste Werkzeug angesehen, Ideen und intuitives Vorgehen sind der beste Weg. 'Outer-directed' sucht Daten und Informationen in der Umgebung. Es wird angenommen, dass wir in eine ideale Welt eingebettet sind und wir dort unser Informationen, Anweisungen und Entscheidungsgrundlagen finden und holen.

sequential time-synchronous time 'Sequential Time' sieht einzelne Ereignisse als separate Elemente, die eines nach dem anderen zeitlich angeordnet sind. Ordnung beruht auf einer serialisierten Anordnung von Aktionen, die nacheinander ablaufen. 'Synchronous Time' sieht Ereignisse als parallel ablaufend, die miteinander synchronisiert und verkoppelt sind. Ordnung beruht auf einer Koordination von vielfältigen Anstrengungen und Aktionen.

5 Beispiel: Einfluss kultureller Differenzen auf Software-Entwicklung

In diesem Abschnitt skizzieren wir einige der Konsequenzen stark voneinander differierenden Werte in den kulturellen Dimensionen. Zur Anschaulichkeit werden wir uns nur auf die Daten von Hofstede&Hofstede [HHM10] berufen, weil dort das Zahlenmaterial konsistenter dargestellt wird.

Wir betrachten die Auswirkungen sowohl auf den Prozess (hier wird es sich meistens um Meinungsverschiedenheiten und Fehlinterpretationen handeln) und auf das Produkt (hier wird es sich meistens unterschiedliche Ansichten über die Gestaltung der Schnittstelle handeln). Es soll nochmals darauf hingewiesen werden, dass derartige Unterschiede für Outsourcing-Projekte von großer Bedeutung sind [KH05b, KH05a, KSW04].

Man muss aber auch sehr deutlich die Schwächen des von Hofstede verwendeten Zahlenmaterials und dessen Interpretation betonen, sowohl was die Aktualität des Materials betrifft als auch die Analysemethoden [SW14, FK07].

Wir werden nur jene drei Regionen näher beleuchten, die in den anderen Beiträgen des Workshops näher diskutiert werden : Österreich, Indien und Ukraine.

Die Diagramme in Abb. 6 stellen für die ausgewählten geographischen Regionen die Werte für die entsprechenden kulturellen Dimensionen dar. Je weiter die individuellen Werte voneinander verschieden sind, desto größer ist die Gefahr einer Dissonanz. Dabei ist prinzipiell der Wert ('score') zu beachten, aber auch das Gewicht dieser Unterschiede im Vergleich zu den anderen geographischen Regionen, das drückt prinzipiell der Rang aus.

Da die Daten für die Ukraine nur für 2 der Dimensionen vorliegen haben, wir wegen der geschichtlichen und geographische Nähe der Ukraine zu Russland, die Werte für Russland für die fehlenden Werte der Ukraine eingesetzt. Die 2 ermittelten Wertepaare lassen vermuten, dass dies zulässig ist.

Die Tabellen (unten) und die Diagramme in Abb. 6 zeigen für Österreich, Indien und Ukraine (von insgesamt 74 Nationen/Regionen) den Punktwert ('score'), von 0 bis 100, und den relativen Platz ('rank'), von dem höchsten Punktwert *absteigend*.

Score	PDI	IDV	MAS	UAI	LTO	IVR
Austria	11	55	79	70	60	63
India	77	48	56	40	51	26
Ukraine	93	39	36	95	86	14
median	63	39	47	70	45	42

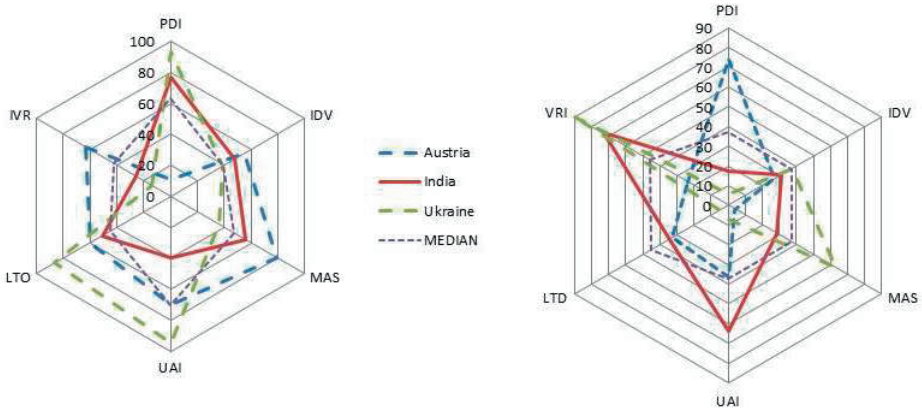
Rank	PDI	IDV	MAS	UAI	LTD	IVR
Austria	74	27	4	37	33	24
India	18	31	29	64	35	73
Ukraine	6	38	63	7	5	90
median	37	37	37	37	45	46

Einige Folgerungen lassen sich daraus ableiten:

PDI - Power Distance Index Die starke Differenz des PDI zwischen Österreich und sowohl Indien als auch Ukraine bedeutet, dass der Meinungs austausch zwischen erfahrenen und jüngeren oder auch Vorgesetzten und Mitarbeitern nicht 'auf Augenhöhe' erfolgen wird [KSW04]. Während im Westen im Prinzip alle gleichgestellt sind, sind in Indien hierarchische Strukturen einzementiert [Kobayashi-05a]. In der werden daher auch ein Produkt mit einer viel rigideren hierarchischen Informations- und Kommunikationsstruktur als Österreicher aufbauen.

IDV Individualism Der im Westen stärker vertretene Individualismus kann das Klima im Team belasten. In der würden Internet-Präsentationen so erzeugen, dass sie vom Benutzer stärker individualisierbar sind (bessere Lokalisierung).

LTO Long Term Orientation Mangelnde LTO (Ukraine) lässt zu befürchten, dass das kurzfristige erreichbare im Vordergrund steht, während hohes LTO auch auf das End-



Punkte (Score) der 3 Nationen

Rang der 3 Nationen

Abb. 6: Kulturelle Unterschiede: Österreich - Indien - Ukraine

ziel des Produktes achtet und der Wartung und Wiederverwendung mehr Augenmerk geschenkt wird.

UAI Uncertainty Avoidance Dies lässt erwarten, dass Beschlüsse stärker dokumentiert und abgesichert werden, was oft auch auf Kosten der Effizienz geht.

IVR Indulgence versus Restraint Niedrige IVR gibt wenig Freiheit für die Entfaltung des Einzelnen, was zum Teil zu Meinungsverschiedenheiten mit Team-Mitgliedern mit hohem IVR (z.B. Österreich) führen kann.

MAS Masculinity Der hohe MAS-Index für Österreich kann bedeuten, dass Inder und besonders Ukrainer den Frauen viel mehr Leitungsaufgaben geben als Österreicher.

6 Kommunikation mit einem Werkzeug ('Lokalisierung')

Anfänglich war die Kommunikation 'mit' einem Werkzeug relativ einseitig. Für Werkzeuge gab es (mehr oder weniger gute) Benutzer-Anleitungen, die (in Form einer Einbahnstraße) von den Entwicklern an die Benutzer weitergegeben werden.

Mit der Computertechnik gewann ein weiterer Kommunikationskanal an Bedeutung: interaktive Kommunikation mit dem Werkzeug. Computer (besonders sogenannte eingebettete Systeme oder software-intensive Systeme) übernehmen immer weitergehende Aufgaben, die früher dem Menschen vorbehalten waren [BWS98]. Zur Erfüllung ihrer Aufgaben ist Informationsaustausch mit Benutzern, in Analogie zu einem Gespräch zwischen Menschen, notwendig.

Das bedeutet, dass die Schnittstellen-Software ähnlichen Anforderungen genügen muss, wie sie für die Kommunikation zwischen Menschen erforderlich ist. Als Folge muss die

Software an die lokalen Gegebenheiten angepasst werden, wir sprechen von "Lokalisierung" [Ch07].

Eine ganze Reihe von Gründen sprechen für eine Lokalisierung von Software-Produkten [Ch08], u.a. technologischer Fortschritt (Geschwindigkeit, Performanz, Displays), logisch anspruchsvollere, benutzerorientierte Anwendungen, allgegenwärtige software-intensive Systeme, vermehrte nicht-verbale Kommunikation zwischen Benutzer und Computer (Graphik, steigende Zahl pseudo-persönliche Schnittstellen (Avatare als Ersatz für Menschen, Abb. 3), internationale Kooperation, globaler Käufermarkt und globales Marketing, steigende Bedeutung des Kontexts der Interaktion, stark gestiegene Erwartungen der Benutzer.

In der Kommunikation zwischen zwei Partnern erwartet man einen gewissen Verhaltenskodex. Man benötigt auch bis zu einem gewissen Grad eine gemeinsame (Basis-)Weltansicht. Der eigentliche Inhalt und die Form der Botschaft ist nur ein Teil der gesamten Kommunikation. In Face-to-face-Situationen spielt die Körpersprache eine große Rolle [Fa71] [La90] [Mo88] [Mo94]. In vielen Einsatzsituationen (besonders im privaten Bereich) erwartet man eine nahtlose Anpassung an die individuelle Kultur und Umgebung [Ch07] ("man kann nicht nicht kommunizieren" [Wa00])

Diese Erwartung überträgt sich auch auf Computerschnittstellen. Benutzer tendieren dazu, einer komplexen, gut gestalteten Computerschnittstelle menschliche Qualitäten zuzuschreiben (Wer hat noch nie seinen Computer beschimpft?) Wir erwarten von einer derartigen Computerschnittstelle "gutes Verhalten ... mit der Sensitivität eines intuitiven, höflichen Butlers" [Mi04]. Je realistischer die Computerschnittstelle gestaltet ist (von simplen Text-Schnittstellen bis zu belebten, interaktiven Darstellungen mit Software Agenten, sogenannte Avatare, desto mehr erwarten die Benutzer die Einhaltung sozialer Konventionen [Mi04].

7 Zusammenfassung

Jeder ist fast überall ein Fremder

In diesem Beitrag haben wir basierend auf [HHM10] eine Klassifizierung von kulturellen Unterschieden nach insgesamt 6 Dimensionen vorgenommen. Weiters haben wir die Niveaus der Kommunikation dargestellt und einige Auswirkungen auf internationale Entwicklungsteams diskutiert.

Aus den Werten für die Dimensionen haben wir für die höherwertigen Kommunikationsaufgaben beispielhaft gewisse Folgerungen für das Verhältnis zwischen Österreich, Indien und Ukraine diskutiert, da Indien und Ukraine klassische Outsourcing-Ziele sind und Österreich eine klassische Outsourcing-Quelle.

Dieser Beitrag soll dazu beitragen, Probleme in internationalen Teams und in Outsourcing-Situationen zu vermeiden. Wir hoffen dadurch einen Beitrag zum besseren Verständnis der Internationalisierung geleistet zu haben

8 Danksagung

Der Autor dankt dem Shaker Verlag, Aachen, für das Recht, Material des Artikels "Kulturelle Unterschiede und ihr Einfluss auf Software-Entwicklungsprozesse und Produkte" [Ch08] wiederverwenden zu dürfen.

Literaturverzeichnis

- [Ad93] Adams, G.: Internationalization and character set standards. *StandardView*, 1(1):31–39, 1993.
- [Be00] Benevolo, L.: *Die Geschichte der Stadt*. Campus Verlag, Frankfurt-New York 2000, 2000.
- [BWS98] Bourges-Waldegg, P.; Scrivener, S.A.R.: Meaning, the central issue in cross-cultural HCI design. *Interaction with Computers* vol. 9 (1998), S. 287–309, 1998.
- [Ch06] *Changing Minds* (ed.): , Trompenaars' and Hampden-Turner's cultural factors. http://changingminds.org/explanations/culture/trompenaars_culture.htm [2005-10-27], 2006.
- [Ch07] Chroust, G.: Software like a courteous butler - Issues of Localization under Cultural Diversity. In: *ISSS 2007. The 51th Annual meeting and Conference for the System Sciences*. Tokyo, Japan, August 5-10, 2007. Curran Associates, Inc. (Mar 2008), 2007.
- [Ch08] Chroust, G.: Kulturelle Unterschiede und ihr Einfluss auf Software-Entwicklungsprozesse und Produkte. In (Höhn, R.; R., Petrasch; O., Linssen, Hrsg.): *Vorgehensmodelle und der Product Life-cycle - Projekt und Betrieb von IT-Lösungen* (15. Workshop der FG WI-VM der GI. Shaker Verlag, Aachen 2008, S. 152–175, 2008.
- [CKN88] Curtis, B.; Krasner, H.; N., Iscoe: A Field Study of the Software Design Process for Large Systems. *Comm ACM*, 31:11:1268–1298, 1988.
- [Da02] Davidson, R.: Cultural Complications of ERP. *Comm. ACM* vol. vol. 45 (2002) no. 7, S. 109–111, 2002.
- [Es00] Esselink, B.: *A Practical Guide to Localization*. John Benjamins Publishing Comp., Amsterdam / Philadelphia 2000, 2000.
- [Fa71] Fast, J.: *Körpersprache*. Rowohlt, Hamburg, 1971.
- [FK07] Froholdt, L.L.; Knudsen, F.: , The Human element in Maritime Accidents and disasters-a matter of communication. <http://www.imla.co/imec/FabienneKnudsen.pdf>, 2007.
- [Ha76] Hall, E.T.: *Beyond Culture*. Doubleday Anchor Books, Garden City, NY, 1976, 1976.
- [HBL02] He, Z.; Bustard, D. W.; Liu, X.: Software internationalisation and localisation: practice and evolution. In: *PPPJ '02: Proceedings of the inaugural conference on the Principles and Practice of programming*, 2002. National University of Ireland, Maynooth, County Kildare, Ireland, S. 89–94, 2002.
- [HHM10] Hofstede, G.; Hofstede, G. J.; Minkov, M.: *Cultures and Organizations - Software of the Mind*, 3rd edition. McGraw-Hill, NY 2010, 2010.

- [HTT00] Hampden-Turner, C.; Trompenaars, F.: Building Cross-Cultural Competence - How to Create Wealth from Conflicting Values. Yale Univ. Press 2000, 2000.
- [IM05] Ishida, R.; Miller, S.K.: , Localization vs. Internationalization. <http://www.w3.org/International/questions/qa-i18n>; [Jan 2005], 2005.
- [KB99] Kunda, D.; Brooks, L.: Human, Social and Organisational influences on Component-Based Software Engineering. In: ICSE 99. IEEE Comp. Society 1999, 1999.
- [KH05a] Kobayashi-Hillary, M.: Outsourcing to India. The Offshore Advantage. Springer Berlin 2005, 2005.
- [KH05b] Kobayashi-Hillary, M.: A Passage to India - Pitfalls that the outsourcing vendor forgot to mention. Quality Assurance, vol. 3 (2005), No. 1, also: <http://www.acmqueue.org/modules.php?name=Content&pa=showpage&pid=283&page=3>, S. 7, 2005.
- [KSW04] Krishna, S.; Sahay, Sundeep; Walsham, Geoff: Managing cross-cultural issues in global software outsourcing. Comm. ACM, 47(4):62–66, 2004.
- [La90] Lauster, P.: Menschenkenntnis: Körpersprache, Mimik und Verhalten. 5. Auflage, ECON Taschenbuch, Düsseldorf 1990, 1990.
- [La06] Lang, Y.: , Marketing Communications, Culture, and Localization. http://www.translate.com/technology/multilingual_standard/marketing_communications.culture.html [Oct. 2006], 2006.
- [Mi04] Miller, C.A, Hrsg. Human-Computer Etiquette: Managing Expectations with Intentional Agents. Comm. ACM vol. 47 (2004), No. 4, 2004.
- [Mo88] Molcho, S.: Körpersprache als Dialog. Mosaik-Verlag 1988, 1988.
- [Mo94] Morris, D.: Bodytalk - A World Guide to Gestures. Jonathan Cape Ltd. London 1994, 1994.
- [Mo06] Molcho, S.: ABC der Körpersprache. H. Hugendubel Verlag, Kreuzlingen/München 2006, 2006.
- [NT95] Nonaka, I.; Takeuchi, H.: The Knowledge Creating Company. Oxford University Press, New York 1995, 1995.
- [Sc01] Schneider, U.: Innovation in der Wechselwirkung der Kulturen. IT'S T.I.M.E., vol. 2001, no. 1, S. 69–72, 2001.
- [SW14] Schmitz, L.; Weber, W.: , Are Hofstede's dimensions valid? A test for measurement invariance of Uncertainty Avoidance. <https://www.google.at/search.....>, 2014.
- [Wa00] Watzlawick, P.: Die Moeglichkeit des Andersseins. Huber-Verlag 2000, 0000.
- [WDH07] Winkler, J.; Dibbern, J.; Heinzl, A.: Der Einfluss kultureller Unterschiede beim IT-Offshoring. Wirtschaftsinformatik, vol. 49 (2007), no. 2, S. 95–103, 2007.

Erweiterung agiler Prozesse durch sozietäre Vorgehensbausteine

Eckhart Hanser¹

Abstract: Im Folgenden wird das neuartige agile Meta-Modell MAP (Meta Agile Process Model) vorgestellt, welches u.a. eine Evaluationsmethode zur Bestimmung der menschlichen Verhaltensweisen (Typen) im agilen Team beinhaltet. MAP ermöglicht es, Projektrollen optimal zu besetzen, was maßgeblich dazu beiträgt, die Projekt- und Produktqualität zu erhöhen. Da MAP zwei Rekursionsebenen des neurokybernetischen Viable System Models (VSM) implementiert, kann gezeigt werden, dass MAP-Teams im Sinne einer systemischen Beschreibung „lebensfähige Systeme“ (viable systems) darstellen, die über entsprechende organisatorische Intelligenz zur Weiterentwicklung des Teams und zur Bewältigung kritischer Projekte verfügen. Die Analyse von Projekten aus der Praxis zeigt, dass MAP den Teamerfolg vorhersagen kann.

Keywords: Meta Agile Process Model (MAP), sozietär, sozial, Teambildung, Teamerfolg, agil, Viable System Model (VSM), Vorgehensmodell, Prozessmodell, Projektmanagement.

1 Verhaltensweisen (Typen) im Team – das Meta Agile Process Model (MAP)

Die Organisationspsychologie weiß seit längerem, dass die richtige „Mischung“ von Denk- und Verhaltensmustern von Teammitgliedern über den Erfolg eines Projekts mitbestimmt. Arbeiten von Baldegger und Gotsman [BG01] oder Belbin [Be04] beschäftigen sich mit diesem Thema. Allerdings sind die für den Erfolg eines Teams notwendigen Verhaltensweisen der Teammitglieder abhängig vom Tätigkeitsfeld des Teams. Ein Managementteam braucht eine andere Mischung der Verhaltensweisen als ein Software-Entwicklungsteam. Diese Lücke schließt der Autor mit seinem 2010 erstmalig veröffentlichten *Meta Agile Process Model* (MAP) [Ha10], welches eine „Landkarte“ der Verhaltensweisen im Software-Entwicklungsteam entwickelt (Abb. 1). Diese „Landkarte“ basiert auf der mehrjährigen Beobachtung erfolgreicher studentischer Teams an der Hochschule des Autors, die in realen Software-Entwicklungsprojekten ihren Erfolg unter Beweis stellen konnten.

¹ Duale Hochschule Baden-Württemberg (DHBW) Lörrach, Kompetenzzentrum für agile IT-Prozesse, Hangstraße 46-50, D-79539 Lörrach, hanser@dhbw-loerrach.de, www.dhbw-loerrach.de/kap.html.

Haupteigenschaften	bereit zur Kommunikation	erfahren	belastbar
visionär		Ideengeber	allg. menschl. Eigenschaften
kreativ			
kostenbewusst		Qualitätsprüfer	
kritisch/hinterfragend			Techn. Problemlöser/Troubleshooter
strukturiert	Kommunikator		
analytisch			
planend		Experte/Fachspezialist	
umsetzungsorientiert			Teamworker

projektarbeitsbezogene menschl. Eigenschaften

Abb. 1: „Landkarte“ (MAP) der verschiedenen menschlichen „Typen“ (Verhaltensweisen) im Entwicklungsprojekt

Abb. 1 ist wie folgt zu „lesen“: In der oberen Zeile von links nach rechts stehen allgemeine Eigenschaften der Teammitglieder, die der Autor in seiner langjährigen Projektstätigkeit als für den Projekterfolg wichtig identifiziert hat. In der Spalte vorne stehen hingegen projektbezogene menschliche Eigenschaften, die weitgehend den von Baldegger identifizierten wichtigen Eigenschaften von Teammitgliedern entsprechen ([Ba04]). Speziell für Entwicklungsteams müssen sie aber um die Eigenschaften *strukturiert* und *umsetzungsorientiert* erweitert werden, während die von Baldegger identifizierte Eigenschaft *erfahren* sinnvollerweise den allgemeinen menschlichen Eigenschaften zugeordnet wird.

Aus den verschiedenen menschlichen Eigenschaften der Teammitglieder leitet MAP „Typen“ von Teammitgliedern ab, die für die spätere Besetzung der Projektrollen bedeutend sind: Kommunikator, Ideengeber, Qualitätsprüfer, Technischer Problemlöser (Troubleshooter), Experte (Fachspezialist) und Teamworker (Entwickler).

Diese Typen ergeben sich aus der Kombination allgemeiner menschlicher Eigenschaften und projektbezogener Eigenschaften der Teammitglieder und überlappen sich zum Teil². Außerdem können sie sich im Laufe der Jahre verändern. Die Diskussion mit Organisationspsychologen [Wa08] ergibt folglich, dass es besser ist, von *Verhaltensweisen* anstatt von „Typen“ zu reden. Damit sind aber Kombinationen grundlegender menschlicher Eigenschaften gemeint und nicht „antrainierte“ Verhaltensmuster³.

Wie in [Ha10] gezeigt wird, entstehen die MAP-Typen bzw. MAP-Verhaltensweisen aus der Analyse eines besonders erfolgreichen Teams des Autors im Jahr 2005 – also aus der Praxis. Alle MAP-Typen sind in diesem Team vorhanden! Es ist aber auch möglich, die

² Diese Überlagerung der Typen (dargestellt als verschiedenfarbige Rechtecke) ist in Abb. 1 gut zu erkennen.

³ Der Einfachheit halber wird im weiteren Text *Typ* wie auch *Verhaltensweise* synonym benutzt.

MAP-Typen aus den 9 Team-Rollen von Belbin herzuleiten [Be07][Be04][Ha15]. Hier muss allerdings beruecksichtigt werden, dass Belbin Managementteams und keine Entwicklungsteams analysiert (Tab. 1).

Team-Rollen nach Belbin	MAP-Typen (-Verhaltensweisen)
Coordinator & Resource investigator	⇒ Kommunikator
Plant („Kreativist“)	⇒ Ideengeber
Completer & Monitor evaluator	⇒ Qualitätsprüfer
Shaper („Gestalter“)	⇒ Technischer Problemlöser (Troubleshooter)
Specialist	⇒ Experte (Spezialist)
Teamworker & Implementer	⇒ Teamworker (Entwickler)

Tab. 1: Abbildung der Team-Rollen nach Belbin auf die MAP-Typen (-Verhaltensweisen)

2 Das MAP-Team als System – Integration in das Viable System Model

Die bisherige Praxis der MAP-Teams mit ihrem sich im Laufe des Projekts anpassenden agilen Prozess zeigt, dass zum Verständnis des neuen Ansatzes nicht der (agile) Prozess im Mittelpunkt stehen sollte, sondern das Team selbst mit all seinen internen und externen Wechselwirkungen. Das MAP-Team ist also ein soziales System, welches über Steuerungs- und Regelungsmechanismen für einen erfolgreichen Projektverlauf verfügt. Diese Mechanismen stehen auf mehreren Ebenen zur Verfügung, sowohl auf der Ebene des Teams als auch auf der Ebene des einzelnen Teammitglieds mit seinen verschiedenen MAP-Verhaltensweisen (also den identifizierten Typen aus Abb. 1).

Mit dem Viable System Model (VSM) von Stafford Beer [Be79][Be81] steht ein geeignetes neurokybernetisches Modell zur Verfügung, um das MAP-Team „systemisch“ zu beschreiben: Während sich die MAP-Teams eines Unternehmens auf Ebene 1 wie andere agile Projektteams in das Unternehmen eingliedern (siehe hierzu [WH12]), kann im Folgenden das MAP-Team im Rahmen des VSM auch auf der Team-Ebene (Ebene 2) und der Ebene 3 der individuellen Verhaltensweisen (MAP-Typen) mit seinen Wechselwirkungen, sowie Steuerungs- und Regelungsmechanismen dargestellt werden.

Das Viable System Model bildet die Basis des „St. Galler Managementkonzepts“ [WH12] (siehe hierzu auch [Sc00]). Es beschreibt ein Unternehmen mit seinen Strukturen und seinen Projektteams als lebensfähiges System und orientiert sich dabei stark am Aufbau des zentralen Nervensystems: Die *Operativen Subsysteme* (im biologischen Organismus wären das z.B. Muskeln oder Organe) interagieren mit ihrer *Umwelt*, überwacht von einem *Metasystem*, welches in der Biologie dem Nervensystem und dem Gehirn entspricht [VS12]. Das Metasystem wird in die *Systeme 2 bis 5* unterteilt, während die operativen Subsysteme als *System 1* bezeichnet werden. Während System 2 die Systeme 1 koordiniert, ist das System 3 das operative Management,

welches durch die Audit- und Monitoringfunktionen des Systems 3* erweitert wird. System 4 interagiert mit der Umwelt und beherbergt das strategische Management, während System 5, das normative Management, für die Wertesystem des Teams und der Teammitglieder zuständig ist.

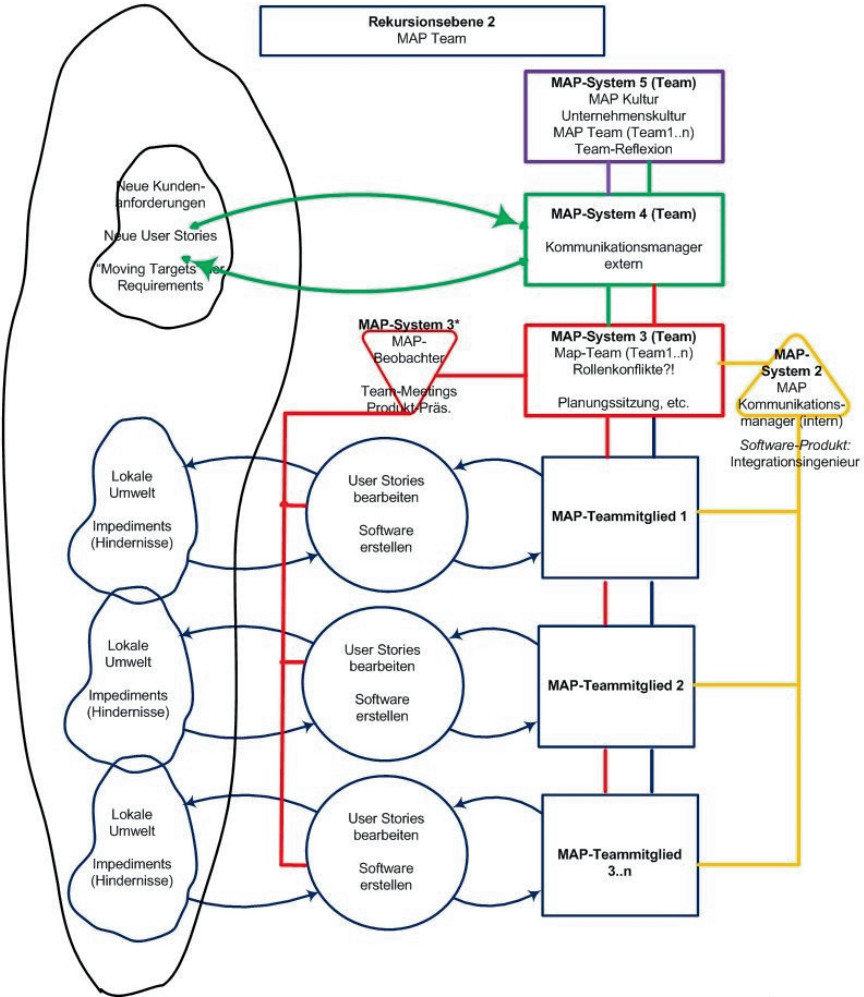


Abb. 2: VSM-Rekursionsebene 2 – MAP Team⁴

Abb. 2 zeigt die Rekursionsebene 2 des Viable System Models [HW12]: Die MAP-Teammitglieder sind die operativen Systeme 1, welche durch das System 2, bestehend

⁴ Abb. 2 wurde gemeinsam mit Jan Wehinger (Volkswagen Konzernforschung, Zukunftsforschung und Trendtransfer) entwickelt [We13].

aus internen Kommunikationsmanagern, koordiniert werden. Bei der Software-Produktion uebernimmt die Koordinationsfunktion des Systems 2 der Integrationsingenieur. System 3, das operative Management, ist das Team selbst, welches sich im Rahmen seiner Planungssitzungen etc. selbst steuert. Unterstuetzt wird es dabei durch den MAP-Beobachter (System 3*), der das Team als Coach betreut. Der oder die externen Kommunikationsmanager halten den Kontakt mit dem Kunden, um neue oder sich aendern Spezifikationen (User Stories) zu identifizieren. Das normative System 5 entspricht der „MAP-Kultur“ des Teams, welche dieses in den Team-Reflexionssitzungen pflegt und weiterentwickelt [Ha15].

Die problemlose Integration von MAP in das Viable System Model⁵ zeigt, dass MAP gut in das VSM „passt“. MAP-Teams bilden also im Sinne des VSM lebensfaehige Systeme, die ueber die „organisatorische Intelligenz“ verfuegen, die notwendig ist, um das Ueberleben auf dem Markt zeitlich und anforderungstechnisch kritischer Projekte zu sichern.

3 MAP in der Praxis

Seit 2004 konnte durch Analyse von studentischen Projekten an der DHBW Loerrach⁶, Teams weiterer Hochschulen, sowie von Unternehmensteams⁷ gezeigt werden, dass sozietaere Erweiterungen von Entwicklungsprojekten gemaeß MAP die Qualitaet von Projekt *und* Produkt erhoehen. Eine Korrelation zwischen dem Projekterfolg und einer guten Teamrollenbesetzung gemaeß MAP ist nachweisbar (siehe Abb. 4). Dabei zeigen die Unternehmensprojekte keine Abweichungen zu den studentischen Projekten! Dies ist auch nicht zu erwarten, da sich die ueblicherweise in agilen Unternehmensprojekten arbeitenden „Young Professionals“ von den Verhaltensweisen und der Ausbildung her meist nicht sehr von den Studierenden in den studentischen Projekten unterscheiden, insbesondere da diese waehrend der Projekte schon in den hoeheren Semestern sind und aufgrund ihres dualen Studiums bereits Praxiserfahrung in Unternehmensprojekten sammeln konnten.

Der groeße Erfolg der untersuchten Projekte beruht auf der konsequenten Umsetzung der von MAP geforderten „Projektzutaten“. Im Vorfeld unterzogen sich die Teammitglieder einer Selbst- und zwei Fremdevaluationen. Daraus wurde die projektspezifische „Landkarte“ (MAP) der Verhaltensweisen im Team ermittelt und die Projektrollen besetzt.

⁵ ohne Anpassung von MAP!

⁶ die neueren davon im Unternehmensumfeld

⁷ mit in der Summe mehr als 300 Teilnehmern in 12 Jahren

7 Teammitglieder (anonymisiert)

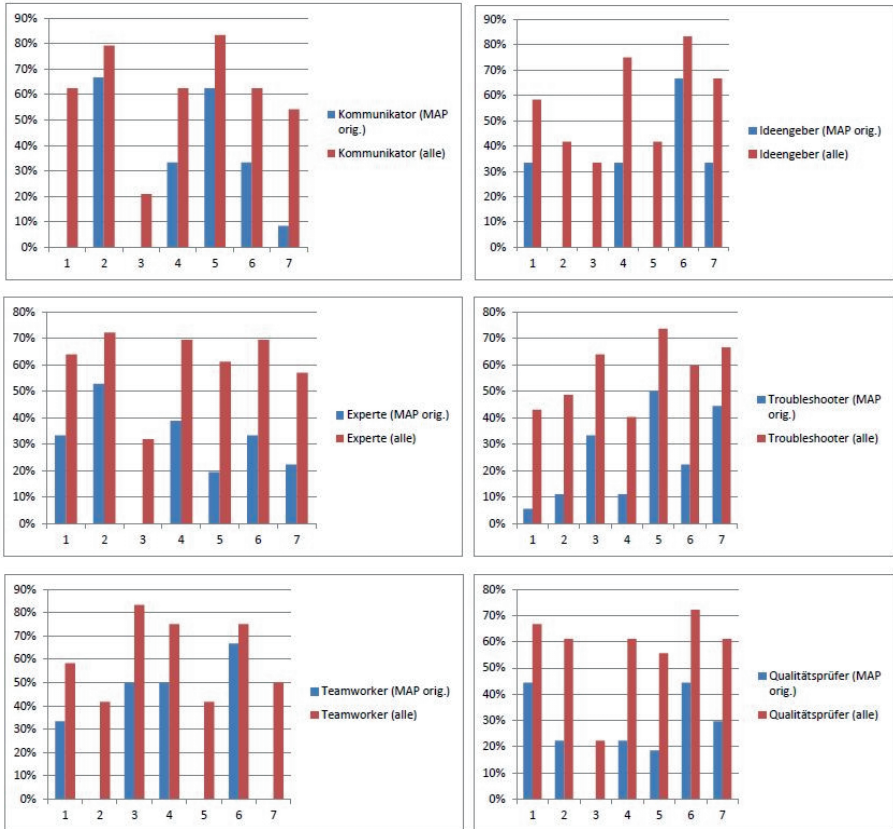


Abb. 3: Verhaltensweisen im Team – Auswertung Projekt 2011

Wie man in Abb. 3 sieht, werden bei der Zuordnung der Verhaltensweisen oder *Typen* zu den Projektmitgliedern zwei getrennte Berechnungen durchgeführt. Deshalb sind jedem Teammitglied (dargestellt durch die Nummern 1 bis 7) zwei Balken zugeordnet. Grundsätzlich ist die Höhe des Balkens ein Maß für die Affinität zum jeweiligen Typ. Der kleinere, helle Balken (links) berücksichtigt dabei nur Ergebnisse aus der Evaluation, wenn sie mindestens in zwei von drei Evaluationen auftauchen. Alles andere wird nicht berücksichtigt. Die gute Zuordnung des Teammitglieds Nr. 2 zum Typ Kommunikator heißt also, dass die entsprechenden projektbezogenen und allgemeinen menschlichen Eigenschaften in zwei bis drei Evaluationen angekreuzt wurden. Einfache Nennungen werden nicht berücksichtigt. Aus der Überdeckung der „Kreuzungsflächen“ der projektbezogenen und der allgemeinen menschlichen Eigenschaften mit den Typen

in der „Landkarte“ der Verhaltensweisen in Abb. 1 berechnet sich die Affinität zum jeweiligen Typ. Anders verhält es sich mit dem höheren, dunklen Balken (rechts). Hier werden alle Nennungen der Eigenschaften berücksichtigt. So bildet sich also pro Teammitglied ein Balkenpaar, welches im Vergleich mit den anderen Teammitgliedern die Zuordnung zum entsprechenden Projekttyp verdeutlicht: Beispielsweise sind die Teammitglieder Nr. 2 und 5 gut als Kommunikatoren geeignet!

Abb. 3 zeigt aber auch Teammitglieder, bei denen die Ausbildung eines Typs nicht sehr ausgeprägt ist. So ist für Teammitglied Nr. 3 bei der Zuordnung zu vier Typen der von der Ausprägung her „schärfere“ linke Balken gleich 0. Der zugehörige rechte, dunkle Balken zeigt viermal eine Höhe zwischen 20% und 30%. Dieses Teammitglied ist also bezüglich der vier Verhaltensweisen Kommunikator, Ideengeber, Experte und Qualitätsprüfer recht ausgeglichen und „unauffällig“. Es zeigt sich aber eine Affinität zum Teammitglied (Teamworker) und – abgeschwächt – zum Troubleshooter.

4 Konkurrierende Teams – Vorhersage des Teamerfolgs mit MAP

Um den Projekterfolg vergleichen zu können, ist es interessant, Teams zu beobachten, die ein Projekt in Konkurrenzsituation unter gleichen Bedingungen zu bewältigen haben. Diese Situation, die im Unternehmen eher selten entsteht, ist in Hochschulprojekten öfter zu beobachten.

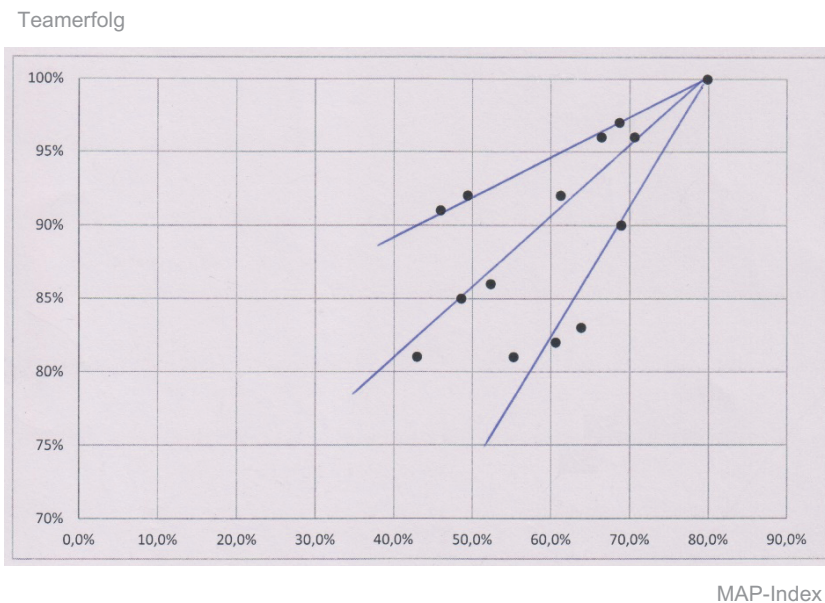


Abb. 4: MAP-Index vs. Teamerfolg – Vergleich von 14 konkurrierenden studentischen Teams der DHBW

Abb. 4 zeigt einen aktuellen Vergleich von 14 konkurrierenden studentischen Teams der DHBW mit derselben Projektaufgabe. Dabei repräsentiert jeder Punkt ein Team. Der Einfachheit halber wird für jedes Team ein „MAP-Index“ (X-Achse) berechnet, der die Werte für die sechs MAP-Verhaltensweisen, die in Abb. 3 beschrieben sind, mit Gewichtungskoeffizienten multipliziert und aufaddiert [Ha15]. Die Y-Achse ist ein Maß für den Projekterfolg. Abb. 4 zeigt deutlich, dass Teams mit einem besseren „MAP-Index“ (rechts auf der X-Achse) in der Regel auch einen größeren Projekterfolg haben. Dieser Zusammenhang wird deutlicher mit dem Level des MAP-Index, also mit der Zunahme der „sozialen Qualität“ des Teams.

Diese „systemische“ Sicht auf das Projektteam zeigt, dass das „Empowerment“ des Teams durch Optimierung seiner Teamstruktur (Maximierung des MAP-Index) in der Projektpraxis wichtiger ist, als die Wahl des Vorgehensmodells. Für das MAP-Team ist das gewählte Vorgehensmodell nur der „Startprozess“, der im Rahmen der MAP-Regeln verändert und angepasst werden darf und soll. Die (zugegebenermaßen provokante) Frage sei erlaubt, inwiefern es dann überhaupt noch Entwicklungsprozesse braucht: Reicht für den Projekterfolg nicht ein in Prozesspraktiken gut geschultes Team, dessen Schlüsselpositionen durch eine MAP-Evaluierung besetzt werden? Gilt also für moderne Entwicklungsteams, dass das System (also die Struktur des Teams) wichtiger ist als der gewählte Prozess? Die bisherigen Ergebnisse des Autors lassen diesen Schluss zu.

Literaturverzeichnis

- [Ba04] Baldegger R.: Erfolgreich im Team. Baldegger-Verlag, 2004.
- [Be79] Beer S.: The Heart of Enterprise. Wiley, Chichester, 1979.
- [Be81] Beer S.: Brain of the Firm. 2nd ed., Wiley, Chichester, 1981.
- [Be04] Belbin R. M.: Management Teams. Elsevier, 1981, 2004.
- [Be07] Belbin R. M.: Team Roles at Work. Elsevier, 1993, 2007.
- [BG01] Baldegger R., Gotsmann L.: Ganzheitliches Projektmanagement. Baldegger Verlag, 2001.
- [Ha10] Hanser E.: Agile Prozesse: Von XP über Scrum bis MAP. Springer, Heidelberg, 2010.
- [Ha15] Hanser E.: Erfolgreiche Teams mit dem Meta Agile Process Model (MAP). In „Agiles Management – Innovative Methoden und Best Practices“. Symposium, 2015.
- [HW12] Hanser E., Wehinger J.: Integration MAP in VSM. Private Notizen, 2012.
- [Sc00] Schwaninger M.: Das Modell Lebensfähiger Systeme - Ein Strukturmodell für organisationale Intelligenz, Lebensfähigkeit und Entwicklung. Universität St. Gallen, www.tranquilla.ch/serveattachment/984e4202df9fe30e5459fa815a35c6dc/D35.pdf, 2000, Stand: Februar 2016.
- [VS12] The VSM Guide, www.esrad.org.uk/resources/vsmg_3/screen.php?page=preface, Stand: Mai 2016.

- [Wa08] Wagenhals K.: Diskussion und private Notizen. interPM 2008, Glashütten.
- [We13] Wehinger J.: Jenseits der Vorgehensmodelle – Projektteams wirksam in Organisationen verankern, das Viable System Model und MAP. In „Vorgehensmodelle – Anspruch und Wirklichkeit“, Lecture Notes in Informatics, 2013.
- [WH12] Wehinger J., Herrmann C.: Das Viable System Model als Bezugsrahmen für Agilität am Beispiel Scrum. In „IT-Projektmanagement 2012+ im Spagat zwischen Industrialisierung und Agilität?“, interPM 2012, Dpunkt.verlag, 2012.

Ein Diversity Framework für Informatik-nahe Projekte

Christina Böhm¹, Renate Motschnig²

Abstract: Soziale Kompetenzen und die Fähigkeit, effektiv in interdisziplinären Projektteams zu arbeiten, sind für erfolgreiche Informatik-nahe Projekte unabdingbar und gewinnen durch die Globalisierung stetig an Bedeutung. Allerdings scheinen diese sozialen Qualitäten aktuell in der Informatik-Ausbildung wie auch in der praktischen Anwendung unterrepräsentiert. Um diesem Defizit zu begegnen, wurde das „Diversity Framework“ entwickelt. Es bündelt Methoden, Techniken und Prozesse, welche dazu dienen, implizite, soziale Faktoren und Verhaltensweisen zu explizieren und eine systematische und praxisnahe Integration solcher Aspekte zu unterstützen.

Keywords: Diversität; soziale Kompetenzen; Framework; Ausbildung; Softwareentwicklung

1 Einleitung

Neben technischen und analytischen Fähigkeiten benötigen SoftwareentwicklerInnen soziale, „softe“ Fähigkeiten um in interdisziplinären Teams effektiv zu arbeiten [Cu14], [JH10]. Zu diesen Fähigkeiten zählen unter anderem effektive Kommunikations-Skills [MN09], [MN14] und der Umgang mit unterschiedlichen Verhaltensweisen und Arbeitsstilen sowie mit weiteren interkulturellen Aspekten.

Wenngleich diese Faktoren die Zusammenarbeit in Teams wesentlich beeinflussen können [AB15], [DH05], [IJ91], gibt es – weder im Forschungsbereich noch in der Praxis – systematische Unterstützung für das Thema [Bö15], [BM15]. Um eine Integration dieser Aspekte in der Praxis und Ausbildung zu ermöglichen wurde das sogenannte „Diversity Framework“ im Rahmen der Doktorarbeit der ersten Autorin dieses Beitrags an der Universität Wien entwickelt [Bö16]. Für den Zweck des Frameworks ist Diversität als die Vielfalt an unterschiedlichen Verhaltensweisen (und darunterliegenden Werten) definiert, die Individuen erfahren, wenn sie in Projektteams zusammenarbeiten. Der Fokus liegt dabei auf nicht-sichtbaren, impliziten Diversitätsaspekten wie beispielsweise unterschiedliche Arbeitsstile und -präferenzen, welche einen Einfluss auf den Projekterfolg haben können. Zu diesen Präferenzen zählen unter anderem wie im Team kommuniziert wird, wie Beziehungen aufgebaut und Entscheidungen getroffen werden, wie Projekte geplant und durchgeführt werden, wie Informationen geteilt werden oder wie Aufgaben bearbeitet werden [Bö16].

Das Diversity Framework unterstützt das Management von impliziten,

¹ Universität Wien, Fakultät für Informatik – Forschungsgruppe CSLEARN (Educational Technologies), Währinger Straße 29, 1090 Wien, christina.boehm@univie.ac.at

² Universität Wien, Fakultät für Informatik – Forschungsgruppe CSLEARN (Educational Technologies), Währinger Straße 29, 1090 Wien, renate.motschnig@univie.ac.at

verhaltensbasierten Unterschieden in Teams. Es bietet einen systematischen, strukturierten Zugang [Bö16] zur Identifizierung von verhaltensbasierten Unterschieden und Gemeinsamkeiten der Projektteammitglieder. Im Gegensatz zu anderen, existierenden Ansätzen fokussiert das Framework auf die Individuen im Team und stellt dadurch den Faktor „Mensch“ [CH01] und dessen Erfahrungen, Fähigkeiten und Präferenzen in den Vordergrund [Bö14], [Mo12]. Darüber hinaus ist das Frameworks generisch und daher sowohl für traditionelle als auch agile (z.B. [Hi04] oder [SB01]) Vorgehensmodelle in Projekten anwendbar und offen für Anpassungen, Tailoring und Erweiterungen [Bö16]. Dies wird durch eine Integration in Form eines eigenen Workflows erreicht, der in den Rational Unified Process (RUP) [Kr04] eingebunden ist.

Dieser Workshop-Beitrag gibt zuerst einen Überblick über das verhaltensbasierte Diversity Framework und beschreibt, wie in 5 Stufen kulturelle, zwischenmenschliche Unterschiede in Projektteams erkannt werden können. Des Weiteren werden Möglichkeiten der Integration des entwickelten Diversity Frameworks in die Ausbildung von SoftwareentwicklerInnen – auf akademischen sowie praxisnahem Level – beschrieben. Darüber hinaus werden Vorschläge für weitere, zukünftige Forschung und Entwicklung angeboten.

2 Das verhaltensbasierte Diversity Framework

Das an der Universität Wien entwickelte Diversity Framework basiert auf qualitativer Forschung sowie auf einer systematischen Literaturstudie und wurde sowohl quantitative als auch qualitativ auf Aspekte wie Relevanz, Vollständigkeit, Usability, Anwendbarkeit und Effizienz validiert. Das Framework besteht aus fünf systematischen Phasen, welche die Anwendung in Praxis und Ausbildung ermöglichen sollen (siehe Abb. 1).

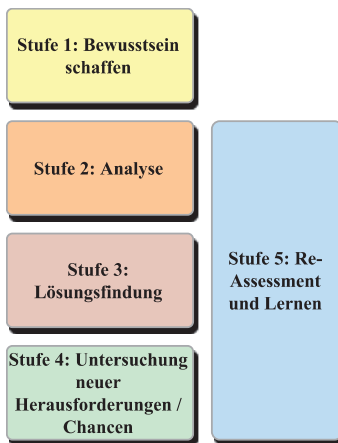


Abb. 1: Die 5 Stufen des Diversity Frameworks

2.1 Stufe 1: Bewusstsein schaffen

In dieser ersten Stufe wird vorerst geklärt, ob eine Anwendung des Frameworks in der jeweiligen Projektsituation überhaupt sinnvoll beziehungsweise möglich ist. Des Weiteren soll ein Bewusstsein für das Thema Diversität im Team sowie ein gemeinsames Verständnis erzeugt werden.

2.2 Stufe 2: Analyse

Die zweite Stufe des Frameworks beschäftigt sich mit der Durchführung einer Diversitäts-Analyse, welche nähere Details zu den Erwartungen, Verhaltensweisen sowie Differenzen und Gemeinsamkeiten hinsichtlich dieser Präferenzen im Team aufdecken soll. Dazu werden in einem ersten Schritt sogenannte „Diversity Features“ diskutiert und Unterschiede identifiziert. Die Diversity Features sind vordefinierte, gruppierte Verhaltenskategorien, welche besonderen Einfluss auf den Projekterfolge haben können. Beispiele für solche Diversity Features sind zum Beispiel: wie im Projekt kommuniziert wird (z.B. direkt versus indirekt); wie Projekte geplant und durchgeführt werden (z.B. agil versus traditionell); wie Informationen geteilt werden (z.B. restriktiv versus offen); wie Aufgaben bearbeitet werden (z.B. parallel versus seriell).

Zur Identifikation von Unterschieden oder Gemeinsamkeiten kann aus verschiedenen Vorgehensweisen gewählt werden. In einer Studie mit Teams hat sich gezeigt, wie wichtig es ist, dass unterschiedliche Optionen angeboten, da selbst bei ähnlichen persönlichen Hintergründen (z.B. Alter, Ausbildung, etc.) die Präferenzen für oder gegen eine bestimmte Vorgehensweise sehr konträr ausfallen kann [Bö16].

1. Eine offene Gruppendiskussion (team-orientiert).
2. Eine Gruppendiskussion, welche durch die Visualisierung der unterschiedlichen Meinungen durch Positionieren von Figuren auf einem Tisch unterstützt wird (team-orientiert).
3. Eine Gruppendiskussion, welche durch die Visualisierung der unterschiedlichen Meinungen durch Positionieren von Personen im Raum unterstützt wird (team-orientiert).
4. Verwenden eines vordefinierten Fragebogens (individuell-orientiert).

Abb. 2 zeigt exemplarisch, wie die Unterschiede in der Analyse dargestellt werden können. Für alle identifizierten Unterschiede wird im Weiteren versucht spezifische Projektsituationen zu finden, in denen diese Unterschiede negative oder positive Implikationen haben könnten.



Abb. 2: Darstellung der Ergebnisse einer Diversitäts-Analyse anhand zweier Diversity Features

2.3 Stufe 3: Lösungsfindung

In der dritten Stufe wird – aufbauend auf der vorangegangenen Analyse – versucht, die Meinungen, wie mit der speziellen Projektsituation umgegangen werden soll, zu sammeln und eine Lösung zu definieren. Diese Lösung beinhaltet konkrete Richtlinien und Vorgehensweisen für diese Projektsituation und kann in den Projektmanagement-Plänen dokumentiert werden.

2.4 Stufe 4: Untersuchung neuer Herausforderungen / Chancen

Während der Projektdurchführung können neue Herausforderungen oder Chancen – wie beispielsweise veränderte Rahmenbedingungen oder Änderungen in der Projektorganisation – auftreten, welche bisher in den Stufen 1-3 noch nicht betrachtet wurden. In der vierten Stufe werden solche Herausforderungen oder Chancen wiederum systematisch analysiert. Dazu werden die Ursachen betrachtet, individuelle Interessen erfasst und eine Lösungsstrategie für Unterschiede erarbeitet.

2.5 Stufe 5: Re-Assessment und Lernen

Diese fünfte Stufe findet parallel zu den Stufen 2-4 statt. Hier werden nun die definierten Lösungen in regelmäßigen Zyklen (oder bei Bedarf ad-hoc) überprüft und angepasst. Außerdem ist an dieses Re-Assessment eine Lernschleife angebunden, damit Lessons Learned direkt für das Projekt, für die individuelle Entwicklung der Teammitglieder und für die Organisation genutzt werden können.

3 Einbindung des Diversity Frameworks in die Informatik-Ausbildung

Wie eingangs beschrieben, ist es erforderlich, dass soziale Qualitäten auch in die Informatik Ausbildung eingebettet werden. Als Beitrag dazu zeigen hier zwei

Möglichkeiten auf, wie das Diversity Frameworks in die Ausbildung von SoftwareentwicklerInnen eingebunden werden kann:

- Integration in akademische Informatik-Curricula

Das Computer Science Curriculum der ACM / IEEE [Jo13] fordert, dass Studierende ganzheitlich ausgebildet werden müssen. Zu dieser Ausbildung zählt nicht nur die Vermittlung von technischen Fakten, sondern auch von sozialen und persönlichen Kompetenzen. Zwei verschiedene Szenarien sind möglich, um das Diversity Framework auf einem akademischen Level zu integrieren:

- Szenario 1: Die Integration in eine Basis-Lehrveranstaltung zum Thema Software Engineering, oder in eine Lehrveranstaltung zu ausgewählten Kapiteln des Software Engineering. Hierfür könnte das Diversitäts-Thema im Rahmen der Softwareentwicklungsmodelle (u.a. Rational Unified Process [Kr04]) eingebracht werden, da moderne Vorgehensmodell bereits auch angrenzende Gebiete (z.B. Projektmanagement) berücksichtigen.
- Szenario 2: Die Integration in eine Basis-Lehrveranstaltung mit einem Fokus auf Projektmanagement und/oder soziale Kompetenzen. Im Gegensatz zu Szenario 1 liegt hier der Fokus auf effektiver Zusammenarbeit in (Projekt-)Teams.

- Integration in internationale Projektmanagement-Standards

Da das Framework an einige Gebiete der wichtigsten, bestehenden Projektmanagement-Standards [Hi12], [In15], [Pr13] anschließt, scheint beispielsweise eine Integration in die bestehenden ‚Knowledge Areas‘ (z.B. Communication Management, Risk Management, Stakeholder Management) des PMBOK® [Pr13] oder eine neue Knowledge Area, welche sich mit sozialen Aspekten auseinandersetzt, sinnvoll.

4 Zusammenfassung

Die Wichtigkeit von sozialen Kompetenzen in Softwareentwicklungs-Projekten scheint in der Ausbildung noch nicht angelangt zu sein. Das hier beschriebene Diversity Framework greift diesen Mangel auf und soll dazu beitragen, professionell mit dem Thema in einer zeitgemäßen Informatik-Ausbildung umgehen zu können.

Das Diversity Framework bietet einen Vielzahl an Vorgehensweisen und Möglichkeiten und ist daher an unterschiedliche Situationen und Teams anpassbar. Des Weiteren hat es sich als sehr hilfreich erwiesen, dass das Diversity Framework in den Rational Unified Process integriert ist, da dies zum einen die Anwendung für Softwareentwicklungs-Projekte vereinfacht, aber auch die Anbindung an die Informatik-Ausbildung vereinfacht.

Weitere Forschung sollte sich der Integration des Diversity Frameworks in die

akademische- sowie praxisnahe Ausbildung widmen. Weiters sind Feldstudien zur Verwendung des Diversity Frameworks in Informatik-Projekten vorgesehen.

Literaturverzeichnis

- [AB15] Amster, R.; Böhm, C.: Culture-based Behavior in Global IT Projects: An Investigative Study with Indian Managers of Outsourcing Projects. *Procedia Computer Science* 64, S. 232-239, 2015.
- [Bö14] Böhm, C.: The Connection between Agile Project Management and Human-Centered Principles. In: *Proceedings of the 7th International Conference on Contemporary Business and the 14th Global Conference on Flexible Systems Management (Joint Conference)*. Curtin University, Singapore, S. 661-666, 2014.
- [Bö15] Böhm, C.: Practical Insights on Managing Diversity in International ICT Projects. In (Sushil; Bhal, T.; Singh, S.P. Hrsg.): *Managing Flexibility: People, Process, Technology and Business*. Springer, India, S. 73-84, 2015.
- [Bö16] Böhm, C.: A Framework for Managing Diversity in ICT Projects. Processes and Techniques for Explicating Soft Facts and Dealing with Behavioral Differences. Dissertation, Universität Wien, Wien, 2016.
- [BM15] Böhm, C.; Motschnig-Pitrik, R.: New Research Perspective on Managing Diversity in International ICT Project Teams. In (Chroust, G.; Sushil Hrsg.): *Systemic Flexibility and Business Agility*. Springer, India, S. 21-31, 2015.
- [CH01] Cockburn, A.; Highsmith, J.: *Agile Software Development: The People Factor*. *Computer* 34/11, S. 131-133, 2001.
- [Cu14] Cukierman, U.R.; Palmieri, J.M.: Soft Skills in Engineering Education: A Practical Experience in an Undergraduate Course. In: *International Conference on Interactive Collaborative Learning (ICL)*. S. 237-242, 2014.
- [DH05] Dunavant, B.M.; Heiss, B.: *Global Diversity 2005. Diversity Best Practices*, Washington, 2005.
- [FI02] Flick, U.: *An Introduction to Qualitative Research*, 2. Auflage, SAGE Publications, London, 2002.
- [Hi04] Highsmith, J.: *Agile Project Management. Creating Innovative Products*, 2. Auflage, Addison-Wesley, Boston, 2004.
- [Hi12] Hinde, D.: *PRINCE2® Study Guide*. John Wiley & Sons, Alameda, 2012.
- [Ho14] Holzer, A.; Bendahan, S.; Cardia, I.V.; Gillet, D.: Early Awareness of Global Issues and Development of Soft Skills in Engineering Education: An Interdisciplinary Approach to Communication. In: *International Conference on Information Technology Based Higher Education and Training (ITHET)*. IEEE, S. 1-6, 2014.
- [In15] International Project Management Association: *ICB 4.0. Individual Competence Baseline for Project, Programme & Portfolio Management*. International Project Management Association, 2015.

- [IJ91] Ives, B.; Jarvenpaa, S.L.: Applications of Global Information Technology: Key Issues for Management. *MIS Quarterly*, 15/1, S. 33-49, 1991.
- [JH10] Jouny, I.; Hornfeck, W.: Teaching Soft Skills to ECE Students. In: *Creating Interdisciplinary Skills for Complex Global Environments*. IEEE, S. 1-12, 2010.
- [Jo13] Joint Task Force on Computing Curricula Association for Computing Machinery (ACM) IEEE Computer Society: *Computer Science Curricula 2013. Curriculum Guidelines for Undergraduate Degree Programs in Computer Science*, ACM / IEEE, 2013.
- [Kr04] Kruchten, P.: *The Rational Unified Process. An Introduction*, 3. Auflage, Pearson Education, Boston, 2004.
- [Ma10] Mayring, P.: *Qualitative Inhaltsanalyse. Grundlagen und Techniken*, 11. Auflage, Beltz-Verlag, Weinheim, Basel, 2010.
- [MN09] Motschnig, R.; Nykl, L.: *Konstruktive Kommunikation. Sich und andere verstehen durch personenzentrierte Interaktion*, Klett-Cotta, Stuttgart, 2009.
- [MN14] Motschnig, R.; Nykl, L.: *Person-centered Communication. Theory, Skills, & Practice*, Open University Press, McGraw Hill, UK, 2014.
- [Mo12] Motschnig, R.: Agile Management needs Agile Persons – Developing Inner Flexibility. In: *Systemic Flexibility and Business Agility – Proceedings of the Twelfth Global Conference of Flexible Systems Management*, S. 281-290, 2012.
- [Pr13] Project Management Institute: *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)*, 5. Auflage, Project Management Institute (PMI), Pennsylvania, 2013.
- [SB01] Schwaber, K.; Beedle, M.: *Agile Development with SCRUM*, Prentice-Hall, New York, 2001.

Independent vs. interdependent identities: Eine theoretische Basis zu kulturellen Unterschieden in der Entwicklung und Nutzung von Software

Walter Renner¹, Shiva Khalili²

Abstract: Im Ausschreibungstext wird festgehalten, dass eine Software die Rolle eines "einfühlsamen, höflichen Butlers" einnehmen solle und es wird gefordert, dass Softwareprodukte den kulturellen Erwartungen und Erfordernissen der User anzupassen seien. Wenn User und "Butler" zufriedenstellend interagieren sollen, stellt sich somit die Frage der kulturellen Identität auf beiden Seiten. Der aktuelle Beitrag stellt die kulturpsychologische Basis zu diesen Postulaten und weiterführenden Entwicklungen bereit. In einer Weiterführung des Konzeptes individualistischer vs. kollektivistischer Gesellschaften präsentierten Markus und Kitayama die Unterscheidung einer "unabhängigen" (independent) Identität in den USA und der EU gegenüber einer "wechselseitig abhängigen" (interdependent) Identität in Asien. Diese Überlegungen werden am Beispiel Iran hinsichtlich des politischen Kontextes und bezüglich der religiös-ideologischen, soziokulturellen und genderbezogenen Konsequenzen und Implikationen veranschaulicht. Anschließend sollen diese Aspekte im Workshop diskutiert werden.

Keywords: Softwareentwicklung, Kultur, Identität, self-construal, independent vs. interdependent, Iran, Kollektivismus und Individualismus

1 Die Anfänge in der Philosophie

Kulturelle Unterschiede und deren Auswirkungen auf das Zusammenleben wurden bereits vom Philosophen Tönnies [Tö87] thematisiert: die ursprüngliche Form der Koexistenz, die *Gemeinschaft*, impliziert real begreifbare Verbindungen zwischen Individuen, welche derselben Dorfgemeinschaft bzw. Großfamilie angehören, wie dies etwa in einem Tiroler Bergbauerndorf des frühen 20. Jahrhunderts oder in einer afrikanischen Großfamilie auch heute noch der Fall ist. Dem stellte Tönnies das Modell der *Gesellschaft* gegenüber: mit der Industrialisierung etablierten sich neue Staatsformen auf der Grundlage abstrakter Vereinbarungen und Regeln, basierend auf Arbeitsteilung, Gesetzen, Versicherungen, Sozialsystemen und anderen Merkmalen komplexer Gesellschaftlichen.

2 Die empirische Basis

¹ Pan-European University, Faculty of Psychology, Tomasikova 20, P. O. Box 12, 82009 Bratislava, Slowakische Republik, walter.renner@aau.at

² Independent Researcher, shivakhalili@yahoo.com

In ähnlicher Weise unterschied Hofstede [Ho84], [Ho01], [HHM10] zwischen kollektivistischen (entsprechend der Gemeinschaft) und individualistischen (entsprechend der Gesellschaft) Kulturen. Anhand der internationalen Datenbasis des IBM-Konzerns konnte Hofstede diese und andere Dimensionen für den Vergleich der länderspezifischen Gegebenheiten für die meisten Staaten der Welt empirisch abbilden. Dadurch wurde das Konzept von *Individualism vs. Collectivism* (I/C) weltweit bekannt und zum Beispiel von Harry C. Triandis, der im kollektivistischen ländlichen Griechenland aufgewachsen und schon früh mit seinen Eltern in die individualistisch geprägten USA emigriert war, aufgegriffen [Tr95].

Sowohl Hofstede als auch Triandis verdeutlichten, dass I/C nicht als kategorial zu trennende Alternativen, sondern als die beiden Enden eines Kontinuums zu verstehen sind, entlang dessen sich Kulturen positionieren lassen. So sind etwa die nördliche EU hinsichtlich ihrer autochthonen Bevölkerung und die weiße U.S.-Mittelschicht heute ganz überwiegend als individualistische Kulturen positioniert, während Minderheiten, wie etwa die türkischstämmige Community in Deutschland und Österreich auch im 21. Jahrhundert noch stark von kollektivistischen Werten geprägt ist.

Auch unter dem Aspekt der Globalisierung sind der arabische Raum und andere moslemische Länder wie der Iran, Afghanistan und die Türkei, der gesamte asiatische Kontinent, sowie Afrika südlich der Sahara bis heute eindeutig kollektivistisch. Dies gilt trotz der sonstigen kulturellen Unterschiede, etwa zwischen Indien, Taiwan und Japan oder zwischen Ägypten, Nigeria oder einer südafrikanischen Stammeskultur.

3 Kulturpsychologie

Shweder [Sh96], [Sh08] und Shweder und Haidt [SH00] differenzierten in ihrer *cultural psychology* hinsichtlich der kulturspezifischen Wertorientierungen zwischen einer (1) *Ethic of Autonomy*, entsprechend individualistischen Werten und einer (2) *Ethic of Community* und (3) einer *Ethic of Divinity*, welche unterschiedlichen Facetten einer kollektivistischen Kultur entsprechen.

Während die *Ethic of Autonomy* lediglich verbietet, was anderen Individuen schadet oder sie in ihren Rechten beschneidet, gelten in der *Ethic of Community* Werte wie Gruppenloyalität, Familienehre, Respekt, Keuschheit und Demut auch dann als allgemein verbindlich, wenn aus deren Verletzung kein objektiver Schaden entsteht. Die *Ethic of Divinity*, die etwa für die Kulturen Südostasiens charakteristisch ist, fokussiert schließlich auf religiös begründete Reinheitsgebote und auf die Hochachtung vor allem Lebendigen, weil etwa Tiere der Sitz einer göttlichen Seele sind.

4 Identität: Independent vs. Interdependent

Während die bisherigen Ansätze auf Eigenheiten der Kulturen abgestellt hatten, fokussierten Markus und Kitayama [MK91] ebenfalls entlang der Linien von I/C auf unterschiedliche "*self-construals*", also Sichtweisen einer Person von sich selbst oder, präziser, von ihrer Identität als "Mitglied der Familie Müller" (*interdependent*) oder als "Frau Gerda Müller" (*independent*).

Eine interdependente Selbstdefinition impliziert im Extremfall, dass "kollektivistische" Personen, die etwa durch Flucht und Migration von ihrer Ursprungsfamilie getrennt sind, das erleben und erleiden, was Eisenbruch [Ei91] als *cultural bereavement* charakterisiert hatte: die bloße Trennung von der identitätsstiftenden Bezugsgruppe hat nicht nur traumatischen Charakter, sondern stellt die Identität der betroffenen Person ernsthaft in Frage.

5 Das Beispiel Iran

Die Bedeutung kultureller Dimensionen wie Individualismus/Kollektivismus und anderer politischer und soziokultureller Merkmale für die IT und Softwareentwicklung können am aktuellen Beispiel der islamischen Republik Iran diskutiert werden.

Politik, Religion und Kultur: kollektivistische Kultur Irans zwischen Islamisierung und Globalisierung

Seit der Islamischen Revolution 1978-79 stellt die theokratische Regierung Irans eine Vereinigung von Macht, Politik und Religion dar. Die Kulturrevolution oder Islamisierung in allen Aspekten der persönlichen (Geschlechtsidentität, Familienleben) und sozialen Lebens (einschließlich Kunst, Wissenschaft und Bildung, Sport, etc.) sind seitdem gefolgt.

Konformität hat als *personal value* [Sc12], der zu Konfliktvermeidung und Sicherheit beiträgt, im Iran Priorität. Diese trägt zur Stärkung der kollektivistischen Kultur, sowie - auf einer niedrigeren Ebene - zur Verstärkung der spezifischen Gruppenidentitäten bei. Zugleich wird dadurch ein Raum für die persönlichen Strebungen und Interessen des Individuums geschaffen.

In dieser Hinsicht werden neue Technologien und Software von vielen Iranern als Chance gesehen, um Zugang zu Information, Wissen und Wissenschaft zu erlangen, aber auch um persönliche Interessen zu folgen, um Verbindungen zu anderen (neuen Gruppenidentitäten) zu herzustellen, um Gelegenheiten für Veränderung und Entwicklung zu schaffen, aber auch, um neue Erfahrungen und Spaß zu erleben.

Heute zeigen die IT-Entwicklung, Produktion und Anwendung im Iran ein Wechselspiel zwischen

(1) Politik und Religion mit der Tendenz die (Veränderungen der) Kultur zu kontrollieren und die Menschen in ihren individuellen Entscheidungen und in ihrer Freiheit einzuschränken - und

(2) der individuellen Sehnsucht der Iraner nach Freiheit in Entscheidung und Meinung und dem Bedürfnis eines freien privaten Bereiches (gemeint ist einen Bereich, in dem die Person sich ohne Regulierungen und Zwänge von Behörden und den damit verbundenen Ängsten ausdrücken darf). Diese Sehnsucht ist im Hinblick auf die sich laufend verändernde Umwelt und im Licht der Globalisierung zu sehen.

Diese Trends sind in den Bemühungen der Islamischen Republik um Zensur und Kontrolle des Internets, Social Media (Twitter, Facebook, YouTube, Google Plus) [Sh14], um Kontrolle der Inhalte von Nachrichten und Bilder, Blogs und sogar der privaten Bereiche wie persönliche E-Mails, mobile Gespräche, Apps und deren Inhalte zu beobachten.

Die iranische Bevölkerung hat trotz dieser Schwierigkeiten und Einschränkungen ein großes Interesse an neuen Technologien gezeigt. Heute hat der Iran mit mehr als 46 Millionen Internet-Nutzern im Nahen Osten, nach Israel den zweithöchsten Prozentsatz der Bevölkerung online vernetzt [An13].

Der gleiche Trend kann - sowohl in dem Bevölkerungsinteresse als auch in der staatlichen Regulierung, Zensur und Kontrolle - im iranischen Gamer Bereich [E116] beobachtet werden. Diese Kontrolle und Zensur im Iran dient der Stabilität des Landes und der Verteidigung und Pflege seiner traditionell-religiösen Kultur sowie der Vermeidung der Ausbreitung unerwünschter Ideen, Kritik und Aktivitäten, die als Gefahr für die nationale Sicherheit und für das religiöse und politische Establishment interpretiert werden. Bereits der Verdacht eines solchen Vorhabens wird hart bestraft.

Es wird angenommen, dass die erforderliche Technologie und Werkzeuge zu diesem Zweck von den westlichen Ländern und Firmen an iranische Behörden verkauft wurden [RC09]. Somit werden der Iran und China als höchstrangige Länder mit Internet-Zensur/Kontrolle und der Iran als eines der Länder mit sehr aktiven regierungsabhängigen Hackern (beschäftigt mit Cyber-Piraterie und Cyber Spionage) genannt [LHS16].

Es werden durch die Regierungsunterstützte IT- und Software-Organisationen viele national-, religions- und kultursensible Spiele und edukative Softwares erzeugt, obwohl die Qualität meistens nicht mit dem internationalen Teil (d.h. mit den im Iran verfügbaren Raubkopien) konkurrieren kann [LHS16]. Die Vielfalt individueller Interessen und Vorlieben sowie das weite Spektrum der iranischen ethnischen Gruppen und diverser religiösen Traditionen wird dabei nicht annähernd abgedeckt.

Das Beispiel Iran deutet auf die Notwendigkeit der Überprüfung und Untersuchung der Risiken und Vorteile der Fokussierung auf kulturelle Identitäten vs. inklusivistisch-universalistischen Identitäten im 21. Jahrhundert im Bereich der IT und Software-Produktion.

Während eine von Ideologie, Religion und Wirtschaft angetriebene Politik einen Fokus auf "eigene" Kultur, Nationalität, Rasse (Themen und Inhalte wie weiße Rasse, westlichen Lebensstil, Amerikaner als Helden in vielen Spielen) und religiöse Identitäten (z.B. ISIS) legt, kann dieser Ansatz bei Usern zur Bildung und/oder Verstärkung von Ingroup/Outgroup Identitäten führen: das Gefühl doppelter Standards, Wettbewerb, und deren Folgen wie Diskriminierung, Aggression und radikale Ansichten werden gefördert.

Identitäten tendieren in der Regel in stabilen Situationen stabil zu bleiben: Aktionen und Erzählungen (narratives) sind vorhersehbar. Änderungen in Kontext, Werkzeuge (Tools: inkl. IT und Software) und Umwelt bieten Möglichkeiten für neue Aktionen und damit Änderungen in Identitäten und Kulturen an.

Wir leben in einer sich schnell verändernden Welt mit Fortschritten in Wissenschaft und Technologie. Wir sind mit einer Vielfalt von Informationen und Möglichkeiten konfrontiert. Daher wird von der Software, erwartet unterschiedlichen Populationen Hilfe und Unterstützung für ihre Aufgaben und Kapazitäten zu geben, wie sie den Anforderungen des 21. Jahrhunderts entsprechen (z.B.: kritisches Denken, Toleranz und die Fähigkeit zur Zusammenarbeit und Koexistenz mit und in der Vielfalt).

Neue Erzählungen, neue Aufgaben, Werkzeuge (Tools), Jobs... können persönliche und Gruppenidentitäten destabilisieren. In dieser Situation haben die Teilnehmer/User die einmalige Gelegenheit, ihre Identitäten neu zu verhandeln (renegotiate) [B110]. Für Software-Entwickler und Anwender in 21. Jahrhundert ist es entscheidend, die Sensibilität dieser Phase der Identität Neuverhandlungen (identity renegotiations) und die Herausforderungen der Mitglieder aller Kulturen in Bezug auf diese Veränderungen und Transformationen der individuellen und kollektiven Identitäten zu berücksichtigen.

Die Risiken und Vorteile der einengenden spezifischen Gruppenidentitäten gegenüber breiteren, universalistischen und globale Identitäten sollte geprüft werden. Es sollte auf der gemeinsamen Achtung für die wissenschaftliche Bildung (inkl. Informationen über alle Kulturen in ihrem evolutionären und Entwicklungskontext) und auf psycho-educativen Programme (wie Apps und Software, die Life-Skill-Trainings inkl. Kritisches Denken und Verstärkung des Bewusstseins über menschliche kognitiven Verzerrungen enthalten; Programme und Spiele, die individuellen moralischen Grundlagen der Fairness und Care/Empathie verstärken etc.) aufgebaut werden. Hinzu sollten eine respektvolle Haltung und das Zelebrieren aller Kulturen als Erbe der modernen Menschen kommen: dadurch können die Ingroup-outgroup Aggression und

Diskriminierung im Zusammenhang mit der Zugehörigkeit zu einer bestimmten Nation, Rasse oder Kultur verringert werden.

IT- und Software-Entwickler können durch spezifische Symbole und Elemente (wie Sprache, Kunst, Musik, Mythologie) diverser Kulturen zur Vielfalt beitragen, statt nur eine, etwa die U.S.-amerikanische Kultur zu favorisieren. Durch die Berücksichtigung breiter Gruppen von Anwendern/User soll die vielfältige Gruppenidentität potentieller User zu den oben erwähnten Programmen und ihren Wirkungen beitragen.

Auf diese Weise kann die IT die kulturellen Potentiale der Anwender/Users hervorheben und Möglichkeiten zur Identitätsfindung bereitstellen: dies betrifft zum Beispiel Gespräche über kulturelle Identität sowie Nacherzählungen der persönlichen Identitätsgeschichten [B110], jedoch auch die Möglichkeit anders zu handeln und dadurch ein geeignetes Klima für die Identität Neuverhandlung zu schaffen.

6 Kulturelle Identität und Software

Beim Programmieren von Software soll diese so gestaltet werden, dass sie gewissermaßen "intuitiv" den Erwartungen, Wahrnehmungs- und Denkgewohnheiten der Userinnen und User folgt. Etwa um die Jahrtausendwende wurde in der Psychologie wie oben dargestellt die Sichtweise populär, dass Gesetzmäßigkeiten nicht – wie man früher naiv annahm - universell gelten, sondern kulturelle Unterschiede und Gemeinsamkeiten berücksichtigt werden müssen.

In gleicher Weise soll bei der Gestaltung von Software eine kulturspezifische Perspektive den zuvor geltenden Universalismus ersetzen, der aus westlichem Ethnozentrismus resultierte.

Die vorausgehenden Überlegungen stellen die Basis für eine ausführliche Diskussion im Workshop dar, aus der wir erste praktisch realisierbare Ansätze für eine solche neue Sichtweise gewinnen wollen.

Literaturverzeichnis

- [An13] Anderson, C.: Iran Infrastructure and Policy report. Small Media. <https://smallmedia.org.uk/sites/default/files/u8/iiipjuly.pdf>, retrieved: 17.06.2016
- [B110] Blume, T.W.: Counseling for Identity Renegotiation. In Identity: An International Journal of Theory and Research, 10:92-105, 2010. Taylor and Francis Group, LLC.
- [Ei91] Eisenbruch, M.: From post-traumatic stress disorder to cultural bereavement. Diagnosis of Southeast Asian refugees. Social Science and Medicine, 33, 6/91, S. 673-680, 1991.

- [El16] Elmjouie, Y. The Game Industry of Iran. <http://www.polygon.com/features/2016/1/14/10757460/the-game-industr-of-iran>, Jan 14 2016. Retrieved 10.06.2016
- [Ho84] Hofstede, G.: Culture's consequences. International differences in work related values (2nd ed.). Sage, 1984.
- [Ho01] Hofstede, G.: Culture's consequences – Comparing values, behaviors, institutions and organizations across nations (2nd ed.). Thousand Oaks, 2001.
- [HHM10] Hofstede, G., Hofstede, G. J., Minkov, N. Cultures and organizations. Software of the mind. Intercultural cooperation and its importance for survival (3rd ed). McGraw Hill, 2010.
- [LHS16] Larson, E., Hurtado, P., Strohm, C. Iranians Hacked From Wall Street to New York Dam, U.S.Sas. March 24, 2016. <http://www.bloomberg.com/news/articles/2016-03-24/u-s-charges-iranian-hackers-in-wall-street-cyberattacks-im6b43tt> / retrieved on June 16, 2016
- [MK91] Markus, H. R., Kitayama, S.: Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98, 2/91, S. 224-253, 1991.
- [RC09] Rhoads, C., Chao, L. : Irans Web Spying Aided By Western Technology, in *The Wall Street Journal*, June 23, 2009. Retrieved: 18.06.2016
- [Sc12] Schwartz, S. H. : An Overview of the Schwartz Theory of Basic Values. *Online Readings in Psychology and Culture*, 2 (1), 2012. <http://dx.doi.org/10.9707/2307-0919.1116>.
- [Sh96] Shweder, R. A.: Thinking through cultures. *Expeditions in cultural psychology* (4th ed). Cambridge University Press, 1996.
- [Sh08] Shweder, R. A. : The cultural psychology of suffering: The many meanings of health in Orissa, India (and elsewhere). *Ethos*, 36, 1/08, Seite 60-77, 2008.
- [SH00] Shweder, R. A., Haidt, J. (2000): The cultural psychology of the emotions: Ancient and new. In (Lewis, M., Haviland-Jones, J.M., Hrsg.): *Handbook of emotions* (2nd ed.), S. 397-414, Guilford Press, 2000.
- [Sh14] Shaheed.A.: Layers of Internet Censorship in Iran. United Nations Special Rapporteur.<http://shaheedoniran.org/english/blog/layers-of-internet-censorship-in-iran/> Stand: 10.06.2016
- [Tö87] Tönnies, F.: *Gemeinschaft und Gesellschaft*. Fues, 1887.
- [Tr95] Triandis, H. C.: *Individualism and collectivism*. Westview Press, 1995.

Communication, Management and Teambuilding Issues in Austrian-Ukrainian Outsourcing Project: 10 Years of Experience and Future Challenges

Walter Khom¹, MykolaTkachuk², Volodymyr Sokol³, Oleksiy Kosmachov⁴, and Vladyslav Sokol⁵

Abstract: Outsourcing von Softwaredienstleistungen gehört heute zu den wichtigsten und stetig wachsenden Wirtschaftsfaktoren. Outsourcing bietet nicht nur Kosteneinsparungspotenziale, sondern vielmehr auch die Möglichkeit des ad hoc Aufbaus und Zugriffs auf Personalressourcen. Jedoch erfordert Outsourcing eine gute Organisation bzw. ein effizientes Management um auch die gewünschten positiven Ergebnisse zu erzielen. Eine der häufigsten Problemfelder bei Software-Outsourcing sind kulturelle Unterschiede und verschiedene Denkweisen zwischen Auftraggeber und Outsourcing-Dienstleister. Das Bewusstsein darüber gehört zu den kritischen Erfolgsfaktoren einer effizienten Kommunikation und eines effizienten Managements, gleichermaßen wie die passende Organisations- und Teamstruktur auf beiden Seiten. Dieser Artikel stellt einen Erfahrungsbericht aus 10 Jahren Software-Outsourcing zwischen Unternehmen in Österreich und der Ukraine dar und behandelt insbesondere aktuelle und künftige Herausforderungen einer solchen Kooperation.

Keywords: Software Outsourcing, Kulturelle Unterschiede, Mehrsprachige Kommunikation, Agiles Projektmanagement, Organisationsstruktur, Teambuilding.

1 Introduction

Nowadays software outsourcing becomes a very important industrial driver in the globalized world, especially for the countries with transition economy. It concerns surely Ukraine where according to some last statistical studies, e.g. in [ITU16], the IT-sector during 2014-2016 is grown ca. in 18 %, and the essential part of Ukrainian IT-companies (ca. 59 %) are dealing actively with outsourcing projects. That is why the problems of effective outsourcing projects management (OPM) and international cooperation become more important for a lot of middle-and small Ukrainian IT-companies [ITU16].

During the last dozen years many publications were issued in the field of outsourcing project management, especially dedicated for definition of *key success factors* (KSF) in

¹bitmedia e-solutions GmbH, Kärntner Straße 337, 8054 Graz, Austria, Walter.Khom@bitmedia.at

² National Technical University “Kharkiv Polytechnic Institute” (NTU “KhPI”), Dept of Software Engineering and Management Information Technologies (SEMIT), Frunze str., 21, 61002 Kharkov, Ukraine, tka.mobile@kpi.kharkov.ua

³ NTU “KhPI”, Dept of SEMIT, Frunze str., 21, 61002 Kharkov, Ukraine, vlad.sokol@gmail.com

⁴ NTU “KhPI”, Dept of SEMIT, Frunze str., 21, 61002 Kharkov, Ukraine, alexey.kosmachev@gmail.com

⁵ Academy Smart LTD, Nauky Ave., 56, 61000 Kharkiv, Ukraine, vladyslav.sokol@academysmart.com.ua

this domain. E.g., the comprehensive review of IEEE sources given in [KK13] lists the following KSF: *Contract Flexibility, Trustworthy Relationship Management, Corporative Knowledge Sharing, Conflict Reconciliation Mechanism* and some others. One of such critical important KSF within the whole OPM-framework is the necessity to deal with cultural differences and mindsets between customer and outsourcing IT-company. This issue is also discussed intensively in a lot of publications (see e.g. in [THT04], [Ba09], [HHM10]), and one of the recognized expert in this area G. Hofstede [HHM10] defines the following four main dimensions to distinguish specific national cultures, namely: *Power Distance, Uncertainty Avoidance, Individualism vs. Collectivism, Masculinity vs. Femininity, Long Term vs. Short Term Orientation, and Indulgence vs. Restrain.*

The objectives of this paper is to report our lessons learned from 10 years of experience in the real-life software outsourcing project between Austrian and Ukrainian IT-companies, and to figure out some future challenges to be met in this cooperation. The rest of this paper is organized in the following way: Section 2 describes our project background, its main features and technological evolution. In Section 3 we present our multi-language agile project management framework, and in Section 4 the competency-centered and university-joined approach to teambuilding in our cooperation is briefly outlined. Section 5 concludes the paper with a short summary and provides an outlook on the next steps to be done in order to improve the presented OPM-framework.

2 Our Project Background, its Main Features and Technological Evolution

The SITOS e-Learning platform project has history from year 2000. This is a learning management system that supports modern e-learning and communication methodologies. SITOS e-Learning platform proposes web platform with responsive design for all conventional devices and allows on-site installation [Si16]. The main goal of the project was to develop a flexible software that may be operable in different environment, easily customizable, with intuitive user interface and low costs of maintenance. Starting from 2006 the SITOS development and maintenance activities are carried out mainly as the outsourcing project by the developer team located in Kharkiv, Ukraine.

In order to rapid react on customer needs, and to provide easy and low project implementation costs the well-proved PHP technology was chosen. The application software is developed to allow flexible system configuration according to the customer requirements. It includes:

- maintain base core functionality of learning management system and collaboration features;
- provide API for software extension;

- design tools for customization of web UI design;
- support multiple versions and configurations according to customer requirements;
- role-based access control to system information resources.

The structural change in software leads to evolution of used system architecture (see Table 1.). It includes development of single sign-on interfaces to access corporative data resources, and Web-services to provide API for peer-to-peer integration with other software e. g. WBTplus Authoring Tool[WB16]. All added features are marked with “+” sign.

System architecture & technologies used / Years	2006-2007	2008-2010	2011-2016
2 – tier client server & PHP	+	-	-
3 – tier client server & PHP	-	+	+
SOA (Web services)	-	-	+

Tab. 1: SITOS-project evolution (architecture & technology used)

It is well known that long lasting software development leads to degradation of software structure that gives problems for maintaining current system configurations and for future extension of the software. Therefore, the approach to software refactoring was proposed, which now is performed with respect to such methods:

- usage of modern general purpose PHP framework[Yi16];
- separation of business logic from data presentation functionality;
- apply of modern UX components and design tools;
- support of development process by providing of electronic tools to document new system features and elaborated software solutions.

This refactoring framework allows quick implementation of requirements changes and integration of software with different systems infrastructures. However, it would be interesting to extend software by making supporting of Experience API [Ex16a]. This standard implies to make Learning Record Storage to save the learning progress of different activities and sharing information of personal learning results.

That gives chance to build social networks based on current software functional capabilities. The main concept of such network is described in [Ex16b]. This kind of network represents semantic web that allows sharing and structuring learning content, personal learning profiles and build learning programs according to individual goals and recommendation and experience of passed courses. This requires shifting in future to multi-tier RESTful Web Services architecture [RR07].

3 Multi-language Agile Project Management Framework

An efficient communication is one of the main KSF for any project, and, of course, it becomes one of the critical important KSF in our OPM framework. Especially, that is true for modern software projects with agile development methodology, e.g. in Scrum-based teams, where the customers and developers are supposed to communicate each other immediately and intensively in all project phases and iterations. During our cooperation, we have gone through three main stages:

a. Development on Ukrainian side, all other activities are on Austrian side. Such approach allowed us to make a fast start, however link between those two entities were weak and therefore performance was not sufficient. Typical customer requirements document was in German, so translation chain looked like German ↔ English ↔ Russian (although our company resides in Ukraine, the major speaking language is Russian), what brings additional “lost in translation” effect. While technical issues usually can be translated with not too much distortion, this aspect is very critical with respect to cultural differences of user interfaces etc.

b. As above with addition of some intermediate project manager on Ukrainian side with German language skills and QA- engineer on Austrian side with Russian language skills. It allowed to reduce translation efforts and to improve communication; however, the Ukrainian project manager was far away from Austrian customers and therefore cannot communicate effectively with them directly. That means there was need to have project management on both sides, which brings additional overhead. There are still some communication issues and for small bug reports it took too much time for communicate than there was actual effort to fix.

c. Rollback to the first structure with addition of QA and Team Lead/PM on the Austrian side, with both Russian and German language skills. This structure fits the best, since there is clear understanding of what to do on both sides, small issues can be handled immediately without sending them to Ukrainian team.

It is to mention that during whole time of SITOS outsourcing development the special attention was paid to the improvement of language skills in the Ukrainian developer team. In the Table 2 the average data related to this issue are shown.

Foreign language skills in Ukrainian team / Years	2006-2009	2010-2013	2014-2016
English	Intermediate	Intermediate	Advanced
German	Beginner	Beginner	Intermediate

Tab. 2: Average data about foreign language skills in Ukrainian project team

The Austrian-Ukrainian cooperation was faced with several issues that ranged from of understanding foreign languages till more sophisticated cultured-specific and mindsets differences which are briefly discussed in the next Section 4,

4 Competency-centered and University-joined Approach to Teambuilding

In order to understand some culture-specific and mindsets differences in our cooperation project it is reasonable to consider the appropriate indicators based on Hofstede's culture model dimensions [HHM10] for both project teams, which are presented in the Table 3.

Country (Team) / Culture Dimensions	Power Distance	Individualism vs. Collectivism	Uncertainty Avoidance	Masculinity (M) vs. Femininity (F)
Austria	small	individualism	weak	more M
Ukraine	large	collectivism	strong	more F

Tab. 3: Main cultural differences between Austrian and Ukrainian teams

It is to mention that another 2 dimensions [HHM10]: *Long Term vs. Short Term Orientation, and Indulgence vs. Restrain* are not presented in this table because they are not so valuable for our analysis taken into account specific features of our research domain: long-term outsourcing software development in small-size team [Si16].

The core of Ukrainian development team was formed initially from postgraduate students on SEMIT Department at the NTU “KhPI” (www.kpi.kharkov.ua), who already had been trained in the universities in Austria, particularly in Alpen-Adria University of Klagenfurt (<http://www.aau.at>) and Carinthia University of Applied Science in Villach (<http://www.fh-kaernten.at>). At the first projects phase the selection of new employees were mostly performed among Ukrainian university students, with obligatory basic knowledge of English and German languages. A special attention was paid to their professional IT - skills, as well as such individual features like communication skills, ability to work under pressure and with deadlines. The importance of development team members in sense of their responsibility and self-confidence (w.r.t. some cultural “gaps” shown in Table 3) can be illustrated with the following example:

There was a problem with one of the staff member (call him B.), because he became inattentive, and was not well enough to perform current project tasks. He repeatedly was criticized, but it did not produce results, and it had even been decided to assign to B. a special supervisor for additional control. Obviously, this way would lead to unnecessary project costs and time missing. Then one of team-managers offered the alternative solution: instead of doing more control and more observations for B., the responsibility to manage a new junior in the project was assigned to him, i.e. the developer B. himself became a supervisor. After some time, he started to perform his duties more carefully, and he not only begun himself to work with better quality, but also he helped new junior more quickly to understand the project tasks, solutions etc.

The approach to teambuilding mentioned above allowed us to improve consequentially the quality of our developer team (see Table 4).

Staff Position (%) / Years	2006-2009	2010-2013	2014-2016
Junior	60%	50%	30%
Middle	25%	15%	40%
Senior	15%	35%	30%

Tab. 4: Qualification changes in Ukrainian project team

5 Conclusions and Future Challenges

The elaborated approaches to teambuilding and project management as well as the evolution of software tools provided an opportunity for successful product development for the past ten years of cooperation between our Austrian and Ukrainian IT-companies. The actual state of communication framework and project development gives us a chance to improve the current product in architectural perspective to move to multi-tier distributed application in order to gain a new quality of software solutions and business issues.

Finalizing our applied research, we can conclude that real understanding and accounting of Hofstede's statements, presented in our case with concrete values in Table 3, help us to provide more efficient and comfortable operation in our multi-language agile project management framework. In future in order to improve our team structure and communication it would be interesting to implement a virtual team structure [Ch10] that allows to attract most perspective specialists to development by proposing a high flexible working schedule.

References

- [ITU16] IT Ukraine: IT-Services and Software R&D in Europe's Rising Tech Nation, Version 1.11, http://www.uadn.net/files/ua_hightech.pdf, State: April 2016.
- [KK13] Khan, A.W.; Khan, S.U.: Critical Success Factors for Offshore Software Outsourcing Contract Management From Vendors' Perspective: An Exploratory Study Using Systematic Literature Review, *IET Softw.*, Vol. 7, Iss. 6, pp. 327–338, 2013.
- [THT04] Trompenaars, F.; Hampden-Turner, C.: *Managing People Across Cultures*. Capstone Publishing Ltd., 2004.
- [Ba09] Beardsell, J.: *Managing culture as critical success factor in outsourcing*, SMC Working Paper Series, Iss. 09/2009.
- [HHM10] Hofstede, G.; Hofstede, G. J.; Minkov, M.: *Cultures and Organizations: Software of the Mind* (Rev. 3rd ed.). New York: McGraw-Hill, 2010.
- [Si16] SITOS Lernmanagementsysteme, <http://www.bitmedia.at/produkte/produktfirmen/e->

- [learning-loesungen/sitos-lernmanagementsystem/](#), State: April 2016
- [WB16] WBTplus Authoring Toolbit media e-solutions, <http://www.bitmedia.at/en/products/produktfirmen/e-learning-solutions/wbt-authoring-tool/>, State: April 2016
- [Yi16] Yii PHP Framework: Best for Web 2.0 Development, <http://www.yiiframework.com/>, State: April 2016
- [Ex16a] Experience API, <https://www.adlnet.gov/adl-research/performance-tracking-analysis/experience-api/>, State: 2016
- [Ex16b] Experience API (xAPI): Potential for Open Educational Resources – Part 3, <http://classroom-aid.com/2016/02/25/experience-api-xapi-potential-for-open-educational-resources-part-3/>, State: February 2016
- [RR07] Richardson, L.; Ruby S.; RESTful Web Services. O'Reilly Media, 2007
- [Ch10] The Challenges of Working in Virtual Teams Survey report, http://www.communicationcache.com/uploads/1/0/8/8/10887248/the_challenges_of_working_in_virtual_teams.pdf, State: 2010.

View Harmonization in Software Processes: from the Idea to QuASE

Vladimir A. Shekhovtsov¹ and Heinrich C. Mayr²

Abstract: Qualitätssicherung in Software Prozessen setzt voraus, dass die Beteiligten ein gleiches Verständnis der zu erfüllenden Anforderungen haben. Ein zentraler Aspekt des Quality-Aware Software Engineering (QuASE) ist daher die Harmonisierung von Stakeholder-Sichten bei der Systementwicklung. Das Papier zeigt auf, wie von der ersten Idee zur Unterstützung des wechselseitigen Verstehens von Qualitätsanforderungen das System QuASE entstand: dieses unterstützt nicht nur ein umfassendes „Understandability Management“, sondern integriert auch Wissen aus unterschiedlichen Ticketing-Systemen, welches damit zugänglich und wiederverwertbar gemacht wird.

Abstract: Quality assurance in software processes requires that the involved parties share a common understanding of the requirements to be addressed. As a result, a central aspect of Quality-Aware Software Engineering (QuASE) is to harmonize the stakeholder views in system development. The paper shows how our approach to QuASE evolved from the initial idea to a complete software system supporting mutual understanding of communicated information: this system not only provides a comprehensive set of understandability management techniques but also integrates knowledge from different ticketing systems.

Keywords: quality-aware software development, stakeholder view harmonization; domain-specific model; ontology; knowledge base.

1 Introduction and motivation

To ensure successful development, the participants in a software process have to be continuously aware of the quality level of the system under development regarding both: (1) the technical quality of the system as well as (2) the fulfilment of the requirements put on it. Whereas (1) can be ensured by relying on defined and proven software engineering methodologies, (2) traditionally suffers from impairments of the mutual understanding of the project stakeholders having different professional and educational background. Consequently, the quality expectations of the project stakeholders have to be collected, aligned before and during their transfer into the system and software requirements, which are expected to guide the architectural design and implementation activities. Moreover, during development continuous or at least periodical reviews and

¹ Alpen-Adria-Universität Klagenfurt, Institute for Applied Informatics, Universitätsstraße 65-67, A-9020 Klagenfurt am Wörthersee, Volodymyr.Shekhovtsov@aau.at

² Alpen-Adria-Universität Klagenfurt, Institute for Applied Informatics, Universitätsstraße 65-67, A-9020 Klagenfurt am Wörthersee, Heinrich.Mayr@aau.at

adjustments have to be performed in co-operation of all stakeholders concerned. We refer to such an approach as *quality-aware software engineering* (QuASE).

The challenge which initially motivated our QuASE research was the difficulty to involve non-IT stakeholders into the quality-aware software development process as they hardly can talk about the quality of the system without experiencing it. As a result, our first research question in a short version was: “*Is it possible to involve non-IT stakeholders into the process of eliciting quality requirements and if so, how?*” The QuASE project³ and the software solution to be implemented as a result of this project were initially intended to target that question by finding ways of making all involved parties in the software process speak the same language while discussing quality of a software under development (*harmonizing stakeholder views on quality*).

The QuASE project was carried out in co-operation with four partner software companies, each of them being active in a different field of application. While we collected the knowledge about stakeholder communications from the representatives of the partner companies, and analyzed their completed projects, we came to the conclusion that the original scope of our project had to be extended: from targeting only quality-related communication to harmonizing communicated stakeholder views on arbitrary aspects of the software under development. The rationale for this extension was that, for small and medium-sized projects supported by issue management (ticketing) systems such as Atlassian Jira, the amount of quality-related communicated information is small as compared to other kinds of information. As a result of this scope extension, the QuASE solution was redesigned to implement flexible view harmonization support. This support is provided on the basis of a knowledge-oriented representation of communicated information collected in project repositories (the databases maintained usually using industrial issue management systems).

In this paper, we track down the evolution of ideas which served as the foundations for the QuASE research, together with the evolution of the QuASE solution itself. Also, we describe the current state of the project and its future prospects.

The paper is organized as follows. In Section 2, we describe the preliminary research on quality conceptualization in user-centered requirements engineering and show the connections between this research and ensuing QuASE ideas. Section 3 outlines the initial problem statement for the QuASE project. Section 4 discusses the further evolution of QuASE ideas by (1) making it model-based with the specifics of the particular deployment site defined by means of a metamodel-based modeling language and (2) extending the scope of the project to cover arbitrary communicated information. The current state of the QuASE project is sketched in Section 5. This section also shows how the generic QuASE solution can be customized to solve the original problem

³ The QuASE project was an FFG Bridge 1 project conducted from 01.03.2013 until 28.02.2015 in cooperation with four Austrian IT small and medium-sized IT companies

of quality view harmonization, it is followed with a description of the related work, the possible directions for the future research and conclusions.

2 Preliminary research

The origins of the QuASE project can be tracked down to research activities that aimed at conceptualizing and managing quality-related information in the context of user-centered requirements engineering. The goal of introducing quality in this context was to facilitate the involvement of non-IT stakeholders into the activities of discussing and negotiating the quality of a system under development.

The following research question has been formulated at that stage: "*How to support and thus better involve non-IT stakeholders into the requirements elicitation and analysis process so that they can address the expected system quality before experiencing the completed system*". To address this question, we analyzed and adapted a predesign approach (i.e., a technique for user-centric requirements modeling [MK98]) to exploit the following research ideas:

1. using aspect orientation to incorporate crosscutting concepts (which can be used to represent quality-related issues such as security or performance) into user centered requirements models for supporting separation of concerns in a problem domain; this research led to establishing the Aspectual Predesign Model [SK05];
2. exemplifying the aspectual predesign model to support modeling qualities of the prospective system; this research led to establishing the Quality-Aware Predesign Model [Ka08, SKM08] which allows for an explicit separation of functional and quality concerns;
3. further specializing the established quality-aware model to represent qualities of software services and/or business processes [Sh11];

The techniques developed at this stage suffered from the following shortcomings:

1. support of the conceptualization of quality-related information only; no guidelines for the integration of these conceptualizations into the software process involving business stakeholders, no means supporting such involvement;
2. no tool support provided, only few practical experience.

To overcome these limitations, we drove our research into the problem area of quality view harmonization and elaborated the initial problem statement for the QuASE project.

3 Quality view harmonization

In investigating the research area of harmonizing stakeholder views on quality, we started from the idea of a model-based involvement of business stakeholders into the software development activities. These initial activities concentrated on the conceptual foundations of the user's process of evaluating the qualities of a system under development; initially the scope was limited to service-oriented systems, but soon this restriction [SMK12a, SM12] was lifted.

Prior to making further conceptualizations and, eventually, proposing the conceptualization of the entire view harmonization process, we collected knowledge from representatives of our partner companies by means of quantitative techniques (questionnaires, structured and free-form interviews etc.) [SMK12b, SMK14].

Based on this collected knowledge we elaborated both initial [SMK13] and extended [SMK14] conceptualizations of the process of harmonizing stakeholder views on quality. We defined a harmonization process as the entirety of interactions between involved process sides targeting at a set of terminology means, quality views, and qualities that satisfies all these process sides [SMK14]. We addressed the following dimensions of quality view harmonization: (1) the quality of the system under development and the related aspects as well as their interrelations, (2) the knowledge about stakeholders, the differences in their perception of quality, and their amount of knowledge about the software quality, (3) the knowledge about the activities of the harmonization process and the capabilities of the process participants needed for performing these activities.

In [SMK14], we also distinguished three levels of defining the harmonization process: (1) *terminology harmonization*, where the stakeholders seek an agreement on a commonly understandable quality-related terminology, i.e., the language constructs used for expressing expectations and opinions on quality; (2) *view harmonization*, where the stakeholders seek an agreement on the sets of objects and the types of their qualities they are interested in, as well as the procedures of assessment, and (3) *quality harmonization*, where the stakeholders seek an agreement on the evaluation schemes and the particular qualities they are interested in.

The conceptualizations served as a basis for developing the initial QuASE proposal [SM13]. We built this proposal around two main ideas which are described below.

The first basic idea was to make the resulting QuASE system implemented as an ontology-driven solution where the ontology incorporates the knowledge about the quality harmonization domain, and serves as a basis for establishing the harmonization process. This ontology was planned to be established based on the domain knowledge collected from the company representatives by means of a standard ontology development approach like presented in [FGJ97]. The ontology was intended to be a sole source of harmonization, to be extendable by including other specialized ontologies such as measurement ontology [BVG06] etc. We initially did not plan to make the ontology

specific for a particular QuASE deployment configuration, in particular, to a specific target IT company. In addition, we did not expect to provide any means allowing the company representatives to modify this ontological knowledge beyond standard ontology development tools such as Protégé⁴.

In [SMK14], for standardization reasons we proposed to base the QuASE ontology on the Unified Foundational Ontology (UFO) [Gu05]. Consequently, the set of QuASE concepts forms a special-purpose ontology, which is derived from the UFO concepts.

The second basic idea was to not restrict our endeavor to the direct terminology harmonization support, e.g., implemented by means of mutually “translating” the communicated information's natural-language fragments stored in the project repositories. Rather, we decided to investigate the extension of our approach by means for exploiting quality-related communication experience represented in the repository in order to support decision making and predicting the behavior of the involved parties based on past quality-related communications.

This led to the objectives of the initial QuASE proposal as follows [SM13]:

1. Acquiring and formalizing domain knowledge about handling quality-related issues in the software process;
2. Collecting the raw information about such issues from the involved parties and converting it into operational knowledge (using available domain knowledge to ensure conversion correctness);
3. Using the collected knowledge for establishing a quality-related communication basis for the involved parties, supporting decision making, reuse of quality-related experience, and the prediction of the future quality-related requirements of the involved parties.

4 Ensuring flexibility and a proper scope

During the first year of the QuASE project another qualitative research was performed in cooperation with the representatives of the partner companies. By this set of interviews we pursued the following goals:

1. To reveal the evidence of the current solution's suitability as outlined in the proposal for the industrial settings, targeting mainly small and medium-sized enterprises; a special focus was put on the opinions about the suitability of the proposed ontology-based view harmonization approach;
2. To collect the opinions and experience about the exact scope of communicated information which the company representatives consider suitable for QuASE.

⁴ <http://protege.stanford.edu>

As a result of these interviews, the purpose-built ontology proved to be too limited for our purposes:

1. The knowledge about quality-related communications that can be collected from the company representatives varies from company to company; this knowledge, while being similarly structured, is different enough to justify being separately configurable for the particular deployment site;
2. The communicated information collected by the companies is stored, as a rule, in their project repositories (e.g. Jira databases); to benefit from the QuASE approach, such information has to be converted into knowledge, which is difficult to implement based on a special-purpose ontology as this ontology was not supposed to contain the instructions for such conversion;
3. The company representatives felt very uncomfortable with the idea of the direct modification of the ontology by knowledge suppliers using low-level ontology editing tools like Protégé.

These limitations led us to the elaboration of an extended set of QuASE use cases to be supported by the software solution [SML14], and to changing the ontology and knowledge base building approach:

1. The ontology building approach was changed in favor of generating the ontology based on a special-purpose model which is developed using a tool-supported metamodel-based visual domain-specific modeling language (*QuASE site DSL* [SMK15a]); this language is supposed to be suitable for use by knowledge suppliers and system configurators;
2. The knowledge base approach has been changed to a model-based one, too: the model was made to contain the mapping between the project repository and the modeling concepts which is then used to generate the individual knowledge base instances complying to the (generated) ontology concepts from the repository data [SM14, SMK15a].

In addition to the ontological matters, the second-stage qualitative research activities also revealed the challenges related to the scope of the information to be dealt with in the QuASE system.

The interviews favored the opinion that it is not productive to limit the scope of the system to harmonizing quality-related communications while targeting IT small-to-medium size enterprises. This is related to the fact that, in the projects typical for such companies, the subset of the quality-related communicated information stored in the project repository is limited as compared to the amount of the information related to functionality and other, non-quality-related factors. The representatives of the companies emphasized that the stakeholders not only refuse to discuss the quality of the system without experiencing it, but often prefer to delegate all quality-related communications to the IT people motivating that decision by lack of immediate interest to such issues and by trusting the developer's judgments at that stage.

To overcome this limitation, the QuASE solution was redesigned to implement flexible view harmonization support based on a knowledge-oriented representation of the communicated information collected in project repositories. The proposed extension was targeting the whole volume of communicated information with the possibility to filter it by topic – with the topics including quality of the prospective system and other aspects of software development.

The base set of the QuASE concepts underlying the QuASE site DSL [SMK15a] and the generated QuASE site ontology were modified to be generic enough to cover different categories of communicated information, the final set of concepts is depicted on Fig.1.

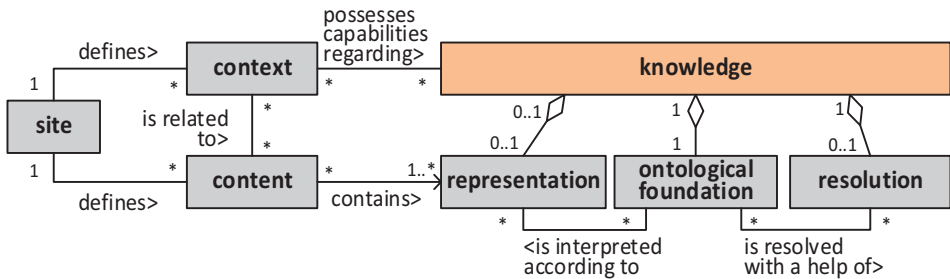


Fig.1. Generic QuASE concepts covering arbitrary communicated information [SMK15a]

This set includes the following concepts [SMK15a]: (1) *site*: owner of the given QuASE installation, e.g. a software provider; (2) *context*: units having particular views on communicated information e.g. projects, organizations, involved stakeholders; (3) *content*: units shaping communicated information e.g. issues/tickets; (4) *knowledge*: units encapsulating communicated knowledge that is a subject of harmonization.

Every QuASE knowledge unit is composed of: (1) *ontological foundation*: a reference to the conceptualization of the particular piece of knowledge through ontological means; (2) *representation*: the representation of the knowledge unit in a format that could be perceived by the communicating parties (e.g. plain text); representation units are also contained in content units; (3) *resolution means*: the means of resolving understandability conflicts related to the particular knowledge unit (e.g. explanations).

Context units possess *capabilities* to deal with knowledge units. In particular, the capabilities could refer to the ability of understanding a given knowledge unit at hand or explaining a knowledge unit using resolution means.

5 Current state of the QuASE research

Every QuASE knowledge unit is composed of: (1) *ontological foundation*: a reference to the conceptualization of the particular piece of knowledge through ontological means;

(2) *representation*: the representation of the knowledge unit in a format that could be perceived by the communicating parties (e.g. plain text); representation units are also contained in content units; (3) *resolution means*: the means of resolving understandability conflicts related to the particular knowledge unit (e.g. explanations).

The QuASE project has been successfully completed in 2015: a comprehensive tool support for view harmonization activities was implemented, which is currently available in a release quality to be used in real pioneer industrial settings; the subsequent release will target on normal professional usage. In this section we outline the current state of the solution: its architecture, the supported scenarios and support techniques, and the ability of the QuASE solution to address original QuASE research goals related to quality view harmonization.

5.1 QuASE system architecture

The current high-level architecture of the QuASE system is depicted in Fig.2. It includes the following components [SMK15a]:

1. A *QuASE site modeling tool* which supports (1) QuASE site DSL for describing the site-specific communication environment by means of the QuASE generic concepts; (2) a means of defining the site configuration; (3) and a means of specifying a mapping between the model and the particular project repository (e.g. Jira database).
2. A *QuASE ontology builder utility* transforming the models expressed in the QuASE site DSL into OWL2 representations of the QuASE site ontology, and the control information for mapping project repositories.
3. A *QuASE knowledge base builder utility* converting the data from project repositories into individuals corresponding to the QuASE site ontology based on the mapping specified by means of the QuASE modeling tool.
4. The interactive web-based *QuASE tool* implementing (1) the end-user support scenarios; (2) the acquisition of the information not found in project repositories from knowledge suppliers (e.g. user-supplied attribute values or the decision information).
5. A *QuASE knowledge base* which is a triple store containing OWL2 representation of site ontology (TBox) and knowledge base individuals (ABox).
6. An *IMS QuASE integration module* which allows using QuASE functionality from the user interface of issue management system such as Atlassian Jira.

A detailed description of the implementation architecture of the QuASE software solution has been published in [Sh15, Sh16b]; it includes server and client components, the security subsystem and the knowledge base support by means of Jena TDB triple store. It is extendable on the server side through web service-based API and on the client

side as web application based on AngularJS framework. To implement Jira integration, it includes the QuASE Jira plugin.

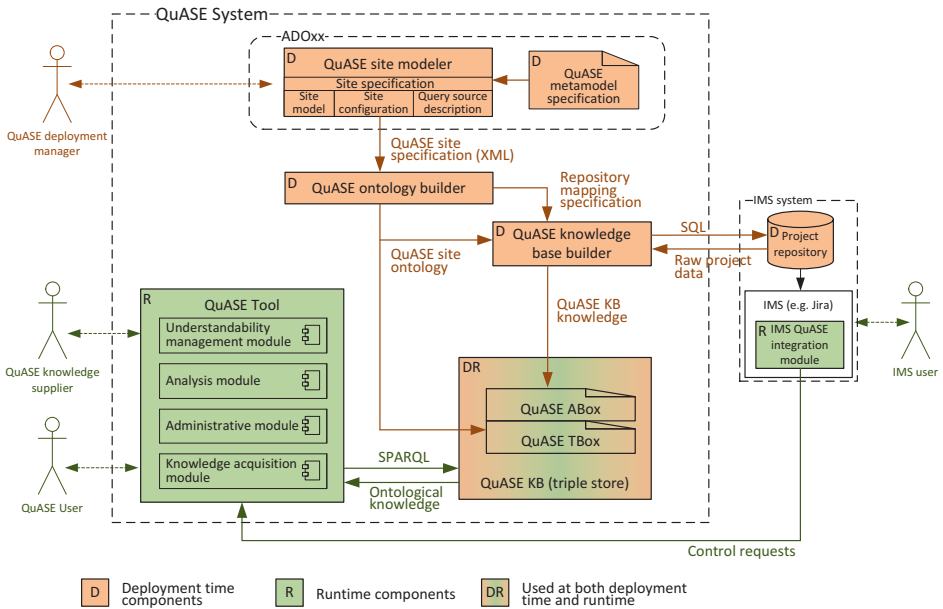


Fig.2. QuASE system architecture (extended version based on [SMK15a])

5.2 QuASE functionality

The complete set of QuASE tool usage scenarios is described in its user guide [Sh16a], it includes understandability management and analytical scenarios.

We established the scope of terminology harmonization support in QuASE in [SM14b] by providing the definition for *understandability management* activities. The implementation of the understandability management support in QuASE [Sh15] applies natural language processing techniques to localize the source of understandability conflicts in communicated information and to provide conflict resolution through terminology translation and issuing targeted explanations.

The current architecture and tool support for analytical activities in QuASE are described in [SMK15b]. The implementation of such support applies machine learning techniques to the set of QuASE knowledge base individuals: (1) similarity search for implementing information reuse; (2) regression analysis for predicting attribute values; (3) hybrid (partially supervised) learning for classification-based recommendation and decision support.

5.3 QuASE as a knowledge-based integrated solution

The advantage of the QuASE approach is that after extracting the data from a project repository, this data is immediately converted into knowledge available for understandability management and analysis. As the mapping is flexible by means of using a special-purpose DSL, this allows large amounts of existing data to be the subject of applying these techniques. The QuASE system can be seen as a bridge which connects end users, the data in project repositories and the (extendable) set of machine learning and natural language processing techniques. These techniques become applicable to the repository data automatically after the repository and communication environment are described by means of the QuASE site DSL.

5.4 Using QuASE for harmonizing quality views

While being a generic solution dealing with arbitrary communicated information, QuASE clearly is also able to address the original problem of harmonizing stakeholder views on quality. To allow this, it includes support for filtering the terminology used for understandability management by the communication domain; the QuASE knowledge base includes a set of concepts corresponding to such domains including the hierarchy of software quality characteristics. By filtering the ontological units based on one of such characteristics, the QuASE tool can be used to support quality view harmonization.

5.5 Tool deployment and usage status

As of June 2016, the QuASE tool has been installed at the sites of two partner companies, and it performed well for their repositories which included up to 26.000 issues. The resulting knowledge base contained up to 1.7 million axioms without significant performance problems. The main usage patterns for the tool are as follows:

1. The end users prefer accessing the system through the Jira integration interface which provides familiar means for arriving to the particular context or content unit;
2. The perceived usefulness of the understandability management capabilities of the tool depends on the quality of the terminological knowledge in the knowledge base, in particular, on the completeness of this knowledge with respect to the particular application domain (e.g. software quality).
3. The perceived usefulness of the analytical capabilities of the tool depends on the completeness of the set of metrics available for analysis. It is recommended to perform qualitative studies in cooperation with the company representatives to define the metrics influencing the development process to be included in the site model.

6 Related Work

A comprehensive review of the related work on quality view harmonization has been published in [SMK14], a review of the work similar in scope to the QuASE project – in [SMK15a].

7 Conclusions and the directions for future research

Based on the description of the evolution of the QuASE approach provided in this paper, we can distinguish the following major turning points in the evolution of the underlying ideas:

1. Advancing from quality conceptualization to the idea of harmonizing quality views;
2. Augmenting harmonization techniques with facilitating reuse of communicated information and supporting decisions based on this information;
3. Moving from harmonizing quality-related views to the generic idea of harmonizing communicated information;
4. Going from establishing a special-purpose ontology without end user support to a (meta-)model-based generic solution.

The QuASE approach is the result of a ten year's evolution of ideas on view harmonization, finally arriving to a comprehensive and flexible software solution which offers different techniques of harmonization support and makes them applicable to the communicated information of different nature stored in industrial project repositories.

In future, we see this approach applied to view harmonization problems occurring in other domains. Also we plan to enhance the understandability management support with more advanced terminology specification and translation techniques.

References

- [BVG06] Bertoa, M., Vallecillo, A., García, F.: An Ontology for Software Measurement. In *Ontologies for Software Engineering and Software Technology*, 175–196, Springer, 2006
- [FGJ97] Fernández, M., Gómez-Pérez, A., Juristo, N.: METHONTOLOGY: From Ontological Art towards Ontological Engineering. In *AAAI Spring Symp. Series*, 33–40, 1997
- [Gu05] Guizzardi, G.: *Ontological foundations for structural conceptual models*. University of Twente, 2005
- [Ka08] Kaschek, R., Kop, C., Shekhovtsov, V.A., Mayr, H.C.: *Towards Simulation-Based*

- Quality Requirements Elicitation: A Position Paper. In REFSQ 2008, Lecture Notes in Computer Science, Vol. 5025, 135-140, Springer, 2008
- [MK98] Mayr, H.C., Kop, C.: Conceptual Predesign - Bridging the Gap between Requirements and Conceptual Design. In ICRE '98, 90-100, IEEE CS Press, 1998
- [Sh11] Shekhovtsov, V.A., Kaschek, R., Kop, C., Mayr, H.C.: Relational service quality modeling. In Milanovic, N. (ed.): Non-Functional Properties in Service Oriented Architecture: Requirements, Models and Methods, 172-193, IGI Global, 2011
- [Sh15] Shekhovtsov, V.A., Mayr, H.C., Ianushkevych, S., Kucko, M., Lubenskyi, V., Strell, S.: Implementing tool support for effective stakeholder communication in software development – a project report. In Ausgewählte Beiträge zur Anwenderkonferenz für Softwarequalität Test und Innovation - ASQT 2014, books@ocg.at, Vol. 310, 45-58, Österreichische Computer Gesellschaft, 2015
- [Sh16a] Shekhovtsov, V.A. QuASE Tool User Guide v.1.3, http://quase-ainf.aau.at/documents/QuASE_User_Guide_1.3.pdf, accessed on 18.06.2016
- [Sh16b] Shekhovtsov, V.A. QuASE Architectural Description v.1.0, http://quase-ainf.aau.at/documents/QuASE_Architecture_1.0.pdf, accessed on 18.06.2016
- [SK05] Shekhovtsov, V.A., Kostanyan, A.: Aspectual predesign. In Information Systems Technology and its Applications - ISTA'2005, Lecture Notes in Informatics, Vol. P-63, 216-226, GI, 2005
- [SKM08] Shekhovtsov, V.A., Kop, C., Mayr, H.C.: Capturing the Semantics of Quality Requirements into an Intermediate Predesign Model. In SIGSAND-EUROPE'2008 Symposium, Lecture Notes in Informatics, Vol. P-129, 25-37, GI, 2008
- [SM12] Shekhovtsov, V.A., Mayr, H.C.: Let Stakeholders Define Quality: A Model-Based Approach. In Qualitätsmanagement und Vorgehensmodelle - 19. Workshop der GI-Fachgruppe Vorgehensmodelle, 101-110, Shaker Verlag GmbH, 2012
- [SM13] Shekhovtsov, V.A., Mayr, H.C.: Towards intelligent handling of quality related issues in software development – a project report. In Ausgewählte Beiträge zur Anwenderkonferenz für Softwarequalität Test und Innovation - ASQT'12, books@ocg.at, Vol. 295, 113-129, Österreichische Computer Gesellschaft, 2013
- [SM14a] Shekhovtsov, V.A., Mayr, H.C.: Managing Quality Related Information in Software Development Processes. In CAiSE-Forum-DC 2014, CEUR Workshop Proceedings, Vol. 1164, 73-80, CEUR-WS.org, 2014
- [SM14b] Shekhovtsov, V.A., Mayr, H.C.: Towards Managing Understandability of Quality-Related Information in Software Development Processes. In Murgante, B., Misra, S., Carlini, M., Torre, C., Nguyen, H.Q., Taniar, D., Apduhan, B., Gervasi, O. (eds.): ICCSA 2014, Proceedings, Part V., Lecture Notes in Computer Science, Vol. 8583, 572-585, Springer, 2014
- [SMK12a] Shekhovtsov, V.A., Mayr, H.C., Kop, C.: Stakeholder Involvement into Quality Definition and Evaluation for Service-Oriented Systems. In USER'12 Workshop at ICSE'12, 49-52, IEEE, 2012
- [SMK12b] Shekhovtsov, V.A., Mayr, H.C., Kop, C.: Acquiring Empirical Knowledge to Support Intelligent Analysis of Quality-Related Issues in Software Development. In Faria, J.P.,

Silva, A., Machado, R.J. (eds.): QUATIC 2012, 153-156, IEEE Press, 2012

- [SMK13] Shekhovtsov, V.A., Mayr, H.C., Kop, C.: Towards Conceptualizing Quality-Related Stakeholder Interactions in Software Development. In Mayr, H.C., Kop, C., Liddle, S., Ginige, A. (eds.): Information Systems: Methods, Models, and Applications, Lecture Notes in Business Information Processing, Vol. 137, 73-86, Springer, 2013
- [SMK14] Shekhovtsov, V.A., Mayr, H.C., Kop, C.: Harmonizing the Quality View of Stakeholders. In Mistrik, I., Bahsoon, R., Eeles, R., Roshandel, R., Stal, M. (eds.): Relating System Quality and Software Architecture, 41-73, Morgan-Kaufmann, 2014
- [SMK15a] Shekhovtsov, V.A., Mayr, H.C., Kop, C.: Facilitating Effective Stakeholder Communication in Software Development Processes. In Nurcan, S., Pimenidis, E. (eds.): Information Systems Engineering in Complex Environments, Lecture Notes in Business Information Processing, Vol. 204, 116-132, Springer International Publishing, 2015
- [SMK15b] Shekhovtsov, V.A., Mayr, H.C., Kucko, M.: Implementing tool support for analyzing stakeholder communications in software development. In Software Testing, Verification and Validation Workshops (ICSTW), 2015 IEEE Eighth International Conference on, 1-10, 2015
- [SML14] Shekhovtsov, V.A., Mayr, H.C., Lubenskyi, V.: QuASE: A Tool Supported Approach to Facilitating Quality-Related Communication in Software Development. In 9th International Conference on the Quality of Information and Communications Technology (QUATIC), 2014, 162-165, 2014

Software Qualität ist wie
Schönheit

Zusammenfassung: Software Qualität ist wie Schönheit

Sabine Wieland¹, Andreas Hartmann²

Entsprechend den von der GI formulierten „Grand Challenges der Informatik“ beschäftigt sich der Workshop mit einer aktuellen Herausforderung der Informatik - der Software Qualität. „Software ist heute allgegenwärtig und kommt im täglichen Leben überall zum Einsatz: in der Kommunikation, der Unterhaltungselektronik, im Haushalt, in der Medizintechnik und in sicherheitskritischen Bereichen.“ [Grand Challenges der Informatik, GI] Der Workshop diskutiert die Besonderheiten von Software, verschiedene Methoden und Werkzeuge die Qualität von Software zu messen und sichtbar zu machen sowie die Auswirkungen von Software Qualität auf die IT Sicherheit. Er adressiert somit zwei in besonderem Maße wichtige Themen der Informatik: Nachhaltigkeit und Sicherheit von Software. „Verlässlichkeit umfasst dabei nicht nur die Frage, ob Software ihre funktionalen Anforderungen erfüllt. Auch Datenschutz und schnelles Erkennen von Fehlfunktionen sind wichtige Aspekte. Für Entwickler etwa wäre die Garantie, dass von ihnen entwickelte Softwaresysteme sensible Daten schützen, ein Qualitätsmerkmal.“

Der Workshop hat zum Ziel, die Zusammenarbeit zur Verknüpfung der qualitätssichernden Maßnahmen zu fördern und voranzutreiben.

- Software Qualitätsmaße
- Methoden zur Software Qualitätsmessung
- Entwicklung von Qualitätssoftware
- Software Sicherheit als wesentliches Qualitätsmaß
- Software Qualität sichtbar machen
- Projektmanagement für Qualitätssoftware
- Nachhaltige Validierung von Software Qualität
- Anforderungen und Umgebungsbedingungen verschiedener Stakeholder

¹ Hochschule für Telekommunikation, Leipzig, wieland@hft-leipzig.de

² Hochschule für Telekommunikation, Leipzig, Andreas.Hartmann@hft-leipzig.de

Beiträge

Titel	Vortrag	Beitragsart
Beurteilung von Software Qualität stärken durch Ontologien und Wiederverwendung von Wissen,;	Vladimir A. Shekhovtsov,	wissenschaftlicher Beitrag
An Integrated Approach to Evaluation of Domain Modeling Methods and Tools for Improvement of Code Reusability in Software Development;	Mykola Tkachuk, Iryna Martinkus , Rustam Gamzayev, Andrii Tkachuk;	wissenschaftlicher Beitrag
Citrus – Integrationsframework mit Demonstration;	Christoph Deppisch;	Praxisbericht und Demonstration
Software Qualität ist wie Schönheit;	Sabine Wieland, Andreas Hartmann;	Erfahrungsbericht

Programmkomitee Vorsitzende

- Prof. Sabine Wieland, HfTL
- Prof. Andreas Hartmann, HfTL

Programmkomitee

- Dr. habil. Roman Englert, T-Systems
- Prof. Andreas Hartmann, HfTL
- Prof. Sebastian Schinzel, FH Münster
- Prof. Sabine Wieland, HfTL
- Prof. Wilhelm Rossak, Friedrich-Schiller-Universität Jena
- Frank Termer, Bitkom
- Aleksandra Sowa, Deutsche Telekom AG

Software Qualität ist wie Schönheit

Sabine Wieland¹, Andreas Hartmann²

Abstract: Globalisierung, Digitalisierung, Industrie 4.0, Internet of Things, Big Data, Cyber Kriminalität, SOA, Cloud Computing und die Energiewende bestimmen den Alltag in Deutschland. Software Qualität wird von jedem Nutzer unterschiedlich empfunden und lässt sich genauso schlecht messen wie Schönheit. Die Gemeinsamkeiten von Software Qualität und Schönheit werden in diesem Artikel vorgestellt und daraus Schlussfolgerungen für die Verbesserung der Software Qualität gezogen. Der Artikel geht außerdem auf Methoden und Regeln zur Entwicklung von Software mit hoher Qualität ein.

Keywords: Software Qualität, Methoden und Regeln für Software Qualität, Software Qualitätsmerkmale, komplexe Software Systeme, Qualitätsmaße

1 Die Herausforderung der Digitalisierung

Mit Industrie 4.0 ist die nächste Ebene der Digitalisierung unserer Gesellschaft erreicht. Jeder Lebensbereich ist von der Digitalisierung betroffen und daher implizit abhängig von der Qualität der eingesetzten Software. Vom intelligenten Haus bis zum Internet der Dinge wird unsere Gesellschaft durch Software beeinflusst und der Mensch hinterlässt mit jeder Aktivität eine digitale Spur – er wird sprichwörtlich zum gläsernen Mensch.

Vielen Unternehmen sind die Sicherheitslücken ihrer IT Systeme nicht bewusst. Dennoch sind ebendiese Sicherheitslücken in den Software Systemen unterschiedlich häufig enthalten. Es genügt schon eine einzige veröffentlichte Schwachstelle und das gesamte System ist angreifbar. Große Datenmengen (Stichwort Big Data) werden erfasst und weltweit übertragen. Die Daten können manipuliert und missbraucht werden, wodurch ein hoher wirtschaftlicher Schaden entstehen kann. Es geht daher nicht nur darum, die Übertragung der Daten abzusichern, sondern auch die Daten verarbeitende Software vor Angriffen zu schützen. Dies gelingt am besten, wenn die Software selbst sicher ist.

Komplexe Softwaresysteme bedienen sich an Komponenten aus Klassenbibliotheken von Drittanbietern bzw. nutzen sie für das Design und die Entwicklung Services, Patterns und Frameworks. Softwaresysteme entstehen auch unter Verwendung von Open-Source-Komponenten. Diese Wiederverwendung hat viele Vorteile für die Entwicklung von komplexen Software Systemen. Bereits mehrfache genutzte und getestete Softwarebausteine können wiederverwendet und zu neuen Systemen

¹ Hochschule für Telekommunikation Leipzig, Gustav-Freytag-Str. 43-45, 04277 Leipzig, wieland@hftl.de

² Hochschule für Telekommunikation Leipzig, Gustav-Freytag-Str. 43-45, 04277 Leipzig, hartmann@hftl.de

kombiniert werden. Hier herrscht eine große Freizügigkeit, die durch verschiedene Lizenzmodelle unterstützt wird. Dabei beachten die Nutzer dieser Softwarekomponenten nicht die eingeschränkten Haftungsbestimmungen der Lizenzmodelle. Unter dem Gesichtspunkt der Haftung muss jede wiederverwendete Softwarekomponente im Zielsystem auf mögliche Fehler und Schwachstellen geprüft werden.

Die professionelle Cyberkriminalität nimmt stetig zu. Die Netzwerkangriffe sind um das Fünffache, von 7 Millionen im Mai auf 36 Millionen im August 2015 gestiegen [BSI14]. Dabei wird das Ziel, kritische Infrastrukturen zu treffen und deren Betrieb zu beeinträchtigen und zu stören, mit professionellen Methoden und hoher krimineller Energie verfolgt.

Im Jahr 2014 wurden laut HPI [Ho15] weltweit fast 6500 Sicherheitsschwachstellen in Software gemeldet – davon alleine 200 Sicherheitslücken für OpenSSL. Das ist positiv, denn jede gemeldete Sicherheitslücke kann und muss behoben werden. Ein zeitnahes Release mit der Fehlerbehebung bei OpenSSL zeigt den professionellen Umgang mit erkannten Sicherheitslücken. Software von Microsoft, Apple, Google, Mozilla, Oracle, HP, Adobe etc. reiht sich mit kritischen Sicherheitslücken in die Statistik [BSI14] ein. Unberücksichtigt bleiben Sicherheitslücken, die zwar bekannt sind, jedoch nicht gemeldet werden.

2 Die aktuelle Software Qualität

Entgegen diesem Hintergrund wird Software nur zu 30% getestet, uralte Software verwendet die nicht mehr gepflegt wird und Software ausgeliefert, bei der einfachste Entwicklungsregeln nicht berücksichtigt wurden.

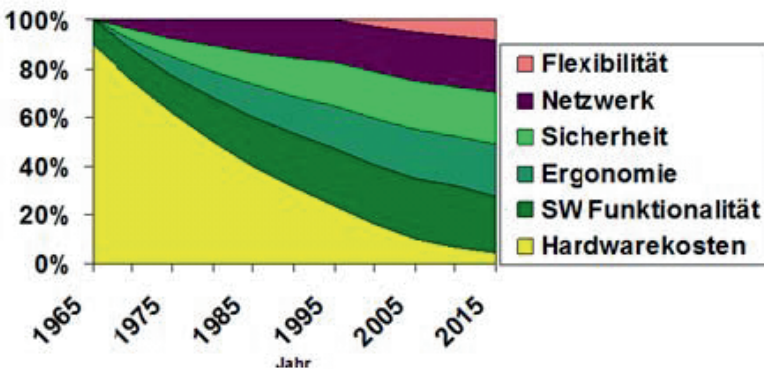


Abb. 1: Entwicklung der Software Komplexität

Allein diese Fakten zeigen, dass das Qualitätsbewusstsein für Software verbessert werden muss. Außerdem müssen geeignete Rahmenbedingungen für die Herstellung von

sicherer Software in allen Bereichen geschaffen werden.

2.1 Software Qualität ist Geschmackssache

Software Qualität wird sehr subjektiv bewertet. Jeder Nutzer stellt andere Anforderungen an die von ihm genutzte Software. Diese Anforderungen werden geprägt durch individuelle Erfahrungen, bereits genutzte Software und durch ein ganz persönliches Nutzerverhalten. Entsprechend der 80/20 Regel verwenden 80% der Nutzer 20% der Funktionalität, 20% der Nutzer ca. 80% der Funktionalität der Software. Nur sehr selten wird die volle Funktionalität von Software genutzt.



Abb. 2: Schönheit [Fo16, Wi16, Ac16]

Verschiedene Stakeholder – meist Gruppen von Interessenten – bestimmen die Qualitätsanforderungen, die an Software gestellt werden. Neben dem Nutzer werden der Auftraggeber und der Entwickler bei jedem Projekt mit berücksichtigt. Das spiegelt sich z.B. auch bei den Projektmanagementrollen im Vorgehen Scrum, dem Scrum Master, dem Product Owner und dem Entwicklerteam, wieder. Andere Stakeholder dagegen werden oft vergessen oder zu spät berücksichtigt. Dazu zählen der Systemadministrator, der Betriebsrat, der Datenschützer oder der Gesetzgeber. Eine rechtzeitige Integration aller Stakeholder in den Entwicklungsprozess ermöglicht die Entwicklung von sicherer und guter Software. Sobald einige Stakeholder nicht berücksichtigt werden, leidet das Immunsystem der Software darunter, vergleichbar mit Bild rechts in Abbildung 2. Ein schwaches Immunsystem ist bei Software an einem erhöhten Testaufwand, erhöhten Wartungsbedarf, vielen Sicherheitsschwachstellen und mangelnder Dokumentation zu erkennen.

Unser Empfinden sagt uns - nur ein gesundes Gesicht ist ein schönes Gesicht. Gesundheit muss jedoch auch gepflegt werden, denn ohne Pflege und gesunde Ernährung kann die Gesundheit nicht erhalten bleiben. Auch hier findet sich eine Parallele zur Software. Software *altert* über die veränderlichen Anforderungen (Changes) und durch Veränderungen in der Umgebung. Erfolgt keine Wartung und Anpassung an die veränderten Bedingungen *veraltet* die Software und wird schließlich unbrauchbar.

Software Qualität ist prozess- und produktbezogen. Eine MS Access Datenbank kann prima im privaten Gebrauch und auch noch für eine Handbibliothek mit 500 Büchern und einem Nutzer funktionieren. MS Access ist aber mit einer Anwendung für mehr als 100 Mitarbeiter, die gelegentlich Einträge vornehmen, überfordert. Hier bestimmt der Prozess, in dem die Software zum Einsatz kommt, die Anforderungen, die erfüllt werden müssen. Zum Beispiel steigen die Sicherheitsanforderungen, wenn personenbezogene Daten verarbeitet werden. Eine flexible Skalierung bei wechselnden Nutzerzahlen wird heute von fast jeder Software verlangt. Auch Echtzeitanforderungen variieren je nach Geschäftsprozess.

Datenbanksysteme, Editoren, Betriebssysteme, Entwicklungssysteme, Data Warehouse und Firmware wurden für ein bestimmtes Anwendungsgebiet entwickelt. Für jedes Anwendungsgebiet gibt es verschiedene Softwareprodukte - bei den Datenbanken z.B. Informix, Oracle, MySQL, Access etc. Jedes dieser Produkte weist Besonderheiten auf, die es einmalig machen. Als Beispiel seien zwei Editoren genannt, die sehr verschieden sind und jeder für sich Vor- und Nachteile ausweist. Der zu Unix gehörende vi ist schlank, robust und sehr leistungsfähig; besitzt aber nur eine rudimentäre Bedienerführung. MS Word dagegen enthält viele verschiedene Funktionen und eine ansprechende Oberfläche, ist jedoch fehleranfällig und vergleichsweise weniger robust.

Da Software immateriell ist, kann sie (theoretisch) bei Benutzung nicht verschleifen. Ihr fehlen wie die physikalischen Grenzen auch mechanische Teile, die durch Reibung verschleifen können. Software verschleißt nur durch Programmierfehler. Zum Beispiel wenn Speicherbereiche nach Nutzung nicht gelöscht und freigegeben werden, wenn ungewollte Wechselbeziehungen zwischen Programmteilen existieren.

Passt eine Software gut in den zu unterstützenden Geschäftsprozess entstehen neue Anforderungen beim Anwender, da angrenzende Geschäftsprozesse ebenfalls bearbeitet werden, oder noch eine Statistik fehlt. Auch während dem Entwicklungsprozess verändert sich die Umwelt, in der die Software zum Einsatz kommt. Ein Betriebssystem Update wird eingespielt, eine Markterweiterung steht an und damit wird eine weitere Sprache benötigt etc. Jeden Monat ändern sich auf diese Art 5% der Anforderungen. Das macht die Software Entwicklung zu einem komplexen, zeitkritischen Prozess.

2.2 Qualitätsmerkmale nach ISO Standard

Software Qualität ist zunächst in der ISO/IEC Norm 9126 [ISO11] mit 6 Qualitätsmerkmalen definiert, siehe Abbildung 3, wobei diese Norm in den Standard ISO/IEC 25000 überführt wurde. Die Definition berücksichtigt auch, dass Software immateriell ist. Software wird wie ein Kunstwerk entwickelt. Jede Software wird als Einzelstück entwickelt. Hier gibt es viele Parallelen zur Baubranche. Jedes Gebäude ist auf Grund seiner Lage, Funktion und Ausführung einzigartig. Genormte Teile vereinfachen die Konstruktion der Gebäude. Auch für das Software Engineering existieren Standards und „genormte Teile“, wie Schnittstellendefinitionen, Klassenbibliotheken etc. Das bedeutet, dass alle Software Services / Module / Klassen /

Methoden etc. nur mittels Nachrichten und Funktionsparameter kommunizieren. Sobald alle Variablen eines Systems private sind und nur von den klasseneigenen Methoden verändert werden können, steigt nicht nur die Wiederverwendbarkeit dieser Services, sondern auch deren Software Qualität und Sicherheit!

Qualitätsmerkmale von Softwaresystemen (ISO 9126)

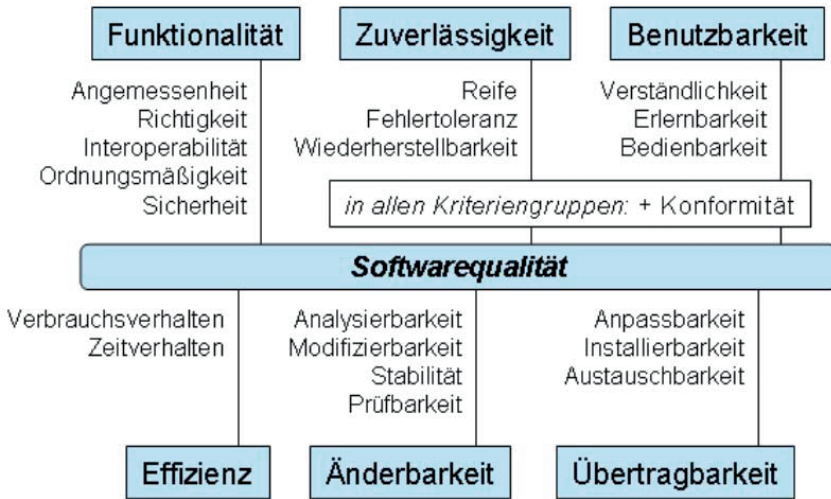


Abb. 3: ISO/IEC Standard 9126 [ISO11]

Es wird zwischen inneren und äußeren Qualitätsmerkmale unterschieden. Die äußeren Qualitätsmerkmale entsprechen den im ISO/IEC Standard 9126 beschriebenen Merkmalen. Sie werden auch vom Nutzer wahrgenommen und bestimmen die Nutzerakzeptanz. Die inneren Qualitätsmerkmale beschreiben die innere Struktur der Software und definieren die Fehleranfälligkeit, Robustheit und Änderbarkeit der Software. Zum Beispiel können objektorientierte Strukturen mit den Metriken nach Chidamber und Kremerer [CK91] gemessen werden. Mit den Metriken WMC (Weighted Methods per Class), RFC (Response for a Class), DIT (Depth of Inheritance Tree), NOC (Number of Children), etc. kann die Komplexität und damit die Fehleranfälligkeit eines Software Systems bestimmt werden. Jede Metrik muss dabei folgende Kriterien [Ba08] erfüllen:

- Objektivität, keine subjektiven Einflüsse sind möglich,
- Zuverlässigkeit, liefert bei Wiederholung das gleiche Ergebnis,
- Validität, lässt eindeutige Rückschlüsse auf die gesuchte Kenngröße zu,
- Normierung, nutzt eine Vergleichsskala,

- Vergleichbarkeit, kann mit anderen Maßen in Relation gesetzt werden,
- Ökonomie, die Messung ist mit geringem Aufwand möglich,
- Nützlichkeit, die Aussage der Messung erfüllt praktische Bedürfnisse.

Die im Standard beschriebenen Teilmerkmale können durch Qualitätsindikatoren bzw. Metriken mess- und bewertbar gemacht werden, wobei quantifizierbare Qualitätsindikatoren mit Hilfe von Qualitätsmaßen gemessen werden. Die im Standard beschriebenen Qualitätsmerkmale werden vom Kunden bzw. Nutzer in den NFA (Nichtfunktionalen Anforderungen) beschrieben. Die Definition von NFA fällt dem Kunden noch schwerer als die Bestimmung von funktionalen Anforderungen (FA). Der Standard ISO/IEC 9126 gibt mit seiner klaren Definition der Qualitätsmerkmale wichtige Hinweise für die Definition von NFA und FA. Checklisten, Schablonen und Metriken helfen bei der Definition der Anforderungen. Nur klar definierte Anforderungen bilden die Grundlage für eine nützliche Qualitätsmessung.

2.3 Spagat zwischen Wissenschaft und Kunst

Soweit die Theorie zur Entwicklung von guter Software. Heutige Software Systeme sind mit Hundertwasserhäusern vergleichbar. Es wird absichtlich von den Standards abgewichen, um die Genialität des Entwicklers darzustellen. Besonders Apps – Software für mobile Geräte – werden von Quereinsteigern entwickelt, die nur wenige Methoden und Regeln der Software Entwicklung kennen und keine entsprechende Ausbildung haben. Ein Beispiel dafür ist das Crowdfunding, bei dem der Betatester durch eine Vielzahl von Nutzern durchgeführt wird. Diese Tests erfolgen meist ohne Testplanung, definierten Testfällen und ausreichender Dokumentation. Das Crowdfunding hat den Vorteil, dass eine Vielzahl von verschiedenen Kombinationen aus Hardware, Betriebssystem, Versionen, Providern etc. getestet werden können. Die Testvielfalt entspricht damit auch den in den Zielgruppen genutzten Konfigurationen. Ein Crowdfunder ist besser als gar kein Test.

Man nehme ein wirklich flexibles, frei orchestrierbares SOA-System, bei dem je nach Bedarf Services ausgetauscht werden. Wie soll die Funktionalität dieses Systems getestet werden, wenn zwischen den einzelnen Services nicht dokumentierte Abhängigkeiten existieren? Die Variantenvielfalt steigt ins Unendliche und dann werden tatsächlich nur 30% des Systems getestet. Software Entwickler sind sehr intelligente, kreative Designer von neuen Systemen, die neben der Funktionalität auch immer das Ego des Entwicklers erfüllen sollen. Künstler – also auch Software Entwickler – halten sich nicht gern an Gebote und Regeln. Diese schränken die Möglichkeiten und Kreativität ein. Daraus entsteht der Spagat zwischen der Anwendung wissenschaftlicher Erkenntnisse und der kreativen Entwicklung von neuem.



Abb. 4: Hundertwasserhaus in Magdeburg [Wi16b]

Der Widerspruch ist auf die besondere Eigenschaft der Software zurückzuführen – Immaterialität. Jede beliebige Software – auch sogenannte Standardsoftware wird nur einmal entwickelt. Danach kann sie prinzipiell ohne Veränderung unendlich oft kopiert werden. Das bedeutet im Umkehrschluss – Software ist immer einmalig und jede Software wird neu durch einen kreativen Prozess entwickelt. Wenn tatsächlich heute jede Software von Grund auf neu entwickelt wird, dann sind Technologien wie SOA und Cloud Computing nicht realisierbar. Denn SOA als auch Cloud Computing basieren auf lose gekoppelten Software Services, die nur mittels Nachrichten kommunizieren.

Wie jede kreative Tätigkeit benötigt auch ein Software Entwicklungsprojekt eine Pause und die Besinnung auf das große Ganze. Das Whisky-Syndrom steht im Zusammenhang mit dem Zeitfaktor bei der Entwicklung von Software-Projekten. Nach dem Motto „Zeit ist Geld“ wird versucht, die Mitarbeiter eines Software-Projektes zum effektiven Arbeiten anzuhalten. Dabei ist die Frage *Was machen die Mitarbeiter eigentlich den ganzen Tag?* sicher berechtigt. Dass es nicht ausreicht, Befehle aneinander zu reihen, sondern dass diese Befehle auch die Aufgabe lösen müssen, kann dabei schnell in Vergessenheit geraten.

Schnell setzt sich beim Management der Irrglaube fest, die Entwickler sind nur produktiv, wenn sie Befehle schreiben, also *codieren*. Damit werden die Erfahrungen von vielen Projekten missachtet die besagen, dass die Implementierung den geringsten Zeitaufwand bei der Software-Entwicklung erfordert, siehe Abb. 5. Einen Befehl zu schreiben dauert keine Minute, den richtigen Befehl zu finden kann insgesamt (mit Anforderungsanalyse und Test) über eine Stunde dauern. Der Name Whisky-Syndrom wurde von der Frage *Why isn't Sam coding yet?* abgeleitet und macht mit einem Augenzwinkern auf das Problem aufmerksam. Dass jeder kreativen Tätigkeit eine Phase des Überlegens, Nachdenkens, aber auch der Recherche und Untersuchung vorausgehen

muss, sollte nicht vergessen werden.

■ Analyse ■ Design ■ Code ■ Test

Traditionell (-1974) Strukturiert (-1982) mit CASE (ab 1989)

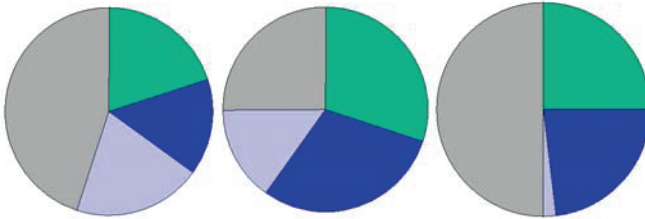


Abb. 5: Zeitaufwand für die Phasen Analyse, Design, Code und Test im SWE Prozess

Dass einfache Gemüter diesen Teil der kreativen Tätigkeit gern vernachlässigen, ist nicht neu. Auch Leonardo da Vinci hatte mit seinen Auftraggebern so seine liebe Not, die ihn zu dem Ausspruch zwang, „dass erhabene Geister bisweilen am meisten schaffen, wenn sie am wenigsten arbeiten“. Nachdenken kann schnell als „In die Luft starren“ – also Nichtstun gedeutet werden, da es nicht durch eine entsprechende Mimik deutlich wird, ganz im Gegenteil. Viele kreative Köpfe machen in den schöpferischsten Phasen ein ausgesprochen dummes Gesicht.

In diesem Ausspruch von Leonardo da Vinci steckt auch, das für eine hervorragende kreative Leistung auch Phasen der Entspannung notwendig sind. Ein Entwickler, der ständig unter Zeitdruck steht, wird in seiner Kreativität und Produktivität schnell nachlassen. Dieses Phänomen wird auch in den sogenannten Startup- Unternehmen beobachtet. Nach mehreren Monaten, manchmal auch Jahren intensivster Arbeit ohne Pause und mit reichlichen Überstunden sind auch die kreativsten Köpfe ausgepowert und brauchen Erholung.

Der nächste Aspekt, der mit zum Whisky-Syndrom gehört, ist die Konzentration auf ein Fachgebiet. Der Blick über den Tellerrand kann zu neuen Erkenntnissen führen. Ein einseitig nur mit Rechnern und dessen Problemen Beschäftigter nutzt keine Erkenntnisse aus anderen Fachgebieten. In der Informatik hat diese Isolation der Fachkräfte und Experten schon verheerende Ausmaße angenommen. Hier kennt der Datenbank-Spezialist die neuesten Erkenntnisse auf dem Gebiet der Betriebssysteme nicht und der Netzwerk-Spezialist hat keine Ahnung, wie ein Datenbank-Management funktioniert. Erst recht werden Erkenntnisse der Randgebiete, wie die Künstliche Intelligenz, übersehen. In der Nutzung möglicher Synergien liegt ein sehr großes Potential, das beachtet werden sollte.

3 Methoden und Regeln für Software Qualität

Das Teufelsquadrat beschreibt die Wechselwirkungen zwischen den Größen Qualität, Quantität, Projektdauer, Projektkosten und Produktivität. Die Größen Qualität, Quantität, Kosten und Zeit stehen in sehr enger Beziehung zueinander. Eine Veränderung einer Größe wirkt sich entsprechend auf die anderen aus.

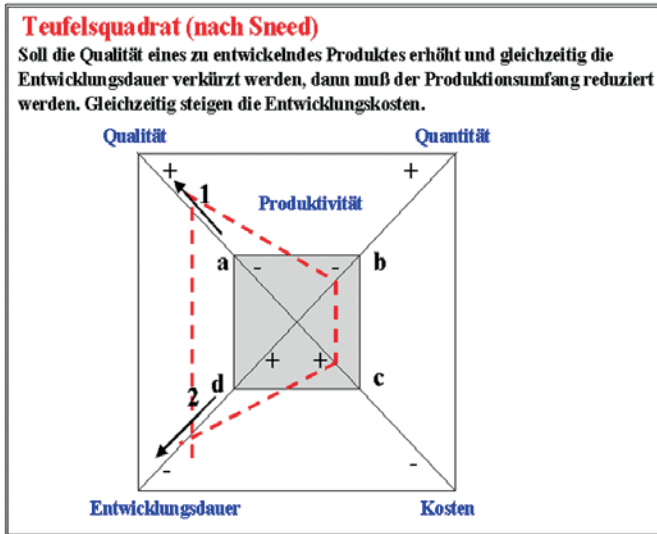


Abb. 6: Teufelsquadrat nach Sneed [Ba09]

Hohe Qualitätsanforderungen können nur mit einem ausreichenden Budget und ausreichender Zeit realisiert werden. Die Quantität wird durch die funktionalen Anforderungen definiert und stellt den Umfang des Software Systems dar. Für die meisten Software-Projekte sind sowohl die Qualitätsanforderungen als auch der Lieferzeitpunkt vom Auftraggeber vorgegeben. Eine entsprechende Kalkulation des Preises für das Softwarepaket (Funktionsumfang und geforderte Qualität) und die anschließende Diskussion mit dem Auftraggeber ist die erste Hürde, die das Projektmanagement bewältigen muss. Ein Missverhältnis zwischen den Anforderungen und den für die Entwicklung zur Verfügung stehenden Zeitraum kann das Softwareprojekt zum Scheitern verurteilen. In einem Softwareprojekt ist es nicht möglich, mit doppelt so vielen Entwicklern die Entwicklungszeit zu halbieren, da der höhere Kommunikationsbedarf zwischen den Entwicklern zu einer Verzögerung führt. „Ein verspätetes Softwareprojekt wird durch die Hinzunahme eines zusätzlichen Entwicklers weiter verzögert.“ [Ba09] Dieses Phänomen wurde in vielen Softwareprojekten beobachtet und vielfach beschrieben. Der neue Entwickler muss mit dem Projekt vertraut gemacht werden, ehe er bei der Entwicklung mithelfen kann. In einer späten Entwicklungsphase ist das mit hohem Aufwand verbunden. Auf den Fortschritt eines

Softwareprojektes wirken sich neue Qualitätsanforderungen und Budgetkürzungen negativ aus.

3.1 Miss es oder vergiss!

Jede Anforderung, die in das Pflichtenheft aufgenommen wird, muss nachweisbar sein! Es muss eine klare Beschreibung vorliegen, wie diese Anforderung nachgewiesen / getestet werden kann. Ist die Anforderung nicht nachweisbar, soll sie nicht in das Pflichtenheft aufgenommen werden. Denn kann die Anforderung nicht nachgewiesen werden, braucht sie auch nicht realisiert werden. **Miss es oder vergiss es!** Um das Formulieren von Anforderungen zu vereinfachen, werden Anforderungsschablonen und UML-Diagramme oder die BPMN Notation verwendet. Je klarer die Anforderung formuliert ist, desto einfacher ist deren Nachweis in einem Test. „Die Anwendung muss performant sein“ oder „Die Anforderung muss beim Datenbankzugriff eine Antwortzeit unter 1 s erfüllen.“ Damit beginnt die Testphase bereits zur Zeit der Analyse, siehe Abb. 7. Während der Testplanung werden für jede Anforderung (Anforderungsgruppe) die geeigneten Testverfahren / Testfälle definiert und beschrieben. Auf dieser Grundlage werden in der Design Phase die Testumgebungen entwickelt und die Testdaten erstellt. Zur Implementierung folgt der Modultest. Vor dem Betrieb der Software erfolgt der Systemtest und während der Wartungsphase erfolgen die Wiederholungstests.

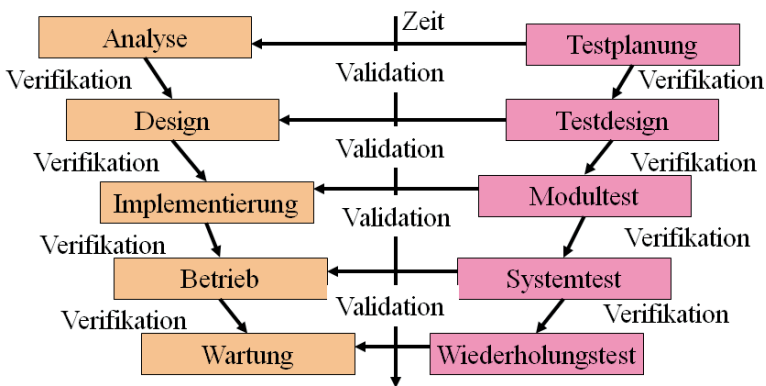


Abb. 7: angepasstes V-Modell

Codezeilen, die nicht getestet wurden sind „easter eggs“ und müssen gelöscht werden, denn diese nicht getesteten Codezeilen stellen Sicherheitslücken dar. Programmelemente die beim Testen der geforderten Funktionen und NFA nicht aufgerufen werden (dead code) müssen gelöscht werden, da sie ein erhöhtes Sicherheitsrisiko darstellen.

3.2 Technische Schulden

Ausprägungsart	Beschreibung	gesund / ungesund
Quelltextschulden	hohe Quelltextkomplexität, Nicht-Einhaltung des festgelegten Programmierstils	ungesund
Architekturschulden	Modularisierbarer Quelltext	ungesund
Dokumentationsschulden	Unverständliche Kommentare, defizitäre Informationsstruktur	ungesund
Testschulden	Nicht getesteter Quelltext (manuelle und Regressionstests)	gesund / ungesund
Schulden durch eine technologische Lücke	Nutzung veralteter Technologien	gesund / ungesund
Sicherheitsschulden	Nicht-Einhaltung der Sicherheitsrichtlinien	ungesund

Abb. 8: technische Schulden [Ba15]

Sobald der Blick primär auf den Liefertermin gerichtet ist und im Projekt notwendige Arbeiten aus Zeitmangel nicht realisiert werden können entstehen sogenannte „technische Schulden“ [KNO12], siehe Abb. 8. Wie bei der Kreditaufnahme müssen auch auf die technischen Schulden Zinsen gezahlt werden. Ein „Point of no return“ wird erreicht, wenn die zu erbringenden Zinsen höher als der aufgenommene Kredit sind. Durch die Dynamik im Software Entwicklungsprozess wachsen die Zinsen von technischen Schulden rasant. Die technischen Schulden müssen ausgeglichen werden, also die nicht realisierten Aufgaben müssen später nachgeholt werden. Erfolgt kein Ausgleich der technischen Schulden, wird die Wartung am System erschwert und die Qualität verschlechtert sich zunehmend (vgl. *Alterung*).

3.3 Das Geheimnisprinzip

Das Geheimnisprinzip fordert die Kapselung der Daten im Modul. Das ist u.a. einfach durch die Definition von privaten Attributen zu realisieren. Die Kopplung zwischen Modulen wird durch die Schnittstellen definiert. Je einfacher die Schnittstelle (Übergabe einfacher Datentypen per Nachricht) desto loser ist die Kopplung zwischen den Modulen. Solange einfache Datentypen (**Datenkopplung**) übergeben werden, ist das Software System gesund. Sobald Datenstrukturen übergeben werden (**Datenstrukturkopplung**) muss die Schnittstelle überprüft werden. Werden wirklich alle Datenelemente der Struktur vom Modul verwendet? Die Datenstruktur muss bei einer Weiterentwicklung oder Verwendung in einem anderen System simuliert werden das schränkt die Wiederverwendbarkeit ein.

Kopplung		wart- bar	wieder- ver- wendbar	Bindung	
Daten	lose	++	++	funktional	fest
Datenstruktur	lose	+	+	funktional oder sequentiell	fest
Kontroll	fest	-	-	kommunizierend oder problembezogen	lose
Hybrid	fest	--	--	problembezogen oder programmstrukturell	lose
global	fest	--	--	zeitlich	lose
Inhalt	fest	--	--	zufällig	lose

Abb. 9: Zusammenhang zwischen Kopplung und Bindung [My74]

Eine starke Kopplung verursacht eine lose Bindung / sehr geringen Zusammenhalt der Module [My74]. Eine starke Kopplung und lose Bindung in den Modulen eines Softwaresystems beruht meist auf Missachtung des Geheimnisprinzips und verursacht eine gegenseitige Beeinflussung der Module. Nach einigen wenigen Erweiterungen und Veränderungen durch Wartung treten nicht nachvollziehbare, kritische Runtime Fehler auf, die nur mit sehr hohem Aufwand gefunden und beseitigt werden können. Das Softwaresystem ist nicht mehr wartungsfähig. Auch die einzelnen Module / Services des Softwaresystems sind für eine Wiederverwendung in einem neuen Software System nicht geeignet, da sonst der Virus der gegenseitigen Beeinflussung in das neue System übertragen wird.

Das Geheimnisprinzip ist der Schlüssel für die Modularisierung da es Voraussetzung für die Abgeschlossenheit der Module und deren gegenseitigen Nichtbeeinflussung ist. Mit einem einfachen Software Qualitätsmesser können die öffentlichen Variablen in einem Softwaresystem ermittelt und mit der Gesamtanzahl der Variablen / Attribute verglichen werden. Viele Open-Source-Projekte weisen hier gravierende Sicherheitslücken auf, denn jedes öffentliche Attribut kann für einen Hackerangriff genutzt werden.

4 Fazit

Das tolle an Software ist: sie verschleißt nicht bei Benutzung – sofern sie frei von technischen Schulden ist! Hier ist sie besonders robust und widerstandsfähig! Bei Software gibt es (per Definition) keine Verschleißteile, die regelmäßig kontrolliert (jährliche Durchsicht) und ausgetauscht werden müssen! Software altert jedoch durch die Veränderungen in der Umwelt, durch die neuen Wünsche der Anwender, durch neue Gesetze und neue Marktsituationen.

Zusammenfassend können wir feststellen: Software Entwicklung mit hohem Qualitätsanspruch ist schwierig – um nicht zu sagen unmöglich!

1. Der Kunde weiß nicht was er will und kann es auch nicht klar ausdrücken.
2. Es gibt etwas über 20 Stakeholder, die die Qualität der Software definieren.
3. Die Anforderungen an die Software ändern sich (5% pro Monat).
4. Software ist immateriell und hat keine physikalischen Grenzen.
5. Projektfortschritt kann nur mittelbar über die Dokumentation gemessen werden.
6. Software Qualität ist prozess- und produktbezogen.
7. Software steht in enger Wechselbeziehung zu ihrer Umgebung (Prozesse, Hard- und Software).
8. Softwareprojekte sind einmalig – jede Software wird nur einmal entwickelt.
9. Software Entwicklung ist Kunst – ein kreativer Prozess.
10. Das Management unterschätzt den kreativen Prozess und unterliegt dem Whiscy-Syndrom.
11. Das Team ist nicht optimal zusammengestellt – es fehlen notwendige Skills
12. Das Geheimnisprinzip wird seit Jahren missachtet!

Wir erschaffen ein Monster! Ein weltumspannendes, das gesamte Leben und die Gesellschaft beherrschendes, von jedem manipulierbares Monster. Dieses Monster können wir auch nicht einfach abschalten, denn dann geht überall das Licht aus... Es gibt nur eine Lösung: Software unter Einhaltung der Methoden und Regeln des Software Engineering entwickeln und schrittweise alle andere Software ablösen. Hierzu sind SOA und Cloud Computing der richtige Weg, denn diese Technologien funktionieren nur mit lose gekoppelten Services!

Literaturverzeichnis

- [Ac16] Acne Vulgaris, Cysic acne on the face, wikipedia, <http://history.amedd.army.mil/booksdocs/wwii/internalmedicinevolIII/chapter20figure85.jpg>, Stand: 25.6.2016
- [Ba15] Bauer, M.; Ermittlung von technischen Schulden unter Verwendung von Metriken in einem agilen Software-Entwicklungsprojekt eines Provisionsabrechnungssystems , Bachelorarbeit, Hochschule für Telekommunikation Leipzig, 2015
- [Ba08] Balzert, H.; Lehrbuch der Software Technik – Softwaremanagement, Spektrum Akademischer Verlag; Auflage: 2. Aufl. 2009
- [Ba09] Balzert, H.; Lehrbuch der Software Technik – Basiskonzepte und Requirements Engineering, Spektrum Akademischer Verlag; Auflage: 3. Aufl. 2009
- [Bo11] Boehm, B.: in Sophist White Paper 2011
- [BSI14] BSI Lagebericht 2014: https://www.allianz-fuer-cybersicherheit.de/SharedDocs/Downloads/DE/BSI/Publikationen/Lageberichte/Lagebericht2014.pdf?__blob=publicationFile

- [CK91] CHIDAMBER, Shyam R. ; KEMERER, Chris F.: *A metrics suite for object oriented design*. Rev. July 1993. Cambridge, Mass. : M.I.T. Center for Information Systems Research, 1993 (Sloan WP 3524-93)
- [EM03] Andreas Essigkrug, Thomas Mey: Rational Unified Process kompakt. Spektrum Akademischer Verlag, Heidelberg u. a. 2003, ISBN 3-8274-1440-7.
- [Fo16] iStockfoto; http://www.fitforfun.de/beauty-wellness/haut-haare/huebsch-schoenheit-ist-messbar-aid_8673.html;
- [Ho15] Hofmann, K.: IT Business, 18.3.2015: CeBIT 2015: Anwendungen mit den meisten Sicherheitslücken
- [ISO11] ISO-Store ISO/IEC 9126-1:2001[1] This standard has been revised by: ISO/IEC 25010:2011
- [KNO12] Philippe Kruchten, Robert L. Nord und Ipek Ozkaya. Technical Debt: From Metaphor to Theory and Practice. IEEE Computer Society, 2012.
- [My74] Glenford J. Myers: Reliable Software through Composite Design. Mason and Lipscomb Publishers, New York 1974.
- [Wi16a] Wikipedia, <https://de.wikipedia.org/wiki/Alter>, Stand 25.6.2016
- [Wi16b] Wikipedia, https://de.wikipedia.org/wiki/Friedensreich_Hundertwasser#/media/File:Magdeburg_Hundertwasserhaus.jpg, Stand 25.6.2016

An Integrated Approach to Evaluation of Domain Modeling Methods and Tools for Improvement of Code Reusability in Software Development

Mykola Tkachuk¹, Iryna Martinkus², Rustam Gamzayev³, and Andrii Tkachuk⁴

Abstract: A domain-driven design (DDD) as a modern approach to improvement of software development quality, especially to support code reuse is considered. To emphasize DDD advantages a 3-level design scheme is proposed which is similar to well-proved 3-level vision about data representation in database development. According to this metaphor the main attention is paid to the phases of logical domain specific modeling (DSM) and to the physical modeling both, with usage of 2 alternative DSM-methods with appropriate CASE-tools: JODA- and ODM approaches respectively. To evaluate their impact on generated code reusability (CR) the software complexity metrics are chosen, and the analytic hierarchy process (AHP) is used to make a final decision about the relationship between CR and DSM.

Keywords: software quality, domain-driven design, domain model method, code reusability, complexity metrics, analytic hierarchy process.

1 Introduction: Research Actuality and Aims

Modern software development is a complex interdisciplinary process, which consists of several interconnected phases, which are quite expensive for large software system projects. In general all traditional and modern software methodologies are supposed to decrease these project costs taking into account some functional and non-functional requirements (or quality attributes) to be met in a target system (see e.g. [So11]). One of the most effective ways to resolve this problem is a reusing of different project solutions (assets): domain knowledge, requirements specifications, software architectures, and finally programming code. This approach is the basis of advanced concepts of software engineering as the creation of software products lines and software factories [GS04], as well as methods of software variability management [CBK13].

Nowadays a Domain-Driven Design (DDD) is considered as a recognized methodology to build a complex software in different application areas with respect to this important challenge: to provide a high level of assets reusability in a given project [Ev03], [AF07], [Re13]. Although main essential advantages and some limitations of DDD are already

¹ NTU “KhPI”, SEMIT Dept, Frunze str., 21, 61002 Kharkiv, Ukraine, tka@kpi.kharkov.ua

² NTU “KhPI”, SEMIT Dept, Frunze str., 21, 61002 Kharkiv, Ukraine, imartinkus@gmail.com

³ NTU “KhPI”, SEMIT Dept, Frunze str., 21, 61002 Kharkiv, Ukraine, rustam.gamzayev@gmail.com

⁴ NTU “KhPI”, SEMIT Dept, Frunze str., 21, 61002 Kharkiv, Ukraine, tkachuk.andrey.polt@gmail.com

discussed intensively in many recent publications, from our point of view the positive core of DDD-methodology can be emphasized once again, if we draw an analogy between DDD-approach to software applications and well-known 3-level vision about data representation in database development [BCN92] (see Fig. 1).

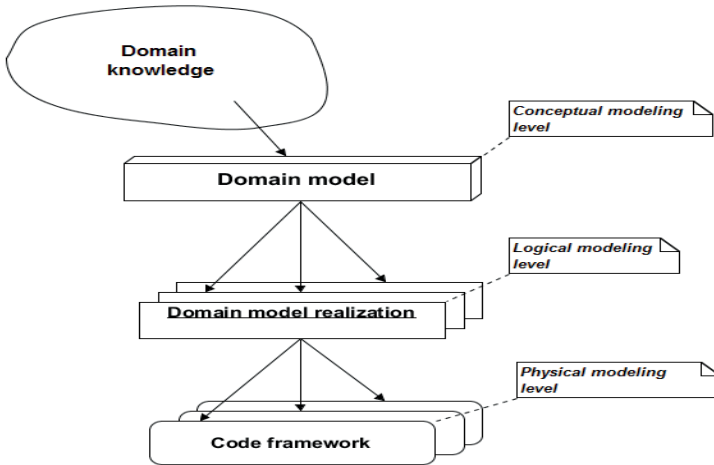


Fig. 1: 3-level scheme in DDD approach

It is to note that according to this vision about DDD-approach for the one and the same domain model (at the conceptual modeling level) a lot of different its realizations (at the logical modeling level) can be constructed, and for each of them an appropriated code framework might be generated finally (at the physical modeling level) using some CASE-tools.

Taking into account the scheme given in Fig. 1 and understanding the fact that there are a lot of different domain-specific modeling (DSM) methods and appropriate CASE-tools, the goal of this paper is to analyze some key capabilities and features of different DSM methods and tools. In this way we are going to propose an integrated approach to their evaluation with respect to prediction of source code reusability and, finally, to improvement of maintainability in a target software system.

2 Domain-specific Modeling and its Impact on Code Reusability. Related Work

Even a briefly overview of recent publications dedicated to code reusability (CR) issues shows that finding a relationship between CR and different impact factors in software development is not a straight forward process. E.g. in [An13] some common reuse design principles are considered, including technological, project's managerial, and even human factors (like experience of developers in different programming languages, etc.).

In [Ta14] the results of the empirical study on CR-trends in open-source software are presented, and a comprehensive collection of reusability metrics are proposed to define the most significant CR-factors. In [Na16] the research is focused on the constructing relationship between extent of CR and values of software complexity (SC) metrics in object-oriented programming. Especially, the following common SC-metrics are used for this purpose

- Depth of Inheritance Tree (DIT)
- Reponses for a Class (RFC)
- Number of Children (NOC)
- Coupling between Object Classes (CBO).
- Weighted Methods per Class (WMC)

Based on experimental studies a set of empirical hypotheses (EH) had been proposed regarding the impact of SC-metrics on CR, namely [Na16]:

- EH1: Better CR can be obtained by moderate value of DIT in every class.
- EH 2: The complexity in code design and CR decreases for maximized values of RFC.
- EH 3: CR degrades for increasing values of NOC related to given class.
- EH 4: An increasing of CBO values does not have much impact over CR
- EH 5: CR declines for increasing values of WMC in every class.

It is necessary to note that besides such empirical assertions like (EH1)-(EH5) concerning relationships between single SC-metrics and an extent of CR, in the most part of publications dedicated to these issues there are no more or less motivated suggestions about the quantitative estimation of correlation between SC metrics and extent of CR. Moreover even in the recent studies mentioned above the correlation between domain modeling methods and complexity of generated code is not analyzed. Hence, it is crucial to identify this correlation because in this way we can make decision about the choosing of an appropriate DSM methods and tools in order to reduce in this way the target implementation costs in DDD-oriented software projects.

There is another important aspect of this investigation. According to [IS16], the 6 main software quality attributes exist, namely: *Functionality, Reliability, Usability, Efficiency, Maintainability and Portability*, and each of them have some sub-characteristics. In case of such quality attribute as *Maintainability* the appropriate list of characteristics includes: *analyzability, changeability, stability, and testability*, where *changeability* characterizes an amount of efforts needed to change a software system to be maintained. From the other hand, in some publications (see e.g. in [AM13]), a correlation between changeability and software complexity metrics is emphasized, especially the CBO and WMC metrics are inversely proportional to the level of changeability in the given software system. From this point of view it can be concluded, that the complexity estimation of code generated basing on DSM finally allows us to make conclusion about such important software quality attribute as its *Maintainability*.

3 Some Domain Modeling Methods and Tools: Short Overvie

3.1 Domain Analysis and DSM methods

During last 10-15 years a lot of different domain analysis (DA) and DSM methods [Fe99], [KT08] were developed. Despite of their differences from the implementation point of view, the most suitable way to classify DA and DSM methods is consider them by type of phases / artifacts to be reused in a software development process. Based on this suggestion the following list of these methods should be considered:

1. DA&DSM methods for software product reuse;
2. DA&DSM methods for software process reuse;
3. DA&DSM methods for software technology reuse;
4. DA&DSM methods for software experience reuse.

Taking into account our main research goal: to identify a relationship between different DSM methods and code reusability (CR) in target application, we further consider more detailed the methods from the group (1). Therefore 2 such methods for software product reuse, namely JODA and ODM [Fe99], were chosen and they are presented briefly below.

Method JODA (Joint integrated avionics Object oriented Domain Analysis) uses object-oriented approach to cover the domain analysis phase, and it includes the following processes:

- Domain data preparation: identification and gathering of appropriate data sources, references and software artifacts which are relevant for a given domain.
- Domain scope definition: elaboration of diagrams for higher-level entities, identify of generalization-specialization, aggregation and other relations within domain, build a domain glossary.
- Domain modeling: identification, definition and modeling several domain scenarios in order to group domain-specific objects and activities to represent them in next domain engineering process.

Method ODM (Organizational Domain Modeling) supports systematically mapping of domain-specific artifacts into reusable assets that can be reused in future software development activities. This approach includes the following phases:

- Plan domain engineering: this one is focused on understanding of stakeholders and defining of domain analysis scope.
- Domain modeling: it concerns collecting and documenting the domain-specific information resources which are relevant for future reusing.

- Domain assets base: the final phase of ODM method that supposes defining the project scope, creating (choosing) system architectures and implementing of physical asset base for the given domain.

In order to support all main phases / activities in any DA&DSM method the appropriate CASE-tool has to be used, and a short overview of them is given in the next paragraph.

3.2 Domain Modeling CASE-tools

Generally, visual modeling tools in software engineering have evolved a lot in recent years. One of the new trends in this domain is the transition from unified modeling environments like UML or SysML [Om10], to some domain-specific modeling (DSM) languages and tools, e.g. WebML, SoaML, and some others [Re13]. These DSM - approaches allow developers to design and to analyze software in terms of target problem domain, and finally to generate source code in different programming languages based on high-level requirements specifications.

It is to mention that exiting CASE-tools for DSM are quite varied in their capabilities, e.g. such wide-used CASE-tools such as: Eclipse Modeling Framework, Rational Rose, FeatureIDE, Visual Paradigm, Actifsource and others [Re13]. To compare them it is necessary to choose a set of criteria, and obviously there are a lot of different ways to define such criteria configurations. Generic enough, and in the same time, a practice-oriented one is the following list of criterion: a possibility to generate code by domain model, a possibility to build model by code, and last but not least: a necessity to have a mandatory license. Taking into account these criteria, exactly Actifsource and Eclipse Modeling Framework (EMF) were chosen for our future research. Both of these CASE-tools are license-free and they support JODA and ODM methods correspondingly.

4 An Integrated Evaluation Approach and Case Study Results

Based on already mentioned points concerning some relationships existing between extent of code reusability (CR) and values of software complexity (SC) metrics we purpose to evaluate the selected DMS methods: JODA and ODM with respect to the CR of the generated programming code. Moreover this approach has to produce an integrated estimation values for extent of CR in target DSM – based applications taking into account an impact of the weighted SC-metrics collection (see in Section 2). To perform this evaluation approach we propose the information technology which is presented in Fig. 2 using IDEF0 notation [ID16]. It consists of 3 fictional Blocks: “A1: DM construction and code generation”, “A2: Code analysis”, “A3: OOP Code Reusability estimation”. Block A1 operates with User stories, obtained from Domain expert and the result of Block A1 is a Domain model and generated source code. After this operation it is possible to perform Code analysis (see Block A2) in order to calculate OOP code complexity metrics (SC-metrics). Finally, Block A3 performs OOP Code reusability estimation and calculates the integrated value of CR-extent.

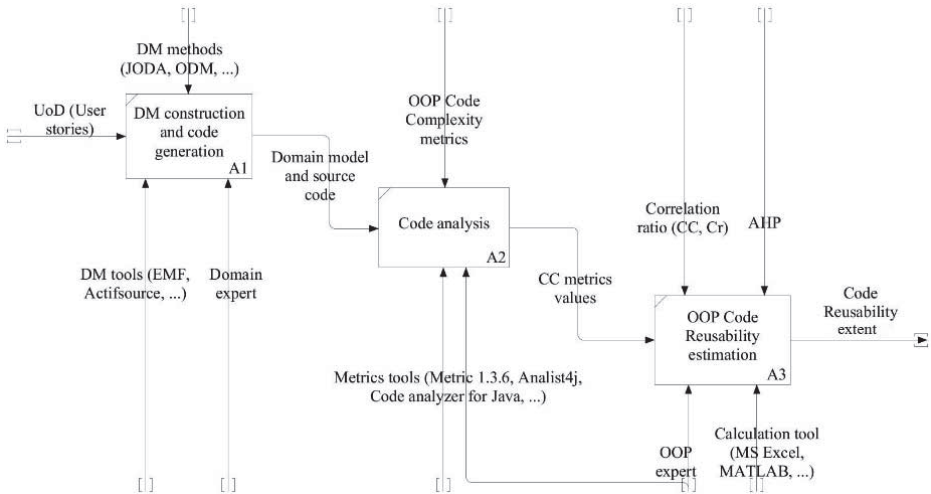


Fig. 2: Information technology to support the evaluation approach

It is necessary to note in our approach to estimate DSM methods we consider the data modeling level only, because at the initial stage of software development, where these methods are used, in fact for the most part of domain entities their methods and interfaces cannot be defined as usually.

According to the common DDD-framework shown in Fig. 1, to apply any DSM method first it is necessary to obtain initial data about target problem domain. One of such way can be representation of problem domain by a collection of user stories, especially if such agile methodologies as Scrum or XP (Extreme Programming) are used in project developer teams [Am16]. This is a high-level definition of system requirements, containing just enough information so that the developers can produce a previous estimate of the effort to implement them. As the case-study the user stories for the simple domain “Students data management in the educational process” were formulated, and they are shown in Table 1.

Iteration	User Stories
1	Maintain student's personal information Maintain student's contact information (address)
2	Define main features of student Identify of student's residence
3	Maintain student's information about his educational career Definition of student's charge

Tab. 1: User stories description

The next step to be done in the proposed evaluation approach is their implementation with usage of 2 chosen DSM methods: JODA and ODM.

4.1 ODM and JODA Implementation

To apply ODM-method *Eclipse Modeling Framework* was used [EMF16]. First we build the logical domain model (see Fig. 3), which represents the initial user stories given in Tab.1.

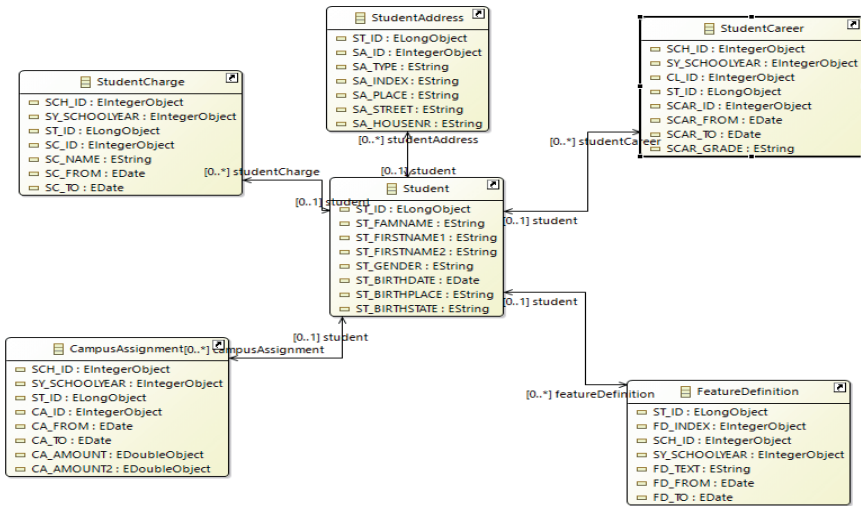


Fig. 3: The domain model in ODM / EMF notation

The next step is to generate Java-source code by this domain model (see Fig. 4).

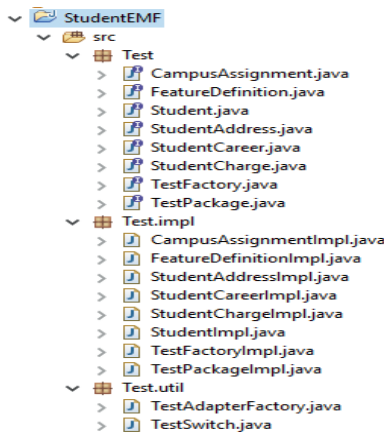


Fig. 4: The package tree generated in EMF CASE-tool

All generated in EMF Java-classes are divided into three packages: *Test* (it includes interfaces), *Test.impl* (contains classes, interfaces, implementation), *Test.util* (utility

classes). The fragment of generated code from one package only is shown below.

```

    /* <!-- begin-user-doc -->
    * <!-- end-user-doc -->
    * @generated
    */
    public Long getST_ID() {
        return sT_ID;
    }
    /**
    * <!-- begin-user-doc -->
    * <!-- end-user-doc -->
    * @generated
    */
    public void setST_ID(Long newST_ID) {
        Long oldST_ID = sT_ID;
        sT_ID = newST_ID;
        if (eNotificationRequired())
            eNotify(new ENotificationImpl(this,
                Notification.SET,
                TestPackage.STUDENT__ST_ID,
                oldST_ID,
                sT_ID));
    }

```

For JODA-implementation the Actifsource CASE-tool was used [Ac16], and the elaborated domain model is shown in Fig.5.

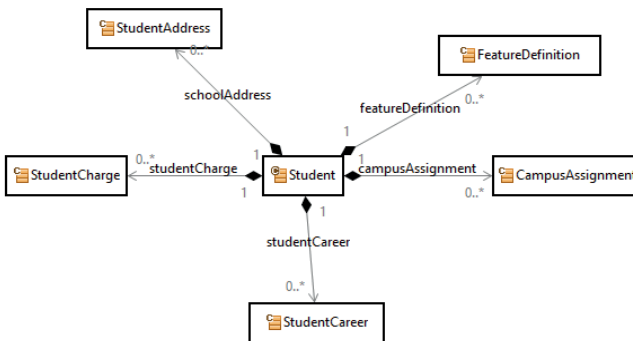


Fig. 5: The domain model in JODA notation

Similar to the previous approach the next step supposes a code generation by this model, and the generated project's package tree is shown in Fig. 6.

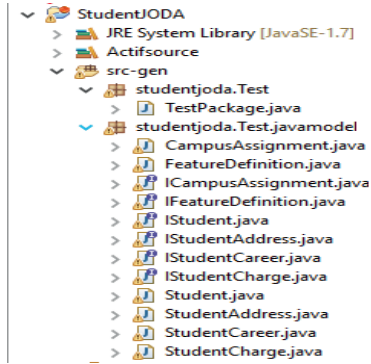


Fig. 6: The package tree generated in Actifsource CASE - tool

The fragment of generated Java-code from one package is shown below.

```
@Override
public java.lang.String selectST_ID() {
    return_getSingleAttribute(java.lang.String.class,
        studentjoda.Test.TestPackage.Student_ST_aE_ID);
}

public void setST_ID(java.lang.String sT_ID) {

    _setSingleAttribute(studentjoda.Test.TestPackage.Student_ST_
        aE_ID, sT_ID);
}
```

The Actifsource-tool creates for the same domain model 2 packages, namely *studentjoda.Test* (contains additional resources) and *studentjoda.Test.javamodel* (includes implementation of classes and interfaces).

The main difference to operate with these facilities is the followings: in Actifsource-tool a source code is generated automatically in case of any changes acquired in the model. By turn, in EMF-tool to generate new code version we need to create a special model-generator utility (see in [EMF16] for more details).

4.2 AHP-based Calculation Scheme and Case Study Results

As already mentioned above, in [Na16] the results of experimental studies are presented, which show the relationship between values of single SC-metrics and the given extent of CR. The fragment of these numerical data is shown in Table 2, where column Value includes the values of SC-metrics, *AVG (Cr)* shows the average values of the CR-extent.

Since our goal is to calculate the integrated value of CR-extent with respect to all

available SC-metrics, we need to define an appropriate coefficient for each such metric. It can be done using the Analytic Hierarchy Process (AHP) method [Sa00].

Metric	Value	AVG (Cr)	Cr/Value
DIT	1	10.46	
	2	25.21	
	
	6	99.11	16,60
RFC	5	21.21	
	10	19.44	
	
	200	95.34	0,46
NOC	0	0.33	
	1	66.67	
	
	6	258.33	43,30
CBO	1	22.22	
	3	21.11	
	
	24	91.73	3,79
WMC	3	40.38	
	5	30.00	
	
	100	105.86	1,05

Tab. 2: CR estimation according to single SC metrics (see in [Na16])

Based on the experimental data presented in [Na16] we can calculate the relative "weight" of one "unit value" of each metric, i.e. the ratio (Cr/Value). It can be concluded, that e.g. the increasing of DIT-metric value to 1 makes CR- index higher, than increasing to 1 of CBO - metric, etc. In this way we are able to determine the relative importance of one metric to others. According to AHP –approach the following assessments scale is used: 1 – equal importance, 3 - moderate importance, 5 - a strong importance, 7 - very strong importance, 9 - extreme importance, and we get the pairwise comparisons for single CR-metrics shown in Table 3.

	WMC	RFC	DIT	NOC	CBO
WMC	3\3	3\1	3\7	3\9	3\5
RFC	1\3	1\1	1\7	1\9	1\5
DIT	7\3	7\1	7\7	7\9	7\5
NOC	9\3	9\1	9\7	9\9	9\5
CBO	5\3	5\1	5\7	5\9	5\5

Tab. 3: Pairwise comparisons

The final AHP estimation values and the weighted coefficients (K) for all CR-metrics are presented in Table 4.

	WMC	RFC	DIT	NOC	CBO	SumRow	K
WMC	1	3	0.42	0.33	0.6	5.35	0,1198
RFC	0.33	1	0.14	0.11	0.2	1.78	0,0398
DIT	2.33	7	1	0.78	1.4	12.51	0,2801
NOC	3	9	1.29	1	1.8	16.09	0,3603
CBO	1.67	5	0.71	0.55	1	8.93	0,2000
Total						44.66	1,0000

Tab. 4: Final AHP estimation values

Thus applying these weighted coefficients to corresponded SC-metrics, we obtain the following formula for integrated value of the CR-extent

$$CR_{extent} = 0,1198 * WMC + 0,0398 * RFC + 0,2801 * DIT + 0,3603 * NOC + 0,2000 * CBO \tag{1}$$

To prove experimentally this approach and to perform the appropriate calculation such tools as Analyst4j, Metrics 1.3.6 plugin for Eclipse, OOMeter, Eclipse Metrics Plug-in 3.4, and some others can be used [LLL08]. These facilities allow not only to calculate average value of each SC-metrics (see Table 5), but also depict them graphically.

Matrices	EMF	Actifsource
WMC	55.00	15.54
RFC	36.70	16,53
DIT	2.9	1.32
NOC	0.44	0.46
CBO	7.80	6.95

Tab. 5: Results of the software metrics' calculation

These values of all CR-metrics for EMF and Actifsource implementations are shown

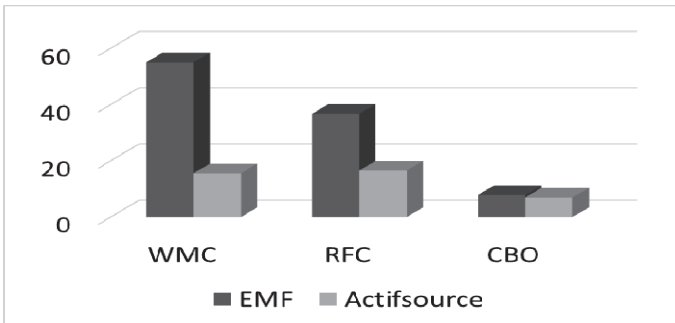


Fig. 7: Results comparison for WMC, RFC and CBO metrics

in Fig. 7 and Fig. 8 respectively.

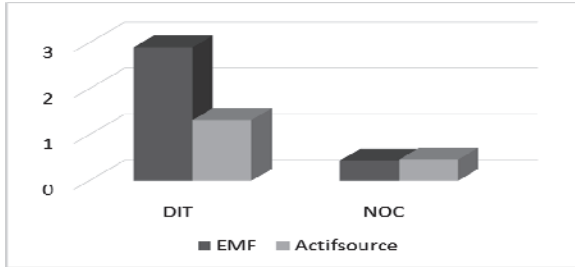


Fig. 8: Results comparison for DIT and NOC metrics

To calculate the integrated estimation value of CR-extent according to formula (1) the special software tool is developed, and the final result of this DSM methods evaluation is shown in Fig. 9.

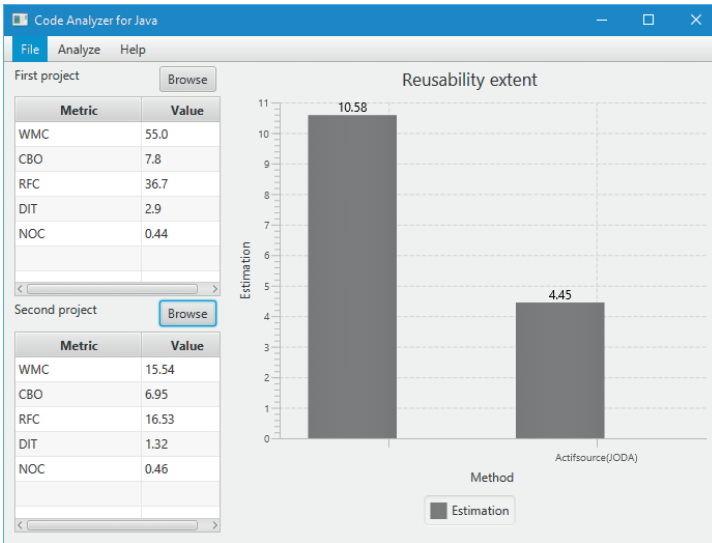


Fig. 9: The prototype main window

The final integrated estimation values for the CR-extent by usage of the chosen DSM-methods: JODA and ODM are the following

$$CR_{extent}(EMF) = 10,58 \tag{2}$$

$$CR_{extent}(Actifsource) = 4,45 \tag{3}$$

Therefore, according to the values in formulas (2) - (3) we can conclude that domain modeling by usage of ODM / EMF tools provides the essential higher extent of code reusability (CR) than by using of JODA method and Actifsource tools.

5 Conclusions and Future Work

In this paper we have considered some essential aspects of domain-driven design (DDD) methodology in modern software engineering. A special attention is paid to comparative analysis of different domain-specific modeling (DSM) methods, and to the appropriate CASE-tools for their implementation. As an important factor for their evaluation the extent of code reusability (CR) is chosen, and the relationship between extent of CR and values of object-oriented software complexity (SC) metrics is emphasized.

Based on this suggestion the integrated approach to evaluation of different DSM / CASE-tools configurations is proposed, which defines the final CR-extent taking into account weighted coefficients for single SC-metrics to be calculated using Analytic Hierarchy Process method. This approach is tested within the elaborated case study domain model, and it allows make decision about the usability of given DSM-methods and tools for reducing of implementation costs in DDD-oriented software projects based on predicted extent of code reusability. Finally, this result can also be considered as a way to improve such important quality characteristic as *Maintainability* for a target software application to be developed using an appropriate DSM-method.

In future we are going to construct the more sophisticated collection of the software complexity metrics, e.g. with metrics of relationships between packages, and to improve the prototype of our software tool to support the proposed evaluation approach.

References

- [Ac16] Actifsource . <http://www.actifsource.com>, accessed on: 15.06.2016
- [AF07] Abel, A.; Floyd, M.: Domain Driven Design Quickly. Lulu.com, 2007.
- [Am16] Ambler, S.W.:Agile/Evolutionary Data Modeling: From Domain Modeling to Physical Modeling., <http://agiledata.org/essays/agileDataModeling.html#DisasterStrikes>, accessed on: 15.06.2016.
- [AM13] Ayalew, Y., Mguni, K.: An Assessment of Changeability of Open Source Software, Computer and Information Science., Vol. 6, No. 3; 2013.
- [An13] Anguswamy, R.: A Study of Factors Affecting the Design and Use of Reusable Components, Software Reuse Lab, Virginia Tech, 2013.
- [Ba06] Balmelli, L. et.al: Model-driven Systems Development. IBM Systems Journal. Vol. 45. P. 569-585, 2006.

- [BCN92] Batini, C., Ceri, S., Navathe, Sh.: Conceptual Database Design: An Entity-Relationship Approach. – Benjamin Publishing Company, 1992.
- [CBK13] Capilla,R., Bosch, J., Kang, K.,: Systems and Software Variability Management, Springer, 2013.
- [EMF16] Eclipse Modeling Framework (EMF). <https://eclipse.org/modeling/emf/>, accessed on: 15.06.2016.
- [Ev03] Evans, E.: Domain-Driven Design: Tackling Complexity in the Heart of Software 1st Edition. Prentice Hall, 2003.
- [Fe99] Ferré, X.: An Evaluation of Domain Analysis Methods.,In 4th CAiSE.IFIP8.1 International Workshop in Evaluation of Modeling Methods in Systems Analysis and Design, P.1-13, 1999.
- [GS04] Greenfield, J.: Short K. Software Factories: Assembling Application with Patterns, Models, Frameworks and Tools, Wiley:-Indianapolis, 2004.
- [ID16] Official Web-site of IDEF Family of Methods, <http://www.idef.com>, accessed on 15.06.2016
- [IS16] ISO 9126 Software Quality Characteristics <http://www.sqa.net/iso9126.html>, accessed on: 15.05.2016.
- [KT08] Kelly, S., Tolvanen, J.: Domain-Specific Modeling: Enabling Full Code Generation. Wiley Computer Society Press. 2008
- [LLL08] Lincke, R., Lundberg, J., Löwe, W.: Comparing Software Metrics Tools, ISSTA '08 Proceedings of the 2008 international symposium on Software testing and analysis, P.131-142, 2008.
- [Na16] Nandakumar, A.N.: Constructing Relationship between Software Metrics and Code Reusability in Object Oriented Design, International Journal of Advanced Computer Science and Applications, Vol. 7, No. 2, 2016.
- [Om10] OMG Unified Modeling Language, Superstructure. Version 2.3. OMG, 2010.
- [Re13] Reinhartz-Berger, I. et al., eds. Domain Engineering: Product Lines, Languages, and Conceptual Models. Heidelberg, Springer, 2013.
- [Sa00] Saaty, T.,L.: Fundamentals of the Analytic Hierarchy Process. RWS Publications, 4922 Ellsworth Avenue, Pittsburgh, PA 15413, 2000.
- [So11] Sommerville, I.: Software Engineering. Addison Wesley, 2011.
- [Ta14] Taibi, F.: Empirical Analysis of the Reusability of Object-Oriented Program Code in Open-Source Software, International Journal of Computer, Electrical, Automation, Control and Information Engineering, Vol.8, No.1, 2014.

Beurteilung von Software Qualität stärken durch Ontologien und Wiederverwendung von Wissen

Stefan Strell¹, Vladimir A. Shekhovtsov² und Heinrich C. Mayr³

Abstract: Angemessene Qualitätsbeurteilung bei Software erreichen wird beeinträchtigt wenn Beteiligte kommunizierte qualitätsbezogene Informationen unterschiedlich wahrnehmen. Wir beabsichtigen diesen Kernpunkt durch Einführung eines breiteren Kontexts von Quality-Aware Software Engineering zu adressieren und diesen Zusammenhang durch Einführung des entsprechenden Software-Toolset, das es erlaubt bereits existierendes Wissen wiederzuverwenden und Werkzeuge zur Hilfe und Förderung der Qualitätsbeurteilung für SoftwareentwicklerInnen zur Verfügung zu stellen, beispielhaft einzuführen. Wir führen das spezifische Beispiel der Entwicklung des Tool-Support für qualitätsbewusste Softwareentwicklung basierend auf den Ergebnissen des Forschungsprojekts QuASE ein. Diese Lösung unterstützt Verständlichkeit und Wiederverwendung von kommunizierten Informationen sowie die Qualität von Entscheidungen basierend auf solchen Informationen durch Verwertung entsprechender Ontologien und Knowledge Bases. In diesem Artikel legen wir den Schwerpunkt auf die Komponenten um Wissensverwaltung des QuASE-Systems, speziell auf seine Knowledge Base (KB), die aus zwei Ontologien besteht: Eine *Site Ontology*, die das Site-spezifische Kommunikationsumfeld definiert, und eine Ontology die das gesamte notwendige Wissen zur Unterstützung der qualitätsbezogenen Kommunikation enthält. Wir beschreiben die gesamte Architektur des Systems und die Komponenten, die für Unterstützung der Wissensverwaltung zuständig sind, im Detail und stellen Informationen über bereits eingesetzte Applikationen dieser Softwarelösung zur Verfügung.

Keywords: Quality-Aware Software Engineering; Software Qualität; Knowledge Reuse; Ontologie; Knowledge Base.

1 Einleitung & Motivation

Seitdem es die Softwareentwicklung als eigenständiges Fachgebiet in der Informatik gibt, ist die Qualität und vor allem die Verbesserung ein oft diskutiertes Thema. Alle ExpertInnen sollten das Ziel haben, Software nicht nur für Maschinen und Computer zu entwickeln, sondern vor allem um den BenutzerInnen, also den Menschen, zu helfen besser zurecht zu kommen. Die Informatik und die Softwareentwicklung sind allerdings keine isolierten Fachgebiete, sondern hier arbeiten zum Teil ExpertInnen in den verschiedensten Bereichen, wie Industrie 4.0 und IoT, zusammen um gemeinsam Ziele,

¹ Alpen-Adria-Universität Klagenfurt, Institute for Applied Informatics, Universitätsstraße 65-67, A-9020 Klagenfurt am Wörthersee, sstrell@edu.aau.at

² Alpen-Adria-Universität Klagenfurt, Institute for Applied Informatics, Universitätsstraße 65-67, A-9020 Klagenfurt am Wörthersee, Volodymyr.Shekhovtsov@aa.u.at

³ Alpen-Adria-Universität Klagenfurt, Institute for Applied Informatics, Universitätsstraße 65-67, A-9020 Klagenfurt am Wörthersee, Heinrich.Mayr@aa.u.at

die sog. „*Grand Challenges*“ [GI16] zu erreichen. Zu diesen zählt nach [GI16] auch die Verlässlichkeit von Software, eines der wichtigen Qualitätskriterien der Softwareentwicklung und zugleich eines der umstrittensten. Nach Hoffmann in [Ho13] gibt es acht Kategorien um Software Qualität einzuteilen, wie in Abb. 1 veranschaulicht.

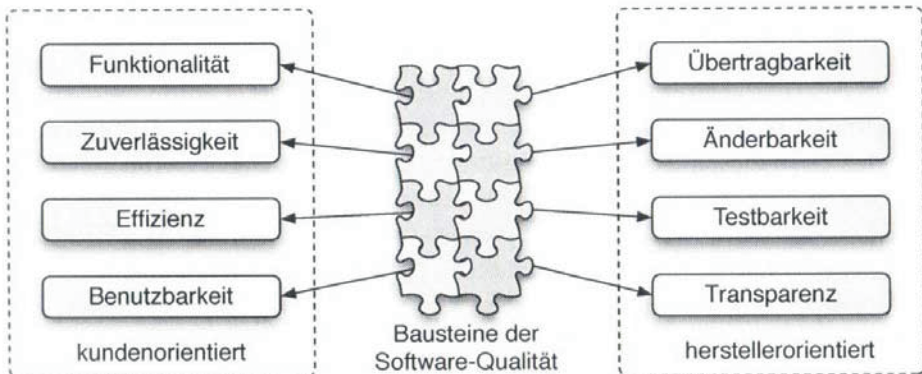


Abb. 1: Qualitätsmerkmale eines Software Produkts (aus [Ho13], S. 7)

Wie in [St16] von einem der Autoren bereits angemerkt gibt es keine eindeutige Definition für Software Qualität. Dieser Umstand macht es mitunter schwierig die Kommunikation zwischen den einzelnen Beteiligten in einem Softwareentwicklungsprojekt auf gleicher Ebene durchzuführen und ist ein wichtiger Aspekt im Forschungsprojekt QuASE⁴. Es wurde in [St16] eine Arbeitsdefinition verwendet, welche in Kapitel 3.1 ersichtlich ist. Das Themengebiet um Software Qualität befasst sich u.a. mit den funktionalen und nicht-funktionalen Anforderungen an ein konkretes Softwareprodukt. Zu den funktionalen Anforderungen werden die spezifischen Funktionen der Anwendung gezählt, also was die Software später auszeichnet (vgl. [St14]). Die nicht-funktionalen Anforderungen sind in der Regel komplexer und nicht einheitlich beschreibbar. Zu diesen zählen Sicherheit, Flexibilität, Usability, Performance und Interoperabilität (vgl. [CP09, St14]). Vor allem diese Anforderungen werden in der klassischen Softwareentwicklung nicht so stark berücksichtigt, bzw. immer wieder neu definiert, allerdings sind sie für die Qualitätsbeurteilung unerlässlich. Um diesen Umstand zu reduzieren und vor allem die Kommunikation zwischen den Beteiligten in einem Softwareprojekt zu verbessern, wurde das Forschungsprojekt QuASE initiiert.

In diesem Artikel zunächst werden die wichtigsten Begriffe definiert, die in der weiteren Folge zur Verwendung kommen und anschließend einleitende Worte über die qualitätsbewusste Softwareentwicklung vorgestellt. In Kapitel 3 wird das Forschungsprojekt QuASE und dessen Proof-Of-Concept erläutert. Kapitel 4 beschäftigt sich mit den Ontologien und der Knowledge Base im Forschungsprojekt QuASE. In

⁴ Offizielle Webseite des Forschungsprojekts: <http://quase-ainf.aau.at/>

Kapitel 5 wird auf weiterführende Literatur verwiesen. Der Artikel schließt mit einer Zusammenfassung und einem Ausblick, sowie einer kurzen Wiederholung zur Beantwortung der Forschungsfrage aus [St16] ab.

2 Quality-Aware Software Engineering

2.1 Definitionen & Grundlagen

Ontologien & Knowledge Base

Der Begriff „Ontologie“ ist im Gegensatz zur „Knowledge Base“ nicht nur in der Informatik zu finden, sondern auch Bestandteil der Wissenschaft der Philosophie. Hier ist es vor allem die Metaphysik, die diesen Begriff geprägt hat (vgl. [Gu05] S. 51 ff., [St16], S. 9 ff.).

Die Knowledge Base (KB) ist das Konzept zur Digitalisierung, Repräsentation und Verwertung von Wissen. Diese werden entweder über komplexe Datenbanken oder mit Hilfe von Ontologien aufgebaut (vgl. [St16], S. 17). Anwendungen, die KBs verwenden, nutzen dafür zumeist komplexe Metriken und SPARQL als Query-Language.

Arbeitsdefinition von Software Qualität

Der Begriff „Software Qualität“ lässt sich nicht konkret mit genau einer Definition beschreiben. Das liegt vor allem an den vielen unterschiedlichen Qualitätsmerkmalen eines Software-Produkts. Denn jede/r SoftwareentwicklerIn und auch jede/r EndbenutzerIn hat eine andere Auffassung, welche Kriterien für ihn bzw. sie als Qualitätsmerkmale in einer Software vorhanden sein müssen. Nach der ISO/IEC-Norm 9126 wird der Begriff Software Qualität wie folgt definiert:

„Software Qualität ist die Gesamtheit der Merkmale und Merkmalswerte eines Software-Produkts, die sich auf dessen Eignung beziehen, festgelegte Erfordernisse zu erfüllen.“ ([Ho13], S. 6, vgl. [ISO00])

Da dies nur eine von vielen Definitionen über den Begriff „Software Qualität“ ist wurde in [St16] eine Arbeitsdefinition erstellt, die auf den gängigen ISO-Standards beruht:

„Die Qualität von Software wird zum einen durch die Fähigkeit zur Ausführung der gestellten Anforderungen und zum anderen durch die effektive und effiziente Projektdurchführung gekennzeichnet.“ ([St16], S. 29)

2.2 Qualitätsbewusste Softwareentwicklung

Die qualitätsbewusste Softwareentwicklung (engl. *Quality-Aware Software Engineering*) hat zum Ziel die Qualitätsbeurteilung in der Softwareentwicklung zu stärken. Vermehrt

wird dieser Ansatz in der Entwicklung von Automatisierungs- und modellbasierten Systemen eingesetzt (vgl. [St16], S. 47, [Ab14]). In [St16] wird der Begriff „Quality-Aware Software Engineering“ folgendermaßen definiert:

„Die qualitätsbewusste Softwareentwicklung unterscheidet sich vom klassischen Software Engineering vor allem dadurch, dass die Software Qualität und die nichtfunktionalen Anforderungen im Vordergrund stehen.“ ([St16], S. 47)

Ziel der qualitätsbewussten Softwareentwicklung und des Forschungsprojekts QuASE ist es, nicht Fehler zu vermeiden, sondern diese früher zu erkennen und aus ihnen besser lernen zu können. Eines der wichtigsten Ziele dieses Ansatzes ist es, *„die Kosten bei der durch ExpertInnen durchgeführten Wartung zu reduzieren“* ([St16], S. 47, vgl. [Ab14]). Hierbei soll nach [St16] und [Ab14] vor allem die Unsicherheit (*uncertainty*) reduziert und aus vorangegangenen Softwareprojekten effizienter gelernt werden (vgl. [St16], S. 47). Um diese Ansätze zu verfolgen werden, wie in [St16] ausführlich beschrieben, vor allem die ISO-Standards 9000 und 9001, sowie eine klare Arbeitsdefinition von Software Qualität hinzugezogen (vgl. [St16], S. 48). In der qualitätsbewussten Softwareentwicklung nutzt man bereits vorhandenes Wissen, das digitalisiert aufbereitet wird, um dieses effizienter zu verwenden. Eine gute Kommunikation unter den Beteiligten und ein klares Entwicklungskonzept zeichnen nach [St16] ebenfalls eine gute Basis dafür aus.

Diese Ansätze nutzt das Forschungsprojekt QuASE, deren Ziel eine Applikation zur Erleichterung der Softwareentwicklung durch eine stärkere Beurteilung von Software Qualität zu entwickeln. In den folgenden Kapiteln werden diese zentralen Themen behandelt und gezeigt, wie Ontologien und Knowledge Bases helfen können, die Qualität von Softwareprojekten zu erhöhen.

3 QuASE-Projekt: Ziele, Architektur & derzeitiger Stand

Das Forschungsprojekt QuASE wurde mit dem Primär-Ziel einen Proof-Of-Concept zu erstellen, der zur Stärkung der Qualitätsbeurteilung durch Wissensverarbeitung dient (vgl. [St16], Kap. 4 ff.). Abb. 2 zeigt das schematische Konzept der Applikation und wie die unterschiedlichen Sichtweisen der Beteiligten eines Softwareprojekts ergänzt und erklärt werden können.

Die automatische Verarbeitung von Wissen mit Knowledge Bases und Ontologien stand hierbei im Vordergrund. Denn nur so können unterschiedlich verwendete Fachbegriffe zusammengeführt und wiederverwendet werden. Die QuASE-Applikation besteht, wie in [St16] dargestellt, aus einer 3-Schichten-Architektur (GUI, Programm-Logik, Daten) (vgl. [St16], S. 59 ff.).

Die Implementierung der Verständlichkeitsverwaltung in QuASE verwendet Methoden zur Verarbeitung von natürlicher Sprache um die Quelle(n) von

Verständlichkeitskonflikten in kommunizierter Information zu lokalisieren, um Konflikte durch terminologische Übersetzung zu reduzieren und um gezielte Erklärungen zu erstellen.

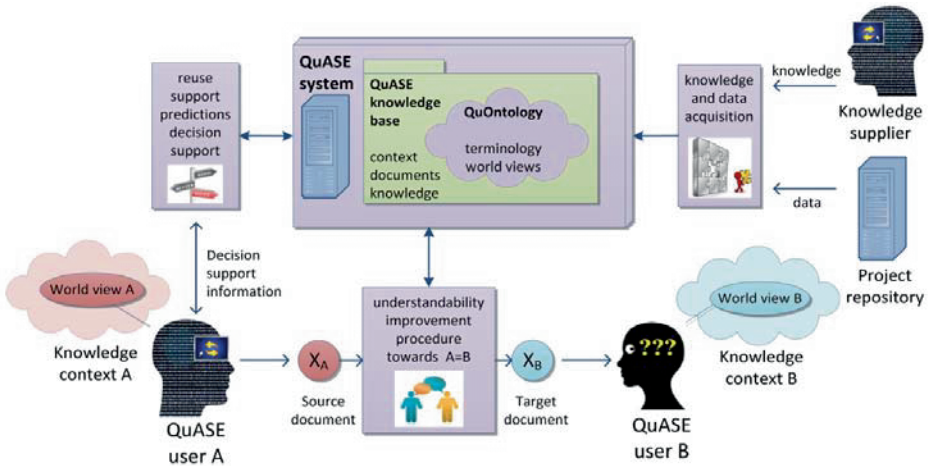


Abb. 2: Das schematische Konzept von QuASE (aus [St16, MS16, SMK15])

Der Teilbereich *Analytical Support* in QuASE verwendet Methoden für maschinelles Lernen für die Menge der Individuals in der QuASE Knowledge Base: Ähnlichkeitssuche für die Wiederverwendung von Informationen, Regressionsrechnung zur Vorhersage der Attribut-Werte und hybrides (teilweise überwachtes) Lernen für auf Klassifizierung basierendes Vorschlagswesen und Entscheidungshilfen.

Der größte Vorteil des QuASE-Ansatzes ist, dass nach Auslesen der Daten aus einem Projekt-Repository, wie z.B. Jira (<http://www.atlassian.com/jira>), diese sofort in Wissen konvertiert werden und für Verständlichkeitsverwaltung und Analysis verfügbar sind. Außerdem erlaubt dies, dadurch dass das Mapping durch die Verwendung einer speziellen DSL flexibel ist, große Mengen von Daten zur Anwendung dieser Methoden.

Das QuASE-System kann als Brücke gesehen werden, die EndbenutzerInnen, die Daten in Projekt-Repositories und die (erweiterbare) Menge von Methoden für maschinelles Lernen und Verarbeitung von natürlicher Sprache verbinden. Diese Methoden werden automatisch nach Beschreibung des Repositories und der Kommunikationsumgebung durch die *QuASE site DSL* verwendbar für die Repository-Daten.

Die QuASE-Architekturkomponenten werden in Abb. 3. Die rot gestrichelte Linie zeigt die Abgrenzung der QuASE-Applikation in seiner Umgebung. Die Applikation beinhaltet Client- und Server-Komponenten. In den folgenden Abschnitten werden diese Komponenten im Detail beschrieben.

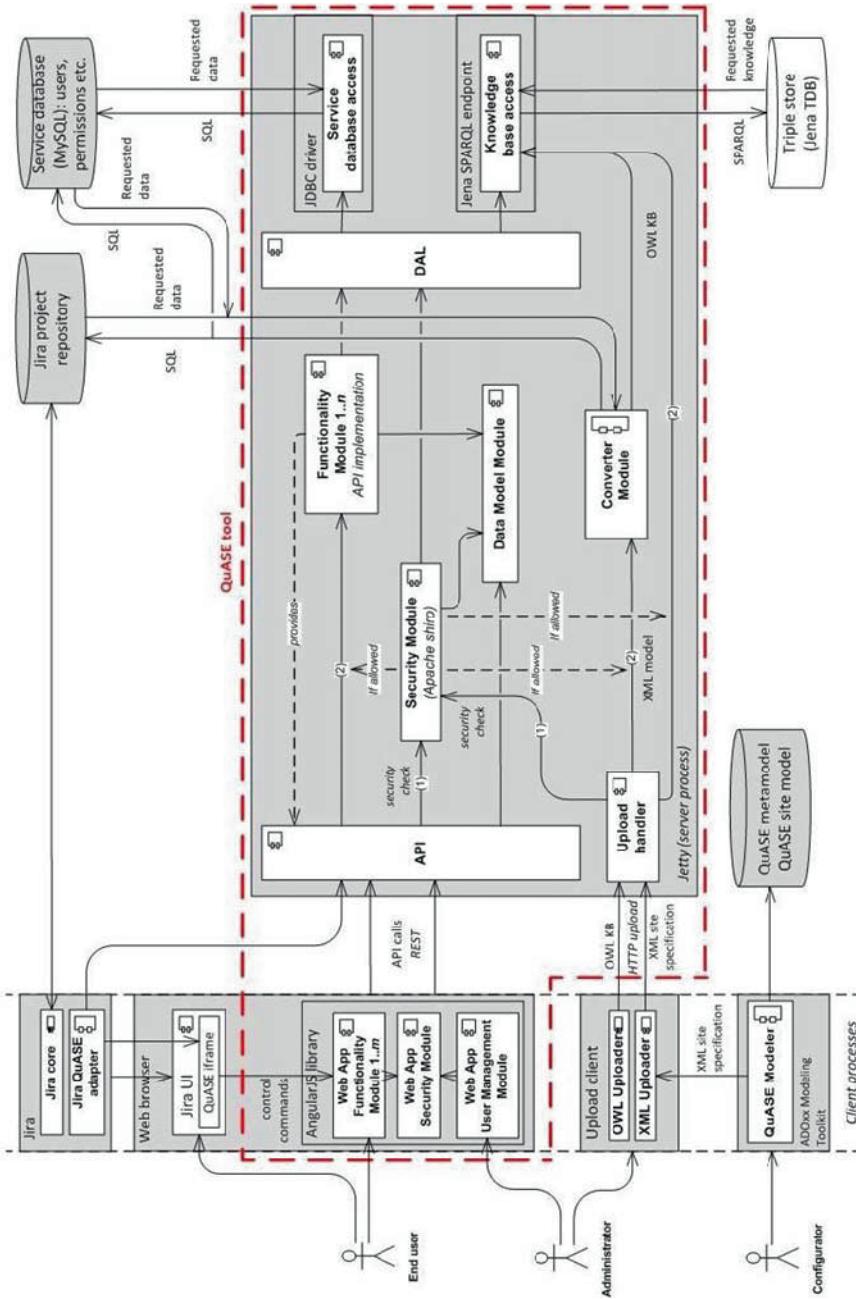


Abb. 3: QuASE System-Architektur

3.1 QuASE Server

Die Hauptfunktionalität der QuASE-Applikation ist im *QuASE Server* implementiert. Es umfasst folgende Funktionalitätsmodule:

1. Sicherheitsmodule;
2. Module für Kontext- und Dokumentenverarbeitung;
3. Verständlichkeitsverwaltungsmodul;
4. Analysis-Module;
5. Module zur Datensammlung;
6. Server-Administrationsmodul.

Diese Module basieren auf den folgenden Komponenten:

1. Die Datenzugriffsschicht (Data Access Layer, DAL) ist durch das DAL Modul implementiert, unterschiedliche Komponenten werden für den Zugriff auf die QuASE knowledge base (gekapselte SPARQL-Queries) und für den Zugriff auf die QuASE service database (gekapselte SQL-Queries) verwendet. Es führt höhere Abstraktionslevels ein, um die Apache Jena TDB und die interne Datenbank, die die Funktionalität komplexe SPARQL-Queries zu konstruieren beinhaltet, zu managen und abzufragen.
2. Die Klassen der Business Logik sind durch das Modul für das Datenmodell (Data Model Module) gekapselt. Der Quellcode zu diesem Modul greift auf externe Speicher durch die DAL zu. Das Datenmodell ist über die Funktionalitätsmodule aufgeteilt. Es stellt eine Menge von Wrappern für Individuals, Daten und Objekt-Properties abgerufen von der Knowledge Base (KB) zur Verfügung.

3.2 Interaktiver QuASE Client

Um EndbenutzerInnen den Zugriff zu ermöglichen ist der QuASE Server durch einen interaktiven *QuASE Client* erreichbar. Der QuASE Client ist in JavaScript implementiert, basierend auf modularer Infrastruktur. Die Infrastruktur beinhaltet den Quellcode zum Einbetten der Security- und API-Zugriffe (API- und Security-Web-App-Modules) und die Menüführung für den Aufruf der Module. Die Menüführung ist mit der Sicherheitsinfrastruktur verbunden, um die Rollen zu spezifizieren, für welche die einzelnen Menü-Items sichtbar bzw. unsichtbar sind.

Die folgenden Kategorien der Web-App-Funktionalitätsmodule sind:

1. Verständlichkeitsverwaltung und -einschätzung;
2. Analysis (Ähnlichkeitssuche, Vorhersage, Empfehlungen und Entscheidungshilfen);

3. Datensammlung (Entscheidungssammlung, Empfehlungsauswertung, Sammlung externer Werte);
4. Web-Administrationsmodul.

3.3 Architektur zur Integration von Jira in QuASE

Die Architektur zur Integration von Jira in QuASE beinhaltet folgende Komponenten:

1. Die Erweiterung der QuASE-Applikation, welche die Befehle von Jira akzeptiert. Diese Befehle beinhalten die Informationen über die aktuell ausgewählte Instanz der Kontext- oder Content-Unit (die Auswahl wird durch den Jira-User über die Jira-UI durchgeführt);
2. Eine Jira-Support-Erweiterung, die die Funktionalität von Jira zur Ausgabe der Kontroll-Anforderung, die zum QuASE-Tool weitergegeben wird, erweitert.

4 Ontologien & Knowledge Base in QuASE

In diesem Kapitel werden zunächst die wichtigsten Begriffe vorgestellt um anschließend auf die Technologien einzugehen, die die Beurteilung der Software Qualität mit Hilfe der Wiederverwendung von Wissen zu stärken. Das Forschungsprojekt QuASE an der Alpen-Adria-Universität Klagenfurt hat diese Technologien als Grundlage für einen Proof-Of-Concept zusammengetragen und eine Software entwickelt die diese zukunftsfähig machen soll.

4.1 QuASE Ontologien

Die Ontologien im Forschungsprojekt QuASE, *QuASE site ontology* und *QuOntology* sind Kernbestandteile der QuASE KB. Die *QuOntology* stellt dabei den wissensspezifischen Teil der Knowledge Base, während die *QuASE site ontology* den kontextspezifischen Teil realisiert (vgl. [St16], S. 67). Die konkrete Umsetzung und Anwendung der Ontologien und QuASE KB werden in [St16] und [SMK15] detailliert dargestellt. Abb. 4 zeigt einen schematischen Aufbau der QuASE Ontologien.

Die *QuASE site ontology* beinhaltet folgende Konzepte (siehe auch [St16], S. 71):

1. *site*: Besitzer der gegebenen QuASE-Installation, z.B. ein Softwareanbieter.
2. *context*: Einheiten (*units*) haben spezielle Sichten auf kommunizierte Informationen, z.B. Projekte, Organisationen und deren Abteilungen, beteiligte Stakeholder

3. *content*: Einheiten gestalten kommunizierte Informationen aus den Projekt-Repositories, wie Issues oder Kommentare in Issues;
4. *knowledge*: Einheiten, die kommuniziertes Wissen kapseln, welche ein Gegenstand der Angleichung sind. (vgl. [St16], S. 71/72)

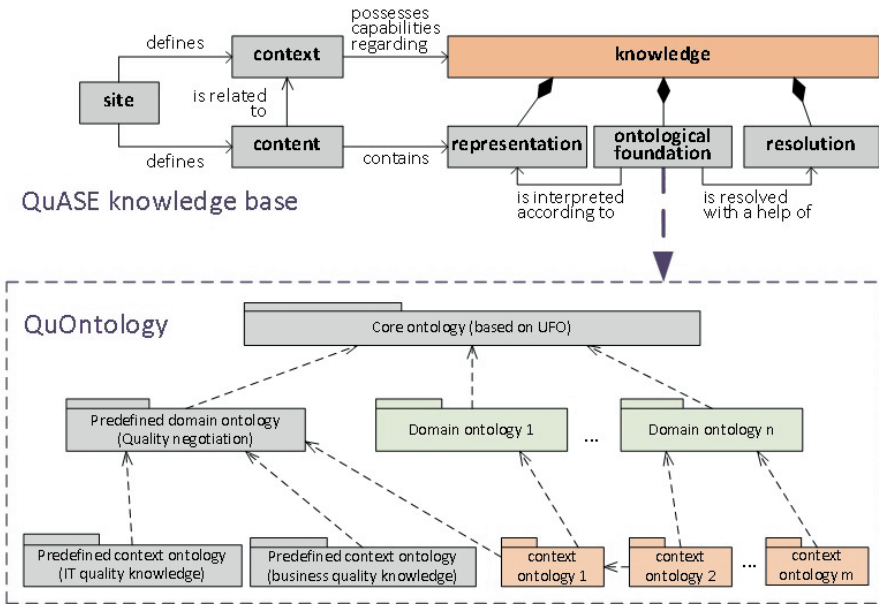


Abb. 3: Konzept der QuASE KB, site ontology & QuOntology (aus [St16, SMK15])

Jede QuASE Knowledge-Unit ist zusammengesetzt aus den folgenden Komponenten:

1. *ontological foundation*: eine Referenz zur Konzeptualisierung der speziellen Teile von Qualität oder Domänenwissen durch die ontologischen Mittel;
2. *representation*: die Repräsentation einer Knowledge Unit in einem Format, das von den Kommunikationspartnern wahrgenommen werden kann (z. B. normaler Text). Representation Units sind in Content Units enthalten.
3. *resolution means*: die Mittel der Auflösung von Verständnis-Konflikten, die zu speziellen Knowledge Units gehören (z.B. Erläuterungen oder externe Referenzen).

Context Units besitzen *capabilities* (deutsch: Funktionen) um mit Knowledge Units umgehen zu können; insbesondere diese Funktionen können sich auf die Fähigkeit des Verstehens einer gegebenen Knowledge Unit (z.B. seine Repräsentation) beziehen; oder eine Knowledge Unit mit einer resolution means erklären. (vgl. [St16], S. 72; [SMK15])

4.2 Wissenswiederverwendung in QuASE

Zur Spezifikation und Wartung der Knowledge Base stellt QuASE folgende Architektur-Komponenten bereit:

1. QuASE-Modellierungstool (*QuASE modeler tool*);
2. Werkzeug für das Hochladen von Modellen (*Model upload tool*);
3. QuASE-Konvertierungsmodul (*QuASE converter module*);
4. QuASE-Komponente zur Unterstützung der Knowledge Base (*QuASE knowledge base support component*).

Diese Komponenten werden in den folgenden Abschnitten beschrieben.

QuASE-Modellierungstool. Dieses Tool, basierend auf dem ADOxx Metamodeling Framework (<http://www.adoxx.org>), erlaubt es ein *QuASE site model* mit einer graphischen Modellierungsschnittstelle zu erstellen und editieren, sowie die Informationen dieses Modells und die spezifischen Verbindungsinformationen für die zur Modellierung bereitgestellte Organisation, um die QuASE KB auf den QuASE Server zu bereitzustellen.

Dieses Werkzeug unterstützt zwei Arten von Diagrammen:

1. QuASE Site-Modell-Diagramm (*QuASE site model diagram*), welches die Struktur des Kommunikationsumfelds und des kommunizierten Wissens beschreibt. Dieses Modell hängt nicht von site-spezifischen Informationen ab (z.B. Typ oder Adresse der Projekt-Repository-Datenbank, die Adresse des QuASE Server, etc.)
2. QuASE-Arbeitsbereichsdiagramm (*QuASE workspace diagram*), welches spezifische Informationen zu einer bestimmten site beinhaltet und diese Informationen zusammen mit den Referenzen zu den passenden Site-Modell-Diagrammen zur Bereitstellung organisiert.

Speziell das QuASE-Arbeitsbereichsdiagramm enthält folgende Elemente:

1. Das Konfigurationselement welches Informationen zu bestimmten Deployment-Site kapselt, dies schließt den Hostnamen und TCP-Port des entfernten QuASE Server, JDBC URLs und benutzerfreundliche Namen zur Verbindung mit dem QuASE Service-DB und die Project Repositories etc. ein;
2. Das Query-Quellelement, welches die spezifischen Informationen für das RDBMS, zuständig für die Implementierung des bestimmten Projekt-Repositories kapselt. Es beinhaltet die Menge der RDBMS-spezifischen Queries, indiziert durch den Namen der Query übertragen durch das Site-Modell. Es können beispielsweise separate Quellen für Oracle- und MySQL-Query-Implementierungen eines bestimmten Repositories existieren, so dass das

Umziehen des Repositories von einem DB-Server zu einem anderen nicht zu einer Änderung des Site-Modells führt;

3. Das Site-Spezifikationselement, welches die Referenzen zur bestimmten Site-Konfiguration und den Query-Quellelementen, sowie das bestimmten Site-Modell beinhaltet, z.B. "die Spezifikation für das Deployment von Unternehmen A (QuASE Server *qqq*) zum Oracle-basierenden Repository bereitgestellt auf Server *xxx*". Dieses Element beinhaltet alle notwendigen Informationen um das Deployment der QuASE KB auf den bestimmten QuASE Server durchzuführen; seine UI stellt den "deploy"-Button zur Initiierung des Deployment-Prozesses zur Verfügung.

Eine Deployment-Aktivität, die vom Modeling-Tool angestoßen wird, bringt den Upload der XML-Datei, die vom ausgewählten Site-Spezifikationselement erzeugt wurde (beinhaltet die Informationen aus Site-Modell, Site-Konfiguration und Query-Quellelement), in den spezifischen QuASE Server, der den Upload Client verwendet.

Upload Client. Der Upload Client ist ein Kommandozeilen-Tool, welches zum QuASE Server verbindet. Es führt die notwendige Authentifizierung durch, und transferiert die bestimmte Datei zu diesem Server. Es ist in der Lage folgende Dateien hochzuladen:

1. OWL Knowledge Base Dateien (z.B. generiert durch Kommandozeilen-Converter),
2. XML Site Spezifikationsdateien (z.B. exportiert durch das QuASE-Modellierungstool).

Der Default-Modus zum Start des Upload Clients ist aus dem QuASE-Modellierungstool. In diesem Fall werden die Modell-Informationen korrespondierend zur aktuellen Site-Spezifikation in eine XML-Datei konvertiert, das mittels des Tools übertragen wird.

Knowledge Base Unterstützung im QuASE Kernel. Einige Architektur-Komponenten unterstützen die QuASE KB, das wichtigste davon ist ein Konverter-Modul (*Converter Module*). Dieses Modul erzeugt die QuASE KB aus dem QuASE Site-Modell, interaktive Daten werden in der QuASE Service-DB gespeichert und die Daten für die Site sind verfügbar aus dem Projekt-Repository (z.B. Jira-Datenbank).

1. Ein Konverter-Modul konvertiert zuerst eine XML-Model-Datei in eine OWL QuASE site ontology, die die Struktur des Kommunikationsumfeld beschreibt;
2. Nach Erstellung der QuASE site ontology sammelt dieses Modul die Daten aus dem Projekt-Repository (z.B. der Jira-DB), sowie die vom Benutzer bereitgestellten Daten, die in der QuASE Service-DB gespeichert sind, und konvertiert diese Daten in die Menge der OWL-Individuals entsprechend zur OWL site ontology. Dies formt die QuASE site knowledge base.

Als ein Resultat besitzt die QuASE Site-KB die wiederherstellbare Eigenschaft: z.B. die

komplette KB kann jederzeit basierend auf dem QuASE Site-Model, den Daten aus der QuASE Service-DB und den Daten aus dem Projekt-Repository rekonstruiert werden.

Es existieren zwei Implementierungen des Konverter-Moduls, die im QuASE-System eingebaut sind:

1. Die Implementierung als ein "In-Process"-Modul für den QuASE Server, aufgerufen durch seine Datei-Upload-Handler-Komponente (dies ist der Default Use-Case für den Aufruf des Konverter-Codes);
2. Die Implementierung als ein separates Kommandozeilen-Werkzeug das remote aufgerufen werden kann. In diesem Fall kann die resultierende OWL KB zum Server hochgeladen werden.

Basierend auf der vom Konverter-Modul generierten OWL KB-Datei initialisiert ein spezielles KB-Support-Modul einen Jena TDB Triple-Store (es ist möglich den Triple-Store von Grund auf zu generieren, sowie inkrementelle Updates durchzuführen).

Andere Komponenten unterstützen die KB folgendermaßen:

1. Ein Teil des DAL-Modul weist SPARQL-basierten KB-Zugriff mit Verbindung zum Modell auf und mit der Möglichkeit zum Schreiben von Ad-hoc benutzerdefinierten SPARQL und die resultierenden Daten auszuführen;
2. Ein UI-Modul implementiert den sog. „Internal Endpoint“, z.B. das interactive Query-Fenster erlaubt dem Benutzer eigens definierte SPARQL Queries am existierenden TDB-Store auszuführen.

Aktueller Stand des Projekts. Die QuASE-Applikation wurde an den Standorten von zwei Partner-Organisationen installiert und in Betrieb genommen. Es arbeitet gut mit deren Repositories, welche bis zu 26000 Issues bzw. Tickets beinhalten. Die resultierende Knowledge Base beinhaltet bis zu 1,7 Million Axiome ohne signifikante Performance-Probleme.

Die Haupt-Verwendungszwecke des Tools sind wie folgt:

1. Die EndbenutzerInnen bevorzugen das System über die Jira-Integrationsschnittstelle zu verwenden, was ihnen bekannte Möglichkeiten zum Erreichen der speziellen Kontext- und Content-Units bietet;
2. Die vermeintliche Nutzlosigkeit der Understandability-Management-Fähigkeiten der Applikation hängt von der Qualität des terminologischen Wissens in der KB (QuOntology) und zuallererst an der Vollständigkeit dieses Wissens in Verbindung zur bestimmten Applikationsdomäne (z. B. Software Qualität) ab.
3. Die vermeintliche Nutzlosigkeit der analytischen Fähigkeiten der Applikation hängen von der Vollständigkeit der Menge der verfügbaren Metriken für die Analyse ab. Es wird empfohlen qualitative Studien in Zusammenarbeit mit den

UnternehmensvertreterInnen durchzuführen, um die Metriken, welche den Entwicklungsprozess beeinflussen, in das Site-Modell zu integrieren.

5 Related Work

Dieser Artikel basiert zumeist auf den wissenschaftlichen Artikeln und Arbeitsunterlagen des Projektteams QuASE (u.a. [SMK15], [Sh15]) der Forschungsgruppe Application Engineering. Zum Thema Software Qualität bieten Wallmüller [Wa01] und Hoffmann [Ho13] weiterführende Informationen. Guizzardi [Gu05] bietet allgemeine Informationen über Ontologien, [SMK15] und [Sh15] bieten eine Übersicht über die konkrete Umsetzung der Ontologien und Knowledge Bases in QuASE. In [Ab14] wird eine weitere Möglichkeit zur Umsetzung einer Softwarelösung innerhalb des Themengebiets der qualitätsbewussten Softwareentwicklung vorgeschlagen.

6 Zusammenfassung & Ausblick

Die Forschungsfrage in [St16] *„Wie können Ontologien und Wissensdatenbanken die Arbeit der Akteure im Softwareentwicklungsprozess erleichtern und helfen Probleme und Missverständnisse zu reduzieren?“* wird ebenda zunächst auf die beiden Kernaussagen *Stärkung der Qualitätsbeurteilung durch Ontologien und KBs* und *wie dadurch Missverständnisse reduziert und die Arbeit der Akteure erleichtert werden* aufgeteilt und separat beantwortet (vgl. [St16], Kap. 4 u. 6). Um die Software Qualität zu verbessern gibt es bereits seit Jahrzehnten verschiedene Ansätze, wie die Autoren in [Ma95] [Wa01], [Ho13] oder [De01] ausführlich beschreiben. Der Einsatz von Ontologien und Wissenswiederverwendung in diesem Spektrum ist jedoch im Zusammenhang mit Software Qualität und dessen Beurteilung ein relativ junges Forschungsgebiet (vgl. [St16] S. 90 ff.). Die Beteiligten in einem Softwareentwicklungsprojekt müssen zumeist vermehrten Aufwand in administrative Tätigkeiten aufwenden, der mit elektronischen Hilfsmitteln deutlich reduziert werden könnte.

Diese Hilfsmittel, zu der auch QuASE gezählt werden darf, können allerdings nicht nur Vorteile bringen, sondern bedürfen auch einem höheren Aufwand durch Einarbeitung und Pflege der Projektdatenbanken, damit auf längere Zeit gesehen die Vorteile deutlich überwiegen können. Zudem können Synergien zu bereits bestehenden Werkzeugen und Softwaresysteme gebildet werden, die ähnlich gelagerte Probleme beschreiben und lösen können. Auch im Hinblick auf die Weiterentwicklung der IT in Industrie 4.0, Internet der Dinge und Automatisierung von Maschinen kann mit solchen Decision-Support-Systemen die Wissenschaft und Forschung vorantreiben und weitere Anwendungsmöglichkeiten gefunden werden (vgl. [St16], S. 84-86, 103 ff.).

Literaturverzeichnis

- [Ab14] Abufouda, M.: "Quality-aware Approach for Engineering Self-adaptive Software Systems". Kaiserslautern, 2014.
- [CP09] Chung, L., do Prado Leite, J.: "On Non-Functional Requirements in Software Engineering", Mylopoulos Festschrift, LNCS 5600, p. 16, 2009.
- [De01] DeMarco, T.: Spielräume. München, Wien: Carl Hanser, 2001.
- [GI16] GI, Gesellschaft für Informatik e.V., <https://www.gi.de/themen/grand-challenges-der-informatik.html>, Stand: 16.04.2016
- [Gu05] Guizzardi, G.: "Ontological foundations for structural conceptual models," PhD Dissertation, Centre for Telematics and Information Technology, University of Twente, Enschede (NED), 2005.
- [Ho13] Hoffmann, D. W.: Software-Qualität, 2nd Edition. Berlin, Heidelberg: Springer, 2013.
- [ISO00] Information technology – Software product quality – Part 1: Quality Model, ISO/IEC ISO/IEC 9126-1:2001, 2000/2001.
- [Ma95] Mayr, H.: "Software Qualität: Nur eine Frage des Software Engineering?," in ISO 9000 - Softwareentwicklung ; Ethik, Analysen, Tools ; Beiträge vom adi QM/IT Expertentreffen 1994, ed. Münster (GER) [u.a.]: Norbert Ruppenthal, 1995, p. 49 ff.
- [MS14] Mayr, H.; Shekhovtsov, V.: QuASE - About the Project, <http://quase-ainf.aau.at/> Stand: 05.05.2016
- [Sh15] V. Shekhovtsov, H.C. Mayr, S. Ianushkevych, M. Kucko, V. Lubenskiy, and S. Strell: Implementing Tool Support for Effective Stakeholder Communication in Software Development – A Project Report. In: Ausgewählte Beiträge zur Anwenderkonferenz für Softwarequalität Test und Innovation - ASQT 2014, books@ocg.at, Vol.310. Österreichische Computer Gesellschaft, Wien, 2015, pp. 45-58.
- [SMK15] V. Shekhovtsov, H.C. Mayr, C. Kop: Facilitating Effective Stakeholder Communication in Software Development Processes. In: Nurcan, S., Pimenidis, E. (eds.): Information Systems Engineering in Complex Environments, Lecture Notes in Business Information Processing 204, Springer, 2015, pp. 116-132
- [St14] Strell, S.: „Der Qualitätsbegriff im Requirements und Software Engineering“ (Seminararbeit), unpublished.
- [St16] Strell, S.: Ontologien in Quality-Aware Software Engineering. Master thesis, Klagenfurt am Wörthersee, 2016.
- [Wa01] Wallmüller, E.: Software-Qualitätsmanagement in der Praxis. München; Carl Hanser, 2001.

Citrus – Integrationsframework mit Demonstration

Christoph Deppisch¹

Abstract: Citrus ist ein Open Source Framework für den automatisierten Test von Schnittstellen einer modernen Enterprise Anwendung. Das Framework ist in Java geschrieben und bietet fertige Komponenten für den Nachrichtenaustausch als Client und Server mit verschiedensten Protokollen und Formaten (z.B. Http REST, SOAP, JMS, FTP, File, Websocket, XML, JSON und vieles mehr). Das Framework bietet eine XML sowie Java DSL für das Schreiben von voll automatisierten Integrationstests, die sich einfach in den Continuous Build Lifecycle einbinden lassen. Der Vortrag zeigt eine kurze Einführung in das Framework und demonstriert die Arbeitsweise anhand eines Live-Code Beispiels.

Keywords: Automatisiertes Testen, Schnittstellentests, Integrationstests, Open Source, Messaging, Microservices

1 Einleitung

Softwarelandschaften bestehen heutzutage nicht mehr aus wenigen schwergewichtigen Monolithen, sondern aus vielen eigenständigen Komponenten und Anwendungen. Die Anzahl der an einem Geschäftsprozess beteiligten Softwarekomponenten steigt zunehmend an. Trends wie Microservices zeigen uns wie man schwergewichtige Alleskönner in mehrere kleinere Services aufteilt und so die Vorteile von mehr Flexibilität und schnelleren Releasezyklen ausschöpfen kann.

Mit diesem Trend hin zur verteilten Anwendung steigt natürlich auch die Anzahl der beteiligten Schnittstellen. Alle Komponenten tauschen Daten über wohl definierte Schnittstellenverträge miteinander aus. Diese nachrichtenbasierten Schnittstellen (asynchron wie synchron) nehmen einen zunehmend essentiellen Platz in der Softwareentwicklung ein.

Nahezu jede Software besitzt Schnittstellen nach Außen zu Partnern und anderen Komponenten im Unternehmen. Diese Kommunikation mit dem zugehörigen Datenaustausch muss automatisiert getestet werden, um eine verlässliche Aussage über die Qualität des gesamten Bildes von mehreren zusammenarbeitenden Services zu treffen.

¹ ConSol Software GmbH, Development, Franziskanerstr. 38, D-81669 München,
christoph.deppisch@consol.de

2 Herausforderungen im Integrationstest

Unit-Tests sind in der Regel klein und schnell in der Ausführung. In voller Isolation zu anderen Einheiten (Klassen, Methoden) wird ein Code-Abschnitt durchlaufen und überprüft. Im Integrationstest ist diese Isolation nicht mehr gegeben. Hier wollen wir gerade die Zusammenarbeit mehrerer Komponenten im Test nachstellen und überprüfen. Daher hat der Integrationstest naturgemäß Abhängigkeiten und eine komplexere Infrastruktur als Grundlage.

Die Software wird zum Beispiel im Vorfeld des Tests in einem Application Server deployed. Eventuell wird ein Message Broker oder ein Webserver benötigt. Schnittstellenpartner stehen während des Tests nicht zur Verfügung und müssen simuliert werden. Das alles stellt Entwickler und Tester vor einige Herausforderungen, ins Besondere wenn an einem Geschäftsprozess mehrere Schnittstellen unterschiedlicher Art beteiligt sind.

3 Citrus Framework

Das Citrus Open Source Framework für den automatisierten Integrationstest hat sich zum Ziel gesetzt die zuvor angesprochenen Herausforderungen im Schnittstellentest zu meistern. Mit fertigen Komponenten für unterschiedliche Nachrichten und Transportwege bietet das Framework alles, um den Datenaustausch zwischen verschiedenen Services zu testen.

Als Client oder Server verbindet sich Citrus mit dem System Under Test (SUT) und verschiebt Nachrichten über die wirklichen Transportwege (z.B. Http, JMS, FTP, ...). Das SUT wird dabei möglichst produktionsnah gestartet und arbeitet mit der zu testenden Schnittstelle. So können die Regeln der Schnittstelle (WSDL, XSD) im Test überprüft werden und auch die fachlich korrekte Arbeitsweise der Services sichergestellt werden.

Citrus ist nämlich in der Lage alle Nachrichten des SUT mit erwarteten Inhalten zu überprüfen. Dabei bietet das Framework fertige Validatoren für XML, JSON, XPath, JsonPath und HTML. Die ausgetauschten Inhalte werden damit überprüft und auf Regression getestet.

Citrus ist als Open Source Framework unter der Apache Lizenz 2.0 frei verfügbar. Nähere Informationen zum Framework selbst sind in [CF16] abrufbar. Der freie Code des Frameworks ist unter [Gi16] verfügbar.

4 Das Beispiel

Der Vortrag demonstriert die Arbeitsweise von Citrus an einem Code Beispiel. Im

Beispiel werden mehrere Schnittstellen einer verteilten Microservices Anwendung in einem automatisierten Test überprüft. Dabei kommen unterschiedliche Transportwege (z.B. JMS, Http REST) zum Einsatz und die Validierung von XML und JSON Nachrichteninhalten wird exemplarisch vorgeführt.

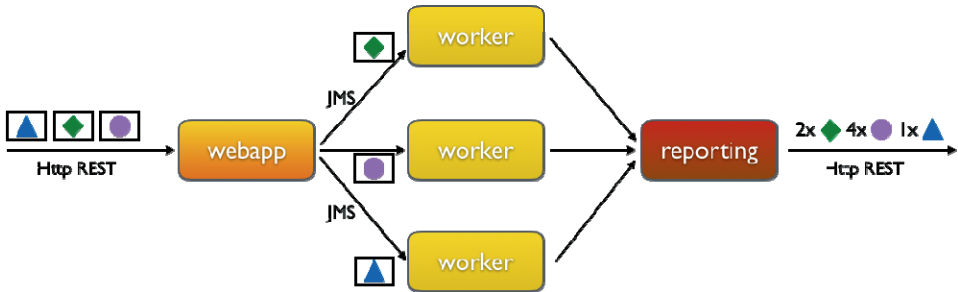


Abb. 1: Beispiel Anwendung

Die zu testende Anwendung besteht aus mehreren Komponenten, welche Nachrichten untereinander austauschen. Eine Web-Anwendung nimmt Aufträge über eine Http REST Schnittstelle entgegen und verteilt diese, je nach Art des Auftrags, an so genannte Worker Komponenten. Nachdem ein Worker seine Arbeit erfolgreich beendet hat, teilt er dies einem zentralen Reporting Server mit, so dass man die Aufträge eines Tages im Reportüberblick verfolgen kann.

Im Vortrag wird ein voll automatisierter Test dieser Vorgänge gezeigt. Der Test baut dabei die benötigte Infrastruktur auf und versendet echte Nachrichten über die Transportwege. Die ausgetauschten Nachrichteninhalte werden automatisch syntaktisch und semantisch überprüft, damit die Integrität aller Schnittstellen im gesamten Bild gewährleistet ist.

5 Über die Autoren

Dipl. Inf (FH) Christoph Deppisch arbeitet als Consultant und Software Architekt bei der ConSol Software GmbH und verfügt über mehr als zehn Jahre Erfahrung bei der Umsetzung großer Enterprise Projekte. Er hat sich auf die Bereiche Middleware Integration und Testautomatisierung spezialisiert und interessiert sich besonders für neue Technologien und Frameworks.

Als aktiver Open Source Entwickler ist Christoph verantwortlich für das Test Framework „Citrus“ und entwickelt dies stetig weiter. In letzter Zeit beschäftigt sich Christoph vor allem mit der Frage welche Einflüsse Microservices und Container auf die Continuous Delivery Pipeline haben.

Literaturverzeichnis

- [CF16] Citrus Framework, <http://citrusframework.org>, Stand: 30.06.2016.
[Gi16] Github, <http://github.com/christophd/citrus>, Stand: 30.06.2016.

Software-Engineering für Multi-Plattform-Umgebungen

Software-Engineering für Multi-Plattform-Umgebungen

Ruth Breu¹, Wolfgang Glock², Gabriele Haller³, Thomas Matzner⁴, Gerhard Müller⁵,
Friederike Nickl⁶, Oliver Wiegert⁷

Entwickler von Softwareanwendungen stehen in zunehmendem Maß vor der Aufgabe, nicht nur für eine homogene Zielplattform (etwa Programmiersprache, Frameworks, Datenbank) zu entwickeln, sondern für mehrere. Anlass dazu kann etwa sein:

- die Anforderung, Teile der Anwendung zur Steigerung der Benutzerfreundlichkeit auf dem Endgerät des Benutzers zu betreiben, Teile davon aus Sicherheitsgründen auf einem Server,
- die Modernisierung von Legacy-Anwendungen durch die Anbindung neuer Benutzeroberflächen,
- Microservices, bei denen die Plattformscheidung den einzelnen Entwicklerteams freigestellt wird, die jedoch nach außen als kohärente Anwendung erscheinen müssen,
- das Zusammenwirken von Anwendungskomponenten, für die unterschiedliche Plattformen technisch sinnvoll sind, etwa operative Daten und Funktionen zusammen mit Business Intelligence,
- die begrenzte Lebensdauer mancher Plattformen und die Unsicherheit, was als nächstes kommt.

Die traditionellen Grundsätze des Software Engineering gehen auf solche Anforderungen kaum ein; es gibt für sie keine geschlossene Lehre. Deshalb wollen wir Überlegungen und Erfahrungen von in der Praxis tätigen Software-Entwicklern und -Architekten zu diesem Aufgabenfeld zusammentragen.

Der Workshop ist eine Veranstaltung des Netzwerks „Software Engineering Live“ (selive.org). Inhaltlicher Schwerpunkt des Netzwerks ist die Diskussion und der Erfahrungsaustausch zwischen Software-Architekten und Methodenverantwortlichen über den Einsatz von Techniken und Methoden des Software Engineering zur Beherrschbarkeit der Komplexität bei der Entwicklung und Evolution großer IT-

¹ Universität Innsbruck, QE Lab, Technikerstraße 21a, 6020 Innsbruck, ruth.breu@uibk.ac.at

² Landeshauptstadt München, Direktorium, Marienplatz 8, 80331 München, wolfgang.glock@muenchen.de

³ gt-muenchen GmbH, Josef-Beiser-Straße 28, 81737 München, info@gt-muenchen.de

⁴ Berater für Systemanalyse, Beltweg 22, 80805 München, tam@tamatzner.de

⁵ TNG Technology Consulting GmbH, Beta-Straße 13, 85774 Unterföhring, gerhard.mueller@tngtech.com

⁶ Swiss Life AG, IT-KAP, Zeppelinstraße 1, 85748 Garching, friederike.nickl@swisslife.de

⁷ iteratec GmbH, Inselkammerstraße 4, 82008 Unterhaching, oliver.wiegert@iteratec.de

Systeme.

Da es sich bei den eingereichten Beiträgen um Impulsvorträge für eine ausführliche Diskussion handelt, wurden keine Papers voller Länge gefordert, sondern lediglich Exposés. Bei der Begutachtung wurde entschieden, ob die Beiträge thematisch und inhaltlich geeignet sind, eine solche Diskussion auszulösen.

Folgende Vorträge wurden angenommen:

- Leonid Kof: One size does not fit all – aber mit wenigen Standardgrößen kann man viel abdecken.
- Thomas Matzner: Architekturmuster für Informationssysteme mit Browser-Oberfläche

Suchmaschinen: Zwischen technischer Entwicklung und gesellschaftlicher Bedeutung

Suchmaschinen: Zwischen technischer Entwicklung und gesellschaftlicher Bedeutung

Olaf Resch¹

Vorwort

Das Internet ist eine wesentliche Informationsquelle, die das Entscheidungsverhalten von Menschen in einem hohen Maße mitbestimmt. Suchmaschinen machen das Auffinden dieser entscheidungsrelevanten Informationen häufig erst möglich und nehmen somit eine Schlüsselposition ein. Suchmaschinen treten jedoch nicht nur als Universal-Suchmaschinen auf, sondern auch in Form von Produktsuchmaschinen oder als Suchfunktion für einzelne Websites oder Intranets. Die Suchergebnisse werden zunehmend auch nicht mehr nur in Form der klassischen SERP (Search Engine Result Page) dargestellt, sondern als Service von anderen Anwendungen konsumiert oder sie bilden die Grundlage für persönliche Assistenten.

Der Workshop zeigt insbesondere die Vielfalt der Suchmaschinenlandschaft. Es werden entsprechende Anwendungen speziell für die Suche in Bildern oder in juristischen Dokumenten und als integraler Bestandteil des Wissensmanagements vorgestellt. Diese Anwendungen verfolgen unterschiedliche Ziele und nutzen eigene Geschäftsmodelle und stellen somit echte hochspezialisierte Alternativen insbesondere zu dem Mainstream-Suchmaschinen dar.

Programmkomitee: Michael Christen, Stephan G. Humer, Arne Kunisch, Christian Kunz, Dirk Lewandowski, Philipp Mayr, Vivien Petras, Olaf Resch, Rainer Rumpel, Wolfgang Sander-Beuermann, Andreas Schmietendorf

¹ Hochschule für Wirtschaft und Recht Berlin, Informationsmanagement und E-Business, Alt-Friedrichsfelde 60, 10315 Berlin, olaf.resch@hwr-berlin.de

Die Suchergebnisseite als Dauerwerbesendung?

Dirk Lewandowski¹

Abstract: Die Trefferdarstellung auf den Suchergebnisseiten von Suchmaschinen wie Google ist in den letzten Jahren immer komplexer geworden; es hat ein Wandel weg von der Darstellung einer Ergebnisliste hin zu einer komplexen Trefferpräsentation, die sich aus unterschiedlichen Elementen wie organischen Ergebnissen, Textanzeigen, verschiedenen Formen von Universal-Search-Ergebnissen und Fakteninformationen zusammensetzt. Im Zusammenspiel mit der erweiterten Funktion von Suchmaschinen, die von reinen Vermittlern zwischen Nutzern und externen Informationen zu Vermittlern und gleichzeitig Inhalteanbietern geworden sind, stellt der Beitrag die Frage nach einer für den Nutzer nachvollziehbaren Kennzeichnung von nicht durch allein „redaktionelle“ Entscheidungen platzierten Treffern und kommt zu dem Schluss, dass heutige Suchergebnisseiten aufgrund der Vielfalt der für den Nutzer nicht erkennbaren Interessen hinter der Platzierung einzelner Treffer(blöcke) analog zu den Dauerwerbesendungen im Fernsehen behandelt werden und entsprechend gekennzeichnet werden sollten.

Keywords: Suchmaschinen, Suchergebnisseite, Google, Universal Search, Search Engine Advertising (SEA), Suchmaschinenoptimierung (SEO)

1 Einleitung

Suchmaschinen haben eine enorme Bedeutung; einerseits für die Vermittlung von Informationen bzw. die Sichtbarkeit von Angeboten im Web, andererseits als tragende Säule des Online-Werbemarkts. Google kommt in Bezug auf seine Marktmacht eine besondere Bedeutung zu, ohne dass bisher allerdings erkennbar ist, dass das Unternehmen aus dieser Position auch eine besondere Verantwortung für sich ableitet [Le16a].

In diesem Beitrag wird die Darstellung auf den Suchergebnisseiten aus der Perspektive der Trennung zwischen „redaktionellem Inhalt“ und Werbung diskutiert. Dabei wird insbesondere berücksichtigt, dass es sich bei Suchwortwerbung um eine besondere Werbeform handelt, die als Ergebnis zu einer Suchanfrage angezeigt wird, und dass Suchmaschinen ein Interesse daran haben, ihre eigenen Inhalte auf den Suchergebnisseiten zu platzieren. Dies ist vor allem von Bedeutung, da sich gezeigt hat, dass Nutzer nur unzureichend in der Lage sind, die Herkunft der Treffer zuverlässig zu erkennen.

Bereits in der Vergangenheit gab es immer wieder Diskussionen über die Trennung von Ergebnissen und Werbung auf den Suchergebnisseiten. In diesem Beitrag werde ich

¹ Hochschule für Angewandte Wissenschaften Hamburg, Department Information, Finkenau 35, 22081 Hamburg, dirk.lewandowski@haw-hamburg.de

allerdings argumentieren, dass die Frage nach einer einfachen Trennung zwischen zwei Arten von Elementen auf den Suchergebnisseiten falsch gestellt ist: Unter den Voraussetzungen einer immer komplexer werdenden Ergebnispräsentation und den mannigfaltigen Einflüssen auf das Zustandekommen der Suchergebnisse (bzw. allgemeiner: der Platzierung der Elemente auf der Suchergebnisseite) ist eine solche Trennung nicht ausreichend. Stattdessen sollte die Suchergebnisseite als Ganzes betrachtet werden: Da Nutzer nicht nachvollziehen können, wie die Platzierung der Elemente zustande kommt und eine größere Transparenz von Seiten der Suchmaschinenbetreiber (vor allem: von Google) nicht zu erwarten ist, sollte die *Suchergebnisseite als Ganzes* analog den Dauerwerbesendungen im Fernsehen gekennzeichnet werden.

2 Stand der Forschung

In den folgenden Abschnitten wird der Stand der Forschung zur Bedeutung der Suchmaschinenwerbung, zu den Elementen der Suchergebnisseiten, zum Auswahlverhalten auf den Ergebnisseiten und zur bevorzugten Darstellung bestimmter Elemente durch die Suchmaschinen beschrieben.

2.1 Suchmaschinenwerbung als bedeutende Säule der Onlinewerbung

Onlinewerbung wird in der Regel in klassische Onlinewerbung (v.a. Banneranzeigen), Suchwortvermarktung und Affiliate-Werbung unterteilt; dabei entfällt auf die Suchwortvermarktung mehr als die Hälfte der Werbeerlöse [KM15]. Bei einem stetig steigenden Suchanfragevolumen ist auch mit einem weiteren Wachstum dieser Werbeform zu rechnen [LE16b]. Suchwortvermarktung zeichnet sich vor allem dadurch aus, dass (1) der Suchende mit seiner Suchanfrage sein Interesse offenbart und die Werbung passend zur Suchanfrage angezeigt wird, (2) die Abrechnung gegenüber dem Anzeigenkunden nach Klicks erfolgt, (3) die Anzeigen textbasiert sind und damit als weniger störend empfunden werden als grafische Werbung, (4) das Versteigerungsverfahren, nach dem die Anzeigenpreise bestimmt werden, transparent ist, und (5) Anzeigenkampagnen leicht selbst erstellt werden können, auch wenn das Budget gering ist (vgl. [LE15], S. 150).

Durch diese Eigenschaften der Suchwortwerbung ergibt sich keine direkte Vergleichbarkeit mit Anzeigen in Zeitschriften oder der Werbung in den Massenmedien. Insofern erscheint auch der im Englischen häufig verwendete Begriff „sponsored links“ als passender als die deutsche Bezeichnung „Anzeigen“, da hier eher zum Ausdruck gebracht wird, dass es sich bei den Anzeigen um eine Form von Suchergebnissen handelt, nur eben um solche, für deren Darstellung bezahlt wurde.

2.2 Elemente der Suchergebnisseite

Auf den Ergebnisseiten der Suchmaschinen haben sich vier Ergebnistypen herausgebildet:

- „Organische Ergebnisse: Dabei handelt es sich um die echten Suchergebnisse, d. h. um aus dem Web-Index mittels Algorithmen generierte und durch die Rankingverfahren in eine Reihenfolge gebrachten Ergebnisse, wobei die Algorithmen alle Dokumente im Index gleich behandeln.
- Werbung (nach Googles Bezeichnung auch AdWords): Hierbei handelt es sich um kontextbasierte Textanzeigen, die passend zu einer Suchanfrage angezeigt werden und in ihrer Darstellung organischen Treffern ähneln (Titel, Beschreibung, URL-Angabe). Die Textanzeigen können daher auch als eine besondere Form von Suchergebnissen angesehen werden.
- Universal-Search-Ergebnisse: Hierunter sind Treffer zu verstehen, die nicht aus dem allgemeinen Web-Index kommen, sondern aus gesondert aufgebauten Kollektionen. Solche Treffer werden in der Regel innerhalb der Liste der organischen Treffer platziert; die Blöcke mit solchen Ergebnissen unterbrechen die Liste der organischen Suchergebnisse. Die Trefferdarstellung weicht in der Regel von der der organischen Ergebnisse ab und ist auf die jeweilige Kollektion angepasst. So werden beispielsweise Video-Ergebnisse mit einem Vorschaubild (anstatt der üblichen textuellen Beschreibung) präsentiert.
- Fakteninformationen: Suchmaschinen zeigen zunehmend direkte Antworten zu geeigneten Suchanfragen an.“ ([Le15] S. 126)

Abb. 1 und Abb. 2 zeigen beispielhaft die Suchergebnisseiten von Google für zwei unterschiedliche Suchanfragen, einmal in der Darstellung auf einem Desktop-Computer, einmal auf einem Mobiltelefon. Diese Seiten werden in der weiteren Diskussion als Beispiele für die beschriebenen Phänomene herangezogen.

Eine interessante Verdeutlichung der Vielzahl der mittlerweile auf den Suchergebnisseiten von Google vorhandenen Elemente zeigt die (künstlich erstellte) “Mega-SERP”.²

² <https://moz.com/blog/mega-serp-a-visual-guide-to-google>

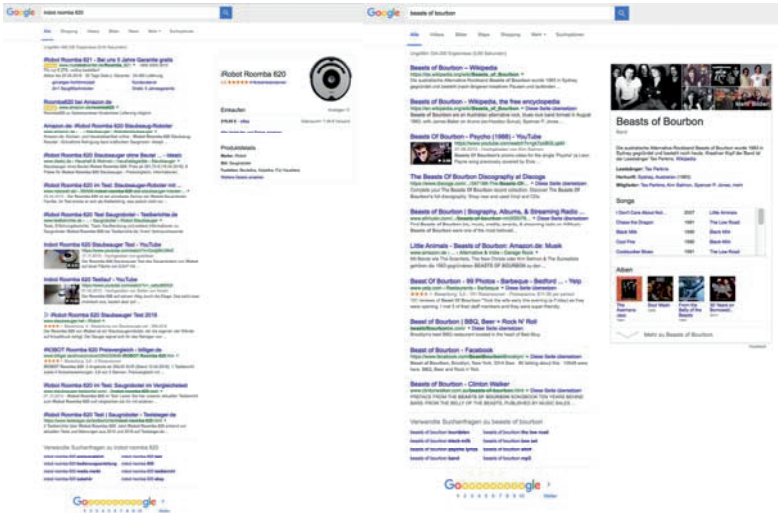


Abb. 1: Suchergebnisseiten von Google für die Suchanfragen „irobot roomba 620“ und „beasts of bourbon“ (19.4.2016)

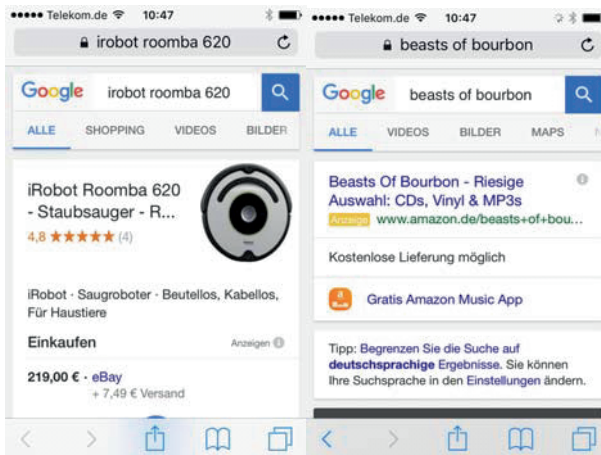


Abb. 2: Mobile Darstellung der Suchergebnisseiten

2.3 Auswahlverhalten auf den Suchergebnisseiten

Zunächst einmal beeinflusst die Darstellung der Suchergebnisseiten das Blickverhalten der Nutzer. Elemente, die auf den vorderen Positionen platziert oder durch grafische

Elemente besonders hervorgehoben sind, ziehen die Blicke der Nutzer verstärkt an (u.a. [Li15], [Pa07]). Ähnliches lässt sich für das Selektionsverhalten feststellen; auch hier gibt es eine starke Präferenz für die vorderen Treffer und solche, die besonders hervorgehoben sind, also v.a. Universal-Search-Ergebnisse. Man kann die Darstellung auf den Trefferseiten daher im Spannungsfeld zwischen einer technischen Unterstützung des Suchenden und einer Manipulation durch den Suchmaschinenbetreiber sehen (vgl. [LKS14]).

2.4 Diskussion um die Bevorzugung bestimmter Suchergebnisse

Die Diskussion um die Darstellungen auf den Suchergebnisseiten bezieht sich im Wesentlichen auf den Vorwurf an Google, seine eigenen Angebote im Rahmen der Universal Search zu bevorzugen. Die Europäische Kommission untersucht bereits seit einigen Jahren die Bevorzugung von Googles Shopping-Ergebnissen vor anderen Anbietern. Allerdings beschränkt sich das Problem nicht auf die Shopping-Ergebnisse, sondern tritt in allen Bereichen zutage, in denen Suchmaschinen selbst vertikale Suchdienste anbieten, die in Konkurrenz zu anderen spezialisierten Suchmaschinen stehen, bzw. in allen Bereichen, in denen Suchmaschinen selbst Inhalte anbieten, anstatt „nur“ als Vermittler zwischen Suchenden und externen Inhalten auftreten.

Das Problem der Bevorzugung eigener Angebote verstärkt sich im Fall von Google aufgrund seiner enormen Marktmacht.

3 Analyse der Ergebnistypen auf den Suchergebnisseiten

Das Beispiel Google Shopping zeigt bereits eine grundlegende Problematik auf, nämlich dass Suchmaschinen nicht mehr nur Vermittler zwischen Nutzern und externen Informationsobjekten sind, sondern selbst Inhalte anbieten, sei dies in direkter Form (bspw. betreibt Google das Videoportal YouTube) oder in indirekter Form, wenn eine Suchmaschine auf eigene vertikale Dienste verweist (wie eben im Fall von Google Shopping). Im Folgenden wird diskutiert, inwieweit es sich bei den unterschiedlichen Ergebnistypen um „echte“ Suchergebnisse handelt.

3.1 Sind die organischen Ergebnisse „organisch“?

Zunächst einmal besteht kein direkter Einfluss externer Inhaltenanbieter auf die organischen Ergebnisse. Es ist also weder möglich, gegen Bezahlung in den Index aufgenommen zu werden (sog. Paid Inclusion, vgl. [LE16b]), noch gegen Bezahlung eine bevorzugte Position im Ranking zu erhalten. Möglichkeiten zur Beeinflussung der organischen Suchergebnisse gibt es dennoch, und zwar in Form von Suchmaschinenoptimierung. Darunter „versteht man alle Maßnahmen, die dazu geeignet sind, die Position von Webseiten im Ranking der Suchmaschinen zu verbessern.“

([Le15] S.161) Es geht hier also um die Optimierung auf der Seite der Anbieter von Inhalten, die schließlich mittels Suchmaschinen aufgefunden werden sollen. Dabei fokussiert Suchmaschinenoptimierung wegen seiner Marktstellung vor allem auf Google, auch wenn die getroffenen Maßnahmen sich auch positiv auf das Ranking in anderen Suchmaschinen auswirken. Gerade Googles Marktmacht sorgt aber auf der einen Seite dafür, dass sich durch Suchmaschinenoptimierung enorme Trafficgewinne erzielen lassen – wer am besten positioniert ist, erhält eine Vielzahl von Klicks –, auf der anderen Seite dafür, dass die Konkurrenz enorm ist, da ja faktisch nur um Platzierungen bei einem einzigen Anbieter konkurriert wird.

Diese Situation hat dazu geführt, dass insbesondere die Ergebnisse zu hochgradig kommerziell ausnutzbaren Suchanfragen durchweg „optimiert“ sind. Klassische Beispiele sind Suchanfragen wie „Kredit ohne Schufa“ oder „Private Krankenversicherung Vergleich“, denen gemeinsam ist, dass sich allein durch die Vermittlung eines entsprechenden Vertrags hohe Umsätze erzielen lassen und auf Seiten des Vermittlers wenig Aufwand nötig bzw. Infrastruktur vorhanden sein muss.

Suchmaschinenoptimierung hat sich zu einer eigenen Branche entwickelt und bildet eine der wesentlichen Säulen des Online-Marketings. Für das Jahr 2016 werden allein für die USA Ausgaben von 65 Milliarden Dollar für Suchmaschinenoptimierung erwartet [Su16]. Ihr Einfluss reicht weit über die Optimierung von Seiten für Produkte und Dienstleistungen hinaus. Zu nennen ist hier vor allem der Nachrichtenbereich, der als Vorreiter in Bezug auf die Optimierung informativer Inhalte gelten kann. Da Nachrichten im Web in aller Regel kostenlos angeboten und durch Werbung refinanziert werden, besteht ein hoher Druck, die Besucherzahlen auf den Nachrichtenangeboten zu erhöhen. Dies hat nicht nur zu einem starken Wettbewerb unter den Nachrichtenanbietern geführt, sondern auch die Art, in der Nachrichten geschrieben und publiziert werden, verändert. Der Titel einer bereits im Jahr 2007 publizierten Studie der Friedrich-Ebert-Stiftung fasst die Tendenzen gut zusammen: „Klicks, Quoten und Reizwörter“ [RS07].

Es zeigt sich, dass Suchmaschinenoptimierung immer weitere Bereiche erfasst und vor allem in Hinblick auf die informativen Inhalte ein Problem darstellt, insbesondere, wenn man die hochgradige Optimierung auf der einen Seite mit der eher naiven Vorstellung der Nutzer von vertrauenswürdigen und „neutralen“ Suchergebnissen [PBR12], s.a. [SMJ14] kontrastiert. Das Potential der Suchmaschinenoptimierung vor allem für den Bereich der Public Relations ist bei weitem noch nicht ausgeschöpft, auch wenn sich schon heute bei „verdächtigen“ Suchanfragen wie „Atomkraft“ eine erhebliche Tendenz zu interessengeleiteter Kommunikation auf den vorderen Ergebnisrängen zeigt. Systematische Untersuchungen dazu stehen allerdings noch aus.

Zusammengefasst lässt sich sagen, dass organische Suchergebnisse zwar nicht direkt durch Bezahlung an die Suchmaschinenbetreiber beeinflusst werden können, dass Einflussmöglichkeiten aber durch „Optimierungsmaßnahmen“, d.h. Anpassungen der eigenen Inhalte an das von der Suchmaschine Gewünschte, bestehen. Während sich auf

der einen Seite eine hochgradig professionalisierte Branche von Suchmaschinenoptimierern herausgebildet hat, steht dieser auf der anderen Seite eine großteils naive Nutzerschaft gegenüber. Während also per Definition im Ranking der organischen Treffer alle Informationsobjekte die gleiche Chance haben, zuvorderst gerankt zu werden, können die Voraussetzungen für ein gutes Ranking durch SEO-Maßnahmen erheblich beeinflusst werden.

3.2 Werden Anzeigen als solche erkannt?

Die Einordnung der Textanzeigen auf den Suchergebnisseiten ist vermeintlich einfach: Anzeigen werden als solche gekennzeichnet und können dadurch von Suchergebnissen unterschieden werden. Allerdings zeigen Studien, dass Nutzer durchaus Schwierigkeiten haben, zwischen Anzeigen und organischen Suchergebnissen zu unterscheiden. Dies wurde bereits 2009 in einer kleineren Studie des Bundesverbands Digitale Wirtschaft gezeigt [Bu09]. Diese Befunde bestätigten sich in einer repräsentativen Studie mit 1.000 Teilnehmern, in der Nutzer auf Suchergebnisseiten jeweils die Werbung bzw. die organischen Ergebnisse markieren sollten [Le16c]. Es wurde gezeigt, dass Nutzer zum einen wenig über die Einnahmequellen der Suchmaschine Google und die Positionierung der Werbung wissen, und zum anderen, dass sie nur zu einem geringen Teil in der Lage waren, zuverlässig zwischen Werbung und organischen Ergebnissen zu unterscheiden. Letzteres zeigte sich noch verstärkt, wenn auf den Suchergebnisseiten Universal-Search-Ergebnisse gezeigt wurden.

Woran liegt nun die mangelnde Unterscheidungsfähigkeit der Nutzer? Es bieten sich verschiedene Erklärungen an:

1. Bei den Anzeigen handelt es sich um eine Form von Suchergebnissen: Anzeigen werden als Antwort auf eine Suchanfrage ausgegeben und können zu dieser Suchanfrage relevant sein. In vielen Fällen mag es für einen Nutzer keine Rolle spielen, ob er auf eine Anzeige oder auf ein organisches Suchergebnis klickt, da beides sein Informationsbedürfnis befriedigen kann.
2. Anzeigen und organische Ergebnisse werden ähnlich dargestellt: Sowohl Anzeigen als auch organische Ergebnisse werden auf den Suchergebnisseiten ähnlich beschrieben, die „Snippets“ enthalten jeweils eine klickbare Überschrift, eine (teils abgekürzte) URL und eine kurze Beschreibung. Auch die farbliche Gestaltung ist jeweils gleich; die einzige Unterscheidung besteht in der Kennzeichnung als „Anzeige“, welche bei Google durch eine gelbe Markierung erfolgt (s. Abb. 1, linkes Bild, und Abb. 2, rechtes Bild).

Insbesondere die mangelnde bzw. die von den Nutzern zumindest teils nicht wahrgenommene Kennzeichnung der Anzeigen war immer wieder Gegenstand von Beschwerden (s. u.a. (‘‘FTC Consumer Protection Staff Updates Agency’s Guidance to Search Engine Industry on the Need to Distinguish Between Advertisements and Search Results | [Ft13])). Obwohl die genauen Effekte unterschiedlicher Kennzeichnungsarten

bislang nicht systematisch untersucht wurden, wurde doch gezeigt, dass zumindest die seit 2014 veränderte Kennzeichnung der Anzeigen bei Google (von einer Schattierung zu der Kennzeichnung mit der gelben Markierung) zu einer Erhöhung der Zahl der Klicks auf die Anzeigen geführt hat [Ed14]. Und letztlich ist es ja das Interesse von Google und anderen Suchmaschinen, möglichst viele Klicks auf die Anzeigen zu generieren: Die Anzeigen stellen das wesentliche – angesichts der Umsatzzahlen kann man auch sagen: das einzig nennenswerte – Geschäftsmodell von Google dar [LE16a]. Jede Erhöhung der Anzeigenklicks führt zu höheren Umsätzen. Solange eine Suchmaschine in der Lage ist, auch über die Anzeigen ihren Nutzern zumindest vermeintlich relevante Ergebnisse zu liefern, bietet es sich an, die weit verbreitete Unkenntnis der Nutzer hinsichtlich der Unterscheidung zwischen Anzeigen und organischen Ergebnissen zumindest zu tolerieren.

3.3 Sind Universal-Search-Ergebnisse „organisch“?

Ähnlich wie bei der Unterscheidung zwischen organischen Ergebnissen und Textanzeigen ergibt sich bei den Universal-Search-Ergebnissen das Problem, dass für manche Platzierungen bezahlt wurde, während andere „neutral“ aus dem jeweiligen vertikalen Index der Suchmaschine generiert wurden. Dazu kommt, dass manche Ergebnistypen, die früher organisch generiert wurden, mittlerweile Anzeigen sind.

Man muss als unterscheiden zwischen (1) organischen Universal-Search-Ergebnissen und (2) bezahlten Universal-Search-Ergebnissen (Universal-Search-Anzeigen).

(1) Bei organischen Universal-Search-Ergebnissen handelt es sich um Treffer bzw. Trefferblöcke auf der Suchergebnisseite, die aus spezialisierten Indexen der Suchmaschine (wie bspw. Nachrichten, Videos, Bilder) generiert wurden, wobei alle Treffer im Index von den Algorithmen der Suchmaschine gleich behandelt werden und keine direkte Bezahlung für die Rankingposition möglich ist. In diesem Fall ergeben sich offensichtlich die gleichen Probleme, wie sie oben bei den organischen Ergebnissen beschrieben wurden; Suchmaschinenoptimierung erfolgt auch für die vertikalen Suchen.

Ein weiteres Problem ergibt sich, wenn Suchmaschinen selbst als Inhabeanbieter auftreten. Dies ist beispielsweise bei Google der Fall, welches das Videoportal YouTube betreibt, dessen Ergebnisse nicht nur in die allgemeine Google-Suche eingehen, sondern auch in Googles Video-Suchmaschine. Zwar enthält letztere weit mehr als nur Videos von YouTube, es ist allerdings zumindest ein Interessenkonflikt zu konstatieren, wenn eine Suchmaschine ihre eigenen Inhalte ranken soll. Oft finden sich auch die exakt gleichen Videos auf unterschiedlichen Plattformen. Aus der Sicht der Suchmaschine wäre zumindest dann eine Bevorzugung der Version von der eigenen Plattform wirtschaftlich sinnvoll, da mit Werbeplatzierungen rund um das Video wiederum Umsätze erzielt werden können.

Bei der Bewertung insbesondere des Verhältnisses von Google und YouTube ergeben sich allerdings erhebliche Schwierigkeiten: Natürlich ist YouTube die beliebteste

Videoplattform – allerdings stellt sich die Frage, ob sie das nicht gerade dadurch werden konnte, dass sie so häufig von Google in den Suchergebnissen (ob organische oder Universal-Search-Ergebnisse) angezeigt wurde.

(2) Bei bezahlten Universal-Search-Ergebnissen handelt es sich um analog zu den Textanzeigen markierte Anzeigen. Sie werden nicht nur bevorzugt platziert, sondern auch bevorzugt dargestellt, indem bspw. Bilder zu Produkten gezeigt werden (s. Abb. 2, linkes Bild). Gerade bei den Shopping-Ergebnissen handelt es sich aber um Ergebnisse, die früher „organisch“ waren, dann aber auf Anzeigen umgestellt wurden. Es ist mehr als fraglich, ob Nutzer diese Umstellung bemerkt haben, auch wenn die Ergebnisse nun als Anzeigen gekennzeichnet werden. Ein Wandel von organischen Universal-Search-Ergebnissen zu Anzeigen dürfte sich für anderen Ergebnisarten wiederholen; durch diesen Wechsel dürfte weitere Verwirrung darüber entstehen, ob ein Ergebnis nun organisch ist oder ob es sich um eine Anzeige handelt.

3.4 Fakteninformationen

Die Anzeige von Fakteninformationen erscheint zunächst einmal unproblematisch, da die Fakten vor allem aus Wikipedia stammen. Allerdings ist offensichtlich, dass, falls Fakteninformationen zukünftig auch aus weiteren Quellen angezeigt werden, aufgrund der (notwendigerweise) geringen Transparenz dieser Ergebnisform Möglichkeiten zur Manipulation von außen und zur Bevorzugung „eigener Fakten“ (im Sinne einer vom Suchmaschinenbetreiber bevorzugten Sicht der Dinge) gegeben sind. Es ist allerdings noch deutlich zu früh, um hier bereits einen Schluss zu ziehen.

4 Fazit und Implikationen

Aus der obigen Darstellung ergibt sich zunächst einmal, dass eine weitere (und tiefergehende) Diskussion der Rolle der Suchmaschinen als Informationsvermittler, vor allem auch in Hinblick auf die Darstellung der Suchergebnisse und die Vermischung unterschiedlicher Trefferarten auf den Suchergebnisseiten, nötig ist.

Aufgrund der beschriebenen Situation sehe ich die bisherige Kennzeichnungen auf den Suchergebnisseiten als nicht zielführend an, da

- Nutzer Anzeigen nicht von organischen Ergebnissen unterscheiden können,
- Nutzer organische Universal-Search-Ergebnisse nicht von bezahlten Universal-Search-Ergebnissen unterscheiden können,
- auf die organischen Ergebnisse ein erheblicher externer Einfluss besteht,
- erhebliche Eigeninteressen der Suchmaschinenbetreiber bestehen, ihre eigenen Inhalte in der allgemeinen Suche zu platzieren.

Der Fehler liegt nun darin, die Suchergebnisse isoliert zu betrachten und allein für eindeutig bezahlte Ergebnisse (also „echte Anzeigen“) eine Kennzeichnung vorzuschreiben. Diese Betrachtungsweise beruht auf einer Analogie zwischen klassischer Werbung und Suchmaschinenwerbung, die so nicht gegeben ist. Ich plädiere stattdessen dafür, auch diejenigen Treffer, für die (direkt oder indirekt) bezahlt wurde, als *Suchergebnisse* zu betrachten. Damit ergäbe sich eine differenziertere Unterscheidung eben nicht nur in Werbung und organische Treffer, sondern es ließe sich beispielsweise im Falle von suchmaschinenoptimierten Treffern auch ein den PR-Beiträgen in Nachrichtenmedien analoger Treffertyp definieren. Weiterhin könnte damit auch die „Eigen-PR“ der Suchmaschinenbetreiber berücksichtigt werden.

Betrachtet man die Suchergebnisseiten in ihrer heutigen Form, so kann man sie eigentlich nur den Dauerwerbesendungen im Fernsehen vergleichen. Dauerwerbesendungen “sind nach § 7 V des Rundfunkstaatsvertrages zulässig, wenn der Werbecharakter erkennbar im Vordergrund steht und die Werbung einen wesentlichen Bestandteil der Sendung darstellt. Sie müssen zu Beginn als Dauerwerbesendungen angekündigt und während ihres gesamten Verlaufs als solche gekennzeichnet werden.“ [Da16].

Dieses Konzept lässt sich auf die Suchergebnisseiten übertragen: Der Rezipient/Nutzer weiß zwar, dass er im Verlauf der Sendung / an bestimmten Stellen mit Werbung bzw. Product Placement konfrontiert werden wird, er weiß allerdings nicht, wann bzw. an welcher Stelle die Werbung bzw. das Product Placement erfolgt, und er weiß auch nicht notwendigerweise, ob eine bestimmtes Produkt bzw. Suchergebnis gezeigt wird, weil dafür bezahlt wurde, oder ob es zum normalen Ablauf der Sendung bzw. der normalen Suchergebnismenge gehört.

Literaturverzeichnis

- [Bu09] Bundesverband Digitale Wirtschaft. (2009). Nutzerverhalten auf Google-Suchergebnisseiten: Eine Eyetracking-Studie im Auftrag des Arbeitskreises Suchmaschinen-Marketing des Bundesverbandes Digitale Wirtschaft (BVDW) e.V.
- [Da16] Dauerwerbesendung. (2016). In Gabler Wirtschaftslexikon. Springer Verlag.
- [Ed14] Edelman, B. (2014). Google’s Advertisement Labeling in 2014. Retrieved from <http://www.benedelman.org/adlabeling/google-colors-oct2014.html>
- [Ft13] FTC Consumer Protection Staff Updates Agency’s Guidance to Search Engine Industry on the Need to Distinguish Between Advertisements and Search Results | Federal Trade Commission. (2013). Retrieved November 17, 2015, from <https://www.ftc.gov/news-events/press-releases/2013/06/ftc-consumer-protection-staff-updates-agencys-guidance-search>
- [KM15] Knapp, D. (IHS), & Marouli, E. (IHS). (2015). Adex Benchmark 2014 European Online Advertising Expenditure, (July).

- [Le15] Lewandowski, D. (2015). *Suchmaschinen verstehen*. Berlin Heidelberg: Springer Vieweg.
- [Le16a] Lewandowski, D. (2016a). Is Google responsible for providing fair and unbiased results? In L. Floridi & M. Taddeo (Eds.), *The Responsibilities of Online Service Providers*. Berlin Heidelberg: Springer.
- [Le16b] Lewandowski, D. (2016b). Status Quo und Entwicklungsperspektiven des Suchmaschinenmarkts. In J. Krone & T. Pellegrini (Eds.), *Handbuch Medienökonomie* (pp. 1–23). Wiesbaden: Springer Fachmedien. doi:10.1007/978-3-658-09632-8
- [Le16c] Lewandowski, D., Kerkmann, F., Rümmele, S., & Sünkler, S. (2016). Search engine ad disclosures. [Submitted].
- [LKS14] Lewandowski, D., Kerkmann, F., & Sünkler, S. (2014). Wie Nutzer im Suchprozess gelenkt werden: Zwischen technischer Unterstützung und interessengeleiteter Darstellung. In B. Stark, D. Dörr, & S. Aufenanger (Eds.), *Die Googleisierung der Informationssuche - Suchmaschinen im Spannungsfeld zwischen Nutzung und Regulierung*. Berlin: De Gruyter.
- [Li15] Liu, Z., Liu, Y., Zhou, K., Zhang, M., & Ma, S. (2015). Influence of Vertical Result in Web Search Examination. In *Proceedings of SIGIR'15, August 09 - 13, 2015, Santiago, Chile*. New York: ACM.
- [Pa07] Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication*, 12(3), 801–823.
- [PBR12] Purcell, K., Brenner, J., & Raine, L. (2012). *Search Engine Use 2012*. Search. Washington, DC.
- [RS07] Range, S., & Schweins, R. (2007). *Klicks, Quoten, Reizwörter: Nachrichten-Sites im Internet – Wie das Web den Journalismus verändert*. Berlin: Friedrich-Ebert-Stiftung.
- [SMJ14] Stark, B., Magin, M., & Jürgens, P. (2014). Navigieren im Netz – Befunde einer qualitativen und quantitativen Nutzerbefragung. In B. Stark, D. Dörr, & S. Aufenanger (Eds.), *Die Googleisierung der Informationssuche - Suchmaschinen im Spannungsfeld zwischen Nutzung und Regulierung* (pp. 20–74). Berlin: De Gruyter.
- [Su16] Sullivan, L. (2016). Report: Companies Will Spend \$65 Billion On SEO In 2016. Media Post. Retrieved May 14, 2016, from <http://www.mediapost.com/publications/article/273956/report-companies-will-spend-65-billion-on-seo-in.html>

Juristische Suchmaschine Lawsearch Enterprise – drei Herausforderungen

Franz Kummer¹ und Blaise Dévaud²

Abstract: Die juristische Suchmaschine Lawsearch Enterprise bewältigt den Informationsfluss einer juristischen Organisation (Anwaltskanzlei, Gericht, Rechtsdienst, Notariat etc.) mit einem Single Point of Search. Drei technische und organisatorische Herausforderungen stehen im Vordergrund: Auswahl und Integration der Informationsquellen; Beschaffung und Verwendung der Metainformationen; Erkennung und Verlinkung der juristischen Referenzen. Der Beitrag zeigt konkrete Lösungsansätze dieser drei Probleme und gibt dabei einen Überblick über die Funktionsweise der juristischen Suchmaschine Lawsearch Enterprise.

Keywords: Lawsearch; Suchmaschine; Recht, Wissensmanagement; Indexierung; Metadaten; Verlinkung; Parsing.

1 Einleitung

Im Jahre 2015 hat die Schweizerische Bundeskanzlei über 6'000 Seiten Rechtsvorschriften publiziert. Dies alleine entspricht 16 DIN A4 Seiten pro Tag. Dazu werden jeden Monat mehrere hundert relevante Gerichtsentscheide, juristische Beiträge in Fachzeitschriften und Monographien veröffentlicht. Ganz zu schweigen von Tausenden von Word, PDF und E-Mail Dateien, die jede juristische Organisation produziert und mehr oder weniger sorgfältig intern ablegt.

Die Bewältigung dieses Informationsflusses ist das Hauptproblem eines effizienten juristischen Wissensmanagements.

Systemisch gesehen besteht der Informationsfluss einer juristischen Organisation aus externen und internen Informationsquellen. Die externen Quellen sind insbesondere die Gesetzgebung, die Rechtsprechung und die juristische Literatur; die internen Quellen bilden die Büro-Dokumente, E-Mails zu den behandelten Rechtsfällen sowie eine allfällige eigene Bibliothek. Alle diesen Quellen sind stark heterogen: Gesetzgebung entsteht auf verschiedenen Staatsebenen; Rechtsprechung wird durch verschiedene Gerichte publiziert; und Literatur – wenn überhaupt digital verfügbar – ist meistens in bezahlten Systemen einzelner Verlage abgelegt (Paywall). Auch die internen Informationen lassen

¹ Jurist; Lehrbeauftragter am Kompetenzzentrum Informatik und Recht, Universität Bern; Gründer und Inhaber der Weblaw AG; Weblaw AG, Laupenstrasse 1, 3008 Bern, Schweiz; www.weblaw.ch; franz.kummer@weblaw.ch.

² Jurist und Wirtschaftsinformatiker; Projektleiter bei der Weblaw AG; Weblaw AG, Laupenstrasse 1, 3008 Bern, Schweiz; www.weblaw.ch; blaise.devaud@weblaw.ch.

sich selten einheitlich verwalten: aus historischen oder operativen Gründen sind die Dokumente in separaten E-Mail-Archiven, Geschäftsverwaltungssystemen, Datei-Ablagen und Bibliothekslösungen verteilt.

In einer dermassen komplizierten Informationslandschaft spielt eine Suchmaschine eine zentrale Rolle.

Die juristische Suchmaschine Lawsearch Enterprise³ der Firma Weblaw AG ist entstanden, um alle wichtigen Informationsquellen einer juristischen Organisation in einer einzigen Suchmaske verfügbar zu machen. Die Benutzer können sich mit diesem Single Point of Search über die (neusten) externen Veröffentlichungen informieren und gleichzeitig in den kompletten internen Archiven relevante Informationen zu den eingehenden Rechtsfällen finden.

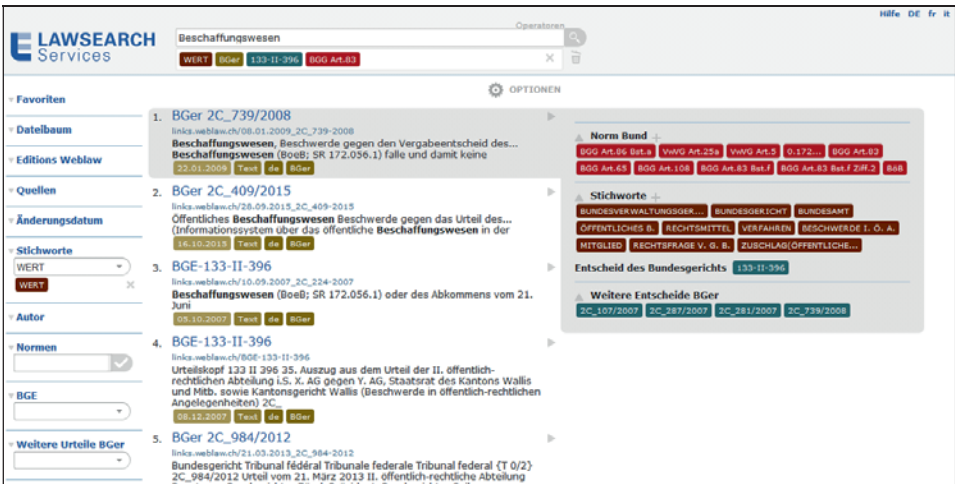


Abb. 1: Beispiel einer Suche mit der juristischen Suchmaschine Lawsearch. Textsuche („Beschaffungswesen“) kann mit einer mehrsprachigen Suche nach Stichworten und Referenzen (Bundesgerichtsentscheide BGE, Bundesnormen) kombiniert werden. In der Mitte werden Suchtreffer mit einem kurzen Textauszug angezeigt; rechts werden zu jedem Treffer relevante Metainformationen (Bundesnormen, Stichwörter, Rechtsprechungsreferenzen) gezeigt.

Die Gestaltung und Entwicklung der juristischen Suchmaschine Lawsearch Enterprise ist aber technisch und organisatorisch anspruchsvoll. Drei spezifische Herausforderungen stehen im Vordergrund: Auswahl und Integration der Informationsquellen; Beschaffung und Verwendung der Metainformationen; Erkennung und Verlinkung der juristischen Referenzen.

³ Die Leser können die Webversion der juristischen Suchmaschine Lawsearch Enterprise bei den folgenden öffentlichen Instanzen testen: <http://www.weblaw.ch/competence/search.html> (Suchseite des Internet-Auftritts der Weblaw AG); <http://lawsearch.gr.ch/le/> (publizierte Gerichtsentscheide des Kantons Graubünden, Schweiz); <https://apps.vs.ch/le/> (publizierte Gerichtsentscheide des Kantons Wallis, Schweiz). Weitere Informationen: <http://www.weblaw.ch/competence/technology/lawsearch-enterprise.html>.

2 Auswahl und Integration der Informationsquellen

Das erste Problem ergibt sich aus der Verteilung in interne und externe Informationsquellen. Die internen Quellen einer juristischen Organisation sind extrem sensibel und dürfen das interne IT-System (bspw. einer Anwaltskanzlei) nicht verlassen. Dies hat zur Folge, dass die Suchmaschine beim Kunden innerhalb von seinem IT-System installiert und eingesetzt werden muss. Die externen Quellen – mehrheitlich frei im Internet verfügbar – sind sehr umfangreich (Hunderttausende von Gesetzen, Gerichtsentscheidungen und Textdokumenten). Die Integration dieser Daten in den internen Suchindex jeder juristischen Organisation würde zu unnötigem und redundantem Ressourcenverbrauch führen (Speicherplatz, Bandbreite, Arbeitslast bei der Indexierung). Aus diesem Grund werden alle externen juristischen Quellen zentral bei der Weblaw AG indexiert und in einem gemeinsamen Suchindex gespeichert. Suchanfragen zu den externen Quellen werden aus jeder organisationsinternen Suchmaschine an den zentralen Suchindex weitergeleitet; die Suche wird im zentralen Index ausgeführt und die organisationsinterne Suchmaschine erhält die Suchresultate zurück. Dieses Vorgehen erlaubt eine effiziente Verwaltung der externen Quellen und respektiert die hohen Sicherheitsanforderungen für die internen Quellen.

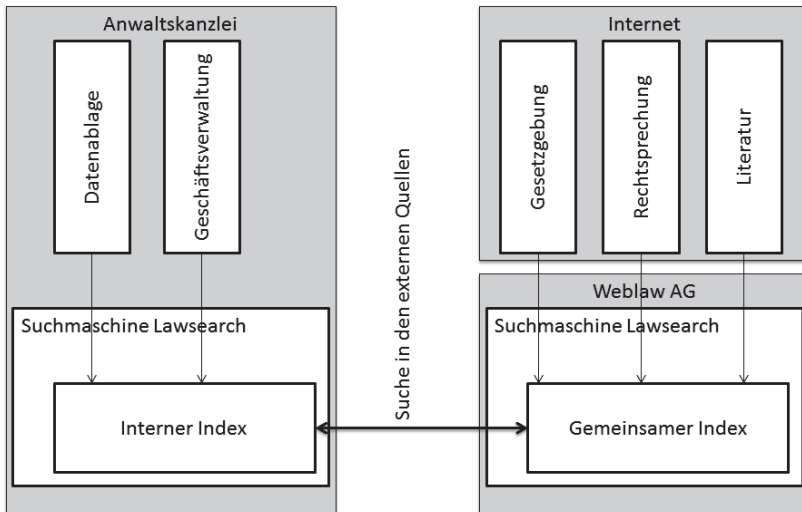


Abb. 2: Die internen Quellen sind intern beim Kunden indexiert; die externen sind zentral bei der Weblaw AG in einem gemeinsamen Index gespeichert. Anfragen zu den externen Quellen werden an den gemeinsamen Index weitergeleitet; die interne Suchmaschine erhält die Resultate zurück.

Sowohl die internen als auch die externen Quellen zeigen weitere spezifische Detailprobleme.

Die internen Informationen werden jeweils aus internen Drittsystemen des Kunden be-

schafft: Filesysteme; Intranet; Sharepoint; Geschäftsverwaltungs- und Zeiterfassungssysteme; E-Mail-Archive etc. Zu jedem dieser Drittsysteme muss eine Schnittstelle bestehen oder erstellt werden, öfters unter Berücksichtigung der Benutzerverwaltung und des Sicherheitskonzepts. Diese Tatsache erhöht die Wartungskomplexität der Suchmaschine. Andererseits können Skalenerträge erreicht werden, wenn mehrere Kunden die gleiche interne Technologie einsetzen.

Bei den externen Informationsquellen aus dem Internet ist im ersten Schritt eine intellektuelle Auswahl gefragt⁴: es sollen vorzugsweise Informationsquellen bestimmt werden, die für alle Juristen von Interesse sind. Quellen aus spezifischen Rechtsgebieten sind eingeschränkt, können aber auf Anfrage aufgenommen werden. Des Weiteren sind die externen Quellen dem üblichen Problem der Internet-Quellen ausgesetzt: die ausgewählten Webseiten/Datenbanken unterlaufen öfters formellen oder inhaltlichen Änderungen, was eine ständige Überwachung und Anpassung der spezifischen Crawler impliziert. Diese Situation ist in der föderalen Schweiz insbesondere akut: jeder Kanton entscheidet unabhängig von den anderen über die Publikation seiner Gesetzgebung und Rechtsprechung. Obwohl gemeinsame technische Standards bereits erarbeitet wurden⁵, finden sie in der Praxis noch kaum Anwendung. Der immense Kundennutzen einer aggregierten Suche in allen heterogenen Rechtsquellen rechtfertigt entsprechend diese hohen Integrationsaufwände.

Eine besondere Art der externen Informationsquellen bilden Inhalte, die die Weblaw AG als juristischer Verlag selber produziert. Sie bestehen aus den Publikationen der Editions Weblaw⁶ (juristische digitale Zeitschriften Jusletter (jusletter.ch), Jusletter IT (jusletter-it.eu), Richterzeitung (richterzeitung.ch) und Archiv für Schweizerisches Abgaberecht (asa.weblaw.ch); juristische Datenbank Banking and Finance (bf.weblaw.ch); digitaler Rechtsprechungskommentar (drsk.weblaw.ch)) sowie aus den bibliographischen Angaben zur juristischen Literatur (biblio.weblaw.ch). Diese letztgenannten Angaben entstammen der juristischen bibliographischen Datenbank⁷ der Weblaw AG.

Alle Quellen sollen effizient und benutzerfreundlich durchgesucht und dargestellt werden. Diese Anforderung wird bei Lawsearch Enterprise durch den Einsatz von Suchreibern erfüllt. Unter dem Volltextsuchfeld werden verschiedene Reiter dargestellt: durch das Klicken auf die einzelnen Tabs werden die Resultate aus den entsprechenden Quellen angezeigt. Somit kann mit einer einzigen Eingabe der Suchbegriffe (und Filter) in allen verfügbaren Quellen das Resultat konsultiert werden. Weiter werden für eine Suche alle internen und externen Quellen detailliert in einem Suchbaum aufgelistet: so lässt sich die Suche auf einzelne Quellen einschränken.

⁴ Eine Liste der aktuellen externen Quellen befindet sich unter http://www.weblaw.ch/competence/technology/lawsearch-enterprise/juristische_quellen.html.

⁵ CHLexML heisst der Standard zur Darstellung von Gesetzen im XML Format; <http://www.svri.ch/de/CHLexML.html>. CHDecML für Rechtsprechung und CHDocML für Literatur befinden sich im Entwurfstadium).

⁶ Siehe dazu die Verlagsinformationen unter http://www.weblaw.ch/competence/editions_weblaw.html.

⁷ http://www.weblaw.ch/competence/editions_weblaw/bibliotheksloesung/rechtsbibliografie.html.

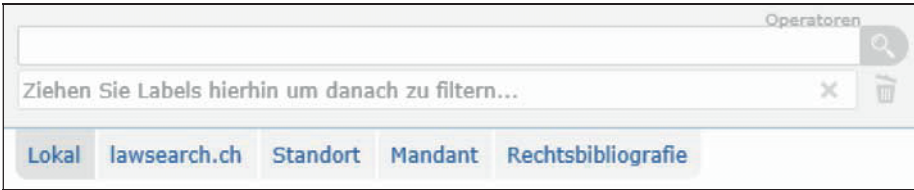


Abb. 3: Die Reiter unter den Suchfeldern erlauben eine Navigation in allen verfügbaren Quellen mit einer einzigen Suchanfrage. Der erste Reiter zeigt die Treffer aus den internen Quellen (Lokal), der zweite die Treffer aus den externen Quellen (lawsearch.ch), und die drei letzten Reiter zeigen die passenden bibliographischen Angaben (jeweils aus der Bibliothek vom eigenen Organisations-Standort; aus allen Bibliotheken der Organisation; aus der gesamten Rechtsbibliografie).

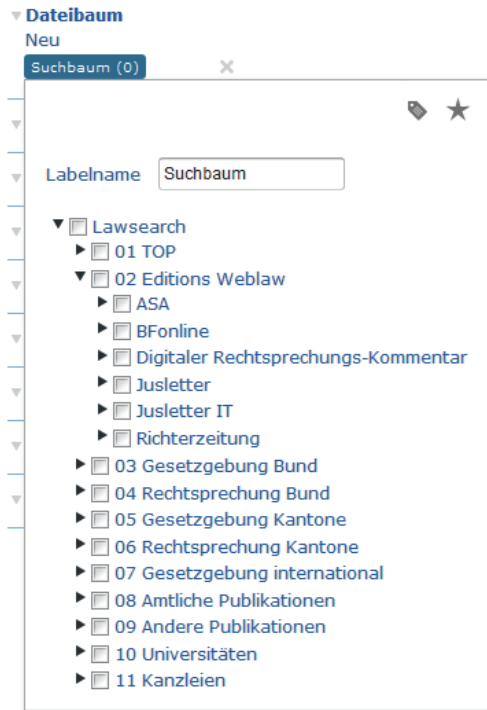


Abb. 4: Auswahl der externen Quellen im Suchbaum.

3 Beschaffung und Verwendung der Metainformationen

Reine Textsuchen in den verfügbaren Dokumenten sind für Juristen nicht ausreichend. Vielmehr benötigen sie bei der täglichen Arbeit die Möglichkeit, die Treffer nach gewissen Metainformationen zu filtern. Diese Metainformationen beziehen sich auf die juristi-

schen (Dokumentenart, Rechtsgebiet, Aktennummer, zuständiges Organ), technischen (Format, Erstellungsdatum, Änderungsdatum) und organisationsbezogenen (Verbindungen zu Projekten und Personen) Eigenschaften des Dokuments.

Bei den internen Dokumenten werden die Metadaten meist direkt über die Schnittstelle zu den Drittsystemen beschafft. Dabei besteht die technische Herausforderung darin, die Metadaten zu den Dokumenten aktuell zu halten. Dies ist insb. dann problematisch, wenn sich Metadaten ändern, ohne dass das Dokument an sich geändert wird. Dieses Problem wird durch einen Push-Mechanismus auf der Seite des Drittsystems gelöst: das Drittsystem benachrichtigt die Suchmaschine periodisch über die letzten Änderungen der Metadaten.

Bei den externen Dokumenten ist eine strukturierte Abfrage der Metadaten mittels eines Web-Services nur selten möglich. In den meisten Fällen müssen die publizierten Dokumente inhaltlich analysiert werden (Parsing), um die relevanten Informationen zu extrahieren. Aufgrund der verschiedenen Formate und Formatierungen der halbstrukturierten Text-Dokumente ist leider keine generische Lösung möglich: für jede Quelle muss ein dediziertes Parsing programmiert werden, was den Arbeitsaufwand bei den externen Quellen weiter erhöht.

Die hohe Anzahl der verfügbaren Quellen resultiert in einer entsprechend hohen Anzahl von verfügbaren Metadaten-Dimensionen. Es ist eine Design- und Usability-Herausforderung, alle Metadaten-Typen effizient und benutzerfreundlich in der Suchmaske verfügbar zu machen. Dieses Problem löst Lawsearch Enterprise mit dem Einsatz von den sog. Labels: es sind Suchfilter, die grafisch dargestellt und via drag-und-drop (oder Doppelklick) ins zweite Suchfeld gezogen werden (siehe dazu Abb. 4 „Ziehen Sie Labels hierhin um danach zu filtern“). Dieses zweite Suchfeld wird auch „Label-drop-space“ genannt). Mit diesem Lösungsansatz können beliebig viele Labels und Label-Typen miteinander kombiniert werden. Im Resultat liegt eine multidimensionale Facetten-Suche vor.



Abb. 5: Mehrere Labels können in der Suche gleichzeitig eingesetzt werden: dies entspricht einer multidimensionalen Facetten-Suche. Die Suche kombiniert den Begriff Beschaffungswesen mit dem mehrsprachigen Deskriptor Wert, der Quelle Bundesgericht, der Bundesgerichtsreferenz 133 II 396 sowie dem mehrsprachigen Erlasszitat Art. 83 (inkl. aller Substufen) BGG (Bundesgerichtsgesetz).

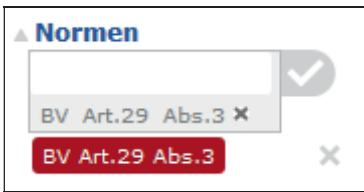


Abb. 6: Labels können beliebig von Hand erstellt werden. Hier wird ein Label erstellt, um alle Dokumente zu finden die den Artikel 29 Absatz 3 der Bundesverfassung zitieren (in drei Sprachversionen). Vorgeschlagen werden nur Referenzen, die sich effektiv im indizierten Dokumentenstamm befinden.



Abb. 7: Metainformationen zu jedem Treffer-Dokument werden als Labels dargestellt (rechts) und können direkt zur Verfeinerung der Suche benutzt werden.

4 Erkennung und Verlinkung der juristischen Referenzen

Lawsearch Enterprise bereichert jedes Dokument mit weiteren juristischen Metadaten (juristische Referenzerkennung, Veredelung). Die Suchmaschine erkennt automatisch die wichtigsten juristischen Referenzen (und Deskriptoren), die im Text des Dokuments erwähnt werden: zitierte Gesetzesartikel, „Fall-Nummern“ von Gerichtsentscheiden und zitierte Literatur aus juristischen Zeitschriften sowie weitere jur. Referenzen.

Die Erkennung erfolgt während der Indexierung der Dokumente und basiert auf dem Einsatz von regulären Ausdrücken. Voraussetzung der korrekten Erkennung ist eine einheitliche Zitierweise. Dies erweist sich in der Praxis öfters als problematisch: nicht alle Referenzen, insb. auf Kantonsebene, sind standardisiert. Des Weiteren kann der Sinn einer Referenz kontextabhängig sein (z.B. wird die gleiche Abkürzung zur Bezeichnung von verschiedenen kantonalen Gesetzen oder von verschiedenen historischen Versionen des gleichen Gesetzes verwendet). Die regulären Ausdrücke verarbeiten diese Abweichungen soweit möglich.

Die erkannten Referenzen werden zu drei Zwecken genutzt.

Sie können als Suchfilter verwendet werden. Dies erlaubt eine fortgeschrittene mehr-

sprachige Suche nach Dokumenten, die einen konkreten Gesetzesartikel oder einen bestimmten Gerichtsentscheid zitieren – diese Suchmöglichkeit wird in der Praxis besonders geschätzt.

Zweitens werden die erkannten Referenzen in einer Arbeitsversion des Dokumentes (sog. Cache-Version) direkt im Text hervorgehoben und verlinkt. Aufgrund der häufigen Änderungen der Internet-Adressen stellt die Verlinkung eine besondere Herausforderung dar. Um eine nachhaltige Verlinkung zu gewährleisten wurde ein Webservice⁸ als Zwischenstufe eingebaut. Jeder automatisch erstellte Link zeigt auf die URL des Webservices mit der Referenz als Zusatzparameter⁹; der Webservice prüft die aktuelle URL für diese Referenz und leitet die Browser-Anfrage entsprechend weiter auf die korrekte Website. Der Webservice erlaubt eine zentrale Verwaltung der juristischen Links und eine zeitnahe, für die Benutzer kaum merkbare Anpassung bei Änderungen.

Drittens werden alle erkannten Referenzen in der Cache-Version gesammelt aufgelistet. Dies bietet eine automatische Zusammenfassung des Dokumentes und ist besonders bei längeren Texten hilfreich.

<p>Gegenstand Art. 9⁹⁹ BV (Beschaffungswesen),</p> <p>Staatsrechtliche Beschwerde gegen das Urteil des Kantonsgerichts Appenzell Innerrhoden, Abteilung Verwaltungsgericht, vom 6. September 2005.</p> <p>Sachverhalt:</p> <p>A. Der Kanton Appenzell Innerrhoden holte im Einladungsverfahren bei zwölf Firmen Offerten für ein elektronisches Archivsystem ein. In die Schlussevaluation kamen vier Anbieter, darunter die Y. _____ AG, St. Gallen, und die X. _____ GmbH, Zürich.</p> <p>Im Auftrag von Landammann und Standeskommission erteilte der Ratschreiber den Zuschlag für das elektronische Archivsystem am 25. Februar 2005 zum Preis von Fr. 243'396.-- an die Y. _____ AG. Die X. _____ GmbH, deren Offerte unberücksichtigt blieb, erhob hiegegen am 6. März 2005 Beschwerde beim Kantonsgericht (Abteilung Verwaltungsgericht). Das Gericht befand, die Beschwerde genüge den gesetzlichen Anforderungen nicht und räumte der X. _____ GmbH Frist ein, um die Rechtschrift zu verbessern. Dies tat die Beschwerdeführerin mit Eingabe vom 22. März 2005, worauf der Kantonsgerichtspräsident der Beschwerde mit Verfügung vom 24. März 2005 zunächst aufschiebende Wirkung gewährte. Diese Verfügung hob er am 12. April 2005 wieder auf und entzog der Beschwerde die aufschiebende Wirkung, wogegen sich die X. _____ GmbH bis vor Bundesgericht zur Wehr setzte. Die am 25. April 2005 gegen diesen Zwischenentscheid des Kantonsgerichtspräsidenten erhobene staatsrechtliche Beschwerde blieb jedoch ohne Erfolg; sie wurde vom Geschäftsverzeichnis abgeschrieben, nachdem sich herausgestellt hatte, dass der Vertrag mit der berücksichtigten Konkurrentin, der Y. _____ AG, bereits am 15./22. April 2005 abgeschlossen worden war (vgl. Beschluss ⁹⁹2P.116/2005 vom 24. Mai 2005, insbesondere E. 1).</p>	<p>BGG: Art.84 Art.86 Art.87 Art.88 Art.156 Art.159</p> <p>BGBM: Art.9</p> <p>Weitere Urteile BGER</p> <p>2P.116/2005 2P.218/2001 2P.294/2005</p> <p>Leitentscheide BGE</p> <p>125-II-86 132-I-86 125-II-518 129-I-173</p>
--	--

Abb. 9: In der Cache-Version (Arbeitsversion) eines Dokumentes werden die juristischen Referenzen erkannt, im Text hervorgehoben, verlinkt und auf der rechten Seite zusammenfassend aufgelistet. Die gesuchten Begriffe werden ebenfalls farblich gekennzeichnet.

⁸ <http://links.weblaw.ch>.

⁹ Z.B. „<http://links.weblaw.ch/de/SR-210 Art.3>“.

5 Fazit

Die juristische Suchmaschine Lawsearch Enterprise wird täglich von hunderten Schweizer Juristen erfolgreich eingesetzt. Zu den Kunden zählen Anwaltskanzleien, Gerichte, Versicherungen und Rechtsdienste vieler Unternehmen. Sie ist ein relevantes Arbeitsinstrument bei der täglichen juristischen Arbeit.

Die gelungene Lösung der drei Herausforderungen hat zu diesem Erfolg wesentlich beigetragen. Die juristische Suchmaschine Lawsearch Enterprise ist eine Schweizer Success Story.

Möglichkeiten und Herausforderungen bei der Verwendung von Bildern in Suchmaschinen

Timo Schulz¹, Björn Zessack², Bengt Rabe³ und Sebastian Kielmann⁴

Abstract: Das Datenvolumen, das Suchmaschinen indexieren, ist in den letzten Jahren rasant angestiegen und zusätzlich zu der klassischen Domäne Text, bekommen Multimedia-Dateien zunehmend eine wichtige Bedeutung. Dazu zählen in erster Linie Bilder, aber auch Sprache und Filme. Zusätzlich zum kontinuierlichen Anstieg des Volumens bedeutet die Erschließung dieser neuen Domänen aber auch, dass neue Verfahren benötigt werden, um Nicht-Text-Daten ausreichend zu beschreiben und sinnvoll durchsuchbar zu machen und das zu der Erschließung der Inhalte auch neue Verfahren für die Indexierung benötigt werden, da im Gegensatz zu Text, Multimedia-Dateien ein wesentlich größeres Datenvolumen beinhalten. Nachfolgend soll eine Übersicht gegeben werden, welches Potenzial eine Einbindung der neuen Domain in Suchmaschinen bringt, aber auch die Herausforderungen und Probleme, wie z.B. die Einbindung in bestehende User-Interfaces und die Mehrdeutigkeit bei der Bestimmung der Nutzerabsicht. Als konkreter Fall wird dabei der Bereich e-Commerce mit Schwerpunkt Shopping betrachtet.

Keywords: Bild, Deep Learning, GPU, e-Commerce

1 Einleitung

Mit der zunehmenden Verbreitung von Breitband-Internetzugängen, nimmt das Datenvolumen im Internet rasant zu, da es immer einfacher möglich ist, mehr und größere Daten schnell auszutauschen und zu präsentieren. Schon 2001 gab es Suchmaschinen, die über 250 Millionen [Sm10] Bilder indexiert hatten und die heutige Zahl dürfte bei über 10 Milliarden liegen. Im Gegensatz zu den Anfängen der Suchmaschinen, wo oft nur Text indexiert wurde, stellt die Indexierung beliebiger Multimedia-Dateien eine deutlich größere Herausforderung dar, da mehr Speicherplatz und mehr Rechenleistung für die Erschließung und Indexierung verwendet wird.

Im Gegensatz zur Textsuche, wo die Eingabe lediglich aus einer Menge Stichworten, bzw. kurzen Sätzen besteht, stellt die Suche in Multimedia-Dateien eine zusätzliche Herausforderung dar, da es mehrere Modalitäten bei der Eingabe geben kann, die mit der Absicht des Nutzers in Einklang gebracht werden müssen. Im Beispiel von Bildern gilt es festzulegen, wie/was genau gesucht werden sollten. Die folgende Liste enthält einige Beispiele:

¹ picalike GmbH, Stresemannstraße 23 D-22769, t.schulz@picalike.com.

² picalike GmbH, Stresemannstraße 23 D-22769, b.zessack@picalike.com.

³ picalike GmbH, Stresemannstraße 23 D-22769, b.rabe@picalike.com.

⁴ picalike GmbH, Stresemannstraße 23 D-22769, s.kielmann@picalike.com.

- Stichworte: visueller Inhalt ist mit Schlagworten verknüpft
- Freitext: Text wird aufbereitet und in einem semantischen Raum transformiert
- Skizze: Erfordert Hochladen und Skelett-Indexierung der Bilder
- Beispielbild: Semantische, globale Ähnlichkeitssuche
- Kombination mehrerer Möglichkeiten

Des Weiteren gibt es verschiedene Anwendungsfälle der Bildersuche:

- Exakte Suche: Finde alle Bilder, die z.B. eine bestimmte Person/Objekt zeigen
- Suche von (Fast-)Duplikaten: Finde alle Bilder, beliebig nah an der Referenz sind
- Konzept-Suche: Finde alle Bilder, die zu einem Konzept passen (Rotes Ballkleid mit Rüschen und Punkten)

Im weiteren beziehen wir uns auf die Konzept-Suche anhand eines Beispielbildes. Für diesen Fall ist unsere These, dass ein vollständiges, maschinelles Bildverständnis einer reinen Textsuche oft überlegen ist. Weiterhin ist eine Ähnlichkeitssuche auf Bildern, gerade im Bereich Mode, einer reinen Textsuche in den meisten Fällen überlegen, da Produktbeschreibungen begrenzt, bzw. fehleranfällig sind und damit das Suchvokabular begrenzt ist. Zeigt z.B. ein Bild eine Bluse mit blauen, auffälligen Knöpfen, aber keine Erwähnung in der textuellen Beschreibung, kann dieses Detail für eine Suche nicht verwendet werden.

2 Nachteile der klassischen Textsuche für die Bildersuche

Bilder spielen im Internet eine immer größere Rolle. Gerade im e-Commerce gibt es Bereiche wie Mode, in denen Entscheidungen oft rein visuell, bzw. ästhetischer Natur sind. Ohne eine genaue Erfassung der Bilddaten kann eine Suche aus Sicht von Nutzern nicht zufriedenstellend durchgeführt werden, da der Erfolg von einer vollständigen, manuellen Annotation des Bildmaterials abhängt. Das heißt, es muss eine Suche auf konzeptioneller Ebene möglich sein, die über eine reine Stichwortsuche hinausgeht. Ein Vorteil ist, dass zur Beschreibung der Konzepte das Bild selbst verwendet werden kann und es somit nicht notwendig ist, dass eine manuelle Annotation für ein Bild vorliegt.

In der Vergangenheit, wurde oft der umliegende oder Anker-Text verwendet, um Bildinhalte auf Seiten zu beschreiben[He09, Me09, FC09]. Der Ansatz ist allerdings nicht zuverlässig, da es keine Garantie gibt, dass der Text eine ausreichende Relevanz zu dem Bild hat. Auch werden Mehrdeutigkeiten wie Jaguar (Auto oder Raubkatze) auf dieser Ebene nicht aufgelöst und es ist möglich, dass nur ein kleiner Aspekt des Bildes beschrieben wird. Generell ist es bei manueller Annotation auch möglich, dass sich die

Relevanz nachträglich oder zeitlich ändert. Dass in frühen Zeiten vermieden wurde, Bildinhalte zu analysieren, ist auch der mangelnden damaligen Leistung der CPUs geschuldet, da selbst moderne Verfahren ohne GPU schnell an ihre Grenzen geraten, wenn es darum geht Millionen von Bildern mittels Deep-Learning-Verfahren zu verarbeiten.

Ein anderes Problem bei der manuellen Annotation von Bildern ist die Mehrdeutigkeit von Metadaten durch individuelle Schemata, aber auch Details wie die Nutzung verschiedener Sprachen, wodurch katalogübergreifende Lösungen oft nicht durchführbar sind, oder nur mit erheblichem Mehraufwand. Auch ist nicht auszuschließen, dass Daten fehlerhaft oder unvollständig vorliegen, was gravierende Einschränkungen bei der Suche durch Nutzer haben kann. Insgesamt ist eine Erschließung des Bildmaterials durch eine manuelle Annotation fehleranfällig und zeitaufwendig und nicht zeitnah möglich. Eine einheitliche Taxonomie wäre in jedem Fall von Vorteil, um alternative Fachbegriffe und unscharfe Formulierungen zu behandeln, aber auch um unterschiedliche Konzepte klar abzugrenzen.

Allerdings kann es auch bei einer direkten Verwendung des Bildes eine Taxonomie notwendig sein, um eine konsistente Darstellung zu ermöglichen. Für eine mögliche Stichwortsuche in Bildkonzepten ist es somit notwendig, Eingaben von Nutzer direkt in die Taxonomie zu überführen, z.B. durch eine automatische Vervollständigung der Eingabe, was allerdings voraussetzt, dass Mehrdeutigkeiten, wie Sprache, bereits auf anderer Ebene behandelt wurden. Hierbei wäre eine Möglichkeit, die Konzepte für das Bildmaterial sprachneutral zu codieren und dann für jede Sprache ein Mapping anzulegen, was allerdings nicht immer konsistent ist, da Konzepte in einigen Sprachen sehr vielfältig beschrieben werden können.

3 Vorteile durch die Verwendung von Bilddaten

Da ein Bild unabhängig von Sprache selbst-beschreibend ist, also alle relevanten Informationen enthält, um ein Objekt zu beschreiben, besteht die Herausforderung darin, eine geeignete Methode zu finden, die rohen Pixelwerte in eine konzeptionelle Sicht zu überführen. In der klassischen Bildverarbeitung, vor dem Siegeszug von Deep Learning, wurden Bilder oft durch lokale Features beschrieben, wie z.B. HoG[DT05], SIFT[Lo04], SURF[Ba08] oder globale Beschreibungen wie GIST[Do09], die dann in eine kompakte Repräsentation überführt wurden. Da diese Verfahren allerdings nicht für einen speziellen Zweck optimiert wurden, sind die Ergebnisse oft nicht zufriedenstellend gewesen. Seit der Beschleunigung durch GPUs werden deshalb immer öfter Modelle verwendet, die gezielt für einen speziellen Zweck trainiert werden können. Der Vorteil hierbei ist, dass Bildfeatures nicht mehr generisch sind, sondern passend zum Problem erzeugt werden. Allerdings sind auch bei Deep Learning die meisten Verfahren auf Annotationen angewiesen, so dass Veränderungen bei Taxonomien eine Anpassung und ein erneutes Training der Modelle zur Folge haben können. Weiterhin ist ein Abwägen notwendig, um ein Mittelmaß zwischen Genauigkeit und vorhandenen Ressourcen zu finden, da selbst mit Deep Learning die Extraktion von allen relevanten Bildinformationen in einer

angemessenen Zeitspanne ein Thema für zukünftige Forschung ist. Bei der Klassifikation haben die Modelle allerdings schon übermenschliche Genauigkeit[Ru14] erreicht und durch einen großen Label-Raum, zum Beispiel ImageNet[De09], sind die Features solche Modelle zum Teil vielseitig einsetzbar, wodurch die Einführung neuer Attribute nicht zu einem kompletten neuen Training führt, sondern stückweise Erweiterungen erlauben. Es soll nun ein konkreter Fall für eine Ähnlichkeitssuche betrachtet werden.

3.1 Use Case

Mode Im Fall von einer e-Commerce Shopping-Seite gehen wir von einer Seite aus, die eine einfache Suche nach Stichworten oder Artikelnummer erlaubt. Von Interesse ist allerdings nur der Fall, dass ein Nutzer nur eine sehr grobe Vorstellung hat, was er eigentlich sucht. Das könnte z.B. ein Paar neue Sneaker sein, wobei allerdings noch nicht klar ist, was für genaue Eigenschaften das Produkt haben soll. Ähnlich der Textsuche, ist ein Startpunkt notwendig, der dann verfeinert wird. Dies könnte durch die Auswahl der Ober-Kategorie geschehen, oder aber durch die Auswahl von vordefinierten Einsprungspunkten, wie z.B. Trends. Je nach Ziel ist eine gewisse Vielfalt der Ergebnisse notwendig, um dem Nutzer seine Möglichkeiten aufzuzeigen. So ist es möglich, dass der Nutzer ein (globales) Konzept gefunden hat, dass er weiter verfeinern möchte, ohne aber vordefinierte Filter, wie Marke, zu verwenden. Das ausgewählte Bild fungiert dann als Referenz, um ähnliche Produkte zu finden, die verwandte Konzepte enthalten. Wie genau die Bewertung der Bilder stattfindet ist variabel, es kann z.B. Form/Farbe/Textur unterschiedlich bewertet werden, was ein klarer Vorteil gegenüber textlicher Beschreibung ist, da sich z.B. die Wahrnehmung von Farben nur schwer mit festen Filtern abbilden lässt. Genauso ist eine vollständige Aufzählung sinnvoller Texturen für eine spezielle Suche oft überhaupt nicht möglich. Obwohl bei Modebildern oft nur ein Objekt abgebildet ist, kann die Absicht eines Nutzers nicht ohne explizites Feedback inferiert werden. Eine Verfeinerung der Suche ist nämlich nur dann möglich, wenn klar ist, welches Detail auf dem Bild für die weitere Suche relevant ist, z.B. hoher Absatz, aber keine Schnürsenkel. Die Entscheidungen bei jeder Iteration können ebenfalls als Feedback in die Suche einfließen, Stichwort Ranking. Am Ende der Suche wurden die Produkte dann hoffentlich so eingegrenzt, dass ein Nutzer leicht seinen Favoriten wählen kann.

4 Zusammenfassung

Die These, dass die inhaltliche Verwendung von Bilddaten einen Mehrwert für nutzergetriebene Suchen darstellt, wurde exemplarisch belegt. Allerdings ist der Mehrwert stark davon abhängig, wie vollständig das Bild konzeptionell durch Features beschrieben werden kann. Hierzu ist einerseits ein gutes Modell notwendig, andererseits eine ausreichende, manuelle, Annotationen der Trainingsdaten. Das heißt, die Verwendung der Bildmodalität setzt weiterhin Text als Modalität voraus, um Bilder zu beschreiben. Des weiteren stellt die Verwendung von Bilddaten auch neue Anforderungen an

Suchmaschinen selbst dar, da es notwendig ist, große Mengen an Bilddaten kompakt zu beschreiben, ohne große Einbußen bei der Qualität, diese zu speichern und schnell abrufbar zu halten. Im speziellen Fall von Katalogbildern, die durch den Fokus auf nur ein Produkt, sehr oft keine Mehrdeutigkeit erlauben, ist eine Bestimmung der Nutzerabsicht einfacher, allerdings kann dieser Fall nicht auf natürliche Bilder verallgemeinert werden, was dazu führt, dass nicht immer klar ist, ob ein spezielles Konzept vom Benutzer bevorzugt wird. Das wirft die Frage auf, wie Suchmaschinen bei Verwendung von Bildern als Eingabe eine Verfeinerung oder Korrektur der Suchabsicht erlauben. Auch ist eine weitere Herausforderung in Hinsicht auf die Ressourcen, der kontinuierliche Bildupload von Nutzer, wenn betrachtet wird, dass eine normale Suchanfrage im Schnitt nur wenige kurze Worte enthält, wobei selbst ein kleines Farb-Bild (3*256*256) eine Größe von 192 KB hat. Auch wenn die Ressource Bild in den vergangenen Jahren sehr stark an Bedeutung gewonnen hat und die Fortschritte, zum Teil durch Deep Learning, sehr eindrucksvoll sind, bleiben noch viele offene Fragen und Herausforderungen, um die Modalität Bild in Suchmaschinen einzubinden.

Literaturverzeichnis

- [Ba08] Bay, Herbert; Ess, Andreas; Tuytelaars, Tinne; Van Gool, Luc: Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.*, 110(3):346–359, Juni 2008.
- [De09] Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L.: ImageNet: A Large-Scale Hierarchical Image Database. In: *CVPR09*. 2009.
- [Do09] Douze, Matthijs; Jégou, Hervé; Sandhawalia, Harsimrat; Amsaleg, Laurent; Schmid, Cordelia: Evaluation of GIST Descriptors for Web-scale Image Search. In: *Proceedings of the ACM International Conference on Image and Video Retrieval. CIVR '09*, ACM, New York, NY, USA, S. 19:1–19:8, 2009.
- [DT05] Dalal, Navneet; Triggs, Bill: Histograms of Oriented Gradients for Human Detection. In: *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*. *CVPR '05*, IEEE Computer Society, Washington, DC, USA, S. 886–893, 2005.
- [FC09] Filipe, Joaquim; Cordeiro, José, Hrsg. *WEBIST 2009 - Proceedings of the Fifth International Conference on Web Information Systems and Technologies*, Lisbon, Portugal, March 23-26, 2009. INSTICC Press, 2009.
- [He09] Hearst, Marti A.: *Search User Interfaces*. Cambridge University Press, 1. Auflage, 2009.
- [Lo04] Lowe, David G.: Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vision*, 60(2):91–110, November 2004.
- [Me09] Metaxas, Panagiotis Takis: On the Evolution of Search Engine Rankings. In: (Filipe; Cordeiro, J.) [FC09], S. 200–207.

- [Ru14] Russakovsky, Olga; Deng, Jia; Su, Hao; Krause, Jonathan; Satheesh, Sanjeev; Ma, Sean; Huang, Zhiheng; Karpathy, Andrej; Khosla, Aditya; Bernstein, Michael S.; Berg, Alexander C.; Li, Fei-Fei: ImageNet Large Scale Visual Recognition Challenge. CoRR, abs/1409.0575, 2014.

- [Sm10] Smith, Nate: , Ooh! Ahh! Google Images presents a nicer way to surf the visual web. <https://googleblog.blogspot.de/2010/07/ooh-ahh-google-images-presents-nicer.html>, July 2010.

Sprachgesteuerte Websuche im Test – Ein Erfahrungsbericht aus der Perspektive von Auftraggeber, Auftragnehmer und Durchführendem

Sebastian Sünkler¹, Friederike Kerkmann¹, Sonja Quirnbach² und Sebastian Schultheiß¹

Abstract: Beschrieben wird ein studentisches Projekt, das im Auftrag eines Telekommunikationsunternehmens die sprachgesteuerte Websuche vier verschiedener Systeme evaluierte und aus den Erkenntnissen Empfehlungen für die Gestaltung einer optimalen Voice Web Search ableitete. Der Beitrag versucht, die unterschiedlichen Perspektiven und Erwartungen von Auftraggeber, Projektleitung und durchführenden Studierenden zu beleuchten und miteinander in Beziehung zu bringen. Mit diesem Erfahrungsbericht sollen zum einen die konkreten Ergebnisse in ihren Grundzügen vorgestellt, zum anderen die unterschiedlichen Blickwinkel und Anforderungen der Beteiligten reflektiert sowie die wissenschaftlichen und forschungspraktischen Herausforderungen des innovativen Themas dargestellt werden.

Keywords: Websuche, Sprachsteuerung, Smartphone, Google Now, Apple Siri, Microsoft Cortana, Amazon Fire OS, Usability, Evaluation, Suchmaschinen

1 Einleitung

Die Popularität der Sprachsteuerung hat durch den rasanten Anstieg der Absatzzahlen von Smartphones stark zugenommen und wird genutzt, um Anrufe aufzubauen, Textnachrichten zu verfassen, Apps zu öffnen oder Websuchen durchzuführen. Aktuelle Studien zeigen, dass jeder zweite Deutsche die Sprachsteuerung anwendet [Bi16] und in den USA knapp 40% der Besitzer auf die Eingabe von Befehlen per Sprache zurückgreifen [PA16]. Wie effektiv, effizient und zufriedenstellend aber arbeiten derzeitige Systeme, speziell im Bereich der Websuche? Wie gut funktioniert „googeln“, wenn der Nutzer seine Suchanfrage nicht schreibt, sondern spricht? Diesen Fragen ist ein studentisches Projekt nachgegangen, das im Auftrag eines Telekommunikationsunternehmens die sprachgesteuerte Websuche von vier verschiedenen Systemen evaluierte und aus den gewonnenen Erkenntnissen Handlungsempfehlungen zur Gestaltung einer eigenen Voice Search Anwendung für den Auftraggeber ableitete.

In diesem Beitrag wird versucht, die drei unterschiedlichen Perspektiven (1) Auftraggeber, (2) Auftragnehmer bzw. Projektleitung und (3) Durchführender auf das

¹ Hochschule für Angewandte Wissenschaften, Fakultät Design, Medien und Information, Department Information, Finkenau 35, 22081 Hamburg, vorname.nachname@haw-hamburg.de

² Deutsche Telekom AG, Products & Innovation, T-Online-Allee 1, 64295 Darmstadt, s.quirnbach@telekom.de

Thema, das Projekt sowie die Ergebnisse zu beleuchten. Mit diesem Erfahrungsbericht sollen zum einen die konkreten Ergebnisse in die Fachcommunity transportiert, zum anderen die unterschiedlichen Blickwinkel und Anforderungen der Beteiligten reflektiert werden. Gleichzeitig sollen auch die wissenschaftlichen und forschungspraktischen Herausforderungen eines innovativen Themas wie die Websuche per Sprachsteuerung, für das es kaum etablierte Maße und Methoden gibt, diskutiert werden.

Der Beitrag gliedert sich in drei Abschnitte: Zunächst werden als Rahmen das Projekt sowie die Grundlagen der sprachgesteuerten Websuche vorgestellt. Im zweiten Abschnitt beschreiben nacheinander der Auftraggeber, die Auftragnehmer bzw. Projektleiter und ein Student, der mit der Durchführung beauftragt war, ihre jeweiligen Erfahrungen und Einschätzungen im Hinblick auf die Zielsetzung, die Durchführung und die Ergebnisse. Der Erfahrungsbericht endet mit einem gemeinsamen Fazit, das diese unterschiedlichen Perspektiven bündelt, die Chancen und Herausforderungen der Auseinandersetzung mit der sprachgesteuerten Websuche zusammenfasst und weitere Forschungsbedarfe konkretisiert.

2 Das Projekt „Ok Google... The End of Search as we know it“

Das Projekt „Ok Google... The End of Search as we know it. Sprachgesteuerte Suche im Test“ fand als studiengangübergreifendes Projekt der beiden Studiengänge *Medien und Information* und *Bibliotheks- und Informationsmanagement* am Department Information der Hochschule für Angewandte Wissenschaften (HAW) Hamburg im Wintersemester 2015/16 statt. Auftraggeber war die Deutsche Telekom AG; teilgenommen haben 15 Studierende beider Studiengänge unter der Leitung von zwei Projektleitern.

Die Zielsetzung bestand darin, die drei marktführenden Systeme zur sprachgesteuerten Websuche *Google Now*, *Apple Siri* und *Microsoft Cortana* sowie das inzwischen nicht mehr erhältliche *Amazon Fire OS* auf ihre Stärken und Schwächen hin zu evaluieren. Ausgehend davon sollten Handlungsempfehlungen für die Entwicklung einer sprachgesteuerten Websuche abgeleitet und eine optimierte Oberfläche konzipiert werden. Die Ergebnisse wurden in einem Abschlussbericht aufgearbeitet und dem Auftraggeber in einem Workshop vorgestellt und dort gemeinsam mit diesem weiterentwickelt.

3 Sprachgesteuerte Websuche

Die Geschichte der modernen sprachgesteuerten Websuche ist jung. Sowohl *Google* als auch *Apple* veröffentlichten im Jahr 2010 ihre Dienste *Google Voice Search* und *Apple Siri*. *Microsoft Cortana* wurde 2014 als Bestandteil von Windows 8.1 auf den Markt gebracht. Die drei genannten Systeme sind momentan die Marktführer in ihrem Segment, da sie integrierter Bestandteil der Betriebssysteme der Anbieter sind. Sie

nutzen dabei vergleichbare Technologien zur Verarbeitung von Spracheingaben und sind selbstlernende Systeme, die auf Cloud-Server-Architekturen zugreifen, um die Eingaben per Stimme durch ihre Nutzer auszuwerten und Websuchen durchzuführen. In der praktischen Anwendung bedeutet dies, dass ein Nutzer eine Suchanfrage einspricht, die durch Websuchmaschinen verarbeitet wird und die Ergebnisse dem Nutzer anschließend präsentiert werden. Die Darstellung und Verarbeitung hängen dabei von der Integration der Websuchmaschinen in das Betriebssystem des Gerätes ab. So wird die akustische Suchanfrage entweder direkt in Text umgewandelt und an eine Suchmaschine weitergeleitet oder es findet eine Aufbereitung der Suchergebnisse bezogen auf das System statt.

Alle diese Anbieter setzen dabei auf Suchmaschinentechnologien, die sie entweder selbst entwickelt haben (*Google, Microsoft Bing*) oder verwenden eine Kombination vorhandener Suchtechnologien. *Apple* nutzt *Google, Bing, Yandex* und weitere Suchmaschinen. Die Ergebnispräsentation unterscheidet sich bei den Anbietern dahingehend, dass *Google* auch bei Suchanfragen per Sprache eine Darstellung analog zur traditionellen Search Engine Result Page (SERP) bei textuellen Suchanfragen nutzt, während *Microsoft* und *Apple* in diesem Fall jeweils ein eigenes Layout der Ergebnisse präsentieren. Vereinzelt unterscheidet sich die Ergebnispräsentation auch dadurch, dass bestimmte Suchanfragen durch Spracheingabe auch akustisch beantwortet werden. Zusätzlich hebt sich die sprachgesteuerte Websuche von der traditionellen Suche dadurch ab, dass die Systeme zum Teil in der Lage sind, mit dem Anwender „zu sprechen“, indem sie in den Dialog mit ihm treten und sich so auf vorangegangene Suchen beziehen können (z. B. „Wie heißt der Präsident der Vereinigten Staaten?“ → Antwort → „Wie groß ist er?“ → Antwort).

Das Fire Phone von *Amazon*, dessen Sprachsteuerung keinen eigenen Namen trägt und von anderen sprachbasierten Amazon-Diensten wie bspw. *Alexa* abzugrenzen ist, ist inzwischen nicht mehr käuflich erhältlich. Das Gerät, das durch Netlock ausschließlich an das Netz der Telekom gebunden war, dennoch in den Test einzubeziehen, war Wunsch des Auftraggebers.

Eine detaillierte Darstellung des derzeitigen Entwicklungsstandes und Funktionsumfangs zu allen Geräten findet sich im Projektbericht, der unter <http://searchstudies.org/ok-google/> online zur Verfügung steht.

4 Perspektive des Auftraggebers

4.1 Zielsetzung und Erwartungen an das Projekt

Innerhalb meiner Arbeit in der Produktentwicklung von innovativen Informationstechnologien bei der Deutsche Telekom AG werde ich zunehmend mit technisches Endgeräten konfrontiert, die

- nur ein minimalgroßes Interface haben, wie Smartwatches,
- eine intelligente Schnittstelle zum Kunden bekommen, deren Benutzerschnittstellen noch nicht standardisiert sind bzw. sich erst entwickeln, wie Geräte rund um Smarthome oder Connected Car,
- mit neuen Steuerungsmöglichkeiten zur Mensch-Maschine-Kommunikation ausgestattet werden, wie Geste oder Sprache.

Diese Endgeräte sind für die Nutzer noch neu, weil sie zum einen noch nicht in deren Lebensalltag angekommen sind und zum anderen andere bzw. bisher ungelernete Benutzerschnittstellen aufweisen. Es gibt also eine große Veränderung der Benutzerschnittstellen. Das heißt, sie werden vielfältiger werden und nicht jede wird in der gewohnten Art erscheinen und ein klassisches grafisches interaktives Frontend haben und über Tastatur, Maus oder Berührung steuerbar sein. Festzuhalten ist aber, dass die neuen Endgeräte eine Suchfunktion besitzen bzw. über diese auch Suchen über verschiedene Indexe möglich sind; sei es eine Websuche oder eine Suche über eine Applikation in einem Katalog.

Eine schnelle Entwicklung vollzieht sich zur Zeit insbesondere im Bereich der Sprache als Interaktions- und Steuerungselement zwischen Mensch und Maschine, die zunehmend ausgereifter und intelligenter werden, z. B. *Siri* oder *Amazon Echo*.

Als Folge darauf war es für mich naheliegend zu forschen, inwieweit es unter Nutzern schon verbreitet ist, auch die Websuche über die Sprache zu bedienen. Aus diversen selbst durchgeführten Nutzerstudien zur Sprachsteuerung weiß ich, dass es noch nicht selbstverständlich ist, die Sprache als Steuerungselement zu nutzen. Die meisten Befragten gaben dabei an, Sprache bei Navigationssystemen im Auto, bei Telefonhotlines oder maximal bei Sprachnachrichten über *WhatsApp* zu nutzen.

Dies gab mir den Anlass aufzuspüren, wie Nutzer aktuell die Spracheingabe verwenden, wie sie damit suchen und wie ihre Erwartungshaltung ist. Da ich davon ausgegangen bin, dass es Qualitätsunterschiede zwischen den Anbietern der sprachgesteuerten Websuchen von *Google Now*, *Apple Siri*, *Microsoft Cortana* und *Amazon Fire OS* gibt, war auch ein Benchmark dieser Systeme als Grundlage nahe. Die marktführenden Systeme sollten auf ihre Stärken und Schwächen evaluiert werden. Ein weiterer Fokus war für mich zu erfahren, wie Nutzer nach einer Sucheingabe mit den Ergebnissen interagieren.

Ein besonderer Schwerpunkt war dabei auch die Präsentation der Suchergebnisse und wie Nutzer mit diesen umgehen. Hier besteht die grundlegende Frage darin, den Medienbruch Sprache und visuelle Ausgabe der Ergebnisse gleichzeitig in einem nutzerfreundlichen Kontext zusammenzubringen. Eine der Stärken von Smartphones ist, dass sie in einem mobilen Kontext von unterwegs verwendet werden können und der Nutzer jederzeit online ist. Es wurde in der Studie deshalb davon ausgegangen, dass sich die Nutzer geographisch bewegen und ihre Suchmotivation innerhalb eines bestimmten Kontexts heraus entsteht (bspw. abgebildet über die sog. Micro-Moments, von Google definierte Suchintentionen im mobilen Kontext, s. Tab. 2).

4.2 Methodendiskussion

Schon innerhalb meiner eigenen durchgeführten Recherche zum Thema fiel mir auf, dass es kaum öffentlich verfügbare Studien gibt, die sich mit dem Nutzerverhalten im Zusammenhang mit der Sprachsteuerung befassen. Das hatte sich dann auch bei der begleitenden Recherche der Studierenden bestätigt. Umso mehr wurden wir alle gleich schon zu Beginn des Projektes darin bestätigt, zu diesem Thema zu forschen.

Mehr denn je sind die Einschätzungen zur User Experience zur Spracheingabe in Bezug auf Websuchen oder Suchen allgemein noch ungenau und subjektiv. Daraus eine allgemeingültige Ableitung zu treffen, ist dann die Herausforderung für weitere Studien. Sprache ist ein sehr individuelles Kommunikationsmedium, das grundlegend von den soziokulturellen Bedingungen der Nutzer abhängt. Hierzu gehört beispielsweise auch das verwendete Vokabular zu Suche, Satzbau und vor allem das Wissen um die Sprachbefehle und die Auswirkungen auf den Erfolg der Suche.

Aufbauend auf meinen Studien rund um Nutzerevaluierung in Suchmaschinen [Qu12] ist hier eine Erweiterung der anzuwendenden Evaluationsmethoden und Einschätzung der User Experience notwendig. Es sollten Kriterien identifiziert werden, um ein optimales System zu konzipieren und darauf aufbauend auch um dieses Systeme von Beginn an in der Gebrauchstauglichkeit und letztlich seinem Erfolg beurteilen zu können.

Im Folgenden zwei Fragestellungen, welche die Komplexität der Nutzung über die Sprache verdeutlichen:

1. Welche akustischen Ankerpunkte bekommt der Nutzer zur Sprachausgabe mitgeliefert, wenn zu einer Suchanfrage viele Ergebnisse gefunden werden und die ersten fünf Suchergebnisse von der Liste vorgelesen werden? Wie viele Ergebnisse kann der Nutzer für sich im Kopf behalten und sich erinnern, um dann das für ihn passende auszusuchen?
2. Wie kann der Nutzer effektiv mit den Ergebnissen arbeiten, um komplexe Aktionen miteinander zu verknüpfen (bspw. in Form von Formulierungen wie „Suche mit den neuen Trailer von Ice Age 4 auf Deutsch, ohne Werbung und sende ihn per Whats App an meine Schwester.“).

Hier gibt es in Sachen Nutzerverhalten und User Experience noch großes Entwicklungs- und Forschungspotential, insbesondere da bis dato noch keine Standards bezüglich Sprachsteuerung vorhanden sind.

Aus der Studie kam heraus, dass sich die Probanden eine intelligente Nutzerführung seitens des Systems wünschen. Deswegen wurden *Google* und *Siri* etwas besser bewertet, da sie den Probanden intelligenter erschienen. Zudem wird eine stärkere Emotionalisierung der Sprache bevorzugt, denn es gefällt wenn das System lustig und natürlich reagiert, wie das bei *Siri* der Fall ist. Ein natürliches Kommunikationsverhalten wird also als besser empfunden, als eine Stimme, die mechanisch und künstlich klingt.

4.3 Einschätzung der Ergebnisse

Zusammenfassend zeigt die Studie, dass die Bedienung der Geräte über Sprache vielen Nutzern noch sehr fremd ist. Teilweise war die Nutzung der Spracheingabe in öffentlichen Raum für die Probanden sogar unangenehm, da es für sie schon in ihre Privatsphäre eindringt, wenn sie in der Öffentlichkeit ihre Suchanfragen preisgeben. Hinter jeder Sucheingabe befinden sich persönliche Vorlieben und Motivationen, die man eben nicht gern öffentlich bekannt geben möchte.

Eine andere Frage ist, wie die Nutzerführung aussieht. Einerseits finden Probanden es hilfreich und erwarten es, wenn mehrere Ergebnisse vorgelesen werden. Andererseits stellt sich die Frage danach, inwieweit man sich auch mehrere Ergebnisse merken kann, um diese zu reflektieren und das passende auszusuchen. Die intelligente Nutzerführung wird bei der Suche über die Sprachsteuerung somit einen großen Fokus bekommen.

Es gibt also noch viele offene Fragestellungen, die bisher weder im Kontext der Nutzung noch aus der Perspektive der Nutzer detaillierter untersucht worden sind. Interessant in diesem Zusammenhang sind auch die Fachgebiete Suchmaschinenoptimierung (Search Engine Optimization, SEO) und Online Marketing (OM). Beispielsweise wäre interessant herauszufinden, wie spezielle Topergebnisse oder Anzeigen über die Sprache ausgegeben werden können. Es ist aktuell noch kein Wissen vorhanden, um über die Möglichkeiten und vor allem Chancen von SEO oder OM in Bezug zur Sprache Aussagen zu machen.

Die Nutzung der Suche über die Sprache ist somit dato noch nicht im Fokus der UX-Betrachtung. Deshalb kann noch keine Handlungsempfehlung bezüglich User Experience getroffen werden. Jedoch können erste Tendenzen abgeleitet werden, insbesondere die Erwartungshaltung der Nutzer gegenüber dem System. Diese Studie stellt somit Weichen und kann als Grundlage für weitere Evaluationen herangezogen werden. Ich hoffe mir natürlich aus meiner Sicht, dass weitere Projekte diesbezüglich stattfinden.

5 Perspektive der Projektleiter

5.1 Zielsetzung und Erwartungen an das Projekt

Der allgemein formulierte Projektauftrag, Stärken und Schwächen der Systeme zu erheben, musste für die praktische Durchführung auf konkret zu beurteilende Qualitätskriterien heruntergebrochen werden. Analog zu den Qualitätsdimensionen traditioneller Desktop-Suchmaschinen nach [LH07] sollten daher auch die Voice Web Search Systeme anhand ihrer Treffer-Qualität, Recherche-Qualität sowie Usability beurteilt werden. Die Index-Qualität, die üblicherweise den vierten Qualitätsaspekt von Suchmaschinen definiert, wurde nicht gesondert betrachtet, da zu solchen Evaluierungen

Aspekte der Vollständigkeit, Abdeckung des Webs und der Aktualität untersucht werden müssten, die zum einen methodisch durch die Dynamik des Webs schwer zu messen und zum anderen für den genannten Anwendungsfall nicht relevant waren. Die Treffer-Qualität wurde dabei durch die Retrievaleffektivität, den Einsatz von Personalisierungsmaßnahmen sowie die Präsentation der Ergebnisse abgebildet, die Recherche-Qualität über die Verarbeitung der Suchanfragen und die Dialogfähigkeit der Systeme. Die Beurteilung der Usability erfolgte anhand des Zugangs, der Bedienung, der Barrierefreiheit sowie der Bereitstellung von Hilfen.

Zur Evaluation wurde den Studierenden ein Multimethoden-Ansatz bestehend aus verschiedenen experten- und nutzerbasierten Verfahren vorgegeben. Diese Methodenvielfalt diente sowohl dem konkreten Erkenntnisgewinn als auch dem Kennenlernen der Geräte und Systeme. Gleichzeitig bestand der Anspruch darin, den Studierenden ein breites Methodenrepertoire an die Hand zu geben. Tabelle 1 zeigt die eingesetzten Methoden sowie ihre jeweilige Zielsetzung in der Übersicht:

	Methode	Ziel inhaltlich	Ziel methodisch
		Qualitätsdimension	
expertenbasiert unter Laborbedingungen	Literaturrecherche	State of the Art	Kennenlernen des Systems/Geräts, Einüben von wissenschaftlichem Arbeiten
	Heuristik	Identifikation von ‚bedeutenden‘ / ‚katastrophalen‘ Usability-Problemen, Vergleich der Geräte	Kennenlernen des Systems/Geräts, Einüben der Methode
		Usability	
	Szenariobasierter Walkthrough	Identifikation von Usability-Problemen aus Nutzersicht, Hineinversetzen in verschiedene Nutzertypen und Nutzungsszenarien	Kennenlernen der jeweils anderen Systeme/Geräte, Einüben der Methode + Persona-Entwicklung
		Usability	
	Retrievaltest	quantitative Analyse der Retrievaleffektivität unter Laborbedingungen, Vergleich der Geräte	Einüben der Methode
Treffer-Qualität			
nutzerbasiert unter Realbedingungen	Vorabbefragung	Erfassung demographische Daten und Vorerfahrung	Einüben der Methode, „Angst vorm Nutzer verlieren“
	Teilnehmende Beobachtung	Identifikation von Stärken und Schwächen unter realen Bedingungen	Einüben der Methode
		Recherche-Qualität, Treffer-Qualität, Usability	
	Retrospective Thinking Aloud	Identifikation von Stärken und Schwächen unter realen Bedingungen	Einüben der Methode
		Recherche-Qualität, Treffer-Qualität, Usability	
SUS-Fragebogen	quantitative Analyse der	Einüben der Methode	

		Gebrauchstauglichkeit, Vergleich der Geräte	Einüben der Methode
		Usability	
	Nachbefragung	Identifikation von Stärken und Schwächen in der Rückschau	
		Recherche-Qualität, Treffer-Qualität, Usability	

Tab. 1: Eingesetzte Methoden und Ziele

Dem Walkthrough, dem Retrievaltest sowie der Nutzerstudie zugrunde lagen Suchanfragen, die sich jeweils an den von *Google* formulierten Micro-Moments³ [Go15] orientierten, um so die verschiedenen Suchintentionen, die im mobilen Kontext auftreten können, gleichmäßig zu berücksichtigen. Tabelle 2 zeigt die vier genutzten Micro-Moments in der Übersicht:

Micro-Moment	Suchintention	Beispiel
WANT-TO-KNOW	informationsorientierte Suchanfragen	Ein Nutzer will sich über ein Produkt informieren, das er in diesem Moment in einer TV-Werbung sieht.
WANT-TO-GO	lokale Suche	Ein Nutzer sucht nach einem italienischen Restaurant in seiner Nähe.
WANT-TO-DO	Howto-Suche	Ein Nutzer sucht nach einem Rezept für einen Apfelkuchen.
WANT-TO-BUY	transaktionsorientierte Suche	Ein Nutzer befindet sich einem Geschäft und nutzt die Suche als Entscheidungshilfe beim Produktkauf.

Tab. 2: Micro-Moments

5.2 Methodendiskussion

Anwendungen der Sprachsteuerung lassen sich bei der praktischen Nutzung nicht immer trennscharf voneinander unterscheiden; auch die Abgrenzung der im Rahmen des Projektes relevanten Websuche von allgemeiner Suchfunktion gestaltete sich nicht einfach. Häufig bildet die Suche den Zugang zu der eigentlichen Anwendung. Dies trifft insbesondere auf die Nutzung von Personal Assistants zu, die Anwender bei der Erfüllung von Aufgaben unterstützen sollen. Z. B. wird vor dem Aufbau eines Anrufs eine Suche in der Kontaktliste auf dem Smartphone durchgeführt. Dieser Prozess zählt zu den Voice-Search-Anwendungen der Sprachsteuerung ohne Bezug auf eine Websuche. Voice-Search-Anwendungen lassen sich von Voice Web Search dadurch abgrenzen, dass Voice-Search sich sowohl auf Offline- und Online-Inhalte bezieht, dabei aber nicht zwangsweise als Websuche agiert. Voice Web Search hingegen bezieht sich ausschließlich auf die Inhalte aus Websuchmaschinen und ist damit unabhängig von den

³ Mobile Suchen sind im Gegensatz zur Suche an stationären Geräten praktisch von überall möglich und damit stärker von Spontaneität und Ungeduld der Nutzer getrieben. Dieses Verhalten bilden die Micro-Moments ab, indem sie die verschiedenen Momente, in denen ein Nutzer mit seinem Smartphone suchen kann, und die sich daraus ergebenden Suchbedürfnisse beschreiben.

gespeicherten Inhalten auf dem Gerät. Diese inhaltliche und sprachliche Abgrenzung war im Projekt ebenso wie die Differenzierung von Gerät vs. System nicht immer einfach und blieb auch bis zum Projektabschluss z.T. unklar – Ein Zustand, der bisweilen zu Unsicherheit und Unschärfe führte.

Die methodische Herausforderung des Vorhabens bestand im Fehlen etablierter Standardmethoden für die Evaluation von Voice Web Search bzw. in deren (Nicht-)Übertragbarkeit. So existiert bislang keine Heuristik für diesen Bereich. Ansätze, wie die Guidelines zur Navigation und Orientierung auf Smartphones aus der Grundlagenstudie von [eR14] bilden zwar eine gute Ausgangsbasis für mobile Websites, sind jedoch nicht auf Sprachsteuerung und Websuche übertragbar. Mangels etablierter Heuristiken wurde eine eigene Heuristik entwickelt, die folgende potentielle Problemkategorien abprüfte: Zugang/Einstieg, Bedienung, Suchanfrageverarbeitung, Personalisierung, Barrierefreiheit, Ergebnispräsentation und Hilfestellung [SK16]. Die Überprüfung der Tragfähigkeit dieser selbstentwickelten Heuristik steht noch aus. Auch bei der Durchführung des Retrievaltests mussten gegenüber der klassischen Variante Einschränkungen in Kauf genommen werden: So waren den Juroren die Systeme bei der Bewertung bekannt. Außerdem variierte die Darstellung bzw. Ausgabe der Ergebnisse (akustisch vs. herkömmliche Trefferdarstellung mit Ergebnisliste) so stark zwischen den Systemen, dass eine Aussage über die Retrievaleffektivität auf Dokumentenebene nicht möglich war.

Im Rahmen der Nutzertests blieb der Grundsatz, Probanden die Anwendung ausschließlich mit dem Endgerät testen zu lassen, dass sie auch selbst besitzen (bzw. bestenfalls ihr eigenes Gerät zur Studie mitzubringen) [eR13], unberücksichtigt. Hintergrund war zum einen die geringe Verbreitung des Microsoft- und Fire-Phones, zum anderen hätten Verzerrungen durch Personalisierungsmaßnahmen auf dem eigenen Gerät nicht systematisch ausgeschlossen werden können.

5.3 Einschätzung der Ergebnisse

Der genutzte Methodenmix war ein Versuch, etablierte Methoden der Usability- und Suchmaschinen-Forschung für ein neues Anwendungsfeld zu adaptieren und miteinander zu kombinieren. Dabei kam es zu den beschriebenen Methodenschwächen und zum Teil nur bedingt aussagekräftigen Ergebnissen. Stärke und Schwäche von Usability-Evaluierungen ist die überwiegende Nutzung „weicher“, subjektiver qualitativer Usability-Kriterien. Das zeigt sich auch in den Projektergebnissen, die nach der Durchführung der Tests keinen eindeutigen Sieger festmachen können, wobei Google von allen getesteten Systemen noch die besten Ergebnisse erzielte. Gründe dafür liegen zum einen in den „weichen“, wenig greifbaren Maßen, zum anderen in der Anzahl der Evaluatoren, die dieselben Sachverhalte möglicherweise unterschiedlich beurteilten. Schwierigkeiten ergaben sich auch durch eine Affinität der Studierenden und Probanden zu bestimmten Systemen, was eine objektive Beurteilung teilweise einschränkte.

Dennoch konnten die Ergebnisse der verschiedenen Untersuchungsschritte genutzt werden, um daraus Erfolgskriterien abzuleiten, die eine nutzerfreundliche Voice Web Search kennzeichnen. Diese Kriterien können herangezogen werden, um bestehende Systeme in ihrer Gebrauchstauglichkeit zu beurteilen oder um ein optimiertes System zu konzipieren. Tabelle 3 zeigt die gesammelten Erfolgskriterien in der Übersicht:

Zugang/ Einstieg	Be- dienung	Suchanfrage- verarbeitung	Personali- sierung	Barriere- freiheit	Ergebnis- präsentation	Hilfestellung
Eindeutig als Sprachsteuerung erkennbar	Lange Sprechpausen möglich	Akzente und Dialekte werden korrekt erkannt	Standort wird bei der Suchanfrage einbezogen	Grafiken in den Ergebnissen werden in ausreichender Größe dargestellt	Sprachausgabe bei Faktentrecherche	Fehlermeldungen sind hilfreich
	Brauchbare Ergebnisse bei Stichwortsuchen	Englische Anfragen werde korrekt erkannt	Ausgabesprache wird der Eingabesprache angepasst	Die Schriftgröße bei den Ergebnissen ist ausreichend	Akustisches Feedback zu den Ergebnissen nach der Suchanfrageverarbeitung	Aktive Unterstützung bei der Bedienung
	Interaktion mit Suchergebnissen per Sprache möglich	Unbekannte Abkürzungen können interpretiert werden		Kontrast bei der Ergebnisdarstellung ist angemessen	Sprachausgabe direkt zu den Suchergebnissen wird unterstützt	Links zur Hilfe sind vorhanden und gut sichtbar
	Suchverarbeitung per Sprache beenden	Korrekte Erkennung der Anfrage		Akustische Wiedergabe aller Suchergebnisse	Passende Darstellungen zu den Micro-Moments	Hilfestellung in Textform wird angeboten
	Unterstützung von Dialogen	Autokorrektur wird unterstützt		Schriftvergrößerung wird unterstützt	Die Quellen der Suchergebnisse sind immer erkennbar	Hilfe bei Null-Treffer-Seiten und nicht verstandenen Anfragen
	Echtzeitfeedback bei der Verarbeitung				KWIC wird unterstützt	
	Zuverlässige Erkennung bei lauten Umgebungsgläuschen				Die Trefferbeschreibungen sind verständlich	
					Die Ergebnisse sind auf dem aktuellen Stand	

Tab. 3: Erfolgskriterien von Voice Web Search Anwendungen [SK16]

Für die praktische Arbeit mit den Ergebnissen ergeben sich aber andere Anforderungen als an die wissenschaftliche Aussagekraft. Für eine Sicherstellung der wissenschaftlichen Aussagekraft ist die Entwicklung eines Forschungsdesigns mit Berücksichtigung der Reliabilität, Validität und Objektivität der Studien notwendig, was aber nicht im Fokus der Projektarbeit stand. Vielmehr ging es darum festzulegen mit dem Einsatz wissenschaftlichen Methoden praxistaugliche Ergebnisse generiert werden. Dabei fand eine ständige Abwägung zwischen Wissenschaftlichkeit und Pragmatismus statt, die zu den teilweise unscharfen Bewertungs- und Erfolgskriterien von Voice Search Anwendungen führten. Die Projektergebnisse bieten in jedem Fall einen nutzbaren ersten Einblick zu den Stärken und Schwächen bestehender Anwendungen der sprachgesteuerten Websuche. Die Handlungsempfehlungen sind ein brauchbarer Einstieg bei der Gestaltung und Entwicklung eines eigenen Dienstes.

6 Perspektive des Durchführenden

6.1 Zielsetzung und Erwartungen an das Projekt

Als Studierender stieg ich mit einer relativ klaren Erwartungshaltung in das Projekt ein. Diese wurde insbesondere durch den hohen Innovationsgehalt des Themas der sprachgesteuerten Websuche geprägt. Denn insbesondere aufgrund fehlender etablierter Evaluationsmethoden war von Beginn an deutlich, dass ein gewisser Handlungsspielraum gegeben war.

Zur privaten Vorbereitung machte ich mich zunächst mit wesentlichen Funktionen der sprachgesteuerten Websuche vertraut, die ich als Privatanwender bislang nicht verwendet hatte. Somit war die Zielsetzung an das Projekt zweigeteilt. So wollte ich die auf dem Markt verfügbaren Systeme einerseits grundlegend kennenlernen. Insbesondere von Interesse war allerdings die Anwendung diverser Methoden zur Evaluation der Systeme, worauf im Folgenden eingegangen werden wird.

6.2 Methodendiskussion

Der unter 5.1 beschriebene Multimethoden-Ansatz erwies sich für uns Studierende als Gewinn. Vor deren Durchführung konnten wir die verschiedenen Methoden zunächst kennenlernen, indem wir diese auf ihre Einsetzbarkeit hin überprüft haben. Ausschlaggebend für den Lernerfolg war schließlich die unmittelbare Anwendung, die in experten- und nutzerbasierte Evaluation unterteilt war.

Die expertenbasierte Evaluation bestand aus Literaturrecherche, Heuristik, szenariobasiertem Walkthrough sowie einem Retrievaltest. Diese war aus zweierlei Gründen wichtig. Zum einen waren die Voice Web Search Systeme den meisten Studierenden in ihren Funktionsweisen unbekannt. Insbesondere die Durchführung von Recherche und Heuristik haben demnach sichergestellt, dass sich jeder Projektteilnehmer umfangreich mit den jeweiligen Geräten beschäftigt hat, bevor die Nutzertests begannen. Des Weiteren diente die expertenbasierte Evaluation als wichtige Grundlage für die Nutzertests. Es konnten bereits vorab bestimmte Schwächen der Systeme ausgemacht werden, denen in den Nutzertests besondere Beachtung geschenkt worden ist. So wurde während der Durchführung der Heuristik ermittelt, dass die Systeme oftmals nicht abwarten, bis der Nutzer die Suchanfrage zu Ende gesprochen hat. Anhand der Nutzertests konnte dieser Eindruck bestätigt werden.

Die Nutzertests können als Kernbestandteil der gesamten Evaluation bezeichnet werden. Umrahmt von einer Vor- und Nachbefragung wurden die Probanden während der Tests gebeten, ihre Eindrücke zu schildern (Retrospective Thinking Aloud) und zum Schluss einen Fragebogen zur Gebrauchstauglichkeit (System Usability Scale, SUS [Br86]) auszufüllen.

Für die Tests konnten Probanden akquiriert werden, die ein sehr heterogenes Probandenfeld abbildeten. So war der jüngste Proband 8, die älteste 46 Jahre alt. Diese Vielfalt stellte sich schon während der einzelnen Durchläufe als vorteilhaft heraus, weil sich die unterschiedlichen Altersgruppen in der Benutzung der sprachgesteuerten Websuche teilweise stark unterscheiden haben. Da reale Anwender solcher Systeme ebenso divers sind, ist dieser Aspekt der Evaluation als besonders positiv hervorzuheben.

Als Resultat der Evaluation wurden Empfehlungen für die Nutzerführung einer optimalen sprachgesteuerten Websuche ermittelt. Dafür wurde per Flussdiagramm ein Suchprozess erstellt, angefangen von der Sucheingabe bis hin zur Ergebnisauswahl. Abbildung 1 zeigt beispielhaft den ersten Schritt des Sucheinstiegs innerhalb einer optimalen Voice Web Search Anwendung.

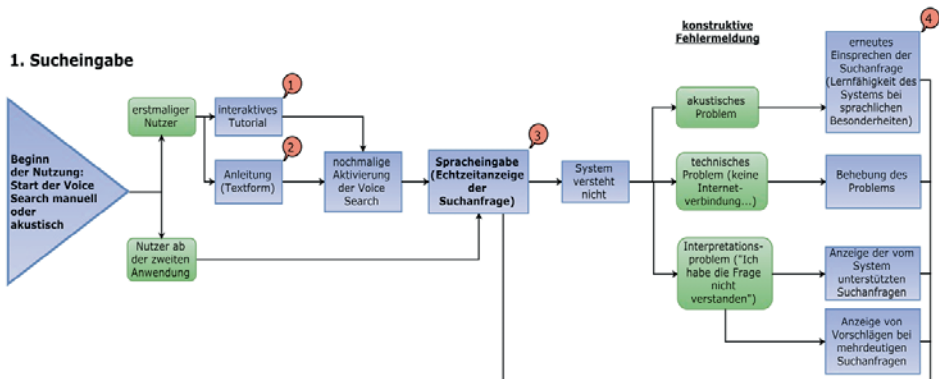


Abb. 1: Sucheinstieg innerhalb einer optimalen Voice Search [SK16]

Zusätzlich wurden mögliche Nutzerinterfaces skizziert, die einzelne Schritte innerhalb des Prozesses visualisieren sollten. Sämtliche dieser Empfehlungen wurden gruppenübergreifend erarbeitet. Dadurch vereinigten sich die Evaluierungsergebnisse aller inhaltlichen Arbeitsgruppen und es wurde deutlich, wie ähnlich die Erfahrungen aus Experten- sowie Nutzersicht waren. Abbildung 2 illustriert die Oberflächen für ein interaktives Tutorial sowie eine schriftliche Anleitung für den Nutzer, der die sprachgesteuerte Websuche das erste Mal nutzt.



Abb. 2: Sucheinstieg innerhalb einer optimalen Voice Search [SK16]

6.3 Einschätzung der Ergebnisse

Die schon unter 5.3 erwähnten „weichen“ Usability-Kriterien erschwerten auch für uns Studierende die Durchführung und Auswertung der einzelnen Methoden.

Dies äußerte sich bereits in den Heuristiken, die Usability-Probleme identifizieren und gewichten sollten. Hat beispielsweise eine Gruppe fehlende Funktionen zur Barrierefreiheit als „Usability-Katastrophe“ gewichtet, könnte eine andere Gruppe diese Gewichtung sehr unterschiedlich vorgenommen haben, obwohl deren System identische Mängel aufweist. So kann ein Vergleich der einzelnen Geräte auch lediglich als grobe Einschätzung betrachtet werden.

Bei der Konzipierung der optimalen Voice Web Search hingegen waren sich alle Gruppen zügig einig, obwohl als Grundlage dieselben, subjektiven Ergebnisse verwendet worden sind. Dies kann dadurch begründet werden, dass die grundsätzlichen Stärken und Schwächen der Systeme lediglich in ihren Ausprägungen, nicht aber in ihrem Vorhandensein zur Diskussion standen.

Eine interessante Fragestellung für die Zukunft wäre es zu ermitteln, wie Probanden auf einen Prototyp der konzipierten Voice Web Search reagieren würden. Dies wurde selbstverständlich nicht durch die Zielsetzung des Projekts abgedeckt, könnte aber die vorliegenden Erkenntnisse aufgreifen und diese in funktionierende Prototypen umsetzen.

7 Fazit

Das Projekt kombinierte Methoden der Usability- und Suchmaschinenevaluierung ebenso wie Grundlagen des prototypischen Designs und passte diese jeweils auf mobile Endgeräte und die Steuerung durch Sprache an. Dabei basierten diese Adaptionen überwiegend auf eigenen, praxisgetriebenen Überlegungen und weniger auf wissenschaftlich bestätigten Ansätzen. Dies zeigt die Chancen und Herausforderungen, die die Auseinandersetzung mit einem innovativen Thema, für das es keine Erfahrungen gibt, mit sich bringt: Zum einen beinhaltet ein solches Projekt eine große Freiheit, in dem die Methoden, Vorgehensweisen und Anwendungsbezüge mangels Standards frei gewählt werden können. Zum anderen jedoch existiert kein in der Wissenschaft oder Praxis geprüfter Handlungsrahmen, an dem sich das Vorhaben orientieren kann. Gründe für eine fehlende wissenschaftliche Grundlage zu der systematischen Evaluierung der sprachgesteuerten Suche liegen in der jungen, innovativen Technologie bzw. der erst jetzt relevanten Verbreitung vorangetrieben durch die Verfügbarkeit und den hohen Absatz von smarten Devices. Wie auch in anderen Bereichen der Forschung und Suchmaschinenforschung lassen sich solche Daten erst erheben, wenn sich solche Technologien etablieren. Bisherige Studien zur Nutzung werden momentan durch die Marktführer selber durchgeführt. Die Methoden und die Rohdaten stehen der Wissenschaft aber nicht zur Verfügung.

Damit zeigt sich erheblicher weiterer Forschungsbedarf: Auf Methodeneben werden bspw. eine auf ihre Tragfähigkeit hin überprüfte Heuristik benötigt, Ansätze, wie akustische und herkömmliche Treffer gleichermaßen in ihrer Effektivität beurteilt werden können oder Erkenntnisse über die Auswirkungen des retrospektiven Lauten Denkens oder den Einfluss eines Fremdgerätes im Rahmen einer Nutzerstudie. Auch der Frage, wie Akustik beschrieben und dargestellt werden kann, stellt sich.

Auf inhaltlicher Ebene lässt sich zunächst der Bedarf an einem allgemein anerkannten Begriffsverständnis festhalten. Die in diesem Zusammenhang relevanten Elemente (Anbieter / System / Gerät, Suche / Web-Suche / Sprachsuche, Persönlicher Assistent / Sprachsteuerung etc.) müssen eindeutig voneinander differenziert und beschrieben werden. Weiterer Forschungsbedarf ergibt sich im Hinblick auf Fragen der Personalisierung, des Datenschutzes oder der tatsächlichen Barrierefreiheit für Menschen mit Behinderungen.

Trotz der beschriebenen Schwächen und weiterhin offenen Fragen kann das Projekt als erfolgreich bezeichnet werden. Die identifizierten Erfolgskriterien bilden einen soliden Ausgangspunkt für die weitergehende Auseinandersetzung mit der sprachgesteuerten

Websuche und die Weiterentwicklung entsprechender Anwendungen. Die Studierenden, die bis zum Projekt weder über Erfahrung in der sprachgesteuerten Websuche noch in der Durchführung von Usability- oder Suchmaschinenevaluationen verfügten, konnten von der vermittelten Methodenvielfalt profitieren und für das Thema der sprachgesteuerten Websuche sensibilisiert werden.

8 Literaturverzeichnis

- [Bi16] Bitkom: Sprachsteuerung setzt sich bei Smartphones durch, <https://www.bitkom.org/Presse/Presseinformation/Sprachsteuerung-setzt-sich-bei-Smartphones-durch.html>, 2016.
- [Br86] Brook, J.: SUS – A quick and dirty usability scale, http://cui.unige.ch/isi/icle-wiki/_media/ipm:test-suschart.pdf, 2016.
- [eR14] eResult, <http://www.eresult.de/ux-wissen/forschungsbeitraege/einzelansicht/news/mobile-web-navigation-navigieren-und-orientieren-auf-dem-smartphone/>, Stand: 11.04.2016
- [eR13] eResult, <http://www.eresult.de/ux-wissen/forschungsbeitraege/einzelansicht/news/usability-testing-von-responsive-webdesigns/>, Stand: 11.04.2016
- [Go15] Google, <https://think.storage.googleapis.com/images/micromoments-guide-to-winning-shift-to-mobile-download.pdf>, Stand: 01.04.2016
- [LH07] Lewandowski, D.; Höchstötter, N.: Qualitätsmessung bei Suchmaschinen. System- und nutzerbezogene Evaluationsmaße, Informatik Spektrum 30(2007)3.
- [PA16] Parks Associates: Almost 40% of U.S. smartphone owners use voice recognition software, <http://www.parksassociates.com/blog/article/pr-01272016>, 2016.
- [Qu12] Quirnbach, S.: Suchmaschinen User Experience, Usability und nutzerzentrierte Website-Gestaltung, Springer-Vieweg, 2012
- [SK16] Sünkler, S.; Kerkmann, F.: Ok Google... The End of Search as we know it. Projektbericht zum Studienprojekt „Ok Google... The End of Search as we know it. Sprachgesteuerte Suche im Test“, <http://searchstudies.org/ok-google/>, Stand: 04.05.2016

Sinnsuche noch günstiger: wie Google an Inbegriffen scheitert


Irene Teich¹, Peter Schnupp²

Abstract: Die semantischen Erkennungseigenschaften auch hochentwickelter Suchmaschinen sind noch immer unbefriedigend. Der Aufsatz spürt den Gründen nach und findet sie in der mangelnden Berücksichtigung des heutigen „Wissens über Wissen“: der Konstruktiven Epistemologie, Wittgensteins „Sprachspielen“ und den daraus abgeleiteten „Inbegriffen“ als Wortsymbole, die abhängig von dem jeweiligen Kontext („Wirklichkeit“ des Benutzers) gänzlich unterschiedliche Bedeutungen haben können. Es wird gezeigt, dass in diesen Fällen die derzeitige Chomskysche Linguistik als Grundlage des Analyse- und Suchvorgangs zwar nicht theoretisch aber praktisch nicht verwendbar ist. Eine brauchbare Alternative bietet eine assoziative sprachliche „Mustererkennung“ auf der Basis der *Wittgensteinschen* Linguistik. Deren Implementierung und die ersten marktreifen Produkte werden kurz beschrieben.

Keywords: Semantische Suche, Konstruktivismus, Wirklichkeit, Chomsky, Wittgenstein, Sprachspiele, Linguistik.

1 Einleitung

Eine Google-Recherche nach „Sinnsuche“ überraschte kürzlich neben einigen hunderttausend anderen Ergebnissen mit dem Resultat



Sinnsuche noch günstiger -
Preisvergleich ist unsere Stärke!
www.guenstiger.de
Sinnsuche noch günstiger.

Dass es sich hier nicht um ein Standard-Suchergebnis sondern wohl um eine danebengegangene „semantische Recherche“ für eine *Adword*-Anwendung handelt, macht den Unsinn nicht sinniger sondern stellt ganz im Gegenteil eine grundlegenden Mangel jeder suchwort-orientierten Recherche heraus.

Dieser Beitrag erklärt die wissenschaftstheoretisch sehr tiefen Grundlagen dieser generellen Schwäche sowie ein bereits implementiertes und auf dem Markt eingeführtes, völlig anders arbeitendes Konzept, das sich an der seit fast einem halben Jahrhundert dahinschwindenden, assoziativ-wissensbasierten Literaturrecherche umfassend gebildeter Bibliothekare und Dokumentare orientiert. Es beruht auf der Konstruktiven

¹ AIQ.go Enterprises GmbH & Co. KG, Zentrale, Im Mediapark 5, 50670 Köln, teich@aiq-go.de

² AIQ.go Enterprises GmbH & Co. KG, Entwicklung, Senftenberger Straße 19, 02977 Hoyerswerda, schnupp@aiq-go.de

Epistemologie von *Watzlawick* und *Rupert Lay*. Zur Implementierung der Grundsoftware sowie der ersten, bereits auf dem Markt eingeführten Applikationen verwendet es anstelle der bisher in der Informatik praktisch ausschließlich verwendeten Linguistik von *Chomsky* diejenige von Wittgenstein.

2 Konstruktive Epistemologie, Wirklichkeiten, Sprachspiele, Inbegriffe

Um diese Implementierung zu verstehen, ist es notwendig, kurz einige Begriffe und Ergebnisse von Nachbarwissenschaften (Epistemologie, Konstruktivismus, *Wittgensteins* „Philosophischer Linguistik“, und der von *Rupert Lay* als Folge unterschiedlicher Wirklichkeiten eingeführten „Inbegriffe“) zu referieren, die bis jetzt in der Informatik kaum bekannt sind und noch weniger verwendet werden.

Beginnen wir also mit dem Konstruktivismus. Nach der in den letzten Jahrzehnten vor allem von *Watzlawick* und *Rupert Lay* [Wa2005, La2015a, La2015b] entwickelten *Konstruktiven Epistemologie* ist die von der (ersten) Aufklärung propagierte Auffassung, dass alle „vernünftigen“ Menschen die gleiche, sich an der *Realität* orientierende *Wirklichkeit* besitzen oder zumindest anstreben, schlicht falsch – eine Einsicht, die übrigens schon Sokrates in den platonischen Dialogen immer wieder demonstrierte.

Stattdessen *konstruiert* sich jeder Mensch *Wirklichkeiten* aus seinen Erfahrungen, Wünschen, Überzeugungen sowie, vor allem, seinen sozialen Umgebungen und deren Sprachgemeinschaften. *Wittgenstein* nennt sie *Sprachspiele*. Aus diesen stammen die *Begriffe* sowie die sie bezeichnenden *Wortsymbole*. Und da ein Mensch bestenfalls anstreben kann, dass seine Wirklichkeiten der unbekannteren Realität entsprechen, sind auch durch dasselbe Wortsymbol benannte Begriffe semantisch sehr unterschiedlich – eine Erkenntnis, die schon durch *Thomas von Aquin*³ formuliert aber in den Jahrhunderten dazwischen wieder vergessen wurde: Alles Erkante wird im Erkennenden nach der Weise des Erkennenden erkannt.

Wem all dies zu abstrakt oder zu spekulativ erscheint, können zwei inzwischen leider ebenfalls vergriffene Bücher empfohlen werden. Beide beschreiben sehr unterschiedliche Wirklichkeiten und die in ihnen entwickelten Sprachspiele – sofern dieser harmlose Namen als wissenschaftlicher Term auch in dem dort beschriebenen Kontext gestattet ist: in beiden Büchern ist dieser das zum Glück nur 12 Jahre währende aber auf Tausend Jahre geplante Dritte Reich, dessen damals „Weltanschauung“ genannte Wirklichkeit und die spezielle Sprache, die *Victor Klemperer* [Kl1947] im Buchtitel als *Lingua Tertii Imperii* bezeichnet. Als jüdischer Philologe, der diese zwölf Jahre vor allem auch durch die Loyalität seiner nicht-jüdischen Ehefrau in Dresden überlebte und seine sprachliche Umwelt protokollierte und analysierte, liefert sein Buch

³ *Cognitum autem est in cognoscente secundum modum cognoscentis* (Thomas von Aquin, *Summa theologiae*, I, q.12, a. 4)

vor allem sehr gute Beispiele für die Umwertung von Wortbedeutungen. Diese vor inzwischen siebzig Jahren gewonnene Einsicht ist das Beste den Verfassern bekannte Beispiel für ein erst vor kurzem von *Rupert Lay* (wieder-) entdecktes Konzept des heutigen Konstruktivismus, die „Inbegriffe“.

Auch das zweite dieser Bücher entstand unmittelbar nach dem Ende des Dritten Reichs. Der Wiesbadener Philologe *Heinrich Reichert* [Re1948] beschäftigt sich in ihm mit zwei sehr unterschiedlichen Wirklichkeiten und den daraus entstandenen Sprachspielen: dem des römischen Rechts-, Geschäfts- und Alltagslebens („Human und Urban“, wie der Titel der zweiten Auflage seines Buchs lautet) im Gegensatz zur Weltanschauung des gerade vergangenen und unmittelbar erlebten Dritten Reichs. Während *Klemperer* seine Analysen weitgehend auf einzelne Worte und damit auf die Herausarbeitung von „Inbegriffen“ in der Sprache des Dritten Reichs konzentriert, arbeitet *Reichert* die fundamentalen Unterschiede der römischen Wirklichkeit und der nationalsozialistischen Weltanschauung aus den „Lateinischen Sentenzen“ (dem ursprünglichen Titel seines Buchs) heraus, genau der semantische Ansatz, den Wittgenstein in seiner Linguistik verfolgt und fordert.

Die gerade eingeführten *Inbegriffe* spielen als Bezeichnung für viele, oft sehr unterschiedlich konstruierte Begriffe unterschiedlicher Wirklichkeiten in der konstruktiven Erkenntnislehre eine zentrale Rolle. Die *Sinnsuche* aus unserem Titelbeispiel gehört ebenso dazu wie die Mehrzahl der für jeden intellektuellen Diskurs besonders wichtigen Worte wie *Gerechtigkeit*, *Freiheit*, *Bildung*, ... Eine „naive“ wortorientierte Suche in großen Wissensbasen oder gar im gesamten Internet muss daran zwangsläufig scheitern.

Wesentlich ist, dass die Vielfalt der in einem Inbegriff enthaltenen und verstandenen Begriffe dem suchenden Menschen noch nicht einmal bewusst ist. Damit ist er denn auch der Interpretation eines solchen Inbegriffs durch eine Suchmaschine hilflos ausgeliefert, es sei denn, die Ergebnisse sind so offensichtlich unsinnig wie die obige „Sinnsuche“.

Die Lösung liegt in einem anderen Begriffs- und Sprachmodell für die Computer-Linguistik. Ein entsprechendes, wissens technologisches Objektmodell und eine darauf beruhende „konstruktivistischen Recherche-Software“ sind inzwischen implementiert.

3 Inbegriffe und Wirklichkeiten

Die Implementierung einer Ontologie und einer befriedigenden Semantischen Recherche über einer großen, vielfältigen Textbasis wird umso schwieriger und schließlich hoffnungslos, je unterschiedlicher die Wirklichkeiten und die (In-)Begriffsdefinitionen der Autoren einerseits und der Recherchierenden andererseits sind. Wenn ein Inbegriff schon je nach dem ihm verwendenden Sprecher sehr unterschiedliche Inhalte haben kann: was hindert mich dann, ihn einfach einmal zu verwenden, wenn mir gerade keine

passende Bezeichnung einfällt? Etwa die „Gerechtigkeit“ – da sie leider ein Inbegriff ist, schützt sie niemand davor, dass sich ein Politiker, ein Wirtschaftler, ein Philosoph auf sie beruft, wenn ihm die richtigen Worte fehlen? *„Gewöhnlich glaubt der Mensch, wenn er nur Worte hört, es müsse sich dabei doch auch was denken lassen“* sagt Mephisto in der Hexenküche von Faust I. So dass sein Schöpfer, *Goethe*, durchaus auch den Anspruch auf die Erfindung der Inbegriffe erheben könnte, wenn er es denn nötig hätte.

Nun aber zu Alltagsanfragen, wie sie täglich zu hunderten an Google oder andere Suchmaschinen gestellt werden. Und vor allem auch zu den auch heute schon von ihnen verlangten und auch geleisteten semantischen Interpretationen von Anfragen und dem „Sinn“ der in ihnen enthaltenen Suchworte. Zuerst ein einfaches Beispiel:

Mercedeshändler Köln.

Natürlich geht der Anfragende davon aus, dass die Suchmaschine bei der Aufbereitung und Verarbeitung der Anfrage einige einfache semantische Umformungen und Ergänzungen vornimmt. Sie sollte also das erste Wort in seine beiden Bestandteile zerlegen. Und sie sollte vermutlich den (In)Begriff „Händler“ auch durch einige andere Inhalte wie „Vertretung“, „Vertreter“, „Niederlassung“ ergänzen. Ob sich in „seiner Wirklichkeit“ ein bestimmter Autohändler lieber als Niederlassung sieht, und dass dies für ihn aus rechtlichen, geschäftlichen oder sonstigen Gründen vielleicht sogar sehr wichtig ist, interessiert den prospektiven Autokäufer nicht. Damit sind diese beiden Wirklichkeiten auch problemlos mit einander verträglich. Und es ist auch völlig unerheblich, ob diese Unterschiede jemandem bewusst werden oder nicht.

Nicht ganz so einfach ist es mit dem ersten Teilwort, „Mercedes“. Will der Anfragende tatsächlich unbedingt einen Mercedes, oder ist das für ihn nur ein Sammelbegriff für ein Premium-Auto, so dass eine semantische Suche ihm auch Audi-, BMW- oder Jaguar-Händler nachweisen sollte? Und wenn nicht – wie ist es mit einer „Daimler-Vertretung“? Hier lohnt es sich schon eher, vor der Konstruktion und Implementierung eines semantischen Recherchesystems, über Unterschiede in Wirklichkeiten und Inbegriffen nachzudenken. Aber „falsche“ Entscheidungen sind nicht sehr folgenreich. Werden nach „Mercedes“, vielleicht mit einer etwas geringeren Bewertung durch den Suchalgorithmus, auch andere Premium-Marken gesucht und als Ergebnisse ausgegeben, wird dies den Anfragenden nicht allzu sehr stören, selbst wenn in seiner Wirklichkeit ein Mercedes „das einzige ordentliche Auto“ wäre.

Und umgekehrt – beschränkt sich die Suchmaschine ausschließlich auf Mercedes, kann der Recherchierende ja einfach alle ihn interessierenden Marken in seiner Anfrage nennen und sich darauf verlassen, dass diese bei der Verarbeitung als Alternativen interpretiert werden. Oder er kann einfach für jede ihm sonst noch in den Sinn kommende Automarke eine entsprechende Anfrage nachreichen. Wobei es aber wichtig ist, wie schnell er die Antwort auf die Recherche erhält. Sind es Sekunden, so ist das überhaupt kein Problem. Dehnen sich aber die Antwortzeiten auf Stunden, Tage, Wochen aus, wie es in vor-digitalen Jahrhunderten immer und heute noch bei anspruchsvollen Recherchen etwa für Gerichtsgutachter oder in der Wissenschaft die

Regel ist, gilt das nicht mehr. Fehlinterpretationen, oder in unserem Modell Fehlanpassungen von Wirklichkeiten und Inbegriffen, können dann nicht mehr „praktisch umsonst“ durch erneute Anfragen korrigiert werden. Wir werden darauf noch zurückkommen.

Zuerst aber eine kleine Veränderung der Suchanfrage:

Gotteshäuser Köln.

Jetzt ist es plötzlich nicht mehr uninteressant, was in den Wirklichkeiten des Anfragenden und der Autoren der Dokumente in der Wissensbasis ein „Gotteshaus“ ist. Der Kölner Dom wird vermutlich von allen als solcher verstanden werden, genauso wie die Pfarrkirche einer christlichen Gemeinde in einer Kölner Vorstadt. Aber wie ist es mit einer Synagoge oder einer Moschee? Mit einer Meditations-Stätte einer Sekte oder in einem Krankenhaus? Mit einer kleinen Waldkapelle irgendwo im Grünen oder einem Versammlungs- und Verwaltungshaus von Scientologen? Oder der Ruine einer Klosterkirche aus dem Mittelalter?

Schon wird es bedeutsam, wie Vertreter unterschiedlicher Wirklichkeiten diesen Inbegriff interpretieren. Und es sind nicht nur emotionale Verärgerungen, wenn etwa jemand mit einer mohammedanischen Wirklichkeit seine Moschee nicht als Gotteshaus anerkannt sieht. Auch „sachlich“ kann für ihn, ebenso wie die Vertreter anderer Wirklichkeiten, die Klärung dieser semantischen Frage sehr wichtig und relevant sein: etwa wenn eine Gemeinde die Baugenehmigung für eine Moschee verweigert und es darum geht, ob dies gegen im Grundgesetz garantierte Freiheiten verstößt.

Man könnte nun versuchen, das Problem einfach auf die Ontologen abzuwälzen, welche die semantischen Definitionen für die Suchmaschine erstellen: sollen die doch bitte „Gotteshaus“ so definieren, dass alle vernünftigen Wirklichkeiten sich in dieser Definition wiederfinden. Diese Aufgabe ist unlösbar, aus verschiedenen Gründen.

Aber, wie man in einem anderen, ebenfalls empfehlenswerten Buch von Rupert Lay über den Konstruktivismus [La2015b], nachlesen kann, sind wertende Attribute wie „vernünftig“ oder „richtig“ auf Wirklichkeiten nicht anwendbar. Das ergibt sich schon daraus, dass das einzige Maß dafür die Realitätsnähe wäre. Und dass ein Mensch die Realität nicht erkennen kann, weiß man seit mehr als zweitausend Jahren. Die einzige, leidlich „vernünftig“ einer Wirklichkeit zuschreibbare Eigenschaft ist „nützlich“ im ethischen Sinn. Aber die stellt sich leider, wenn überhaupt, erst im Nachhinein an den Folgen heraus: dass der Nationalsozialismus und seine Wirklichkeit nützlich waren, wird angesichts des von ihm angerichteten Elends wohl kaum noch ein „vernünftiger“ Mensch behaupten.

Aber setzen wir uns einmal über diese ganze Wirklichkeits-Theorie hinweg und nehmen als Informatiker einfach einmal an, wir müssten den Begriff „Gotteshaus“ in das Objektmodell einer Ontologie einbauen, vielleicht mit einer Parametrisierung der zugrundeliegenden Wirklichkeit als „christlich“ „mohammedanisch“, „agnostisch“ ...

Wir werden vermutlich damit beginnen, dass wir ihn, ähnlich wie den „Mercedeshändler“ im ersten Beispiel, in die Begriffe „Gott“ und „Haus“ zerlegen. Und wenn wir das „Haus“ nicht unmittelbar als „Wohnhaus“ definieren sondern ein Gebäude, das irgendwie „Gott“ gewidmet ist oder ihm verbundenen Funktionen oder Handlungen dient, werden uns vermutlich Vertreter aller drei betroffenen Wirklichkeiten folgen.

Aber was meint „Gott“? Ist das ein Inbegriff? Vermutlich nicht – was der derzeitige Konstruktivismus dazu meint, werden wir am Ende dieses Abschnitt noch kurz ansprechen. Vorerst einmal: wie „konstruieren“ wir einen Gottesbegriff. Schon da gibt es die ersten Probleme: einige religiöse Wirklichkeiten, die mohammedanische ebenso wie die jüdische, verbieten dies sogar ausdrücklich. Also müssen wir einen anderen Weg suchen, zum Beispiel die Übernahme des abrahamitischen Gottesbegriffs, den zumindest Christen, Juden und Mohammedaner gleichermaßen anerkennen. Auch Agnostiker finden ihn meist akzeptabel: wenn es „in Wirklichkeit“ keinen Gott gibt, ist es auch gleichgültig, wie andere Wirklichkeiten ein solches Abstraktum konstruieren. Und das gilt auch für viele anderen religionsartigen Gemeinschaften – auch für sie gibt es entweder keinen Gott im abrahamitischen Sinn, oder es gibt viele Götter, und da kommt es, auf die Eigenschaften von einem von ihnen nicht mehr besonders an.

Doch wir stoßen auf andere Schwierigkeiten. Zum einen schließt die Beschränkung auf den abrahamitischen Gott einige der obigen Prätendenten auf die Bezeichnung Gotteshaus aus. Das wäre noch zu ertragen. Aber für einen Juden oder einen Mohammedaner ist zwar Gott Vater abrahamitisch, aber nicht Jesus. Im Verständnis eines Mohammedaners ist er allenfalls der zweitgrößte Prophet nach Mohammed. Und für einen Juden ist er wohl noch nicht einmal das. Noch schlimmer steht es um den Heiligen Geist. Sogar aus Sicht mancher Christen ist er nur ein Konstrukt, das Kaiser Konstantin 325 während des Konzils zu Nikäa einführte, um Angesichts der Anerkennung von Jesus als „Sohn Gottes“ den Anspruch des Christentums als monotheistischer Religion durch eine typische Kompromiss-Lösung aufrechterhalten zu können: man erfindet einen dritten Begriff und fasst sie als „Dreieinigkeit“ zu verschiedenen Erscheinungen des gleichen Gotts zusammen.

Wenn auch dies einem Agnostiker vielleicht noch gleichgültig ist, kann man es einem in einer mohammedanischen oder jüdischen Wirklichkeit lebenden Menschen nicht übelnehmen, wenn er dies bei aller Toleranz nicht nachvollziehen kann. Und wenn kaum bestreitbar ist, dass im Kölner Dom neben Gott Vater auch Jesus und der Heilige Geist verehrt werden – disqualifiziert es ihn nicht vielleicht als Gotteshaus?

Aus einem halben Jahrhundert Erfahrung in der Software-Entwicklung und in der Künstlichen Intelligenz: so etwas kann nur scheitern, schon in der Konzeption unter dem herkömmlichen Linguistik- und Objektmodell. Wir können nicht gegen die Einsichten des Konstruktivismus und die Inbegriffe angehen, sondern wir müssen unmittelbar auf beidem aufbauen.

Ach ja: ist „Gott“ eigentlich ein normaler Begriff, wie „Auto“, „Haus“, oder auch „Mercedes“? Sicher nicht – dazu sind die Bedeutungen in den verschiedenen

Wirklichkeiten viel zu unterschiedlich. Ist es also ein Inbegriff? Auch das ist er nicht: viele, vor allem religiöse Wirklichkeiten akzeptieren weder viele verschiedene, durch das Wort bezeichnete aber unterschiedliche Begriffe. Das ist viel zu nahe an Vielgötterei oder zumindest Konstrukten wie der christlichen Dreieinigkeit. Und einen Gottes-Inbegriff zu konstruieren verbieten, wie bereits gesagt, viele Religionen mit gutem Grund. Dass es nicht funktionieren kann, haben wir ja gerade gesehen.

Der moderne Konstruktivismus [La2015a] führt deshalb eine weitere Begriffsklasse ein, welche er als „Urbegriffe“ bezeichnet. Er nimmt an, dass sie in allen oder zumindest den meisten Menschen entweder von Geburt an oder in seiner frühen Sozialisation angelegt und dann in seine Wirklichkeiten eingebaut werden. Neben „Gott“ gehören dazu vor allem emotional besetzte Begriffe wie „Mutter“, „Liebe“, „Gewissen“. Ob er sie dann im Laufe seines Lebens irgendwie rationalisiert oder immer „aus dem Bauch heraus“ erfühlt, hängt von seiner Persönlichkeit, seiner Bildung und seiner Einbettungen in Kommunikations- und Lebensgemeinschaften ab. Auch hier sind, wie bei allen Wirklichkeiten „richtig“, „falsch“ oder „wahr“ schlicht keine anwendbaren Attribute.

Auch auf diese Urbegriffe und den Umgang mit ihnen könnte wieder Goethe die Urheberrechte anmelden. In Herrmann und Dorothea sagt er von den „Kindern“, auch einem solchen Urbegriff: „So wie Gott sie uns gab, so muss man sie haben und lieben.“ Und wir Nachfolger sollten keine Ontologie, keine Semantik und keine semantische Recherche akzeptieren, die nicht mit Gott, Kindern, Müttern und ähnlichen Urbegriffen umgehen kann ...

4 Konstruktive Wissensrecherche

Repetieren wir kurz die wesentlichen Probleme einer Recherche. Sie soll sich nicht auf ein oder wenige Suchworte beschränken. Statt dessen soll wenigstens insoweit „semantisch“ sein, dass sie, wie in unserem obigen Beispiel, etwa zusammengesetzte oder flektierte Worte in die Stammformen der Bestandteile zerlegen und leidlich mit mehreren Interpretationen eines solchen Textelements umgehen kann. Eine Aufgabe ist, mit verschiedenen, aus einem Inbegriff wirklichkeitsabhängig zu extrahierenden Bedeutungen umzugehen. Entsprechend hat sie die einfachere, aber ähnliche Aufgabe der Unterscheidung von Synonymen, etwa für das von *Wittgenstein* als Beispiel verwendete Wort „König“, das einmal die Schachfigur und ein andermal das Staatsoberhaupt einer Monarchie zu erfüllen. Wie *Wittgenstein* bemerkt, ist aus dem Einzelwort nicht zu ersehen, was gemeint ist. Findet man es in einem längeren Text, so ist meist sofort klar, was gemeint ist. In aller Regel weiß der Leser, ob er gerade einen Text über Schach, über Politik oder in einem Lifestyle-Blatt liest. Aus dem jeweiligen Kontext, dem Sprachspiel der besprochenen Wirklichkeit, weiß er dann auch, welcher König gemeint ist.

Natürlich kann auch Google nicht erkennen, was für einen König es suchen soll, wenn man bloß das einzelne Wort eingibt. Vielleicht kann eine Semantik-Software durch

Annahme aus Bezügen zu anderen eingegebenen Worten eine mehr oder weniger sichere Klassifizierung vornehmen. Ist das zweite Suchwort „Turm“ oder „Dame“, meint der Recherchierende vermutlich die Schachfigur, und ist es „Rochade“, so ist diese Interpretation so gut wie sicher. Aber kann man vom Benutzer die Eingabe der „richtigen“ Zweit- oder Dritt-Suchwörter dem Benutzer überlassen?

Ist es nicht besser, zu den alten, bewährten Rechercheverfahren zurückzukehren. Handelte es sich um komplexere Rechercheaufgaben in kaum oder gar nicht bekannten „Wirklichkeiten“, ging man zu seinem Bibliothekar oder Dokumentar. Und natürlich warf man dem nicht einfach ein oder zwei Worte an den Kopf, etwa „Gotteshäuser Köln“ oder die „Sinnsuche“ aus dem Titel. Sondern man gab ihm ein oder mehrere Texte und bat ihn um ähnliches, verwandtes Material. Das konnte eine Kundenanfrage sein, Zeitungsartikel, Gerichtsurteile – was immer die gerade anliegende oder vermutete Wirklichkeit so hergab. Nach ein paar Tagen, oder in schwierigen Fällen ein paar Wochen legte ein guter Dokumentar seinem Auftraggeber die Ergebnisse seiner „semantischen Recherche“ vor: Dokumente als Kopien oder Sonderdrucke, Zettel mit den wichtigsten Begriffen, ihrer Bedeutung im aktuellen Kontext und Referenzen auf die Fundstellen, „Teaser“, also kurze, relevante Textauszüge – ähnlich unseren obigen Goethe-Zitaten.

Wie das unser „Digitaler Dokumentar 2.0“ auf dem Bildschirm als Ergebnis einer Recherche nach Eingabe eines Suchtexts tut, zeigt Abbildung 1.



Abb. 1 Das Rechercheergebnis des digitalen Dokumentars

Natürlich braucht er dazu nicht Tagen oder Wochen sondern einige Sekunden, wie wir es

von unseren Computern und dem Internet gewohnt sind. Die Fundobjekte sind kleine viereckige Knöpfe, die dann das jeweilige Fundobjekt bei Darüberfahren mit der Maus oder nach einem Mausklick anzeigen oder aufblättern.

Schon haben wir den digitalen Dokumentar erfunden – wir müssen ihn nur noch implementieren. Aber da stoßen wir auf praktische Probleme, die wir bereits oben theoretisch abgehandelt haben. Listen wir sie auf, bevor im nächsten Abschnitt kurz ihre Lösung beschreiben.

- Die Software muss „irgendwie“ aus den Eingabetexten neben den Inhalten und Aussagen des Verfassers auch die „Wirklichkeit“ erkennen und entnehmen, in und aus der Autor seinen Text verfasst hat.
- Sie soll aus der Dokumenten- oder Wissensbasis die Dokumente finden, „assoziiieren“, die nicht nur den Inhalten sondern auch der Wirklichkeit des Suchtexts am besten entsprechen.
- Dabei muss sie diejenigen Begriffe – vor allem auch als Begriffsinhalte der Inbegriffe – extrahieren, welche diese Assoziation nahelegen und verursachen.
- Sie sollte dem Auftraggeber nicht nur diese Begriffe sondern auch Referenzen auf ihre Herkunft und ergänzende Informationen liefern, welche dem Benutzer helfen, diese Begriffe sowohl im Kontext der Wirklichkeiten, sowohl die der Dokumentenautoren als auch seiner eigenen zu verstehen und zu bewerten.
- Und schließlich sollte sie dem Benutzer nicht nur Begriffe sondern auch kurze Textausschnitte vorlegen, welche beim Verständnis der gefundenen Dokumente und den Auffassungen und Informationen ihrer Verfasser am besten helfen.

Wenn wir behaupten würden, wir hätten den Digitalen Dokumentar entweder top-down von der geeigneten Benutzeroberfläche oder bottom-up aus einem entsprechenden Objektmodell nach den Regeln der Softwaretechnologie entworfen und realisiert – der Leser würde es uns doch nicht glauben. Deshalb beschränken wir uns lieber darauf, auf welche Weise wir die einzelnen Komponenten fanden und zusammenstellten.

5 Der Digitale Dokumentar

Beim Versuch, diese Teilleistungen des Digitalen Dokumentars durch ein Objektmodell und seine Methoden softwaremäßig zu implementieren, stößt man auf mancherlei ungelöste Probleme. Die meisten entstehen durch den Konstruktivismus als Modell für die Wissensgewinnung, Wissensanalyse, Wissensverdichtung und Darstellung des dem Auftraggeber vorgelegten Rechercheergebnisses, das wir als „Wissensobjekt“ bezeichnen und dessen Visualisierung Abb. 1 zeigte. Die meisten dieser Schwierigkeiten brachte die Akzeptanz gleichberechtigter aber oft sehr unterschiedlicher (aber als solche noch nicht einmal erkannter) Wirklichkeiten und entsprechend der sehr unterschiedlichen individuellen Begriffsinhalte der von der Sprachgemeinschaft zu ihrer

Konstruktion und Beschreibung verwendeten Wortsymbole, der Inbegriffe. Manche konnten mit bereits existierenden Modellen und Techniken, allenfalls nach leichten Anpassungen, behandelt und gelöst werden. Andere benötigten, zumindest nach unserer Einschätzung, das völlige Verlassen der eingeführten Grundlagen (der „Paradigmen“ nach *Thomas Kuhn* [Ku1991] und Übergang zu einem ganz neuen Konzept. Dies gilt vor allem für das verwendete Linguistik-Modell und, eng damit verbunden, der Ontologie zur Beschreibung der wesentlichen Begriffe. Hier adoptierten wir anstelle der in die Informatik allgemein eingeführten *Chomsky*-Linguistik diejenige von *Wittgenstein*.

5.1 Erkennung und Analyse von Kontexten und „Wirklichkeiten“

Zuerst einmal: die uns interessierenden Wissensbasen sind nicht „Big Data“, also keine großen, verteilten und im Wesentlichen relational organisierten Datenbanken. Sie sind eher „Big Wisdom“, unformatierte Texte in zahlreichen Nationalsprachen und – wie wir inzwischen von Wittgenstein gelernt haben – noch weit zahlreicheren „Sprachspielen“. Die Grundlage dafür ist deshalb die Erkennung und Analyse von Satz- und Wortstrukturen und ihrer Zusammenhänge. *Chomsky* verwendet dazu eine im Wesentlichen lineare Syntaxanalyse, wobei die Klassifizierung in Wortarten und deren (mutmaßliche) Bedeutungen durch Nah-Kontexte erfolgt.

Wittgenstein behauptet und belegt statt dessen, dass dies zumindest dann nicht ausreicht, wenn man Wortsymbole im Kontext von Sprachspielen und der dadurch induzierten Vielfalt der Inbegriffe (in heutiger Sprechweise) semantisch verstehen und verarbeiten will. Dazu braucht man längere Phrasen und Sätze als primäre Elemente des Analyseprozesses, die dann einerseits der ersten Erkennung und Klassifizierung der in ihnen verwendeten Wortsymbole dienen und andererseits durch das über den ganzen Text aufgespannte Netz dieser Teilelemente, der „Sätze“, die genaue semantische Diskriminierung der Wortsymbole ermöglichen, also im obigen Beispiel erkennen lassen, von was von einem „König“ die Rede ist.

Auf den ersten Blick ist nicht zu sehen, wieso eine *Chomsky*-Linguistik nach entsprechender Erweiterung und Ausbau nicht auch diese, von Wittgenstein gestellte Erkennungsaufgabe leisten kann. Auch wir haben daran nicht gezweifelt und wurden darin sogar bestärkt, als wir auf ein Buch namens „Grammar of English Grammars“ stießen, eine von *Goold Brown* [Br1851] verfasste „Metagrammatik“ des Englischen, die wir deshalb auch in [TS2016] ausführlicher referieren. Es ist eine ausführliche, verbale aber trotzdem in den für uns wichtigsten Abschnitten sehr formale Anleitung, wie man eine englische Grammatik so schreiben soll, dass nach ihr entweder ein korrekter Text in dieser Sprache erstellt oder ein vorhandener analysiert („parsed“) werden kann, und zwar nicht nur syntaktisch sondern, soweit möglich, auch semantisch korrekt.

Das Buch ist in mehrerer Hinsicht unglaublich. Zum einen könnte nach ihm ein leidlich geschickter Softwareentwickler, zwar mit beträchtlichem Zeitaufwand aber ohne große Probleme einen Parser für englische Texte in einem bestimmten Sprachspiel schreiben – „könnte“, weil selbst in der Heimatstadt von *Goold Brown*, New York, Computer und

Software-Technologen erst über einhundert Jahre später Einzug hielten: das Buch wurde 1851 veröffentlicht. Und zeitaufwendig wäre es, selbst bei der exzellenten Vorbereitung, auch – das Buch enthält weit über 2000 Seiten, sämtlich mühelos und zeitweise sogar direkt kurzweilig zu lesen, aber trotzdem eine recht umfangreiche „Spezifikation“. Das hätte uns noch nicht abgeschreckt. Aber das immer wieder, wenigstens für die wichtigsten natürlichen Sprachen und für jede Wirklichkeit und die von ihr induzierten Sprachspiele – also letztlich mehrfach für jeden betroffenen Menschen ... nichts gegen *Chomsky* als theoretische Alternative, aber praktisch kam er für uns nicht mehr in Frage, nachdem wir das realisiert hatten.

5.2 Assoziative Recherche mit konventionellen Suchmaschinen

Was aber dann? Alles, von der Basissoftware bis zur Benutzeroberfläche neu zu konzipieren und zu implementieren kam nicht in Frage, selbst wenn uns ein großes Unternehmen oder ein großer Wagnisfinanzierer etliche Millionen zur Verfügung gestellt hätte. 50 Jahre Programmiererfahrung lehrten den älteren der beiden Autoren, dass die ideale Größe eines Entwicklerteams eine Person ist, oder allenfalls zwei, falls einer von ihnen an einen Baum fährt oder etwas ähnlich Unsinniges tut.

Also mussten wir uns die einzelnen Komponenten zusammensuchen, möglichst eingeführte Industriesoftware oder bewährte Standard-Techniken, und die wenigen verfügbaren Mannjahre für den programmtechnischen Kitt dazwischen aufsparen. Das meiste machte auch keine Probleme: die Einbettung in Microsoft-Office und vor allem *Word*, sowie Ben Shneiderman's „TreeMaps“ [Sh1998]. Bloß – wo gab es eine für unsere Zwecke geeignete Suchmaschine, für das Internet sowie proprietäre Wissensbasen. Eine, die sich nicht nur „semantisch“ nannte, sondern auch wenigstens einige der obigen Grundfunktionalitäten wie die Aufspaltung von zusammengesetzten Worten in die Komponenten, möglichst mit Erkennung von deren Grundformen und die Angabe von Synonymen, wie etwa „Verkauf“ oder „Niederlassung“ neben „Handlung“ in unserem Mercedes-Beispiel? Und das mit einem „Verstehen“ von Wirklichkeiten, Sprachspielen und Inbegriffen bei der Auswahl, der Assoziation von Dokumenten in der Datenbasis zu einem eingegebenen Suchtext?

Es gab sie nicht. Und ein entsprechend semantisches Google in vielleicht zwei Jahren neu zu schreiben – das trauten sich die beiden Programmierer unseres Entwicklerteams auch bei aller Selbstüberschätzung nun doch nicht zu. Aber sie waren erfahren genug, Programmierung einfach durch Software-Architektur zu ersetzen:

- einen Präprozessor, der aus dem vorgelegten Text einen wortorientierten Suchausdruck extrahierte, welchen eine verfügbare Suchmaschine – in unserem Fall Microsoft Bing – gerade noch verarbeiten konnte,
- einen Compiler, welcher die zurückgelieferten Fundtexten (ähnlich denen von Google) parste und so gut wie möglich in die Struktur aus Dokumenten, Begriffsbeschreibungen und „Semantische Splitter“ übersetzt, die wir schon oben

als „Wissensobjekt“ bezeichneten, und

- eine proprietäre Implementierung von Ben Shneiderman's TreeMaps, neben dem Compiler für die Bing-Ausgabe die einzige wirklich neu konzipierte und geschriebene Komponente des Gesamtsystems.

Erfreulicherweise reichen die etwa 30 Suchworte die Bing in einer Anfrage verarbeiten kann zusammen mit den gebotenen Möglichkeiten ihrer Verknüpfung durch logische Operatoren (OR, AND, NOT) sowie der eingebauten, wenn auch beschränkten semantischen Unterstützung aus, durch eine entsprechende Aufbereitung der Eingabe und Nachbearbeitung der Ergebnis-Textstruktur aus, eine sicher nicht vollkommene aber in vielen Fällen durchaus brauchbare „Wittgenstein“-Assoziation der Dokumente in der Wissensbasis zu erreichen.

5.3 Begriffsextraktion und -analyse

Der Mechanismus beruht im Wesentlichen auf dem schon oben beschriebenen, mustererkennung-ähnlichen Phänomen, dass in vielen Fällen einfach das gleichzeitige Vorkommen bestimmter Worte in einem Text ausreicht, ihn und die diesen Worten zugeordneten Begriffe einem Sprachspiel zuzuordnen: in *Wittgensteins* Beispiel waren dies „König“, „Turm“, „Bauer“, „Rochade“, „matt“ oder ähnliche, um das Thema als „Schach“ und den „König“ als Schachfigur zu identifizieren.

Wenn vor allem in sehr großen Dokumentenbasen und durch die Beschränkung auf etwa dreißig Suchworte die vollautomatische Erstellung der Suchanfrage und die Möglichkeiten der Nachbearbeitung durch den Compiler nicht ausreichen, gibt es zwei Wege, die Ergebnisse nachhaltig zu verbessern. Der Benutzer kann durch die Eingabe zusätzlicher Wort-Bezeichner für das „Thema“, seine „Absicht“ oder „unerwünschte Dokumente“ dem System Hinweise für die Suche und deren Aufbereitung geben. Der Präprozessor baut sie in den hauptsächlich aus Worten des vorgegebenen Recherchetexts bestehenden Suchausdruck mit geeigneten logischen Operatoren ein.

Zum anderen erwies es sich als erstaunlich produktiv, einfach die Recherche in wiederholten Schritten zu verfeinern: entweder indem der Benutzer die ihm besonders zusagenden Objekte (Dokumente, Begriffe, Semantische Splitter) in eigene Wissensobjekte sammelt, die dann persistent in seiner persönlichen Wissensbasis gesammelt werden. Oder er kann diese besonders guten Treffer, seien es Dokumente, Begriffe oder Textsplitter, einfach als Ausgangstexte für eine erneute Recherche verwenden. Auf diese Weise reichern sich die semantischen Kriterien zur Auswahl eines bestimmten Sprachspiels immer mehr an, und die nächsten, durch die Suche und den Compiler erzeugten Wissensobjekte entsprechen immer mehr dem Sprachspiel und der aktuellen Wirklichkeit des Benutzers.

Wie schon oben erwähnt, ist dies nur praktikabel, weil die Software für eine gründliche, semantische Recherche eben nicht Tage oder Wochen sondern nur noch Sekunden

braucht. Dies ist der eine, wesentliche Fortschritt der neuen Recherche-Software.

Den zweiten entdeckten wir rein zufällig: weil die linguistische Analyse hier nicht auf einer Syntax-Analyse aufbaut sondern diese nach *Wittgenstein* durch eine „linguistische Mustererkennung“ ersetzt, funktioniert sie unabhängig von der natürlichen Sprache der Such- und Zieldokumente. Wir erprobten dies bereits, ohne jede Änderung der Software, mit deutschen, englischen, spanischen und russischen Eingabe-Texten. Und mit chinesischen, aber dort konnten wir die Qualität der gefundenen Ergebnisse auf Grund unserer ungenügenden sprachlichen Kompetenz nicht nachprüfen ...

5.4 Der WisARD als integriertes Software-Hardware-Produkt

Inzwischen gibt es diesen Digitalen Dokumentar nicht nur als Basissoftware zur Implementierung von Anwendungen auf PCs, in lokalen Netzen oder in der Cloud.



Abb. 2 WisARD: der fertig konfigurierte Digitale Dokumentar

Es gibt ihn auch unter dem Namen WisARD (Wissens-Akquisition, Recherche und Darstellung) auf dem Laptop AIQ.Research, fertig installiert vor allem für Studenten, Wissenschaftler und Wissensarbeiter in der Politik oder Wirtschaft.

Und es gibt weiter ausgebaute Systeme, zur konsensorientierten Entscheidungsfindung in Teams oder Diskussionsgruppen von Personen auch mit unterschiedlichen Wirklichkeiten. Oder zur Entscheidungsorientierten Projektplanung und Steuerung nach Hubbard [Hu2014]. Oder zur Einrichtung, der laufenden Fortschreibung und optimalen Nutzung der Qualitätsdokumentation nach der ISO 9001:2015. Aber über diese Applikationen und deren Möglichkeiten werden andere Aufsätze berichten.

Literaturverzeichnis

- [Br1851] Brown, G.: Grammar of English Grammars, 1851, Kindle Edition, Amazon Media EU S.a.r.l., 2011.
- [Kl1947] Klemperer, V.: LTI – Notizbuch eines Philologen. Reclam Verlag, Leipzig, 1947.
- [Re1948] Reichert, H. G.: Lateinische Sentenzen, Essays. Dieterich'sche Verlagsbuchhandlung, Wiesbaden, 1948.
- [Sh1998] Shneiderman, B.: Treemaps for space-constrained visualization of hierarchies, 1998/2014, <http://www.cs.umd.edu/hcil/treemap-history/>.
- [Hu2014] Hubbard, Douglas W.: How to Measure Anything, 3rd Edition, Wiley, Hoboken, N.J. 2014.
- [TS2016] Teich, I.; Schnupp, P.: Semantical Grammar. AIQ.go, Köln, Kindle Edition 2016.
- [La2015a] Lay, R.: Konstruktivismus, Grundlagen der Wissenstechnologie, Band 3, Amazon Kindle, 2015.
- [La2015b] Lay, R.: Die zweite Aufklärung – Einführung in den Konstruktivismus, Verlagshaus Monsenstein und Vannerat, Münster 2015.
- [Ku1991] Kuhn, Thomas: Die Struktur wissenschaftlicher Revolutionen. Suhrkamp, Frankfurt am Main, 1991.
- [Wa2005] Watzlawick, Paul: Wie wirklich ist die Wirklichkeit - Wahn, Täuschung, Verstehen; Piper, München, 2005. Original 1976.

Visualisierung von Suche im Dialog

Olaf Resch¹

Abstract: Internetsuchen sind ein wesentlicher Teil des täglichen Lebens aber auch ein Teil von anspruchsvoller professioneller Recherchearbeit geworden. Meist werden für die alltägliche Suche und die professionelle Recherche die gleichen Werkzeuge eingesetzt. Der Beitrag argumentiert, dass eine Suche immer einen Dialog darstellt und dieser Dialog bewusst für eine spezielle Zielsetzung gestaltet werden muss und demnach für anspruchsvolle Recherchen anders ablaufen muss als bei der alltäglichen Suche.

Keywords: Internet, Suchmaschine, Recherche, Dialog.

1 Einleitung

Etwas im Internet zu suchen und zu finden wird häufig mit den Mainstream-Suchmaschinen assoziiert, was beispielsweise durch das Synonym *googeln* zum Ausdruck kommt. Suche ist aber natürlich viel mehr und so wird auch in Shops oder in Intranets nach Informationen gesucht. Diese Arbeit stellt einen Ansatz vor, der Suche als Dialog begreift. Dieser Dialog muss explizit und unter Beachtung der jeweiligen Rahmenparameter gestaltet werden, um ein optimales Ergebnis für alle Beteiligten zu erzielen.

Zur schrittweisen Konkretisierung dieses Ansatzes wird zunächst der Aspekt der Suche als Dialog näher beleuchtet, danach werden einige Konzepte vorgestellt, die für die Gestaltung konkreter Dialoge hilfreich sind und im Anschluss erfolgt eine exemplarische Beschreibung einer Suche als Dialog anhand der Anwendung serioussersuch.de. Es handelt sich bei dem vorliegenden Beitrag um einen Erfahrungsbericht und nicht um eine umfassende wissenschaftliche Arbeit, insbesondere fehlt die Diskussion der relevanten Literaturlbasis und es werden lediglich einige zentrale Basiswerke zitiert.

2 Suche als Dialog

Die Suche – und das beinhaltet im Folgenden auch das Finden – wird hier als zielorientierter Prozess der Informationsbeschaffung begriffen. Dabei wird die Auffassung vertreten, dass dieser Prozess immer einen Dialog darstellt. Beispielsweise sucht das Kind nach Ostereiern, indem es im Dialog mit seiner Umwelt steht, ganz banal

¹ Hochschule für Wirtschaft und Recht Berlin, Informationsmanagement und E-Business, Alt-Friedrichsfelde 60, 10315 Berlin, olaf.resch@hwr-berlin.de

z. B. *unter diesem Kissen ist also kein Osterei*, über das *heiß* und *kalt* seiner Eltern bis zu komplexeren Erwägungen wie *diese Vase sieht nach einem guten Versteck aus*. Der Suchprozess ist somit von allen Dialogpartnern abhängig. Das Kind will die Eier finden und weist dafür eine mehr oder weniger ausgeprägte Kompetenz und Geduld auf und die Eltern können die Eier so verstecken, dass sie schwieriger oder leichter zu finden sind und eventuell noch weitere Ziele verfolgen, z. B. die schönsten Eier beim Klavier oder in der Schulmappe verstecken.

Bei der Suche im Internet läuft es nicht anders und es handelt sich immer um einen Dialog; zwischen dem Kaufinteressenten und dem Webshop, dem Googler und Google, dem Studi., der seine Thesis schreibt und Springerlink usw. Dieser Dialog kann aber sehr wohl und sehr bewusst eher Nachfrager- oder Anbieter-seitig gestaltet werden. Ein Beispiel für einen extrem Anbieter-seitigen Dialog ist die Startseite eines Webshops. In diesem Dialog beschränkt sich der Anteil des Nachfragers darauf, eben diesen Webshop und nicht einen anderen zu besuchen. Der Webshop bestimmt dagegen vollständig über das was und wie der Präsentation. Bei der klassischen Web-Suche ist der Anteil des Nachfragers am Dialog natürlich wesentlich höher, da der Nachfrager die Suchmaschine wählt, die Suchbegriffe eingibt und ggf. verfeinert und selber entscheidet, wie er mit dem Suchergebnis umgeht. Aber auch in diesem Fall existieren Möglichkeiten, den Dialog zu gestalten. Ganz banal geschieht dies zunächst durch das Ranking, das darüber bestimmt, was weiter oben steht und damit einfacher gefunden wird. Es gibt darüber hinausgehend aber auch noch viele weitere Möglichkeiten, den Dialog zu bestimmen und somit im Sinne des Informationsanbieters zu lenken. Eine besonders wichtige Rolle spielt dabei die visuelle Gestaltung der Dialogelemente, die meist unter dem Stichwort User Interface (UI) thematisiert wird und die den Schwerpunkt der nachfolgenden Ausarbeitung bildet.

3 Rahmenbedingungen der Gestaltung von Suchdialogen

3.1 Usability und User Experience

Die Qualität der Interaktion einer Website über ihr User Interface mit ihren Nutzern wird häufig unter den Stichworten Usability und User Experience (UX) thematisiert. Dabei beschreibt die Usability, wie gut die Website nutzbar ist. User Experience geht darüber hinaus und behandelt neben der Usability weitere Aspekte, wie die Funktionalität und die emotionale Wirkung einer Website. [Mo04] Eine (leicht) abweichende Auffassung sieht Usability als den Teil der direkten Interaktion und User Experience zusätzlich auch als die vor- und nachgelagerten Teile, den Erwartungen an diese Interaktion sowie deren Nachwirkungen. [DI11]

Teilweise stehen Elemente der User Experience sogar in Widerspruch zueinander, beispielsweise fördert es eine besondere User Experience, wenn dem Nutzer mehr Funktionen zur Verfügung stehen, insbesondere, wenn diese Funktionen neu sind und

ein Alleinstellungsmerkmal gegenüber anderen Angeboten darstellen. Auf der anderen Seite sind mehr und neuere Funktionen auch schwieriger zu nutzen, verschlechtern somit zunächst die Usability und damit auch die User Experience. Solche Widersprüche müssen bei der Gestaltung von Dialogen beachtet werden, indem der Usability bei neueren Funktionen eine besondere Aufmerksamkeit zuteilwird.

3.2 Personas, Proto-Personas und User Stories

Als Personas werden hier im Sinne von *Cooper* stilisierte Nutzer bezeichnet [Co04], z. B. *Joe Doe ist 32 Jahre alt und macht ein Aufbaustudium zum Interaktionsdesigner. Er sucht nach neuen Fachbegriffen.....* „Richtige“ Personas müssen in einem aufwändigen Prozess durch Interviews mit (potentiellen) Nutzern und anschließender Generalisierung abgeleitet werden. Eine einfache Form von Personas stellen die sogenannten Proto-Personas dar, die nicht auf der Basis von tatsächlichen, sondern lediglich auf angenommenen Merkmalen beruhen. Personas dienen der Eingrenzung des Dialoges auf bestimmte Zielgruppen. Je mehr Personas beachtet werden sollen, desto differenzierter muss die Gestaltung des Dialoges erfolgen.

Zur Präzisierung der Interaktion von Personas mit dem System finden z. B. User Stories Verwendung. User Stories stammen aus der agilen Softwareentwicklung und sind ein einfaches Mittel zur Spezifikation von Anwendungen. [We99] Dabei wird aus der Sichtweise von Nutzern in knappen Sätzen beschrieben, was diese mit dem System machen, z. B. *als Nutzer kann ich nach einer exakten Wortkombination suchen.*

3.3 Machine Cues

Als Machine Cues bezeichnet man Hinweise, die aus der technischen Umgebung abgeleitet werden können. [Cl16] Sehr häufig findet z. B. die IP Adresse zur Ermittlung des Standortes Verwendung. Dafür gibt es insbesondere bei mobilen Endgeräten aber auch noch exaktere Hinweise. Machine Cues können als Non-verbale Kommunikation interpretiert werden und in diesem Sinne auch den Dialog mitbestimmen.

3.4 Multi-sided vs. Single-sided

Auch das Geschäftsmodell einer Anwendung spielt für die Gestaltung des Dialoges eine Rolle. Als Single-sided wird ein Modell bezeichnet, bei dem lediglich ein Akteur beachtet werden muss, bei einem Multi-sided-Geschäftsmodell dagegen mehrere. [OP09] Sofern ein und derselbe Akteur eine Leistung in Anspruch nimmt und gleichzeitig dafür bezahlt, handelt es sich meist um Single-sided-, im Falle von werbefinanzierten Angeboten, bei denen neben den Interessen der primären Leistungsempfänger auch die Interessen der Werbetreibenden beachtet werden müssen, dagegen meist um Multi-sided-Geschäftsmodelle.

4 Gestaltung eines Suchdialoges für serioussearch.de

serioussearch.de ist eine Suchanwendung, die auf der Basis des Bing-Indexes sowie der Dienste einiger weiterer Informationsanbieter tiefere Recherchen ermöglicht. [Re14] Im Folgenden wird serioussearch.de als Suchmaschine bezeichnet, was den Bing-Index mit einbezieht. Zielgruppe sind Studierende, Wissenschaftler und andere Rechercheprofis. Da serioussearch.de den genutzten Index nicht verändern kann, muss die Kernfunktionalität der tiefergehenden Recherche durch eine besondere Form der Präsentation realisiert werden.

4.1 Rahmenbedingungen

Eine besondere User Experience ist für das Ziel der tiefergehenden Recherche zentral. Dazu werden einige Funktionen angeboten, die den Nutzern nicht aus den Mainstream-Suchmaschinen bekannt sind und deren Gebrauch auch nicht selbsterklärend ist. Die intendierte emotionale Wirkung der Anwendung ist es, Ruhe zu vermitteln und die Dialogbereitschaft zu fördern, da tiefergehende Recherchen Zeit und die aktive Mitarbeit der Nutzer benötigen. Ein positives Beispiel ist Google-News, das wesentlich stärker zu einer Auseinandersetzung mit den Inhalten einlädt als die Google-Suche.

Als sehr einfache Proto-Personas können Studenten, Journalisten und Wissenschaftler angenommen werden. Allen gemeinsam ist der Wunsch nach einer tiefergehenden, nachvollziehbaren und dokumentierten Recherche. Alle kennen sich gut mit dem Internet als Recherchefeld aus und haben eine erhebliche Erwartungshaltung, die insbesondere durch die Mainstream-Anwendungen geprägt ist. Zu unterscheiden sind aber neue und erfahrene Benutzer, da diese ganz andere Anforderungen an die Usability stellen.

Mögliche Machine Cues sind der Ort, z. B. zur Voreinstellung der Sprache und die Bildschirmgröße sowie die technischen Interaktionsmöglichkeiten, insbesondere das Vorhandensein eines Touch-Screens, um mithilfe dieser Informationen die UI-Elemente optimal anordnen zu können. Weiterhin wäre es interessant festzustellen, welchen Zugang der Nutzer zu kostenpflichtigen Informationsanbietern, z. B. SpringerLink hat, damit diese Informationsanbieter bei der Recherche Berücksichtigung finden können.

Neben den primären Nutzern von serioussearch.de sind zusätzlich auch die Informationslieferanten zu bedenken, von denen die Anwendung abhängig ist. Dafür gibt es teilweise rechtliche Gründe, teilweise ist es aber auch der Fairness geschuldet, insbesondere dann, wenn die Informationen kostenfrei geliefert werden. serioussearch.de ist somit eine Multi-sided Anwendung. Microsoft als Betreiber des Bing-Index stellt erhebliche Anforderungen an die Art der Präsentation und bestimmt somit den Dialog wesentlich mit. Aber auch die anderen Anbieter, z. B. die Wikipedia, das DOAJ und Open Street Map stellen spezifische Anforderungen an die Art, wie ihre Informationen präsentiert werden dürfen. Es stellt daher eine wesentliche Herausforderung bei der

Dialoggestaltung dar, die Anforderungen der primären Nutzer und die der Informationslieferanten miteinander abzugleichen.

4.2 Dialoggestaltung

Zur Beschreibung des Dialoges finden einfache User Stories Verwendung. Im Folgenden wird der Dialog für einige exemplarische Situationen in einem typischen Dialogverlauf beschrieben.

Aufruf

S: Als Anwender möchte ich ein Werkzeug für tiefere Internetrecherchen nutzen.

A: Die Anwendung ist konsequent auf tiefere Recherchen ausgerichtet und kann diese Nachfrage auch am besten bedienen. Für alltägliche Suchen kann [serioussearch.de](#) zwar ebenfalls verwendet werden, dafür stehen aber auch genug andere Suchmaschinen, auch im nicht Mainstream-Bereich zur Verfügung. Dies wird durch Hinweise und entsprechende Funktionalitäten auch deutlich zum Ausdruck gebracht, um keine falschen Erwartungen zu befördern.

Neuer vs. erfahrener Nutzer

S: Als neuer Nutzer möchte ich gerne viele Erklärungen zu den UI-Elementen und aktiv auf interessante Funktionen hingewiesen werden.

S: Als erfahrener Nutzer möchte ich möglichst wenig Ablenkung bei meiner Recherche.

A: Die Anwendung ist so konzipiert, dass sie möglichst intuitiv zu bedienen ist. Bei Funktionen, die herkömmliche Suchmaschinen nicht bieten, sind Hinweise und Erklärungen aber durchaus hilfreich. Diese werden durch erfahrene Nutzer allerdings als störend empfunden. [serioussearch.de](#) verlässt sich an dieser Stelle auf die Selbsteinschätzung der Nutzer und zeigt auf Nachfrage eine sehr umfangreiche Benutzerführung an, die jedoch immer wieder ausgeschaltet werden kann.

Erweiterte Suchoptionen

S: Für eine optimale Recherche stelle ich oft möglichst präzise Suchanfragen.

A: Der Bereich für erweiterte Optionen ermöglicht die üblichen Einstellungen. Da eine möglichst präzise Nachfrage eine möglichst zutreffende Informationsbereitstellung erleichtert, wird der Nutzer zu dieser Präzisierung ermutigt. Die Präzisierung wird durch entsprechende UI-Elemente unterstützt, spiegelt sich aber transparent im Eingabefeld wider und kann dort direkt bearbeitet werden, um den Dialog zu fördern.

Feld zur Eingabe von Suchbegriffen

S: Als Rechercheprofi wähle ich jeden Suchbegriff wohldefiniert.

A: serioussearch.de unterstützt diesen wichtigen Teil des Dialoges zur Formulierung durch ein großes und prominent platziertes Eingabefeld. Bei der Suche wird zwischen groß- und kleingeschriebenen Begriffen unterschieden.

Autocomplete

S: Als Nutzer hilft es mir, Begriffe hinsichtlich ihrer Schreibweise und Bedeutung zu überprüfen.

A: Der Autocomplete zeigt abhängig von der jeweiligen Eingabe eine Liste relevanter Begriffe. Die Liste wird durch Existenz und Relevanz ermittelt. Anders als bei den Mainstream-Suchmaschinen werden hier nicht die Begriffe gelistet, nach denen andere gesucht haben, sondern Begriffe, zu denen ein Wikipedia Eintrag existiert. Sobald der Anwender sein Interesse an einem Begriff durch eine Fokussierung kommuniziert, zeigt die Anwendung eine kurze Begriffserklärung zu diesem Begriff an.

Suchbuttons

S: Als Anwender will ich teilweise ganz neu suchen und teilweise einen bestehenden Ergebnisraum um neue Ergebnisse erweitern.

A: Der Anwender kann dies durch Betätigen eines der beiden Buttons kommunizieren. Bei einer Erweiterung bleiben die bisherigen Einträge erhalten und alle Operationen erstrecken sich auf die alten und die neuen Ergebnisse.

Suchergebnisse

S: Als Anwender will ich im Internet recherchieren.

S: Als Anwender will ich auch Hinweise auf weiterführende Informationen erhalten.

A: Aktuell werden Ergebnisse für die Web-Suche, für Basis-Infos und wissenschaftliche Artikel ermittelt. Sofern Ergebnisse für die einzelnen Bereiche gefunden werden, legt serioussearch.de eine eigene Slide und ein Icon im Navigationsbereich an, mithilfe derer eine Slide aufgerufen werden kann. Der Nutzer kann die Suche aber auch auf eine reine Web-Suche beschränken.

Web-Suche: Die Suchbegriffe und eventuelle Verfeinerungen über die erweiterten Suchoptionen werden an den Bing-Index weitergegeben. Das Ergebnis entspricht zunächst der klassischen SERP (Search Engine Result Page) der Mainstream-Suchmaschinen. Für eine einzelne Suche werden maximal 50 Ergebnisse angezeigt. Allerdings kann der Nutzer diese Ergebnismenge per

Anschlussuche oder durch Erweiterung mit neuen Suchbegriffen beliebig ausdehnen. Bei der Darstellung der Ergebnisse sind immer auch die Vorgaben des Bing-Index zu beachten.

Basis-Infos: serioussearch.de ermittelt Basis-Infos und sofern vorhanden auch Vorschaubilder und eine Karte für zentrale Suchbegriffe. Auch hier sind die Vorgaben der Informationslieferanten zu beachten.

Research To Go (wissenschaftliche Artikel): Je nach Nutzer wird für wissenschaftliche Artikel nur im DOAJ (Digital Open Access Journals) Verzeichnis oder zusätzlich auch in proprietären Datenbanken, wie z. B. Springerlink recherchiert. Um die Berechtigung des Nutzers abzuleiten, werden Machine Cues verwendet. Auch hier sind die Vorgaben der Informationslieferanten zu beachten.

Tree-View: Der Tree-View ist eine zusammenhängende Darstellung aller Suchergebnisse. Der Nutzer erhält dadurch interaktiv Antworten auf Fragen, wie z. B.: Welche Schlüsselwörter hängen mit einem Suchergebnis zusammen? oder: Welche wissenschaftlichen Artikel erweitern ein bestimmtes Suchergebnis? Auch hier sind die Vorgaben der Informationslieferanten zu beachten.

Details vs. Übersicht

S: Für Recherchen benötige ich teilweise eine Übersicht über die Suchergebnisse und teilweise Detailinformationen.

A: serioussearch.de ermöglicht es beim Tree-View bei Bedarf auch Details anzuzeigen. Bei den Ergebnissen der Web-Suche werden die Details standardmäßig angezeigt, lassen sich auf Nachfrage aber auch ausblenden.

Filter

S: Für tieferegehende Recherchen möchte ich den Ergebnisraum filtern können.

A: Für den Tree-View und die Web-Ergebnisse steht ein Keyword-basierter Filter zur Verfügung. Dieser ermöglicht eine mehrstufige Filterung. Die notwendige Relevanz der angezeigten Keywords lässt sich ebenfalls mehrstufig einstellen. Zusätzlich bietet der Filter die Möglichkeit, sich kurze Erklärungen zu den einzelnen Keywords anzuzeigen. Wenn die aktuelle Ergebnismenge zur Befriedigung der Suchanfrage nicht ausreicht, kann der Filter zur Initiierung von Anschlussuchen verwendet werden.

Blacklist

S: Als Anwender möchte ich Webseiten ausblenden können, die für mich irrelevant sind.

A: Die Anwendung erlaubt es, eine Blacklist mit irrelevanten Webseiten zu

führen. Dort aufgeführte Seiten werden für die aktuelle Suche ausgefiltert und bei zukünftigen Suchen nicht mehr berücksichtigt. Die Blacklist lässt sich als Cookie auch für zukünftige Besuche von serioussearch.de speichern.

Suche im Dialog mit anderen

S: Bei Recherchen möchte ich auch mit anderen zusammenarbeiten können.

A: serioussearch.de unterstützt die Suche im Team durch eine Art Screen-Sharing. Nachdem das Team sich per Einladung und Annahme formiert hat, kann immer genau ein Mitglied ein „Muschelhorn“ an sich nehmen und damit die Anzeige aller Team-Mitglieder mitbestimmen.

5 Fazit

Jede Suche kann als Dialog zwischen einem oder mehreren Suchenden und der Umwelt aufgefasst werden. Bei Suchen im Internet wird diese Umwelt wesentlich von der verwendeten Suchmaschine geprägt. Diese kann den Dialog mehr oder weniger stark bestimmen und sie kann das aus unterschiedlichen Gründen tun. Aktuell ist die Visualisierung von Ein- und Ausgaben das Hauptinstrument der Dialoggestaltung. Das kann sich in Zukunft aber durchaus ändern, wenn Spracheingaben und -ausgaben eine größere Rolle spielen.

Auf jeden Fall muss der Dialog mit einem oder auch mit mehreren Nutzern unter Beachtung der jeweiligen Zielsetzung und Rahmenparameter explizit gestaltet werden. Somit kann auch davon ausgegangen werden, dass eine Universal-Suchmaschine, die jegliche Zielsetzung bedienen will, stets suboptimale Ergebnisse liefern wird.

Der vorliegende Beitrag wirft wesentlich mehr Frage auf, als er beantwortet. Insbesondere fehlen wirklich verlässliche Informationen darüber, wie Menschen Suchmaschinen für unterschiedliche Problemstellungen nutzen und welchen Einfluss Erwartungen, z.B. aufgrund eines Markenimages auf die Bewertung von Ergebnissen haben. Es ist auch nicht wirklich bekannt, ob und wie externe Faktoren beispielsweise Zeitdruck den Rechercheprozess mitbestimmen. Tatsächlich ist der Komplex Suche im Vergleich zu seiner Bedeutung in der öffentlichen Forschung noch stark unterrepräsentiert.

Literaturverzeichnis

[Cl16] Clark, J.A.: Anticipatory Design: Improving Search UX using Query Analysis and Machine Cues. Weave Journal of Library User Experience Issue 4/Vol 1. DOI: <http://dx.doi.org/10.3998/weave.12535642.0001.402>.

[Co04] Cooper, A.: The Inmates are Running the Asylum: Why high-tech products drive us

crazy. Sams Publishing, 2004.

- [DI11] DIN EN ISO 9241-210:2011-01, Ergonomie der Mensch-System-Interaktion - Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme.
- [Mo04] Morville, P.: User Experience, http://semanticstudios.com/user_experience_design, Stand: 25.6.2016.
- [OP09] Osterwalder, A.; Pigneur, Y.: Business Model Generation, Selbstverlag, 2009.
- [Re14] Resch, O.: Internetsuche für wissenschaftliche Arbeiten am Beispiel von serioussearch.de. In (Barton, T. et al Hrsg.): Angewandte Forschung in der Wirtschaftsinformatik Prozesse, Technologie, Anwendungen, Systeme und Management, Verlag News & Media, Berlin, S. 188-196, 2014.
- [We99] Wells, D.: User Stories, <http://www.extremeprogramming.org/rules/userstories.html>, Stand: 25.6.2016.

Value-Sensitive Design of
Internet-Based Services:
Towards an Integration of
Technology and Values

Value-Sensitive Design of Internet-Based Services: Towards an Integration of Technology and Values

Hannes Hartenstein¹, Alexander Maedche² and Martina Zitterbart³

Dear readers,

This is the first time the workshop on „Value-Sensitive Design of Internet-Based Services: Towards an Integration of Technology and Values” takes place as part of the German “INFORMATIK 2016” conference in Klagenfurt on September 29th 2016. The workshop brings together researchers that deal with questions of value-sensitive design and delivery of Internet-based services from different disciplinary and interdisciplinary perspectives.

Internet-based services are increasingly important for our society. We consume them on a daily basis for many different purposes, comprising chat applications, social networks, shopping and the like. With the Internet of Things and advanced applications such as smart cities and smart home we are expecting huge potentials for a more sustainable society at large, covering the economical, ecological and the social dimension. However, the provisioning of Internet-based services comes with conflicting interests between the different stakeholder groups of providers, consumers and governance actors. A prominent example of such a conflict is using consumer data as the “currency” for the consumption of Internet-based services, e.g. delivered via mobile Apps. Resulting conflicts between consumers and providers occur on a continuous basis, as personal data is delivered to the provider and, thus, compromises the privacy of the consumer. However, mostly the consumer is left without any choice and, in this sense self-determination of the individual consumer is limited. Governance actors, again, have a different view which is heavily motivated through societal values, such as human rights and constitutional law or regulations and policy instruments. They can also help finding a balance in the case of interest conflicts between providers and consumers.

At the one-day workshop two invited talks and five papers are presented: Karsten Weber and Peter Reichel provide in their invited talks two different perspectives on value-sensitive design and delivery in the context of information and communication technologies: Karsten Weber focuses on the question why ethics for building good computers is important. Peter Reichel discusses the “Internet of People (IoP)” and specifically perspectives of an anti-copernican revolution in networking research. Marcus Ast and Martin Gaedke deal in their paper about values apply in hyper-

¹ KIT Karlsruhe, hannes.hartenstein@kit.edu

² KIT Karlsruhe, alexander.maedche@kit.edu

³ KIT Karlsruhe, zitterbart@kit.edu

connected societies and present a concept for integrating functional and non-functional values into technology platforms. Teemu Kärkkäinen and Jörg Ott propose in their work a Liberouter network, which decentralizes both the communication infrastructure and services. Neudecker et al. specifically look at access control systems and propose and compare design alternatives for value-sensitive design. Burger et al. present their paper titled “Popping the bubble: principles of information neutrality and counter measures against biased information”. In their work, they investigate information processing mechanisms and their influence on net neutrality as well as suggest counter measures against biased information. Complementing the provider and technology-oriented perspectives, Dorner et al. present an experimental design of a value-oriented decision support system for selecting Internet-based services. Complementing the invited talks and the five paper presentations, a brainstorming session following the “Zukunftswerkstatt” concept is done as part of the workshop. The underlying idea of this session is to jointly develop new ideas or solutions for fundamental problems of designing and delivering value-sensitive Internet-based services.

We hope that the workshop is establishing a good starting point and a community for further extending research in the exciting field of value-sensitive design and delivery of Internet-based services. At this point we would like to thank all participants for their contributions to make this workshop a success: the authors, the program committee and the event organizers.

Thank you!

Hannes Hartenstein, Alexander Maedche and Martina Zitterbart

Value-Sensitive Design in Hyper-Connected Societies

Markus Ast¹, Martin Gaedke²

Abstract: In hyper-connected societies, where the opposite of communication and interaction may unobtrusively change from a human to a bot or another device, underlying services and technologies can easily come into conflict with cultural and social values. In this paper, we discuss how values apply to hyper-connected societies and we elaborate how values could be functionally integrated into platforms that provide and combine both human-to-human, human-to-service and human-to-thing communication and interaction.

Keywords: Value-Sensitive Design, Hyper-Connected Society, Internet of Services, Internet of Things, Human-computer Interaction

1 Introduction

With the dramatic growth of the web, our lives have been fundamentally changed. Our society is increasingly transforming into a networked society. We now expect “instant online access to information, our contacts and goods and services from everywhere in the world”, we participate in multiple communities—often through multiple online personae—, and we “make new acquaintances and business contacts, find nearby friends, and navigate completely new worlds” through the internet[Er15].

Values have a huge influence of how we human beings undertake communication and interaction between ourselves. These values apply not only in face to face communication and interaction of our society, but also when using information and communication technology to communicate and interact with others. However, our connected society is transforming into a hyper-connected society, where communication is ubiquitous, omnipresent, and communication partners might be humans, services, or things. In the hyper-connected society, we communicate and interact with more and more new actors in our daily lives. New actors, like a variety of different services of the Internet of Services or Internet of Things (IoT) devices, and an increasing amount of avatars and bots, i.e., artificial intelligences. It is thereby often not obvious if we currently communicate with a human or a machine, and sometimes communication and interactions switches from a machine to a human in an unobtrusive manner. For certain humans, communication with non-humans, or the switch from human-machine to human-machine communication might violate with their values.

Value is thereby often referred to as the “equivalent monetary worth of a specified sum

¹ Technische Universität Chemnitz, Germany, markus.ast@informatik.tu-chemnitz.de

² Technische Universität Chemnitz, Germany, martin.gaedke@informatik.tu-chemnitz.de

or amount”[Ox10]. However, a more broader meaning of value is “the estimation in which a thing is held according to its real or supposed desirability or utility”, where such “estimation [is] regarded in relation to an individual or group”[Ox10]. The more specific meaning of the term *Value* in the context of this paper, is “the generally accepted or personally held judgement of what is valuable and important in life”[Ox10]. Such values are for instance: *privacy* [Ro04], *ownership* and *property* [LB00] and *autonomy*[Hi91].

In this paper, we want to start the discussion of how values apply to hyper-connected societies, i.e., to communication and interaction between humans, services and things. We will point out pitfalls when designing systems that support hyper-connected societies and elaborate possibilities of how to integrate values into systems to have actual effects on functionality.

2 Related Work

Designing information and communication technology in a way that takes values into account, is not a new topic. The so called value-sensitive design (VCD) “is a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process”[FKB08]. The research area of value-sensitive design emerged because values can “seemingly conflict with economic goals and can be difficult to articulate clearly and to translate into meaningful processes and designs”[Fr96]. Besides other notable work in VCD, like [FHN09] and [Ho07], there are also efforts about shifting the focus of HCI towards a value-centered design process[Co04].

The Business Motivation Model (BMM) describes two directives that are similar to how values apply in organizations. These two directives are *Business Rules* and *Business Policies*. A Business Policy “is a Directive that is not directly enforceable whose purpose is to govern or guide the enterprise” and a Business Rule “is a Directive, intended to govern, guide, or influence business behavior, in support of Business Policy that has been formulated in response to an Opportunity, Threat, Strength, or Weakness”[Om15].

3 Values in Hyper-Connected Societies

More and more communication and interactions in our networked society are not only human to human anymore. Human to machine and also machine to human interactions are increasing. Borders between human-human and human-machine are thereby often not clear and changing even during an ongoing interaction in an unobtrusive manner. A recent example is Facebook’s personal assistance *M*. Rather than always responding to assistance queries using an artificial intelligence (AI) analyzes of the query, Facebook’s *M* has a team of humans behind the scene. This team can jump in and help out on queries, where AI is not yet well trained. That is, for the user, it is not always clear if

they are talking to the AI or to a human being. Furthermore, during a conversation, the opposite can change from human to AI, or the other way around.

Another example is GitHub, where multiple bots have emerged over the years to take over some repeating tasks, like checking whether a Contributor License Agreements has been signed for incoming pull requests, or like automatically upgrading dependencies and running tests accordingly to ensure that these upgrades do not break the application. Occasionally, this leads to having bots talking to bots.

With technology like the Facebook Messenger Platform³, which allows to easily integrate services into Facebook's messenger, it is to be expected that human to machine communication is increasing rapidly in the next couple of years. This makes it even more important to take values into account, when designing such services and integrations.

Current research focuses on taking values during the design process into account. However, we think values should also be integrated into actual operation of systems and as such paving the way for a culture of a hyper-connected society. According to [Fr96], two common issues that concern values are: user autonomy and freedom from bias. They further identified four aspects of systems that can promote or undermine user autonomy: System Capability, System Complexity, Misrepresentation of the System and System Fluidity. The other concern of freedom from bias is about avoiding biased technology, i.e., to avoid to "systematically and unfairly discriminate [...] certain individuals or groups of individuals in favor of others"[Fr96]. Such values can be supported by improving implementation-related parts of the application.

However, in a hyper-connected society, platforms are not monolithic applications anymore, but rather compositions and integrations of many third party services and devices/things. That is, such third party services are provided as is. The influence of the compliance with values is limited to how the integration of such services is done and represented. Furthermore, values are often not dependent on functionality. We therefore propose to differentiate values into *functional values* and *non-functional values*, which are introduced in detail in Section 4.

It is also important that when implementing values into a system, they must not be used to enforce a certain behavior. Because then we are talking about rules and not about values anymore. The difference between rules-oriented and values-oriented culture is how much emphasis is placed upon controlling human thoughts and behavior[JS06]. Rules are more about governing behavior, while values are about guiding behaviors. For a healthy and thriving culture, an optimal blend of both rules and values is necessary[JS06]. While some rules are necessary, other rules that would discriminate a certain user group, may not necessary and are better replace by values. That is, there are always functionalities and especially security related features, that have to be enforced

³ <http://messengerplatform.fb.com>

by rule. Values instead are not enforced, but supported, which also enables the coexistence of individuals with different values.

4 Functional and Non-functional Values

Values are often not dependent on functionality, but rather on the level of effort that is put into something. For this reason, we propose to differentiate values into *functional values* and *non-functional values*, where both kinds can be implemented as follows.

Functional Values are values that provide practicable guidance. That is, functional values can be automated inside a system. However, the functionality that automates guidance of functional values must not be enforced. That is, instead of enforcing a certain functionality, implementations of functional values can provide recommendations of how a technology or software is used. Taking for example a value of *transparency*, and a platform where employees inside an organization can manage their goals. When defining goals, read access to them can be set as either public or private. Since, in this case, transparency is a value and not a rule, both options are still available. However, upon creating a private goal, a dialog may appear asking to reconsider the decision of creating a private instead of public goal.

Non-Functional Values are values that are not directly practicable, but still guide and steer decisions and actions of an organization. Such values can be integrated into a system as measurements. These measurements can thereby be about the system itself or about the behavior of the organization that provides the system. Taking for example the value of being responsive to customer inquiries. By measuring response time and whether inquiries are answered at all, the non-functional value of being responsive could be observed. In case the compliance with such a non-functional value is, according to the measurements, not satisfying, further steps can be taken inside an organization.

We are working on integrating functional and non-functional values into goal-oriented composition of apps, services and things based on the SmartComposition[KWG14] approach. In the center of goal-oriented composition is a means and ends structure, which is partially as shown in Fig. 1. In this structure, means are hierarchically broken down into Mission, Strategy, Operation and Task. Respectively, ends are broken down into Intent, Goal, Purpose and Result. Means thereby depict the course of action (COA) that is taken to achieve the desired effects, which in return is described by corresponding ends. All ends have success measurements attached to them. This enables to track their level of accomplishment. Furthermore, it allows to distinguish between both a project's completion (through progress in means) and success (through progress in ends).

Fig. 1 shows, besides the means and ends structure, how functional and non-functional values can be integrated. Functional values serve as guiding force for how activities are actually conducted to achieve desired effects that are described through ends. It is thereby only a guiding force, which recommends and promotes certain behaviors. That

is, it does not enforce a certain behavior. Non-functional values provide the basis for functional values. They are used to govern how things are approached. They are something we want to achieve, but which is not within our direct control. For some non-functional values it is possible to measure them, i.e., measure their compliance. Doing so, could be considered a good practice, because it shows the motivation behind such values and also shows how such a value may cascade to other ends. This, in return, makes an overall organization's vision more visible. The resulting integration of functional values and non-functional values can be—respectively—compared to Business Rules and Business Polices of the Business Motivation Model[Om15].

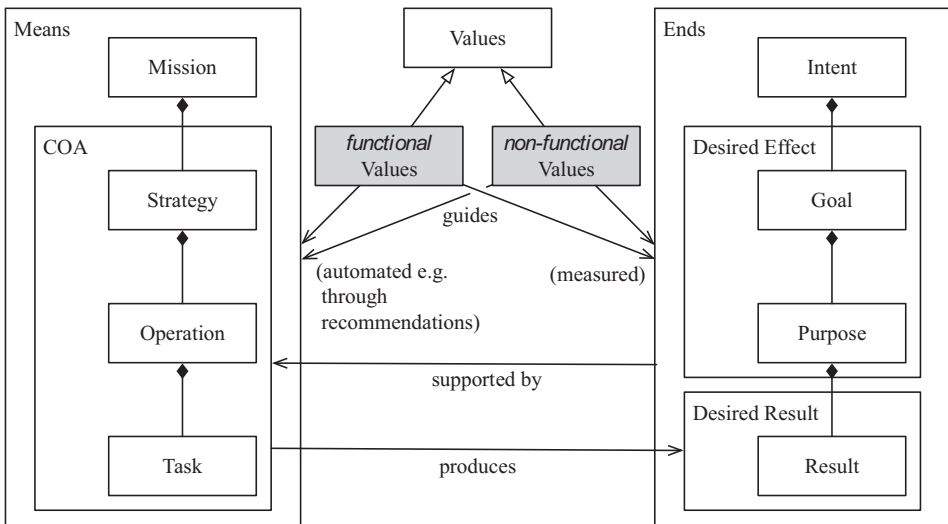


Fig. 1: Integration of Function and Non-function Values on Means and Ends

5 Conclusion

Values in hyper-connected societies are an important topic. Especially the increase in human to machine communication poses new challenges in regard of human values. Values are thereby not as easy to actual integrate in systems as rules are, because they are not to be enforced on everyone. We therefore proposed to distinguish between functional and non-functional values and introduced first possible means of implementing values into software platforms.

By integration functional and non-functional values into our model of means and ends, we are able to improve the foundation of a goal-oriented composition. Because doing so, can in the end, enable us to use the values and vision of an organization to derive an appropriate course of action and help to both complete this course of action, but also to achieve the underlying desired effect, i.e., achieve actual success. Building this on top of

SmartComposition, provides this functionality in an end-user friendly way to enable non-technical personae to apply this approach.

Certainly, more thinking is needed for being able integrate all possible values in software platforms in generic way. Also, the impact on human to machine communication to values has to be examined further. Both these challenges will be subject of our future work.

Acknowledgement: The authors gratefully acknowledge funding by the DFG (GRK 1780/1).

References

- [Co04] Cockton, G.: Value-centred HCI. Presented at the NordiCHI '04: Proceedings of the third Nordic conference on Human-computer interaction, New York, New York, USA October, 2004.
- [Er15] Ericsson AB: Understanding the Networked Society – new logics for an age of empowerment. 1–13, 2015.
- [FHN09] Flanagan, M., Howe, D.C., Nissenbaum, H.: *Embodying Values in Technology: Theory and Practice*. Cambridge University Press, Cambridge, 2009.
- [FKB08] Friedman, B., Kahn, P.H., Borning, A.: *Value Sensitive Design and Information Systems*. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2008.
- [Fr96] Friedman, B.: Value-sensitive design. *interactions*. 3, 1996.
- [Hi91] Hill, T.E.: *Autonomy and Self-Respect*. 1991.
- [Ho07] van den Hoven, J.: ICT and Value Sensitive Design. In: *The Information Society: Innovation, Legitimacy, Ethics and Democracy In honor of Professor Jacques Berleur s.j.* pp. 67–72. Springer US, Boston, MA, 2007.
- [JS06] Jones, M.J., Sanghi, S.: *Driving Excellence*. 2006.
- [KWG14] Krug, M., Wiedemann, F., Gaedke, M.: SmartComposition: A Component-Based Approach for Creating Multi-screen Mashups. *ICWE*. 8541, 236–253, 2014.
- [LB00] Lipinski, T.A., Britz, J.: Rethinking the ownership of information in the 21st century: Ethical implications. *Ethics and Information Technology*. 2, 49–71, 2000.
- [Om15] Object Management Group: *Business Motivation Model*. 1–114, 2015.
- [Ox10] Oxford University Press: *Oxford English Dictionary*. 2010.
- [Ro04] Roessler, B.: *The Value of Privacy*. Linz, 2004.

The Case for Liberouters

Teemu Kärkkäinen¹ Jörg Ott²

Abstract: In this paper we argue that a set of fundamental forces has driven modern communication services to become highly centralized. This centralization has enabled globe spanning communication infrastructure and services to be built, but has also enabled powerful actors to monitor and control communications on an unprecedented scale. As a counterbalancing force we propose a *Liberouter* network, which decentralizes both the communication infrastructure and services.

Keywords: Decentralization, centralization, Liberouter.

1 Introduction

Communications are one of a small number of fundamental functions necessary for us humans to function as a society. It is how business is done, how ideas spread, and how knowledge and information necessary for learning is transmitted. Throughout history, we have invented systems and technologies to provide increasingly effective and efficient means of communication.

Starting with the invention of complex language some 100,000 years ago and the appearance of cave paintings about 40,000 years ago, the invention of writing systems around 4000 BCE, alphabets in 2000 BCE, the postal system in the 6th century BCE and the printing press in 1440, and culminating in modern inventions such as the telegraph and the telephone in the 19th century, radio communications in the early 20th century, and finally the invention of the Internet in the late 20th century.

Over the past few decades the Internet has become *the* communication substrate, quickly taking over other forms of communication: We use Skype instead of phone calls, Netflix instead of TV, news sites instead of newspapers, Spotify instead of the radio, email instead of letters and social media instead of meeting friends at coffee shops.

In this paper we argue that within the development of the Internet infrastructure and services we can see a strong trend towards *centralization* of technology, ownership and control. This in turn has enabled powerful actors to monitor and influence communications between people on a scale never seen before in human history.

We further argue that this trend must be counterbalanced by a complementary, fully decentralized network, and show how our *Liberouter* network achieves this.

¹ Technical University of Munich, kaerkkae@cs.tum.edu

² Technical University of Munich, ott@cs.tum.edu

The contribution of this paper is the argument *why* the Liberouter network is important in a world with ubiquitous Internet connectivity – for technical details of our system we refer the reader to our previous publications [KO14].

2 Centralization of Communications

We present our analysis of the centralization tendencies in Section 2, and describe how our distributed system design provides a decentralized alternative in Section 3.

While both the Internet and the Web are decentralized architectures, in practice both have shown a strong tendency towards centralization. This is a natural result of *technological*, *economic* and *societal* forces acting on the communication *infrastructure* and *services*. As a consequence of the centralization, powerful actors are able to apply pressure and exert control over the people’s communications and the interactions on an unprecedented scale.

From the *technological* perspective, systems are typically scaled up by building hierarchical structures (e.g., trees), which results in the higher levels of the hierarchy acting as centralization points. In terms of infrastructure, last-mile ISPs are centralized points over the consumers, while their transit ISPs and peering points are centralized points over them. In addition, typical modern large scale communication services are technically distributed among many data centers, but are all under the control of a single company. This is in contrast to early Internet services and protocols, such as email (SMTP) and Usenet (NNTP), which were designed to federate independently owned and operated servers.

From the *economic* perspective, the capital requirements of building the physical infrastructure are driving centralization, as only the largest actors can raise enough capital. For example, Vodafone, a European mobile operator, spent over 32 billion euros over three years on infrastructure and spectrum rights [Vo16]. With services, the demand-side economies of scale, or “network effects,” have driven centralization. Each segment is dominated by one, or a few, services; e.g., Google for search and Facebook for social networking. As a result, there is no incentive for interoperating with competitors and companies instead build closed, centralized systems. This extends to native applications through the Internet distribution channels, such as Apple’s App Store, Google’s Play Store, and Valve’s Steam – each holding a near monopoly of control in their market segments.

From the *societal* perspective, the centralized communication systems are an effective tool for enforcing rules and regulations. Especially since much of the discourse has moved to these centralized platforms from previously decentralized mediums, e.g., from pamphleteering to blogging. The important aspect is not the content of the rules – which may be perfectly justifiable – but that the rules are decided centrally and applied to everyone. This is a dangerous power to delegate to a central authority, especially when it comes to the content of the communications, because it enables mass scale imposition of ideologies.

While the centralization has created globe spanning infrastructure and services, it has also been used to apply mass control over the users. In the crudest form of control, governments completely cut off their population from the Internet, such as the outages during the Arab

Spring [De11]. More sophisticated forms of control seek to censor certain topics or information through infrastructure based filtering or direct content removal. Freedom House, a watchdog organization, found that 34% of all Internet users live in countries that have cut off communications, and 61% in countries that censor criticism of those in power [Fr16]. Finally, the large privately owned services have no free speech protections and can act as all-powerful censors either for business reasons or after being compelled by governments; e.g., Germany compelling Facebook, Google and Twitter to police “hate speech” [BB16] on their platforms.

3 Decentralization via Liberouters

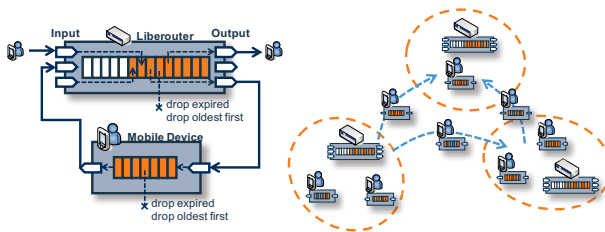


Fig. 1: The Liberouter system.

We argue that to tackle the problems caused by the centralization, a system is needed that decentralizes the communication *infrastructure* and *services* on the *technical*, *economic*, and *societal* axes. The goal is not to replace the current networks, but to augment them with a parallel, decentralized medium for communication, which can serve as a balancing force and a safety valve for the centralization.

Our solution, *Liberouters* [KO14], is a combination of opportunistic networking software with mobile phones and inexpensive embedded devices like the Raspberry Pi, as shown in Fig. 1. Left side of the figure shows the micro-level functionality of the system; a stationary Liberouter device acts as a mailbox to which mobile devices connect to pick up and drop off messages. Although each Liberouter can serve nearby mobile users independently, multiple Liberouters will cooperate to form a macro-level network as shown on the right side of the figure. Services are built as mobile applications that use the message dissemination service provided by the network to communicate directly with each other.

The design decentralizes both the infrastructure and service technologies. First, the infrastructure is made up from a dynamic set of Liberouter devices and mobile phones installed and carried by the users. The messages replicate in the network randomly through multiple paths over short lived contacts between the nodes. Second, the services running on the network are built as directly communicating mobile applications, without centrally controlled backends. Further, the messaging is open by default, i.e., the messages created by one application can by default be read by any other application.

The economic centralizing forces can be overcome by the low cost of the network components. The Liberouter devices cost about 50 euros; cheap enough to be deployed by individuals. The cost will further decline as cheaper embedded Linux platforms (e.g., Raspberry

Pi Zero) come into the market. The low cost and size of the devices can enable novel infrastructure deployment models, e.g., concealing devices in public spaces where they will serve passing mobile phones. And because the services are developed as mobile apps, the service developers do not need expensive backend systems required by web services.

The design also decentralizes the system along the societal axis. An individual, or a group, deploying the Liberouter devices is in control of the rules for using and accessing the deployment. Communities can still set their own standards on what is acceptable behavior or speech, but cannot impose those standards on others against their will; neither can governments nor powerful private entities.

We believe the Liberouter network solves many of the problems created by centralization: There is no central switch to turn off the network, instead each Liberouter device would have to be found and turned off individually. The messages do not travel along centralized links where the traffic could be filtered or monitored and there are no centralized service operators that could be compelled to censor content. To monitor the communications, the attacker needs to be physically in the proximity of the devices, and even then the temporally decoupled publish/subscribe communication model does not reveal any communication relationships. To censor a message, all copies from all the devices and mobile phones would need to be deleted, otherwise the information will continue to live and replicate.

The decentralized system is an alternative set of trade-offs compared to classical Internet architectures, and as such has its own sets of limitations. For example, the asynchronous messaging model does not work for real-time applications or applications requiring low latency end-to-end interactions. And since the messages spread via human carriers, the data dissemination is likely to be geographically constrained rather than global. This is why we propose such a system as a *complement* rather than an *alternative* to the Internet.

4 Conclusion

In this paper we have made our case for a fully distributed Liberouter network, which can serve to balance the currently popular centralized networks and services, and the potential dangers created by them. We believe this balance can lead to a more sustainable and open society, as well as new innovative services and business models.

We are working to provide everything required for anyone to build their own Liberouter devices via the Liberouter website³. This includes system images for various embedded Linux devices with all the software required to act as a Liberouter device. We further provide various applications, including messaging, photo sharing and experience sharing that use the Liberouter network. We are also exploring various deployment opportunities where our system could provide communication services for people in situations where normal Internet access is not available.

³ <http://liberouter.mobi>

References

- [BB16] BBC News: Facebook, Google and Twitter agree German hate speech deal, <http://www.bbc.com/news/world-europe-35105003>, Accessed: 26.5.2016.
- [De11] Dainotti, Alberto; et al.: Analysis of Country-wide Internet Outages Caused by Censorship. In: Proc. of ACM SIGCOMM IMC. 2011.
- [Fr16] Freedom House: Freedom on the Net 2015. 2016.
- [KO14] Kärkkäinen, Teemu; Ott, Jörg: Liberouter: Towards Autonomous Neighborhood Networking. In: Proc. of IEEE WONS. 2014.
- [Vo16] Vodafone Group Plc: Annual Report 2015. 2016.

A Comparison of Value Sensitive Design and Sustainability Design

Stefanie Betz, Andreas Fritsch¹

Abstract: Research and practice in different disciplines try to address technical, social, economical and environmental issues in software systems. This includes not only direct effects of a system and its features but also long-term and cumulative effects that are only shown over time. Nevertheless, in traditional software engineering these effects are usually not taken into consideration. This makes it hard to assess these long-term and cumulative impacts of a system. Therefore, solutions are needed to help software engineers to understand and assess the systemic effects of decisions taken in requirements engineering and systems development. In this paper we discuss and analyze two existing approaches, which provide concepts and methods for this task: Value Sensitive Design (VSD) and Sustainability Design (SD). The analysis shows that the two approaches have conceptual similarities such as the different sustainability dimensions (in SD) versus values (in VSD) and hierarchy of effects (in SD) versus direct and indirect stakeholders (in VSD). Altogether, SD provides a more holistic and structured conceptualization, while VSD – looking back at a longer tradition – can provide a range of experiences when it comes to integrating values in software engineering.

Keywords: Value Sensitive Design, Sustainability Design, Software Engineering, Sustainability, Value

1 Introduction

Software is the backbone of modern society as software systems are increasingly embedded in our society [Ch16]. Therefore, software designers need to be aware of the implication of the socio-technical systems they build. However, in traditional software engineering practice the long-term and cumulative impacts of a system are usually not taken into account and treated in isolation [Be15b]. For example, when designing a groupware system that supports knowledge sharing – balancing of human aspects like privacy and reputation of the involved individuals might well decide whether the software is eventually accepted by its intended users [Mi07]. Or think about the design of a procurement system, where a traditional approach would focus on financial return and supplier compliance and define the requirements accordingly. A holistic approach could integrate further aspects like providing visibility of a product's carbon footprint [Be15a]. One can imagine that this decision has far reaching and accumulative effects on the ecological performance of the respective organization. This might in turn affect consumer perception and thereby the long-term financial performance.

These examples illustrate that the discipline of Software Engineering needs to cope with multidimensional and long-term impacts of systems that have traditionally received little

¹ Karlsruhe Institute of Technology, Institute of Applied Informatics and Formal Description Methods (AIFB), Building 11.40, D-76128 Karlsruhe, {firstname}.{lastname}@kit.edu

attention [Be15b]. Thus, concepts and methods are needed to support understanding of possible effects of software systems and to provide a common ground for research and practice.

Several approaches have been proposed to consider (and balance) technical, social, individual, economical, and environmental concerns in software engineering (e.g. [Fr99], [HL07], [Be14], [Ra15]). Also, a number of conferences and workshops are emerging in this research area, e.g. ICT4S (<http://ict4s.org>), the SEIS Track at ICSE (<http://2016.icse.cs.txstate.edu/seis>) or the RE4Susy Workshop at the RE (<http://web.csulb.edu/~bpenzens/re4susy/>). This paper focuses on two approaches that explicitly provide concepts and methods to support understanding and provide a common ground: one is called Value Sensitive Design [Fr99] and the other Sustainability Design [Be14]. In the following, we are going to shortly introduce them and then discuss their similarities and differences. We believe that either of the approaches has its specific strengths (exemplary case studies versus a structured conceptual framework) and hope that a juxtaposition may inspire synergetic advancements in both communities and related research fields.

2 Value Sensitive Design

The term Value Sensitive Design (VSD) was first coined by Friedman to describe an emphasis of human values in technology design projects [Fr99]. It has since become a “branded term” for specific strategies and techniques concerned with human values [DN15] in technology design. In 2015, Davis and Nathan [DN15] published a summary of the field from which we draw in order to outline VSD.

According to VSD, a technology influences humanity in an “emergent and relational process” [DN15, p. 15]. The influential factors of this process are the design of the technology (e.g. product features), the context of its use and the involved stakeholders. A core assumption herein is that technology products can be improved by identifying and addressing human values in the design. Friedman defines the term *human value* as “what a person or group of people consider important in life” [FKB06, p. 349]. This broad notion is accompanied by a non-exclusive list of specific values named Human Welfare, Ownership and Property, Privacy, Freedom from Bias, Universal Usability, Trust, Autonomy, Informed Consent, Accountability, Courtesy, Identity, Calmness and Environmental Sustainability [FKB06]. Davis and Nathan [DN15] describe four *core commitments* of VSD. These can be seen as characteristics that may distinguish VSD from other design approaches. According to the authors, these characteristics are **proactive stance**, **interactional perspective**, **direct and indirect stakeholders** and **tripartite methodology**. VSD is **proactive** in that it asks researchers to consider human values during the design of a technology – rather than merely criticizing and analyzing existing technologies. The **interactional perspective** is taken by acknowledging that technology and values influence each other bidirectionally: on the one hand, a technology’s design supports (or hinders) certain values and on the other hand, the usage of a technology is dependent on people’s values. Furthermore, VSD distinguishes between **direct stakeholders** and **indirect stakeholders**. Direct stakehold-

ers are those who use a product or technology. Those who do not interact directly with a product or technology, but are influenced by others' use are indirect stakeholders.

The **tripartite methodology** of VSD consists of *conceptual, empirical and technical* “investigations” [FK02]. These investigations are seen as “iterative” and “integrative”. During *conceptual investigations*, the affected stakeholders and values affected by a technology's use are to be identified. *Empirical investigations* apply methods like surveys or questionnaires (among others) to examine the relationship of technologies and their stakeholders with respect to human values. Lastly, *technical investigations* focus on the relationship between specific technological features and values – a concrete approach might be the design of a software that explicitly supports one of the human values listed above.

Several methods have been developed to support VSD [DN15]. Most of them are adaptations of design or social science methods, like for example value-oriented mock-ups or value-oriented semi-structured interviews. Further methods were described by Davis and Nathan [DN15] as values-oriented analyses (*Direct and Indirect Stakeholder Analysis, Value Dams and Flows*) and a values-oriented toolkit (*Envisioning Criteria and Cards*).

In the following, we shortly illustrate some of the described concepts and methods using our first introductory example of a groupware system for the purpose of knowledge sharing (see [Mi07]). The system CodeCOOP was developed together with an industry partner to support software engineering knowledge sharing. During development of the system conceptual and empirical VSD-methods were employed: During *conceptual investigations*, *direct stakeholders* were identified by their roles as (among others) those who submit questions to the systems or those who answer questions. Examples for *indirect stakeholders* are managers or executives of the firm. The *empirical investigation* method *Value Dams and Flows* served to identify several tensions between critical values like privacy, reputation or trust. One outcome was for example that “queriers” fear that their reputation could be harmed when they ask a poor question. This insight led to the implementation of a feature that allows editing of posts to correct errors and improve quality.

3 Sustainability Design

Sustainability Design (SD) in the context of software engineering has been coined by Becker et al. in the Karlskrona Manifesto of Sustainability Design [Be14]. The manifesto is the central paper of this approach providing the main concepts, which are presented in the following. The core definition of *sustainability* that has been adopted for the manifesto is the simple and common “capacity to endure” [Be15b].

The main concepts the manifesto is presenting to approach sustainability are the **five dimensions** and the **three orders of effects**. The **five dimensions** are (1) individual sustainability, which aims at maintaining individual human resources (e.g. health, education), (2) social sustainability, which aims at preserving and improving the societies in their solidarity and services, (3) environmental sustainability, which aims at preserving the natural resources, (4) economical sustainability, which aims at retaining capital and added value, and (5) technical sustainability, which aims at maintaining and evolving information, sys-

tems, and infrastructure. These dimensions can be in conflict with each other [Be15c]. Thus, being technologically sustainable may have a negative effect on the economical sustainability.

The **three orders of effects** are: (1) the direct effects of the software system development and use; (2) the enabling effects that result from the ongoing use of the software system, and (3) the systemic changes caused by the long-time software system usage on a larger scale [Be15c]. The manifesto also follows the idea of not presenting concrete techniques but rather a set of *principles and commitments* for SD [Be14]. These are for example that sustainability is systemic, multidimensional and applies to a system and its wider context. Moreover, sustainability needs to be addressed interdisciplinarily and requires action from several levels, which interact with each other. Finally, sustainability is independent of the purpose of the system, requires long-term thinking, and can be achieved without cutting the needs of the future generations.

The manifesto ends with several suggestions for different stakeholders how to get started to achieve sustainable design. Researchers could identify and discuss research questions, customers and users can put the concern on the table and try to use sustainable products, education may revise the curricula and codes of ethics to include sustainability and finally, software practitioners should raise awareness and try to identify the effects on the different dimensions [Be14].

There exists some papers providing initial methods and techniques (e.g. *system scoping*, *stakeholder participation*, *stakeholder impact analysis*, *goal modelling*) to support sustainability design (see for example [Be15a], [Be15c], [Ch15]). These papers are focusing on requirements engineering as the authors of the manifesto see “requirements as the key to sustainability” [Be15a].

We now go back to our second initial example, a procurement system (see [Be15a]), to illustrate some of the listed methods, dimensions and effects: Discussing the purpose of the project the project team assesses possibilities to support sustainability development of the company emphasizing the effects that the procurement system can have on sustainability in all dimensions (*system scoping*). For example, the system can visualize the carbon footprint of products and facilitate the choice of providers who apply sustainable practice in the environmental dimension. Another action the project team conducts is to extend the number of stakeholders using a *stakeholder impact analysis*. Possible stakeholders involve for example local supplier representatives, service delivery organizations, process analysts, the CTO, and the strategic planning group. Surrogate stakeholder can be introduced to keep the number of stakeholders manageable. The visualization of the carbon footprint of a product within the procurement system qualifies as a *direct effect* along the environmental dimension of sustainability. This in turn enable users of the system to buy products with low carbon footprints (*enabling effect* and *economic dimension*). Taking a long-term perspective, this can lead to the systemic effect of reduced carbon footprints (*systemic effect* and *environmental dimension*) (see [Be15a]).

4 Discussion

First of all, one can identify several similarities between VSD and SD. The focus of both approaches is on "thinking in a broader context", when designing technology. Also, both approaches advocate the idea that technology has an impact on humanity. Thus, they not only take into account immediate effects and technical requirements but also think about effects in multiple dimensions for different stakeholders. Interestingly, both approaches use the idea of commitments for their design approaches rather than providing concrete techniques. But the inherent concept of sustainability design is much broader and more holistic than value sensitive design. Of course, this is already indicated when only looking at the names of the two approaches (value versus sustainability). Thus, SD is emphasizing the long-term effects of systems over time. Additionally, VSD is only emphasizing the "human values" such as human welfare, trust, privacy etc. and environmental sustainability. SD, in turn, has the already mentioned broader focus including not only individual, social, and environmental aspects but also technical and economical ones. This might help to not only integrate e.g. social or environmental aspects into software design, but also to identify and address tensions between these aspects and more "traditional" foci like efficiency and cost.

The sustainability dimensions can be understood as a way of structuring different kinds of values: Penzenstadler and Femmer propose a generic model for sustainability where values represent dimensions of sustainability [PF13]. In this view, the value privacy for example would be an aspect of the individual dimension of sustainability.

In this context it is also worth to mention, that the authors of the VSD approach have not determined and defined the human values for VSD. They only provided a seemingly rather arbitrary list of values. However, both approaches acknowledge systems thinking, the "hierarchy" of effects and the possible conflicts between different dimensions or values. Although, systems thinking is not explicitly mentioned in VSD and the hierarchy of the effects is only mentioned indirectly (and up to the second level) by referring to direct and indirect stakeholders.

Finally, there is one big difference of the two approaches and their use as a common ground for sustainability design. The manifesto is explicitly written to provide a definition and common ground for SD. VSD has evolved more and is based on different publications. We have provided a table on page 272 that provides a summary of the described characteristics of VSD and SD.

	Value Sensitive Design	Sustainability Design
Values/Dimensions	Human Welfare Ownership and Property Privacy Freedom from Bias Universal Usability Trust Autonomy Informed Consent Accountability Courtesy Identity Calmness Environmental Sustainability	Social Sustainability Economical Sustainability Environmental Sustainability Individual Sustainability Technical Sustainability
Impact	direct stakeholders indirect stakeholders	direct effects enabling effects systemic effects
Principles/Commitments	proactive stance interactional perspective direct/indirect stakeholders tripartite methodology	Sustainability is systemic Sust. has multiple dimensions Sust. transcends multiple disciplines Sust. is a concern independent of the purpose of the system Sust. applies to both a system and its wider contexts Sust. requires action on multiple levels System visibility is a necessary precondition and enabler for sust. design Sust. requires long-term thinking It is possible to meet the needs of future generations without sacrificing the prosperity of the current generation

Tab. 1: Value Sensitive Design versus Sustainability Design

5 Conclusion

We have presented VSD and SD as two approaches with somewhat similar objectives. Altogether, the described differences amount to VSD following a bottom up and SD a top down approach. The first started out with design projects and the task to consider human values in the process. And the latter was initialized with a definition of the challenge to be resolved (sustainability in software design) and is now applying these conceptualizations in design processes. Hereby, SD is more structured and holistic (e.g. by defining the five dimensions of sustainability as a structuring means for different values, where VSD has so far considered a seemingly arbitrary list of human values). However, VSD is looking at a longer tradition and can provide a range of experiences when it comes to integrating values in software engineering.

We invite researchers leaning towards either of the presented approaches to share experiences and ideas: both approaches offer concepts and methods to address related issues and both promise to offer valuable insights in order to improve the design of software systems. We think it is important that software engineers are aware of the major role software plays in our society and that they are responsible for long term effects of the system they engineer (see also [Be15a]). Future steps for us as researchers and software engineers are to further implement the concepts we presented in current software engineering practices and apply these in exemplary case studies. So, for example one could imagine to develop templates for requirements documentation and sustainability aware modelling languages.

References

- [Be14] Becker, Christoph; Chitchyan, Ruzanna; Duboc, Leticia; Easterbrook, Steve; Mahaux, Martin; Penzenstadler, Birgit; Rodriguez-Navas, Guillermo; Salinesi, Camille; Seyff, Norbert; Venters, Colin; Betz, Stefanie; , The Karlskrona manifesto for sustainability design, 2014.
- [Be15a] Becker, Christoph; Betz, Stefanie; Chitchyan, Ruzanna; Duboc, Leticia; Easterbrook, Steve M.; Penzenstadler, Birgit; Seyff, Norbert; Venters, Colin C.; Kocak, Sedef Akinli: Requirements: The key to sustainability. *IEEE Software*, (1):1–1, 2015.
- [Be15b] Becker, Christoph; Chitchyan, Ruzanna; Duboc, Leticia; Easterbrook, Steve; Penzenstadler, Birgit; Seyff, Norbert; Venters, Colin C.: Sustainability design and software: The Karlskrona manifesto. In: *Proceedings of the 37th International Conference on Software Engineering-Volume 2*. IEEE Press, pp. 467–476, 2015.
- [Be15c] Betz, Stefanie; Becker, Christoph; Chitchyan, Ruzanna; Duboc, Leticia; Easterbrook, Steve; Penzenstadler, Birgit; Seyff, Norbert; Venters, Colin: Sustainability debt: A metaphor to support sustainability design decisions. In: *Fourth International Workshop on Requirements Engineering for Sustainable Systems (RE4SuSy)*. Ottawa, Canada, August 2015.
- [Ch15] Chitchyan, Ruzanna; Betz, Stefanie; Duboc, Leticia; Penzenstadler, Birgit; Easterbrook, Steve; Ponsard, Christophe; Venters, Colin: *Evidencing Sustainability Design through Examples*. 2015.

- [Ch16] Chitchyan, Ruzanna; Becker, Christoph; Betz, Stefanie; Duboc, Leticia; Penzenstadler, Birgit; Seyff, Norbert; Venters, Colin: Sustainability Design in Requirements Engineering: State of Practice. In: ICSE'16: 38th International Conference on Software Engineering. Austin, Texas, USA, May 2016.
- [DN15] Davis, Janet; Nathan, Lisa P.: Value Sensitive Design: Applications, Adaptations, and Critiques. In (van den Hoven, Jeroen; Vermaas, Pieter E.; van de Poel, Ibo, eds): Handbook of Ethics, Values, and Technological Design, pp. 11–40. Springer Netherlands, Dordrecht, 2015.
- [FK02] Friedman, Batya; Kahn, Peter H. JR: Human values, ethics, and design. In: The human-computer interaction handbook. L. Erlbaum Associates Inc., pp. 1177–1201, 2002.
- [FKB06] Friedman, Batya; Kahn, Peter H. JR; Borning, Alan: Value Sensitive Design and Information Systems. Human-computer interaction in management information systems: foundations, pp. 348–372, 2006.
- [Fr99] Friedman, Batya: Value-sensitive design: A research agenda for information technology. Technical report, 1999.
- [HL07] Hochheiser, Harry; Lazar, Jonathan: HCI and Societal Issues: A Framework for Engagement. International Journal of Human-Computer Interaction, 23(3):339–374, December 2007.
- [Mi07] Miller, Jessica K.; Friedman, Batya; Jancke, Gavin; Gill, Brian: Value tensions in design: the value sensitive design, development, and appropriation of a corporation's groupware system. In: Proceedings of the 2007 international ACM conference on Supporting group work. ACM, pp. 281–290, 2007.
- [PF13] Penzenstadler, Birgit; Femmer, Henning: A generic model for sustainability with process- and product-specific instances. In: Proceedings of the 2013 Workshop on Green In Software Engineering. Fukuoka, Japan, p. 3, March 2013.
- [Ra15] Rashid, A.; Moore, K.; May-Chahal, C.; Chitchyan, R.: Managing Emergent Ethical Concerns for Software Engineering in Society. In: 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering. volume 2, pp. 523–526, May 2015.

Consideration of Values in the Design of Access Control Systems

Till Neudecker,¹ Arsen Hayrapetyan,² Alexander Degitz,¹ Philipp Andelfinger¹

Abstract: Since access control systems codify many of the power structures that govern everyday life, the design of such systems has direct ramifications w.r.t. moral values held by the system's designers, users, or operators. As an alternative to a design process based solely on functional or economical requirements, "value-sensitive design" has been proposed as a structured approach to produce systems that are congruent with given sets of moral values. However, the literature has pointed out the lack of methods for handling tradeoffs between values that may limit the practical utility of the approach. In this position paper, we explore the value-sensitive design of an access control system in a data sharing scenario. To this end, we step through the analysis and evaluation of design alternatives from a purely qualitative consideration to a simple formalization that enables discussion and comparison of designs with respect to tradeoffs between values. While any final design decision depends on value judgments by the stakeholders, we believe that by making value judgments explicit, the formalization can substantiate design discussions and lead to more satisfying designs.

Keywords: value-sensitive design, access control, data sharing

1 Introduction

Many of the interpersonal and institutional hierarchies and power structures present in everyday life can be understood in terms of being granted or denied access to certain resources. On a technical level, the specification and enforcement of access policies is performed using access control systems. Due to this central role, access control systems must be designed carefully so that – in addition to satisfying functional requirements – their operation does not violate moral values.

Value-sensitive design [Fr96] is "a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process" [Fr13]. Value-sensitive design is descriptive as opposed to normative [MH11], i.e., the approach supports the analysis of systems, but does not provide means for resolving conflicts between values or handling tradeoffs when considering multiple system design alternatives [MH11].

In this paper, we explore a systematic approach of designing IT systems that adhere to given sets of moral values. On the example of access control in two variants of a data sharing scenario, we step through a design process from a qualitative level to a consideration of

¹ Karlsruhe Institute of Technology, Steinbuch Centre of Computing and Institute of Telematics, Engesserstraße 2, 76131 Karlsruhe, {till.neudecker, alexander.degitz, philipp.andelfinger}@kit.edu

² Karlsruhe Institute of Technology, Steinbuch Centre of Computing, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, arsen.hayrapetyan@kit.edu

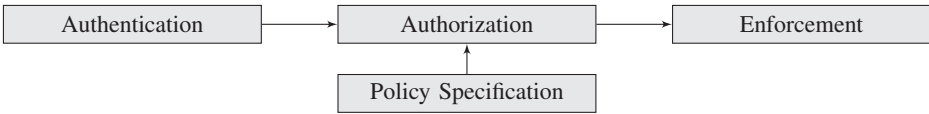


Fig. 1: Access Control Process.

moral values and finally towards a simple formalization of the consideration of values. The aim of the formalization is to make tradeoffs between values explicit and to substantiate discussions of design alternatives among stakeholders.

Our main contributions are as follows: we propose a mapping between the “Objective, Model, Architecture, Mechanism” (OM-AM) framework from the access control literature [Sa00] and the “Values, Norms, Design Requirements” framework from value-sensitive design [VdP13]. The mapping clarifies the role of different “design requirements”. Further, we analyze key implications of design alternatives for access control on the example of a data sharing scenario with respect to trust relationships, norms, and values. Finally, we propose a simple formalization for evaluation of system design alternatives with respect to moral values. The formalization is applied to the data sharing scenario to illustrate how the users’ prioritization of values lead to different access control system designs.

2 Related Work

2.1 Access Control

Access control is a security element which “is concerned with determining the allowed activities of legitimate users, mediating every attempt by a user to access a resource in the system” [VCH06]. An access control system usually consists of an entry point, authentication, authorization and access control policy enforcement services. The entry point provides the user with the means to present their security credentials for authentication. The authentication service is responsible for establishing the identity of the user based on the credentials presented. The authorization service is responsible for verifying if the authenticated user has the necessary rights to access the resource. Finally, the policy enforcement service is responsible for applying the authorization decision, either allowing or denying the access to the protected resource accordingly. The access control process consists of the following major steps (Fig. 1):

- **Policy specification:** The resource owner specifies with whom to share the resource, usually by assigning access rights (e.g., read, write) to identities (e.g., user names, email addresses). However, access rights can also be assigned to groups or based on attributes.
- **Authentication & Attribute assertion:** Before the user can access the resource, the user’s identity must be established using an authentication mechanism (e.g., password based, using certificates, smart-card). Optionally, the identity is then linked to attributes.
- **Authorization:** Based on the formalized policy specification and the identity or attribute information a binary (yes/no) authorization decision has to be made that encodes whether the user should be allowed to access a resource.
- **Enforcement:** Depending on the authorization decision, the access is allowed or not. This is performed by the resource provider as the user attempts an access.

The “Objective, Model, Architecture-Mechanism” (OM-AM) framework was proposed by Sandhu [Sa00] as a tool for engineering authority and trust relationships across organizations and individuals. The framework consists of four layers: the Objective and Model layers specify the security objectives, requirements and tradeoffs, while the Architecture and Mechanism layers describe the means to address those requirements. In [Sa00], OM-AM is applied to articulate the corresponding aspects of the role-based access control models. In [PS01], for example, the framework is applied to the Usage control system.

There are a number of authentication, authorization and access control policy specification frameworks, e.g., Security Assertion Markup Language (SAML) for exchanging authentication and authorization data between security domain; X.509 for the Public Key Infrastructure; OAuth for authorization and secure delegated access, and eXtensible Access Control Markup Language (XACML) for access control policy specification.

2.2 Frameworks for Considering Values in System Design

When considering “values”, we focus on “human values”, i.e., “what a person or group of people consider important in life” [Fr13]. Here, we disregard frameworks focusing on economical consideration such as “business value” or “stakeholder value” [Co05, Bo06].

Socio-technical system design (STSD) is a set of approaches for a “joint optimization of the social and technical systems” [Mu06]. Baxtor et al. [BS11] provided a survey of STSD methods and noted the lack of methods to guide system synthesis. VENUS [GH14] is a research project building on STSD to design socially aware systems in the field of ubiquitous computing: normative propositions are refined to more specific normative criteria a design can be judged against. Now, technological requirements are formulated that adhere to the normative criteria and finally, design proposals are made.

Value-sensitive design (VSD) [Fr96] is “a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process” [Fr13]. Due to the significant research activities in the field of VSD in the past years, we base our analysis of access control systems on VSD concepts. In [Fr13], an iterative tripartite methodology is described that comprises *conceptual* (identification of stakeholders, clarification of terms), *empirical* (assessment of the human context by observation and measurement) and *technical* (analysis and design of systems in light of the stakeholders’ values) investigations. Our work focuses on conceptual and technical investigations by analyzing the trust relationships and resulting considerations of values in the context of an access control system design. Manders-Huits notes that VSD should be complemented with an ethical theory to enable systematic resolution of tradeoffs between values [MH11]. To structure the consideration of values in VSD, Van de Poel [VdP13] proposes a layered hierarchy of *values*, *norms*, and *design requirements*. Values are “what a person or group of people consider important in life”, e.g., “autonomy”. Norms are “prescriptions or restrictions on action” [VdP13], e.g., “only administrators may grant access rights”. Design requirements are technical properties of the envisioned system. Typically, there are many options in determining elements of lower layers of the

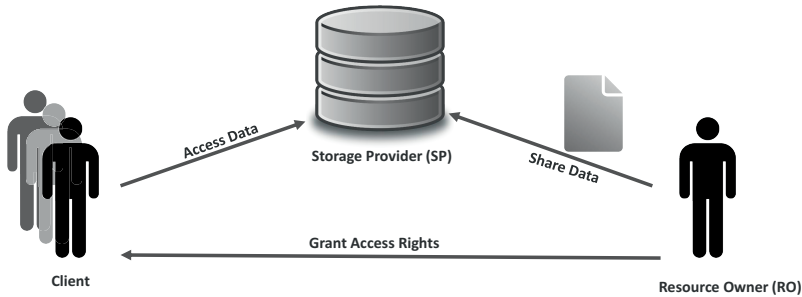


Fig. 2: Sharing Outsourced Data Scenario.

hierarchy that conform with higher layers. However, elements on lower layers can hinder or promote elements on higher layers. In a recent work in VSD, Puylaert has investigated the values of different stakeholders in the design of automated vehicles [Pu16].

We base our analysis on concrete values taken from Schwartz' refinement [Sc12b] of his theory of basic individual values [Sc12a], one result of which is a set of 19 values that can to some degree be considered as interculturally recognized, e.g., “self-direction – action”, i.e., “the freedom to determine one’s own actions” and “security – personal”, i.e., “safety in one’s immediate environment”. Depending on context, it might be appropriate to determine relevant values according to further sources, e.g., human rights [OB16].

3 Sharing Outsourced Data Scenario and Design Process

We will now first describe a **scenario** of sharing outsourced data that we will consider in the following. This description characterizes the core functionality of the data sharing system and also makes assumptions that limit the possible design space: A user (*Resource Owner*, RO) wants to share files with one or more other known users (*Client*, CL). We assume that the RO stores the files at a dedicated *Storage Provider* (SP), such as Dropbox. Although there are numerous ways of sharing files over the Internet (peer-to-peer networks, email, own web servers, etc.), we acknowledge the fact that using a dedicated storage provider has become the predominant method of personal file sharing and, therefore, exclude listed alternatives from further analysis. Regardless of the concrete realization of this scenario, the access control system performs all tasks described in Section 2.1. It is important that the system design supports the users' values, e.g., if the sharing system only allows sharing of data with the public or not at all, undesired accesses by third parties may occur.

In order to obtain the relationships between values, norms and design requirements in a given scenario, we propose a **design process** that is depicted in Fig. 3 consists of three steps. **Step 1** is the exploration of the technically feasible design space based on the scenario description (see above) and the state of the art. This step is exemplarily executed in Section 5 and, contrary to the subsequent steps, can be regarded as being based on techni-

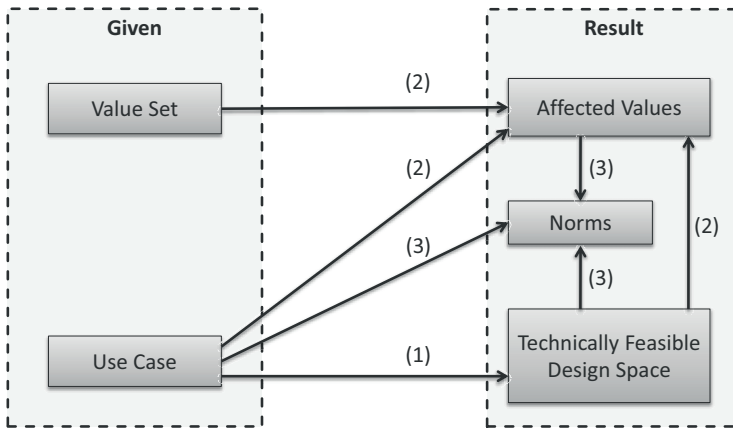


Fig. 3: Proposed Design Process.

cal aspects alone, i.e., independent of subjective judgements. In **step 2**, the set of possibly *affected values* is derived from a given set of values (in our case from [Sc12a]) by an intuition-led process, in which effects of design choices on values are gathered. Above's example of a system design that only allows sharing of data with the public or not at all intuitively affects certain values of a user. Additionally, the affected values depend on the use case, e.g., it makes a difference whether the data to be shared contains weather information or clinical data.

Finally, **step 3** leads to the relevant norms, which link concrete design requirements to abstract values (cf. Fig. 5) by providing an answer to the question *why* a certain value is affected by a specific design requirement and how strong the impact is. This step depends strongly on subjective judgments by the stakeholders and is also dependent on the use case. Although we do not provide a more objective method for this step, we argue that our process makes divergent subjective judgments explicit and hence enables discussion. In the remainder of the paper, we will sketch the described process for our example scenario. In order to proceed with the first step, we need to refine the notion of *design requirements* with respect to access control systems. To this end, we will now propose a mapping that enables us to use fine-grained descriptions of access control systems within the framework of value sensitive design.

4 Value-Sensitive Design and the OM-AM Framework

This section presents a mapping between the values hierarchy from [VdP13] and the OM-AM framework. The former focuses on non-technical notions of values and norms, while the latter underlines the technical aspects of the design requirements. The combination gives a more detailed and technical picture of the system design based on the values. The OM-AM layers articulate the following aspects of an access control system:

- **Objective:** The security objective of the system expressed as a set of policies to be achieved. An example of a security objective is "no unauthorized access to the resources" (e.g., access is denied to any entity accessing the files owned by the RO, for which she has not granted the access rights) in the scenario presented in Section 3.
- **Model:** The mathematical formalization of the security objectives. For the above scenario, the ABAC model is an example of a model.
- **Architecture:** The description of the functionality of the system's logical components and their interdependencies. The architecture typically includes components responsible for authentication, authorization and policy enforcement. For the above scenario, the architecture would include authentication and authorization servers responsible for the authentication of the RO and CL, and making authorization decisions for file access requests, as well as a policy enforcement component at the storage provider.
- **Mechanism:** The protocols and software implementations of the functionality of architecture components. For the above scenario, the mechanism could be the OAuth protocol to exchange authentication and authorization information.

We propose the following mapping between OM-AM and the VSD hierarchy (cf. Fig 4):

- **Values:** no equivalent in OM-AM. Since OM-AM is a technical framework whose goal is to articulate the aspects of an access control system with a specific security objective (or set of security objectives), OM-AM has no layer corresponding to the high-level abstract Value notion in the values hierarchy.
- **Norms:** correspond to *Objectives* in OM-AM. Both Norms and Objectives specify the domain-specific goals to be achieved. In the scenario presented in Section 3, the general norm is "no unauthorized access is allowed"
- **Design Requirements:** subsumes Model, Architecture and Mechanism:
 - *Model:* The Model layer specifies formally the *goals* to be strived for to achieve the objective. For the scenario from Section 3, the model is ABAC, and it specifies what it means to allow "no unauthorized access": to deny access to a user who does not possess the attributes required for an access.
 - *Architecture:* This layer specifies the scope of the norm, defining what is meant by the statement "the authentication and authorization are enforced in a trustful way." For the data sharing scenario, it describes the components enforcing authentication and authorization as well as the trust relationships between those components.
 - *Mechanism:* The Mechanism layer specifies the actions to achieve the aims formalized by the Model. For the data sharing scenario, the OAuth framework defines the enforcement protocols and the format of authentication and authorization data.

Value Sensitive Design	OM-AM
Values	
Norms	Objectives
Design Requirements	Model
	Architecture
	Mechanism

Fig. 4: The proposed mapping between the VSD hierarchy and the OM-AM framework.

We will see in Section 5 that an analysis of design requirements in terms of model, architecture and mechanism enables a more fine-grained analysis of design decisions.

5 Qualitative Analysis of Design Alternatives

The data outsourcing scenario presented in Section 3 describes the core functionality of the system to be designed. As the access control process (cf. Section 2.1) neither specifies which tasks are performed by whom, nor how interactions between the tasks should be implemented, we will now explore these degrees of freedom in order to establish the set of possible *designs*. Each design has several *characteristics*, such as the used protocol or the system architecture. It must be possible to evaluate each characteristic against *design requirements*, i.e., answer the question whether a given characteristic satisfies a design requirement or not, or whether it is independent of the design requirement. As shown in Section 4, design requirements can be on the abstraction level of models, architectures or mechanisms. Our exploration follows the OM-AM model by walking through the layers and identifying characteristics for each layer.

The access control **model** defines how access policies can be specified. Hence, it also affects the authorization, as the formalization of the policy must be evaluable by the authorization point. It can also impose constraints on the authentication, e.g., the authentication must provide the clients' attributes in order to use an attribute based access control model. Finally, the protocols within the access control process must provide the required expressiveness in order to transfer the policy specification or attributes. Characteristics that affect the model layer include, therefore, the expressiveness of the used protocol and the capabilities of the authentication and authorization point.

The **architecture** of the AC system is the most influential layer, as the architecture specifies each entity's tasks. It is clear that the policy enforcement must be performed by the storage provider and that the policy specification must be performed by the resource owner.⁴ Authentication and Authorization, however, can be done by either one, or by another party, such as an *identity provider* (IdP), which has a large effect on the overall system.

Table 1 gives an overview of the characteristics of different architecture designs. Whenever a task is outsourced by the RO to a third party, the RO trusts the third party in some regard.

⁴ The resource owner could encrypt the files to be shared and, therefore, cryptographically enforce access control. However, as this would require additional key distribution, we focus only on logical access control here.

Tab. 1: Characteristics of architecture level design alternatives. ✓: fulfilled, X: non-fulfilled

		Authentication			Authorization		
		by IdP	by SP	by RO	by IdP	by SP	by RO
Trust	File access possible by...	IdP, SP	IdP, SP	SP	IdP, SP	IdP, SP	SP
	Access monitored by...	IdP, SP	IdP, SP	SP	IdP, SP	IdP, SP	SP
	Policies known to...	N/A	N/A	N/A	IdP(, SP)	SP	✓
Functionality	Free choice of AC model	N/A	N/A	N/A	X	X	✓
	No IDM Overhead	✓	✓/X	✓/X	N/A	N/A	N/A
Non-functional	Availability	✓	✓	X	✓	✓	X
	Cost	✓	✓	X	✓	✓	X
	Performance	✓	✓	X	✓	✓	X

For example, if authorization is done by a third-party IdP, this IdP could impersonate clients and access outsourced data. Third parties can also learn something about the Client or RO, e.g., by monitoring the access or analyzing the policies. In all these cases, the RO trusts the third party, that it does not perform these malicious actions. Depending on the architecture design, the required trust in third parties varies.

In terms of functionality, the architecture layer overlaps with the model layer, because the choice of the access control model is limited to the access control model offered by the authorization point. Hence, a RO requiring full control of the access control model should not outsource the authorization step. Another functional characteristic is the possibility to rely on an already established identity management system. In most cases of personal data sharing, outsourcing the authentication to a provider with a large user base (e.g., Google or Facebook) decreases the overhead for identity management (IDM). In cases of academic or commercial data sharing, all clients might already have an account at the RO's organization. Non-functional characteristics include performance, cost and availability, which can be satisfied to a higher degree by specialized providers. The **mechanisms** used in an AC system also primarily impact non-functional characteristics. For instance, authentication mechanisms could apply symmetric or asymmetric cryptography, resulting in different processing delays due to different complexity of calculations.

The performed qualitative assessment of the core aspects of the system design is the first step (cf. Sec. 3) of our design process. It also demonstrates the challenge of a comprehensive assessment. Hence, we described the design in terms of the “Values, Norms, Design Requirements” hierarchy (cf. Fig. 5). In the figure, we focus on the design decision whether the Authentication and Authorization (A&A) process is outsourced to a third party. We identify three norms that are clearly affected by the design decision: First, **minimizing the monetary cost** for the resource owner: we assume that it is less expensive to outsource A&A than it is to host the necessary infrastructure oneself. Second, **specification of arbitrary access control policies**: in case A&A is outsourced, the resource owner may need to confine herself to the types of access control policies allowed by a third party. Third, **monitoring of enforcement**: in case the RO hosts the A&A infrastructure herself, she is able to monitor all steps of the A&A process, which may be infeasible in case of outsourcing.

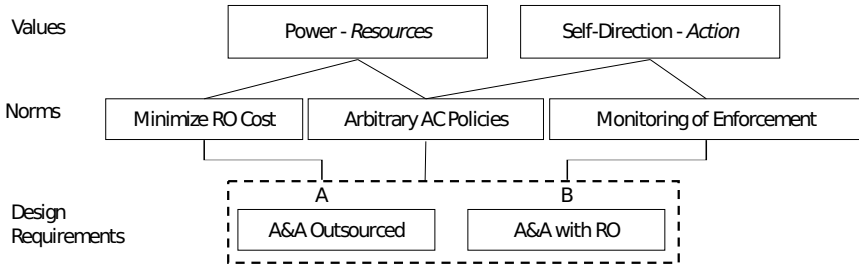


Fig. 5: Values, norms and design requirements in the private data sharing scenario.

We consider the following values (from [Sc12b]): First, **Power - Resources**, i.e., “Power through control of material and social resources”. Here, Power - Resources is affected by the monetary cost of a design. Second, **Self-Direction - Action**, i.e., “Freedom to determine one’s own actions”. Here, Self-Direction - Action is affected by the possibility to constrain or react to the behavior of users accessing the shared data. The consideration of values and norms gives an overview of the effects of different design decisions, but does not allow for design decisions in case multiple different designs are possible. To support design decisions, in the following, we move towards a quantitative assessment.

6 An Attempt at Formalization

Van de Poel studied conditions for aggregating value scores to determine an overall most desirable design [VdP15]. Such a procedure depends on the ability to compare value scores across design alternatives on a common scale (“value commensurability”). Further, Van de Poel analyzed concrete procedures to evaluate designs: in *cost-benefit analysis*, considerations of values are expressed in terms of monetary units. Van de Poel notes a number of issues with cost-benefit analysis, e.g. the non-linearity in the utility of money, which is shared by the approach of *direct trade-offs*, where designs are scored on an individual scale for each value. To compare alternatives, there must be a way to compare scores across values. Like cost-benefit analysis, direct trade-offs assumes that “a loss in one value can always be compensated by a gain in another value”. Counter-examples to this have been called *taboo trade-offs* [Te03]. In the *maximin* approach, the design with the highest score is selected according to each design’s lowest-scoring value. Using *satisficing*, a design is evaluated only w.r.t. whether a certain minimum score with respect to each relevant value is exceeded, which does not suggest a “best” design. Van de Poel proposes an overall design process in which satisficing is performed first, and which, if morally acceptable designs can be identified, proceeds to selecting the “best” design using cost-benefit analysis, direct tradeoffs, or maximin. In the following, we assume that satisficing has been applied and morally unacceptable designs discarded. The comparison is based on direct trade-offs.

We formalize the analysis of a set of design alternatives as follows: a *design analysis* is characterized by a tuple (R, N, V, I_R, I_N, W_V) , where R is a set of design requirements, N is a set of norms, and V is a set of values. $I_R : R \times N \rightarrow \mathbb{R}$ reflects the impact of design requirements on norms, if fulfilled. $I_N : N \times V \rightarrow \mathbb{R}$ reflects the impact of norms on values,

if fulfilled. For a design requirement r and a norm n , $I_R(r, n) > 0$ iff r adheres to or supports the norm n . $I_R(r, n) < 0$ iff r hinders or violates n . $I_R(r, n) = 0$, if r has no bearing on n . $I_N(n, v)$ is defined analogously for each pair of a norm n and a value v . $W_V : V \rightarrow \mathbb{R}_+$ reflects the importance of each value in the context of the considered system.

To evaluate a (partial) design w.r.t. the considered values, we first select a conflict-free set of design requirements $D \subseteq R$. To determine the value score of a design, it may be sufficient to consider those design requirements with large impact on the value score. We assume that in a real-world setting, stakeholders will iteratively extend the *design analysis* with values, norms and design requirements they consider to be relevant. Now we can evaluate the design's adherence to the considered values by calculating the design's score:

$$S(D) := \sum_{d \in D, n \in N, v \in V} I_R(d; n) I_N(n; v) W_V(v).$$

By comparing the scores of all feasible designs, the design that best fulfills the considered values can be determined. Meaningful comparisons require that the weights on each level of the “Values, Norms, Design Requirements” hierarchy are chosen so that they represent the relative impacts consistently. Certainly, some situations may require a more expressive language to comfortably express intricacies of the consideration of values:

- Adherence to one design requirement r_1 may be sufficient to fully adhere to a norm n , so that additional adherence to a design requirement r_2 has no further effect. A solution would be to specify maximum scores that each norm can attain.
- Values may be more or less important w.r.t. different aspects of a design. In such cases, values should be specified further to differentiate the relevant aspects.
- Our formalization scores values and norms linearly, e.g., diminishing returns w.r.t. a certain value or norm cannot be specified. As solution would be to allow for arbitrary functions that determine the score w.r.t. a given value based on its linear score.

Although the proposed extensions may result in an analysis process that more accurately reflects the underlying considerations of values, the goal of making explicit the reasoning that guides design decisions still suggests settling for the simplest suitable formalization.

7 Making Design Decisions

Now, using our formalization that enables scoring of different designs according to the stakeholders' value judgments, we determine value scores with respect to three assumed users. In addition to the values *Power - Resources* and *Self-Direction - Action* introduced in Section 5, we consider the value *Face*, i.e., “Security and power through maintaining one's public image and avoiding humiliation” [Sc12b]. In our scenario, we consider *Face* to be affected by the possibility to monitor who accesses the shared pieces of data, and by the availability of the resources to the users. Figure 6 shows our consideration of values,

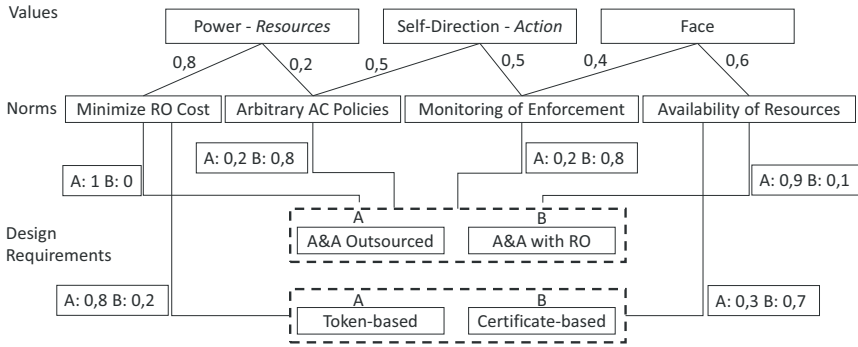


Fig. 6: Scoring the congruence of a design with values considered in the private data sharing scenario.

norms and design requirements. It represents an example for a value-sensitive discussion about design requirements. The considered norms and impacts can be derived in step 3 of the design process (cf. Section 3) and were exemplarily chosen here. Since the value "Power-Resources" has more influence on the norm "Minimize RO Cost" than on the norm "Arbitrary AC Policies", the edge weights differ significantly (0.8 vs. 0.2). The design requirements in question are the following:

A. Whether to outsource the Authentication and Authorization (A&A) process to a third party: a discussion of the implications of this aspect was given in Section 5.

B. Whether to use a token-based or a certificate-based A&A mechanism: in a token-based A&A approach like OAuth 2.0, the IdP and the authorization server do not manage client connections, but only issue tokens, which the clients send to the next server according to the used protocol. Token-based mechanisms tend to be less resource-intensive, since the need to store connection information is avoided. For this reason, the edge weight between "Minimize RO Cost" and "Token-based" (A) is 0.8, compared to the weight for "Certificate-based" (B) 0.2. In contrast, Certificate-based A&A approaches not only need to maintain a public key infrastructure, but usually also need to manage client connections, which are redirected from the IdP to the corresponding authorization server. However, certificate-based approaches tend to be less error-prone, since connections are managed by the server at all times. Therefore the edge weight from "Availability of Resources" to "Certificate-based" (B) is 0.7, whereas the weight for "Token-based" is only 0.3.

Figure 5 depicts a quantitative analysis of the impacts of the discussed design requirements according to the proposed formalization. Edges are assigned weights representing how well a value is represented by a norm or how well a norm is represented by a design requirement, enabling quantitative discussions about the degree to which a design requirement fulfills a set of values. As an example, we consider three assumed users:

User 1 is only interested in the monetary cost of sharing her data and thus only considers the value *Power-resources* (importance 1.0). The score in Figure 6 suggests that user 1 should outsource A&A to a third party provider and that she should use token-based A&A.

Tab. 2: Three users with differing scores with respect to their values for the data sharing scenario.

	Score (1)	Score (2)	Design Decision
User 1: Money (1;0;0)	A: 0.84 B: 0.16	A: 0.64 B: 0.16	A - A
User 2: Availability (0.1;0.1;0.8)	A: 0.6 B: 0.4	A: 0.21 B: 0.35	A - B
User 3: Self-direction (0.2;0.7;0.1)	A: 0.37 B: 0.63	A: 0.15 B: 0.07	B - A

User 2 is mostly interested in availability, so he sets the highest importance to the value *Face* (importance 0.8). He expects to “lose face” and possibly business partners if his shared data is not available. User 2 also considers the cost and his ability to set up and monitor customized rules (both importance 0.1). With these weights set, User 2 should also outsource A&A, but use a certificate-based A&A approach, which are usually more reliable.

User 3 is highly privacy-aware. She wants to customize and monitor the enforcement of access rights herself through the value of *Self-direction* (importance 0.7). She is also interested in a cost-efficient solution (importance 0.2) while still considering the availability of her data to a small amount (importance 0.1). With these weights set, User 3 should not choose to outsource the A&A process. Like User 1, she should choose a Token-based approach, because she valued the cost of the solution more than the availability of her data.

Table 2 summarizes the value scores and selected designs for the three assumed users.

8 Summary and Discussion

In this paper, we considered the value-sensitive design of access control systems. As a first step, we performed a purely qualitative and informal analysis of an assumed access control scenario. In the analysis, extending the “Values, Norms, Design Requirements” hierarchy from value-sensitive design with the Object-Model-Architecture-Mechanism framework from the access control literature enabled a more fine-grained consideration of design requirements. Based on existing proposed design procedures, we introduced a simple formalization that enables a numerical scoring of the congruence of system designs with the stakeholders’ values. The formalization considers the effects of design decisions on the considered values, making the value judgments in the design process more explicit. We calculated scores reflecting the congruence with the values held by three assumed types of users of the envisioned access control system with distinct prioritizations of values. The per-user scoring demonstrated that the design of even a relatively simple access control system can be affected strongly by value judgments of the stakeholders.

The presented analysis highlights some of the key challenges on the path towards a value-sensitive analysis and design of real-world systems:

- While technically feasible system designs can be derived from the requirements defined by a given use case based on technical properties of the envisioned system, determining the largely subjective impacts of designs on norms and values poses larger difficulties. Similarly to approaches for determining security metrics w.r.t. IT environments, quantification and comparison is a challenging issue.

- In the value-sensitive design process sketched in Section 3, a major challenge lies in determining norms on a suitable level of abstraction, based on previously identified design requirements and values. An abstraction and generalization of the design requirements, e.g., formally using domain-specific languages, and a specification of the relevant values in the given context, e.g., in the form of concrete laws or policies, could aid in bridging the remaining gap between technical and ethical considerations. A similar problem is given in policy-based management, where generic policies must be translated to concrete actions.
- In real-world IT environments, the consideration of values may not necessarily be a priority in the design process. Instead, administrators and users are frequently confronted with existing systems that reflect the implicit value judgments made at the time the system was designed. Further, in a previous design process, the compatibility with existing IT environments may be the predominant decision factor. In such cases, we argue that a systematic evaluation of the existing system in light of the values held by the stakeholders' may enable a clearer analysis and critique of the system and could suggest steps to align the system more closely with the stakeholders' values.
- A possible focus of future work could be the question of scoring designs with respect to *privacy*-related values. For certain types of values, a prioritization can be performed in terms of well-defined numerical weights, e.g., using probabilities or established currencies. For other values, the selection of suitable weights seems less clear. Particularly in the case of access control systems, values related to privacy can come into focus. It may not be possible to enumerate or evaluate the potential consequences of unintended data flows, although in many scenarios such data flows are clearly undesirable. This is related to the general problem of justification of privacy concerns [So07]. Possibly, numerical weights associated with privacy-related values may therefore only be understood as expressing subjective prioritizations in relation to other values.
- We think that the benefits of more explicit consideration of values in system design can only be proven in a real-world setting. User studies such as [Pu16] might give pointers to improve the process of determining value-congruent system designs systematically.
- Finally, some of the steps in resolving conflicts and tradeoffs between values in the system design process seem to have natural counterparts in mathematical logic and optimization: *cost-benefit analysis* and *direct trade-offs* can be considered as mathematical optimization problems. Further, it seems natural to formalize problems of *satisficing* as satisfiability problems from mathematical logic. Hence, similarly to approaches in IT security [Kö15], given a formal description of the impacts of different design requirements on the relevant values for a desired system functionality, it may be possible to generate or parametrize designs in an automated fashion to accommodate varying scenarios or users.

References

- [Bo06] Boehm, Barry W: Value-Based Software Engineering: Overview and Agenda. In: Value-based software engineering, pp. 3–14. Springer, 2006.
- [BS11] Baxter, Gordon; Sommerville, Ian: Socio-Technical Systems: From Design Methods to Systems Engineering. *Interacting with computers*, 23(1):4–17, 2011.

- [Co05] Cockton, Gilbert: A Development Framework for Value-Centred Design. In: CHI'05 extended abstracts on Human factors in computing systems. ACM, pp. 1292–1295, 2005.
- [Fr96] Friedman, Batya: Value-Sensitive Design. *Interactions*, 3(6):16–23, December 1996.
- [Fr13] Friedman, Batya; Kahn, Peter H.; Borning, Alan; Hultdtgren, Alina: Early Engagement and New Technologies: Opening up the Laboratory. In (Doorn, Neelke; Schuurbiens, Daan; van de Poel, Ibo; Gorman, E. Michael, eds): *The Handbook of Information and Computer Ethics*, chapter Value Sensitive Design and Information Systems, pp. 55–95. Springer Netherlands, Dordrech, 2013.
- [GH14] Geihs, Kurt; Hoffmann, Holger: Socio-Technical Design of Ubiquitous Computing Systems. chapter A Research Agenda for the Socio-Technical Design of Ubiquitous Computing Systems, pp. 3–18. Springer International Publishing, Cham, 2014.
- [Kö15] Köhler, Jens: Tunable Security for Deployable Data Outsourcing. KIT Scientific Publishing, 2015. Dissertation, Department of Computer Science.
- [MH11] Manders-Huits, Noëmi: What Values in Design? The Challenge of Incorporating Moral Values into Design. *Science and engineering ethics*, 17(2):271–287, 2011.
- [Mu06] Mumford, Enid: The Story of Socio-Technical Design: Reflections on its Successes, Failures and Potential. *Information Systems Journal*, 16(4):317–342, 2006.
- [OB16] Orwat, Carsten; Bless, Roland: Values and Networks – Steps Toward Exploring the Relationships. *Computer Communication Review*, 2016. Editorial Note.
- [PS01] Park, Jaehong; Sandhu, Ravi: Towards an Engineering Framework for Usage Control and Digital Rights Management. 2001.
- [Pu16] Puylaert, S.A.A.: Social Desirability and Mobility Impacts of Early Forms of Automated Vehicles. Master's thesis, Delft University of Technology, 2016.
- [Sa00] Sandhu, Ravi: Engineering Authority and Trust in Cyberspace: The OM-AM and RBAC Way. In: *Proceedings of the fifth ACM workshop on Role-based access control*. ACM, pp. 111–119, 2000.
- [Sc12a] Schwartz, Shalom H: An Overview of the Schwartz Theory of Basic Values. *Online Readings in Psychology and Culture*, 2(1):11, 2012.
- [Sc12b] Schwartz, Shalom H; Cieciuch, Jan; Vecchione, Michele; Davidov, Eldad; Fischer, Ronald; Beierlein, Constanze; Ramos, Alice; Verkasalo, Markku; Lönnqvist, Jan-Erik; Demirutku, Kursad et al.: Refining the Theory of Basic Individual Values. *Journal of personality and social psychology*, 103(4):663, 2012.
- [So07] Solove, Daniel J: 'I've Got Nothing to Hide' and Other Misunderstandings of Privacy. *San Diego law review*, 44:745, 2007.
- [Te03] Tetlock, Philip E: Thinking the Unthinkable: Sacred Values and Taboo Cognitions. *Trends in cognitive sciences*, 7(7):320–324, 2003.
- [VCH06] Vincent C. Hu, David F. Ferraiolo, D. Rick Kuhn: *Assessment of Access Control Systems*. 2006.
- [VdP13] Van de Poel, Ibo: Translating Values into Design Requirements. In: *Philosophy and engineering: Reflections on practice, principles and process*, pp. 253–266. Springer, 2013.
- [VdP15] Van de Poel, Ibo: Conflicting Values in Design for Values. *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*, pp. 89–116, 2015.

Towards designing individual value-oriented decision support for selecting internet-based services

Verena Dorner¹, Stefan Morana¹, Alexander Maedche¹, Christof Weinhardt¹

Abstract: Including individuals' value systems is seldom considered in the current body of decision support research, although individuals' values have a strong impact on decision-making. We propose that adapting decision support to include individuals' values increases decision quality and decreases decision effort. We propose to extend decision support for choosing internet-based services such that services are matched not only regarding individuals' preferences for search and experience attributes, but that services also match their value systems. By helping users identify these matching services, we aim to enable them to participate in greater parts of digital life in accordance with their individual values. In this research-in-progress paper, we introduce the conceptual foundations and describe a preliminary research design. The results will serve as a guideline for designing value-based decision support in the next step of our research project.

Keywords: individual values, design, decision support, internet-based Services

1 Introduction

Since the number of internet-based services keeps multiplying, and business models often rely on the (re)sale and analysis of personalized data, the question of how users can keep control over their digital life in terms of its coherence with their individual value system [Sc92] becomes more and more important. Specifically, choosing from an ever-larger range of internet-based services increases decision time as well as effort, results in different cognitive-affective responses from different individuals [RG14], and increases the likelihood of information overload which in turn has adverse effects on decision quality. Decision support systems (DSS) support individuals in their decision-making. DSS provide decisional advice [TA01] to enable faster, better, and easier decision making [LD00]. They also serve to decrease cognitive effort [SC12] and avoid information overload [LS12] during decision-making. Most research on the behavior of users in DSS and the design of DSS focuses on decision tasks and decision objects described in terms of search and experience attributes [XB07]. One dimension of decision-making that is usually not addressed is how decision tasks and objects relate to the user's individual value system. Anecdotal real-life evidence indicates, however, that individual value systems play a large role in shaping decisions. This is particularly interesting in the domain of internet-based services, with successful services such as instant messaging services WhatsApp and Snapchat as well as online social networks like Facebook. For instance,

¹ Karlsruhe Institute of Technology, Institute of Information Systems and Marketing, Fritz-Erler-Straße 23, 76131 Karlsruhe, {forename.surname}@kit.edu

individuals with a high demand of self-direction [Sc92], may require a high degree of control over their personal data and refuse to use such services as gather huge amount of personal data. If such alternatives are not readily available, participation in certain parts of digital life may be restricted. This reduces the benefits of such services both for the individual and – especially in the instance of services that rely on network effects – for other service users. Assuming that preferences for values can be expressed and measured similarly to (performance) attributes of decision objects, the DSS needs to be extended to add individual values to the list of relevant attributes describing a service. In this research-in-progress paper, we present key conceptual foundations and a preliminary research design addressing the research question. In our overall research project, we aim to understand, **first**, whether and how explicitly including individual values as decision criteria results in different cognitive-affective responses and possibly decreases effort (despite increasing the number of attributes) due to increased (potential) coherence of service and value system in cases where services matching the individual’s value system exist. **Second**, we intent to develop design guidelines for integrating individual value attributes into DSS for internet-based service selection.

2 Conceptual Foundations

We base our understanding and measurement of individual values on the popular Theory of Universals in Values [Sc92] which takes values to be “*conceptions of the desirable that influence the ways people select action and evaluate event*” [SB87]. Universal values arise from basic underlying goals (e.g. fulfilment of biological needs, welfare and preservation of groups) and can be classified in ten dimensions placed on a circle: Power - Achievement - Hedonism - Stimulation - Self-Direction - Universalism - Benevolence - Tradition - Conformity - Security - Power. The positions on the circle indicate whether values are compatible (close proximity) or conflicting (opposing) in terms of how they “*guide [...] perceptions, preferences, and behaviors*” [BD16]. While these values are universal across individuals and cultures, individuals differs with regard to the relative importance they accords each value [BB15]. In addition, the decision situation must be accounted for: different situations lead to large and systematic differences in individual decision-making. We base our research on recent research that identifies eight situational cues that affect individual behaviour most strongly [RG14]. Research in DSS has developed a number of methods for measuring preferences for decision objects and attribute weights [SD15]. However, these methods focus on experience and search attribute, not values.

3 Preliminary Research Design

For the overall research project, we follow the design science research (DSR) approach [HM04] in order to derive guidelines, in the form of design knowledge, for the design of value-based decision support for selecting internet-based services. In a first step, we address the first part of our research project in a laboratory experiment and intend to

conduct a field study in a later stage of the research project. For the laboratory experiment, we develop two DSS for service selection (treatments) which participants use in a between-subject repeated measures experimental design. One system lets the participants enter self-reported weights for values based on a classical approach (such as TRADEOFF [KM91]), the other based on the principles of a product configurator [SD15]. Participants are then asked to rate a set of 10 services on a scale from 1 (worst) to 10 (best). The services are shown in an ordered list based on participant input.

The participants use the DSS multiple times for choosing different services shown in randomized order within treatments. We choose different services because specific situational characteristics elicit different cognitive-affective responses from different individuals [RG14]. We use within-subject design because we hope to be able to identify cross-situationally stable effects of individual values on choice behavior. One such situation is the choice of a fitness app, which includes strong situational cues for positivity (potentially enjoyable experience), negativity (potentially frustrating, e.g. fitness goals are missed) and duty (e.g., adherence to training plan). Another is the choice of a dating app, which includes strong cues for mating (potential romantic partners are present), sociality (potentially develop close personal relationships), adversity (potentially being criticized or rejected), and deception (potentially untruthful statements, pictures etc.). A third situation is the choice of a brain-training app that includes strong cues for intellect (cognitive effort is required). After each situation, participants are asked to fill in the DIAMONDS standardized questionnaire [RG14] and the NASA TLX [HA06] for measuring cognitive effort. In both treatments, participants are asked to fill in the Universal Values standardized questionnaire [Sc92] prior to the experiment and indicate their satisfaction [BH01] with the DSS after the experiment.

4 Expected Contribution

With this research, we intend to shed light on how explicitly including individual values as decision criteria reduces effort, increases decision quality, and satisfaction. We compare two attribute weight elicitation methods in order to determine which requires less user effort. Following the DSR approach, we aim to derive valuable design knowledge for integrating value attributes into DSS for service selection as next step. In addition, we hope to be able to identify types of users that exhibit similarities in their behavior and link these similarities back to idiosyncrasies in their values.

In our future work, we plan to investigate ways of increasing the usability of the DSS by collecting information on individual value systems automatically. Manual input is time-consuming and tends to reduce perceived usability and usefulness of information systems. One possible solution to this issue is applying machine learning techniques to users' search and choice histories [RF10]. We plan to evaluate the DSS, including such solutions, in a field study, where we can supplement users' observed behavior and inputs with contextual data such as current location or travel status. A field study setting will also afford us the

opportunity to examine to which degree prices and users' willingness-to-pay [ZS13] as well as perceived quality [VS15] affect the selection of internet-based services.

5 Bibliography

- [BB15] Borg, I.; Bardi, A.; Schwartz, S. H.: Does the Value Circle Exist Within Persons or Only Across Persons? In *Journal of personality*, 2015.
- [BD16] Borg, I.; Dobewall, H.; Aavik, T.: Personal values and their structure under universal and lexical approaches. In *Personality and Individual Differences*, 2016, 96; pp. 70–77.
- [BH01] Bhattacharjee, A.: Understanding Information Systems Continuance: An Expectation-Confirmation Model. In *MIS Quarterly*, 2001, 25; pp. 351–370.
- [HA06] Hart, S. G.: Nasa-Task Load Index (NASA-TLX); 20 Years Later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2006, 50; pp. 904–908.
- [HM04] Hevner, A. R. et al.: Design Science in Information Systems Research. In *MIS Quarterly*, 2004, 28; pp. 75–105.
- [KM91] Kahn, B. E.; Meyer, R. J.: Consumer multiattribute judgments under attribute-weight uncertainty. In *Journal of Consumer Research*, 1991; pp. 508–522.
- [LD00] Limayem, M.; DeSanctis, G.: Providing Decisional Guidance for Multicriteria Decision Making in Groups. In *Information Systems Research*, 2000, 11; pp. 386–401.
- [LS12] Lankton, N. K.; Speier, C.; Wilson, E. V.: Internet-Based Knowledge Acquisition: Task Complexity and Performance. In *Decision Support Systems*, 2012, 53; pp. 55–65.
- [RF10] Rendle, S.; Freudenthaler, C.; Schmidt-Thieme, L.: Factorizing personalized Markov chains for next-basket recommendation. In (Rappa, M. et al. Eds.): *the 19th international conference*, 2010; p. 811.
- [RG14] Rauthmann, J. F. et al.: The Situational Eight DIAMONDS: a taxonomy of major dimensions of situation characteristics. In *Journal of Personality and Social Psychology*, 2014, 107; pp. 677–718.
- [SB87] Schwartz, S. H.; Bilsky, W.: Toward a universal psychological structure of human values. In *Journal of Personality and Social Psychology*, 1987, 53; pp. 550–562.
- [SC12] Shen, M. et al.: Emergency Management Information Systems: Could Decision Makers be Supported in Choosing Display Formats? In *Decision Support Systems*, 2012, 52; pp. 318–330.
- [Sc92] Schwartz, S. H.: Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries: *Advances in Experimental Social Psychology* Volume 25. Elsevier, 1992; pp. 1–65.
- [SD15] Schuhbeck, V.; Dorner, V.: The Effect of Defaults and Task Difficulty on Consumer

Satisfaction - Implications for Value Co-Creation Processes: *Wirtschaftsinformatik Proceedings 2015*. Paper 98, 2015.

- [TA01] Turban, E.; Aronson, J. E.: *Decision Support Systems and Intelligent Systems*. Prentice Hall, Upper Saddle River, NJ, 2001.
- [VS15] Varela, M. et al.: QoE in the Web: A dance of design and performance: 2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX), 2015; pp. 1–7.
- [XB07] Xiao, B.; Benbasat, I.: E-Commerce Product Recommendation Agents: Use, Characteristics, and Impact. In *MIS Quarterly*, 2007, 31; p 137–209.
- [ZS13] Zwickl, P.; Sackl, A.; Reichl, P.: Market entrance, user interaction and willingness-to-pay: Exploring fundamentals of QoE-based charging for VoD services: 2013 IEEE Global Communications Conference (GLOBECOM 2013); pp. 1310–1316.

Principles of Information Neutrality and Counter Measures Against Biased Information

Valentin Burger¹, Matthias Hirth¹, Tobias Hoßfeld² and Phuoc Tran-Gia¹

Abstract: The Internet serves as ubiquitous, prompt source of information in our society. Due to the giant amount of information, the access to the information relies on information processing mechanisms as implemented in search engines, product recommendation systems, or online social networks. As a result of the engineered preprocessing mechanism, the retrieved information is biased and does not represent a neutral view on the available information, while end-users often are not aware of this bias. In this article, we define the term “Information Neutrality”, review current principles of information processing in the Internet and discuss influence factors hindering Information Neutrality as well as appropriate countermeasures. The main contribution of this article is to raise the awareness of Information Neutrality as an emerging key challenge in the Internet and to potentially consider Information Neutrality as a factor in the value-sensitive design of Internet-based services.

Keywords: Information Neutrality, information filters, recommendation, filter bubble, information bias, personalization.

1 The Internet as Information Source

In today’s Internet users face the problem to cope with a too large amount of information. Current developments, like the increasing amount of location based data in 5G, sensor networks in smart cities and on wearable devices, and the ability to store and process data using big data techniques, eventually increases the volume of information available. The huge mass of information makes it nearly impossible to find items, which are personally most relevant. Back in October 2012, already more than one billion people were active on Facebook. Imagine how to find potential friends out of one billion people without any specialized means.

To make the selection easier for end-users, mechanisms are needed that generate manageable subsets from this large mass. To provide every individual user with information that is of interest, the subsets must contain suggestions that are tailored on the users’ preferences. The volume of the subset has to be reduced and thus information may be aggregated. However, there is an obvious trade-off between the degree of aggregation and the information contained. Hence, there is a need for systems assisting the user to find relevant information, e.g., by recommendations.

¹ University of Würzburg, Institute of Computer Science, Chair of Communication Networks, Germany
{firstname.surname}@uni-wuerzburg.de

² University of Duisburg-Essen, Chair of Modeling of Adaptive Systems, Essen, Germany
tobias.hossfeld@uni-due.de

Systems assisting users to find relevant information often base on the users past behaviour, which causes the user to end in a positive feedback loop based on its history. It gets hard for the user to find new eye-opening information, the chance for serendipity disappears and the user gets stuck in a so called “filter bubble” [Pa11]. This means web services do not treat users neutrally, but recommend only information tailored on the user. To address this issue, we define the term “Information Neutrality”.

Information Neutrality is the principle to treat all information provided (by a service) equally. The information provided, after being processed by an information neutral service, is the same for every user requesting it, independent of the user’s attributes, including, e.g., origin, history or personal preferences and independent of the financial or influential interest of the service provider, as well as independent of the timeliness of information.

Sociologists are worried about the limited scope of users in the Internet. Books like “Filter Bubble” [Pa11] or “Siren Songs and Amish Children” address this problem of Information Neutrality. Leading Internet companies already react to these concerns. Companies like Google or Facebook improve their algorithms to avoid filter bubbles or provide means to access neutral information. Nevertheless, user profiles and tailored information will always be present, since customer oriented ads are very profitable.

Computer scientists as well as software engineers and companies have ethical responsibility developing and designing information neutral algorithms. Supported by recommendations, the large mass only propagates emotional issues and topics, suppressing serious topics. Appropriate algorithms can help to identify serious topics or counter discrimination by also showing contributions of minorities.

Examples for such mechanisms are already developed in related work. In [Ka12] an algorithm is developed that enhances neutrality of recommendations towards a viewpoint specified by a user. This is achieved by prohibiting the use of some information for the purpose of making recommendation by privacy policies. In [Ma12] semantic-based recommendation systems are used to increase the serendipity in different use-cases by enriching the data with user activities, interests and other meta data for a better alignment of user profiles. [Ng14] measure the filter bubble effect in terms of content diversity received by individual end-users and determine the impact of collaborative filtering-based recommender system, showing that recommendations have a positive effect on the user experience.

The bias on information processed by computer systems has already been analysed in the past [FN96]. In [IN00] the bias of search engines is investigated, showing that certain sites are favoured, which narrows the scope of the web and counters its values such as unrestricted access and widespread information. Investigating the bias of such mechanisms is especially important today, due to the vast amount of data and widely used and more sophisticated algorithms in today’s Internet.

Developers need to provide interfaces to enable the user to decide on the amount of tailored information confronted. This allows end-users to determine the degree of Information Neutrality and to adjust the trade-off between tailored information vs. neutral information.

In this work, we identify factors that influence Information Neutrality and present techniques that help end-users escaping the filter bubble and guidelines for providers to improve their service to be information neutral. Finally, we discuss key questions and challenges that have to be faced to achieve a good trade-off between privacy and usability of services.

2 Definition of Neutrality in the Internet

To get a clear definition of Information Neutrality we need to distinguish it from different kinds of neutrality concerning the Internet. Please note that we do not intend to evaluate the existing neutrality concepts, because there are still ongoing debates whether (and which kind of) neutrality is beneficial for which stakeholders. Neutrality in the Internet is often used with different meanings. In order to clarify our understanding of neutrality in the Internet, we distinguish the following four different types:

1. Network Neutrality also known as Net or Internet Neutrality: Equal treatment of all Internet traffic.
2. Transport Neutrality: Message exchange independent of transport protocol.
3. Search Neutrality: Even-handed treatment in search results.
4. Information Neutrality: Equal treatment of information (includes Search Neutrality).

Strict Network Neutrality “prohibits Internet service providers from speeding up, slowing down or blocking Internet traffic based on its source, ownership or destination.” Network Neutrality ensures that users and applications are treated equally in a way that no user is discriminated depending on its origin or network usage. Hence, the access to the Internet is not limited depending on the location of the connection and on the application or platform used. [BEa] shows that Network Neutrality is violated by all providers in Europe by prioritizing throttling and blocking traffic. Peer-to-peer traffic, for instance, is throttled or blocked in many transport networks and Voice-over-IP traffic is blocked in certain mobile networks. Hence, non-profitable traffic, or traffic that produces transit costs or that is produced by applications that replace the services offered by providers, is blocked by providers. Throttling and blocking is commonly realized by deep packet inspection. A network neutral service forwards bits with equal priority, no matter to which flow, user, or operator they belong. An example for a mechanism that enables Network Neutrality is the transmission control protocol (TCP). The available bandwidth on a link is shared equally among concurring flows. However, emerging technologies like OpenFlow and software defined networking aim to dynamically control and prioritize flows by software. This new paradigm will drastically limit Network Neutrality, but will also open new business models and efficient resource allocation and utilization. Network Neutrality is also a part of

Internet governance, which tries to develop shared norms and rules that regulate the usage of the Internet among governments that can have very different interests.

Transport Neutrality is a concept that allows message exchange of web services independent of the transport. This is realized by a distinction between payload and headers, so that application protocol information can be captured without creating dependencies on transport protocols. Sender and receiver have to share an interpretation of headers placed in the message. For example, web services using the Simple Object Access Protocol (SOAP) are transport neutral, since the SOAP envelope provides a framework for separating a payload from accompanying headers. In contrary to transport and Network Neutrality that consider the data transmission, Information Neutrality considers the bias put on information by presenting information filtered based on user profile and personal preferences, history of user, location of user, timelines, etc.

An existing concept is Search Neutrality that considers even-handed treatment in search results. Search Neutrality is defined as “the principle that search engines should have no editorial policies other than that their results be comprehensive, impartial and based solely on relevance.” In contrast to Search Neutrality, Information Neutrality requires neutral information while consuming a service without explicitly searching for information, e.g. feeds in on news aggregators, recommendation of new friends, or hints on related products.

The system bias introduced by big providers of information systems poses a threat on equal treatment of users, businesses, and opinions in the Internet. Service providers determine what users read, consume, use, and purchase online. Hence, service providers have a huge influence, which can be abused to shape public opinions and to gain competitive advantages. Hence, this also has an impact on business and support monopoly. To maintain a fair market and free opinion making Information Neutrality matters. Regarding this, the European Commission has already raised an issue concerning the anti-competitive behavior by Google in changing the shopping search results for price comparison from a vertical service to directly monetized ads.

To be applicable to recent developments and applications in the Internet, the common layer model has to be extended considering the user and its social interactions as shown in in Figure 1. Depicted is the network stack on layers 1 to 7 as defined in the OSI model. Layers 8 and 9 are added corresponding to end-users and the social network. Layer 8 considers communication among end-users and provides interfaces to end-user as the Quality of Experience perceived or the personal preferences as well as personal information provided. The social interaction between end-users and their organization in social networks provides a vast amount of information, which is subsumed in layer 9. Structural information reveals the organization of end-users in communities and the formation of clusters where information flows within as well as across communities. Interfaces to the social network layer 9 are necessary to get access to the information provided, which is highly valuable for emerging business models. The different types of neutrality identified touch different layers of this extended model. Depending on the considered scope of Network Neutrality it can concern everything from the physical layer to the user layer, corresponding to equal treatment of bits and to equal treatment of flows of different users. Information Neutrality considers equal treatment of information provided in social networks, filter bubbles due

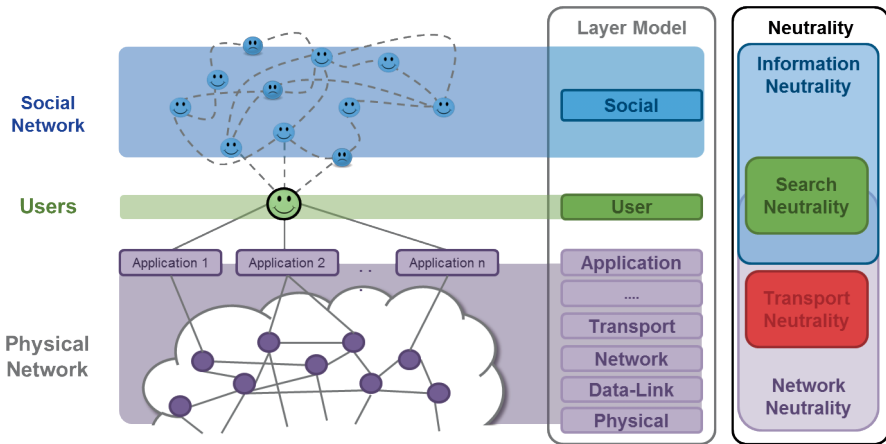


Fig. 1: Layer Model and Neutrality in the Internet.

to tailored information based on user profiles as well as the bias put on information by application interfaces.

3 Information Processing Mechanisms

There are different information processing mechanisms, which influence the Information Neutrality derived in the Internet.

The most basic mechanism, which leads to non-neutral information, is *information filtering*. In general, it is used to derive a small manageable subset from a very large set, which could not be processed in a decent amount of time. There are different policies to filter information, for example on most websites content can be filtered by popularity or recency. By confronting the user with filtered information, other topics are dashed and a bias is put on the information presented. Hence, the information the user derives is non-neutral. Simple mechanisms are collaborative and content based filters. An example for collaborative filtering is Amazon's product recommendation: "Customers Who Bought Items in Your Recent History Also Bought". It relies on the assumption that there is a high probability that a user who bought a specific product is also interested in products other users bought, who bought the same product. Content based filtering differs from collaborative filtering in using attributes of the items instead of using the collaborative intersection.

Secondly, *tailored information* is filtered information tailored to the user's profile. For example, to increase the click-through rate on advertisements and products, user profiles are set up, to identify the users' interest and to provide users information in which they are likely interested. The problem is that the information provided is based on the user profile, which is a reflection of the user's past activity. The information seen is similar to the information consumed in the past. This prevents discovering new unexpected topics and reduces the chance to find interesting.

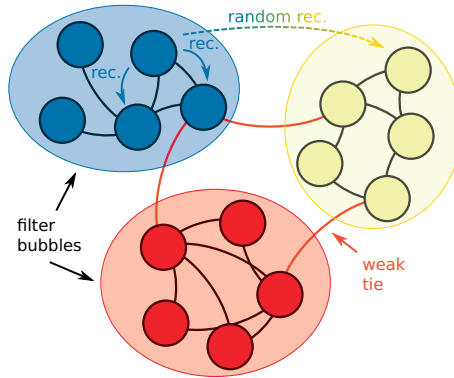


Fig. 2: Filter bubbles in a network with recommended items.

Thirdly, there are means to provide the user with an *aggregation/condensation of information*, e.g., summaries of different blogs, articles, products, which help him to get an overview of a large collection of objects belonging to a topic that he can process in an affordable amount of time. Given a large set of similar information objects, these techniques summarize the essence or locate the source of information. However, interesting additional information, e.g. details, thoughts or comments on a discussed topic, might be suppressed.

Finally, there are *recommendations*. In recommendation systems, attributes of users, their history, as well as recency and the network structure are used to predict which friend, product, video the user is going to connect to/consume next. Hence, a subset of objects is selected based on all considered attributes of the user and the surrounding network. This puts a bias on the recommended objects and therefore fades out different objects. Further, it supports segmentation of communities by strengthening ties between objects of shared interest. It also reduces the potential of building bridges to different communities, since until now, most recommendation metrics foster strongly connected components.

Consider for example the mechanism recommending friends on Facebook. The algorithm behind is based on a supervised random walk [BL11] that aims to predict which links in the networks are going to evolve in the near future. A rich set of attributes is used to guide a random walker on the network graph by adding strength to the edges. The strength of the edge determines the probability of the random walker to follow an edge. A learning task with the goal to find a function assigning strengths to the edges is used to make the random walker visit nodes more likely that the user is going to interact with. The underlying optimization problem cannot be solved exactly, because of the large amount of users in the network. Therefore, approximations are used and only a subset of nodes can be considered. In the case of Facebook only friends-of-friends are considered, which count already 40k on average. Figure 2 shows a network with different items, which can be persons, products, news articles. Links connect associated items, for example, by having the same interest or being purchased together. In the network are different clusters with items that are similar and are relevant with high probability. If a mechanism only recommends similar items that are directly connected to the initial item, it is hard to escape these filter

bubbles. If random items are recommended, they might be in another cluster and the filter bubble can be escaped. Another possibility to escape the filter bubble is to recommend items connected to weak ties. Weak ties are links that connect different clusters. According to [Gr73] most interesting information is derived via weak ties, hence from people of different communities, which are only slightly in contact.

4 Non-Algorithmic Influence Factors on Information Neutrality

Besides the different information processing mechanisms, there are also non-algorithmic influence factors on Information Neutrality.

4.1 Application Interfaces

A technical, yet not algorithmic, influence factor on Information Neutrality are application interfaces. Application interfaces can have an influence on Information Neutrality, because they determine how the information is presented to the user. A simple example is the formatting of headlines in newspapers. Articles with large and bold headline immediately draw more attention even if they might not be as relevant and interesting to the readers as other articles with headlines that are not so prominent. This immediately puts a bias on how the available information is perceived. In an online scenario, the arrangement of search results and advertisements determines what a user is likely to click. Google places customer oriented ads as first results on its sites, because users click on them with high probability. To protect customers, Google had to change this and make advertisements clearly recognizable. The number of results per page can also have an influence on what a user consumes. Recent work shows that most users only click on first three pages of Google Search. Top sites like YouTube, Facebook or Amazon frequently change their interface. For instance, the interface of YouTube changed from the subscriptions being the front-page to recommended videos being on the front-page. The videos seen on first sight are no longer the videos from subscribed channels, but also recommended videos based on your profile.

4.2 Social Relationships and Proximity

Social networks use recommendation systems not only to suggest friends, but also to decide which posts users are confronted on first sight. Different topics, which are not presented on the front-page, are suppressed. Facebook for example filters the news on the front-page. Algorithms calculate and decide which of your friends or stories are interesting for you and show only these activities. This service helps to manage the large mass of posts by friends and their activities. However, if the user does not claim interest to a person, it might disappear from the feed and the user might totally forget about her or him. Further, users are only exposed with impressions of their social environment, which hardens their point of view and offers less space for revising their opinions. This can also

harden fronts between opposing parties. Further on, emotional issues and topics are more likely to be propagated by the large mass, suppressing serious topics.

Proximity has a high impact on our social relationships and the communities we join determine our environment. Users are interested in people and events that are locally close. Hence, there is also a bias on these topics. The authors of [LYW15] call this phenomenon the “majority illusion” and show that behavior that is globally rare may be systematically overrepresented in the local neighborhoods of many people, i.e., among their friends.

4.3 External Influence Factors

Information reaches the users from both, mass media and from personal social network, meaning the friends that the users meet in person. Myers et al. [MZL12] investigated information diffusion of the OSN Twitter, they find that “only about 71% of the information volume in Twitter can be attributed to network diffusion, and the remaining 29% is due to external events and factors outside the network.” That means that about one third of the information users get is from external influences.

5 Guidelines Towards Neutral Information

Both service providers and users can counter tailored information. Service providers can add means to disable information filters or make them adjustable to the personal preferences. End-users have several possibilities to overcome tailored information, even if providers do not directly support this. In the following, first, guidelines for end-users and second, guidelines for service providers are described.

Tailored information can be avoided and recommendations can just be ignored. A user can visit an alternative web site handling information in a more neutral manner. Being aware of weak ties, a user can target information sources accordingly. Such sources can be Facebook walls of friends that are not in the everyday clique, that you met while travelling, that are of different culture or religion. However, for most users it is hard to assess, which web service is most neutral.

Internet and services can be accessed anonymously to prevent recommendation systems to set up user profiles. One prominent tool providing anonymous access is “Tor”, which is a network of virtual tunnels developed with the U.S. Navy. Its primary purpose is protecting government communications. One of its use cases is preventing websites from learning location or browsing habits.

The virtual tunnels are routed via three proxies, so called relays, that change for every connection. Each relay has a different key for decryption and removes the corresponding decryption layer of the message. The layered encryption referred to as onion routing. This hides the origin and prevents the traceability of the request, but also adds a delay depending on the location of the relay, c.f. Figure 3, which can be a problem for non-delay tolerant

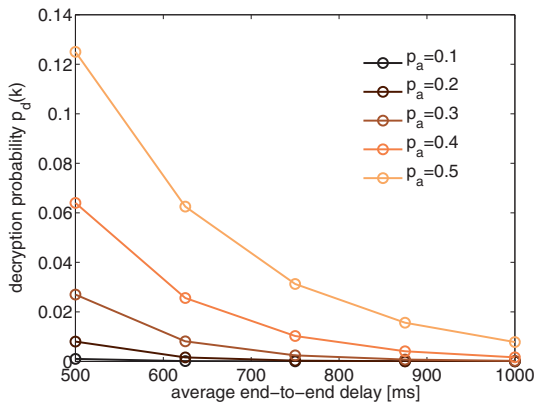


Fig. 3: Trade-off between decryption probability and average end-to-end delay for different attack probabilities. If the attack probability p_a is low, the decryption probability is low independent of the end-to-end delay. If p_a is high, a high delay has to be tolerated for a secure connection. Hence, anonymity can either be bought with waiting time or by deploying a high number of relays.

applications. An attacker can still decrypt the message, if it controls all relays on the route. So adding more relays on the route increases the probability that at least one relay is not controlled by the attacker.

Another mean towards information neutrality considers tracking cookies. In order to track users, web services store cookies, with user IDs on your local machine. These tracking cookies can automatically be blocked or deleted to avoid tracking. Tools like “NoScript”⁵ or “AdBlock”⁶ can block scripts on websites that read the cookie to forward the information. Modern browsers offer a “Private Browsing” function, which allows running a clean browsing session that does not store cookies or deletes them after closing the session. However, there are very persistent cookies like “evercookie”⁷ that store the cookie data in several types of storage mechanisms and recreate the cookie, if any of the stored cookie types has been removed. A website that lets you check how safe your browser is against tracking is provided by “Panopticlick”⁸.

In general, to get out of the filter bubble one can simply look beyond the front page of websites. An example is the Facebook front page. The default setting for the Facebook news feed is “Top Stories”. The “Top Stories” presented are selected by a filtering algorithm according to the user’s interests, recent activity, and contacts. To get rid of this filter the setting can be switched to “Top Stories”, which shows the unfiltered and neutral version of the news feed. In [LYW15] an interactive method to visualize the personalized filtering is presented, which provides awareness of the filter bubble in online social networks.

⁵ <https://noscript.net/>

⁶ <https://adblockplus.org/>

⁷ <https://samy.pl/evercookie/>

⁸ <https://panopticlick.eff.org/tracker>

It would not be necessary to use tools like “TrackMeNot”⁹ if service providers add means to disable information filters or make them adjustable to fit the personal preferences. In the following we give examples for guidelines that can be adopted by service provider for fostering neutral information.

A profile is visible as soon as some queries concerning a certain topic are used more frequently. Tools like “TrackMeNot” inject random search queries, so that every topic has equal weight, to destroy the user profile. A service provider can add random recommendations to provide neutral information. In the end there will be always user profiles constructed to increase profits of the commercial industry. For instance online social network users’ web histories are tracked even if they are not logged in. This is realized by tracking cookies that are integrated in the popular share-buttons. These share-buttons are distributed on websites all over the Internet.

Service providers can use more sophisticated recommendation mechanism. In [Zh12] a tool for music recommendation “Auralist” is developed, which aims to mimic the actions of a trusted friend or expert. Such accuracy, diversity, novelty and serendipity of the recommendations shall be improved simultaneously and balanced. Therefore, they investigate three algorithms, namely Latent Dirichlet Allocation, Listener Diversity and Declustering. The algorithms are briefly described in the following.

Latent Dirichlet Allocation clusters users with similar preferences. Composition vectors define the listener base of each artist. A similarity metric for the composition vectors is then used to calculate recommendations for a user. Listener Diversity promotes artists with diverse listener communities. Such users are encouraged to explore beyond a given niche. Declustering directly addresses Information Neutrality by identifying “music bubbles” in form of similar artist clusters and counteracts them by recommending least clustered or “boring” items for a user. The result of the combination of all three algorithms especially enhances serendipity in the recommended music.

Soylent [Beb] is an extension for Microsoft Word that uses a crowdsourcing platform. One of the functionalities of Soylent is text shortening, which can be used to filter the most valuable information out of a document. The key concept of Soylent is to divide the text-shortening task in different subtasks that can be submitted to a crowdsourcing platform. The subtasks are identifying paragraphs that can be shortened, shortening the paragraphs and proofreading the shortened paragraphs. A number of different workers perform each step to crosscheck results to assure a certain quality of the condensed text.

Summly is a news app for smartphones that condenses news to 400 characters in readable sentences, such that it keeps the main message. The algorithm identifies, extracts, and combines the most important sentences of the article. The shortened news text with less than 400 characters fits on a smartphone display and contains enough information to get a quick overview. A longer summary or the original article is linked for detailed information. Summly uses a genetic algorithms and machine learning for text summarization. The algorithm learns from the well-formed structure of news articles.

⁹ <https://cs.nyu.edu/trackmenot/>

6 Key Questions and Challenges

Information neutrality is important for end-users as well as service providers. Its characteristics are dependent on every single stakeholder. First, convenience and the possibility to get a quick and good selection from a vast set of items in a short time are important for end-users. But, end-users need to be aware that the information presented is filtered and that there is a bias on the results. The question is, to which extent information is filtered and how to show the bias and that the information is filtered?

Second, it is also important for end-users to discover new items and topics to experience serendipity. This can be accomplished by better mechanisms or random results. Hence, the question is how to improve mechanisms and which amount of random results to add?

Further, users must be able to decide on how tailored the information presented is. A user might just want to access neutral results, to get the unbiased view. Therefore means, like sliders that range from “neutral” to “personal” could be provided by service providers, so that the user can adjust according to its personal preferences. The question is how to measure the degree of information neutrality and how to provide means to let the user decide on the degree presented?

Metrics are necessary in order to measure the degree of information neutrality. A problem to define appropriate metrics is to define which information can be considered as totally neutral. A value determining the information neutrality of a service could for example be measured by the amount of data used from user profiles to process the information.

Users have to be aware that the personal information they provide is of high value for service provider. Service providers may not abuse the capital they get, which is the information base collected from all users. The information base can be used for opinion shaping as well as provisioning of non-competitive services. Public opinion has an impact especially on democratic societies. The information base and the access to many users can be abused to form the opinion according to the interest of the company. Provisioning of non-competitive services helps huge companies to further strengthen their position. This threatens the free market and will lead to monopolies of the giants. But companies can also take the chance and use their means to identify serious topics or counter discrimination by also showing contributions of minorities. The challenge is to provision fair, neutral and inspirational services, Companies as well as developers of information systems need to be aware of that. Information neutrality has to be considered by companies, organizations and government to establish policies.

7 Conclusion

Like all commercial companies, information providers seek to maximize their revenue. For web services, the revenue is often tightly connected to the click through rates on the offered products or content. Good recommendations and customized content can increase the click through rate of less than 1% to more than 5%. To enable customized content, user data has to be accumulated. However, users of web services do not know and cannot

control what is collected. It is not clear how the collected data is processed and exploited by companies and how much they profit from it. Hence, there is a value of the personal data we provide, which is not rewarded by service providers.

The massive data accumulation may not only have severe consequences on information neutrality, but also on the privacy of the user. The information collected by recommendation systems is personal and can be highly sensitive. To protect the users' privacy it has to be clear, which personal data may be collected, who may access personal data and how it may be used.

Algorithmic driven user profiles limit the information neutrality as they preserve models of every user. People change and so do the models have to change. Falsifiability of the models has to be considered to train them efficiently. Web services must experiment with their users and present them new and random content to even refine the users profile, keep it up to date and help the user encounter serendipity.

There is a trade-off between the privacy and serendipity level of a service and its usability. Strong user profiles and tailored recommendations lack privacy and serendipity, but provide a good set of results quickly. Interfaces have to be provided that make the bias on the information visible to the user and let the user decide on the degree of neutrality in the information provided. It is part of future work to investigate how to enable user empowerment in web services.

References

- [BEa] BEREC: , A view of traffic management and other practices resulting in restrictions to the open Internet in Europe, 2012.
- [Beb] Bernstein, Michael: , It's People! Meet Soylent, the Crowdsourced Copy Editor.
- [BL11] Backstrom, Lars; Leskovec, Jure: Supervised random walks: predicting and recommending links in social networks. In: Proceedings of the fourth ACM international conference on Web search and data mining. ACM, pp. 635–644, 2011.
- [FN96] Friedman, Batya; Nissenbaum, Helen: Bias in computer systems. *ACM Transactions on Information Systems (TOIS)*, 14(3):330–347, 1996.
- [Gr73] Granovetter, Mark S: The strength of weak ties. *American journal of sociology*, pp. 1360–1380, 1973.
- [IN00] Introna, Lucas D; Nissenbaum, Helen: Shaping the Web: Why the politics of search engines matters. *The information society*, 16(3):169–185, 2000.
- [Ka12] Kamishima, Toshihiro; Akaho, Shotaro; Asoh, Hideki; Sakuma, Jun: Enhancement of the Neutrality in Recommendation. In: *DecisionsRecSys*. pp. 8–14, 2012.
- [LYW15] Lerman, Kristina; Yan, Xiaoran; Wu, Xin-Zeng: The majority illusion in social networks. arXiv preprint arXiv:1506.03022, 2015.
- [Ma12] Maccatrozzo, Valentina: Burst the filter bubble: using semantic web to enable serendipity. In: *The Semantic Web–ISWC 2012*, pp. 391–398. Springer, 2012.

- [MZL12] Myers, Seth A; Zhu, Chenguang; Leskovec, Jure: Information diffusion and external influence in networks. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 33–41, 2012.
- [Ng14] Nguyen, Tien T.; Hui, Pik-Mai; Harper, F. Maxwell; Terveen, Loren; Konstan, Joseph A.: Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity. In: Proceedings of the 23rd International Conference on World Wide Web. WWW '14, ACM, New York, NY, USA, pp. 677–686, 2014.
- [Pa11] Pariser, Eli: The filter bubble: What the Internet is hiding from you. Penguin UK, 2011.
- [Zh12] Zhang, Yuan Cao; Séaghdha, Diarmuid Ó; Quercia, Daniele; Jambor, Tamas: Auralist: introducing serendipity into music recommendation. In: Proceedings of the fifth ACM international conference on Web search and data mining. ACM, pp. 13–22, 2012.

2nd International Workshop on
Big Data, Smart Data and
Semantic Technologies
BDS DST 2016

2nd International Workshop on Big Data, Smart Data and Semantic Technologies – BDS DST 2016

Nico Roedder¹, Stefan Zander¹, Natalja Kleiner¹, Suad Sejdovic¹, Stefan Jähnichen², Rudi Studer³

In modern societies, almost every individual is using technology for entertainment, communication or business purposes. Depending on roles and goals, as well on the application domain, the variety of data people work with is becoming more and more complex. At the same time, the growing technical possibilities to gather and aggregate multi-modal data from sensors and different services allow for evidence-based information systems enabling both humans and machines to make well-informed decisions.

These large and complex data sets, commonly characterized by high data volume, high variety of the data types and data sources, high velocity of the incoming data and the expected information output (real-time requirement) as well as the uncertainty about the veracity of the data, are known as Big Data. These characteristics make it difficult to process the data using existing data management applications and traditional information technologies. On the other hand, when processed and analyzed properly, Big Data is transformed into “Smart” Data and might carry huge amounts of useful information, which was not accessible beforehand and allows for better-founded, more robust predictions and improved decision-making in almost any domain. That is why new predictive and prescriptive analytic approaches are continuously increasing in importance.

Besides new analytic approaches, novel information technologies such as semantic technologies are necessary in order to exploit the full potential of the gathered data. Unlike traditional information technology where the meaning of data and their relationships are predefined and “hard-wired” into data formats and applications, semantic technologies encode meanings explicitly and independent from concrete formats and application logic. This enables machines and people likewise to understand, share, and reason over semantically represented data. Semantic technologies provide an abstraction layer on top of existing ICT infrastructures and facilitate the interrelation and integration of data, content, and processes in meaningful ways, which is very important when dealing with high amounts of heterogeneous data.

¹ FZI Resarch Center for Information Technology, Haid-und-Neu-Str. 10-14, D-76131 Karlsruhe, Germany, {lastname}@fzi.de

² FZI Resarch Center for Information Technology, Friedrichstraße 60, D-10117 Berlin, Germany, jaehnichen@fzi.de

³ Karlsruhe Institute of Technology (KIT), Englerstraße 11, D-76131 Karlsruhe, Germany

We believe that using Big and Smart Data as well as methods and tools based on semantic technologies will provide more transparency, enable precise and well-founded decisions and improve planning processes, which will result in more efficient and user-centric processes and systems.

We are happy that the submissions to BDS DST 2016 reflect our sentiment and as they were innovative and from a wide variety of domains. Sixteen reviewers from eleven institutions, ranging from different fields of industry and research, selected the best submissions. In addition, we are pleased that submissions from ongoing projects of the technology program “Smart Data – Data Innovations” by the Federal Ministry for Economic Affairs and Energy in Germany have made it into the selected contributions of our workshop.

Program Committee

Dr. Andreas Abecker	Informationssysteme GmbH, Karlsruhe, Germany
Christoph Bier	Fraunhofer IOSB, Karlsruhe, Germany
Martin Birkmeier	FIR e.V. an der RWTH Aachen, Aachen, Germany
Dr. Simone Braun	CAS Software AG, Karlsruhe, Germany
Jan Froese	Kuehne + Nagel (AG & Co.) KG, Hamburg, Germany
Dr. Mark Hefke	CAS Software AG, Karlsruhe, Germany
Dr. Martin Junghans	IBM Deutschland GmbH, Stuttgart, Germany
Dr. Paul Karänke	TU München, Munich, Germany
Marco Lewandowski	BIBA Universität Bremen, Bremen, Germany
Prof. Dr. Stefan Nickel	Karlsruhe Institute of Technology, Karlsruhe, Germany
Prof. Dr. Jens Nimis	Hochschule Karlsruhe, Karlsruhe, Germany
Dr. Gerald Ristow	Software AG, Darmstadt, Germany
Fabian Schmidt	Software AG, Darmstadt, Germany
Maximilian Schreieck	TU München, Munich, Germany
Dr. Manuel Wiesche	TU München, Munich, Germany
Dr. Thomas Usländer	Fraunhofer IOSB, Karlsruhe, Germany

Semantic Object-Oriented Programming (SOOP)

Florian Weber¹ Andreas Bihlmaier² Heinz Wörn³

Abstract: Currently, the interaction between ontologies and general purpose programming languages mostly consists of an indirect mapping between the ontology and the programming language. The language and libraries are often basically used as a programmable ontology editor. In this paper we present a different approach that creates an ontology from regular data objects in a (statically typed) mainstream language (C++). Because in our case the mapping is going from object-oriented programming (OOP) language to the ontology, we can avoid many problems that approaches going into the opposite direction have to face. Furthermore, the interaction between the domains becomes more straight forward and can be mostly kept free of uncommon language constructs. The declarative programming paradigm on the one hand and the (object-oriented) imperative paradigm on the other hand are in this manner combined into the Semantic Object-Oriented Programming (SOOP) paradigm. As a result, SOOP allows programmers to directly use semantic technologies, especially reasoners, for their usual language objects from within C++.

Keywords: Ontology, Object-Oriented Programming (OOP), Semantic Reasoner, C++

1 Introduction

An ontology is a formal, explicit specification of a shared conceptualization [Gr92][SBF98]. This means that it is a suitable way to represent data for theorem-provers and reasoners, so that they can algorithmically search answers for questions posed in the ontology language. For example, a subject matter, modeled using the ontology, can be checked for consistency. Similarly, the set of individuals satisfying a formula can be extracted.

Since two different domains are brought together here, semantic technologies on the one hand and object-oriented programming (OOP) on the other, we must define some terms (cf. Tab. 1): *Atom* or *individual* refers to a logical individual, i.e. an element of the universe of discourse. An *axiom* is a logical statement that is either true or false. *Predicate* and *quantor* are used in the same manner as in predicate calculus and related logical formalisms. Also *variable* is used in the sense of predicate calculus, more specifically it is synonymous with bound variables.⁴ The terms *class* and *object* have the same meaning as in OOP, they refer to the definition of a data type and its instances. *Entity* is used in both domains, it can refer to an individual (semantic domain) or to objects (OOP domain) that represent the individuals within the programming language.

¹ Karlsruhe Institute of Technology (KIT), Institute for Anthropomatics and Robotics (IAR) - Intelligent Process Control and Robotics (IPR), Engler-Bunte-Ring 8, 76131 Karlsruhe, Germany, uagws@student.kit.edu

² KIT, IAR-IPR, andreas.bihlmaier@kit.edu

³ KIT, IAR-IPR, woern@kit.edu

⁴ In the context of OOP source code, C++ *variable* is used explicitly to denote program variables.

As will be explained in the following, entity does not only connect the domains terminologically, but is the core concept that facilitates semantic object-oriented programming (SOOP).

Tab. 1: Illustration of how the vocabulary used here is assigned to the OOP and semantic domain.

OOP	Shared	Ontologies
Class	Entity	Atom, Individual
Object		Axiom, Variable
		Predicate, Quantor

The most common method to fill such an ontology with data currently is some kind of ontology editor or ontology API. However, even the APIs that allow to define ontologies from within programming languages, don't allow a direct combination of existing program data structures with the ontology. This results in a strong division between the raw data and the semantic data. The raw data is easily accessible from the programming language (e.g. printing, sorting or direct calculations). The semantic data, presented in the ontology, is easily accessible to reasoners, but falls short with regards to usability in the programming language.

One example for this kind of gap is the OWL-API. While the semantic data is well accessible from Java, the Java objects that represent semantic entities have only a single Java type - and thus have no semantics from the programming language point of view: Entities have types like `OWLLiteral` or `OWLNamedIndividual`; this means that the static type-checker is unable to verify the absence of many obvious errors such as an email address being assigned to a username or a company to vehicle.

Our goal was to find a new approach that brings ontologies together with mainstream programming languages, while avoiding the loss of type information. Thereby, creating a deeper integration compared to existing methods. To avoid the common problems that result from ontologies having greater expressive power compared to object-oriented programming (OOP) languages (as pointed out by [Or07]), especially statically typed ones, we decided to reverse the mapping direction: Instead of creating ontologies with some kind of standalone ontology editor and a subsequent import of the resulting ontology description into the programming language or reasoner, our approach goes directly from OOP to a reasoner. For this purpose, we extended or wrap programming language data structures in such a way that they can be used as semantic entities and combined with existing ontologies. Since ontologies have more expressive power than OOP languages, they can easily represent all the semantic connections that exist in OOP. We call this novel deep integration between semantic technologies and OOP languages *Semantic Object-Oriented Programming (SOOP)*.

The remainder of this paper is structured as follows: In the following section an overview of existing integration approaches between ontologies and OOP is provided. Section 3 details the Semantic Object-Oriented Programming (SOOP) concept and its implementation in C++. Afterwards, the benefits of our integration approach for semantic technologies with OOP are illustrated in an exemplary use case (section 4). The impact of SOOP for this

use case and a general discussion follows in section 5. In the final section, open questions and directions of further research are addressed.

2 State of the Art

Nowadays ontologies are primarily created with specific ontology-editors. Particularly well known in this context is Protégé [Kn04]. With regards to the use from within programming languages, there exists a relatively wide range of approaches, some of which we will present in more detail.

However, first, we will describe those approaches in which entities of an ontology are represented in an OOP by a set of generic classes. The aim here is not to create exact representations of the ontology within the OOP-language, but to provide a programmable tool for editing them. These approaches are also called “indirect approaches” [Pu08]. The most important software libraries to mention here are the OWL API [HB11] and Jena [Je13] from the Apache project.

The OWL API is a Java library, which allows access to OWL triples with a comparatively low level of abstraction. Furthermore, it provides a general interface for provers, allowing to easily exchange them with each other. The API provides both the ability to ask questions by means of queries and to edit the ontology. Changes made through the library can be written back to permanent storage. The OWL API is used as backend by Protégé. Apache Jena follows a similar approach, but offers a higher level of abstraction. The basic idea of Jena is to represent data and types of ontologies as objects of classes derived from `OntResource` (for instance `resource`) and to use these through generic methods. The obvious downside of both OWL API and Jena is the lack of typing with regards to the ontology objects in the Java environment and the strong focus on the use of strings instead of native language objects with types.

Moreover, it is also possible to realize simple queries in Jena with native Java (for example to iterate over the classes of an individual). However, the recommended method for complex requests is to use the query language SPARQL (SPARQL Protocol And RDF Query Language). SPARQL is a language that is standardized by the W3C and modeled after SQL (Structured Query Language). It allows to formulate relatively complex semantic requests, which are then answered by a prover.

At the other end of the existing approaches spectrum, there are attempts to translate the data from ontologies into regular OOP languages, which is also known as the “direct approach” [Pu08]. Unfortunately, there are some fundamental problems with this kind of translation, due to the fact that ontologies are much more general and expressive than most OOP languages. Nevertheless a couple of interesting approaches can be found in literature and are discussed in the following.

ActiveRDF [Or07] for example bypasses many of the common limitations by using Ruby as its implementation language instead of the otherwise widespread Java. As a result many of the common problems posed by a static type system disappear, since Ruby is not only

dynamically typed, but also allows to catch any attempt to use non-existing functions. Yet, it is exactly this dynamic type system that implies several downsides for this approach. Most notably is the lack of basic (compile-time) sanity checks that static type systems provide.

The approach “SWCLOS” presented in “OWL vs. Object Oriented Programming” [KAH05] also uses a dynamically typed Language (Lisp + CLOS). Unlike our work, they concentrate on how a program that uses ontologies can be typed to represent semantic structures, whereas we emphasize the semantics of dynamically created objects.

The final related proposal that tries to unite ontologies and object-oriented programming, can be found in “Integrating Object-Oriented and Ontological Representations: A Case study in Java and OWL” [Pu08]. There an attempted is made to combine direct and indirect representations by making the gap between the ontologies and the OOP data “moveable”. Nevertheless, in contrast to SOOP, there are still two “worlds” with some entities only existing as individuals in the semantic domain or as objects in the OOP domain.

3 SOOP

The fundamental idea behind SOOP is to create or populate an ontology from within C++ using regular C++ classes and objects.

This means that we perform the data mapping into the opposite direction compared to the direction of the approaches described in section 2. As an advantage of this direction, many of the usual problems that result from the attempt to map data from a more powerful data representation into a less powerful one are no longer an issue. See table 2 for a summary. Furthermore, the SOOP approach allows us not only to preserve static typing, but also to make active use of the type system in a way that profits the ontology. It is possible to use the C++ entities as regular C++ types, also outside of the ontology context.

Existing knowledge in form of an ontology could be imported into SOOP in two ways: One method directly imports the ontology into the reasoner from a textual representation such as SUO-KIF, i.e. entities from the ontology do not exist as OOP classes and objects. The other method generates OOP code from the ontology⁵, which in turn uses SOOP as described below.

⁵ The constraints described in Tab. 2 apply here.

Tab. 2: Ontologies provide a much more powerful and dynamic representation of data compared to object-oriented programming languages. Therefore it is easy to translate data from OOP languages into an ontology, but basically impossible to do it the other way round (left side according to [Or07]).

Ontology \rightarrow OOP	OOP \rightarrow Ontology
Mapping the dynamic relations of entities into a static type system is basically impossible.	Mapping the type relations of a static type system into a dynamic ontology is trivial.
In ontologies, arbitrary objects may be linked using arbitrary relations. This can be hardly translated into object attributes.	Attribute connections are relatively simple in OOP languages and easy to map into an ontology.
An ontology is a hypergraph that can naturally represent very complex relations. It is not in general possible to map this to the acyclic inheritance graphs of OOP languages, even if these support multiple inheritance.	The directed, acyclic inheritance graphs of OOP languages are always representable, since they represent subgraphs with respect to the much more general hypergraphs of ontologies.
A central mechanism of ontologies is the dynamic evolution of its structure. However, this can result in significant problems, if it is to be translated into a static language.	The constant inheritance structures of OOP languages can easily be transformed into dynamic ontological structures.

3.1 General concept

In order to structure the SOOP concept, we divided it into the following submodules:

- A prover module that handles the communicating with a theorem prover (Z3).
- An ontology module that manages the knowledge and creates queries for the prover module.
- A module that implements the base class of all entities that are backed by a C++ object: `entity`.
- A module that implements formulas in a way that preserves their semantics in a usable manner, while providing a simple syntax in C++ that closely resembles predicate calculus.

The `entity` base class stores a pointer to the associated ontology and an integer that serves as unique ID within the ontology. Due to the unclear semantics of copying⁶, no copy-constructor is provided. In contrast, move operations are allowed and simply inform the ontology of the new address of the moved entity. Every class that inherits from `entity` can be used as an entity of the ontology context. Thereby, it is fully supported for these classes to have regular data members and methods, whose meaning or even existence is

⁶ See the discussion in section 6.

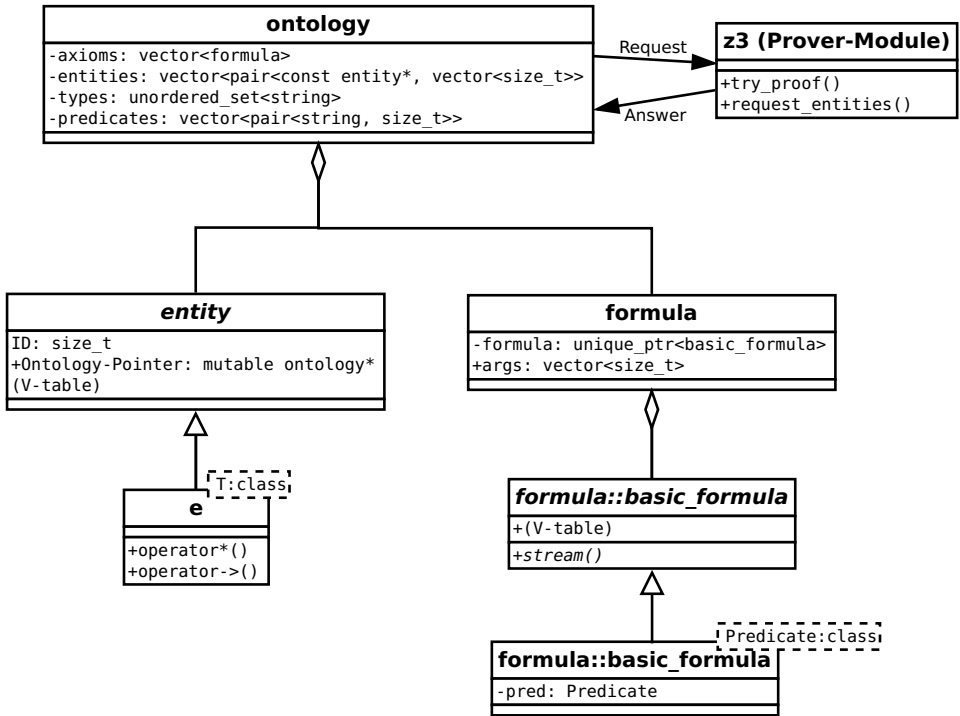


Fig. 1: Overall SOOP architecture in terms of modules and their relations.

not formalized within the ontology. Furthermore, it allows regular usage of the data with normal language facilities, such as sorting or printing functions.

In order to also support the use of already existing types, a class template `e` is provided that wraps the existing types in such a way that the wrapped version (`e<T>`) inherits from `entity` and is therefore usable from within the ontology.

Predicates in SOOP consist of two parts: First, a class template, whose instantiations represent saturated predicates and contain its arguments. Second, a function template that instantiates the class template. By convention the function template is named identically to the predicate (e.g. `f`) and the class template receives `_t` as a suffix (e.g. `f_t`). Through nesting of such predicates, it is possible to create more complicated expressions and types. If `f`, `g` and `h` are for instance predicates and `a` and `b` are entities that are instances of the types `A` and `B`, then `f(g(a), h(a))` will result in the type `f_t<g_t<A>, h_t>` for the entire expression. In order to store different formulas in mixed containers, a wrapper that makes use of type erasure⁷ was created.

⁷ Type erasure is a technique that allows to treat independent types as if they had a common base class. This is accomplished by a base-class `base` with virtual methods for the desired behavior and a template `derived<T>` that inherits from it and implements them via `T`.

The `ontology` class stores a list of all C++ objects, which are at the same time entities of this ontology. The same class also stores the list of all axioms associated with this ontology. Requests whether a specific set of axioms is satisfiable are processed by creating a full textual representation of the entire ontology, which is then sent to the prover module, whose answer is returned to the caller in C++. Notably, it is not only possible to check for satisfiability of a formula, but also *to request the set of variable assignments in the model and return the actual C++ objects that are associated with them*. Requests for these actual entities are processed similar to the satisfiability check, except that an additional function, which maps all entities to their SOOP ID is defined; the prover is then asked for the value of that function when receiving a suitable entity. These IDs are then used to find the corresponding C++ object, by means of a mapping to its address that is known by the ontology.

In order to get the full power of description logics, quantifiers and variables are needed in addition. The approach that we choose here is to create unique types for all variables by providing a class-template `variable` that receives the identifier as template argument and a class `bound_vars` that receives a variadic number of variables in its constructor and internally saves a textual representation of this list as a string. Thus, it is possible to implement quantifiers in the same manner as predicates with the exception that the former receive an instance of `bound_vars` as their first constructor argument.

A downside of this approach can be seen in the requirement, to define variables before it is possible to use them (at least if a clean syntax is desired). We believe that this is an acceptable trade-off for now, as the `variable` types are empty and can therefore be used for global constants without much overhead.

Combined these features allow the easy creation of complex formulas without requiring a syntax that is unusual for either description logics or C++. Stating for instance that a predicate `f` is transitive can be done with the following valid C++ code:
`forall({x, y, z}, implies(and_(f(x, y), f(y, z)), f(x, z)))`.

3.2 Implementation

We implemented SOOP in C++14 in “modern C++” style.⁸

The basic idea for a predicate `foo` is to provide a class template `foo_t` and a helper function `foo`:

```
template<typename T>
class foo_t: soop::is_predicate {
public:
    foo_t(const T& arg): arg{arg} {}
    void collect_entities(std::vector<std::size_t>& ids);
    void stream(std::ostream& out,
```

⁸ It is available as free software at <https://gitlab.com/FJW/soop>.


```

        const std::vector<std::string>& names) const;
private:
    T arg;
};

template<typename T>
auto foo(const T& arg)
-> foo_t<soop::to_bound_type<T>> {
    return {arg}
}

```

Here `soop::is_predicate` is an empty type that is used by the implementation to dispatch printing and collecting arguments to the correct function templates. The methods `collect_entities` and `stream` are needed to allow simple management of different kinds of formulas and will be explained later in this section. The basic idea behind this model is to allow the user to nest predicates such that each predicate-template would be instated with its arguments types, which in turn might be instantiated predicate-templates.

In order to store different kinds of such formulas in the same container, we created another class `formula` that uses type erasure to store arbitrary formulas while providing uniform access to them. This is where the `collect_entities` and `stream`-methods come into play: In order to keep the information about which entities are used inside the formula easily accessible and editable, their IDs are collected by the former in a depth-first tree traversal and memorized in a regular vector outside the erased type. Printing a formula is accomplished by converting the collected IDs to their textual representations and passing them into the `stream` method. This in turn also traverses the tree depth-first and uses the IDs in all places where entities were collected earlier.

Since the definition of predicates does require a lot of boilerplate code, SOOP provides four macros that allow the definition of a new predicate in just one short statement:

```

SOOP_MAKE_TYPECHECKED_RENAMED_PREDICATE(Id, Name, Rank, ...)
SOOP_MAKE_RENAMED_PREDICATE(Id, Name, Rank)
SOOP_MAKE_TYPECHECKED_PREDICATE(Id, Rank, ...)
SOOP_MAKE_PREDICATE(Id, Rank)

```

These macros allow to create predicates with arbitrary (even unlimited) rank. Optionally, it is even possible to check the types of the predicate arguments in order to benefit from type safety offered by a typed OOP language. The later three macros are defined in terms of the first universal one, but pick sensible defaults for the other arguments.

3.3 A minimalistic SOOP example

In order to illustrate the direct integration between semantic information processing and native OOP in SOOP, a minimal complete SOOP program that will print “Max Mustermann is a parent.” follows:

```
#include <iostream>
#include <string>

#include <soop/onto.hpp>

class person {
public:
    person(std::string name): name{std::move(name)} {}
    std::string name;
};
using person_e = soop::e<person>;

SOOP_MAKE_TYPECHECKED_PREDICATE(parent_of, 2,
                                person_e, person_e)

int main() {
    using namespace preds;
    using namespace soop::preds;
    soop::ontology o{};
    o.add_type<person_e>();
    o.add_predicate<parent_of_t>();
    soop::variable<'p'> p;
    soop::variable<'c'> c;
    // parents and children are persons:
    o.add_axiom(forall({p, c}, implies(parent_of(p, c), and_(
        instance_of(p, soop::type<person_e>),
        instance_of(c, soop::type<person_e>))));
    // parents are not their childrens' children:
    o.add_axiom(forall({p, c}, implies(parent_of(p, c),
        not_(parent_of(c, p))));

    person_e max{o, "Max Mustermann"};
    person_e erika{o, "Erika Mustermann"};
    o.add_axiom(preds::parent_of(max, erika));

    soop::variable<'s'> s;
    const auto& parent = std::get<0>(
        o.request_entities<person_e>(
            exists({c}, parent_of(s, c)), s));
```

```

        std::cout << parent->name << " is a parent.\n";
    }

```

4 Exemplary Use Case: Conference Planner

In order to test the usability and practicality of SOOP in a realistic scenario, we created a small conference planner. The motivation for this particular use case is that the problem is easy to understand, realistic and NP-complete.

The basic model that we choose is the following: There are rooms, timeslots, speakers and talks. Every speaker may participate in more than one talk and every talk may have multiple speakers. Obviously, no two talks can be in a single room at the same time and no speaker can hold more than one talk in the same timeslot.

First, we create four C++ classes `talk`, `speaker`, `room` and `slot` that all inherit from `entity`. All of these contain data represented in the usual C++ manner, i.e. without any relation to SOOP. Most of this data does not have a semantic meaning in the use case, but is still important for the final software, such as a description of the talks or a name for the speakers. Only selected fields were given semantics by connecting them to the ontology using SOOP.

This is accomplished by two custom predicates:

```

SOOP_MAKE_TYPECHECKED_PREDICATE(is_speaker_of, 2,
                                speaker, talk);
SOOP_MAKE_TYPECHECKED_PREDICATE(talk_assignment, 4,
                                talk, speaker, room, slot);

```

A `make_ontology` function uses these classes and predicates to create an ontology with all the required types and axioms. In order to define the axioms, one function exists for each axiom, which is returned as a formula. One typical example of these functions is shown in the following. It defines that every talk must be held in one room and one slot and that it must be the only talk in that room during that slot:

```

soop::formula uniqueness_of_talks() {
    return forall({t1,t2,s1,s2,r1,r2,s11,s12},
        implies(
            and_(
                talk_assignment(t1,s1,r1,s11),
                talk_assignment(t2,s2,r2,s12)),
            equal(
                equal(t1,t2),
                and_(equal(r1,r2), equal(s11,s12))))));
}

```

After parsing the data about talks, speakers, rooms and slots from a plain text file, the data is stored in containers and axioms are associated to it: Which speakers participate in which talks and all talks are distinct from each other (instead of just different objects referring to the same entity):

```

auto o = make_ontology();
auto data = read_dataset(file, o);

for (const auto& talk : data.talks) {
    for (const auto speaker_id: talk->speaker_ids()) {
        o.add_axiom(preds::is_speaker_of(
            data.speakers.at(speaker_id), talk));
    }
}

o.add_axiom(soop::preds::distinct_range(
    data.talks.begin(), data.talks.end()));

```

Given this initialisation of the ontology, it is possible to request an assignment for all talks to rooms and time slots. The prover ensures that all constraints - as specified by the axioms - hold. Otherwise the request is not satisfiable and an exception is thrown. It is noteworthy that the returned answers contain references to the actual C++ data with the correct types and can thus be used as all other C++ objects:

```

for (const auto& talk : data.talks) {
    const auto& speaker = data.speakers.at(
        talk->speaker_ids().front());

    auto solution = o.request_entities<room, slot>(
        talk_assignment(talk, speaker, r, sl), r, sl);
    const auto& used_room = std::get<0>(solution);
    const auto& used_slot = std::get<1>(solution);

    std::cout << talk->title()
        << ": in room #" << used_room->number()
        << ", in slot #" << used_slot->time() << '\n';

    o.add_axiom(talk_assignment(talk, speaker,
        used_room, used_slot));
}

```

In the above example code, the assignments are requested sequentially. Thus, it is important to add the selected assignment back to the ontology as axioms. Otherwise, further requests may return assignments that conflict with earlier requested ones.

5 Evaluation and Discussion

While some details and most of the axioms were left out, the above provides an impression of the look and feel when using SOOP (as implemented in modern C++). In principal, there are two alternatives to SOOP with very different properties, which is why this evaluation will be split into two parts: The comparison to alternative uses of ontologies and the comparison to approaches that don't use ontologies at all.

With regards to the first point of comparison, SOOP integrates ontologies and regular OOP languages much tighter than existing approaches without requiring idioms that seem out of place for either. The expressive power of the provided description-logic is large enough to describe even complicated problems (such as the NP-complete timetable assignment problem), but still ensures that every statement can be represented in C++. Furthermore, SOOP never requires to compromise type safety or to deal explicitly with translating data from the ontology (usually strings) into actual OOP objects.

Excluding functionality such as parsing the input that are always necessary, creating a working software with SOOP for the use case required less than one hundred lines of code for the actual program logic. From our point of view, the effort is less or equal to the effort necessary with other ontology APIs and it is far less compared to the effort for a solution without using semantic technologies.

This brings us to the comparison with unaugmented OOP languages. The main work when working with SOOP is formalizing the problem, which is a step that is not strictly necessary otherwise. However, this seems a small price to pay, since formalization is always advisable when tackling complex problems, if a correct solution is desired. Using SOOP to formalize the problem has the added advantage that it is easier to spot incomplete formalizations as simple test cases may produce obviously wrong results. Since everything is performed within C++ test cases can be generated in the same manner and with the same tools as for normal unit testing.

With regards to the disadvantages, it is important to note that the theorem prover Z3 currently used in our proof of concept implementation was not designed for use cases such as SOOP. Thus, it can be expected that significant performance improvements are possible in both the prover interface and selection of a different reasoner. Similarly, the scalability of the SOOP approach is largely determined by the employed solver. In the current SOOP implementation, no constraint is imposed on the expressiveness of the axioms in first-order predicate calculus. However, even rather constrained domain logics are able to represent the minimal required expressiveness for OOP relations (cf. Tab. 2).

Another restriction on the usability of SOOP is that formalizing problems can be surprisingly difficult to do correctly, which means that an imperative implementation may be easier for many straightforward problems. However, the later argument applies to ontologies and semantic technologies in general, which often show their full potential only in domains of a certain minimum level of complexity.

One advantage of SOOP certainly lies in the fact that it enables to easily mix imperative solutions with ontological reasoning. Thus, different aspects of the application at hand can be solved in the domain (imperative OOP or declarative semantic reasoning), which is most suitable - or at least feels that way to the application programmer. This also fits nicely with C++ being a multiparadigm language.

Furthermore, libraries using SOOP present a shared conceptualization on two levels: From an OOP perspective, the SOOP classes can be connected to user defined types by inheritance and composition and thus reused just as non-SOOP libraries. At the same time, due to the connection between the OOP types and an ontology in SOOP, different libraries derived from the same base library are also compatible on a semantic level. Thus, a “foundational” SOOP library, directly derived from an upper ontology, has the potential to render, thus far incompatible, OOP libraries at least interoperable.

A much more in depth explanation of the C++ implementation and an extended discussion can be found in [We16].⁹

6 Directions of Further Research

During the course of the presented work, we created a working research prototype in C++ and solved several conceptual and implementation issues, but there are still several open topics and a lot of room for further research into SOOP:

The biggest potential for improvement appears to be in the area of static typing: While we implemented a method to state type requirements for predicates, variables are currently still largely untyped. This leads to all the dangers of untyped languages and also means that semantic information available in the OOP type system might not be provided to the reasoner.

Related to the topic of type information is the lack of checks for shadowing variables inside a formula, which results in unexpected answers by the reasoner. All the required information should be available at compile-time and we are convinced that these checks could be performed algorithmically.

Another interesting open question is how to approach copies of entities. In general there are three possible semantic meanings:

- Copy and original are independent of each other and the copy inherits no properties from the original.
- Copy and original are the *same* entity and always share all semantics (flat copy).
- Copy and original are different entities with the same semantics, that are copied as well when creating the copy (deep copy).

⁹ The PDF manuscript (German) is available at <http://florianjw.de/ba/soop.pdf>

For each behavior there are cases where this particular behavior is desirable and others where it isn't. For example, the elements in a buffer may even switch desirable semantics during the run of a program. To avoid picking a wrong default, we disabled copying in the current implementation, but want to point out that the architecture supports all of these three possibilities.

References

- [Gr92] Gruber, Thomas R.: A Translation Approach to Portable Ontology Specifications. 1992.
- [HB11] Horridge, Matthew; Bechhofer, Sean: The OWL API: A Java API for OWL Ontologies. *Semant. web*, 2(1):11–21, January 2011.
- [Je13] Jena, Apache: Apache jena. jena.apache.org [Online]. Available: <http://jena.apache.org> [Accessed: Mar. 20, 2014], 2013.
- [KAH05] Koide, Seiji; Aasman, Jans; Haflich, Steve: OWL vs. object oriented programming. In: the 4th International Semantic Web Conference (ISWC 2005), Workshop on Semantic Web Enabled Software Engineering (SWESE). Citeseer, 2005.
- [Kn04] Knublauch, Holger; Fergerson, Ray W.; Noy, Natalya F.; Musen, Mark A.: The Protg OWL plugin: An open development environment for semantic web applications. Springer, pp. 229–243, 2004.
- [Or07] Oren, Eyal; Delbru, Renaud; Gerke, Sebastian; Haller, Armin; Decker, Stefan: ActiveRDF: Object-oriented Semantic Web Programming. In: Proceedings of the 16th International Conference on World Wide Web. WWW '07, ACM, New York, NY, USA, pp. 817–824, 2007.
- [Pu08] Puleston, Colin; Parsia, Bijan; Cunningham, James; Rector, Alan: The Semantic Web - ISWC 2008: 7th International Semantic Web Conference, ISWC 2008, Karlsruhe, Germany, October 26-30, 2008. Proceedings. Springer Berlin Heidelberg, Berlin, Heidelberg, chapter Integrating Object-Oriented and Ontological Representations: A Case Study in Java and OWL, pp. 130–145, 2008.
- [SBF98] Studer, Rudi; Benjamins, V.Richard; Fensel, Dieter: Knowledge engineering: Principles and methods. *Data and Knowledge Engineering*, 25(1-2):161 – 197, 1998.
- [We16] Weber, Florian: Semantische Objektorientierte Programmierung. Bachelor thesis, Karlsruhe Institute of Technology, Germany, 2016. <http://florianjw.de/ba/soop.pdf>.

In-Store Customer Analytics – Metriken & Reifegradszenarien zur Erfassung physischer Kundenkontakte im stationären Einzelhandel

Lars Bollweg¹, Richard Lackes², Markus Siepermann³ und Peter Weber⁴

Abstract: Im Gegensatz zu der im Online-Handel etablierten Erfassung und Auswertung der Kundenkontaktdaten über den Browser (Web Analytics) vernachlässigt der stationäre Handel die Analyse der physischen Kundenkontakte im Ladenlokal (In-Store Customer Analytics) bisher weitgehend und fokussiert sich stattdessen auf die Auswertung von Transaktionsdaten. Dabei hat sich gerade die Erfassung und Auswertung der transaktionsvorgelagerten Kundenkontaktdaten zu einem zentralen Instrument der systematischen Weiterentwicklung und Verbesserung des Online-Handels entwickelt, so z. B. zur Landingpage- und Conversion-Optimierung. Die Analyse von Kundenkontaktdaten lässt entsprechend auch große Potenziale für den stationären Handel vermuten. Der vorliegende Beitrag systematisiert vor diesem Hintergrund auf Basis einer Literaturanalyse die mess- und übertragbaren In-Store Customer-Metriken und setzt diese in einer Kundenerfassungs-Reifegrad-Matrix in Bezug zu den für die Erfassung zur Verfügung stehenden Technologien.

Keywords: In-Store Customer Analytics, Metriken, Kundenerfassungs-Reifegrad-Matrix

1 Einleitung

Im Gegensatz zum stationären Handel ist die Erfassung und Auswertung von Kundenkontaktdaten im Online-Handel (Web Analytics) etablierte Praxis [DFK02]. Die online erhobenen Kundenkontaktdaten werden mit den Transaktionsdaten im Shopsystem abgeglichen und so z.B. die erzielten Verkäufe (Transaktionen) mit den potenziellen Verkäufen (Besucher auf der Webseite) in Beziehung gesetzt (sog. Conversion Rate). Diese Web-Analytics verfolgen das Ziel, durch eine datenbasierte und damit bedarfsgerechte Anpassung der Seitenstruktur, des Designs, des Angebots, der Preisstrategien usw. die Anzahl der Verkäufe zu erhöhen, und gehören faktisch zum Tagesgeschäft in jedem professionell betriebenen Online-Handel [DML11]. Im Gegensatz dazu sind die Kundenkontakte im Ladenlokal für den stationären Einzelhandel in weiten Teilen eine Black-Box [Ge14]. Auch wenn bereits seit Jahrzehnten Strategien zur Analyse der Kundenkontakte im Handel entwickelt und untersucht werden (z.B. über manuelle Zählungen, Testkunden und Eye-Trackingverfahren oder auch Modellierungsansätze wie das Blueprinting) [Gr68] [BD95] [NJ99] [Bi10], hat sich eine (teil-)automatisierte Erfassung

¹ Fachhochschule Südwestfalen, Lübecker Ring 2, 59494 Soest, bollweg.lars@fh-swf.de

² TU Dortmund, Otto-Hahn-Str. 12, 44227 Dortmund, richard.lackes@tu-dortmund.de

³ TU Dortmund, Otto-Hahn-Str. 12, 44227 Dortmund, markus.siepermann@tu-dortmund.de

⁴ Fachhochschule Südwestfalen, Lübecker Ring 2, 59494 Soest, weber.peter@fh-swf.de

und Analyse von Kundenkontaktdaten (In-Store Customer Analytics) im stationären Handel bisher nicht durchsetzen können. Ganz im Gegenteil, Händler und Forschung konzentrieren sich mit Bezug zum stationären Handel primär auf die reine Auswertung von Transaktionsdaten und vernachlässigen damit die Möglichkeit, über die Einbeziehung von Kundenkontaktdaten ihren Analyse-Horizont voll auszuschöpfen [BK12] [Su15]. Dies mag auf der einen Seite daran liegen, dass wir uns in einer Phase des Übergangs befinden, in der umfassende Lösungen noch kostenintensiv, fehleranfällig und bisher nur schwer in die bestehenden Infrastrukturen zu integrieren sind [GC05]. Auf der anderen Seite aber stehen bereits eine Vielzahl von Technologien mit unterschiedlichen Reifegraden zur Erhebung und Auswertung von Kundenkontaktdaten im Ladenlokal zur Verfügung. Es fehlt jedoch derzeit noch an einer strukturierten, technologieübergreifenden Auseinandersetzung mit den Grundlagen der Kundenkontaktdatenerfassung und -analyse. Vor diesem Hintergrund widmet sich der vorliegende Beitrag mit Hilfe einer strukturierten Literaturliteraturanalyse der Aufarbeitung der Grundlagen zu Fragestellungen der In-Store Customer Analytics und der Beantwortung der folgenden Forschungsfragen:

- RQ1:** *Welche bestehenden Technologien zur Kundenkontaktdatenerfassung im Ladenlokal stehen dem Handel zur Verfügung?*
- RQ2:** *Welche In-Store Customer Metriken können im Ladenlokal erfasst werden?*
- RQ3:** *Welche Erfassungstechnologien können welche In-Store Customer Analytics Metriken erfassen?*

Dieser Beitrag betrachtet die Erfassungstechnologien autonom und konzentriert sich auf die von Kundenkontakten ableitbaren Metriken. Verknüpfungen zu Transaktionsdaten werden bewusst nicht betrachtet, um den Fokus auf die Kundendatenerfassung zu erhalten. Der Beitrag gliedert sich wie folgt: In Abschnitt 2 wird in einem ersten Teil die Strukturierung der Web Analytics-Metriken (nach der Web Analytics Association [We16]) diskutiert, um im weiteren Verlauf des Beitrages eine den identifizierten Kategorien angepasste Sammlung von In-Store Customer Analytics Metriken abzuleiten. Im zweiten Teil wird mit Hilfe einer strukturierten Literaturliteraturanalyse der aktuelle Stand der Forschung zur Kundenkontaktdatenerfassung erhoben und die bereits im Einsatz befindlichen Technologien, sowie die damit messbaren In-Store Customer Metriken werden kategorisiert und erfasst. In Abschnitt 3 werden die erhobenen Erfassungstechnologien darauf aufbauend unter Berücksichtigung ihrer individuellen technologischen Entwicklung strukturiert und in Bezug auf ihre Leistungsfähigkeit diskutiert. In Abschnitt 4 werden die erhobenen In-Store Customer Analytics Metriken zur Komplexitätsreduktion auf eine Sammlung von Kernmetriken reduziert, um diese in Abschnitt 5 in einer Kundenerfassungs-Reifegrad-Matrix zusammenzuführen und damit einen strukturierten Überblick über die verfügbaren Erfassungstechnologien und die über sie erfassbaren Metriken zu bieten. In Abschnitt 6 werden die resultierenden Implikationen für die Praxis diskutiert, bevor der Beitrag in Abschnitt 7 mit einem Fazit und einem Ausblick abgeschlossen wird.

2 Literaturanalyse

Die hier durchgeführte Literaturanalyse gliedert sich in zwei Teile. Der erste Teil widmet sich den in der Web-Analyse bereits etablierten Web Analytics-Metriken, um die Grundlagenkategorien für eine Sammlung an In-Store Customer Analytics-Metriken zu entwickeln. Der zweite Teil ist eine strukturierte Literaturanalyse, welche auf die Kategorien des ersten Teils aufbaut und unter deren Berücksichtigung Literatur zu den Erfassungstechnologien systematisiert.

2.1 Web Analytics-Metriken

Während die Kundendatenerfassung im Ladenlokal noch ein Schattendasein führt, ist die Web Analyse bereits hochgradig organisiert und die vorhandenen Metriken sind standardisiert und strukturiert [TB03] [Gu13]. Die dafür verantwortliche Organisation, die Web Analytics Association (WAA) [We16], veröffentlicht auf ihrer Homepage eine Sammlung an strukturierten Web-Metriken und Definitionen, welche als weltweiter Standard gelten. Die WAA unterscheidet zwei Metrikarten, „Count“, als direkt gemessene Zahl (wie z.B. die Anzahl der Webseiten Besucher), und „Ratio“, als abgeleitete Metrik, die sich aus anderen Werten errechnen lässt (wie z.B. durchschnittliche Dauer pro Ladenbesuch). Die 22 vorgegebenen Metriken selbst werden in sechs Kategorien unterteilt, von denen nicht alle auf In-Store Customer Analytics-Metriken übertragbar sind.

1. Building Block	2. Visit Characterization	3. Visitor Characterization	4. Engagement	5. Conversion	6. Miscellaneous
Page	Entry Page	New Visitor	Page Exit Ratio	Conversion	Hit
Page View	Landing Page	Returning Visitor	Single Page Visits	Conversion Rate	Impressions
Visits	Exit Page	Repeat Visitor	Bounce Rate		
Unique Visitor	Visit Duration	Visitor Referer	Page views per Visit		
Event	Referrer	Visits per Visitor			
	Session Referrer	Recency			
	Click-Through	Frequency			
	Click-Through-Rate				

Tabelle 1: Web Analytics-Kategorien und -Metriken [We16]

Die Kategorie 5 bezieht sich auf eine Verrechnung von Kundendaten und Transaktionsdaten und wird deshalb in der Betrachtung dieses Beitrages zurückgestellt. Die Metriken der Kategorie 6 sind in ihrer Definition nicht auf das Ladenlokal zu übertragen und fallen deshalb ebenfalls aus der Betrachtung. Überträgt man die verbleibenden vier Web Metrik-Kategorien, so lassen zwei In-Store Metrik-Kategorien ableiten:

1. Standort-Metriken der Kategorie „Visit Characterization“ und
2. Verhaltens-Metriken der Kategorie „Engagement“

Die Web Analytics Kategorien „Building Block“ und „Visitor Characterization“ zeigen darüber hinaus aber auf, dass die abgeleiteten Kategorien noch um eine Erkennungsgrad-

Dimension erweitert werden müssen, da sie sowohl individualisierte als auch identifizierte Standort- wie auch Verhaltensmetriken beinhalten. Für Metriken wie „New, Returning oder Repeat Visitor“ bedarf es entsprechend einer identifizierten Kundenerkennung, während Metriken wie „Visit“ schon mit einer individualisierten Erfassung erhoben werden können.

2.2 In-Store Customer Analytics-Metriken

Die folgende strukturierte Literaturanalyse mit dem Ziel der Identifikation von verfügbaren Technologien und Metriken zur Kundendatenerfassung im Ladenlokal, umfasst eine Suche von Journal- und Konferenzbeiträgen aus den Jahren 2000 bis 2016. Verwendet wurden jeweils einzeln die folgenden Suchwörter: Retail Analytics, Traffic Analytics, Operations Analytics, Behaviour Analytics, Customer Analytics, In-Store Analytics, Video-Analytics, In-Store Metrics, Shoppers Intelligence, Path Analysis, Traffic Analysis und Costumer Counting. Gesucht wurde in den Datenbanken von EbscoHost, IEEE Xplore, Springer Link und Google Scholar. Da Google Scholar sich mit Einträgen aus den anderen genannten und untersuchten Datenbanken überschneidet, wurden die Doppelungen vorab aussortiert. Des Weiteren wurden nur Journal- und Konferenzbeiträge berücksichtigt, welche einen direkten Zugang per PDF Download ermöglichten. Die Suche ergab einen Korpus von 284 Beiträgen. Der gesamte Literaturkorpus wurde durch eine Analyse der Titel und Abstracts auf 42 relevante Beiträge reduziert. Diese wurden vollständig betrachtet und es konnten abschließend 18 Beiträge ermittelt werden, welche die Basis für die hier durchgeführte Literaturanalyse bilden (Tabelle 3).

	EBSCOHost	IEEE Xplore	Springer Link	Google Scholar
Ergebnisse Total: 284	48	96	28	112
Nach Titel & Abstract: 42	12	14	7	9
Nach vollständiger Analyse: 18	1	6	4	7

Tabelle 2: Literatúrauswahl

Die technologieneutral durchgeführte Suche führte ausschließlich zu technologiebezogenen Veröffentlichungen, von denen ein Großteil die Informationsgewinnung für den stationären Handel, wenn überhaupt, dann nur als Nebenprodukt aufführt. Dies deutet darauf hin, dass sich die Forschung bisher noch nicht eingehend mit dem Thema In-Store Customer Analytics auseinandergesetzt hat. Die Betrachtung der diskutierten Technologien in der Literatur lässt auf zwei Grunderfassungskategorien schließen: 1) direkt messende Technologien, welche den Kunden und seine Handlungen direkt erfassen, und 2) Proxy-Technologien, welche den Kunden mit Hilfe einer Instanz erfassen (z.B. über Einkaufswagen, Smartphones, etc.).

Die Kundenerfassung beider Kategorien (direkt und proxy) kann dabei in drei Graden erfolgen: 1) nicht individualisiert (Kunde kann nicht verfolgt und wiedererkannt werden), 2) individualisiert (Kunde kann verfolgt und wiedererkannt werden) und 3) identifiziert (Kunde kann verfolgt, wiedererkannt und eindeutig identifiziert werden).

Nr.	Author	Technologie	Alternativen	Metrik Kategorien	Kundenerkennung	In-Store Metrics
1.	Sorensen (2003)	RFID	RTLS	Multiple Locations	Proxy Individualized	Density, Speed of Purchasing, Quadrants, Speed of Shopping
2.	Hong et al. (2004)	WIFI	-	Multiple Locations	Proxy Individualized	-
3.	Li-Qun Xu (2007)	Video-Analysis (CCTV)	Optical Sensors	Multiple Locations, Action Event	Individualized	Crowd Density, Customer Counting
4.	Senior et al. (2007)	Video-Analysis (CCTV)	Infrared-Beams, Pressure Pads	Multiple Locations, Action Event	Individualized, Identified	Customer-Counting
5.	Bolliger (2008)	GSM, Bluetooth, WIFI	-	Multiple Locations	Proxy Individualized, Identified	-
6.	Yada (2009)	RFID	-	Multiple Locations	Proxy Individualized	Shopping Time, Staying Time in Sales Areas,
7.	Bourimi et al. (2011)	GSM, Bluetooth, WIFI	RFID, Indoor GPS, CCTV, Photo Sensors	Multiple Location	Proxy Individualized, Identified	-
8.	Blecker et al. (2011)	RFID	CCTV,	Multiple Locations, Action Event	Proxy Individualized	Visited Product Zones, Contact Instances, Physical Movement of Goods
9.	Rai et al. (2011)	Video-Analysis (CCTV)	WiFi, RFID	Multiple Locations	Proxy Individualized	Shopping Time, Staying Time in Sales Areas,
10.	Takai et al. (2012)	RFID	-	Multiple Locations	Proxy Individualized	Shopping Time, Staying Time in Sales Areas,
11.	Cai (2014)	WIFI	-	Multiple Locations	Proxy Individualized, Identified	-
12.	Conell et al. (2013)	Video-Analysis (CCTV)	-	Multiple Locations, Interaction Event	Individualized	People Counting, Conversion Rate, Buying Time and Staying Time, Cart Localization, Basket Size, Line Counting
13.	Rallapalli et al. (2014)	Smart Glasses	CCTV, WiFi	Multiple Locations, Interaction Event	Proxy Identified	Shopping Behavior: Walking, Dwelling, Gazing, Reaching out
14.	Yaeli et al. (2014)	WIFI	GPS, RFID, WiFi, Bluetooth	Multiple Locations	Proxy Identified	Store Zone, Store Visit, Zone Transition, Zone Visit Time, Store Visit, Unique / Repeat Customer, Store Exit Time, Store Enter Time, Time per Zone, Visitors to Store,
15.	Zeng et al. (2015)	WIFI	-	Multiple Locations	Proxy Identified	Walking Fast / Slow, Staying Time
16.	Deva et al. (2015)	WIFI	-	Multiple Locations	Proxy Identified	Returning Visitors, Visit Frequency, Visited Zones, Buying and Staying Time
17.	Pierdicca et al. (2015)	Beacon Technology	-	Multiple Locations	Proxy Individualized	Total Number of People, Avg. Visiting Time, People Passing by, Avg. Group Number, Interactions
18.	Liciotti et al. (2015)	Video – Analysis (CCTV)	-	Multiple Locations, Interaction Event	Individualized	Visitors, Visitors of a Zone, Interaction with Shelf / with Person / with Products, Duration of Interactions, Avg. Interaction Time

Tabelle 3: Literaturanalyse

Neben dem „Wie“ der Messung kann auch der Gegenstand der Messung, also das „Was“, als Gliederungskriterium (s.g. Metriken-Kategorie) für die zu vergleichenden Technologien herangezogen werden. Eine erste Gruppe von Technologien ist darauf ausgelegt, Standortdaten (Standort-Metriken) zu erfassen, was mit einzelnen (Single Location) oder mehreren (Multiple Locations) Messpunkten möglich ist. Während die „Single Location-Erfassung“ nur einem An-Aus Signal gleicht, sind mit der Erhebung von „Multiple Locations“ komplexere Auswertungen (z.B. Pfad-Analysen, Heatmaps) möglich. Die zweite Gruppe von Technologien befasst sich mit der Erhebung des physischen, über die reine Bewegung auf der Fläche hinausgehenden Kundenverhaltens (Verhaltens-Metriken). Auch diese Verhaltens erfassung lässt sich wiederum in zwei Unterkategorien gliedern: Die Messung von 1) individuellen Handlungen (Actions), wie z.B. Laufen, Warten, Betrachten, Greifen und 2) die Messung von Interaktionen (Interactions), wie z.B. von Gesprächen mit dem Personal (Interaktion mit Personal) oder das Heben von Produkten (Interaktion mit Produkten). Zur Auswertung und Interpretation der genannten Metrik-Kategorien bedarf es einer definierten Sammlung spezifischer Metriken, welche in Abschnitt 4 abgeleitet und diskutiert wird.

3 Technologien zur Erfassung physischer Kundenkontakte

Zur Erfassung physischer Kundenkontakte steht dem Handel mittlerweile eine Vielzahl an Technologien zur Verfügung. In der durchgeführten Literaturanalyse wurden acht Alternativen identifiziert, die Bandbreite reicht dabei von einfachen Lösungen wie Infrarot-Lichtschranken bis zu komplexen Lösungen wie Video-Systemen. In diesem Beitrag werden die Technologien ausschließlich autonom untersucht und mögliche Potenziale aus einer Vernetzung und die Gestaltung von Hybrid-Lösungen lediglich im Ausblick angesprochen. In der abgebildeten Tabelle 4 wird der jeweilige Leistungsumfang unter Berücksichtigung der Art der Messung (direkt / proxy), der erheblichen Metrik-Kategorien und des Kundenerkennungsgrades zusammenfassend dargestellt. Im weiteren Verlauf werden die aufgeführten Technologien kurz auf Vor- und Nachteile hin diskutiert.

		Metrik-Kategorien					
		Messung	Standort-Metriken		Verhaltens-Metriken		Kundenerkennung
			Single Location	Multiple Location	Action	Interaction	
Sensorik	Lichtschranken	Direkt	Ja	Nein	Nein	Nein	Nicht Individualisiert
	Druckplatten	Direkt	Ja	Nein	Nein	Nein	Nicht Individualisiert
	Optische Sensoren	Direkt	Ja	Ja	Nein	Nein	Individualisiert
Proxy-Technologien	RFID	Proxy	Ja	Ja	Ja	Nein	Identifiziert
	Beacon Technologie	Proxy	Ja	Ja	Ja	Nein	Identifiziert
	WIFI, GSM, Bluetooth	Proxy	Ja	Ja	Ja	Nein	Identifiziert
	Smart Glasses	Proxy	Ja	Ja	Ja	Ja	Identifiziert
Tracking-systeme	Video-Systeme	Direkt	Ja	Ja	Ja	Ja	Identifiziert

Tabelle 4: Überblick In-Store Customer Analytics Erfassungstechnologien

3.1 Rudimentäre Sensoren

Der Einsatz von Lichtschranken und Druckplatten zur Erfassung von Kundenkontaktdaten stellt die rudimentärste Form der Sensorik zur Kundenerfassung dar. Vorteile: Niedrige Kosten. Nachteile: Ausgelöst durch eine Unterbrechung des Lichtstrahls oder durch ein wahrgenommenes Gewicht geben diese Lösungen einen unreflektierten Zählwert aus, der durch seine Einfachheit einer hohen Ungenauigkeit unterliegt. Gruppen, Angestellte, Dienstleister, spielende Kinder im Eingangs- und Ausgangsbereich sowie Tiere werden beispielsweise ungefiltert mitgezählt. Diese Technologien sind zudem nicht in der Lage, Personen individualisiert zu verfolgen. Lichtschranken und Druckplatten eignen sich daher eher zur Trenderfassung, jedoch nicht zur konkreten Zählung von Kunden [Se07].

3.2 Optische Sensoren

Die Bandbreite der unter „optische Sensoren“ zusammengefassten Technologien reicht von Bewegungssensoren bis hin zu Wärmebildkameras. Vorteile: Die Technologien sind in höheren Entwicklungsstufen in der Lage, Personen im Ladenlokal individualisiert zu verfolgen und Analysen über deren Bewegungsabläufe zu generieren (z.B. Pfad-Analysen, Heatmaps). Da Sensoren-Systeme nicht in der Lage sind, Personen eindeutig zu identifizieren, bestehen dabei kaum datenschutzrechtliche Probleme. Nachteile: Die Gruppenerfassung ist für diese Sensoren oft noch eine Herausforderung, und es ist zu beachten, dass die entsprechenden Systeme autonom und unabhängig vom Kunden agieren und dieser somit keine Möglichkeit hat, sich der Messung zu entziehen [Xu07].

3.3 Proxy-Technologien

Die Gruppe der Proxy-Technologien umfasst Messungen über Wifi Netzwerke, RFID Chips, den Einsatz von Smart Glasses oder Smartphones mit Indoor GPS. Der jeweilige Leistungsumfang der einzelnen Technologien ist vielfältig [Bo09]. Vorteile: Über die Möglichkeit, Kunden eindeutig zu identifizieren, bieten sich durch die Vernetzungsmöglichkeiten z.B. mit Transaktions- und weiteren Kundendaten Möglichkeiten zur Personalisierung des Einkaufserlebnisses und zur Entwicklung digitaler Services, die auf ein erkanntes Kundenverhalten direkt reagieren können. Nachteile: Es muss berücksichtigt werden, dass das jeweilige Kundenverhalten nicht direkt, sondern lediglich indirekt über Proxys als Repräsentation des Kunden gemessen wird. Entsprechend kann es zu Messlücken kommen, wenn der Kunde den Erfassungsraum des Proxys verlässt, sich z.B. vom RFID-Einkaufswagen entfernt, sein Smartphone nicht dabei oder ausgeschaltet hat [So03] [Ca14] [Ya14].

3.4 Trackingsysteme

Video-Systeme stellen die komplexeste Gruppe an Technologien zur Erfassung von Kundenkontaktdaten im Ladenlokal dar. Getrieben durch die Entwicklungen in der Überwachungsindustrie [Xu07] sind Videosysteme heute in der Lage, Kunden zu zählen, ihre Pfade im Ladenlokal auszuwerten und ihre Handlungen (Actions) und Interaktionen (Interactions) zu erfassen. Vorteile: Sie können Gruppen analysieren und relevante von irrelevanten Messgegenständen unterscheiden (z.B. Menschen von Tieren). Gesichtserkennungssoftware, Gestik- und Mimikinterpretation führen zu einer zumindest potenziell erreichbaren, extrem hohen Datendichte, die sich immer stärker der Qualität einer direkten Beobachtung durch Personal angleicht [Li15]. Nachteil: Video-Systeme in der Kundenanalyse stehen unter datenschutzrechtlichen Gesichtspunkten in der Kritik [KPP12] [Co12].

4 In-Store Customer Analytics-Metriken

Um die bislang sehr technologiefokussierte Sicht auf die physische Kundenkontaktdatenerfassung mehr auf die für Händler bedeutende Informationsgewinnung durch Kundenkontaktdaten zu lenken, wird im Folgenden eine Sammlung an In-Store Customer Analytics-Metriken entwickelt. Eine Metrik lässt sich in vier Bausteine unterteilen: 1) die Kernmetrik, die Aufschluss über den Untersuchungsgegenstand gibt und somit den Hauptbestandteil darstellt. Die weiteren Bausteine werden als Präfix vor die Kernmetrik gesetzt, um diese weiter zu definieren. Da wären 2) das Individualisierungspräfix, welches Rückschlüsse auf identifizierte Kunden zulässt, 3) das Zählpräfix, welches die Art der Messung definiert (Anzahl, Häufigkeit, Zeitraum zur letzten Messung, Dauer) und dessen Einsatzmöglichkeit vom Metrik Typ abhängt (Count, Ratio oder Count / Ratio), sowie 4) das Verhältnispräfix, welches die gemessene Metrik in ein zeitliches Verhältnis zur Gesamterfassung setzt (Durschnitt pro Tag / Monat / Jahr) (siehe Tabelle 5). Eine Kernmetrik bedingt ein Zähl-Präfix, wohingegen Individualisierungs-Präfix und Verhältnis-Präfix optional sind. Im Folgenden werden zum Zwecke der Komplexitätsreduzierung nicht alle identifizierbaren In-Store Customer Analytics-Metriken aufgeführt, sondern nur die zugrundeliegenden Kernmetriken. Diese repräsentieren damit in Abhängigkeit des Metrik Typs jeweils weitere, ableitbare Metriken. Zur Ableitung kann man den in Tabelle 5 aufgeführten Metrik-Baukasten nutzen.

4*	3	2*	1
Verhältnispräfix	Zählpräfix (Metrik Typ)	Individual- präfix	Kernmetrik
Average / day	Number of (Count)	Unique	z.B. Visitor (Count)
Average / month	Frequency of (Count /Ratio)	New	
Average / year	Recency of (Count /Ratio)	Returning	
	Time per (Ratio)	Repeat	

Lese-Beispiel: Average (4) Number (3) of Unique (2) Visitors (1)

Tabelle 5: Metrik-Baukasten | * = optional

Die gesamte Sammlung umfasst 20 in der Literaturanalyse und den Web-Analytics Metriken identifizierte Kernmetriken, welche insgesamt 1248 Varianten der In-Store Customer Analytics Metriken repräsentieren. In Tabelle 6 sind die identifizierten Kernmetriken aufgeführt und den jeweiligen Metrik-Kategorien der verfügbaren Technologien und den Kundenerfassungskategorien zugeordnet. Im nächsten Kapitel wird eine Kundenerfassungs-Reifegrad-Matrix entwickelt, welche die bisher erarbeiteten Bausteine zusammenführt und einen Überblick über die verfügbaren Technologien und die von ihnen erfassbaren Kern-Metriken bietet.

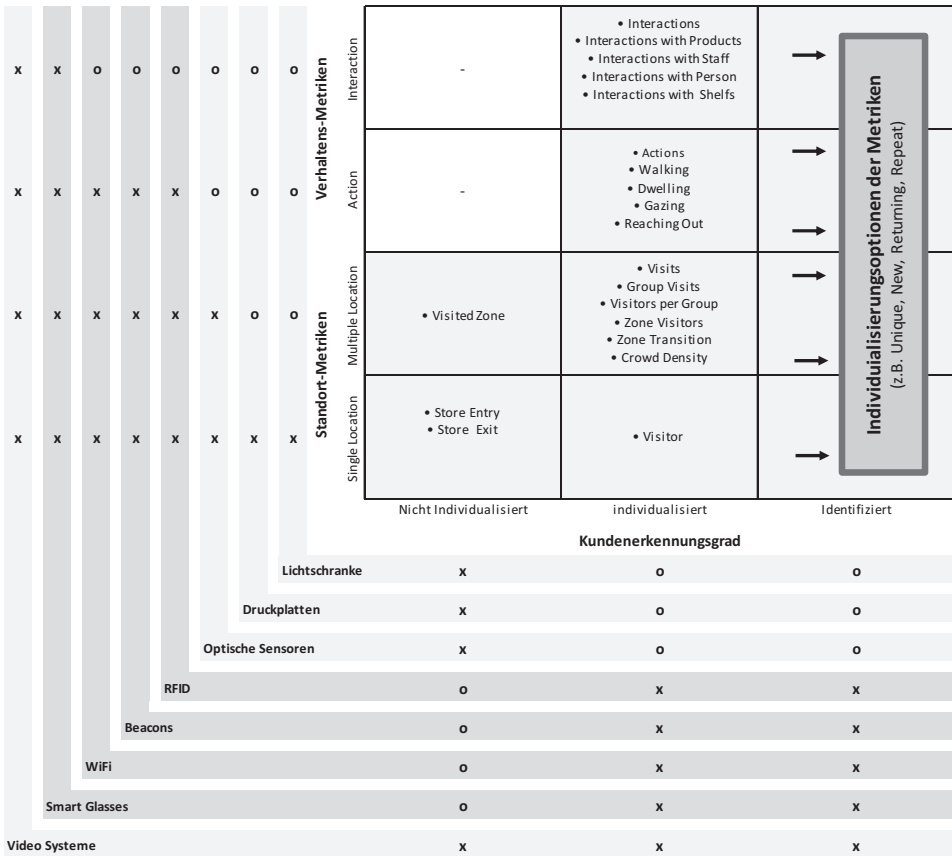
Nr.	Kernmetrik	Metriktyp	Single Location	Multiple Location	Action	Interaction	Kundenerfassung
1.	Store Entry	Count	x				Nicht Individualisiert
2.	Store Exit	Count	x				Nicht Individualisiert
3.	Visitor	Count	x				Individualisiert
4.	Visited Zone	Count / Ratio	o	x			Nicht Individualisiert
5.	Visits	Count / Ratio	o	x			Individualisiert
6.	Group Visits	Count / Ratio	o	x			Individualisiert
7.	Visitors per Group	Count	o	x			Individualisiert
8.	Zone Visitors	Count	o	x			Individualisiert
9.	Zone Transitions	Count	o	x			Individualisiert
10.	Crowd Density	Count	o	x			Individualisiert
11.	Actions	Count / Ratio	o	o	x		Individualisiert
12.	Walking	Count / Ratio	o	o	x		Individualisiert
13.	Dwelling	Count / Ratio	o	o	x		Individualisiert
14.	Gazing	Count / Ratio	o	o	x		Individualisiert
15.	Reaching out	Count / Ratio	o	o	x		Individualisiert
16.	Interactions	Count / Ratio	o	o	o	x	Individualisiert
17.	Interactions with Products	Count / Ratio	o	o	o	x	Individualisiert
18.	Interactions with Staff	Count / Ratio	o	o	o	x	Individualisiert
19.	Interaction with Person	Count / Ratio	o	o	o	x	Individualisiert
20.	Interactions with Shelves	Count / Ratio	o	o	o	x	Individualisiert

Tabelle 6: In-Store Customer Analytics Kernmetriken

5 Kundenerfassungs-Reifegrad-Matrix

Stationäre Händler stehen vor einer Vielzahl an Herausforderungen. Eine besteht darin, mit dem Online-Handel Schritt zu halten. Dem wachsenden Vorsprung in der Erfassung von Online-Kundenkontaktdaten für die Optimierung des Angebots und der Onlineshop-Oberflächen wird dabei gegenwärtig noch wenig Beachtung geschenkt. Die folgende Kundenerfassungs-Reifegrad-Matrix zeigt zu den im Laufe des Beitrags entwickelten Standort- und Verhaltens-Metriken (unterteilt in Single und Multiple Location, Actions und Interactions – von rudimentärer Standorterfassung bis zur komplexen Verhaltens-erfassung) und Kundenerkennungsgraden (Nicht individualisiert, individualisiert und identifiziert) den jeweiligen Leistungsumfang der Erfassungstechnologien auf und erweitert diese Betrachtung mit einer Zuordnung der identifizierten Kernmetriken zu den möglichen Erfassungsräumen. Die Kernmetriken werden in der Matrix in die Erfassungsräume mit dem niedrigsten Grad der möglichen Erfassung eingetragen und stehen damit den höheren Reifegraden – im Sinne einer Vererbung – auch zur Verfügung. So wird z.B. für die Messung eines „Visitors“ minimal eine individuelle Ein-Punkt-Standortmessung benötigt. Ein komplexeres System, das mehrere Standorte erfassen kann und in der Lage

ist Kunden zu identifizieren, ermöglicht auch die Erfassung der Metrik „Visitor“. In der folgenden Grafik sind die Proxy-Technologien mit einem dunkleren Grauton gekennzeichnet. Ein „x“ bedeutet, dass die Technologie den entsprechenden Leistungsumfang besitzt. Ein „o“ schließt den entsprechenden Leistungsumfang aus. Die weißen Flächen in der Matrix weisen darauf hin, dass die Erfassung von verhaltensbezogenen Metriken (Actions, Interactions) erst bei einer individualisierten oder identifizierenden Kundenerkennung möglich wird. Nicht individualisierte Messungen z.B. mit Lichtschranken oder Druckplatten können nur Standort-Metriken erfassen und sind nicht in der Lage, physische Aktionen (z.B. greifen) oder Interaktionen (z.B. Gespräche mit dem Personal) von Kunden zu erfassen und eindeutig zu identifizieren.



Grafik 1: Kundenerfassungs-Reifegrad-Matrix

Stellt man nun den Informationsgewinn anstatt der Technologien in den Fokus der Betrachtung, wird sichtbar, dass der Großteil der identifizierten Kernmetriken von Systemen mit einer individuellen Kundenerkennung geleistet werden kann. Zudem zeigt sich,

dass die Systeme mit einer identifizierenden Kundenerkennung bei der reinen Kundendatenerfassung nur geringfügige Vorteile bieten. Dies ändert sich allerdings, sobald eine Vernetzung der Systeme mit weiteren Daten(-banken) vorgenommen wird.

6 Implikationen

Die Kundendatenerfassung ist in der digitalen Welt längst etablierte Praxis. Stationären Händlern bietet die Kundendatenerfassung ein breites Spektrum an Weiterentwicklungsmöglichkeiten, wie z.B. einen Informationsgewinn in Bezug auf das transaktionsvorgelagerte Kundenverhalten. Diese Informationen könnten Händler nutzen, um das Angebot im Ladenlokal gezielter auf die Nachfrage auszurichten, um die Einrichtung und die Produktpräsentation den tatsächlichen erfassten Laufwegen und dem erkennbaren Kundenverhalten anzupassen, oder um den Einfluss von Marketingaktivitäten auf die Anzahl der Besucher im Ladenlokal zu überprüfen. Darüber hinaus würden Daten für eine verbesserte Personalplanung, z.B. in Bezug auf Kapazitäten und Einsatzbereiche verfügbar [CM15]. Ein weiterer interessanter Aspekt ist der im Online-Handel zu beobachtende Zusammenhang, dass der über eine automatisierte Erfassung permanent gewonnene Strom an Kundenkontaktdaten auch ein kontinuierliches Hinterfragen, Nachjustieren und Weiterentwickeln der Angebote und Services nach sich zieht. Und obwohl die rudimentärsten Datenerfassungstechnologien mit Ungenauigkeiten und Messschwierigkeiten zu kämpfen haben, stellen auch sie bereits einen erheblichen Informationsgewinn in Aussicht. Eine Lichtschranke in Verbindung mit Transaktionsdaten wäre z.B. dazu in der Lage, eine Conversion-Rate auf Trendbasis der Besuche im Ladenlokal zu erstellen. Große Probleme bereiten fast allen Technologien jedoch Gruppen von Kunden, die gleichzeitig ins Geschäft kommen. Auch die problematische Differenzierung zwischen Kunden und Personal, zwischen Kindern und Erwachsenen sowie zwischen Menschen und Tieren muss im Hinblick auf die Messergebnisse berücksichtigt werden. Trotzdem bietet sich Händlern die Möglichkeit, über eine skalierte Einführung der Erfassungstechnologien in die Analyse von Kundenkontaktdaten hinauszuwachsen. Der volle Analyseumfang ist bis heute nur mit hochentwickelten Video-Analyse-Systemen möglich. Für die größte Anzahl der In-Store Customer Analytics Metriken reichen aber schon Systeme, die eine individualisierte Kundenerkennung bieten, wie z.B. optische Sensoren-Systeme, welche unter datenschutzrechtlichen Aspekten weniger problematisch sind.

7 Fazit und Ausblick

Mit Hilfe einer strukturierten Literaturanalyse wurden die formulierten Forschungsfragen wie folgt beantwortet: 1) Es konnten acht verfügbare Technologien zur Erfassung von Kundenkontaktdaten identifiziert werden. Lichtschranken und Druckplatten lassen sich als eine ungenaue, dafür aber einfache Lösung charakterisieren. Optische Sensoren-Lösungen bieten einen hohen Informationsgewinn, aber keine Möglichkeit zur Identifi-

zierung und damit zur Personalisierung des Einkaufserlebnisses. Proxy-Lösungen wie RFID, WiFi, Beacon-Technology, Smartphone-Lösungen und Smart Glasses bieten einen großen Anwendungsraum, sind aber als indirekte Messmethoden mit Messlücken verbunden. Videosysteme bieten den größten Leistungsumfang, sind aber auch gleichzeitig besonders komplex und gehen auch unter datenschutzrechtlichen Gesichtspunkten mit erheblichen Herausforderungen einher. 2) Es konnten 20 In-Store Customer Analytics Kernmetriken (Tabelle 6) identifiziert werden, von denen mit Hilfe eines Metrik-Baukasten (Tabelle 5) insgesamt 1248 In-Store Customer Analytics-Metriken abgeleitet werden konnten. 3) Mit der Kundenerfassungs-Reifegrad-Matrix wurden schließlich die im Laufe des Beitrags entwickelten Bausteine zusammengeführt und in eine Übersicht zum Leistungsumfang der bestehenden Technologien und zu der jeweils realisierbaren Erfassung von In-Store Customer Kernmetriken überführt. Dieser Überblick soll Händler dabei unterstützen, Implementierungsentscheidungen technologieunabhängig und mit Fokus auf die avisierten Informationsgewinne treffen zu können.

In diesem Beitrag wurde der Fokus auf die autonome Untersuchung der Erfassungstechnologien gelegt. Potenziale, die sich durch die Vernetzung mit Transaktionsdaten oder durch Hybrid-Lösungen (Verbindung von Sensoren mit Proxy-Lösungen) ergeben, wurden ausgeklammert. Hybrid-Lösungen haben das Potenzial, die angesprochenen Messlücken der Proxy-Technologien zu überwinden, um unter gleichzeitiger Berücksichtigung des Datenschutzes personalisierte Angebote zu entwickeln. Darüber hinaus ist die Vernetzung der Kundenkontaktdaten mit Transaktions- und Kundendaten äußerst interessant in Bezug auf eine mögliche Anreicherung des stationären, physischen Einkaufsverhaltens mit digitalen Services. Die Kundenkontaktdatenerfassung ist der Grundstock, auf dem In-Store Empfehlungs- und Werbesysteme aufsetzen können. Sie bietet die Möglichkeit, neben Standortdaten (Location-based Services) auch Rückschlüsse aus dem Kundenverhalten zu ziehen. [Fz15] Stationäre Maßnahmen zur Kundenkontaktdatenanalyse sind also nicht nur unter dem Gesichtspunkt eines Anschlusshaltens mit dem Online-Handel zu betrachten, sondern sie stellen auch neue Geschäfts- und Service-Modelle in Aussicht, die vom reinen Onlinehandel schwer zu kopieren wären. Hierzu müssten jedoch zunächst die Qualität der erhobenen Daten verbessert und zum Beispiel die oben genannten Stördaten sauber gefiltert werden. Unter Berücksichtigung von Datenschutzbestimmungen müssten Schnittstellen für Hybrid-Lösungen sowie für eine Vernetzung der Messdaten mit Transaktions- und weiteren Kundendaten geschaffen werden. Es wäre zudem erforderlich, die im vorliegenden Beitrag identifizierten Metriken bzw. Metrik-Kategorien, um die Kategorie der Kontexterfassung zu erweitern. Die Erfassung von Kontextdaten erfordert eine integrierte Vernetzung aller zugänglichen Systeme, um über die Kombination von Verhaltens- und historischen Daten Informationen über den Kontext des Kundenverhaltens liefern zu können. Vor diesem Hintergrund soll in einem nächsten Schritt die Implementierung und Anwendung der entsprechenden Systeme in Praxisprojekten in den Vordergrund unserer Untersuchungen rücken, wobei der Fokus auf inhabergeführte stationäre Einzelhändler gelegt wird. Darauf aufbauend können in einem Folgeschritt die Potenziale einer Vernetzung mit Transaktions- und weiteren Kundendaten in einem Omni-Channel Retailing Ansatz in Augenschein genommen werden.

8 Literaturverzeichnis

- [BD95] Babin, B.J.; Darden, W.R.: Consumer self-regulation in a retail environment. *Journal of Retailing*, 71(1), S. 47-70, 1995.
- [Bi10] Bijmolt, T. H.; Leeﬂang, P. S.; Block, F.; Eisenbeiss, M.; Hardie, B.G.; Lemmens, A.; Saffert, P.: Analytics for customer engagement. *Journal of Service Research*, 13(3), S. 341-356, 2010.
- [BK12] Berman, S. J.; Kesterson-Townes, L.: Connecting with the digital customer of the future. *Strategy & Leadership*, 40(6), S. 29-35, 2012.
- [Bo08] Bolliger, P.: Redpin-adaptive, zero-configuration indoor localization through user collaboration. In *Proceedings of the first ACM international workshop on Mobile entity localization and tracking in GPS-less environments*, S. 55-60, ACM. 2008.
- [Bo09] Bolliger, P.; Partridge, K.; Chu, M.; Langheinrich, M.: Improving Location Fingerprinting through Motion Detection and Asynchronous Interval Labeling. In *Location and Context Awareness*, S. 37-51, 2009. Springer Berlin Heidelberg.
- [Bo11] Bourimi, M.; Mau, G.; Steinmann, S.; Klein, D.; Templin, S.; Kesdogan, D.; Schramm-Klein, H.: A Privacy-Respecting Indoor Localization Approach for Identifying Shopper Paths by Using End-Users Mobile Devices. In *Information Technology: New Generations (ITNG)*, Eighth International Conference, S. 139-144, IEEE. 2011.
- [BRT11] Blecker, T.; Rasch, C.; Teichert, T.: Prospects for PoS Market Research with RFID Technology: Examination of Consumers' In-Store Shopping Processes. In *European Retail Research*, S. 47-62, 2011 Gabler Verlag.
- [Ca14] Cai, D.: A retail application based on indoor location with grid estimations. In *Computer, Information and Telecommunication Systems (CITS)*, 2014 International Conference on, S. 1-4, IEEE. 2014.
- [Co12] Connell, J.; Fan, Q.; Gabbur, P.; Haas, N.; Pankanti, S.; Trinh, H.: Retail video analytics: an overview and survey. In *IS&T/SPIE Electronic Imaging*, S. 86630X-86630X, 2013. International Society for Optics and Photonics.
- [CM15] Chen, L.; Mersereau, A.J.: Analytics for Operational Visibility in the Retail Store: The Cases of Censored Demand and Inventory Record Inaccuracy. In *Retail Supply Chain Management*, S. 79-112, Springer US. 2015.
- [DFK02] Devaraj, S., Fan, M.; Kohli, R.: Antecedents of B2C Channel Satisfaction and Preference. *Information Systems Research*, 13(3), S. 316-333, 2002.
- [DML11] Davenport, T. H.; D'Amboise, L. D.; D'Amboise, J.: Know What Your Customers Want Before They Do. *Harvard Business review*, 89(12), S. 84-92, 2011.
- [DR15] Deva, B.; Ruppel, P.: Location Analytics as a Service: Providing Insights for Heterogeneous Spatiotemporal Data. In *Web Services (ICWS)*, 2015 IEEE International Conference on, S. 353-360, IEEE. 2015.
- [Fz15] Fang, Z.; et al.: Contemporaneous and delayed sales impact of location-based mobile promotions. *Information Systems Research* 26.3, S. 552-564, 2015.

- [GC05] Gagnon, J. L.; Chu, J.J.: Retail in 2010: a world of extremes. *Strategy & Leadership*, 33(5), S. 13-23, 2005.
- [Ge14] Germann, F.; Lilien, G. L.; Fiedler, L.; Kraus, M.: Do Retailers Benefit from Deploying Customer Analytics? *Journal of Retailing*, 90(4), S. 587-593, 2014.
- [Gr68] Granbois, D.H.: Improving the Study of Customer In-Store Behavior. *The Journal of Marketing*, S. 28-33, 1968.
- [Gu13] Gupta, R.; Mehta, K., Bhavsar, K.; Joshi, H.: Mobile web analytics. *International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE)*, 2(3), S. 288, 2013.
- [HL04] Hong, J. I.; Landay, J. A.: An architecture for privacy-sensitive ubiquitous computing. In *Proceedings of the 2nd international conference on Mobile systems, applications, and services*, S. 177-189, ACM. 2004.
- [KPP12] Kenny, R.; Pierce, J.; Pye, G.: Ethical considerations and guidelines in web analytics and digital marketing: a retail case study. In *AiCE 2012: Proceedings of the 6th Australian Institute of Computer Ethics conference 2012*, S. 5-12, 2012. Australian Institute of Computer Ethics.
- [Li15] Liciotti, D.; Contigiani, M.; Frontoni, E.; Mancini, A.; Zingaretti, P.; Placidi, V.: Shopper Analytics: a Customer Activity Recognition System Using a Distributed RGB-D Camera Network. *Video Analytics for Audience Measurement Lecture Notes in Computer Science*, 2015 arXiv preprint arXiv:1508.06853.
- [NJ99] Naumann, E.; Jackson, D.W.: One more time: how do you satisfy customers? *Business Horizons*, 42(3), S. 71-76, 1999.
- [Pi15] Pierdicca, R., et al.: Low cost embedded system for increasing retail environment intelligence. *Multimedia & Expo Workshops (ICMEW)*, 2015 IEEE International Conference on. IEEE, 2015.
- [Ra14] Rallapalli, S.; Ganesan, A.; Chintalapudi, K.; Padmanabhan, V. N.; Qiu, L.: Enabling physical analytics in retail stores using smart glasses. In *Proceedings of the 20th annual international conference on Mobile computing and networking*, S. 115-126, 2014.
- [RJK11] Rai, H. G.; Jonna, K.; Krishna, P. R.: Video analytics solution for tracking customer locations in retail shopping malls. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, S. 773-776, 2011. ACM.
- [Se07] Senior, A. W.; Brown, L.; Hampapur, A.; Shu, C. F.; Zhai, Y.; Feris, R. S.; Carlson, C.: Video analytics for Retail. *2007 IEEE Conference on Advanced Video and Signal Based Surveillance*, 2007.
- [So03] Sorensen, H.: The science of shopping. *Marketing Research*, 15(3), S. 30-35, 2003.
- [Su15] Suresh, S. (2015). Application of Retail Analytics Using Association Rule Mining In Data Mining Techniques With Respect To Retail Supermarket* A. Pappu Rajan.
- [TB03] Teltzrow, M.; Berendt, B.: Web-Usage-Based Success Metrics for Multi-Channel Businesses. In *Proc. of the WebKDD Workshop on Web Mining and Web Usage Analysis*, S. 17-27, 2003.

-
- [TY12] Takai, K.; Yada, K.: A framework for analysis of the effect of time on shopping behavior. *J Intell Inf Syst Journal of Intelligent Information Systems*, 2012.
- [We16] Web Analytics Association – Web Analytics Definitions, http://www.digitalanalyticsassociation.org/Files/PDF_standards/WebAnalyticsDefinitionsVol1.pdf, 29.04.2016.
- [Xu07] Xu, L. Q.: Issues in video analytics and surveillance systems: Research/prototyping vs. applications/user requirements. In *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on S. 10-14, 2007. IEEE*.
- [Ya09] Yada, K.: String analysis technique for shopping path in a supermarket. *J Intell Inf Syst Journal of Intelligent Information Systems*, 2009.
- [Ya14] Yaeli, A.; Bak, P.; Feigenblat, G.; Nadler, S.; Roitman, H.; Saadoun, G.; Sandbank, T.: Understanding customer behavior using indoor location analysis and visualization. *IBM Journal of Research and Development*, 58(5/6), S. 3-1, 2014.
- [ZPM15] Zeng, Y.; Pathak, P. H.; Mohapatra, P.: Analyzing Shopper's Behavior through WiFi Signals. In *Proceedings of the 2nd workshop on Workshop on Physical Analytics*, S. 13-18, ACM. 2015.

Secure Parallel Processing of Big Data Using Order-Preserving Encryption on Google BigQuery

Timo Schindler¹, Christoph Skornia²

Abstract: With the increase of centralization of resources in IT-infrastructure and the growing amount of cloud services, database management systems (DBMS) will be more and more outsourced to Infrastructure-as-a-Service (IaaS) providers. The outsourcing of entire databases, or the computation power for processing Big Data to an external provider also means that the provider has full access to the information contained in the database. In this article we propose a feasible solution with Order-Preserving Encryption (OPE) and further, state of the art, encryption methods to sort and process Big Data on external resources without exposing the unencrypted data to the IaaS provider. We also introduce a proof-of-concept client for Google BigQuery as example IaaS Provider.

Keywords: Order-Preserving Encryption, Big Data, Google BigQuery, Encryption, NoSQL

1 Introduction to Order Preserving Encryption

Cloud Computing has reached to be one of the cornerstones in IT-Infrastructure during the last years. Especially the outsourcing of databases is one main service and has been proposed in several publications. [Ag12, Am16, Br08, Cu11, DAEA13, Go16b, Mi16] The idea to use complex networking and computing infrastructure as a service is reasonable but has its limitations. [HIM02, Ha02] An external platform of data storage has to be treated as untrusted. Encryption is a powerful technology for protecting the confidentiality of the data stored but needs to be decrypted for processing. One approach is to use encryption which allows operations on encrypted data. Fully Homomorphic Encryption (FHE) and Order-Preserving Symmetric Encryption (OPE) are relevant approaches to solve this dilemma with encryption algorithms. As OPE maintains the order of the encrypted data obtained, data can be compared on the encrypted system and are thus sorted. Since the comparison of data is already sufficient to run a significant amount of common operations on the remote database system, this method fulfills two important prerequisites for the outsourcing of DBMS: Data can not only be securely stored, but also processed on a remote system. Because of this quality, OPE is primarily used in databases for processing SQL queries over encrypted data [PLZ13, Ag04, GZ07, Ha02, KAK10, LE09, LW12, LW13, Po11, LILD12]. The concept of OPE is subject of research since a number of years and secure algorithms were found. [PLZ13, Po11, Ma15] The challenge for researchers is to develop a feasible solution for relevant use cases of secure data processing. Some of the most recent aspects are considered and evaluated in this work.

¹ OTH Regensburg, Fakultät Informatik und Mathematik, Labor für Informationssicherheit, Prüfeninger Str. 58, 93049 Regensburg, timo.schindler@oth-regensburg.de

² OTH Regensburg, Fakultät Informatik und Mathematik, Labor für Informationssicherheit, Prüfeninger Str. 58, 93049 Regensburg, christoph.skornia@oth-regensburg.de

In the context of the outsourcing process of computing power and services to external, untrusted systems, Fully Homomorphic Encryption is an alternative to OPE. In contrast to OPE, FHE offers the advantage that the homomorphic property of the encrypted data is retained. Thus, an OPE scheme only guarantees that comparison's plaintext space ($x > y$) has the same result as those obtained in the encrypted space ($Enc(x) > Enc(y)$). In FHE, computation of more sequential operations on encrypted data are possible. Recent work on FHE has shown it is, in principal, possible the perform arbitrary computations over encrypted data [Ge09], the performance overheads are prohibitively high, on the order of 10^9 times [PLZ13, GHS12]. Both procedures have their specific use case and can be used together. In this work we focus on OPE as we are mainly interested in the option to sort data on the untrusted IaaS-Platform.

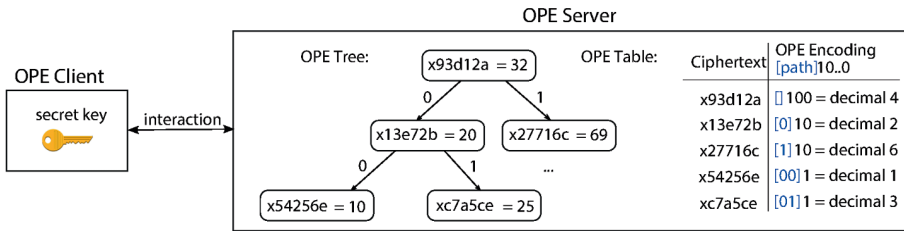


Fig. 1: Datastructure overview of mOPE [PLZ13]

Different OPE algorithms with specific characteristics are known [Ag04, Bo09, BCO11, PLZ13]. Mutable Order-Preserving Encoding (mOPE), a special form of OPE, is introduced by Ada Popa et. al. in [PLZ13]. With mOPE, the data on the server is encoded in a binary search tree (see fig. 1). The server provides the encrypted data stored in a binary tree and the OPE encoding path. The plaintext shown behind the encrypted cipher text (see fig. 1) is not known to the server and is provided with the purpose of the readers' information.

For our implementation we use AES, an existing deterministic encryption algorithm with a constant initialization vector. The encrypted data is represented in a binary tree, in such a way that the right child node is growing and the left one is decreasing. For all commercial IaaS NoSQL providers, it is not possible to implement own methods and algorithms for storing data. mOPE can be modified and used in a IaaS NoSQL scenario without losing security or ordering qualities. As proven in [PLZ13] the only information revealed of the encrypted data is the order and hence the needed minimum of additional exposure. Due to the high security and simplicity this algorithm has been chosen for the following application scenario. The objective of this work is to propose a way to use the secure OPE algorithm mOPE on a cloud platform. The adaption of the mOPE algorithm is not possible without changing the way, data will be encrypted and sent to the cloud database due to the fact that cloud databases works in a different manner. In this work we discuss different challenges and possible solutions processing Big Data in a fast way by many clients.

2 Using Order-Preserving Encryption on Big Data

There are different reasons for storing big amounts of data on potential untrusted databases, such as autonomous cars, intelligent homes or smart grids. However, Big Data is not restricted to new, expensive applications of big corporations, but Big Data can be acquired, also from small companies [TN14]. Big Data, being generally unstructured and heterogeneous, is extremely complex to deal with via traditional approaches, and requires real-time or almost real-time analysis. [YK15] In our scenario, Big Data needs to be accessed from many applications or users and be processed in a fast way. This is a widely used scenario for any kind of data. To work with Big Data also includes the use of NoSQL Databases, because of the major advantages in processing a high amount of data in a fast way. A NoSQL database system is a database without a relational data model. NoSQL is perfectly suited for processing Big Data, because of the horizontal scaling. Contrary to vertical scaling (e.g. SQL-Server), where a single node needs to be upgraded to get more computation power, horizontal scaling simply needs more nodes. [Ca11] This model is cost efficient and can be combined with an IaaS model. Thus, extremely large data can be processed in a fast way. Due to the use of programming models like MapReduce [DG08], which is used in several NoSQL Databases, the processing of Big Data is feasible possible. With OPE on a scalable cost efficient NoSQL resource, no expensive infrastructure is needed and the external resource is capable of a secure analysis of Big Data.

3 The Infrastructure of Mutable Order-Preserving Encryption on Google BigQuery

To order and encrypt data the plaintext has to be known to the system encrypting it. Because this work can not be outsourced to an untrusted service, we use a central proxy configuration. Figure 2 depicts our concept for order preserving encryption in an IaaS infrastructure. The concept of encryption includes an encryption proxy, an NoSQL-like IaaS and n clients. The encryption proxy gets a Big Data chunk as input. The proxy orders and encrypts the data by using private keys. In this concept, it is also possible to use different encryption algorithms and keys for different columns of the data. The encryption proxy in between the Big Data and the IaaS provider is necessary, so no additional information besides the order will be revealed to any external resources. The keys, used to encrypt and decrypt the data, will be kept secret. The sorted and encrypted data will then be sent to the external service. Once the data is uploaded, many clients can perform analyses on it. In general, there is nearly no limitation on the clients or the complexity of the queries, because of the horizontal scaling of NoSQL Databases. The clients themselves can have different privileges. In the example configuration in figure 2, the client can have access to the secret keys. This is not necessary to perform queries on the data. If they do have the secret keys of the encryption proxy, the full range of analysis is possible. It may be possible that encrypted and unencrypted data are present in a single dataset. A client who does not possess the secret keys, can still perform analysis on unencrypted data. This scenario is conceivable, if some of the data in a dataset is public or available unencrypted. The unencrypted data can then complement the encrypted data by providing additional information

and might be useful to perform other analyses. This concept allows several other security models and it is also possible to extend the concept for using different keys for different clients. With this concept, we have created a way to use the powerful computation power of the cloud-service Google BigQuery by many clients but not revealing information about the data, which will be stored in the database. Furthermore, due to the use of different keys, a highly granular rights management for the clients is also possible which can access different information without using or implementing any rights- management on the IaaS provider side.

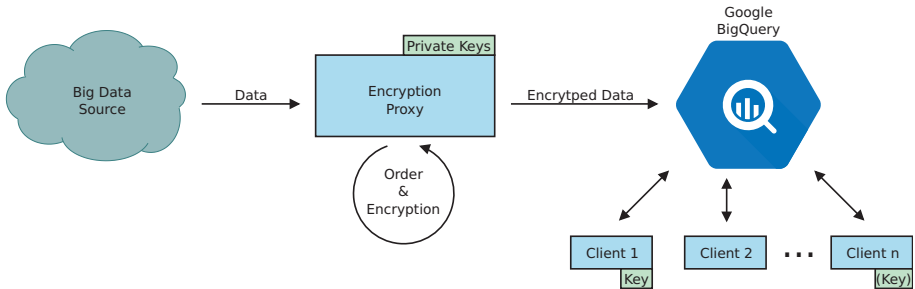


Fig. 2: Concept of Order-Preserving Encryption on Google BigData as example IaaS provider

3.1 Limitations on Order-Preserving Encryption for Big Data

Big Data usually implies that the amount of data cannot be processed by a single machine as the proxy in the described concept (see fig. 2). Also, the Big Data chunk will usually grow over time. This limitation does not only apply in this scenario, but is present every time, the order of encrypted data is needed. If new data will be added, the already encrypted and sorted data needs to be read from the external resource, decrypted, and sorted again. Outsourcing this operation to the external IaaS Provider does also reveal the private key to this provider. For security reasons, this approach is not possible. The concept we use offers the possibility to access data very fast because of the high power of the IaaS provider but still needs high computational power or time for encrypting the data on the encryption proxy.

3.2 Solving the Limitations

To solve the limitation on ordering and encrypting Big Data, we propose different approaches:

- **Using a high sorting range:** A resorting and reencryption is just needed if the plain text space size (domain) reaches a collision in the mapped sorting items (range). Then a new item cannot be sorted to the right position because of creating a collision. This can be prevented by choosing a much higher range-space. A resorting and reencryption is more unlikely.

- **Partial encryption:** In many cases, Big Data consists of a mixture of public and private data. Only when private data and public data are brought into relationship, the public data is interesting for a potential attacker. A distinction has to be drawn between data that should be kept secret and data, which is public available. This can lead to a major decrease of the encrypted data by simply using an intelligent encryption schema on the Big Data.
- **Use of diverse encryption algorithms:** One data chunk consists of various kinds of data, with different claims on encryption. In order to know the right algorithm, the plaintext data needs to be analysed. If, for example, sorting is irrelevant because a data chunk contains the name of patients, this data can be encrypted with an probabilistic algorithm. This procedure can be parallelised and will affect the resources of the encryption proxy less.
- **Separating data chunks on server and merging it on client side:** We proposed an import of the Big Data as chunks. Assuming the returned queries results are significant smaller than the particular data chunk, the uploaded data chunks can be uploaded in different tables. A query then will be applied to the different chunks and the client will merge the much smaller results locally. By using this approach it is not necessary to reencrypt already uploaded chunks. If the fragmentation grows over time, a garbage collection can merge data chunks, reencrypt those and upload it again to the service to lower the tables on server side.³

Additional approaches for solving the limitation can be the use of local NoSQL structures for the encryption and sorting proxy or the decreasing of the amount of data by just using the needed columns for OPE. [Sc16]

4 Google BigQuery as Example IaaS Provider

Google BigQuery is a system, designed to perform SQL-Like statements over Big Data. Google BigQuery can query $1 \frac{TB}{s}$ [TN14] and returns the result of the particular SQL statement. To achieve this, BigQuery uses different techniques of storing and processing data. They are based on several main systems, like BigTable, a forerunner of the NoSQL Database used at Google, or Megastore, a geographically replicated, consistent NoSQL-type datastore. Megastore uses the Paxos algorithm to ensure consistent reads and writes. [TN14]

The cloud service of Google utilises the Dremel Engine, a distributed SQL query engine, to perform complex queries over Big Data. [Me10] This engine uses two technologies to achieve the goal of $1 \frac{TB}{s}$: Colossus, a large, parallel, distributed file system and ColumnIO, a storage format which arranges the data in a manner that makes it easier to query over this data. [TN14]

The Dremel Engine makes use of the Dremel Serving Tree algorithm to run the query on a distributed system. [Me10] The possible parallelisation of SQL-like queries is depending

³ This solution is only adequate if Big Data chunks are uploaded and the fragmentation is not too high.

on the complexity of these queries. In most operations, the SQL query will be distributed to many workers (shards) and mixers. These different nodes will process the SQL query in parallel and are returning the result.

5 Encrypted BigQuery Command Line Interface

Google BigQuery offers a powerful asynchronous API to access the resources provided by Google. In addition, Google has released a beta client written in Python which is using encryption based on [Po11] called "Encrypted Big Query Command-Line Tool" (ebq). [Go16a]

To provide data for the CLI, the plaintext table data needs to be stored either as comma-separated values (CSV) or JavaScript Object Notation (JSON) file. The definition, which encryption should be used on which data, needs to be provided in a second scheme file. This extended table schema is provided by the user, where the extended schema is a modification of the BigQuery scheme. The unmodified ebq-client supports Paillier's homomorphic encryption algorithm [Pa99], as well as probabilistic, pseudonym and different searchword algorithms. [Sc16] To support Order-Preserving encryption, the client has been modified and a new encryption scheme is implemented. The CLI is used as encryption proxy and is sorting the data, using "timsort", a hybrid stable sorting algorithm, with a best case performance of $O(n)$, average and worst case performance of $O(n \log n)$. The plaintext will be encrypted using AES with CBC/PKCS5 Padding. The client uses a simple implementation as proof of concept. The encrypted data and the decimal order value will be written together to a column separated by a special character. In addition to for research purpose, the order is also written to a separate column.

All encrypted data is stored in a temporary file. If the data-encryption and -sorting is finished, the file will be uploaded by using the API of Google BigQuery. The private key will not be transmitted to the external source. BigQuery uses a RESTful API and the Transport-Layer- Security (TLS) protocol to communicate encrypted with the client. The packages sent to the API are lossless compressed with gzip using the deflate algorithm. By using the compression, the data transfer can be sped up by +50-75%, depending on the size and the quality of the data.

We have implemented Order-Preserving Encryption based on the mOPE Algorithm. Based on the modification, the ebq client is now capable of using a high range of different encryption algorithms including OPE for Big Data.

The in-depth analysis of the modified beta client in [Sc16] has been investigated on different components of the client to detect possible impediments of the encryption. We have reviewed the resource and time consumption of uploading, query processing, sorting and encrypting. The analysis has proven that the bottle neck for the encrypted BigQuery client is the encryption of new data before uploading it to the external service. Due to the compression, the upload time for new data is relatively low, compared to the time consumption of the encryption. For sampling we used simulated credit card information with a sample size from 10^3 to 10^7 samples. Comparing the time consumption for ordering and encrypt-

ing the samples on the encryption proxy to uploading and preparing the data on Google BigQuery, has shown that the API runtime period is about 100 times faster than the proxy. These samples also have been tested with complex SELECT and ORDER BY statements. The response time for queries was in any test case less than 10 seconds⁴, even when returning 10^6 query results.

Considering Big Data, we have also evaluated that standard state of the art client laptop⁵ can work with encrypted Big Data. We focused on the encryption as bottle neck for resource consumption. For this examination we have extracted the encryption method used in our modified ebq client and sampled it with random 16 digit integer values, simulating credit card numbers. Table 1 shows a linear correlation in the equivalent file size of the unencrypted data in order to the sample size as well as a linear correlation in time consumption for the encryption in order to the sample size. The sample scenario confirmed our expectation of extremely fast responses once the data is uploaded to the external system but also shows, that basic workstations can work with an extreme high amount of data in a reasonable time.

Sample Size	Digits	Equivalent File Size	Time Consumption
1.000	16	25.9 kB	0.0034 s
10.000	16	268.9 kB	0.0214 s
100.000	16	2.8 MB	0.2035 s
1.000.000	16	28.9 MB	2.0432 s
10.000.000	16	298.9 MB	21.7537 s

Tab. 1: Sample sizes for the encryption validation

6 Conclusion, Future Work and Discussion

We demonstrated that it is possible to use external resources without decreasing the level of security more than the theoretically necessary minimum of Order-Preserving Encryption. With the imperative requirement of revealing the order of the encrypted data, it is possible to work with the data on the external resource, but to maintain the secrecy of private data. With the introduction of a new concept, which depicts to work with encryption on an external resource and by using the modified encrypted BigQuery client, it is feasible to encrypt and sort Big Data using any symmetric encryption algorithm. We have also combined widely used algorithms and state of the art encryption to work with modern Infrastructure-as-a-Service environment. OPE is still a new technology and should be used with care but offers already a feasible way to protect confidentiality. Further work will focus on overcoming the remaining limitations, using the client as encryption proxy. We

⁴ The higher time consumption is caused by the asynchronous API of Google BigQuery. The client will send an asynchronous call to the GBQ API. Depending on the load of the API at the moment, the calls will be processed one by one.

⁵ Intel Core i7-5600 4x2.60GHz; 16 GB RAM; Arch Linux (Kernel: 4.6.2); Python 2.7.11

plan to implement recent algorithm to hide the user's query distribution and by that making it even harder to reveal information besides the order.

With our solution it is possible to handle modern Big Data use cases. Even if, in first place, data is not considered as Big. If a system collecting timestamped GPS data produces a million records a day, it might not be Big Data. In three years, however, the system will have created a billion records. Data processing over a long period of time is important to detect the development or periodic trends. Often, the amount or the length back in time is important and will produce better results the more data sets are available. [TN14] This example shows, that a slower encryption but a fast analysis on the data by many clients is feasible and convenient for modern databases.

We have proposed different encryption algorithms which can be used with our concept besides Order-Preserving Encryption. Different data in a dataset does need different encryption algorithms depending on the projected queries and operations. This does have an impact on the design of the database scheme and can increase data security. With the modified ebq-client, the full range of algorithms and therefore operations is possible.

Recent work by Mavroforakis et. al. [Ma15] has revisited the Modular Order-Preserving Encryption (MOPE) Algorithm [BCO11]. MOPE is slightly different to mOPE, but is also possible for an external database service to gain more information besides the order by observing the user's queries. In [Ma15] the author discusses three contributions to hide the user's query distribution by mixing it with another distribution. The described methods are possible, but are not implemented yet in the proof of concept client of this work.

For high security needs, it is also possible to use Oblivious RAM (ORAM). ORAM is a technique that hides all information about which positions in an outsourced database are accessed by the client, by continually shuffling around and re-encrypting the data. [St13, GO96, Ma15]. Considering [Ma15] ORAM is less efficient and the proposed solutions more sufficient for OPE.

References

- [Ag04] Agrawal, Rakesh; Kiernan, Jerry; Srikant, Ramakrishnan; Xu, Yirong: Order Preserving Encryption for Numeric Data. In: Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data. SIGMOD '04, ACM, New York, NY, USA, pp. 563–574, 2004.
- [Ag12] Agrawal, Divyakant; Abbadi, Amr El; Ooi, Beng Chin; Das, Sudipto; Elmore, Aaron J.: The Evolving Landscape of Data Management in the Cloud. *Int. J. Comput. Sci. Eng.*, 7(1):2–16, March 2012.
- [Am16] Amazon: , Amazon RDS. <http://aws.amazon.com/rds/>, 2016. [Online; Visited 21. Jun 2016].
- [BCO11] Boldyreva, Alexandra; Chenette, Nathan; O'Neill, Adam: Order-preserving Encryption Revisited: Improved Security Analysis and Alternative Solutions. In: Proceedings of the 31st Annual Conference on Advances in Cryptology. CRYPTO'11, Springer-Verlag, Berlin, Heidelberg, pp. 578–595, 2011.

- [Bo09] Boldyreva, Alexandra; Chenette, Nathan; Lee, Younho; O'Neill, Adam: Order-Preserving Symmetric Encryption. In: Proceedings of the 28th Annual International Conference on Advances in Cryptology: The Theory and Applications of Cryptographic Techniques. EUROCRYPT '09, Springer-Verlag, Berlin, Heidelberg, pp. 224–241, 2009.
- [Br08] Brantner, Matthias; Florescu, Daniela; Graf, David; Kossmann, Donald; Kraska, Tim: Building a Database on S3. In: Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data. SIGMOD '08, ACM, New York, NY, USA, pp. 251–264, 2008.
- [Ca11] Cattell, Rick: Scalable SQL and NoSQL Data Stores. SIGMOD Rec., 2011.
- [Cu11] Curino, Carlo; Jones, Evan P. C.; Popa, Raluca Ada; Malviya, Nirmesh; Wu, Eugene; Madden, Sam; Balakrishnan, Hari; Zeldovich, Nikolai: Relational Cloud: A Database-as-a-Service for the Cloud. 5th Biennial Conference on Innovative Data Systems Research, CIDR, 2011.
- [DAEA13] Das, Sudipto; Agrawal, Divyakant; El Abbadi, Amr; ElasTraS: An Elastic, Scalable, and Self-managing Transactional Database for the Cloud. ACM Trans. Database Syst., 38(1):5:1–5:45, April 2013.
- [DG08] Dean, Jeffrey; Ghemawat, Sanjay: MapReduce: Simplified Data Processing on Large Clusters. Commun. ACM, 2008.
- [Ge09] Gentry, Craig: Fully Homomorphic Encryption Using Ideal Lattices. In: Proceedings of the Forty-first Annual ACM Symposium on Theory of Computing. STOC '09, ACM, New York, NY, USA, pp. 169–178, 2009.
- [GHS12] Gentry, Craig; Halevi, Shai; Smart, Nigel P.: Homomorphic Evaluation of the AES Circuit. In (Safavi-Naini, Reihaneh; Canetti, Ran, eds): Advances in Cryptology – CRYPTO 2012: 32nd Annual Cryptology Conference, Santa Barbara, CA, USA, August 19–23, 2012. Proceedings. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 850–867, 2012.
- [GO96] Goldreich, Oded; Ostrovsky, Rafail: Software Protection and Simulation on Oblivious RAMs. J. ACM, 43(3):431–473, May 1996.
- [Go16a] Google Inc.: , encrypted-bigquery-client. <https://github.com/google/encrypted-bigquery-client>, 2016. [Online; Visited 29. Apr 2016], Commit: 7e46deaca522a1bb1e67c3eefe55805bccfd2284.
- [Go16b] Google Inc.: , Google Cloud SQL. <https://cloud.google.com/products/cloud-sql>, 2016. [Online; Visited 21. Jun 2016].
- [GZ07] Ge, T.; Zdonik, S.: Fast, Secure Encryption for Indexing in a Column-Oriented DBMS. In: 2007 IEEE 23rd International Conference on Data Engineering. pp. 676–685, April 2007.
- [Ha02] Hacigümüş, Hakan; Iyer, Bala; Li, Chen; Mehrotra, Sharad: Executing SQL over Encrypted Data in the Database-service-provider Model. In: Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data. SIGMOD '02, ACM, New York, NY, USA, pp. 216–227, 2002.
- [HIM02] Hacigümüş, Hakan; Iyer, B.; Mehrotra, S.: Providing database as a service. In: Data Engineering, 2002. Proceedings. 18th International Conference on. pp. 29–38, 2002.

- [KAK10] Kadhemi, Hasan; Amagasa, Toshiyuki; Kitagawa, Hiroyuki: A Secure and Efficient Order Preserving Encryption Scheme for Relational Databases. Conference: KMIS 2010 - Proceedings of the International Conference on Knowledge Management and Information Sharing, Valencia, Spain, 2010.
- [LE09] LEE, Seungmin; PARK, Tae-Jun; LEE, Donghyeok; NAM, Taekyong; KIM, Sehun: Chaotic Order Preserving Encryption for Efficient and Secure Queries on Databases. IEICE Transactions on Information and Systems, E92.D(11):2207–2217, 2009.
- [LILD12] Liangliang, Xiao; I-Ling, Yen; Dung, T. Huynh: Extending Order Preserving Encryption for Multi-User Systems. Cryptology ePrint Archive, Report 2012/192, 2012.
- [LW12] Liu, D.; Wang, S.: Programmable Order-Preserving Secure Index for Encrypted Database Query. In: Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on. pp. 502–509, June 2012.
- [LW13] Liu, Dongxi; Wang, Shenlu: Nonlinear order preserving index for encrypted database query in service cloud environments. Concurrency and Computation: Practice and Experience, 25(13):1967–1984, 2013.
- [Ma15] Mavroforakis, Charalampos; Chenette, Nathan; O’Neill, Adam; Kollios, George; Canetti, Ran: Modular Order-Preserving Encryption, Revisited. In: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data. SIGMOD ’15, ACM, New York, NY, USA, pp. 763–777, 2015.
- [Me10] Melnik, Sergey; Gubarev, Andrey; Long, Jing Jing; Romer, Geoffrey; Shivakumar, Shiva; Tolton, Matt; Vassilakis, Theo: Dremel: Interactive Analysis of Web-scale Datasets. Proc. VLDB Endow., 2010.
- [Mi16] Microsoft: , SQL Azure. <http://www.windowsazure.com/en-us/develop/net/fundamentals/cloud-storage/>, 2016. [Online; Visited 21. Jun 2016].
- [Pa99] Paillier, Pascal: Public-Key Cryptosystems Based on Composite Degree Residuosity Classes. Advances in Cryptology - EUROCRYPT 99, 1999.
- [PLZ13] Popa, Raluca Ada; Li, Frank H.; Zeldovich, Nikolai: An Ideal-Security Protocol for Order-Preserving Encoding. IEEE Symposium of Security and Privacy, 2013.
- [Po11] Popa, Raluca Ada; Redfield, Catherine M. S.; Zeldovich, Nikolai; Balakrishnan, Hari: CryptDB: Protecting Confidentiality with Encrypted Query Processing. In: Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles. SOSP ’11, ACM, New York, NY, USA, pp. 85–100, 2011.
- [Sc16] Schindler, Timo: Secure Parallel Processing of Big Data Using Order-Preserving Encryption on Google BigQuery. Master’s thesis, OTH Regensburg, 2016.
- [St13] Stefanov, Emil; van Dijk, Marten; Shi, Elaine; Fletcher, Christopher; Ren, Ling; Yu, Xiangyao; Devadas, Srinivas: Path ORAM: An Extremely Simple Oblivious RAM Protocol. In: Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security. CCS ’13, ACM, New York, NY, USA, pp. 299–310, 2013.
- [TN14] Tigani, Jordan; Naidu, Siddhartha: Google BigQuery Analytics. Wiley, 2014.
- [YK15] Yin, S.; Kaynak, O.: Big Data for Modern Industry: Challenges and Trends [Point of View]. Proceedings of the IEEE, 103(2):143–146, Feb 2015.

Big Data Architecture for the Semantic Analysis of Complex Events in Manufacturing

Marco F. Huber¹, Martin Voigt² and Axel-Cyrille Ngonga Ngomo³

Abstract: Today's production processes are monitored densely via a myriad of sensors. Appropriately processed, the data carries the potential to further increase automation, to early detect failures, and thus, to reduce costs. However, these opportunities can only be exploited if the storage and processing system is capable to deal with massive data in (near) real-time. The distributed and cloud-based architecture proposed in this paper addresses the needs of the manufacturing and plant industry in terms of the processing of a huge amount of complex event data from process monitoring. It combines machine learning and semantic technologies, which allows not only the automatic detection of process failures and their root causes, it also renders these findings in a human-interpretable way thanks to the semantification. By means of a real-world use case, where large-scale printing machines are monitored, we can demonstrate the capabilities of the proposed architecture.

Keywords: Industry 4.0, Semantic Web, Big Data, Machine Learning, Root-Cause Analysis

1 Introduction

Manufacturing and plant engineering is the economically strongest industry sector in Germany with an annual turnover of 212 billion Euros in 2014 and 218 billion Euros in 2015 and more than one million employees [VDM16]. Its success originates from increasingly sophisticated production machines that facilitate the execution of previously manual tasks in an automated and safe manner. To determine the current operational state with high precision, these machines are armed with a plurality of sensors. Similar to modern airliners state-of-the-art manufacturing plants like sheetfed printing machines or injection molding machines produce several gigabytes or even terabytes of sensor data per day. However, manufacturers are increasingly challenged by managing the incoming flood of data and by obtaining insights from the data in order to react timely on events like failures of the production processes. According to [MCB⁺11], the advent of methods and technologies for analyzing this so-called big data affords the opportunity to reduce production costs by up to 50%. Big data analytics are also considered as one of the key drivers of the forth industrial revolution, also known as Industry 4.0 [LKY14].

In order to gain these benefits, a data value chain needs to be applied. This chain commonly comprises the following steps: 1) data acquisition, 2) data analysis, 3) data curation, 4) data storage, and 5) data usage. For their implementation, very general architecture principles

¹ USU Software AG, Ruppurrer Str. 1, 76137 Karlsruhe, Germany, marco.huber@ieee.org

² Ontos GmbH, Wurzner Str. 154a, 04318 Leipzig, martin.voigt@ontos.com

³ University of Leipzig, Augustusplatz 10, 04109 Leipzig, Germany, ngonga@informatik.uni-leipzig.de

such as the Lambda⁴ and Kappa⁵ architectures are available. However, none of them fits perfectly for the scenario of event processing from plants, which requires both data analytics being scalable in terms of data volume and velocity as well as human-interpretable results. The main reason besides the continuously evolving zoo of big data technologies⁶ is the usage of standards and technologies from the Semantic Web domain, e. g., RDF⁷, SPARQL⁸, and triple stores. Thus, in order to setup our data value chain, we need to fulfill two objectives regarding the architecture: 1) Combine general big data technologies like Apache Spark⁹ with the Semantic Web stack, and 2) enhance the Semantic Web-based tools to handle big data in their volume, velocity and variety.

We introduce a big data architecture for the analysis of event data from manufacturing processes. This architecture is currently being developed within the SAKE project¹⁰, which is funded by the German Federal Ministry for Economic Affairs and Energy (BMWi). The main goals of the project and the architecture are to facilitate the timely detection and data-driven prediction of failures from event data. To cope with the potentially large amount of data, the architecture utilizes state-of-the-art distributed cloud-based big data technologies such as NoSQL databases for data storage and Apache Spark for data analysis and machine learning. In contrast to existing solutions, data storage and machine learning are compliant with Semantic Web standards. By these means, the following problems of processing big data identified in [vKRS⁺13] and being relevant for the manufacturing and plant industry can be solved: 1) ensuring unique data semantics, 2) scalable machine learning, and 3) generation of human-interpretable analytics.

The paper is structured as follows. In Sec. 2, we elaborate on the design goals being addressed by the proposed architecture. Furthermore, an architecture overview is given. The applicability of this architecture is demonstrated by means of a real-world manufacturing use case in Sec. 3. Conclusions and an outlook to future work complete the paper.

2 Architecture Overview

The guiding theme behind the development of the SAKE architecture can be phrased as:

Facilitate analytics for big data streams in such a way that the end user can understand the results and can reproduce their creation.

Under consideration of the aforementioned data value chain, this theme resulted in the modular architecture depicted in Fig. 1. The architecture consists of the following three layers: The *acquisition layer* is mainly based on Semantic Web technologies and, thus, lays the foundation for human interpretability. Link discovery and structured machine learning are part of the *analytics layer* and allow for the automatic detection of patterns, relations, and failures in event data streams across data sources. The most visible layer for the end users, i. e., the *application layer*, provides intuitive user interfaces and dashboards for a straightforward and gradual exploration and analysis of data.

⁴ <http://lambda-architecture.net/> ⁵ <https://www.oreilly.com/ideas/questioning-the-lambda-architecture>

⁶ <https://www.linkedin.com/pulse/100-open-source-big-data-architecture-papers-anil-madan>

⁷ <https://www.w3.org/TR/rdf11-concepts/>

⁸ <https://www.w3.org/TR/sparql11-overview/>

⁹ <http://spark.apache.org/> ¹⁰ www.sake-projekt.de

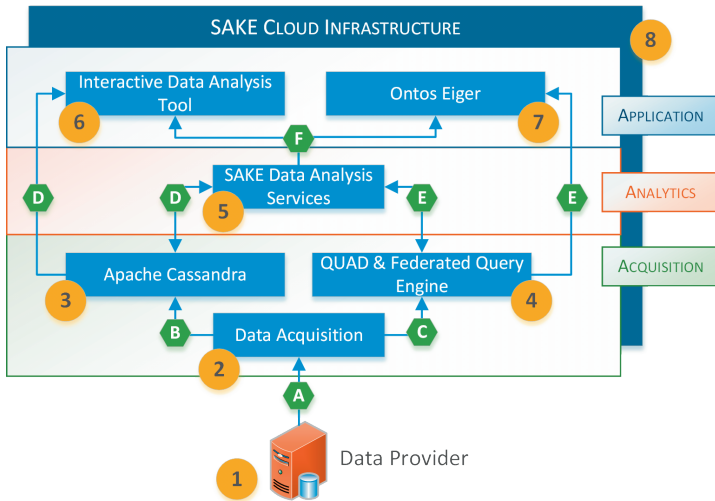


Fig. 1: High-level overview of the SAKE architecture.

These layers run in a distributed manner in the cloud (see Fig. 1-(8)). This is a necessary design decision given the large amount of data common in today's manufacturing processes. In order to stream the data into the platform (Fig. 1-(A)), we rely on Apache Kafka¹¹. In the platform, the communication between all components is in general flexible and loosely coupled by using the REST paradigm. The semantic RDF data is written and queried by using SPARQL (Fig. 1-(C), (E) & (F)) whilst the query language CQL¹² is used for Cassandra (Fig. 1-(B) & (D)). We also developed a novel SPARQL connector¹³ for Apache Spark that eases the querying and usage of RDF data in the analytics layer.

In the following, we give a detailed description of the individual layers and their modules.

2.1 Acquisition Layer

The acquisition layer builds the interface between the data provider (Fig. 1-(1)) and the SAKE system. It carries out the explication of all required RDF for the further processing and ensures its availability to subsequent modules. To achieve this goal, it implements a three-step process: The *preprocessing* contains all data-cleansing steps to prepare the input before distributing it into the cluster. The *parsing* contains several steps that are required to extract the RDF statements required from the input. Finally, the *storage* stage allows to persist the generated model.

Preprocessing Plant machines commonly provide data in two different formats: 1) as compressed XML/CSV files and 2) as data streams. For the first case, tests revealed that

¹¹ <http://kafka.apache.org/>

¹² <https://cassandra.apache.org/doc/cql3/CQL-2.2.html>

¹³ <https://github.com/USU-Research/spark-sparql-connector>

working with compressed and XML files in current distributed frameworks like Apache Spark is not straight-forward, unstable and inefficient because they generally work with line-based input files. For working with data streams, we rely on Apache Kafka to collect and prepare the data on-the-fly. The preprocessing of the data comprises the following tasks: 1) getting or receiving the data from the remote host, 2) uncompress the data if required, 3) transform the data to line-based formats, and 4) ensure UTF-8-conformity.

Parsing In the next step, the data from preprocessing is parsed into a generic event model. It is based on the SAKE event ontology, which formalizes generic concepts, e. g., *Event*, *Timeline* and their time-related relations, as well as domain specific context information, e. g., modules of the printing machines. The workload could be easily parallelized and thus, distributed in a data processing cluster like Apache Spark. The order of the events is kept by their timestamps and provenance ID. Each node in the cluster parses a subset of the input into an in-memory model, which is finally used to create the required output. SAKE provide different parsers, e. g., for XML or CSV, which can be extended for specific use cases through an extensible plug-in mechanism. This core step of the data acquisition requires the most knowledge from domain experts.

Storage In the last step, the generated data model is stored in the data sinks. As shown in Fig. 1, we currently support two different storage solutions: Apache Cassandra ③ is especially used to store numeric data whereas an RDF store (QUAD) ④ is our premiere data sink. The latter is used in order to store all data alongside the defined SAKE ontology. Due to the enormous data size (multiple terabytes) the QUAD instances [PPD⁺13] are distributed in the cluster. For an efficient access to the data, we developed a concept of federated query engine (Fig. 1-④) with a cost-based query optimization.

2.2 Analytics Layer

The analytics layer aims to detect failures and anomalies in the event data. For this purpose, machine learning is employed, where training data comprising events of a regular execution of the manufacturing process are used for learning a probabilistic “normal model” for instance via kernel density estimation. A significant deviation of events of the incoming data stream from this model allows identifying an anomalous situation.

Detecting anomalous behavior is merely a first step. To avoid future failures, it is important to understand the causes of failures for effective counter measures. Accordingly, methods for the provision of drill-downs to root cause become necessary. Such analyses employ structured machine learning, where we learn classifiers from labeled event sequences. The steps involved to achieve this goal are:

Link Discovery First, we link events within and across machine data streams. Therewith, we aim to ensure that the subsequent machine learning process can learn complex OWL class expressions. Given that events can be regarded as entities that exist within a certain time interval, we rely on 1) the functionality offered by the LIMES framework for link discovery [Ngo12], 2) a reduction of Allen’s algebra [All83] to atomic relations to achieve

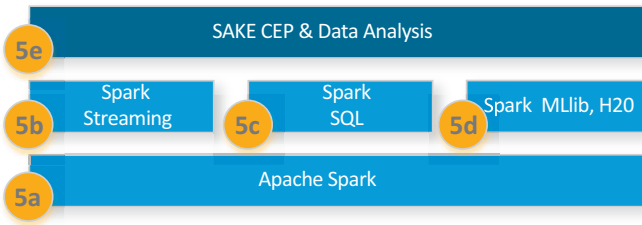


Fig. 2: Detailed view of the SAKE Data Analysis Services based on Apache Spark.

highly scalable link discovery and 3) an efficient planner based on [Ngo14] to scale up to linking tens of thousands of events per minute and processor kernel.

Structured Machine Learning The results of the acquisition and the link discovery are a network of linked events. We learn event descriptions by using a refinement-operator-based approach [Leh09]. The idea here is to start with positive and negative examples for anomalous events and to explore the RDF network around these events (including related events) to find the most generic network patterns that describe the positive examples and do not describe the negative examples.

The SAKE Data Analysis Services Fig. 1-⑤, which form the core module of the analytics layer, are depicted in more detail in Fig. 2. The aforementioned anomaly detection and root-cause analysis capabilities are part of the the SAKE CEP & Data Analysis module ⑤e, where CEP stands for complex event processing and facilitates the application of analytics like anomaly detection on streaming event data. This module sits on top of an Apache Spark stack ⑤a–⑤d. Apache Spark is an open-source cluster computing framework. Similar to MapReduce [DG08], it provides a simplified programming interface for performing distributed computing. In contrast to MapReduce, Spark is especially well suited for iterative algorithms, which are common in machine learning.

While the Spark core ⑤a is designed for batch processing, our analytics also need to process event streams. Thus, we utilize the Spark Streaming module ⑤b in addition. Spark SQL ⑤c supports processing of structured data via a table-like data format called DataFrame, which is preferred by several implementations of machine learning algorithms. These are part of module ⑤d, which comprises the machine learning libraries Spark MLlib and H2O¹⁴.

2.3 Application Layer

At the top layer of the SAKE architecture we combine automatic verbalization techniques with application specific dashboards and reporting tools. The Interactive Data Analysis Tools Fig. 1-⑥ essentially provide web applications that are customized to the needs of the end user. They provide functionality for data exploration, dashboards for live process monitoring, and reporting. Therefore, we employ and extend notebook environments like

¹⁴ <http://www.h2o.ai>

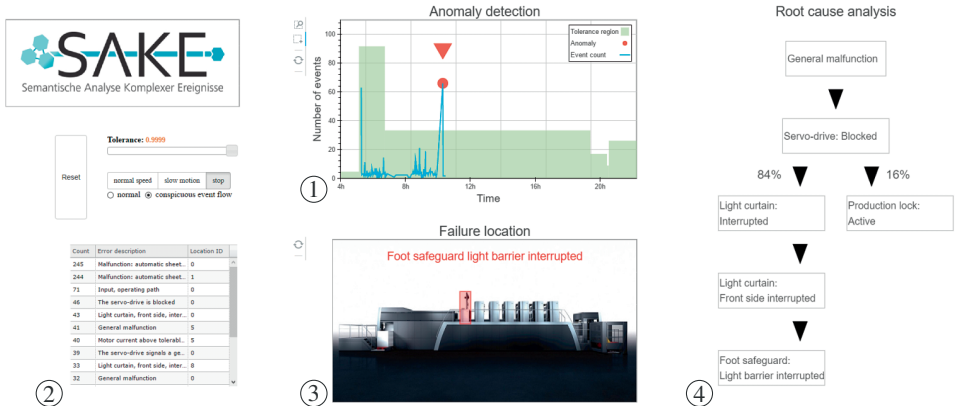


Fig. 3: Web application for analyzing anomalies of printing machines. ① Anomaly detection in event streams, ② event statistics, ③ visualization of failure location, ④ root-cause analysis.

Jupyter¹⁵ and JavaScript libraries like D3¹⁶ or Bokeh¹⁷. This combination turned out to be especially suitable for supporting explorative analysis and interactive visualization. Furthermore, we use the extendable, web-based Linked Data suite Ontos Eiger (Fig. 1-⑦) in order to create user-driver dashboards for the semantic data. Depending on the given use case, domain experts but also managers can easily add and configure required widgets to visualize and understand the outcome of the analytics services.

One of the outcomes of these services, especially the structured machine learning, are OWL Class expressions¹⁸. Given that most users of the SAKE platform are not versed in Semantic Web technologies, we provide a verbalization framework for OWL Class expressions based SPARQL2NL framework [NNBU⁺13]: SemWeb2NL¹⁹ uses a bottom-up approach to verbalize expressions. Atomic expressions are verbalized using verbalization patterns. Complex expressions are verbalized by combining the results of the verbalization of the corresponding atoms via linguistic rules. These verbalizations can then be used in the user dashboards.

3 Use Case: Processing Events from Printing Machines

Besides the design and development of a big data architecture, the SAKE project also aims for applying the architecture to real-world manufacturing processes. In this section, we describe the current status of a use case, in which the SAKE architecture is used for monitoring event data of a sheetfed printing machine of Heidelberger Druckmaschinen AG (HDM)²⁰. HDM is a project member and the considered printing machine depicted in Fig. 3-③ is running at the printing facility of one of HDM's customers.

¹⁵ <http://jupyter.org> ¹⁶ <http://d3js.org> ¹⁷ <http://bokeh.pydata.org/> ¹⁸ <https://www.w3.org/TR/owl12-syntax/>
¹⁹ <https://github.com/AKSW/SemWeb2NL/> ²⁰ <http://www.heidelberg.com/>

On average, printing machines of this type produce approximately four million events per day, where events are represented in XML and comprise for instance the status of the current printing job, safety issues, sensor readings, alerts, etc. Based on these events, we are interested in detecting anomalies leading to an unplanned reboot of the printing machine. Given this use case, the layers of the SAKE architecture and its modules are customized as follows:

Acquisition layer The XML-based event data is pushed into the SAKE architecture and processed as explained in Sec. 2.1. Since the causes for the anomalies is unknown, all available events are extracted and “semantified”. By also obtaining explicit information from plain text log messages, the final data size grows by a factor of eight compared to the size of the plain log files.

Analytics layer The first step is the link discovery across the events generated by the machine. We generate links according to the first seven relations of Allen’s algebra (all other relations can be derived from them). The resulting network of RDF resources and a set of examples (positive and negative) are then forwarded to the DL-Learner. The refinement operator underlying CELOE [LABT11] is then used to derive the corresponding OWL Class expression. For anomaly detection historic events form a training set for learning a “normal model” of the printing machine. Metrics on the frequency of the incoming event stream are then compared against this model.

Application layer We developed a Web application based on Bokeh, which among others visualizes the incoming event stream and detected anomalies (Fig. 3-①). For anomalous situations, the web frontend provides information about all events in a certain time window (Fig. 3-②), the location of the machine part probably causing the anomaly (Fig. 3-③), and linked events that are considered likely to trace back to the root-cause (Fig. 3-④).

4 Conclusion and Future Work

The proposed SAKE architecture reflects the needs of manufacturing and plant engineering in the Industry 4.0 era by combining the latest technology for analyzing big data streams with Semantic Web standards. In doing so, it enables scalable machine learning together with human comprehensible results. Thanks to its modular structure, the architecture can be customized easily to a given manufacturing process as was shown exemplary for the monitoring of a large-scale printing machine.

While the basic design of the architecture is considered complete, there is still room for further developments on the module level. For instance, we plan to exploit unsupervised or at least semi-supervised machine learning for root-cause analysis in order to minimize the effort of event labeling. We also plan to extend our works on large-scale link discovery [NH16], planning [Ngo14] and machine learning [Leh09] to big data processing frameworks such as Spark. Finally, we plan to apply the SAKE architecture also to use cases aside the manufacturing industry like in IT monitoring in order to identify and close gaps towards a more universal architecture for the semantics analysis of complex events.

Acknowledgement

This work was partially supported by the BMWi project SAKE (Grant No. 01MD15006).

References

- [All83] James F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.
- [DG08] Jeffrey Dean and Sanjay Ghemawat. MapReduce: Simplified Data Processing on Large Clusters. *Magazine Communications of the ACM*, 51(1):107–113, January 2008.
- [LABT11] Jens Lehmann, Sören Auer, Lorenz Bühmann, and Sebastian Tramp. Class expression learning for ontology engineering. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(1):71–81, 2011.
- [Leh09] Jens Lehmann. DL-Learner: learning concepts in description logics. *The Journal of Machine Learning Research*, 10:2639–2642, 2009.
- [LKY14] Jay Lee, Hung-An Kao, and Shanhu Yang. Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment. *Procedia CIRP*, 16:3–8, 2014.
- [MCB⁺11] James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. Big data: The next frontier for innovation, competition, and productivity. Report, McKinsey Global Institute, June 2011.
- [Ngo12] Axel-Cyrille Ngonga Ngomo. On link discovery using a hybrid approach. *Journal on Data Semantics*, 1(4):203–217, 2012.
- [Ngo14] Axel-Cyrille Ngonga Ngomo. Helios—execution optimization for link discovery. In *The Semantic Web—ISWC 2014*, pages 17–32. Springer, 2014.
- [NH16] Axel-Cyrille Ngonga Ngomo and Mofeed Hassan. The Lazy Traveling Salesman – Memory Management for Large-Scale Link Discovery. In *Proceedings of the Extended Semantic Web Conference*, 2016.
- [NNBU⁺13] Axel-Cyrille Ngonga Ngomo, Lorenz Bühmann, Christina Unger, Jens Lehmann, and Daniel Gerber. Sorry, i don’t speak SPARQL: translating SPARQL queries into natural language. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 977–988, 2013.
- [PPD⁺13] Alexander Potocki, Anton Polukhin, Grigory Drobyazko, Daniel Hladky, Victor Klintsov, and Jörg Unbehauen. OntoQuad: Native High-Speed RDF DBMS for Semantic Web. In *Knowledge Engineering and the Semantic Web*, volume 394, pages 117–131. Springer Berlin Heidelberg, 2013.
- [VDM16] Maschinenbau in Zahl und Bild. Annual report, VDMA, 2016.
- [vKRS⁺13] Tim van Kasteren, Herman Ravkin, Martin Strohbach, Mario Lischka, Miguel Tinte, Tomas Pariente, Tilman Becker, Axel Ngonga, Klaus Lyko, Sebastian Hellmann, Mohamed Morsey, Philipp Frischmuth, Ivan Ermilov, Michael Martin, Amrapali Zaveri, Sarven Capadisli, Edward Curry, Andre Freitas, Nur Aini Rakhmawati, Umair ul Hassan, Aftab Iqbal, Anna Karpinska, Syzyon Danielczyk, Pablo Mendes, John Domingue, Anna Fensel, and Andreas Thalhammer. Consolidated Technical White Papers. Project report, Big Data Public Private Forum (BIG), 2013.

Predicting the Availability of Parking Spaces with Publicly Available Data

Christoph Pflügler¹, Thomas Köhn¹, Maximilian Schreieck¹, Manuel Wiesche¹ und Helmut Krcmar¹

Abstract: Searching for parking spaces on the street causes a significant part of the urban traffic and results in extra costs for the drivers in terms of time and fuel consumption. Existing approaches to predict the availability of parking spaces have significant drawbacks as they are either expensive or rely on the users' information. This article deals with the prediction of the parking situation based on publicly available data that can be accessed cost-efficiently. Suitable categories of data are identified based on a literature review. Subsequently, a prototypical system that employs a neural network is implemented. The relevance of the different categories of data is evaluated based on 2,779 real world records. The results show that weekday, time of the day, location, and temperature have a significant impact on the prediction; whereas events, traffic, vacation time and rainfall are only of secondary importance. This article categorizes existing solutions to support finding parking spaces and shows that publicly available information can provide a good starting point for the prediction of the availability of parking spaces.

Keywords: Smart Parking, Public Data, Smart City, Parking Prediction, Parking, Neuronal Network

1 Introduction

The volume of traffic in urban regions is continuously increasing [MAZ15] [Ba05]. Additionally, to increasing traffic congestion and longer travel times, health risks for the residents rise [ZB13], [Ch15], [Ra14]. A significant portion of the traffic in urban environments is caused by searching for parking spaces [Sh06]. According to Giuffrè et al. [GST12] the search for parking space is responsible for up to 40% of the total traffic within cities. Apart from the increase of the own consumption of fuel, the search for parking spaces will also negatively impact other road users' consumption. Parking seekers often drive slowly and cause the traffic flow behind them to slow down [In14]. According to Le Fauconnier and Gantelet [LG06] on average it takes 12 minutes to find a free parking space.

In the literature and on the market there are a number of applications and concepts that aim to simplify the search for a free parking space. These applications are usually based either on user data or data from sensors in parking spaces. Therefore, they are either dependent on data entered by the users or they have high setup costs due to the installation of sensors in the parking spaces. A system based on publicly available data that predicts

¹ Technical University of Munich, Chair for Information Systems, Boltzmannstrasse 3, D-85748 Garching, {christoph.pfluegler|thomas.koehn|maximilian.schreieck|wiesche|krcmar}@in.tum.de

the availability of parking by means of an intelligent algorithm would not have the mentioned disadvantages. Up to now, it has not yet been investigated which data is suitable for the prediction of parking spaces within a city.

Based on a literature review, weather, time, traffic, events and locations are identified as possibly relevant categories of data. It is important that the data is publicly available to ensure its independence and sustainability. Subsequently, the data is integrated in a prototypical solution based on a neural network to predict the availability of free parking spaces within a certain area. The data is evaluated in terms of its contribution to the improvement of the predictions. The results show that weekday, time of the day, location, and temperature have a significant impact on the prediction, whereas events, traffic, vacation time and rainfall are only of secondary importance. The training data was gathered in the city of Munich, Germany, where the parking situation is a significant transport problem [CM16].

The Design Science Research (DSR) approach was selected for this article; therefore, the paper is organized according to the framework of Gregor and Hevner [GH13]. The first section provides an overview of existing solutions for the identification of available parking spaces. In the next section, we present a literature review on the categories of data that can influence the parking situation and are publicly accessible. Consecutively, the methodology and the exemplary implementation are described. Subsequently, an exemplary data set is used to evaluate the prototypical implementation. The article concludes with a discussion of the results and a summary of the contribution.

2 Existing solutions to support the search for parking spaces

For the identification of existing solutions in addition to the literature, the market has been systematically studied with a focus on existing concepts and applications. The search was concentrated on systems that support the search for parking spaces on the street. Other kinds of parking such as parking garages or closed parking were not considered. The identified concepts and applications can be divided into three categories: systems based on user data, systems based on parking data and systems based on publicly available data.

2.1 Systems based on user data

A variety of systems work on the basis of data that is provided by users (crowd sourcing). Two types of applications can be provided: Free parking spaces explicitly reported by the users or implicitly recognized by an application installed on the smartphone of the users. The explicitly reported kind is directly activated by the user. The implicit recognition is carried out by the application using the data of the user but without his participation [NEM13]. In both types, users can view the reported free parking spaces and reserve them via a website or a mobile application (app). Applications of this type are widely used and successful in the market. Examples of established applications based on explicit user data

are: ParkoPedia [PA15c], ParkMünchen [PM15], Parkonaut [PA15b] and ParkTAG [PA15d]. Systems based on the implicit recognition of users' data are not yet widely used. A system based on implicit user data is Pocket Parker of Nandugudi et al. [NNC14]. This system measures the variations of the speed of the user and identifies the occupied parking space. Similarly, the system Park Sense [NEM13] measures how fast wireless networks in the vicinity of the user change to derive the users' speed. This data is used to identify when the user has left a parking space. Other example of a system with implicit recognition is the cooperation between the car manufacturer BMW and INRIX which evaluates user data to generate parking information [In15]

Systems based on user data are generally the most widely used and have the advantage of low construction and operation costs [NEM13]. Disadvantages of this type of systems are that a large amount of users is needed to generate sufficient data to provide adequate parking information [CSR12], [Gr15]. This effectively represents a paradox, since such applications are used by users only if sufficient parking information is displayed, but the parking information can be generated only after the users provide the data [DRH11]. Moreover, the quality of the parking information is not ensured. Since the parking spaces are not monitored in real time, it is not guaranteed that these are still available when other users reach them and want to park [Ma10].

2.2 Systems based on parking data

Systems in this category are based on parking spaces equipped with sensors. The sensors detect whether a parking space is currently occupied and reports it to a central database. Various companies offer this kind of solutions [SI15], [GE15], [SP15]. Concrete concepts are also presented by Sujith et al. [Su14], McNeal [Mc13] and Seong-Eun et al. [Se08]. The advantage of this type of systems is that the parking information is very precise and available in real time [NEM13]. The main disadvantage that can hinder the deployment of these technologies is shown by Nandugudi et al. [NNC14]: the equipment necessary for one single parking lot costs around US\$ 2,500. There are also additional costs for cabling, interfaces and communication systems. The SFpark project in San Francisco endowed a total of 8,622 parking sensors and costed a total of about US\$ 18 million [Mc13].

2.3 Systems based on publicly available data

The two previously presented types of systems bring significant disadvantages. Either they have very high initial and operational costs (systems based on parking data) or there is a strong dependence on explicit or implicit user data (systems based on users' data). A system that analyzes publicly available data to generate predictions for the present, would be cost-effective and not as dependent on a user base. The focus on publicly available data also ensures a higher independence. A system that is already used in practice and is not described in the literature, is the system ParkNav. ParkNav combines data of different categories and, using an algorithm, predicts probability values for the availability of

parking spaces [Gr15], [PA15a]. However, it has not been studied in literature which data categories are useful for such a system.

3 Relevant categories of publicly available data

For the development of a system based on publicly available data, it is necessary to identify possible influencing factors to find a parking space. To this end, a literature review was conducted with the aim to identify categories of data that influence the parking situation and are publicly available. The following categories were identified and are summarized in Table 1.

According to Yang et al. [YLW03] weather information is of central importance, since weather significantly affects the current traffic behavior and the traffic flow's intensity. Greengard [Gr15] agrees with this argument and specifically considers rainfall and temperature as important influencing factors for parking prediction.

According to Zehe et al. [Ze07], the time of the day is of great importance. Their results show that holidays, weekdays and time of the day lead to different parking situations. Greengard [Gr15] has a similar conclusion, his work argues that the traffic volume is different during holidays and vacation time as more parking spaces will be searched, making it more difficult to find a parking space. In general, the time can be divided into five data sets, which can impact the traffic volume: time of the day, day, month, bank holidays and vacation time.

Another factor that may influence the availability of parking spaces are events [KMP12], [YLW03]. Events, such as concerts or matches, cause a significant increase in the volume of traffic, consequently more parking spaces will be searched and the search for parking spaces will be more difficult.

Shin and Jun [SJ14] and Yang et al. [YLW03] argue that traffic information is an important factor for the availability of parking spaces. A generally higher traffic volume makes it harder to find a parking space, because more parking spaces will be searched.

Another important factor influencing the car parking is identified by Mathur et al. [Ma10]. They describe that parking availability often depends on the location. The chance to find a parking space generally is higher in certain locations than in others. Giuffrè et al. [GST12] use the example of Milan where the parking situation differs strongly from one area of the city to another. To integrate the impact of the location, it is helpful to divide the urban area into equal rectangular cells. Mathur et al. [Ma10] divided the city of San Francisco into cells of 175x190 meters, because this size ensures that most of the time only one street segment lies within a cell.

Database category	Data	Data source example
Time	Time of the day	Coordinated Universal Time (UCT) Server
	Weekday	
	Month	
	Bank holidays	
	Vacation time	
Weather	Temperature	Deutscher Wetterdienst [DW15]
	Rainfall	openweatherAPI [OW15]
Traffic	Traffic situation	Bing Maps [MI15]
Event	Event locations	eventful [Ev15]
Users' request for parking	Location of users' request	GPS coordinates of the app's user

Table 1: Overview of the categories of data, the data and possible data sources

4 Implementation and evaluation of a prototypical system for the prediction of the parking situation

To develop the prototypical implementation and the integration of the different data sets, we followed the DSR approach [GH13], [He04]. The procedure is based on the DSR process proposed by Peffers et al. [Pe07]. It starts with the identification of existing approaches as well as solutions and their disadvantages (chapter 2). The implementation is based on the insights from this analysis: we chose a system that uses publicly available data to predict the availability of parking spaces and identified relevant data sources (chapter 3). Based on this, the exemplary implementation by means of a neural network is described in this chapter followed by an evaluation of the system. The system and its components are shown in Figure 1, the components will be described in detail below. For the evaluation of the data categories, training data was recorded during the period from July to September 2015 in the downtown area of Munich with a specially created web app. In total, 2,779 records for parking availability in the downtown area were collected. In about 23% of the cases at least one parking space was available within a cell. The web app has been designed for being used while going for a run within the city or walking the dog. Therefore it had a very simple interface. The data collection has been conducted by the authors and by voluntarily participating students.

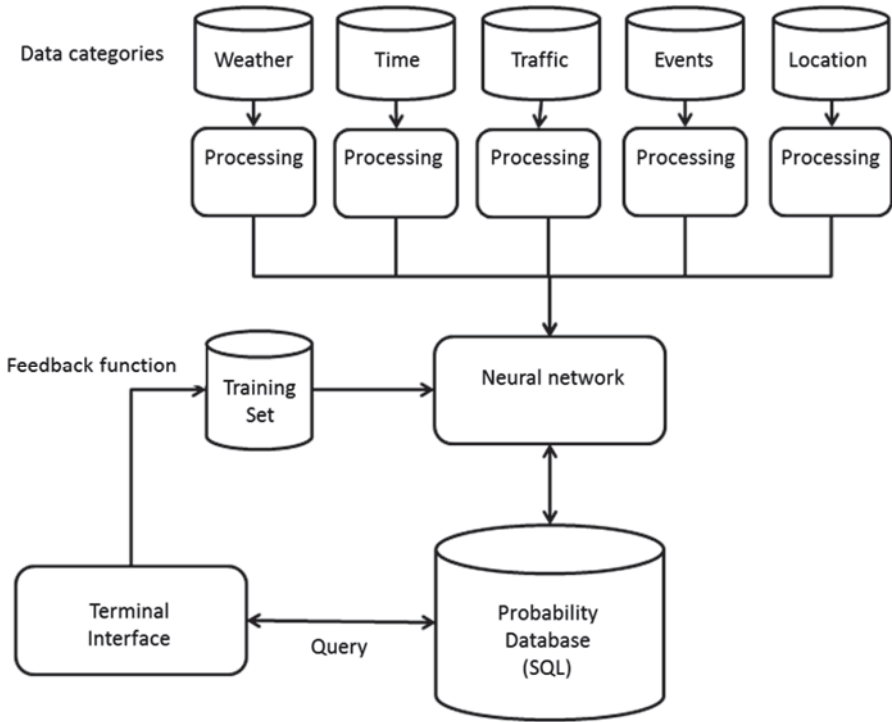


Figure 1: Exemplary implementation by means of a neural network.

4.1 Data preparation

The previously discussed categories of data were integrated as input components of the system (Figure 1). These were designed as autonomous and independent components, because categories of data can be easily included or excluded in the prediction. These components gather the information from the appropriate data sources and make sure that the resulting data is processed appropriately for the subsequent neural network. Some datasets change less frequently than others and it is not necessary to re-query them, therefore these components use the temporary memory.

To process the location data, the urban area of Munich (area within the “Mittlere Ring”) was divided into equal rectangular cells according to Mathur et al. [Ma10] as shown in Figure 2. The grid has a size of 74,429m x 111,193m, as these scales relate to 0.001 degrees in latitude and longitude. Using this cell size ensures that usually only one street segment lies in one cell. Latitude and longitude coordinates were used for segmentation, since they facilitate the identification of relevant cells for a given position.

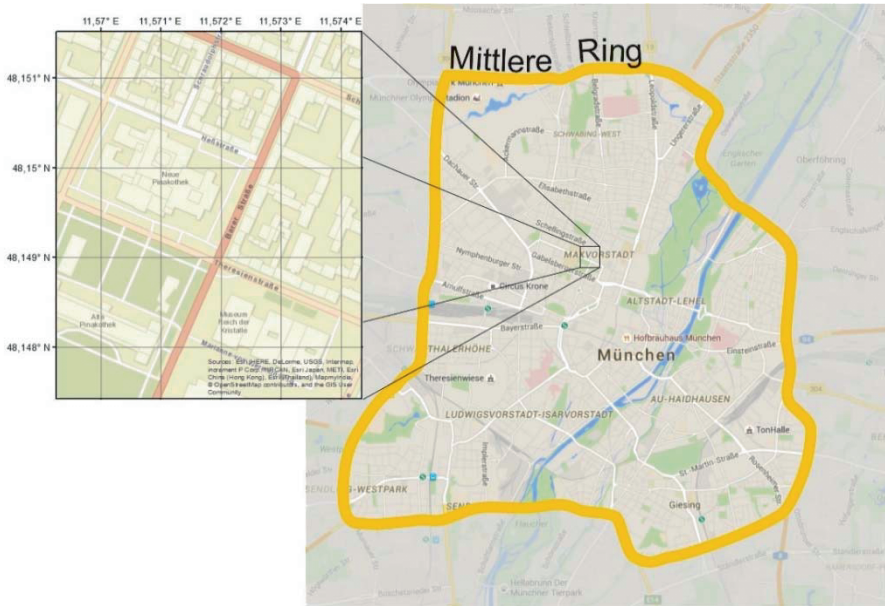


Figure 2: Grid applied to the urban area of Munich

4.2 Neural network

Several reasons support the usage of a neural network as a prediction algorithm. First, neural networks are particularly suitable for predicting events where little or nothing is known about the underlying relationships and features of the events, but enough training data or observation values are available. The information needed can be derived from the training data [ZEH14], [YLW03]. Second, neural networks have the advantage of continuous learning. The training dataset can be easily complemented by other data, so that a continuous improvement of the prediction is possible. Third, neural networks offer the possibility to consider non-linear relationships and interaction effects through the integration of hidden layers [ZEH98], [Ze07].

The number of input neurons of the neural network corresponds to the number of the data from the data sources. It has to be noted that the use of the continuous influences of certain data might not be useful and the input values should be considered as individual neurons. For example, the input value “weekday” should not be used as a continuous value from 1 to 7, since this would mean that the last weekdays (5, 6 or 7) are associated to a higher weight than the first ones (1,2 and 3). Instead, each weekday should be integrated as a single binary neuron. This procedure was used for all the data for which continuity is not meaningful. This is the case for the following neurons: Weekday (7 neurons); Month (12 neurons); Time (24 neurons); Location (1,323 neurons, one for each cell area); Temperature (12 neurons, by establishing a scale of ≤ -20 °C to ≥ 40 °C and dividing the

values in steps of 5 ° C). The remaining data sources are represented by one single neuron each, in which the continuous influence can be taken (traffic intensity, rainfall, i.e.) or the input value is only binary (event, holiday, vacation time, i.e.). Overall, we consider 1383 input neurons. As output neuron a single neuron is used, which represents the predicted value. The amount of neurons and intermediate layers depends of the complexity of the decision [ZEH98], [He08]. There are no rules how a neural network exactly should look, but it is recommended to try out several attempts and choose the one with the lowest error rate [ZEH98], [Sc97]. As a guideline, it can be assumed that no more than two intermediate levels are needed and that the amount of neurons should be between the number of input and output neurons [He08].

4.3 Training data set

The feedback function presented in the architectural model (Figure 1), makes it possible to enhance the training set of the neural network continuously with new training values. The feedback function also ensures that the user can enter information about the correctness of the prediction after receiving a prediction on the parking situation. By providing this information, the prediction can steadily increase its accuracy.

For the collection of an initial training dataset a separate web application has been implemented. Users open the app and start a tour. Every 5 seconds, the current position and the current time are recorded. If they see a free parking space, they press a button. If a user passes through a cell without reporting a free parking space, it is declared as occupied, otherwise as free. The web application was designed for users that go for a run or walk their dog within the city.

4.4 Probability database

To ensure the fastest possible response time of the system, the predicted values are not separated from each request, but are calculated in time intervals and stored in the probability database. The division in cells of the urban area mentioned above is applied here, for the downtown of Munich, resulting in 1,323 cells. The cells are displayed in the database and a probability of availability is assigned to each cell, which states how likely it is to find a free parking space in this area. As the prediction algorithm continuously calculates values for the cells, it is ensured that the database always contains up-to-date values. The calculation of values for all 1,323 cells takes about 20 minutes on modern hardware.

4.5 Evaluation of the data categories

Regarding to the construction of the neural network, the best results were found with two intermediate layers and 25 and eight neurons on each of them. The network has been tested with different activation and training functions. The best results were achieved with the

sigmoid activation function and the “Resilient Backpropagation” training method. In the 2,779 records collected on training, the neuronal network reaches an average square error of 0.16321. Regarding to the training iterations the prediction model was configured so that it stops automatically right after 100 iterations.

For the significance analysis of single neurons or single data sources, the change of Mean Square Error (MSE) can be used, which determine the significance of individual or groups of neurons by removing them. The results (Table 2) indicate that the weekday, location, temperature and time of the day improve significantly the prediction whereas events like traffic, holidays and rainfall have only a secondary relevance.

Data	Error value	Difference to optimal	Normalized (ordered)
Weekday removed	0,17686	-0,01365	186,3225
Location removed	0,17554	-0,01233	152,0289
Temperature removed	0,17252	-0,00931	86,6761
Time of the day removed	0,16972	-0,00651	42,3801
Month removed	0,16777	-0,00456	20,7936
Vacation time removed	0,16523	-0,00202	4,0804
Rainfall removed	0,16481	-0,0016	2,56
Events removed	0,16188	0,00133	1,7689
Holidays removed	0,16397	-0,00076	0,5776
Traffic state removed	0,16369	-0,00048	0,2304

Table 2: Results of the significance analysis of neurons

4.6 Discussion

First, the systems based on publicly available data will be discussed. These have the advantage that they are less expensive than systems based on parking data. They are not

dependent on a large user base either. They still need real data to train the neural network, however far fewer data is required in comparison with the systems based on user's data. Furthermore, the training data can also be derived from the past. By means of the presented feedback function, the system ensures that the training dataset is continually expanding. Thereby, new pattern in the parking availability are automatically recognized.

The evaluation of the data shows that the time, the location and the weather have the greatest impact. Interestingly, these data are relatively easy to obtain. The time - with its dimensions of week, time, month, holiday season, holiday - can be generated from the current system time. The location is also easy to obtain. It corresponds to the cell in which a user searches for a parking space. The weather, especially the temperature, affects the parking situation significantly and is also easily accessible. Various data providers such as the German Weather Service [DW15] provide this information free of charge.

Unlike the data discussed above, the prediction accuracy is only slightly improved by data of traffic and events. This information is more difficult to access than the data described above and it could not contribute significantly to the quality of to the prediction of the parking situation. Similarly, additional metadata about the events could not be retrieved from the events database in a structured way. For instance, the event's audience size could not be obtained.

Overall, the evaluation indicates that the prediction of the parking situation can be conducted with publicly available information. For each data category, it should be evaluated whether the quality increase of the prediction is worth the effort to integrate the source.

The results of the presented evaluation are limited insofar that despite the active diversification of the collected training data, some data are primarily homogeneous. In this regards, time data can be distributed over several time ranges, though the data collection took place only in the months of July, August and September 2015. However, due to this period, weather data is largely consistent, and does not include low temperatures. For further improvement of the accuracy of the prediction, it is necessary to gather additional training data over a longer period of time, at least of the whole year, and recording days with warm and cold temperatures, as well as peak and off-peak hours. Another limitation is that the model presented has been evaluated only in one city, therefore city-based effects cannot be identified and excluded.

Future research should focus on these limitations. The authors plan to evaluate the categories of data with larger data sets extended to several years and enhanced by explicit user data. Furthermore, future evaluation could compare the different systems used to support the search for parking spaces as this has not yet been done in literature.

This work extends existing research by a classification of systems that support the search for parking spaces. It also shows that there are few approaches based on publicly already available data despite their advantages over other systems. Furthermore, the evaluation of

individual data sources shows that simple data is a good starting point for the prediction of the parking situation.

4.7 Conclusion

This paper has been motivated by the effort needed to search for parking spaces on the roadside nowadays in urban environments. First, existing approaches have been researched and categorized based on the data sources they use: data from parking sensors, from users of the system or from publicly available sources. For the systems based on publicly available data, it is not clear which kind of data should be included, therefore the relevant data categories were identified. Subsequently, a prototypical system was implemented by means of a neural network. This served as a basis for the evaluation of the different categories of data. The results suggest that easily available data – i.e. location, time, weather – have the biggest influence for the accuracy of the prediction. This contribution therefore serves as a basis for future systems that predict the parking situation more efficiently.

Acknowledgements

We thank the German Federal Ministry for Economic Affairs and Energy for funding this research as part of the project 01MD15001D (ExCELL).

5 References

- [Ba05] Banister, D. (2005). *Unsustainable Transport: City Transport in the new Century*. New York: Routledge.
- [CSR12] Chen, X., Santos-Neto, E., & Ripeanu, M. (2012). Crowdsourcing for on-street smart parking. Paper presented at the Proceedings of the second ACM international symposium on Design and analysis of intelligent vehicular networks and applications, Paphos, Cyprus.
- [Ch15] Chikaraishi, M., Fujiwara, A., Kaneko, S., Poumanyvong, P., Komatsu, S., & Kalugin, A. (2015). The moderating effects of urbanization on carbon dioxide emissions: A latent class modeling approach. *Technological Forecasting and Social Change*, 90, Part A, 302-317.
- [CM16] City of Munich. (2016). *Parking Zones, Residential Parking Permits*. Retrieved from <http://www.muenchen.de/int/en/traffic/cars-and-traffic/parking-zones.html>
- [DW15] Deutscher Wetterdienst. (2015). *Klimadaten - online - frei*. Retrieved from <http://www.dwd.de/>
- [DRH11] Doan, A., Ramakrishnan, R., & Halevy, A. Y. (2011). Crowdsourcing systems on the

- World-Wide Web. *Communications of the ACM*, 54(4), 86-96.
- [Ev15] eventful. (2015). Concert & Event API. Retrieved from <http://api.eventful.com/>
- [GE15] General Electric. (2015). Intelligent Environments for Cities. Retrieved from <http://www.gelighting.com/LightingWeb/na/intelligent-environments/intelligent-cities/index.jsp>
- [GST12] Giuffrè, T., Siniscalchi, S. M., & Tesoriere, G. (2012). A Novel Architecture of Parking Management for Smart Cities. *Procedia - Social and Behavioral Sciences*, 53, 16-28.
- [Gr15] Greengard, S. (2015). Between the lines. *Communications of the ACM*, 58(6), 15-17.
- [GH13] Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. 37(2), 337-355.
- [He08] Heaton, J. (2008). *Introduction to Neural Networks with Java* (2nd ed.). Chesterfield: Heaton Research Inc.
- [He04] Hevner, A., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS quarterly*, 28(1), 75-105.
- [In14] Inci, E. (2014). A review of the economics of parking. *Economics of Transportation*, 4(1-2), 50-63.
- [In15] Inrix. (2015). BMW N.A. Becomes Latest Auto Maker to Select INRIX. Retrieved from <http://inrix.com/bmw-n-a-becomes-latest-auto-maker-to-select-inrix/>
- [KMP12] Kimms, A., Maassen, K. C., & Pottbäcker, S. (2012). Guiding traffic in the case of big events with spot checks on traffic and additional parking space requirements. *Central European Journal of Operations Research*, 20(4), 755-773.
- [LG06] Le Fauconnier, A., & Gantelet, E. (2006). The time looking for a parking space: strategies, associated nuisances and stakes of parking management in France. Paper presented at the Proceedings of the European Transport Conference (ETC) 2006, Strasbourg, France.
- [Ma10] Mathur, S., Jin, T., Kasturirangan, N., Chandrasekaran, J., Xue, W., Gruteser, M., & Trappe, W. (2010). ParkNet: drive-by sensing of road-side parking statistics. Paper presented at the Proceedings of the 8th international conference on Mobile systems, applications, and services, San Francisco, California, USA.
- [Mc13] McNeal, M. (2013). If Parking Spots Could Talk. *Marketing Insights*, 25(3), 6-7.
- [MI15] Microsoft. (2015). Choose your Bing Maps API. Retrieved from <http://www.microsoft.com/maps/choose-your-bing-maps-API.aspx>
- [MAZ15] Moeinaddini, M., Asadi-Shekari, Z., & Zaly Shah, M. (2015). An urban mobility index for evaluating and reducing private motorized trips. *Measurement*, 63, 30-40.
- [NNC14] Nandugudi, A., Ki, T., Nuessle, C., & Challen, G. (2014). PocketParker: pocket sourcing parking lot availability. Paper presented at the Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Seattle, Washington.

- [NEM13] Nawaz, S., Efstratiou, C., & Mascolo, C. (2013). ParkSense: a smartphone based sensing system for on-street parking. Paper presented at the Proceedings of the 19th annual international conference on Mobile computing & networking, Miami, Florida, USA.
- [OW15] openweatherAPI. (2015). API - für Ihre Website. Retrieved from http://www.wetter.com/apps_und_mehr/website/api/dokumentation/
- [PM15] ParkMünchen. (2015). Parkplatz-App von ParkMünchen. Retrieved from <http://www.parkmuenchen.de/>
- [PA15a] ParkNav. (2015). ParkNav: Street Parking Reinvented. Retrieved from <http://www.parknav.com/>
- [PA15b] Parkonaut. (2015). Parking Becomes Simple & Smart. Retrieved from <http://www.parkonaut.de/>
- [PA15c] ParkoPedia. (2015). Helping you Park! Retrieved from <http://www.parkopedia.de/>
- [PA15d] ParkTAG. (2015). See Which Space Vacates. Retrieved from <http://parktag.mobi/>
- [Pe07] Peffers, K., Tuunanen, T., Rothenberger, M., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45-77.
- [Ra14] Rakowska, A., Wong, K. C., Townsend, T., Chan, K. L., Westerdahl, D., Ng, S., . . . Ning, Z. (2014). Impact of traffic volume and composition on the air quality and pedestrian exposure in urban street canyon. *Atmospheric Environment*, 98, 260-270.
- [Sc97] Scherer, A. (1997). *Neuronale Netze: Grundlagen und Anwendungen*. Braunschweig: Vieweg.
- [Se08] Seong-Eun, Y., Poh Kit, C., Taehong, K., Jonggu, K., Daeyoung, K., Changsub, S., . . . Byungtae, J. (2008). PGS: Parking Guidance System based on wireless sensor network. Paper presented at the 3rd International Symposium on Wireless Pervasive Computing.
- [SJ14] Shin, J.-H., & Jun, H.-B. (2014). A study on smart parking guidance algorithm. *Transportation Research Part C: Emerging Technologies*, 44, 299-317. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0968090X14001077>
- [Sh06] Shoup, D. C. (2006). Cruising for parking. *Transport Policy*, 13(6), 479-486.
- [SI15] Siemens Mobility. (2015). Vision der intelligenten Stadt – Bewegung durch Innovation. Retrieved from <http://www.mobility.siemens.com/mobility/global/de/nahverkehr/strassenverkehr/integrated-smart-parking-solution/seiten/integrated-smart-parking-solution.aspx>
- [SP15] SmartParking. (2015). Smart Parking - A global parking business. Retrieved from www.smartparking.com
- [Su14] Sujith, S. M., Yacine, A., Quan, Z. S., & Zakaria, M. (2014). Building sustainable parking lots with the Web of Things. *Personal Ubiquitous Computing*, 18(4), 895-907.

- [YLW03] Yang, Z., Liu, H., & Wang, X. (2003). The research on the key technologies for improving efficiency of parking guidance system. Paper presented at the Intelligent Transportation Systems, 2003. Proceedings. 2003 IEEE.
- [Ze07] Zehe, W., Jinggang, Y., Jiangtao, L., & Xiuhua, Z. (2007). Study on the Control Strategy of Parking Guidance System. Paper presented at the International Conference on Service Systems and Service Management, 2007.
- [ZEH98] Zhang, G., Eddy Patuwo, B., & Y. Hu, M. (1998). Forecasting with artificial neural networks:: The state of the art. *International Journal of Forecasting*, 14(1), 35-62.
- [ZB13] Zhang, K., & Batterman, S. (2013). Air pollution and health risks due to vehicle traffic. *Science of The Total Environment*, 450–451, 307-316.

A Concept of Crowdsourced Delivery for Small Local Shops

Maximilian Schreieck¹, Christoph Pflügler¹, Christoph Dehner², Samit Vaidya², Simon Bönisch², Manuel Wiesche¹, and Helmut Krcmar¹

Abstract: Small and medium-sized companies, in particular small local shops, suffer from increased competition via e-commerce channels, e.g. by Amazon and its same day delivery services. Consequently, local shops need delivery options to remain competitive. However, available delivery options such as express delivery are too expensive. A crowd-based delivery platform that leverages free capacity of cars from craftsman and nursing services as well as individuals can be one solution to organize affordable deliveries for small local shops. In this paper, we develop a first version of such a transportation platform within an action research study. Based on the early version of the platform, we discover smart data related challenges that need to be solved to establish the platform successfully. These challenges cover (1) identifying matches between the drivers' routes and transportation requests within a large data set, (2) governing the different parties participating in the platform ecosystem, (3) incentivizing users to join to overcome the "chicken or egg" problem, (4) establishing trust among platform users, and (5) ensuring privacy of user data. The solutions to these challenges will not only facilitate the implementation of a transportation platform for small local shops, they will also contribute to other platform-based business.

Keywords: crowd-sourced delivery, local shops, transportation, goods, action design research, agile development.

1 Motivation

With big online retailers like Amazon and even originally brick-and-mortar retailers like Walmart offering their goods online, the sales in e-commerce are growing from year to year. This can also be seen in figure 1, which even forecasts nearly a doubling in global business-to-consumer (B2C) e-commerce sales from 1233 billion dollars in the year 2013 to 2356 billion dollars in the year 2018 [SDC16].

¹ Technische Universität München, Lehrstuhl für Wirtschaftsinformatik, Boltzmannstraße 3, 85748 Garching, {maximilian.schreieck|christoph.pfluegler|wiesche|krcmar}@in.tum.de

² Technische Universität München, Arcisstraße 21, 80333 München, dehner@in.tum.de, {samit.vaidya|simon.boenisch}@tum.de

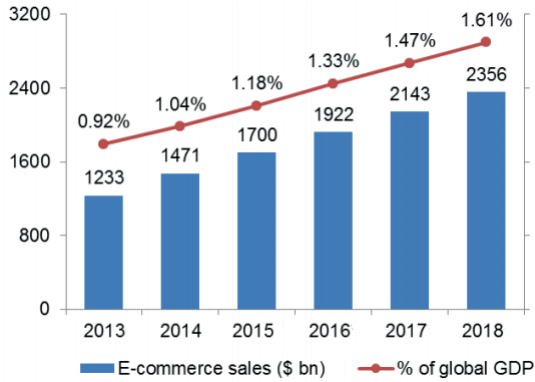


Figure 1: Global B2C e-commerce sales and percentage of GDP (in billion U.S. dollars) [SDC16]

Besides from the absolute numbers rising, the graphic also shows that the sales in e-commerce grow faster than the global gross domestic product (GDP), resulting in an increase of the percentage of the e-commerce sales relative to the GDP. This means that the importance of e-commerce for a B2C business is rising. At the same time, the delivery times of products are getting shorter. An example for this is Amazon: After offering next-day delivery in 2007, in some metropolitan areas the company is now even offering same-day delivery for goods that were ordered in the forenoon [Am15]. To compete with the big online retailers, small- and medium-sized businesses and in particular small local shops also need to be able to offer quick deliveries to the end customer. This however poses a problem to smaller businesses as maintaining an appropriate delivery network is cost-intensive.

At the same time, a huge number of cars are driving through the city and still have capacity for transporting goods. Individuals but also companies that have business processes within cities (e.g. craftsman or nursing services) could leverage these free capacities by accepting delivery requests from local shops organized by a transportation platform. The platform would allow drivers to reduce the costs of their trip by utilizing otherwise empty space in their vehicle while at the same time allowing shop owners or others to perform quick deliveries at an economically reasonable price. This idea appears especially promising due to a similar concept being used successfully in the food delivery market. Businesses like foodora offer food delivery for restaurants that do not have an own delivery service and cannot deal with the logistical overhead of having one [Fo16]. This business model seems to work for foodora, as the company's revenues are rising every year and the enterprise is currently expanding into new markets.

In this research-in-progress study, we chose an action design research approach following Sein et al. [SH16] to develop and evaluate a first version of a crowd-based delivery platform. By evaluating the first prototype, we identify challenges related to the smart data context that need to be solved in the subsequent progress of the research project. These challenges comprise smart algorithms but also aspects such as platform

governance and privacy.

2 Logistics of Small Local Shops

Before proposing a research method, a concept and outcome, this chapter gives some background information about the situation of small local shop owners at the moment. A focus will be the question, how they cope with the increasing importance of e-commerce.

2.1 E-Commerce

The importance of e-commerce is rapidly increasing. Seidenschwarz et al. [Se14] give evidence that the amount of e-commerce has increased significantly and is prognosticated to growing further. In Germany, a large share of the e-commerce business is held by eBay and Amazon. However, the major part is not processed by the companies themselves. As a response to the increasing importance of e-commerce, a lot of midsize companies use the marketplaces of eBay and Amazon to sell their products online.

In future, the amount of e-commerce will continue to increase. Ewert, Kallenbrunnen [EK14] analyze the potential of e-commerce and present auspicious results. In combination with optimized product portfolios and an appropriate degree of specialization e-commerce contains great potential in some branches. Based on that they predict an increase in e-commerce, if businesses seize this relatively new and promising though very technical business area. Such a structural change in the sales process has an effect to various sectors. The research approach of this paper focuses on the consequences and challenges of increasing e-commerce for a delivery system of local shop owners.

2.2 Delivery Options

When it comes to delivery tasks, small and medium-sized businesses and in particular local shops have different options. The most intuitive possibility is to deliver goods on their own. However, such businesses often are too small to implement an efficient delivery network. They cannot leverage economics of scale. As a consequence, delivery is inefficient and time consuming and causes unbearable disadvantages and costs. A solution in this situation could be to outsource the delivery by hiring a professional delivery company. An advantage of this strategy is that it shifts the responsibility and effort to another party and is therefore easy to use for a small shop. However, this comes along with either too long delivery times or comparatively high costs. Standard package delivery e.g. is cheap but takes too long to compete with offers like Amazon's premium delivery. Express delivery is a really fast service but in comparison very expensive.

A third new and innovative alternative for this problem would be crowd-based delivery platforms. Such structures have been approved in the food delivery business and could be transferred to general transportations as well. UberRush [Ub16] represents one first example of a similar crowd-based platform in the New York City Area, San Francisco and Chicago. Following the third approach with a crowd-based solution, this paper analyzes aspects and challenges of such a delivery platform in Germany.

3 Research Method

The project was executed according to Sein et al.'s methodology action design research (ADR) from 2011 [SH16]. The artifact that was aimed at is a software architecture for implementing a platform for crowd-based goods transportation sharing. However, this study comprises only a first version of the platform, as the user requirements for such a platform are not yet completely explored and therefore additional steps have to be taken before the final artifact can be produced. This is also the reason why Action Design Research was the chosen research approach, as it contains a generic schema that features a research process which uses early end-user tests to refine the initial requirements. The schema can be seen in figure 2.

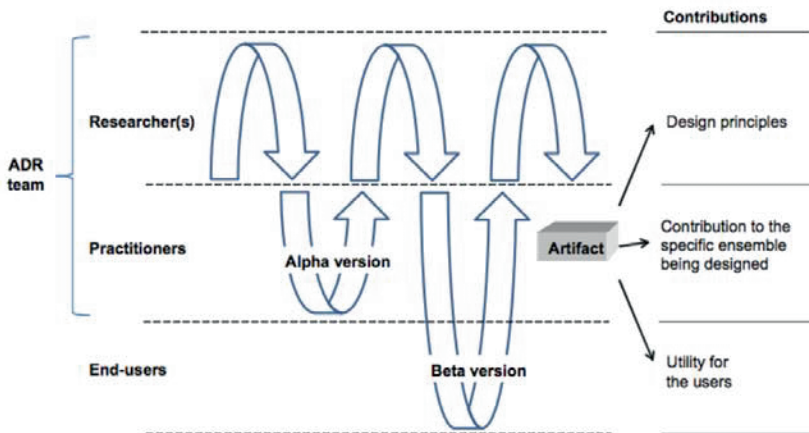


Figure 2: The generic schema for IT-dominant BIE (building, intervention and evaluation) [SH16]

Therefore, the aim of this paper was the development of an alpha version of a prototype application which features the basic functionality for a crowd-based delivery platform. By implementing the platform, we built on an existing solution for ride-sharing that was developed as ride-sharing service for students at Technical University of Munich and has been provided as open source project [Tu16].

The final prototype will then be taken into lab and/or field studies to assess the user acceptance and usability, gather feedback and find out about advanced feature requests.

Based on that information the application is refined and after multiple design cycles, a final application could be developed in a next step, based on the source code of this prototype. As the requirements were not fixed before starting the development, an agile development model was chosen to allow for changes in requirements during the development phase. The method chosen was Scrum, as it is one of the major agile development models, has a big community that can provide additional support and some team members already had experience with it. To follow the procedure of Scrum, we started with getting the requirements right without deviating from the core idea. These requirements were translated into user stories and put in the product backlog for later assignment. Since the Scrum method is based on the concept of sprint based development, we had a development period of two weeks for each assigned user story. During the sprint meetings each user story would be given an estimation point which would represent the time and complexity that it would take for the changes to be realized. After every two weeks (sprint period), a scrum meeting was organized where every team member would report on what they were able to achieve since the last sprint, what they will be doing during the upcoming sprint and if there were any blockers for their development tasks. In the course of these scrum meetings, the requirements would be discussed and changes to the requirements or new requirements would be added to the product backlog.

4 Description of Concept

The main purpose of the software application developed in this project is to provide small and medium-sized businesses, in particular small local shops, with an opportunity to transport their goods in an optimal way via the communication between them and potential transportation drivers (individuals or businesses such as craftsman or nursing services). There are many features to ease the communication between the two parties. For the explanation of the features we use the following terms: requester (the person who creates a transportation request) and transportation driver (who can choose to accept the transportation request created by the requester). The software application we developed consists of two parts, one for each user group: A web application targets the requester (e.g. owners of small local shops) while a mobile application (iOS) targets the transportation driver (people who deliver the packages). The main features are as follows:

Both the iOS and the web application require the registration for both the requester as well as the transportation driver. The registration feature helps the user to be affiliated with the application as a requester or a transportation driver. This process includes providing the basic details of the user. Once the user is registered, then he/she will get an email as a confirmation for registration and a link to verify their account. The verification feature makes sure only the verified users are allowed to use the features provided by the applications. If the requester has a verified account, then the user can login with his/her credentials with the web application and add a transportation request

that meets his/her requirements. The users can add details about their transportation request which includes important details like the dimensions of the transportation goods they want to send, the pickup and destination location, the deadline and the amount for the successful delivery. If the transportation driver has a verified account, he/she can login with his/her credentials with the iOS application and view the transportation requests nearby as well as details regarding the request. The user has the choice to view the ratings of the requester and to accept the request if it meets his/her requirements. Once the transportation request posted by the requester is accepted by the transportation driver, a notification email will be sent to the requester. This will confirm that the transportation driver has accepted to deliver the good. The whole workflow of the package delivery can be viewed in figure 3.

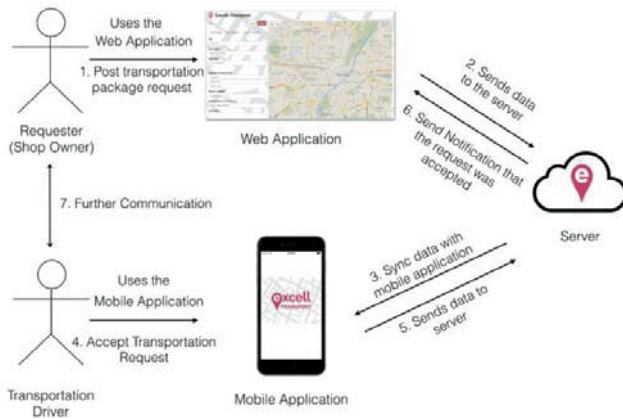


Figure 3: Workflow of the package delivery

5 Discussion and Challenges for Future Research

Having seen the conceptual background and the features of this transportation platform, we will discuss the potential and challenges of this solution. Judging the potential of an innovation is most commonly difficult in advance. However, there are arguments indicating great potential for a crowd-based transportation platform on a city level. Delivery of goods is gaining more and more importance resulting in shorter delivery cycles and complex logistic networks. Same day delivery is a new service, which is used by companies to outperform competitors. Whereas companies like Amazon are able to offer same-day-deliveries to its customers due to their size, smaller local competitors are not. By cooperating in a crowd-based transportation platform, small and medium-sized businesses and in particular small local shops can compensate this inequality.

Another circumstance supporting the potential of a crowd-based transportation platform is the great number of start-ups with similar business ideas. Tech.co [Te15] for example

presents the five international start-ups Entrusters, PiggyBee, Packmule, Shipizy and Jib.li focussing on crowdsourced shipping solutions. The many different start-ups and the diverse set of approaches can be taken as evidence that there is potential in crowd based transportation solutions and a lot of people have already recognized this potential. Compared to the mentioned start-up, our solution follows a similar approach, but has a different orientation. The focus on short deliveries within the radius of one city and on the German market is a new alignment and generates therefore potential, which has not yet been covered.

Opposed to the potential of the solution, several challenges need to be overcome to develop a successful crowd-based-delivery platform. These challenges are related to the fact that a delivery platform gathers data to coordinate delivery processes via a platform. The challenges are summarized in table 1.

First, a **smart matching algorithm** will be necessary to match the routes of drivers with transportation requests. The algorithm needs to process a large number of routes and requests in real-time. An inverted-index data structure to store the information of routes can be used to reduce calculation times. More detailed explanations of such an algorithm are given in [Sc16]. It provides an implementation and evaluation of a dynamic and performant ridesharing service, matching ride offers and requests in real-time. In the further course of the study, those algorithmic features will be incorporated into the existing platform.

The second challenge for such a platform to coordinate crowdsourced transportations will be to incentivize a critical amount of users to benefit from **network effects**. With regard to the transportation platform only a big enough number of users guarantees the necessary local coverage of transportation requests and drivers for a successful running. Strategies and important issues in this context are elaborated in [Ei06].

A key step is for example to attract publicity in the first roll-out phase. It is very likely that there will be an overweight of requests compared to transportation offers at the beginning, resulting in unanswered transportation requests. Partnering with inter-trade organizations to guarantee that all transportation requests are met could then help to address the target group effectively and to overcome the time period until a critical number of users is reached.

Another important factor for the success of such a transportation platform is the quality of the UI (User Interface). Only if the layout appeals the users, the platform will cause a positive impression and has the chance of success. A/B-Testing as described in [Jo06] is an effective method to evaluate different UI designs.

Platform governance is the third challenge and comprises how a flourishing platform ecosystem can be created by the platform owner [Ti14]. [Sc16_2] structures current literature about design and governance of platform ecosystems with regard to different perspectives and concepts. The paper identifies the definition of roles, pricing, boundary resources and openness as most relevant issues. In the further development of our

platform, all these topics need to be considered. For example, when deciding on the platforms' pricing, one needs to pay respect to both platform owner and transportation provider and find an agreement that provides enough incentives for the transportation provider but still allows an economical platform operation. Besides that, in particular the degree of openness of the platform and the decentralization of decision processes needs to be balanced with control mechanisms.

Privacy as fourth challenge is critical as users share location data. Privacy therefore needs to be considered already in the implementation phase with a privacy-by-design approach [Sc10]. The seven key principles of privacy of design according to [Ca10] are proactive not reactive, privacy as the default, privacy embedded into design, full functionality, end-to-end security, visibility and transparency and respect for user privacy. Taking this framework as guidance, those principles of privacy by design will be applied to our transportation platform in a next step. A concrete example for that can be found in the context of doing data analysis on user location data in order to optimize the matching process. By hashing personal information at first, anonymity of users during the data analysis phase is ensured. Furthermore, transparency of the data storage and processing is necessary.

Finally, **trust** among the platform users needs to be established as the requester gives his goods to the transportation driver. Therefore, the security and resistance to criminal actions of the transportation platform is essential for the success. Since up to some extend crowd-based approaches always rely on trust in the crowd, this risk cannot be prevented completely. However, approaches to minimise it are for sure possible and range from regulative measures like the mandatory record of an ID card to crowd-based mechanisms like ratings. For delivery goods with high value, an insurance and money return policy could also be introduced. This would make the transportation requester more comfortable by compensating financial losses in the case of improper handling or loss of the delivery good.

No.	Challenge	Possible solutions to be evaluated
1	Smart matching algorithm	<ul style="list-style-type: none"> • Real-time matching • Inverted index structure
2	Leverage network effects	<ul style="list-style-type: none"> • Publicity from the very beginning • Partnering • High quality UI
3	Platform governance	<ul style="list-style-type: none"> • Decentralization • Openness • Control mechanisms
4	Data privacy	<ul style="list-style-type: none"> • Privacy by design • Transparency
5	Trust	<ul style="list-style-type: none"> • Mandatory registration with ID

-
- Ratings
 - Insurance and money return policy
-

Table 1: Challenges for the further development of the transportation platform

6 Conclusion

This project has produced a first prototype of a transportation platform. Based on this current state a two-pronged approach should be adopted. On the one hand, since the current prototype contains all core features of a crowd-based transportation platform the current version of the prototype will be evaluated to decide which features need to be implemented. A follow-up project will therefore be a field study to present the solution to shop owners and other representatives of the target group in Munich. In subsequent interviews and surveys feedback from this target group will be gained. On the other hand, the five challenges identified in this research-in-process study need to be addressed and the solutions can directly be tested in the field study. The outcome can provide valuable information about positive and negative aspects, priorities of intended users and possible fields of improvement. The results do not only contribute to the project at hand but support researchers and practitioners that implement data-driven platform ecosystems.

Acknowledgements

We thank the German Federal Ministry for Economic Affairs and Energy for funding this research as part of the project 01MD15001D (ExCELL).

References

- [Am15] Amazon (2015): <http://amazon-presse.de/Service/Suche/Pressedetail.html?pid=20be0644-afb8-4bcc-809a-8f2e8e40c71c> (accessed 17.04.2016)
- [Ca10] Cavoukian, A. (2010): Privacy by design. The 7 Foundation Principles. Implementation and Mapping of Fair Information Practices. Information and Privacy Commissioner of Ontario
- [Ei06] Eisenmann, T. R., Parker, G., Alstyne, M. W. V. (2006): Strategies for Two-Sided Markets, Harvard Business Review, October 2006
- [EK14] Ewert, C., Kallenbrunnen, K., (2014): Potenziale des E-Commerce für den Einzelhandel. Hochschule Karlsruhe - Technik und Wirtschaft. merkeZin Heft 5, Februar 2014, http://www.w.hs-karlsruhe.de/markezin/links/Markezin_Heft5_A4Klaus%20Kallenbrunnen.pdf
- [Fo16] Foodora (2016): <http://www.bloomberg.com/news/articles/2016-02-19/delivery-hero->

- s-foodora-expands-in-europe-as-rivalry-intensifies (accessed 17.04.2016)
- [Jo06] Jones, M., Marsden, G. (2006): *Mobile Interaction Design*. Wiley
- [Sc10] Schaar, P. (2010). Privacy by design. *Identity in the Information Society*, 3(2), 267-274.
- [Sc16] Schrieck, M., Safetli, H., Siddiqui, S. A., Pflüger, C., Wiesche, M. Krcmar, H (2016): A Matching Algorithm for Dynamic Ridesharing. *International Scientific Conference on Mobility and Transport Transforming Urban Mobility*, mobil.TUM 2016
- [Sc16_2] Schrieck, M., Wiesche, M., Krcmar, H. (2016): Design and Governance of Platform Ecosystems – Key Concepts and Issues for Future Research. *Twenty-Fourth European Conference on Information Systems (ECIS)*
- [SDC16] Schöder, D., Ding, F. and Campos, J.K. (2016): The Impact of E-commerce Development on Urban Logistics Sustainability. *Open Journal of Social Sciences*, 4, 1-6. <http://dx.doi.org/10.4236/jss.2016.43001>
- [Se14] Seidenschwarz, H., Weinfurtner, X., Stahl, E., Wittmann, G. (2014): E-Commerce-Strategien für den mittelständischen Einzelhandel. Bayerisches Staatsministerium für Wirtschaft und Medien, Energie und Technologie. http://www.stmwi.bayern.de/fileadmin/user_upload/stmwivt/Publikationen/2014/E-Commerce-Strategien_fuer_den_mittelstaendischen_Einzelhandel_ungesichert.pdf
- [SH16] Sein, M.K., Henfridsson, O., Puroo, S., Rossi, M. and Lindgren, R. (2016): Action Design Research. *MIS Quarterly* Vol. 35 No. 1 pp. 37-56
- [Te15] Tech.co (2015): <http://tech.co/international-crowdsourced-shipping-startups-2015-03> (accessed 21.04.2016)
- [Ti14] Tiwana, A. (2014). *Platform Ecosystems: Aligning Architecture, Governance, and Strategy*: Morgan Kaufmann.
- [Tu16] TUMitfahrer (2016): Open Source Github Repository. <https://github.com/Mobility-Services-Lab>. (accessed 17.04.2016)
- [Ub16] UberRush (2016): <https://rush.uber.com/how-it-works> (accessed 17.04.2016)

Weighted Aggregation in the Domain of Crowd-Based Road Condition Monitoring

Kevin Laubis¹, Viliam Simko², Christof Weinhardt³

Abstract: This paper focuses on crowd-based road condition monitoring using smart devices, such as smartphones and evaluates different strategies for aggregating multiple measurements (arithmetic mean and weighted means using R^2 and $RMSE$) for predicting the longitudinal road roughness. The results confirm that aggregating predictions from single drives leads to a higher model performance. This has been expected and confirms the intuition. The overall R^2 could be increased from 0.69 to 0.75 on average and the $NRMSE$ could be decreased from 9% to 8% on average. However, contrary to the intuition, the results show that weighted aggregations of single predictions should be avoided, which is consistent with previous findings in other domains, such as financial forecasting.

Keywords: Crowd-based sensing, road condition monitoring, international roughness index, predictive road maintenance, weighted aggregation, ensemble learning

1 Introduction

Road roughness is one of the most important attributes that gives valuable insights into road condition and driving comfort. Thus, metrics such as the International Roughness Index (IRI) [SGQ86] are considered in most pavement management systems for planning cost efficient road maintenance actions. Nowadays, road authorities rely on measurements from special-purpose vehicles equipped with lasers and further highly precise sensors for sensing the road's profile. Furthermore, specially-trained technicians are required for performing these measurements. A bottleneck in assets and human resources leads to measurements performed on a coarse granular temporal basis. In case of the German's federal road network, this results in monitoring intervals of four years. With regard to road maintenance, this leads to a reactive approach, which directs resources towards road segments that already reached a critical condition.

Using smart devices from drivers and passengers, it is possible to measure and analyze vehicle's vibration and thus, to estimate the road's roughness magnitude. This is a promising alternative to the current way of monitoring the road condition. The inertial measurement units (IMUs) in smart devices allows for a near-real-time assessment of road condition. However, the low accuracy of such sensors, versatile suspension systems, different

¹ FZI Research Center for Information Technology, Information Process Engineering (IPE), Haid-und-Neu-Straße 10–14, 76131 Karlsruhe, laubis@fzi.de

² FZI Research Center for Information Technology, Information Process Engineering (IPE), Haid-und-Neu-Straße 10–14, 76131 Karlsruhe, simko@fzi.de

³ Karlsruhe Institute of Technology (KIT), Institute of Information Systems and Marketing (IISM), Fritz-Erler-Straße 23, 76131 Karlsruhe, weinhardt@kit.edu

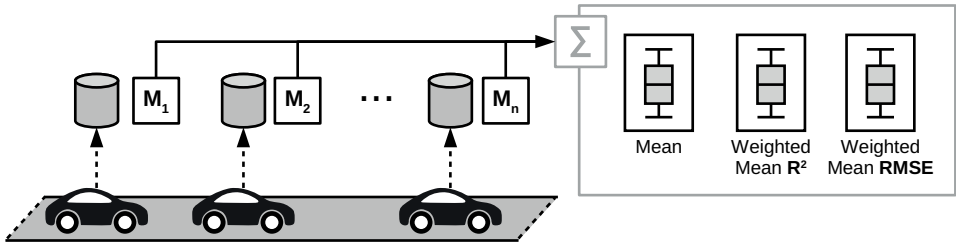


Fig. 1: Outline of aggregating crowd-based road condition measurements from multiple cars

placements of smart devices in the car and other factors lead to a lower prediction accuracy compared to well-calibrated laser-based measurements. For overcoming this decrease in accuracy, multiple approaches for aggregating measurements from several cars are compared in this paper. The outline of this idea is depicted in Fig. 1.

Assuming uncorrelated prediction errors of the single car's predictions, the unweighted mean is expected to reduce the variance component of the errors and thus, increase the prediction accuracy. However, it is not clear to what extent the accuracy can be increased. While it might seem plausible to use weighting schemas based on the model performance when aggregating the results instead of applying a simple arithmetic mean, it has been shown that the arithmetical mean often outperforms a more sophisticated weighting [SW09]. This is true, since a weighted aggregation could increase the prediction error because of increasing the variance component of the error. On the other hand, weighting could reduce the bias component of the prediction error. Thus, it has to be investigated empirically, how weighted aggregation functions perform against the unweighted mean.

For determining, to what extent aggregation of single car predictions can increase the model's accuracy and whether weighting of the single car's prediction is worth an implementation, the focus of this work is to develop and evaluate the extent to increase the performance of crowd-based road roughness estimations by aggregating estimations from multiple cars. We apply unweighted and weighted aggregation methods to a dataset gathered by multiple car drives and laser-based road profile measurements as a ground truth to answer the following research questions:

1. To what extent does the aggregation of crowd-based road roughness measurements from multiple cars increase the model performance?
2. How does the application of weighted aggregation methods instead of unweighted mean affect the model performance?

The paper is structured as follows: The next section 2 summarizes related work. Section 3 describes the data gathered and analyzed from a single car point-of-view followed by the sections 4 and 5, which compare the results from three different aggregation methods. Finally, a conclusion is provided in section 6 and future work is discussed in section 7.

2 Related Work

Investigating road conditions with smart wearable devices or other devices attached to the vehicle, which are able to determine the vehicle's vibrations was goal of several studies [Bh12, Du14, Er08, MPR08, Ni14, Pe11]. They mainly rely on machine learning approaches and on physical models representing a car together with its suspension system ([Ja14]).

The authors of the study [Ni14] show a way of applying supervised machine learning algorithms for predicting the road roughness based on the vertical acceleration provided by high precision sensors. They built models for different road roughness metrics, but did not aggregate multiple measurements.

In [MPR08], the fact that the smart devices could be placed at different locations in the car and with different orientations is considered in addition. This is handled by applying a virtual reorientation of the device's axes. Nevertheless, predictions from single cars were just considered separately.

An approach for determining road segments with single anomalies, such as potholes and bumps was developed in [Pe11]. The authors also applied machine learning algorithms and indicated the single car's performances. An aggregation of multiple measurements was not performed.

A prominent paper in this field is [Er08]. It describes a machine learning approach for also detecting potholes with a fleet of smartphone-equipped taxis. For getting robust results, the pothole candidates from single cars were geo-spatially clustered. However, the performance increase by applying this aggregation function was not investigated.

In none of these studies, the effect of considering measurements from multiple vehicles was investigated. Some of them considered multiple measurements in terms of geo-spatial clustering of singular road anomalies [Er08], but the effect on the overall model performance compared to single measurements was not examined. Since the effect of multiple measurements was not determined yet, likewise the performance of different aggregation functions was not determined in the road condition monitoring domain.

3 Single Car Models

For applying and evaluating different aggregation functions, a set of single-car road roughness prediction models⁴ is considered in this study. Each model, predicts the IRI of road segments based on the built-in IMU sensors in a smartphone attached to the dashboard of a car.

The models consider speed of the car as measured by the GPS module (~ 1 Hz), the 3-axis accelerometer (~ 50 Hz) and the 3-axis gyroscope sensors (~ 50 Hz) of the smartphones. The minimum, maximum, standard deviation, variance and the root mean square

⁴ Prediction models based on random forests [Br01] are considered in this study.

sensor	aggregation	number of features
GPS velocity	min., max., SD, var., RMS	5
accelerometer (3-axis)	min., max., SD, var., RMS, CWT for 5 bands	30
gyroscope (3-axis)	min., max., SD, var., RMS, CWT for 5 bands	30

Tab. 1: Features extracted from the smartphone's IMU sensors

of each 100 m road segment are extracted and considered as features of the prediction model. From the acceleration and gyroscope measurements, the continuous wavelet transformation (CWT) [TC98] is computed using the *biwavelet* R package [GG16]. From the bias-corrected wavelet power spectrum, the wavelengths of 5.5 m, 31 m, 124 m, 351 m and 703 m are selected as additional features. As shown in Tab. 1 this results in a set of 65 features in total.

The car models are trained and evaluated on a 4.8 km road link for which the actual road profile is provided by the Institute of Highway and Railroad Engineering at the Karlsruhe Institute of Technology (ISE/KIT). Seven independent car drives are performed and their measurements are matched to the ground truth road roughness for training and testing the corresponding seven prediction models. The models are tuned separately for increasing the coefficient of determination R^2 . Next to this performance metric, the root mean square error ($RMSE$) and its normalization to the spread of all IRI values measured in this study ($NRMSE$) are determined. The $NRMSE$ is defined as follows:

Def. 1 (Normalized root mean square error). *Let $RMSE \in \mathbb{R}^+$ be the root mean square error of a prediction model. Let $y_{max} \in \mathbb{R}^+$ and $y_{min} \in \mathbb{R}^+$ be the maximum and minimum of the actual laser-based measured IRI values of the considered road link.*

The normalization of the $RMSE$ is defined as follows:

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}}$$

The $NRMSE$ was chosen in addition to the $RMSE$ for better interpretability of the results with regard to its implications. Since both values are proportional, a particular weighting leads to the same performance of the aggregation functions. Thus, the $RMSE$ is chosen as the weight in section 4 and the $NRMSE$ values are considered for discussing the results in section 5.

The performance metrics of the single car models considered in this study are shown in Tab. 2. The last four rows provide summary statistics of the performance metrics for an easy comparison with the performances of the aggregated methods described in the following section.

drive	R^2	$RMSE$	$NRMSE$
1	0.6296	0.2063	0.0975
2	0.5947	0.2184	0.1022
3	0.6604	0.2089	0.0977
4	0.7607	0.1697	0.0794
5	0.7899	0.1601	0.0749
6	0.6967	0.1931	0.0903
7	0.7240	0.1924	0.0900
max.	0.7899	0.2184	0.1022
mean	0.6937	0.1927	0.0902
median	0.6967	0.1931	0.0903
min.	0.5947	0.1601	0.0749

Tab. 2: Out of sample performance of single predictions

4 Aggregation Methods

The aggregation of multiple single car predictions is performed in basically two different ways. The baseline aggregation function is an unweighted arithmetic mean of the single predictions. It is formally defined in the following Def. 2.

Def. 2 (Aggregation by unweighted mean). *Let M_1, \dots, M_n be prediction models. Each model M_i is a function that maps an n -dimensional feature vector $x \in \mathbb{R}^n$ to a real outcome, i.e. $M_i : \mathbb{R}^n \mapsto \mathbb{R}$.*

The predictions $\forall i = 1, \dots, n : M_i(x) = \hat{y}_i$ can be combined using a simple arithmetic mean $\bar{M}(x)$ as follows:

$$\bar{M}(x) = \frac{\sum_{i=1}^n \hat{y}_i}{n}$$

Def. 3 (Aggregation by weighted mean). *Let M_1, \dots, M_n be prediction models with corresponding weights $W = \{w_1, \dots, w_n\}$. Each model M_i is a function that maps an n -dimensional feature vector $x \in \mathbb{R}^n$ to a real outcome, i.e. $M_i : \mathbb{R}^n \mapsto \mathbb{R}$.*

The predictions $\forall i = 1, \dots, n : M_i(x) = \hat{y}_i$ can be combined using a weighted mean $\bar{M}_W(x)$ as follows:

$$\bar{M}_W(x) = \frac{\sum_{i=1}^n \hat{y}_i w_i}{\sum_{i=1}^n w_i}$$

Now, as weights, the R^2 and $RMSE$ can be used for car models M_1, \dots, M_n which yields aggregation functions: \bar{M}_{R^2} and \bar{M}_{RMSE} . The weighted aggregation functions are applied to determine whether attaching importance to more performant models has a positive effect on the prediction performance of the aggregate.

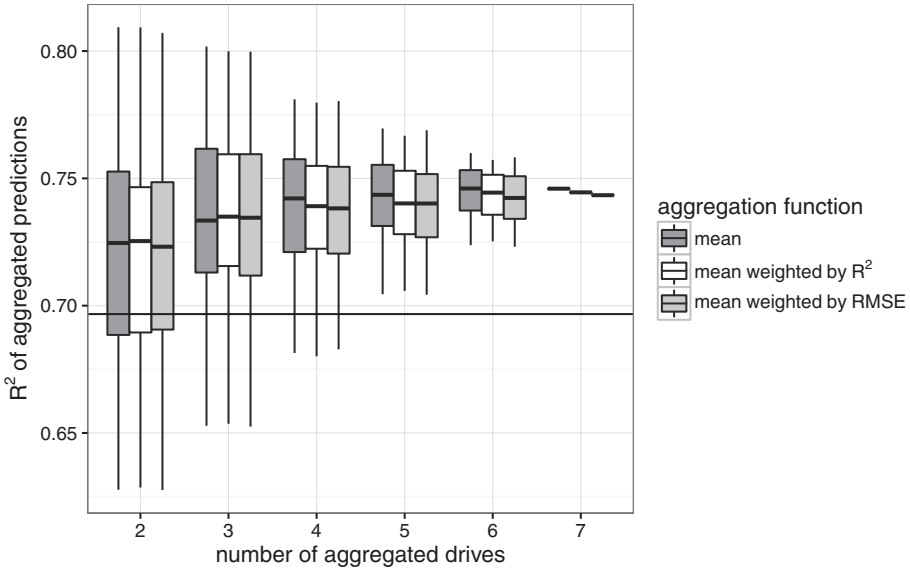


Fig. 2: Out of sample R^2 of aggregated predictions

5 Evaluation

For determining the effect of the aggregation functions and for comparing the performance of the different aggregation functions empirically, they are applied to the models resulting from the drives described in section 3. Since there were $n = 7$ separate drives performed, the combinations from $k = 2, \dots, 7$ drives are considered. This leads to sample sizes of: $\binom{n}{k} = \{21, 35, 35, 21, 7, 1\}$

The distributions of R^2 over the driving combinations are shown in Fig. 2. For each aggregation function (\bar{M} , \bar{M}_{R^2} and \bar{M}_{RMSE}) a separate boxplot is shown.

The median R^2 of single drive’s predictions (see also Tab. 2) is indicated by a horizontal line. Even for combining two drives, all aggregation functions achieve a significantly better median performance (R^2 and $RMSE$) compared to its single drive baseline at a significance level of at least 5%. Increasing the number of considered drives likewise increases the coefficient of determination constantly for each aggregation function. Comparing the performance of the aggregation functions shows that there is just a minor difference between them. However, while considering four and more drives, the unweighted mean outperforms the weighted ones.

A comparison regarding the $NRMSE$ is given in Fig. 3. Similar to the R^2 scenario, a constant decrease of the $NRMSE$ is achieved by considering more drives. Furthermore, the unweighted mean aggregation has a lower median $NRMSE$ than the weighted aggregations except for the case of two drives.

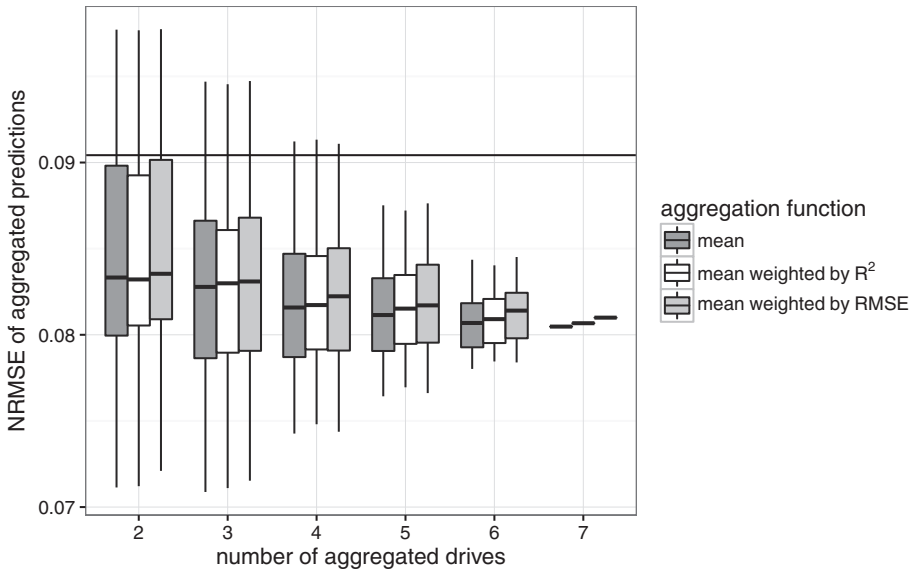


Fig. 3: Out of sample $NRMSE$ of aggregated predictions

The mean R^2 and mean $NRMSE$ over all aggregation combinations for the considered aggregation functions and for the considered number of drives are given in Tab. 3. If there was a significant performance decrease (decrease in R^2 or increase in $NRMSE$) of using a weighted mean aggregation instead of using the unweighted mean aggregation, it is indicated at the corresponding mean performance of the weighted aggregation. Since the first row indicates the baseline with no aggregation, there is no difference between the aggregation functions. Likewise, referring to the number of possible combinations, the seventh row does not contain tests on significant differences in performance. Except for combinations of drives less or equal to three, the mean performance of the R^2 weighted aggregations are worse than the unweighted aggregations. Regarding the mean performance of the $NRMSE$ the unweighted aggregation outperforms the weighted ones for all considered numbers of drives. Even though the absolute differences are minor, the decreases compared to the unweighted mean function are mostly statistically significant even for the small sample sizes. This indicates that applying a weighted aggregation increases the variance error component to a higher extent than decreasing the bias error component. A vertical comparison of the performance metrics provided in Tab. 3 was discussed based on Fig. 2 and Fig. 3.

6 Conclusion

For answering the research questions (1) to what extent the aggregation of crowd-based road roughness measurements from multiple cars does increase the model performance and (2) how the application of weighted aggregation methods instead of unweighted mean does affect the overall performance, three different aggregation methods (arithmetic mean

drives	mean		mean weighted by R^2		mean weighted by $RMSE$	
	R^2	$NRMSE$	R^2	$NRMSE$	R^2	$NRMSE$
1	0.6937	0.0902	0.6937	0.0902	0.6937	0.0902
2	0.7234	0.0846	0.7227	0.0846	0.7223*	0.0849***
3	0.7338	0.0827	0.7327**	0.0828	0.7320***	0.0831***
4	0.7391	0.0817	0.7379***	0.0819**	0.7369***	0.0822***
5	0.7423	0.0811	0.7409***	0.0813**	0.7399***	0.0816***
6	0.7444	0.0808	0.7430**	0.0809*	0.7419***	0.0813***
7	0.7460	0.0805	0.7445	0.0807	0.7434	0.0810

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Tab. 3: Means of R^2 and $NRMSE$ over aggregation combinations for considered aggregation functions (unweighted mean, mean weighted by R^2 and mean weighted by $RMSE$) and for considered number of drives. Significant performance decreases compared to unweighted mean aggregations are indicated.

and weighted means using R^2 and $RMSE$) are applied to seven single car models. Each of the car models is a random forest trained on sensor data collected using a smartphone while driving a distance of 4.8 km.

The results confirm that aggregating predictions from single drives leads to a higher model performance. This has been expected and confirms the intuition. Thus, the R^2 could be increased from 0.69 to 0.75 on average and the $NRMSE$ could be decreased from 9% to 8% on average. In other words, real-time predictive road maintenance gets better with increasing number of participants.

Contrary to the intuition, our results also show that weighting aggregations of single predictions should be avoided. This is consistent with the results of the study [SW09], which describes similar findings in the financial forecasting domain. From a technical point-of-view, this allows a simpler and thus, more efficient implementation.

7 Outlook

It has to be mentioned that all measurements are performed on a homogeneous and recently paved road link. Thus, it is not clear whether the results are valid for other road types and for a wider IRI range. Beside extending the analysis by a broader set of road segments and by further car and sensor types, a step will be to investigate whether the results are reasonable for road conditions other than IRI as well. Furthermore, it is intended to determine how the real-time road condition monitoring affects the road maintenance from a managerial point of view. It has to be investigated to what amount less accurate models could be applied, while still being economically beneficial for road authorities. It should be determined whether single car predictions are already sufficiently accurate for planning maintenance actions and if not, it is going to be determined whether the accuracy increase achieved by aggregation reaches a sufficient accuracy level.

Knowing this economic value of a crowd-based road roughness monitoring for road authorities and for road users a business model can be tailored. Such a model should also encompass incentive mechanisms for motivating drivers to participate in the crowd-based system.

References

- [Bh12] Bhoraskar, Ravi; Vankadhara, Nagamanoj; Raman, Bhaskaran; Kulkarni, Purushottam; Wolverine: Traffic and road condition estimation using smartphone sensors. *IEEE*, pp. 1–6, jan 2012.
- [Br01] Breiman, Leo: Random forests. *Machine learning*, 45(1):5–32, oct 2001.
- [Du14] Du, Yuchuan; Liu, Chenglong; Wu, Difei; Jiang, Shengchuan: Measurement of International Roughness Index by Using Z-Axis Accelerometers and GPS. *Mathematical Problems in Engineering*, 2014:1–10, 2014.
- [Er08] Eriksson, Jakob; Girod, Lewis; Hull, Bret; Newton, Ryan; Madden, Samuel; Balakrishnan, Hari: The pothole patrol: using a mobile sensor network for road surface monitoring. In: *Proceeding of the 6th international conference on Mobile systems, applications, and services - MobiSys '08*. ACM Press, New York, New York, USA, p. 29, 2008.
- [GGS16] Gouhier, Tarik; Grinsted, Aslak; Simko, Viliam: . R package "biwavelet": Conduct univariate and bivariate wavelet analyses, 2016. (Version 0.20.4).
- [Ja14] Jazar, Reza N.: *Vehicle Dynamics*. Springer New York, New York, NY, 2 edition, 2014.
- [MPR08] Mohan, Prashanth; Padmanabhan, Venkata N.; Ramjee, Ramachandran: Nericell: Rich Monitoring of Road and Traffic Conditions using Mobile Smartphones. In: *Proceedings of the 6th ACM conference on Embedded network sensor systems - SenSys '08*. ACM Press, New York, New York, USA, p. 323, 2008.
- [Ni14] Nitsche, P.; Van Geem, C.; Stütz, R.; Mocanu, I.; Sjögren, L: Monitoring ride quality on roads with existing sensors in passenger cars. In: *Proceedings of the 26th ARRB Conference*. Sydney, pp. 1–13, 2014.
- [Pe11] Perttunen, Mikko; Mazhelis, Oleksiy; Cong, Fengyu; Kauppila, Mikko; Leppänen, Teemu; Kantola, Jouni; Collin, Jussi; Pirttikangas, Susanna; Haverinen, Janne; Ristaniemi, Tapani; Riekkki, Jukka: Distributed Road Surface Condition Monitoring Using Mobile Phones. In: *Lecture Notes in Computer Science*, pp. 64–78. 2011.
- [SGQ86] Sayers, Michael W.; Gillespie, Thomas D.; Queiroz, Cesar A V.: *The International Road Roughness Experiment – Establishing Correlation and a Calibration Standard for Measurements*. Technical report, Washington, D.C., 1986.
- [SW09] Smith, Jeremy; Wallis, Kenneth F.: A Simple Explanation of the Forecast Combination Puzzle. *Oxford Bulletin of Economics and Statistics*, 71(3):331–355, jun 2009.
- [TC98] Torrence, Christopher; Compo, Gilbert P.: A practical guide to wavelet analysis. *Bulletin of the American Meteorological society*, 79(1):61–78, jan 1998.

SEO4OLAP – Search Engine Optimized Presentation of Statistical Linked Data

Daniel Breucker¹, Benedikt Kämpgen¹ and Andreas Harth²

Abstract: Statistical data is published online by a variety of organizations, such as *Eurostat*, the statistical office of the European Union. Some of these datasets are available as Linked Data which has great potential for future research or commercial exploitation. Yet, search engines are struggling to index statistical Linked Data and therefore, normal web users have difficulties to access it. In this paper, we present the approach *SEO4OLAP*, which generates search engine optimized webpages for every possible view on statistical datasets. Statistical datasets are usually based on a Multidimensional Model (MDM) and can therefore be queried by means of Online Analytical Processing (OLAP). We present a new OLAP query model which allows queries to be represented by clean URIs. We also present a mathematical model to compute the overall number of possible views of a dataset. We evaluate our approach by publishing two datasets from *Eurostat* with our *SEO4OLAP* system. We observe that our webpages are indexed by search engines and conclude that the system can have benefits for data publishers, web users and search engine providers.

Keywords: Search Engine Optimization (SEO); Online Analytical Processing (OLAP); Statistical Linked Data

1 Introduction

Statistical data is published online by a variety of organizations, among them public agencies and governmental institutions. The statistical office of the European Union *Eurostat* is one of these institutions. Eurostat publishes statistics such as about trade, population, economy and finance inside the EU. The data can be accessed by web services in the SDMX format, downloaded directly or explored by an online pivot table.

Recent work ([KOH12], [Sa12]) has focused on leveraging statistical data on the Web by means of Semantic Web technologies in conjunction with Online Analytical Processing (OLAP). This has been fostered by the W3C, since they recommended the RDF Data Cube Vocabulary (QB) for modelling multidimensional data (especially SDMX) in the Resource Description Framework (RDF).

Based on the current trend, we assume that in the near future organizations are going to publish statistical Linked Data directly. For the moment, Linked Data wrappers exist for

¹ FZI Research Center for Information Technology, Haid-und-Neu-Str. 10-14, 76131 Karlsruhe, breucker@fzi.de, kaempgen@fzi.de

² Institute AIFB, Karlsruhe Institute of Technology, Englerstr. 11, 76131 Karlsruhe, harth@kit.edu

some of those statistical datasets. In the case of *Eurostat*, the corresponding Linked Data wrapper is *Estatwrap* by *OntologyCentral*³.

Linked Data technologies and the RDF Data Cube Vocabulary allow to publish statistical data in a standardized format, interlink different datasets and retrieve information with the standard query language SPARQL. This has great potential for many different fields. Empirical experiments could be published and reused by other research parties. New services on open data may evolve, leading to new business opportunities.

Despite the potential of statistical Linked Data, human friendly interfaces to explore and interact with it are still suspect to ongoing research ([Ho13], [Mu14]). The SPARQL query language is the most common way to query Linked Data. Since SPARQL can only be used by experts in the field of semantic technologies, the information published in the Semantic Web is not available for usual web surfers.

A common way to retrieve information from the Web is the usage of search engines. Since search engines are focusing on HTML-based content, Linked Data is usually not included in search results. Neither are SDMX representations and views created using JavaScript pages.

This leads to the following problem: Even though statistical data is published online, search engines are not able to retrieve corresponding results. In this paper, we analyse how statistical Linked Data can be published to allow search engines to properly index statistical datasets. We present our approach *SEO4OLAP* which generates search engine optimized (SEO) webpages for all possible facts of a dataset. Once these landing pages are published, search engines should be able to crawl the content. From this approach the following research questions can be derived and will be answered:

- Is SEO4OLAP improving the status quo?
- What is the computation complexity of this problem? How many webpages are created depending on the number of dimensions, measures and dimension members?
- How can semantic mark-up technologies such as *Schema.org*⁴ contribute to this?

The remainder of this invited⁵ paper is structured as follows. Section 2 presents our approach *SEO4OLAP*, including both the abstract design of the system and a mathematical model to compute the total number of possible views. In section 3, the approach is evaluated. We implemented a concrete system in Java and published two datasets from *Eurostat*. We compare how search engines rank our pages in comparison to the original data source from *Eurostat*. After a discussion of our findings in section 4, we present related work in section 5. Finally, a conclusion with a discussion on future work is

³ <http://estatwrap.ontologycentral.com/>

⁴ <http://schema.org/>

⁵ This is an invited contribution from the BigGIS project for the BDSST 2016 workshop.

presented in section 6.

2 Approach – Creating Landing Pages for Cubes

The goal of *SEO4OLAP* is to generate search engine optimized webpages for every possible view of a statistical linked dataset modelled in the RDF Data Cube Vocabulary. The schematic process of *SEO4OLAP* is illustrated in Figure 1. The system receives a dataset as input, analyses the data cube schema, computes all possible OLAP-queries and generates webpages for every result with corresponding keywords.

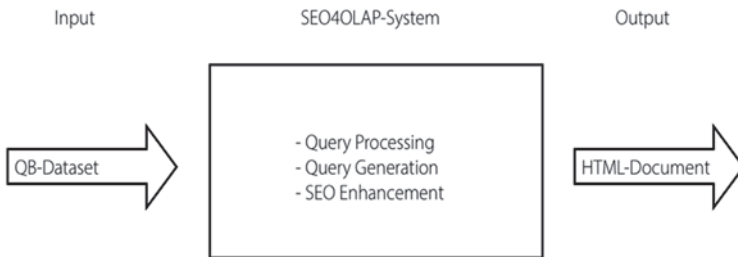


Fig. 1: Schematic process of *SEO4OLAP*

Facts from data cubes can be retrieved through OLAP-queries. Since these facts are represented as a webpage, queries are performed via HTTP GET requests. Therefore, we developed a new OLAP query model which can be represented by pretty URLs.

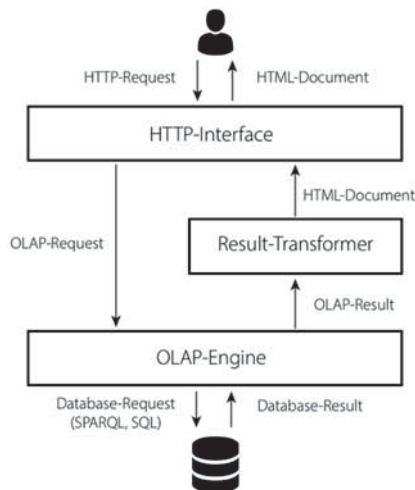


Fig. 2: Query processing of *SEO4OLAP*

The query processing workflow is illustrated by Figure 2. Incoming OLAP-queries are transformed into a series of OLAP-operations, a so called *Logical OLAP Query Plan*. The plan is executed by an OLAP-engine which transforms the OLAP-operations into native database queries, e.g. using SPARQL. Database results are then enhanced with corresponding keywords, e.g. the labels of the selected measure and the diced dimension members. The result is then returned as an HTML page.

The URLs of the facts have to be made accessible to humans and crawlers in order to be found by search engines. This can be achieved by applying a link structure which allows to reach every fact of the dataset. We recommend to add links from fact-webpages to near neighbours. A near neighbour is an OLAP-query with one different parameter. Another possibility to help search engines crawling all webpages is to provide a sitemap. A sitemap is a list of all URLs of a website.

There are various ways to define OLAP-queries and express them with query languages. As an example, MDX allows to specifically define how OLAP-results should be displayed in a pivot table. For our use case, this is too complex. Search engines prefer clean and readable URLs for their ranking. Therefore, we developed a new query model which is presented in the following. First, the parameters are described and afterwards used in our URL-scheme.

We adapt the concept of subcube queries from Kämpgen et al. [KOH12] and define it as per Definition 1.

Definition 1 (Subcube Query): *A subcube query on a certain cube is represented as a tuple (Measures2Project, Dimensions2Keep, Members2Dice), with Measures2Project consisting of identifiers of measures to be projected, Dimensions2Keep consisting of identifiers of dimensions to be kept and Members2Dice consisting of identifiers of members to be diced. Dimensions that are not represented by a member in Members2Dice or part of Dimensions2Keep, are sliced.*

We propose this model because we believe it is a good fit for our use case of SEO. The main advantage is that instead of all dimensions, only relevant parameters have to be set. This means that the query directly defines which measures, dimensions and members should be retrieved. In comparison to Kämpgen et al. [KOH12], less parameters are needed.

As an example, consider a query on an employment dataset asking for the absolute employment number and the employment rate of Germany per year, disregarding the gender. A subcube query for this request would be the following: (`{employment absolute, employment rate}`, `{date}`, `{germany}`). We assume, that the dimension *Date* only has members on the level year. The dimension *Gender* is sliced, since neither men or women are defined as *Member2Dice* nor is it part of *Dimensions2Keep*.

A subcube query as defined per Definition 1 can be submitted via HTTP by setting the cube identifier and the three parameters *Measures2Project*, *Dimensions2Keep* and *Members2Dice*. A typical HTTP GET request for our scenario has the following structure:

```
http://baseUri/ep?cube=id1&measure=id2&dimension=id3&member=id4
```

An API based on this scheme would work just fine. The problem is that search engines prefer clean and readable URLs consisting of keywords. Therefore, we developed a URL-scheme which enables the same functionality but has a clean appearance with keywords used as identifiers. It is presented in the following:

```
http://baseUri/cubeId/pattern/id1/id2/...
```

The pattern consists of three digits, defining the number of used parameters per group. The following identifiers are slash-separated and in the order of the pattern. As an example, the pattern 122 means that the first identifier is a measure, the following two are dimensions and the last two are members. In the following some example URLs are presented for better understanding.

- <http://example.org/employment/211/absolute/rate/date/germany>:
A URL for our previous example.
- <http://example.org/employment/112/rate/date/women/france>:
A URL asking for the employment rate of women in France per year.
- <http://example.org/population/111/absolute/date/poland>:
A URL asking for population numbers in Poland per year. This is a different data cube than the previous.

A cube's size and thus the number of possible views is exponentially dependent on the number of dimensions. The maximum number of possible views can be computed with Formula 1. Every dimension has d_i possible members plus 2 extra values: the implicit ALL-member, which is a slice and the implicit Zero-Member, when no value is selected; m is the number of measures, n is the number of dimensions. The meaning of d_i , m and n apply for all formulas in this chapter.

$$MaxViews = m \times \prod_{i=1}^n (d_i + 2) \quad (1)$$

Formula 1 underlies the following restrictions:

- Only one measure is displayed per view. It has to be noted that this already reduces the potential overall number of views a lot.
- A dimension can be set to one member, be sliced (All-Member) or not set (Zero-Member). Multiple members of the same dimension are not regarded.

- Implicit aggregated members of higher levels are neglected. Only explicit members are regarded.

Due to the exponential growth of the problem, the computational effort to generate separate webpages for every possible view can be enormous. From a SEO and a user experience perspective, it is questionable whether all possible views are necessary. Therefore, we propose two restrictions to our model:

- The *Dice Dimensionality* (DiceDim) restricts the maximum number of dimensions which can be diced.
- The *Number of free Dimensions* (FreeDim) defines the maximum number of dimensions that are neither diced nor sliced, thus free.

Depending on these two restrictions, the total number of possible views can be computed by Formula 2.

$$Views_{DiceDim,FreeDim} = m \times \sum_{d=0}^{DiceDim} (FreeDimFactor_{FreeDim} \times DiceDimFactor) \tag{2}$$

with

$$FreeDimFactor_{FreeDim} = \begin{cases} \sum_{s=0}^{FreeDim} \frac{(n-d)!}{s!(n-d-s)!} & \text{with } \frac{(n-d)!}{s!(n-d-s)!} = 0 \\ & \text{for } (n-d-s) < 0 \end{cases} \tag{3}$$

and

$$DiceDimFactor = \begin{cases} 1 & \text{if } DiceDim = 0 \\ \sum_{i_{n-(d-1)}=1}^{n-(d-1)} d_{i_{n-(d-1)}} \times \dots \times \sum_{i_{n-(d-1)+j}=i_{n-(d-1)+(j-1)}+1}^{n-(d-1)+j} d_{i_{n-(d-1)+j}} \times \dots \times \sum_{i_n=i_{n-1}+1}^n d_{i_n} & \text{if } DiceDim > 0 \end{cases} \tag{4}$$

The *DiceDimFactor* computes the number of possible dice combinations for a given Dice Dimensionality. Every possible combination can be shown with a different set of free dimensions. Therefore, it is multiplied by the *FreeDimFactor*.

In order to convey a better understanding of how Formula 2 has to be applied, we present it with both parameters *DiceDim* and *FreeDim* set to 2 in Formula 5. The formula applies for all $n \geq 4$.

$$\begin{aligned}
Views_{2,2} = m \times & \left(\left(\frac{(n-2)(n-3)}{2} + n - 1 \right) \times \sum_{i=1}^{n-1} d_i \sum_{j=i+1}^n d_j \right) \\
& + \left(\frac{(n-1)(n-2)}{2} + n \right) \times \sum_{i=1}^n d_i \\
& + \left(\frac{n(n-1)}{2} + n + 1 \right)
\end{aligned} \tag{5}$$

The maximum number of possible views is massively decreased by the restrictions *DiceDim* and *FreeDim* for high dimensional datasets. In our URL-scheme, these restrictions can be regarded by defining a maximum pattern. The maximum pattern of Formula 5 is “X22”. It means that *X measures2project*, two *dimensions2keep* and two *members2dice* are possible.

We evaluated the formulas numerically and received the same result for Formula 1 and 2 by setting the restrictions to *n*. We also implemented an algorithm which generates a list of all possible URLs and received an overall number of links as computed by Formula 2. Therefore, we conclude the correctness of the formulas.

3 SEO Evaluation

The previous chapter explained the conceptual model of our approach *SEO4OLAP*. In order to evaluate it, we implemented a concrete system and published two datasets from *Eurostat*. The system is based on Java and deployed on a Google App Engine. In order to evaluate OLAP queries over statistical Linked Data, we used the library OLAP4LD by Kämpgen and Harth [KH14]. The source code of our implementation (for more information, see [Br16]) is published at

<https://github.com/dbreucker/seo4olap>.

For our evaluation, we focused on datasets from Eurostat which were wrapped as Linked Data by Estatwrap. We discovered some challenges that have to be addressed for automated query generation. This is mainly due to the fact that real-world datasets differ from the modelling approaches as intended by the Data Cube Vocabulary (QB) or SDMX. The major issues are presented in the following.

1. **Measure-Dimensions:** In many datasets by *Eurostat* a dimension is used to specify the indicator of the measure. This means that a measure is modelled as a dimension, a *Measure-Dimension*. We understand that this may be due to a transformation process, e.g. by converting a table into SDMX, and therefore a convenient solution for data publishers. Nevertheless, from a conceptual point of view of a Multidimensional Data Model, this is not intended and causes problems for OLAP-operations. First, instead of a Projection on a measure, a Dice on the *Measure-Dimension* has to be performed. Second, the *Measure-Dimension* cannot be sliced, since an aggregation would cause implausible values. Since QB offers the possibility

to explicitly declare such *Measure-Dimensions*, we conclude that this is a common practice in real-world datasets. Nevertheless, the Linked Data from *Estatwrap* does not make use of this QB-feature.

2. *Slice-Members*: Some dimensions contain members which are, from a conceptual point of view, aggregations to a higher level. As an example, at *Eurostat* we often find three values for the dimension *Gender*: *Female*, *Male* and *Total*. In consequence, a slice on *Gender* would aggregate all three members and thus lead to wrong values. The member *Total* represents the correct values. Therefore, we define such members as *Slice-Members*. A *Slice-Member* is a member of a dimension, which represents the aggregation of the dimension. We understand that publishers of datasets may have good reasons for this. Nevertheless, this is a challenge for automated OLAP-query generation, since such dimensions cannot be sliced. Instead of a slice, a dice on the *Slice-Member* has to be performed for generating correct values.

In order to evaluate our approach from a SEO perspective, we measured how our generated webpages rank in search engine results in comparison to the original website of *Eurostat*. In the following, we present our evaluation method and our findings. Our assessed evaluation data is published at

<https://github.com/dbreucker/seo4olap-evaluation>.

For the purpose of this evaluation, we acquired the domain <http://open-statistics.org>. On open-statistics.org, we published two datasets from *Eurostat*. The first is about employment statistics mainly in European countries per year and gender. The second contains information about the gross domestic product in European countries per year. They were first published by us on 22nd of December 2015. The sitemap was submitted to Google on the same day. Since it takes a couple of weeks until new pages rank well, we waited until the 6th of March 2016 for this evaluation. Besides setting two links from other websites, we did not do any SEO off-page optimization techniques in order to strengthen the PageRank or TrustRank of open-statistics.org.

In this evaluation, we measured the search engine rank depending on different keywords for our website open-statistics.org and two benchmarks. The baseline is an online pivot table by *Eurostat*, which allows users to explore the dataset. The second benchmark is a website by *Eurostat* which describes a small subset of facts included in the datasets. It contains a lot of text and therefore many keywords.

Since Google is by far the most used search engine, we only measured the Google rank. The rank assessment was done by the software CuteRank. By using the software, we guarantee that the rank is not influenced by a personal search profile. Ranks higher than 100 are set to 100 for mathematical aggregation.

We defined a set of 84 different search queries, which are grouped by four main keywords. The main keywords were: "employment", "employment rate", "gross domestic product" and "gdp". So there were two main keywords for each dataset. We considered multiple words as one keyword. Since we wanted to analyse whether the number of keywords per

query affects the search engine rank, we further grouped the query-set by the number of included keywords. As an example, the set includes the query "gdp per capita hungary per year", which has the main keyword "gdp" and three other keywords "per capita", "hungary" and "per year", thus four keywords in total.

We aggregated the values by building the average value per main keyword depending on the number of keywords. Since queries with a value of 100 strongly influence the average value, we calculated two average values: a normal one including such outliers and a clean one excluding such outliers.

Figure 3 illustrates the Google ranks for the cleaned averages per main keyword depending on the number of keywords. Our own published datasets on open-statistics.org are marked with a dotted line. The corresponding *Eurostat* landing pages as benchmark are illustrated in the same colour with a normal line. The baseline benchmark was never found for any query and is therefore always on top at rank 100.

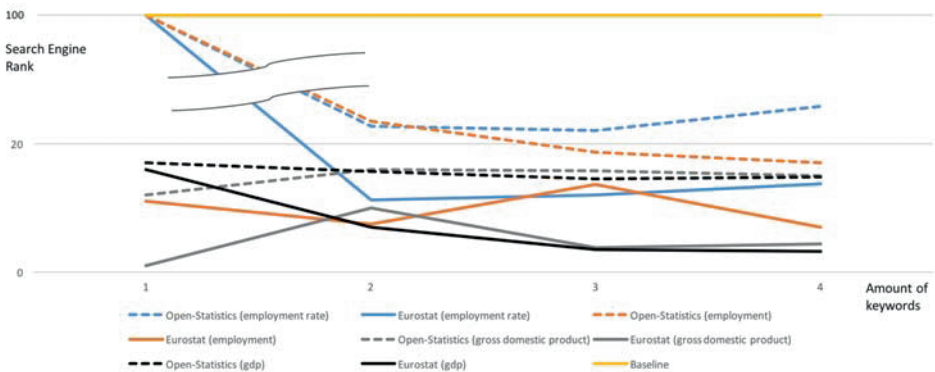


Fig. 3: Cleaned average Google ranks per main keyword for open-statistics.org and benchmarks

Prior to the assessment, we formulated hypotheses, that are presented and discussed in the following:

- Webpages generated by *SEO4OLAP* are indexed by search engines and retrieved for specific queries: The assessment shows that this thesis holds. Even though we created a lot of highly similar webpages, the according landing page for its specified keywords is in most cases retrieved by Google. This shows that the approach to generate websites for a high number of views per dataset is feasible.
- The more specific a query is, the better our search engine results are: Our approach is to generate a lot of views per dataset and thereby produce very specific landing pages for longtail queries, i.e., searches for very specific information that only few websites offer. Since a query consisting of more keywords is more specific and thereby has less competition, we assumed that our pages rank better with more keywords. This thesis cannot be confirmed by the assessed data. As soon as two

keywords are involved in a query, the rankings do not improve significantly with further keywords.

- *SEO4OLAP* pages rank better than the benchmark: This thesis holds depending on the benchmark. The baseline, i.e. the pivot table containing the dataset, is never found by Google for our defined queries. In comparison to this, our pages rank significantly better. On the contrary, the manually generated landing pages by *Eurostat* rank very well and on average always better than our pages.

4 Discussion

The idea of our approach *SEO4OLAP* is to generate custom SEO-landing pages for every possible view of a data cube. The evaluation of our implementation shows that, by doing this, *SEO4OLAP* was able to achieve better rankings than the baseline. In the case of *Eurostat*, the baseline was an online pivot table showing the dataset. In consequence, our approach is an improvement for data publishers who simply publish their data without further efforts to manually describe their data on landing pages. This leads to the following benefits:

- Webpages created by *SEO4OLAP* are found by Google for the corresponding keywords. This can be a new source of user traffic for dataset publishers.
- Single facts or views of a dataset are presented in a human readable representation. Thereby, facts of the Semantic Web are made accessible to normal web surfers.
- *SEO4OLAP* allows to reference specific facts in an HTML representation by a URL, whereas before, one could only refer to an entire dataset. As an example, this is an advantage for researcher who want to reference a statistic source.

In the following, some aspects regarding our approach and our evaluation are discussed.

Besides the baseline, we also benchmarked our approach against manually created landing pages by *Eurostat*. Our evaluation shows that on average, we were not able to achieve better rankings than this benchmark. It has to be noted that *Eurostat* has a high authority domain which is trusted by Google. Our rankings would assumably be better, if we had done SEO off-page optimization to gain domain trust and PageRank. Nevertheless, we can only speculate whether we could have beaten the benchmark by applying these techniques. But we can derive a recommendation for data publishers. If the goal of using *SEO4OLAP* is to gain a new source of traffic, we recommend to apply off-page optimization techniques in order to achieve rankings on the first search engine result page.

In contrast to manually generating descriptive landing pages, our approach allows to automatically generate potentially thousands of webpages. From a SEO perspective, this is a longtail strategy, since these pages are optimized for very specific search queries with low competition. As mentioned in our evaluation, we assumed that the more keywords are

added to a query, the more specific the query and thus the better the rank. We observed that, once a query includes two keywords, adding further keywords does not improve the rank. An explanation would be that queries consisting of two keywords are already very specific in the domain of statistical facts. Thus, the competition for these keywords is already very low. At this point, our pages compete against pages which are not optimized for this exact set of keywords, but instead have a higher PageRank or TrustRank. Therefore, a further specification is not improving the result rank.

In our evaluation, we only published small datasets with a maximum of three dimensions, in order to test our approach at first with reduced complexity. Therefore, we do not know how search engines react to high dimensional datasets with more than 100.000 pages for one dataset. We leave the SEO-evaluation of high dimensional datasets to future research.

SEO4OLAP converts machine readable data into human readable webpages. Search engines would benefit, if the data on webpages was also provided in a machine readable manner. Therefore, semantic mark-up technologies such as *Schema.org* were developed. In our implementation, we added *Schema.org* mark-up in order to semantically describe the content of our webpages. The main intention of doing this, was to achieve better search engine rankings. To the best of our knowledge, *Schema.org* does not provide means to properly describe statistic facts. It is possible to describe a dataset, the publisher, the publishing date and many more. However, single statistical facts cannot be described. We think our approach and also statistical data itself would benefit, if such functionality would be added to *Schema.org*.

5 Related Work

This is, to the best of our knowledge, the first attempt to generate search engine optimized landing pages from data cubes, published as Linked Data. The related work can be grouped in 1) the visualization and human-friendly interaction with Linked Data and 2) studies about Search Engine Optimization. In the following, we discuss some of those studies.

Since Linked Data has a self-describing data format, it has the benefit of being machine-readable, thus allowing machine interpretation, e.g. by search engines such as the Google Knowledge Graph. On the contrary, even though the amount of published Linked Data is growing on a fast scale, there is still a lot of research to be conducted on how non-experts can interact with this data. One approach is described by Hoeffler [Ho13]. He introduces different tools, that allow the analysis and visualization of Linked Data without the knowledge of SPARQL or other semantic technologies. Those tools, further explained by Sabol et al [Sa14] are e.g. the CODE Query Wizard and the CODE Visualization Wizard (Vis Wizard). A similar approach is presented by Salas et al. [Sa12] with CubeViz. The presented studies analyse, transform or visualize RDF Data Cubes for direct user interaction. In contrast, we focus on a presentation optimized for search engines.

The other related field deals with Search Engine Optimization. A lot of research in this area is conducted by major SEO-Agencies, who drive experiments in order to further understand how website rankings can be improved. Since they rely on a competitive advantage, it can be concluded that only a fraction of these experiments are made public. Nevertheless, some academic studies are available. Contrary to our approach, they analyse which techniques influence the search engine ranking; we apply these techniques for our approach. Beel et al. [BGW09] analyse how the ranking of academic papers can be improved. They present some advice for optimization within papers, but do not evaluate whether their approach is successful. Shih et al. [SCC13] describe an empirical evaluation on how basic SEO-techniques correspond to rankings and Malaga [Ma09] evaluates the effect of Web 2.0 techniques.

6 Conclusion & Future Work

We have presented our approach *SEO4OLAP* to generate search engine optimized landing pages from arbitrary data cubes, modelled in the RDF Data Cube Vocabulary. We developed a new OLAP query model which can be represented by a clean URL-scheme and therefore allows to be submitted via HTTP. We illustrated a system architecture which is able to process such OLAP-requests by transforming them into *Logical OLAP Query Plans*, executes them on an OLAP-engine and enhances results with corresponding keywords.

We presented two formulas to calculate the number of possible views depending on the number of members per dimension. The first formula calculates the upper bound of possible views and is exponentially growing with the number of dimensions; the second formula introduces two restrictions in order to decrease the overall number. Both formulas were numerically verified.

To evaluate our approach from a SEO perspective, we implemented a *SEO4OLAP* system in Java and published two datasets from *Eurostat* which resulted in 614 generated landing pages. We evaluated how well our pages are found by search engines in comparison to the dataset source, as well as to manually generated websites from *Eurostat*.

As a conclusion, it can be stated that our approach is feasible in practice and has benefits for data publishers, search engine providers and web users. First, single facts of a dataset are represented as HTML and are thus referenceable and human readable; Second, facts can be found by search engines which results in a new traffic source for data publishers; Third, the system provides an interface for non-experts to statistical Linked Data.

As future work, we plan to publish more data cubes, e.g., all 5,000 datasets from Eurostat; performance optimisations will be necessary to generate and query several GB of RDF data. Also, we plan to investigate the possibility to automatically increase the usefulness of our landing pages by providing additional information from the Semantic Web, e.g., the more commonly used dimension member label “Women” instead of “Female”.

Acknowledgements. This work was supported by the German Ministry of Education and Research (BMBF) within the BigGIS project (Ref. 01IS14012A).

References

- [BGW09] Beel, J.; Gipp, B.; Wilde, E.: “Academic Search Engine Optimization (aseo) Optimizing Scholarly Literature for Google Scholar and Co.” *Journal of scholarly publishing*, vol. 41, no. 2, pp. 176–190, 2009.
- [Br16] Breucker, D.: “Search Engine optimized presentation of Statistical Linked Data”. Master Thesis at Karlsruhe Institute of Technology, 2016.
- [Ho13] Hoefler, P.: “Linked data interfaces for non-expert users,” in *The Semantic Web: Semantics and Big Data*. Springer, pp. 702–706, 2013.
- [KH14] Kämpgen, B.; Harth, A.: “OLAP4LD - A Framework for Building Analysis Applications over Governmental Statistics” *ESWC 2014 Posters & Demo session*, Springer, Mai, 2014.
- [KOH12] Kämpgen, B.; O’Riain, S.; Harth, A.: “Interacting with statistical linked data via OLAP operations,” in *The Semantic Web: ESWC 2012 Satellite Events*. Springer, pp. 87–101, 2012.
- [Ma09] Malaga, R.: “Web 2.0 Techniques for search engine optimization: Two case studies,” *Review of Business Research*, vol. 9, no. 1, pp. 132–139, 2009.
- [Mu14] Mutlu, B. et.al.: “Suggesting visualisations for published data,” *Proceedings of IVAPP*, pp. 267–275, 2014.
- [Sa12] Salas, P. E. R. et.al.: “Publishing statistical data on the web,” in *Semantic Computing (ICSC), 2012 IEEE Sixth International Conference on*. IEEE, pp. 285–292, 2012.
- [Sa14] Sabol, V. et.al.: G. “Discovery and visual analysis of linked data for humans,” in *The Semantic Web–ISWC 2014*. Springer, pp. 309–324, 2014.
- [SCC13] Shih, B.-Y.; Chen, C.-Y.; Chen, Z.-S.: “An empirical study of an internet marketing strategy for search engine optimization”, *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 23, no. 6, pp. 528–540, 2013.

Recht und Technik – Datenschutz im Diskurs

Recht und Technik: Datenschutz im Diskurs

Rüdiger Grimm¹, Christoph Sorge², Indra Spiecker genannt Döhmann³

Sowohl Informatiker als auch Juristen arbeiten daran, Datenschutz und IT-Sicherheit zu verbessern oder zumindest Verschlechterungen zu verhindern. Dazu gehören einerseits ein technikorientiertes Verständnis der Gefahrenlagen und die Entwicklung technikbasierter Lösungsmöglichkeiten auf der Basis der Wertungen des Datenschutzrechts. Andererseits muss auch die Auslegung der Rechtsnormen des Datenschutzes mit der gesellschaftlichen und technischen Entwicklung Schritt halten. Der Workshop, der 2016 bereits zum dritten Mal stattfindet, hat sich daher zum Ziel gesetzt, den Dialog zwischen Vertretern beider Disziplinen zu fördern.

Die insgesamt neun Einreichungen wurden von jeweils drei unabhängigen Mitgliedern des Programmkomitees blind begutachtet. Nach einer inhaltlichen Auseinandersetzung mit den Beiträgen mit Bezug auf ihre Qualität und auf das Ziel des Workshops wurden fünf Beiträge zur Veröffentlichung angenommen.

Die angenommenen Beiträge lassen sich in zwei Bereiche einordnen: Dem Bereich der *Anwendungen* lassen sich die Beiträge von *Volkman et al. (Digitale Tarnkappe: Anonymisierung in Videoaufnahmen)*, *Wagner et al. (Privacy-Dashcam ?- Datenschutzfreundliche Dashcams durch Erzwingen externer Anonymisierung)* und *Jensen et al. (Datenschutz im Fahrzeug der Zukunft: Vernetzt, Autonom, Elektrisch)* zuordnen. Zwei weitere Beiträge – von den Autoren *Bier (Datenschutzziele im Konflikt: Eine Metrik für Unverkettbarkeit als Hilfestellung für den Betroffenen)* und *Pollmann und Kipker (Eingeschränkte Selbstbestimmung im Onlineverkehr: Stärkung der Einwilligungserklärung durch Einführung vorformulierter Datenschutzbestimmungen)* – widmen sich Grundprinzipien des Datenschutzes.

Allen angenommenen Beiträgen gemeinsam ist ihr Potential zum interdisziplinären Diskurs – zum Teil sind bereits unter den Autoren Vertreter beider angesprochener Disziplinen vertreten sowie weiterer.

Die Organisatoren bedanken sich bei allen Autoren für die Einreichung hochwertiger Beiträge. Darüber hinaus gilt unser Dank den Gutachtern, die sich binnen kurzer Frist ausführlich mit den Einreichungen auseinandergesetzt und wohlbegründete Stellungnahmen abgegeben haben. Ohne ihre Unterstützung wäre der Workshop nicht möglich geworden.

¹ Universität Koblenz-Landau und Fraunhofer SIT, Darmstadt

² Universität des Saarlandes

³ Goethe-Universität Frankfurt

Programmkomitee

Vorsitz: Prof. Dr. Rüdiger Grimm, Universität Koblenz-Landau und Fraunhofer SIT, Darmstadt; Prof. Dr. Christoph Sorge, Universität des Saarlandes; Prof. Dr. Indra Spiecker genannt Döhmann, Goethe-Universität Frankfurt

Weitere Mitglieder:

- Prof. Dr. Matthias Bäcker, KIT
- Prof. Dr. Jens-Matthias Bohli, Hochschule Mannheim
- Dr. Katharina Bräunlich, Universität Koblenz-Landau
- Prof. Dr. Johannes Caspar, Hamburgischer Beauftragter für Datenschutz und Informationsfreiheit
- Prof. Dr. Mark Cole, Universität Luxemburg
- Dr. Stefan Drewes, DPA Drewes Privacy Advice GmbH
- Thorsten Feldmann, Rechtsanwalt
- Prof. Dr. Nils Gruschka, Fachhochschule Kiel
- Prof. Christoph Gusy, Universität Bielefeld
- Marit Hansen, Landesbeauftragte für den Datenschutz Schleswig-Holstein
- Niko Härting, Rechtsanwalt
- Michael Herfert, Fraunhofer SIT, Darmstadt
- Dr. Walter Hötendorfer, Universität Wien
- Thomas Kahler, Datenschutzbeauftragter DiBA
- Prof. Dr. Jan Henrik Klement, Universität des Saarlandes
- Prof. Dr. Artus Krohn-Grimberghe, Universität Paderborn
- Prof. Dr. Kai von Lewinski, Universität Passau
- Dr. Thomas Lapp, Rechtsanwalt
- Dr. Mario Lischka, DZ Bank
- Dr. Ronald Petrlic, Referent beim Landesbeauftragten für den Datenschutz Baden-Württemberg
- Dr. Philipp Richter, Universität Kassel
- Prof. Dr. Burkhard Schafer, Universität Edinburgh
- Prof. Dr. Erich Schweighofer, Universität Wien

Digitale Tarnkappe: Anonymisierung in Videoaufnahmen

Sebastian Volkmann¹, Linus Feiten¹, Christian Zimmermann¹, Sebastian Sester¹, Laura Wehle¹ und Bernd Becker¹

Abstract: Videoüberwachung ist heute allgegenwärtig. Sie dient dazu, Delikte im Nachhinein aufzuklären, zur Echtzeit-Überwachung oder zur Abschreckung. Darüber hinaus gibt es aber auch wirtschaftliche Interessen für eine Videoüberwachung und automatische Erfassung von Personen – z.B. zur Erstellung von Kundenprofilen und somit zur Analyse von Kaufverhalten. Dem gegenüber stehen Rechtsansprüche sowie ethische und gesellschaftliche Grundnormen, etwa dass Menschen nicht unter Generalverdacht gestellt oder ohne Zustimmung aufgezeichnet werden dürfen. In dieser Arbeit wird ein technischer Lösungsansatz behandelt, der eine flexible Handhabung der Videoüberwachung erlaubt. Es werden in diesem Zusammenhang neben der technischen Umsetzung auch ökonomische, ethische und juristische Fragen betrachtet. Der Lösungsansatz besteht darin, Personen auf Videoaufnahmen durch ein kryptographisches Verfahren unkenntlich zu machen, noch bevor die Aufnahmen die Kamera-Elektronik verlassen. Nur mittels eines geheimen kryptographischen Schlüssels können einzelne Zeit- und Bildbereiche einer Aufnahme wieder deanonymisiert werden, wodurch rechtlichen wie ethischen Bedenken Rechnung getragen werden kann. In kommerziellen Szenarien erlaubt es diese *digitale Tarnkappe*, dass Kunden z.B. im Rahmen eines Prämien-Programms freiwillig auf Anonymisierung verzichten. Während in der Literatur der Informatik bereits seit längerem Technologien für solche System beschrieben werden, werden in dieser Arbeit Wege gezeigt, wie dessen Einbettung in die Gesellschaft wirklich realisiert werden könnte.

Keywords: Videoüberwachung, Datenschutz, Transparenz, Privatsphäre

1 Einleitung

Videoüberwachung ist heute aus dem Lebensalltag größerer Städte nicht mehr wegzudenken. Staatliche Einrichtungen, Unternehmen und Privatpersonen versprechen sich in erster Linie einen Zuwachs an Sicherheit, indem die Aufnahmen in Echtzeit ausgewertet oder aufgezeichnet werden können, um bei Delikten im Nachhinein die Täter zu identifizieren. Da die Echtzeit-Auswertung ab einer gewissen Anzahl von Kameras nicht mehr effektiv von menschlichem Personal durchführbar ist, gibt es Bestrebungen, die Analyse automatisiert durch Bilderkennungsalgorithmen durchführen zu lassen, wie es sich z.B. das kontrovers diskutierte EU-Forschungsprojekt INDECT zum Ziel gesetzt hat.

Aber nicht nur zur Herstellung von mehr Sicherheit kann Videoüberwachung eingesetzt werden. Durch automatische Bildanalyse rückt die Technik auch in das Interesse privater Wirtschaftsunternehmen. Wie heute bereits das Such- und Click-Verhalten von Internetnutzern gesammelt und analysiert wird, um daraus Nutzerprofile für unter anderem personalisierte Werbung oder Angebotsoptimierung zu generieren, kann dies auch mit Videoaufnahmen geschehen. Durch das *Tracken* von Kunden eines Kaufhauses könnte der

¹Universität Freiburg, Centre for Security and Society, Bertoldstraße 17, 79085 Freiburg, {feiten, sesters, wehle, becker}@informatik.uni-freiburg.de, sebastian.volkmann@philosophie.uni-freiburg.de, zimmermann@iig.uni-freiburg.de

Betreiber z.B. die Anordnung der Produktregale optimieren – ganz im Sinne der heute schon im Internet geläufigen Empfehlungen: „Personen die ... gekauft haben, interessieren sich auch für ...“. Per Smartphone-App, Bildschirmen im Kaufhaus oder sogar eingebautem Display im Einkaufswagen der Zukunft können dem Kunden dann personalisierte Werbung oder Rabattangebote angezeigt werden.

Trotz gesetzlicher Normen dazu, unter welchen Bedingungen Videoüberwachung stattfinden darf, gibt es eine andauernde gesellschaftliche Debatte darüber, inwieweit diese mit einer freiheitlichen Gesellschaft vereinbar ist. Man betrachte z.B. die ‚Big Brother Awards‘, welche mit Anspielung auf George Orwells Roman ‚1984‘ seit 1999 weltweit von Bürgerrechtlern an Akteure verliehen werden, die als besonders schädlich für die persönliche Privatsphäre empfunden wurden. Einen direkten Bezug zu Videoüberwachung gab es z.B. bei den deutschen ‚Preisträgern‘ der Jahre 2000 (Deutsche Bahn, Überwachung von Bahnsteigen), 2004 (Lidl, Überwachung von Mitarbeitern) und 2013 (Uni Paderborn, Überwachung von Hörsälen und Rechnerpools). Am 20.4.2016 entschied zudem das Bundesverfassungsgericht, dass die Ermächtigung zur verdeckten Erhebung personenbezogener Daten im BKA-Gesetz – etwa durch heimliche Videoüberwachung von Privaträumen – in dieser weitreichenden Form nicht Verfassungsgemäß ist und forderte die Sichtung des erhobenen Materials durch eine unabhängige Datenschutzinstanz.

In dieser Arbeit wird ein technischer Ansatz behandelt, der die divergierenden Interessen von Videoüberwachungsbetreibern und um ihre Privatsphäre besorgten Bürgern bzw. Kunden miteinander vereinbart: die *digitale Tarnkappe* (DTK). Dieses eigentlich generelle Konzept [BMP12] wird hier auf Videoaufnahmen angewandt, indem Personen auf Videoaufnahmen automatisch unkenntlich gemacht werden, was jedoch im Nachhinein auch wieder aufgehoben werden kann, wenn eine unabhängige *Schlüsselinstanz* den dafür nötigen kryptographischen Schlüssel freigibt. Für ein kommerzielles Szenario ist es dabei möglich, dass Kunden freiwillig auf ihre Anonymisierung verzichten. Abschnitt 2 beschreibt die Funktionsweise dieser DTK aus technischer Sicht, während Abschnitt 3 ihre ökonomischen Anwendungsmöglichkeiten aufzeigt. Abschnitt 4 unternimmt eine ethische Folgenabschätzung, während Abschnitt 5 die DTK in den Kontext der heutigen deutschen Gesetzeslage setzt. Jede dieser Betrachtungen kann aufgrund der Platzbeschränkung dieser Arbeit nur in aller Kürze erfolgen. Die Besonderheit unseres Beitrags liegt vor allem in der Kombination von technischem Konzept und multidisziplinärer Betrachtungsweise. Abschnitt 6 beschließt die Arbeit mit einer Diskussion und Ausblick. Dabei wird auch das über Videoaufnahmen hinausgehende generelle Konzept DTK beschrieben.

2 Informatischer Lösungsansatz

Technologien zur Anonymisierung in Überwachungsvideos werden in der technischen Literatur seit längerem diskutiert; z.B. [Se05, DE06, CPV06, Sc07, Hu14, PLCFR15]. Die meisten dieser Ansätze detektieren automatisch bestimmte Bereiche in Überwachungsvideos – wie menschliche Silhouetten, Gesichter oder Nummernschilder – und schwärzen diese unumkehrbar aus. Dabei ist oft das Ziel, dass ein Computeralgorithmus ohne menschliches Zutun erkennt, ob sich auf dem Kamerabild etwas ‚überwachungswürdiges‘ ereignet, um diesen Bereich dann *nicht* auszuschwärzen bzw. einen menschlichen Beobachter zu

alarmieren. Einige Ansätze (z.B. [Ch09, CKM09, CCK13]) bieten auch die Möglichkeit, die Ausschwärzung eines Bereichs nachträglich wieder aufzuheben.

Dies ist eine essentielle Funktion der Videoanonymisierung, wie sie hier vorgeschlagen wird. Eine automatische Erkennung von ‚überwachungswürdigen‘ gegenüber ‚harmlosen‘ Videosequenzen ist jedoch nicht erforderlich. In der einfachsten Variante kann grundsätzlich alles ausgeschwärtzt werden, was sich im Bild bewegt, wofür dann kein komplizierterer Bilderkennungsalgorithmus nötig ist. Dennoch kann die hier beschriebene Methode prinzipiell um jeden beliebigen komplizierten solchen Algorithmus erweitert werden, da dies der Funktionsweise keinen Abbruch tun würde. Wie die auszuschwärtzenden Bereiche bestimmt werden, steht nicht im Vordergrund dieser Arbeit sondern viel mehr das Verfahren, mit dem die Schlüssel zur Deanonymisierung erstellt und verwaltet werden.

Die Videoaufnahmen werden wie gewöhnlich von einer Kamera gemacht, wobei deren Elektronik so erweitert ist, dass die aufgezeichneten Rohdaten den internen Speicher der Kamera niemals verlassen. Durch Methoden der Hardware-Security bzw. Trusted-Hardware kann die Kamera so gesichert werden, dass es nahezu unmöglich ist, den internen Speicher von außen abzugreifen. Ein solcher Angriff würde es erfordern, nicht nur das Gehäuse der Kamera zu öffnen, das auf klassische Art verplombt sein kann, sondern auch die Kamera in ein Labor zu bringen, wo komplizierte invasive Eingriffe in die Elektronik durchgeführt werden müssten. Zur Abwehr auch solcher Angriffe gibt es mannigfaltige Maßnahmen wie z.B. *tamper-sensing Meshes* [An06] oder *Physically Unclonable Functions* (PUFs) [RDK11]. Erstere sorgen dafür, dass bei einem invasiven Angriff eine Sicherung zerstört wird, die daraufhin die Funktion des Chips deaktiviert. Letztere erzeugen eine eindeutige Chip-Signatur, die durch invasive Angriffe unwiederbringlich verfälscht wird. Eine solche PUF-Signatur kann somit dazu verwendet werden, sicherzustellen, dass eine Kamera mit DTK-Funktionalität nicht manipuliert oder durch eine scheinbar identische ausgetauscht wurde. Die PUF-Signatur des Kamerachips wird dabei nie nach außen kommuniziert, sondern innerhalb des Chips zur Erstellung des kryptographischen *Private-Keys* eines asymmetrischen Kryptosystems [PP10] verwendet. Der DTK-Chip der Kamera kommuniziert lediglich den zu diesem Private-Key gehörenden Public-Key nach außen. Eine mit diesem Public-Key verschlüsselte Anfrage kann nur von dem DTK-Chip plausibel beantwortet werden, wenn er im Besitz des Private-Key ist. Für physikalische Absicherungen wie Verplombung oder Trusted-Hardware gibt es zwar stets theoretische Möglichkeiten, diese auch zu umgehen, aber die Kosten für einen Angriff können durch Gegenmaßnahmen wie die oben genannten beliebig in die Höhe getrieben werden. Wie hoch der Aufwand für einen Angreifer sein soll, kann je nach Sensitivität des Szenarios angepasst werden. Eine bedeutende Herausforderung wird es sein, die ‚Chain of Trust‘ so zu gestalten, dass jeder von der Videoüberwachung betroffene diesen nachvollziehen und ihm vertrauen kann.

Im Folgenden wird die Hardware der Kamera daher als sicher angenommen, d.h., die von ihr gemachten *Rohdaten-Aufnahmen* sind außerhalb ihres DTK-Chips nicht abgreifbar. Stattdessen werden nur Daten nach außen gegeben, in denen Personen bereits anonymisiert wurden. Dies geschieht dadurch, dass zunächst die zu anonymisierenden Bildbereiche (Regions of Interest, ROIs) in jedem Kamerabild in Echtzeit erkannt werden. Das leistet eine auf dem DTK-Chip arbeitende Bildverarbeitung, die – wie bereits erwähnt – beliebig

kompliziert sein kann. Bereits heute gibt es Algorithmen, mit denen Menschen automatisch in Videos erkannt werden können [DT05, ARS09, WSS10, Be12]. Sollten sich diese Ansätze allerdings als noch zu unzuverlässig oder rechenaufwändig erweisen um in Echtzeit auf in der Kamera eingebetteter Hardware zu laufen, können die ROIs auch durch denkbar simple Algorithmen bestimmt werden; beispielsweise durch die einfache Detektion von Bewegungen im Bild.

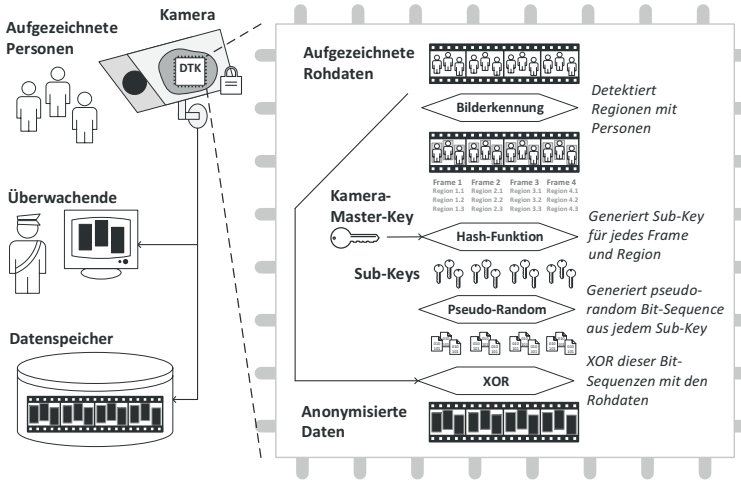


Abb. 1: Der schematische Aufbau des DTK Kamera-Systems.

Abbildung 1 zeigt, wie die Anonymisierung schematisch abläuft. Sie besteht nicht bloß in der Schwärzung eines Bereiches. Damit wären die Aufnahmen zur Strafverfolgung oder kommerziellen Auswertung nutzlos. Stattdessen werden die zu anonymisierenden Bildbereiche kryptographisch verschlüsselt. Jede DTK-Kamera hat ihren eigenen digitalen *Kamera-Master-Key* (KMK), welcher der Verschlüsselung zugrunde liegt und Teil der gesicherten Kamera-Hardware ist. Aus diesem KMK wird für jeden Zeitabschnitt und jeden zu anonymisierenden Bildbereich ein einzigartiger *Sub-Key* generiert. Dies geschieht über eine *Hash-Funktion* [Sh11], welche als Eingabe den KMK sowie die Zeitmarke und die Koordinaten des zu anonymisierenden Bildbereichs bekommt. Starke *Hash-Funktionen* haben die Eigenschaft, dass sich aus ihrem Ausgabewert nicht die Eingabe berechnen lässt. D.h. aus einem *Sub-Key* lässt sich nicht wieder der KMK berechnen. Für die Verschlüsselung, d.h. die Anonymisierung eines Bildbereichs wird aus dem jeweiligen *Sub-Key* ein *Stream-Cipher* generiert, der mit den Rohdaten des zu anonymisierenden Bereiches über ein logisches XOR verknüpft wird. Dies ermöglicht es, dass der gleiche *Sub-Key* auch wieder für die Entschlüsselung, d.h. die Deanonymisierung, verwendet werden kann.

Wie Abbildung 1 zeigt, werden außerhalb der Kamera nur anonymisierte Aufnahmen abgespeichert. Sollte der *Auswertende* es für nötig befinden, eine Person zu deanonymisieren, kann er dies bei der unabhängigen *Schlüsselinstanz* beantragen, die im Besitz des KMK ist (siehe Abbildung 2). Findet diese die Deanonymisierungsanfrage berechtigt, wendet sie ihren KMK zusammen mit den vom *Auswertenden* angegebenen Zeitspanne und Bildbe-

reich auf die *Hash-Funktion* an, um damit die nötigen *Sub-Keys* zu erstellen. Diese kann der *Auswertende* nun lediglich zur Deanonimisierung der von ihm angeforderten Zeitspanne und Bildbereiche verwenden.

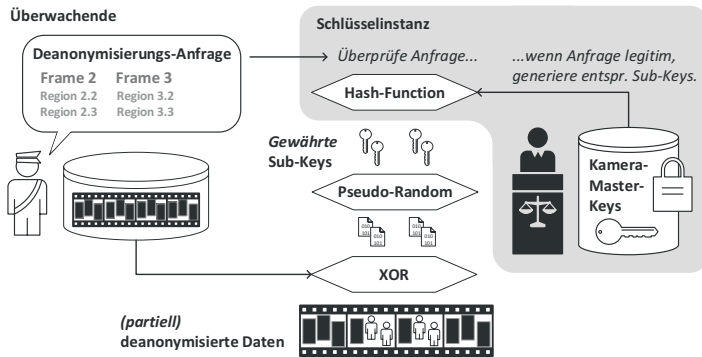


Abb. 2: Deanonimisierung ist nur durch die von der Schlüsselinstanz bewilligten *Sub-Keys* möglich.

Auf den Abbildungen ist bereits angedeutet, dass nicht alle Personen anonymisiert werden. In einem kommerziellen Szenario könnten Personen gegen eine Prämie freiwillig auf ihre Anonymisierung verzichten, und es dem Betreiber somit erlauben, sie auf den Videoaufnahmen zu *tracken*. Wie diese Personen sich dem DTK-System zu erkennen geben können und wie die ökonomischen Anwendungsmöglichkeiten sind, wird in Abschnitt 3 erörtert.



Abb. 3: Ein erster Proof-of-Concept der DTK.

Für einen ersten Proof-of-Concept wurde ein DTK-System als Opt-Out-Variante implementiert. D.h., anstatt dass standardmäßig alle Personen anonymisiert werden und nur solche mit bestimmter Kennzeichnung nicht, war es umgekehrt: Bildbereiche, die sich um eine Infrarot-LED herum befanden, wurden anonymisiert. Grund dieser vorläufigen Vereinfachung war, dass von einer automatischen Identifizierung der ROIs zunächst abstrahiert und der Fokus auf die Daten-Verschlüsselung gelegt wurde. Abbildung 3 zeigt den Aufbau der LED sowie den Anonymisierungseffekt auf einem Kamerabild. Durch eine adäquate ROI-Bestimmung lässt sich dieser Prototyp in die oben beschriebene Opt-In-Variante überführen.

3 Ökonomische Anwendungsmöglichkeiten

DTK-Systeme können nicht nur dazu verwendet werden, bei der Verbrechensprävention bzw. -aufklärung durch Videoüberwachung die Privatheit der Betroffenen zu schützen. Sie ermöglichen es darüber hinaus auch, ökonomisch motivierte Videoüberwachung privatsachhaltend umzusetzen. Im Folgenden sollen die Anwendungsmöglichkeiten des vorgestellten System im Kontext von Kundenverhaltensanalysen in videoüberwachten Kaufhäusern untersucht werden.

Bereits seit langem verwenden Betreiber von Kaufhäusern Videosysteme nicht nur zur Abschreckung von Kaufhausdieben und zur Beweisfindung im Falle von Delikten innerhalb der Kaufhäuser. Videosysteme bieten zusätzlich die Möglichkeit, Kundenbewegungen und sogar die Blickrichtungen der Kunden detailliert nachzuverfolgen [LH08, Li07]. Dies ermöglicht es Kaufhausbetreibern, Erkenntnisse zu gewinnen, die bei der Ladenplanung, wie auch bei der Durchführung von Werbeaktionen, gewinnbringend angewendet werden können. Die in Deutschland hohen Datenschutzbedenken der Bevölkerung stellen allerdings ein Hindernis bei der Einführung und Nutzung solcher Analysemethoden dar. Das hier vorgestellte DTK-System kann verwendet werden, um die Datenschutzbedenken der Kunden zu adressieren und sicherzustellen, dass nur das Verhalten von Kunden, die videogestützten Analysen ihres Verhaltens zugestimmt haben, ausgewertet wird.

Das hier vorgestellte DTK-System kann analog und komplementär zu den momentan weit verbreiteten ‚Bonuspunkt Karten‘ verwendet werden, um einerseits Verhaltensanalysen auf Kunden einzuschränken, die solchen Analysen zugestimmt haben und andererseits, um Kunden Anreize zu bieten, solchen Analysen zuzustimmen. Dabei bieten sich zwei Optionen. Im momentanen Implementierungsstand des vorgestellten DTK-Systems kann es eingesetzt werden, um Kunden eine leicht umzusetzende Möglichkeit des Opt-Out aus der Überwachung und Analyse ihres Verhaltens zu bieten. Dabei können Kunden durch das Tragen eines entsprechenden Signals, bspw. an der Kleidung oder am Einkaufswagen, signalisieren, dass sie keine Überwachung ihrer Bewegung wünschen. Ein solcher Ansatz würde allerdings nicht in Einklang mit den Forderungen der zukünftig zu beachtenden europäischen Datenschutzgrundverordnung stehen, die einen ‚privacy by design‘-Ansatz vorschreibt [Eu12, Art. 23]. Um dieser Anforderung zu genügen, kann das DTK-System allerdings so implementiert werden, dass standardmäßig die komplette Videoaufnahme verschlüsselt wird und nur Bereiche entschlüsselt werden, in denen ein entsprechendes Signal aufgefangen werden kann. Denkbar wäre hier eine Infrarot-LED, deren Lichtimpulse einen der Person zuordenbaren Code an die Kameras senden. Unabhängig von Kameras könnte auch eine Funktechnik eingesetzt werden, durch welche z.B. die ‚Bonuspunkte-Smartcard‘ einer Person verfolgt würde. Wenn diese Positionsdaten dem DTK-Algorithmus zur Verfügung stehen, kann die Anonymisierung der Person in den entsprechenden Bildbereichen aufgehoben werden.

Ein zu lösendes Problem hierbei besteht in der Wahl einer geeigneten *Schlüsselinstanz*. Des Weiteren muss das vorhandene System erweitert werden, um sicherzustellen, dass im Falle gleichzeitiger Aufnahme von Kunden die in die Analyse ihres Verhaltens eingewilligt haben und solchen, die dies nicht getan haben, der Opt-Out des einen Kunden höher priorisiert wird als der Opt-In des anderen.

4 Ethische Folgenabschätzung

Ethische Technikfolgenabschätzung (TA) zielt darauf ab, insbesondere nicht-intendierte Folgewirkungen technischer Entwicklungen in Bezug auf Grundwerte zu bewerten und dieses Wissen einer (oftmals politischen) Entscheidungsfindung zuzuführen [Ot05]. Ethische TA erarbeitet dabei normative Planungsgrößen [Gr99], kann aber auch – wie im Falle der DTK – konstruktiv genutzt werden, um die Technologieentwicklung selbst wiederum normativ zu prägen [Pe99]. In Bezug auf Videoüberwachung öffentlich zugänglicher Räume lassen sich zunächst zwei Grundperspektiven ethischer Argumentation unterscheiden: (1) Nicht-intendierte Folgen können konkrete Individuen unverhältnismäßig beeinträchtigen, insbesondere hinsichtlich ihrer Grundrechte oder anderer gesellschaftlicher Grundwerte; (2) aus gesellschaftspolitischer Sicht können Nebenfolgen einen Trend hin zu einer restriktiven Gesellschaft bestärken. Letzterer Folgentypus ist insbesondere dann wichtig, wenn bestimmbare Individuen jeweils nur verhältnismäßig gering oder vereinzelt direkt beeinträchtigt werden, in Summe aber dennoch ein erhebliches Risiko für den offenen Charakter einer Gesellschaft besteht. Dies könnte durch eine langsame und kaum merkbare aber dennoch beständige Ausweitung von Sicherheitsmaßnahmen geschehen (populärwissenschaftlich als ‚boiling frog‘-Argument bekannt [TZ09]) oder durch eine nachträgliche Ausweitung des Verwendungszwecks („mission creep“-Argument [Pe06]). Eine weitere Gefahr der gesellschaftlichen Restriktion könnte darin bestehen, dass Sicherheitsmaßnahmen für Betroffene völlig intransparent sind und sich die jeweiligen staatlichen und privaten Akteure dadurch Rechenschaftspflichten entziehen können. Wertkonflikte vom ersten Typus sollen in der Folge anhand von zwei in der öffentlichen Debatte sehr präsenten Metaphern [GG00] analysiert werden: Der ‚gläserne Kunde‘ bzw. ‚gläserner Bürger‘, sowie der ‚Generalverdacht‘ unter den sich Betroffene gestellt fühlen können. Für die Besprechung des zweiten Typus von nicht-intendierten Folgewirkungen werden die Metaphern ‚orwellsche‘ und ‚kafkaeske Überwachungsgesellschaft‘ genutzt [Ly09].

Individuumszentrierte Perspektive Bei räumlich begrenzter Überwachung beschreibt die Metapher des gläsernen Kunden vor allem die Angst, dass das Sammeln und Auswerten von Bildinformationen privatwirtschaftlicher Akteure Einblicke in persönliche Einstellungen und Lebenssituationen von Betroffenen erlaubt oder Rückschlüsse auf deren Wünsche und Intentionen gewährt – obwohl diese von den Betroffenen selbst wie auch von der Gesellschaft grundsätzlich als schützenswert betrachtet werden. Die Schutzwürdigkeit bestimmter als ‚privat‘ oder gar ‚intim‘ bestimmter Lebensbereiche gründet dabei einerseits in kulturell geprägten Normen der Zurückhaltung (etwa zu Sexualität oder Krankheit), andererseits aber auch in dem gesellschaftlichen Wunsch, der Gefahr vorzubeugen, dass Einzelne unverhältnismäßig stark benachteiligt werden könnten – etwa aufgrund ihrer finanziellen oder sozialen Situation. Die Metapher des gläsernen Bürgers lässt sich in Abgrenzung hierzu vor allem auf das Sammeln und Auswerten von Informationen durch staatliche Akteure beziehen. Die Schutzwürdigkeit privater Informationen begründet sich dabei zusätzlich darin, dass Einzelne vor dem übermächtigen Zugriff des Staates geschützt werden sollen. Staatliche Videoüberwachung muss daher auch immer in Bezug auf individuelle Abwehrrechte gesehen werden. Das heißt allerdings nicht, dass ein Eingriff in diesen rechtlich geschützten Privatbereich immer illegitim ist. Allerdings sollte dies nicht wahllos geschehen, sondern nur in begründeten Fällen oder mit Zustimmung der Betrof-

fenen – und dies gilt umso mehr, je schutzloser der Betroffene dem Eingriff ausgesetzt ist. Staatliche Eingriffe sind deshalb im Allgemeinen deutlich strenger reglementiert als Eingriffe durch private Akteure. Gesellschaftlich werden in solchen Fällen Eingriffe bewusst zugunsten anderer Werte in Kauf genommen – etwa wenn der Verdacht auf eine Straftat vorliegt, die verfolgt werden soll. Die Metapher des Generalverdachts bezieht sich daher auf die Sorge vor einer anlasslosen Anwendung einer Sicherheitsmaßnahme wie der Videoüberwachung, durch die grundsätzlich jede Person von Eingriffen in schutzwürdige Privatbereiche beeinträchtigt wird – insbesondere bei staatlichen Eingriffen.

Beim klassischen Videobild lassen sich Kunden bzw. Bürger pauschal im Videobild identifizieren (manuell oder auch automatisiert etwa über biometrische Gesichtserkennung oder andere optische Kriterien). Über die Analyse der Bewegungen lassen sich gegebenenfalls automatisiert Rückschlüsse auf Kaufverhalten, Intentionen oder persönliche Lebensumstände ziehen: Wie lange hält sich diese konkrete Kaufhaus-Kundin bei den Kondomen auf, wie häufig besucht sie das Weinregal? Besucht der zufällig auf dem Bild erkannte Nachbar den belebten Platz zu allen Tageszeiten – oder nur außerhalb der regulären Arbeitszeiten? Wie lange unterhält er sich mit dem Prediger – wie lange mit den Leuten vom Wahlkampfstand? Insbesondere durch die Möglichkeiten einer automatisierten Analyse und einer Weiterverwendung der gespeicherten Aufnahmen ergeben sich dabei Beeinträchtigungen schützenswerter Privatbereiche, die gegebenenfalls deutlich darüber hinausgehen, womit Menschen in der Öffentlichkeit ohnehin immer rechnen müssen.

Durch die Unkenntlichmachung der identifizierenden Merkmale auf dem Videobild kann dieser Gefahr effektiv begegnet werden, weil sich die aufgezeichneten Informationen nicht mehr direkt oder zumindest nur noch sehr beschränkt mit konkreten Personen in Verbindung bringen lassen. Abhängig von der konkreten Implementierung der unabhängigen *Schlüsselinstanz* wird ein solcher Eingriff nur ermöglicht, wenn durch sie ein legitimer Grund für diesen Eingriff erkannt wird – etwa zur Nutzung als Beweismittel bei einem Ladendiebstahl oder einer Körperverletzung. Zudem können in einem solchen Fall gegebenenfalls auch nur bestimmte Bildbereiche oder Zeitabschnitte freigegeben werden. Mit Blick auf die im Urteil des Bundesverfassungsgericht zum BKA-Gesetz geforderte unabhängige Datenschutzinstanz zur Sichtung und Freigabe von Videomitschnitten bieten sich ebenfalls mögliche Implementierungen für DTK-Systeme an. Für beide Szenarien zeigt sich somit, dass der Einsatz eines DTK-Systems gegenüber einer klassischen Videoüberwachung die Privatheit Betroffener besser schützen kann und legitime Eingriffe in diesen schutzwürdigen Bereich gezielter vorgenommen werden können.

Gesellschaftszentrierte Perspektive Der Begriff ‚Überwachungsgesellschaft‘ verweist vor allem auf die Allgegenwart von umfassender privater oder staatlicher Beobachtung aller Mitglieder einer Gesellschaft. Wird der Begriff im Kontext von Videoüberwachung gebraucht, so geht es also gerade nicht um vereinzelte Beobachtungen eng umgrenzter Bereiche, sondern um die allmähliche Proliferation dieser Sicherheitsmaßnahme und die Gefahr der schleichenden Vernetzung und Zusammenführung gewonnener Informationen. Auch in diesem Kontext muss staatliche Videoüberwachung immer im Kontext der Begrenzung der staatlichen Übermacht durch die effektive Garantie von Abwehrrechten gesehen werden. In jedem Fall beschreibt die Metapher einer ‚orwellischen Überwachungsgesellschaft‘

aber insbesondere die Sorge, die Entwicklung könnte zu einer Situation führen, in der wir nahezu jederzeit beobachtet werden und nie genau wissen können, wie das eigene Verhalten interpretiert wird oder welche negativen Konsequenzen sich eventuell noch ergeben werden. Gesellschaftspolitisch impliziert dies das Risiko, dass die Wahrnehmung mancher Freiheiten einer Form von Selbstkontrolle zum Opfer fallen könnte: Aus der Angst, gespeicherte Informationen könnten zukünftig zu Nachteilen führen – etwa ein negatives privatwirtschaftliches Bonitätsranking oder die Klassifizierung als Hochrisiko-Fluggast – kann ein Normalisierungsdruck entstehen, der die Offenheit einer Gesellschaft de facto deutlich einschränkt. Die Metapher einer ‚kafkaesken Überwachungsgesellschaft‘ drückt zudem die Angst aus, gegen illegitim erlittene Nachteile effektiv keinerlei Mittel in der Hand zu haben, weil Betroffenen nie ganz klar ist, anhand welcher Informationen und Kriterien nachteilige Entscheidungen getroffen wurden und wie diese angefochten werden können. Durch Weitergabe bzw. Verkauf von Videoüberwachungsdaten ohne Zustimmung der Betroffenen entsteht hier die Gefahr einer umfassenden, kaum noch nachvollziehbaren Zusammenführung und späteren Zweckentfremdung – etwa zur Mustererkennung potentieller Straftäter oder möglicher Versicherungsrisiken.

Bei der klassischen Videoüberwachung lässt sich die Weitergabe, Zusammenführung und Zweckentfremdung gespeicherter Videodaten nur unzureichend einschränken. Auch wenn der Informationspflicht genüge getan und deutlich sichtbar auf die Videoüberwachung hingewiesen wird, kann der Einzelne nur sehr begrenzt absehen, was sich aus einer Zusammenführung von (Bild-)Informationen gewinnen ließe und wofür diese in Zukunft genutzt werden könnten. Zwar lassen sich einige der Risiken durch Selbstbeschränkungen der Betreiber ein Stück weit begrenzen, allerdings ist unklar, worin für private Akteure der Anreiz zu einem solchen Verhalten bestehen soll und welche Sanktionsmöglichkeiten es bei Datenmissbrauch effektiv geben kann. Beim Einsatz eines DTK-Systems lässt sich das Risiko einer Weitergabe, Zusammenführung und Zweckentfremdung der Bilddaten durch die unabhängige *Schlüsselinstanz* hingegen effektiv minimieren – insbesondere, wenn nur jene Bildinformationen entschlüsselt werden, die für den angegebenen Zweck unbedingt notwendig sind. Für private Betreiber einer Videoüberwachungsanlage ergibt sich zudem ein gewisser Anreiz zur Selbstbeschränkung, wenn mit einem erhöhten Schutz privater Kundeninformationen durch eine unabhängige *Schlüsselinstanz* geworben werden kann. Für beide Szenarien zeigt sich somit, dass der Einsatz eines DTK-Systems gegenüber einer klassischen Videoüberwachung die Privatheit der Betroffenen besser schützen kann und Eingriffe in diesen schutzwürdigen Bereich gezielter vorgenommen werden können.

5 Juristische Betrachtung

Gesetzliche Grundlagen für Videoüberwachung sowie Zwecksetzung des Gesetzes
Mit § 6b BDSG existiert eine gesetzliche Grundlage, die der Wahrung des informationellen Selbstbestimmungsrechts durch einen Ausgleich angemessener Interessen Rechnung trägt [Bu00b, S. 92]. Auf der einen Seite steht hierbei das jedem Einzelnen zustehende Allgemeine Persönlichkeitsrecht aus Art. 2 I GG, wobei die bei einer Videoaufzeichnung tangierten Rechte im Wesentlichen das Recht am eigenen Bild, das Recht am gesprochenen Wort und das Recht auf informationelle Selbstbestimmung sind. Ziel des BDSG an sich ist es gem. § 3 BDSG auch, zur Datenvermeidung und Datensparsamkeit beizutragen.

Auf der anderen Seite steht die Pflicht der Gefahrenabwehr und der Strafverfolgung als öffentliche Aufgabe, sowie die Interessen Einzelner am Schutz ihres Eigentums oder anderer gefährdeter Rechtsgüter. § 6b I lässt jedoch die Videoüberwachung nur zum Zweck der Aufgabenerfüllung öffentlicher Stellen (Nr. 1), der Wahrung des Hausrechts (Nr. 2) oder zur Wahrung berechtigter Interessen für konkret festgelegte Zwecke (Nr. 3) zu. Daran zeigt sich zwar das gesetzgeberische Ziel, die Videoüberwachung nur in engen Grenzen zuzulassen, andererseits beschreibt Nr. 3 eine eng auszulegende Generalklausel, die eine solche aufgrund jeglicher objektiv bestehenden rechtlichen, wirtschaftlichen oder ideellen Interessen ermöglicht [Be13, Rn. 17]. Um beiden Seiten angemessen Rechnung zu tragen, ist die Videoüberwachung grundsätzlich verboten und nur unter den im folgenden Absatz zu erläuternden Voraussetzungen zulässig.

Voraussetzungen für eine Videoüberwachung Wie oben genannt, ist gemäß § 6b I BDSG die „Beobachtung öffentlich zugänglicher Räume mit optisch-elektronischen Einrichtungen (Videoüberwachung)“ nur unter den Bedingungen Nr. 1-3 zulässig. Hierbei stellt die Norm zunächst den Anwendungsbereich fest, der sich auf „öffentlich zugängliche Räume“ erstreckt. Dieser weit zu verstehenden Begriff bezeichnet insbesondere öffentliche Plätze, Straßen, für Besucher zugängliche öffentlichen Gebäude, Bahnsteige, Schalterhallen, Museen oder Verkaufsräume [Bu00a, S. 38].

Bei der Beobachtung kommt es nicht darauf an, ob es sich um eine zielgerichtete Beobachtung handelt oder ob diese bloße Nebenfolge ist. Entscheidend ist nur, dass diese von einer gewissen Dauerhaftigkeit geprägt ist [GK14]. Optisch-elektronische Geräte sind alle Geräte, die Bewegtbilder bzw. als solche wahrzunehmende Bildfolgen erzeugen und wahrnehmbar machen können [Be13, Rn. 12]. Umstritten ist, ob der Anwendungsbereich des § 6b BDSG nur bei einer Aufzeichnung eröffnet ist, oder ob auch bloße Beobachtung ohne Speicherung unter § 6b BDSG fällt. Angesichts der klaren Gesetzesbegründung, die darlegt, dass bei § 6b die Beobachtung selbst erfasst ist und es nicht auf das Erfordernis einer Speicherung des Bildmaterials ankommt [Bu01, S. 61], dürfte allerdings klar sein, dass auch das bloße Beobachten an die Voraussetzungen des § 6b BDSG gebunden ist. Denn schon ein solches ist geeignet, beim Bürger ein Gefühl des permanenten Beobachtetseins hervorzurufen, das zu erheblichen Verhaltensänderungen führen kann. Dies stellt einen Eingriff in das allgemeine Persönlichkeitsrecht und die allgemeine Handlungsfreiheit dar.

Voraussetzung für die Installation einer Videokamera ist es, einen der in § 6b I Nr. 1-3 BDSG genannten Zwecke zu erfüllen. Angesichts vorrangiger landessepezifischer Regelungen dürfte (Nr. 1), die Videoüberwachung zur Aufgabenerfüllung öffentlicher Stellen zulässt, eine eher untergeordnete Rolle spielen. Wichtiger ist (Nr. 2), die zur Wahrnehmung des Hausrechts eine Videoüberwachung rechtfertigt. Dieses steht im öffentlichen Raum den öffentlich-rechtlichen Stellen zu, erstreckt sich allerdings auch im Privaten auf durch das Hausrecht erfasste Räume. So steht z.B. einem Kaufhausinhaber innerhalb seiner Geschäftsräume ein Hausrecht zu, das eine Videoüberwachung zulässig macht. (Nr. 3) ist ein Generaltatbestand, der zur Wahrnehmung berechtigter Interessen für konkret festgelegte Zwecke eine Videoüberwachung rechtfertigt. Nach der Gesetzesbegründung dürfen sich nur nicht-öffentliche Stellen auf (Nr. 3) berufen [Bu01, S. 62]. Es genügt jedes rechtliche, ideelle oder wirtschaftliche Interesse, sofern es konkret festgelegt ist.

Zudem muss laut Gesetz eine Erforderlichkeitsprüfung sowie Interessenabwägung erfolgen. Entsprechend des allgemeinen Begriffsverständnisses der Erforderlichkeit setzt diese voraus, dass die Videoüberwachung für den jeweiligen Zweck geeignet ist und kein milderes Mittel, das die Zweckerreichung in gleicher Weise fördern kann, zur Verfügung steht [De10, Rn. 236]. Für die Eignung ist es ausreichend, dass die Zweckerreichung gefördert wird. Berücksichtigt werden muss an dieser Stelle, ob es wirtschaftlich zumutbare Alternativen zur Videoüberwachung gibt, was jedoch aufgrund hoher Personalkosten und der relativ kostengünstiger Technik regelmäßig zu verneinen sein wird [Be13, Rn. 21]. Die Interessenabwägung erfolgt anhand einer Verhältnismäßigkeitsprüfung, die die Interessen der Betroffenen im Vergleich zu der Durchsetzung der in Nr. 1-3 genannten Zwecke unter Berücksichtigung der konkreten Umstände des Einzelfalls betrachtet [Be13, Rn. 22].

Veränderung der juristischen Betrachtung aufgrund der *Digitalen Tarnkappe* (DTK) sowie nötige juristische Gesetzesänderungen Bei Etablierung der DTK stellen sich die Fragen, wie sich die Bewertung aus juristischer Sicht verändern würde und ob gesetzliche Änderungen nötig wären. Zunächst ist festzustellen, dass auch das System an sich an der Tatsache der Beobachtung mittels einer optisch-elektronischen Einrichtung nichts verändert. Die DTK ersetzt nicht die Kamera an sich, sondern ist ein zusätzliches nachgeschaltetes technisches System.

Ein *Tracking* zur Kundenprofilerstellung auf Seiten des Kaufhausinhabers wäre aufgrund wirtschaftlicher Interessen gem. § 6b I Nr.3 BDSG denkbar. § 6b II normiert die Pflicht, den Umstand der Beobachtung und die verantwortliche Stelle durch geeignete Maßnahmen kenntlich zu machen. Eine allgemeine Sichtbarkeit der Kamera reicht jedoch hierfür nicht aus, vielmehr wird eine separate deutliche Kennzeichnung mittels Hinweisschildern verlangt [Be13, Rn. 27]. Würde ein DTK-System installiert, könnte auf einem solchen Hinweisschild zusätzlich eine Internetadresse angegeben sein, über die man auf zugehörige, vorab hinterlegte AGBs zugreifen kann und erfährt, wie man am Prämien-Programm teilnehmen und dabei auf seine Anonymisierung verzichten kann.

Bezüglich der Erforderlichkeit stellt sich die Frage, ob das DTK-System ein gleich geeignetes, milderes Mittel ist. Aufgrund der standardmäßigen Anonymisierung aller Personen, wäre durch die DTK die Persönlichkeitsrechtsverletzung behoben und ein hoher Persönlichkeitsrechtsschutz gewährleistet. Damit wäre die Verwendung einer Kamera samt DTK ein milderes Mittel. Fraglich ist allerdings, ob das System auch gleich geeignet ist. In öffentlich zugänglichen Räumen soll eine Videokamera meist auch der Gefahrenabwehr dienen, insbesondere die Strafverfolgung erleichtern und einen Abschreckungseffekt erzielen. Videoüberwachung ist für die Strafverfolgungsbehörden wichtig, weil Täter identifiziert und ihnen eine Straftat nachgewiesen werden kann. Bei der Speicherung anonymisierter Bildaufnahmen kann eine solche Identifizierung erst nach einer Deanonymisierung erfolgen. Bei konkreten Anhaltspunkten für eine begangene Straftat kann die Polizei einen Schlüssel zur Deanonymisierung anfordern, der mittels einer staatsanwaltlichen Verfügung oder eines richterlichen Beschlusses erteilt werden könnte. Dadurch könnte es zwar vereinzelt zu einer insgesamt längeren Speicherung der Aufnahmen kommen, dies ist hinsichtlich des erhöhten Persönlichkeitsrechtsschutzes jedoch eine hinnehmbare Nebenfolge. Auch die kurzen zeitlichen Verzögerungen bei der Identifizie-

rung der Täter fallen angesichts heutzutage langwieriger Strafprozesse nicht weiter ins Gewicht. Bezüglich des Abschreckungseffektes würden sich keine Änderungen ergeben, da die Anonymität auf den Aufnahmen nur so lange besteht, wie die *Schlüsselinstanz* nicht der Deanonymisierung zugestimmt hätte, und dies einem möglichen Täter bewusst wäre. Das System kann somit aufgrund seiner Gewährung eines hohen Persönlichkeitsrechtsschutzniveaus unter Wahrung der staatlichen Strafverfolgungsinteressen überzeugen. Es bedürfte lediglich einer gesetzlichen Ausgestaltung, die die Herausgabemöglichkeit der Deanonymisierungs-Schlüssel normiert.

6 Diskussion und Ausblick

Wie in der Einleitung erwähnt, ist das Konzept DTK keinesfalls auf Kamerasysteme beschränkt. Das Prinzip verschlüsselt abgespeicherter Daten, die nur durch eine unabhängige *Schlüsselinstanz* freigegeben werden können, lässt sich auf jegliches Szenario anwenden, bei dem sensible personenbezogene Daten anfallen. Das gleiche gilt bedingt für das Prinzip, dass Personen freiwillig ihrer eigenen Deanonymisierung zustimmen. Hierbei ist szenariospezifisch zu prüfen, wie sich diese Personen dem DTK-System zu erkennen geben. Entscheidend für den erfolgreichen Einsatz ist, dass das User-Interface sowohl für potentiell aufgezeichnete Personen, die *Auswertenden* als auch die *Schlüsselinstanz* leicht zu verstehen und intuitiv zugänglich ist.

Ziel dieser Arbeit war es, einen Anstoß für Technologien und multidisziplinäre Auseinandersetzungen zu geben in Richtung von mehr Transparenz und Selbstbestimmung über personenbezogenen Daten. Eine über den Prototyp hinausgehende Nutzung des vorgestellten DTK-Kamerasystems erfordert ausgefeilte automatische Bilderkennungsalgorithmen. Die entsprechende Technik steht oder wird in absehbarer Zukunft zur Verfügung stehen! Es ist den Autoren bewusst, dass eine Ausschwärzung auf Videos keine absolute Anonymität bedeutet, da eine ausgeschwärzte Person z.B. an ihrem Hund erkannt oder bei flächendeckender Überwachung ab ihrer Wohnung verfolgt werden könnte. Wie gezeigt wurde, birgt das Konzept in vielen Szenarien jedoch Vorteile gegenüber klassischer Überwachung, was ein wichtiger Schritt zu zielgenaueren Eingriffen, mehr Bewusstsein und größerer persönlicher Autonomie im Zeitalter von ‚Big Data‘ und Terrorbekämpfung ist.

Literaturverzeichnis

- [An06] Anderson, R.; Bond, M.; Clulow, J.; Skorobogatov, S.: Cryptographic Processors - A Survey. Proceedings of the IEEE, 94(2):357–369, Feb 2006.
- [ARS09] Andriluka, Mykhaylo; Roth, Stefan; Schiele, Bernt: Pictorial structures revisited: People detection and articulated pose estimation. In: Proc. CVPR. IEEE, S. 1014–1021, 2009.
- [Be12] Benenson, Rodrigo; Mathias, Markus; Timofte, Radu; Gool, Luc Van: Pedestrian detection at 100 frames per second. In: Proc. CVPR. IEEE, S. 2903–2910, 2012.
- [Be13] Becker, Thomas: In (Plath, Kai-Uwe, Hrsg.): Kommentar zum BDSG sowie Datenschutzbestimmungen des TMG und TKG. Verlag Dr. Otto Schmidt, 1. Auflage, 2013.

- [BMP12] Becker, Bernd; Müller, Günter; Polian, Ilia: Digital Tarnkappe: Stealth Technology for the Internet of Things. In (Gander, Hans-Helmuth; Perron, Walter; Poscher, Ralf; Riescher, Gisela; Würtenberger, Thomas, Hrsg.): Resilienz in der offenen Gesellschaft. Nomos, 2012.
- [Bu00a] Bundestags-Drucksache 14/4329, 13. Oktober 2000.
- [Bu00b] Bundestags-Drucksache 461/00, 18. August 2000.
- [Bu01] Bundestags-Drucksache 14/5793, 4. April 2001.
- [CCK13] Cichowski, Janusz; Czyżewski, Andrzej; Kostek, Bożena: Visual Data Encryption for Privacy Enhancement in Surveillance Systems. In (Blanc-Talon, Jacques; Kasinski, Andrzej; Philips, Wilfried; Popescu, Dan; Scheunders, Paul, Hrsg.): Advanced Concepts for Intelligent Vision Systems: 15th International Conference, ACIVS 2013, Poznań, Poland, October 28-31, 2013. Proceedings. Springer International Publishing, S. 13–24, 2013.
- [Ch09] Cheung, S.-C.S.; Venkatesh, M.V.; Paruchuri, J.K.; Zhao, J.; Nguyen, T.: Protecting and Managing Privacy Information in Video Surveillance Systems. In (Senior, Andrew, Hrsg.): Protecting Privacy in Video Surveillance. Springer London, London, S. 11–33, 2009.
- [CKM09] Carrillo, Paula; Kalva, Hari; Magliveras, Spyros: Compression Independent Reversible Encryption for Privacy in Video Surveillance. EURASIP J. Inf. Secur., 2009:5:1–5:13, Januar 2009.
- [CPV06] Cucchiara, Rita; Prati, Andrea; Vezzani, Roberto: A system for automatic face obscuration for privacy purposes. Pattern Recognition Letters, 27(15):1809 – 1815, 2006. Vision for Crime Detection and Prevention.
- [DE06] Dufaux, F.; Ebrahimi, T.: Scrambling for Video Surveillance with Privacy. In: 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06). S. 160–160, June 2006.
- [De10] Deterbeck, Steffen: Allgemeines Verwaltungsrecht. Verlag C.H.Beck, 8. Auflage, 2010.
- [DT05] Dalal, Navneet; Triggs, Bill: Histograms of Oriented Gradients for Human Detection. In: Proc. CVPR. IEEE, S. 886–893, 2005.
- [Eu12] Europäische Kommission: Vorschlag für Verordnung des europäischen Parlaments und des Rates zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr (Datenschutz-Grundverordnung). 2012.
- [GG00] Grin, John; Grunwald, Armin, Hrsg. Vision assessment: shaping technology in 21st century society towards a repertoire for technology assessment. Springer, 2000.
- [GK14] Gola, Peter; Klug, Christoph: Zur datenschutzrechtlichen Zulässigkeit der Videoüberwachung. Recht der Datenverarbeitung, S. 65–74, 2014.
- [Gr99] Grunwald, Armin: TA-Verständnis in der Philosophie. In (Bröchler, Stephan; Simonis, Georg; Sundermann, Karsten, Hrsg.): Handbuch Technikfolgenabschätzung, S. 73–81. Edition Sigma, 1999.
- [Hu14] Huber, Matthias; Müller-Quade, Jörn; Nilges, Tobias; Thal, Carolin: A Provably Privacy Preserving Video Surveillance Architecture for an Assisted Living Community. In: 44. Jahrestagung der Gesellschaft für Informatik, Informatik 2014, Big Data - Komplexität meistern, 22.-26. September 2014 in Stuttgart, Deutschland. S. 563–574, 2014.

- [LH08] Leykin, A.; Hammoud, R.: Real-time estimation of human attention field in LWIR and color surveillance videos. In: Proc. CVPRW. IEEE, S. 1–6, 2008.
- [Li07] Liu, Xiaoming; Krahnstoeber, N.; Yu, Ting; Tu, P.: What are customers looking at? In: Proc. AVSS. IEEE, S. 405–410, 2007.
- [Ly09] Lyon, David: Surveillance Studies. An Overview. Polity Press, 2009.
- [Ot05] Ott, Konrad: Technikethik. In (Nida-Rümelin, Julian, Hrsg.): Angewandte Ethik: die Bereichsethiken und ihre theoretische Fundierung: ein Handbuch, S. 568–647. Kröner, 2. Auflage, 2005.
- [Pe99] Petermann, Thomas: Technikfolgen-Abschätzung – Konstituierung und Ausdifferenzierung eines Leitbilds. In (Bröchler, Stephan; Simonis, Georg; Sundermann, Karsten, Hrsg.): Handbuch Technikfolgenabschätzung, S. 17–49. Edition Sigma, 1999.
- [Pe06] Pegarkov, Daniel D.: National Security Issues. Nova Publishers, Januar 2006.
- [PLCFR15] Padilla-Lopez, Jose Ramon; Chaaaraoui, Alexandros Andre; Florez-Revuelta, Francisco: Visual privacy protection methods: A survey. Expert Systems with Applications, 42(9):4177 – 4195, 2015.
- [PP10] Paar, Christof; Pelzl, Jan: Introduction to Public-Key Cryptography. In: Understanding Cryptography. Springer, 2010.
- [RDK11] Ruhrmair, Ulrich; Devadas, Srinivas; Koushanfar, Farinaz: Security based on Physical Unclonability and Disorder. In: Introduction to Hardware Security and Trust. Springer, S. 65–102, 2011.
- [Sc07] Schiff, J.; Meingast, M.; Mulligan, D. K.; Sastry, S.; Goldberg, K.: Respectful cameras: detecting visual markers in real-time to address privacy concerns. In: 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems. S. 971–978, Oct 2007.
- [Se05] Senior, A.; Pankanti, S.; Hampapur, A.; Brown, L.; Tian, Ying-Li; Ekin, A.; Connell, J.; Shu, Chiao Fe; Lu, M.: Enabling video privacy through computer vision. IEEE Security Privacy, 3(3):50–57, May 2005.
- [Sh11] Shi, Zhijie; Ma, Chujiao; Cote, Jordan; Wang, Bing: Hardware Implementation of Hash Functions. In: Introduction to Hardware Security and Trust. Springer, S. 27–50, 2011.
- [TZ09] Trojanow, Ilija; Zeh, Juli: Angriff auf die Freiheit: Sicherheitswahn, Überwachungsstaat und der Abbau bürgerlicher Rechte. Hanser, 2009.
- [WSS10] Walk, Stefan; Schindler, Konrad; Schiele, Bernt: Disparity Statistics for Pedestrian Detection: Combining Appearance, Motion and Stereo. In: Proc. ECCV. IEEE, S. 182–195, 2010.

Privacy-Dashcam – Datenschutzfreundliche Dashcams durch Erzwingen externer Anonymisierung

Paul Georg Wagner,¹ Pascal Birnstill,² Erik Krempel,³ Sebastian Bretthauer,⁴ Jürgen Beyerer⁵

Abstract: Datenschützer sehen in Dashcams Videoüberwachungsanlagen, die von Privatpersonen im öffentlichen Raum betrieben werden und die durch das Aufzeichnen personenbezogener Daten in unverhältnismäßigem Ausmaß in das Recht auf informationelle Selbstbestimmung betroffener Bürger eingreifen. Eine Dashcam kann demnach nur dann datenschutzgerecht sein, wenn sie die Erhebung personenbezogener Daten vermeidet. Dies kann bspw. geschehen, indem mindestens die Gesichter von Passanten sowie KFZ-Kennzeichen automatisch erkannt und unkenntlich gemacht werden. Obwohl Anonymisierungsverfahren mit hinreichend hoher Genauigkeit existieren, kommen diese aufgrund ihres hohen Rechenaufwands nicht für den Einsatz in handelsüblichen Dashcams in Betracht. Dieser Beitrag stellt einen neuen Ansatz vor, der die Anonymisierung von verschlüsselt gespeicherten Dashcam-Videos auf einem separaten Rechner erzwingt, bevor der Benutzer darauf Zugriff erhält. Diese Videos werden auf ihrem Weg vom Speichermedium zu einer Anonymisierungskomponente mittels Methoden der Datenflusskontrolle überwacht, sodass beliebige Dashcams um Datenschutzmechanismen erweitert werden können.

Keywords: Dashcams, Datenschutz, Privacy, Nutzungskontrolle, Datenflusskontrolle

1 Einleitung

Dashcams, die aus einem Fahrzeug heraus das Verkehrsgeschehen aufzeichnen, sind in den letzten Jahren auch in Europa immer populärer geworden. Ihre Benutzer versprechen sich von den Aufzeichnungen eine vereinfachte Schadensregulierung bei Verkehrsunfällen, insbesondere zum Nachweis der eigenen Unschuld. Gleichzeitig ist die Verbreitung von Dashcams von einer datenschutzrechtlichen Kontroverse begleitet, da den Sicherheitsbedürfnissen der Betreiber gewichtige Nachteile gegenüber stehen. So können Dashcams als mobile Videoüberwachungsanlagen betrachtet werden, deren Betrieb durch Privatpersonen im öffentlichen Raum kaum zu legitimieren ist. Für alle derzeit existierenden Arten von Dashcams gilt außerdem, dass eine Erhebung personenbezogener Daten durch im Bild erfasste Fußgänger und KFZ-Kennzeichen angenommen werden muss. Somit liegt ein Eingriff in das Recht auf informationelle Selbstbestimmung der Betroffenen vor, dessen Verhältnismäßigkeit im Hinblick auf die genannten Sicherheitsbedürfnisse der Dashcam-Benutzer umstritten ist. Im Vergleich dazu stellen sogenannte *Crashcams* eine

¹ Fraunhofer IOSB, paul-georg.wagner@iosb.fraunhofer.de

² Fraunhofer IOSB, pascal.birnstill@iosb.fraunhofer.de

³ Fraunhofer IOSB, erik.krempel@iosb.fraunhofer.de

⁴ Zentrum für Angewandte Rechtswissenschaft (ZAR) am Karlsruher Institut für Technologie und an der Goethe-Universität Frankfurt a. M. bei Prof. Dr. Indra Spiecker gen. Döhmman, LL. M. sowie an der dortigen Forschungsstelle Datenschutz, sebastian.bretthauer@kit.edu

⁵ Fraunhofer IOSB & Karlsruher Institut für Technologie (KIT), juergen.beyerer@iosb.fraunhofer.de

datenschutzfreundlichere Methode dar, da diese nur ereignisbezogen aufzeichnen, etwa durch Beschleunigungssensoren ausgelöst. Durch Verwendung eines Ringpuffers, der im Ereignisfall nicht mehr überschrieben wird, kann eine Crashcam Videobilder im Zeitraum von bspw. jeweils einer Minute vor und nach dem Ereignis aufnehmen. Allerdings ist auch bei diesen Kameras eine Erhebung personenbezogener Daten kaum zu vermeiden.

Während in Deutschland noch keine obergerichtliche Entscheidung vorliegt, sind Dashcams in Österreich bereits seit 2012 unzulässig. Weiterhin erklärte das Bundesverwaltungsgericht in Wien 2015 auch eine Crashcam für unzulässig, die zum Schutz der Privatsphäre erfasster Personen mit reduzierter Auflösung betrieben wurde. Das Gericht sah es nicht als erwiesen an, dass die geringe Auflösung der Kamera eine unrechtmäßige Erhebung personenbezogener Daten wirksam verhindere, und erbat explizit eine Stellungnahme des Klägers, weshalb die Kamera datenschutzrelevante Bereiche von Bildern wie etwa Gesichter und KFZ-Kennzeichen nicht unkenntlich mache. Technisch kann festgestellt werden, dass entsprechende Anonymisierungsverfahren existieren, aber für eine Integration in transportable Kameras zu rechenintensiv sind.

Dieser Beitrag stellt einen Ansatz vor, der die zur Anonymisierung benötigte Bildauswertung auf sichere Weise an einen separaten, leistungsfähigen Rechner auslagert. Hierzu wird ein spezielles Speichermedium eingesetzt, das Videos sofort verschlüsselt. Ein Zugriff auf das Speichermedium ist nur über einen Rechner mit Mechanismen der Datenfluss- und Nutzungskontrolle möglich. Diese stellen sicher, dass die Aufzeichnungen der Dashcam vor dem Zugriff durch den Benutzer anonymisiert werden. Weiterhin wird gewährleistet, dass der Benutzer aufgezeichnete Daten weder modifizieren noch löschen kann.

2 Verwandte Arbeiten

Rinner und Winkler⁶ stellen mit dem Prototyp *TrustEYE.M4* eine vertrauenswürdige Kamera mit integrierten Privacy-Enhancing-Technologies (PET) vor. Diese basiert auf einem Raspberry PI, einem preiswerten Miniaturrechner, der die erfassten Daten sofort verarbeitet bevor diese gespeichert oder übertragen werden. Die verfügbare Rechenleistung ist jedoch limitiert und es können nur einfache Verfahren zur Anonymisierung genutzt werden. Ob hierdurch eine Anonymisierung erreicht werden kann, die einen hinreichenden Schutz personenbezogener Bildbereiche bietet, zugleich aber auch für den Nachvollzug von Verkehrsunfällen ausreichendes Material liefert, ist unklar.

Soll die Anonymisierung der Dashcam-Videos an einen separaten Rechner ausgelagert werden, so muss auf ein portables Speichermedium aufgezeichnet werden, das der Benutzer selbst an den vorbereiteten Rechner anschließt. Hierbei muss sichergestellt werden, dass der Benutzer nicht auf die Videodaten zugreifen kann, bevor diese anonymisiert wurden. Mittels Mechanismen der Nutzungskontrolle^{7,8} können Nutzungen von Daten in einer kontrollierten Umgebung kontinuierlich überwacht werden – auch über Systemgrenzen

⁶ Rinner und Winkler, Privacy-protecting Smart Cameras, ICDSC 2015.

⁷ Park und Sandhu, Towards Usage Control Models: Beyond Traditional Access Control, SACMAT 2002.

⁸ Pretschner et al., Distributed Usage Control, Commun. ACM, 49(9), 2006.

hinweg. Bei der Nutzungskontrolle handelt es sich um eine Verallgemeinerung der Zugriffskontrolle, wobei die Kontrolle über Nutzungen von Daten auch nach dem dem initialen Zugriff noch möglich ist. Hierzu ist allerdings eine kontinuierliche Überwachung der zu schützenden Daten erforderlich, weshalb die Nutzungskontrolle darüber hinaus häufig mit einer Datenflusskontrolle erweitert wird. Durch die Datenflusskontrolle können Nutzungsrestriktionen nicht nur in Abhängigkeit von Systemereignissen, sondern ebenso in Form von (un-)zulässigen Datenflüssen formuliert werden.⁹

Zur Anonymisierung datenschutzrelevanter Bildbereiche sind geeignete Bildauswerteargorithmen und Anonymisierungstechniken erforderlich. Hinreichend leistungsstarke Bildauswerteargorithmen, die datenschutzkritische Bildbereiche in Echtzeit detektieren können, sind jedoch im Allgemeinen sehr rechenintensiv und somit nur bedingt für den Einsatz in mobilen Kamerasystemen geeignet. So erreichten Janard und Marurngsith mit einer Local-Binary-Pattern-Gesichtserkennung auf dem Raspberry PI lediglich etwa 17 QVGA-Bilder (320x240 Pixel) pro Sekunde.¹⁰ Auch einige Anonymisierungstechniken sind mit Miniaturrechnern nicht in Echtzeit zu bewerkstelligen. So war für *TrustEYE.M4* eine eigene Implementierung des Privacy-Filters nötig, da sich die Standardalgorithmen der OpenCV-Bibliothek für den Raspberry PI als zu komplex herausgestellt haben.

3 Rechtliche Perspektive

3.1 Unter geltendem Datenschutzrecht

Bereits unter geltendem Datenschutzrecht wird der Einsatz von Dashcams kontrovers diskutiert. Teilweise werden Dashcams als datenschutzrechtlich zulässig erachtet,¹¹ teilweise hingegen als ausdrücklich unzulässig.¹² In der datenschutzrechtlichen Diskussion wird erörtert, ob Datenschutzrecht gar nicht anwendbar sei, falls die Nutzung einer Dashcam eine ausschließlich persönliche oder familiäre Tätigkeit darstellen würde,¹³ ob und inwiefern überhaupt personenbezogene Daten durch die Dashcam-Aufnahme betroffen seien¹⁴ und ob und wie Dashcams generell mit § 6 b BDSG in Einklang gebracht werden können.¹⁵ Eine pauschale Antwort ist jedenfalls nicht möglich, insbesondere unter Berücksichtigung der verschiedenen technischen Gestaltungsmöglichkeiten von Dashcams.¹⁶ Dieser technische Aspekt wird bisher von der Rechtsprechung zu Unrecht nicht berücksichtigt.¹⁷

So kommt der technischen Gestaltung von Dashcams aber gerade eine besonders übertragende Bedeutung bei der datenschutzrechtlichen Bewertung zu. Entscheidend ist die

⁹ Harvan und Pretschner, State-based Usage Control Enforcement with Data Flow Tracking Using System Call Interposition, NSS 2009.

¹⁰ Janard/Marurngsith, Accelerating real-time face detection on a raspberry pi telepresence robot, INTECH 2015.

¹¹ Wirsching, NZV 2016, 13 (16); Knyrim/Trieb, ZD 2014, 547 ff.; Fuchs, ZD 2015, 212 (217).

¹² Ernst, CR 2015, 620 (624); Reibach, DuD 2012, 157 (160); AG München, Urt. v. 13.08.2014 – 345 C.

¹³ Ernst, CR 2015, 620 (622); Reibach, DuD 2012, 157 ff.; Balzer/Nugel, NJW 2014, 1622 (1625); Fuchs, ZD 2015, 212 (214 ff.).

¹⁴ Balzer/Nugel, NJW 2014, 1622 (1625); Fuchs, ZD 2015, 212 (213 ff.).

¹⁵ Ernst, CR 2015, 620 (623); Balzer/Nugel, NJW 2014, 1622 (1626 ff.).

¹⁶ In diese Richtung auch Knyrim/Trieb, ZD 2014, 547 (547).

¹⁷ Knyrim/Trieb, ZD 2014, 547 (547 ff.).

technische Ausgestaltung des Aufnahmesystems, insbesondere was die Auflösung der Bilder angeht, den Speicherzeitpunkt, die Speicherdauer der Aufnahmen, die Zugriffsmöglichkeiten auf die gespeicherten Daten sowie die allgemeine technische Architektur der Dashcam.¹⁸ Eine dauerhafte Speicherung von Klarbildern aus einer Dashcam ist jedenfalls unzulässig, da dies einen besonders schwerwiegenden Eingriff in Persönlichkeitsrechte der von der Dashcam Betroffenen darstellt.¹⁹ Künftig muss deshalb der technischen Spezifikation bei der datenschutzrechtlichen Zulässigkeitsbewertung eine wesentlich größere Bedeutung beigemessen werden. Das gilt bereits ohnehin für das geltende Datenschutzrecht und erst Recht für die ab 2018 kommende Europäische Datenschutzgrundverordnung.

3.2 Unter der kommenden Datenschutz-Grundverordnung (DS-GVO)

Ab dem 25. Mai 2018 gilt die DS-GVO in jedem Mitgliedstaat unmittelbar (Art. 288 AEUV, Art. 99 Abs. 2 DS-GVO), sodass die Datenschutzrichtlinie und die in ihrer Folge bzw. Umsetzung erlassenen nationalen Datenschutzgesetze ihre Geltung verlieren (Art. 94 Abs. 1 DS-GVO). Ausnahmen kommen allerdings dort in Betracht, wo die DS-GVO Öffnungsklauseln bereithält, die den nationalen Staaten dann auch weiterhin eigene Befugnisse einräumen.²⁰ Die datenschutzrechtliche Zulässigkeit von Dashcams richtet sich gleichwohl nach der DS-GVO, da derartige videoüberwachungsähnliche Systeme überwiegend von Privatpersonen genutzt werden - einmal abgesehen von der Nutzung in Polizeiautos -, sodass auch keine Öffnungsklausel für den öffentlichen Bereich in Betracht kommt (vgl. etwa Art. 2 Abs. 2 d) und Art. 6 Abs. 2 DS-GVO). Andere Öffnungsklauseln sind ebenfalls nicht ersichtlich.²¹ Ebenso unterfällt die Nutzung von Dashcams nicht dem „Familienprivileg“ nach Art. 2 Abs. 2 (c) DS-GVO, da die Dashcam-Aufzeichnung nicht ausschließlich zur Ausübung persönlicher oder familiärer Tätigkeit erfolgt, soll sie doch überwiegend als Beweismittel in einem späteren Prozess eingesetzt werden.²² Für die klassische Videoüberwachung hat dies der Europäische Gerichtshof bereits so auch ausdrücklich festgestellt.²³

Damit verbleibt es bei den sehr allgemein gehaltenen Regelungen der DS-GVO, da es an einer spezifischen Videoüberwachungsnorm fehlt.²⁴ In Betracht kommt Art. 6 Abs. 1 (f) DS-GVO, der jedoch nur auf eine Interessenabwägung zwischen dem Dashcam-Betreiber und dem Betroffenen abstellt und keine weiteren eingrenzenden Tatbestandsmerkmale aufweist. Bisherige Regelungen in einzelnen europäischen Mitgliedstaaten zur Videoüberwachung haben jedoch gerade solche expliziten weiteren einschränkenden Tatbestandsmerkmale.²⁵ Damit ist aber fraglich, ob ein derart offen und weit gefasster Erlaubnistatbestand, wie ihn Art. 6 Abs. 1 (f) DS-GVO darstellt, überhaupt hinreichend bestimmt

¹⁸ So bereits Knyrim/Trieb, ZD 2014, 547 (548); Balzer/Nugel, NJW 2014, 1622 (1627).

¹⁹ VG Ansbach, Urt. v. 12.08.2014 – AN 4 K 13.01634 = ZD 2014, 590 (593).

²⁰ Ausführlich hierzu Benecke/Wagner, DVBl 2016, 600 ff.

²¹ Ausführlich Benecke/Wagner, DVBl 2016, 600 (604 Fn. 27).

²² Vgl. hierzu ausführlich Reibach, DuD 2012, 157 ff.

²³ EuGH, Urt. v. 11.12.2014 – C-212/13 = CR 2015, 101 ff. (m. Anm. Bretthauer).

²⁴ Vgl. bereits dazu Bretthauer/Krempel/Birnstill, CR 2015, 239 (242).

²⁵ Vgl. etwa hierzu § 6 b BDSG, § 50 a ff ÖDSG, §§ 16 ff. Litauisches Datenschutzgesetz, § 26 Dänisches Datenschutzgesetz, § 6 a Liechtensteinisches Datenschutzgesetz und §§ 36 ff. Norwegisches Datenschutzgesetz.

ist im Sinne des Art. 52 Abs. 1 EU-Grundrechte-Charta, um eine derartig grundrechtssensible Technologie zu regulieren.²⁶ Der rechtliche Maßstab läuft ausschließlich auf eine reine Interessenabwägung hinaus. Für den Dashcam-Betreiber lässt sich sein Interesse an einer effektiven Beweismittelführung in einem Unfallprozess und einer effektiven Strafverfolgung bei einer möglichen Unfallflucht des Gegners anführen.²⁷ Die von der Dashcam Betroffenen können sich auf den Schutz personenbezogener Daten nach Art. 7, 8 EU-Gr-Charta berufen. In diese Abwägung muss aber ebenso die technische Spezifikation der Dashcam eingestellt werden, sodass datenschutzfreundliche Techniken Berücksichtigung finden können (vgl. Erwägungsgrund 78 DS-GVO und Art. 25 DS-GVO).

Führt man sich die in den einzelnen Mitgliedstaaten derzeit vorhandenen Regelungen zur Videoüberwachung vor Augen und überträgt man die aus dem deutschen Verfassungsrecht bekannten Anforderungen an die Bestimmtheit von Rechtsnormen auf Art. 6 Abs. 1 f) DS-GVO im Hinblick auf den Einsatz von Dashcams, so erweist sich die Regelung als zu unbestimmt, da es neben der Interessenabwägung an weiteren eingrenzenden Tatbestandsmerkmalen fehlt. Die Nutzung von Dashcams ist dann im privaten Bereich jedenfalls nicht mehr von der DS-GVO gedeckt.

4 Systemmodell

Das im Folgenden spezifizierte und formalisierte System arbeitet mit einer handelsüblichen Dashcam und einem Speichermedium, auf welches Video- und Metadaten geschrieben werden. Videos werden solange als klassifizierte Daten behandelt, bis eine entsprechende Softwarekomponente datenschutzrelevante Bereiche entfernt oder anonymisiert hat. Diese Deklassifikation erfolgt auf einem separaten Rechner. Während der Übertragung vom Speichermedium auf die Deklassifikationskomponente müssen die Daten adäquat geschützt werden. Die folgenden Abschnitte gehen auf Schutzziele, Angreifer und Vertrauensbeziehungen im betrachteten Szenario ein.

Schutzziele Schutzgüter sind personenbezogene Merkmale, die in aufgezeichneten Videodaten eingebettet sind, insbesondere Gesichter und KFZ-Kennzeichen. Hinzu kommen Metadaten der Videos, wie etwa Zeitstempel oder GPS-Positionsdaten. Wichtigstes Schutzziel ist die Vertraulichkeit der Videodaten bzw. der darin enthaltenen personenbezogenen Daten. Ein weiteres Schutzziel ist die Integrität der Daten (einschließlich der Metadaten). Sollen diese als Beweismittel vorgebracht werden können, muss sichergestellt sein, dass sie nicht manipuliert wurden. Schließlich muss die Authentizität von Benutzern mit erweiterten Berechtigungen (bspw. Strafverfolgungsbehörden) gewährleistet werden.

²⁶ So bereits Bretthauer/Krempel, in: Schweighofer/Kummer/Hötendorfer (Hrsg.), *Transparenz – Tagungsband des 17. Internationalen Rechtsinformatik Symposions*, 2014, S. 525, 532; zu den Anforderungen nach Art. 52 EU-Grundrechte-Charta im Einzelnen vgl. Rieckhoff, *Der Vorbehalt des Gesetzes im Europarecht*, 2007, S. 155 ff.

²⁷ So etwa Ernst, CR 2015, 620 (623).

Angreifermodell Im Mittelpunkt steht ein Angreifer, der datenschutzrelevante Inhalte aufgezeichneter Videos an eine interessierte dritte Partei weitergeben möchte. Dieser sog. *Privacy-Attacker* hat Zugriff auf die Kamera, das verschlüsselte Speichermedium und den zur Deklassifikation der Videodaten verwendeten Rechner. Gegenüber diesem Angreifer darf das System keine datenschutzrelevanten Inhalte preisgeben. Dies wird insbesondere dadurch erschwert, dass es sich beim *Privacy-Attacker* nicht um einen externen Dritten handelt, sondern um den Benutzer der Dashcam selbst. (Diese ambivalente Betrachtung des Benutzers ist typisch für Einsatzszenarien der Nutzungskontrolle.) Dennoch kann ein Dashcam-System nur datenschutzgerecht sein, wenn es gegenüber dem *Privacy-Attacker* widerstandsfähig ist.

Ein *Modifying-Attacker* verfolgt das Ziel, aufgezeichnete Video- und/oder Metadaten zu seinem Vorteil zu verändern. Schutzmechanismen gegen ihn müssen daher die Integrität der klassifizierten Videos und der Metadaten sicherstellen. Ein Schutz der deklassifizierten Videos oder exportierter Metadaten ist nicht notwendig. Auch der *Modifying-Attacker* hat Zugriff auf das Speichermedium und das Deklassifikationssystem.

Ein ähnliches Ziel verfolgt der sog. *Destructive-Attacker*, der unvorteilhafte Video- bzw. Metadaten zerstören möchte, etwa nach einem selbst verschuldeten Verkehrsunfall. Da eine mechanische Zerstörung des Speichermediums nicht verhindert werden kann, wird dieser Angreifer durch die Annahme beschränkt, dass er Daten für Dritte undetektierbar löschen möchte. Ein Schutz der deklassifizierten Videos oder exportierter Metadaten ist nicht notwendig. Der *Destructive-Attacker* hat Zugriff auf das Speichermedium und das Deklassifikationssystem, sieht jedoch von mechanischer Zerstörung ab.

Vertrauensmodell Im betrachteten Szenario interagieren die im Folgenden beschriebenen Akteure mit dem Dashcam-System. Der *Betreiber* benutzt das System, um bei einem möglichen Verkehrsunfall die Aufzeichnungen der Kamera als Beweismittel zu nutzen. Neben der Nutzung als Betreiber kann er als *Privacy-Attacker*, *Modifying-Attacker* und *Destructive-Attacker* agieren. Da das System dem Betreiber nicht vertraut, ihn sogar als möglichen Angreifer betrachtet, muss er nicht explizit am System authentifiziert werden.

Die *Strafverfolgungsbehörde* hat ein rechtmäßiges Interesse an den aufgezeichneten Videodaten, um diese gegebenenfalls als Beweismittel in einem Prozess vorzubringen. Ihr ist es erlaubt, auf die klassifizierten Videos zuzugreifen. Dies geschieht üblicherweise auf eine richterliche Anordnung hin, da solche Zugriffe einen Eingriff in die Persönlichkeitsrechte erfasster Personen darstellen. Der Strafverfolgungsbehörde wird vollumfänglich vertraut, d.h. ihr ist es gestattet, die Schutzmechanismen des Systems zu deaktivieren. Die Authentifizierung dieses Akteurs gegenüber dem System ist daher essentiell.

Der *Administrator* nimmt das Deklassifikationssystem in Betrieb. An diesem legt er initial die Zugriffsrechte für alle Akteure fest. Dem Administrator muss voll vertraut werden. Im Folgenden wird angenommen, dass der Administrator nur bei der Inbetriebnahme Zugang zum System hat, sodass er für die weitere Sicherheitsanalyse des Systemmodells keine Rolle spielt. Abschnitt 7 geht auf einige grundlegende Vorkehrungen ein, die der Administrator bei der Inbetriebnahme des Deklassifikationssystems treffen muss.

5 Systemspezifikation

In den folgenden Abschnitten wird das konkrete System spezifiziert. Dafür werden die einzelnen Komponenten und ihre Interaktionen untereinander definiert. Schließlich wird eine mögliche Implementierung des entwickelten Systemmodells vorgestellt.

5.1 Systemkomponenten

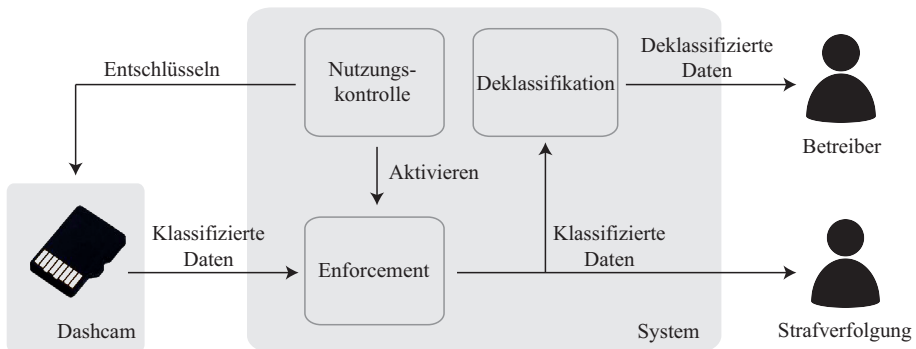


Abb. 1: Realisierung des Systemmodells

Das Systemmodell unterscheidet zwischen der *Kamera* an sich und einem *Speicher*, typischerweise einer SD-Karte, auf dem die erfassten Daten für die weitere Verarbeitung vorgehalten werden. Für das Modell ist es unerheblich, wie Videos mit potentiell datenschutzrelevanten Bereichen in den Speicher gelangen. Von der Kamera selbst wird daher im Folgenden abstrahiert. Der in Abb. 1 dargestellte Ansatz kann somit für alle handelsüblichen Geräte genutzt werden. Der Speicher ist dagegen von herausragender Bedeutung. Da er die klassifizierten Videos speichert, muss er sicherstellen, dass nur dann auf die Daten zugegriffen wird, wenn diese auf dem zugreifenden System geschützt sind.

Die Komponente *Deklassifikation* liest klassifizierte Videos, also Videos mit möglicherweise datenschutzrelevanten Bereichen, und nutzt eine Deklassifikationsfunktion, um eine deklassifizierte Repräsentation der Videos zu erzeugen. Dies geschieht in der Regel durch eine Anonymisierung der datenschutzrelevanten Bildbereiche, etwa durch Schwärzen oder Verpixeln. Nur deklassifizierte Videos dürfen das System verlassen, um dem Benutzer angezeigt zu werden. Die klassifizierten Daten hingegen dürfen die Deklassifikationskomponente keinesfalls verlassen.

Da der einzig rechtmäßige Zugriff auf den Speicher durch die Deklassifikationskomponente erfolgt, müssen alle weiteren Zugriffe darauf verhindert werden. Eine mögliche Lösung hierfür besteht darin, beiden Komponenten ein gemeinsames Geheimnis zu geben und damit die vorhandenen Daten zu verschlüsseln. Ein weitaus flexiblerer Ansatz ist es jedoch, die Deklassifikationskomponente und den Speicher nur lose miteinander zu koppeln. Um dies zu erreichen wird eine weitere Komponente, die *Nutzungskontrolle*, eingeführt. Diese Komponente stellt Datenverbindungen zwischen dem Speicher und beliebig vielen Deklassifikatoren her, vorausgesetzt diese nutzen ausreichend gute Deklassifikationsfunktionen,

um die datenschutzrelevanten Bereiche zu schützen. Authentifiziert sich ein entsprechender Benutzer, in unserem Beispiel eine Strafverfolgungsbehörde, an der Nutzungskontrolle, erhält dieser ebenfalls Zugriff auf die noch nicht anonymisierten Daten. In diesem Fall wird davon ausgegangen, dass die entsprechende Instanz die erhaltenen Daten verantwortungsvoll nutzt und nicht für Unbefugte zugänglich macht.

Den eigentlichen Kern des Systems bildet die *Enforcement*-Komponente. Sie überwacht die vom Speicher in das System fließenden klassifizierten Daten und unterbindet eine missbräuchliche Verwendung innerhalb des Systems. Die Enforcement-Komponente überwacht hierzu sämtliche Datenflüsse, die auf dem System auftreten, und unterbindet diejenigen Operationen, die klassifizierte Daten an den Betreiber weitergeben könnten.

5.2 Interaktion

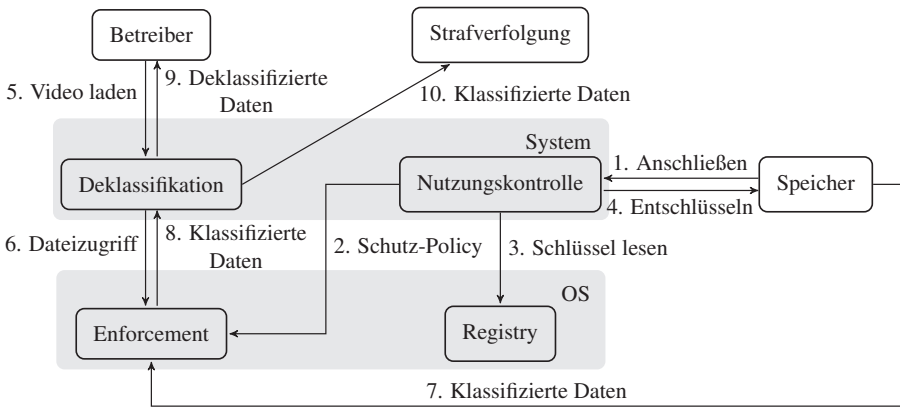


Abb. 2: Überblick über die Systeminteraktion

Abb. 2 zeigt die Interaktion der Komponenten, sobald der Betreiber seinen Speicher an einen Rechner anschließt, auf dem das Deklassifikationssystem aktiv ist. Der Speicher besteht aus einer verschlüsselten SD-Karte, auf der klassifizierte Daten in Form von Dashcam-Videos vorhanden sind. Das Entschlüsseln der SD-Karte geschieht mittels eines zufällig generierten gemeinsamen Geheimnisses (Schlüssel), das dem Betreiber selbst nicht bekannt sein darf und das daher durch die Nutzungskontrolle in einer geschützten Umgebung abgespeichert wird, etwa als verschlüsselter Eintrag in der Registry. Der Speicher kann dementsprechend nur an einem Rechner betrieben werden, der mit einer vertrauenswürdigen Nutzungskontrolle ausgestattet ist. Neben der Verwendung eines symmetrischen Verfahrens zur Verschlüsselung der Speicherkarte ist auch der Einsatz eines asymmetrischen Verfahrens, bei welchem sich der private Schlüssel in der Nutzungskontrolle und der öffentliche Schlüssel auf der Speicherkarte befindet, möglich. Allerdings ist dieses Design mit den meisten verfügbaren asymmetrisch verschlüsselnden Speicherkarten nicht umsetzbar, da der private Schlüssel in der Regel nicht exportiert werden kann.

Ist auf dem Rechner eine Nutzungskontrolle verfügbar, so erkennt diese den verschlüsselten Speicher und aktiviert die Enforcement-Komponente. Diese Aktivierung erfolgt mit Hilfe

einer Schutz-Policy, welche das angeschlossene Speichergerät eindeutig identifiziert. Nach der Aktivierung dieser Schutzmechanismen unterliegen sämtliche Datenflüsse von und auf den Speicher der Kontrolle der Enforcement-Komponente. Daher kann die Nutzungskontrolle nun das gemeinsame Geheimnis verwenden, um den Speicher zu entschlüsseln. Ab diesem Zeitpunkt liegen die klassifizierten Daten auf dem Speicher im Klartext vor und können von beliebigen Systemprozessen ausgelesen werden. Dies zu verhindern obliegt der Enforcement-Komponente, welche sämtliche Lesezugriffe auf den entschlüsselten Speicher überwacht und nur solche mit authentifiziertem Ziel zulässt (d.h. ausgehend von der Deklassifikationskomponente oder der Strafverfolgungsbehörde). Schreibzugriffe auf den Speicher werden ebenfalls untersagt, insbesondere auch für die Strafverfolgungsbehörden. Damit stellt die Enforcement-Komponente sicher, dass die klassifizierten Daten von der Speicherkarte lediglich in die Deklassifikationskomponente und zu Strafverfolgungsbehörden fließen können. Der Betreiber der Dashcam kann die klassifizierten Daten an keiner Stelle abgreifen, da die Deklassifikationskomponente die Videos nur in deklassifizierter Form, also anonymisiert, darstellt. Es wird somit sichergestellt, dass sowohl die Vertraulichkeit als auch die Integrität der klassifizierten Daten geschützt ist. Lediglich der Strafverfolgungsbehörde erlaubt die Enforcement-Komponente nach expliziter Authentifizierung an der Nutzungskontrolle, lesend auf die klassifizierten Daten zuzugreifen.

5.3 Implementierung

Bei der Implementierung des beschriebenen Systems ist insbesondere der Entwurf der Nutzungskontrolle und der Enforcement-Komponente entscheidend. Die Nutzungskontrolle wurde als Windows-Dienst realisiert, welcher stets im Hintergrund auf angeschlossene Speichermodule lauscht, für jene die Schutzmechanismen aktiviert und weitere Steuerungsaufgaben übernimmt. Die Enforcement-Komponente besteht aus einem Dateisystem-Filtreiter für Windows, welcher sich im Kernel registriert, sämtliche Dateizugriffe im System filtert und gegebenenfalls untersagt (vgl. Abb. 2). Dieser Treiber blockiert, sobald die Speicherkarte entschlüsselt wurde, sämtliche unerwünschten Datenflüsse im System.

Als Deklassifikationskomponente wird nun ein Programm benötigt, das in der Lage ist, datenschutzkritische Bereiche der Videos zu detektieren und zu anonymisieren. Durch die lose Kopplung zwischen Nutzungskontrolle und Deklassifikationskomponente kann für diese Aufgabe jedes geeignete Drittprogramm eingesetzt werden.

Weiter ist zu klären, wie die Kamera Videos auf die verschlüsselte Speicherkarte schreiben kann, ohne das zur Ver-/Entschlüsselung benötigte Geheimnis zu kennen. Dies wird durch den sog. *Postbox-Mode* der Karte möglich, der das Schreiben von Daten auch ohne Authentifizierung erlaubt, nicht jedoch das Lesen oder Modifizieren. Die Kamera kann somit Überwachungsvideos aufzeichnen, die zu jedem Zeitpunkt auf der Speicherkarte in verschlüsselter Form aufbewahrt werden. Wird ein asymmetrisches Verschlüsselungsverfahren eingesetzt, entfällt diese Problematik.

Ferner schreiben Dashcams dauerhaft auf die verwendete Speicherkarte. Sobald der Speicher voll ist, werden die ältesten Videos überschrieben. Dies kann durch den Destructive-

Attacker ausgenutzt werden, um unerwünschte Videos zu löschen. Dem Problem der begrenzten Kapazität von Speicherkarten kann wie folgt begegnet werden. Werden ereignisbasiert aufzeichnende Crashcams benutzt, ist der begrenzte Speicherplatz unproblematisch, da auf eine Speicherkarte über 200 Unfallvideos gespeichert werden können. Diese anlassbezogene Speicherung ist auch aus Datenschutzgründen vorzuziehen. Alternativ kann das Deklassifikationssystem um einen Modus erweitert werden, der alle Videos von der Speicherkarte löscht, bei denen die zugehörigen Metadaten keine Auffälligkeiten aufweisen wie etwa außergewöhnliche Beschleunigungen. Somit wird eine volle Speicherkarte wieder nutzbar, ohne dass Videos von Unfällen vom Benutzer gelöscht werden können.

6 Formalisierung

6.1 Datenflussmodell

Zur Formalisierung von Datenflusseigenschaften eignet sich das formale Datenflussmodell von Harvan und Pretschner.²⁸ Dieses wird durch ein Tupel $(D, C, F, \Sigma, \sigma_i, P, A, R)$ beschrieben. D ist die Menge der *Daten*, die vom System überwacht werden. C ist die Menge der *Container* im System, bspw. Dateien, Systemprozesse oder Netzwerkverbindungen. $P \subseteq C$ ist die Menge der *Prozesse*, die Aktionen auslösen können. Prozesse sind Container, da sie Daten im Speicher oder in CPU-Registern vorhalten können. F ist die Menge der *Namen*, über die Container identifiziert werden können, bspw. Dateinamen $F_{fn} \subseteq F$ oder Dateideskriptoren $F_{dsc} \subseteq F$. Der aktuelle Zustand des Modells wird durch $\Sigma = (C \rightarrow 2^D) \times (C \rightarrow 2^C) \times (P \times F \rightarrow C)$ beschrieben und besteht aus drei Relationen. Die *Speicherrelation* $s : C \rightarrow 2^D$ bildet ab, welche Daten in welchen Containern enthalten sind. Die *Aliasrelation* $l : C \rightarrow 2^C$ enthält Container, die implizit aktualisiert werden, sobald Daten in einen anderen Container fließen. Die *Namensrelation* $f : P \times F \rightarrow C$ bildet ab, unter welchen prozessspezifischen Kennungen Container identifizierbar sind. $\sigma_i = (s_i, l_i, f_i) \in \Sigma$ beschreibt den initialen Zustand des Modells. Hier enthält die Speicherrelation die initialen Repräsentationen der überwachten Daten. Die Menge A enthält die im System durchführbaren Aktionen, die Änderungen der Speicher-, Alias-, oder Namensrelation indizieren. Diese Änderungen werden in der (deterministischen) Übergangsrelation $R \subseteq \Sigma \times P \times A \times \Sigma$ beschrieben. Aktualisierungen der Relationen s , l und f werden im Folgenden in der Notation von Harvan und Pretschner²⁷ dargestellt: Sei $m : S \rightarrow T$ eine Relation und $x \in X \subseteq S$ eine Variable. Dann ist $m[x \leftarrow expr]_{x \in X} = m'$ mit $m' : S \rightarrow T$ definiert als

$$m'(y) = \begin{cases} expr & \text{falls } y \in X \\ m(y) & \text{sonst.} \end{cases}$$

Harvan und Pretschner formalisieren explizite Datenflüsse innerhalb des Betriebssystems.²⁷ Modifikationen auf R werden daher durch System-Calls induziert. Da durch Anpassungen der Formalisierungen des *read*- und des *write*-System-Calls das Verhalten des De-

²⁸ Harvan und Pretschner, State-based Usage Control Enforcement with Data Flow Tracking Using System Call Interposition, NSS 2009.

klassifikationssystems spezifiziert werden kann, werden diese im Folgenden in ihrer ursprünglichen Form zitiert.

$$\forall s \in [C \rightarrow 2^D], \forall l \in [C \rightarrow 2^C], \forall f \in [P \times F \rightarrow C], \forall p \in P, \forall e \in F_{dsc} : \quad (1)$$

$$((s, l, f), p, read(e), (s[t \leftarrow s(t) \cup s(f(p, e))]_{t \in I^*(p)}, l, f)) \in R$$

$$\forall s \in [C \rightarrow 2^D], \forall l \in [C \rightarrow 2^C], \forall f \in [P \times F \rightarrow C], \forall p \in P, \forall e \in F_{dsc} : \quad (2)$$

$$((s, l, f), p, write(e), (s[t \leftarrow s(t) \cup s(p)]_{t \in I^*(f(p, e))}, l, f)) \in R$$

Gl. 1 beschreibt die Semantik eines *read*-System-Calls. Dieses Ereignis modifiziert die Speicherrelation s dergestalt, dass jedes Datum aus dem gelesenen Container (konservativ überapproximiert) in die reflexiv-transitive Hülle I^* des aufrufenden Prozess-Containers p fließt. Analog beschreibt Gl. 2 die Semantik von *write*-System-Calls.

Mittels dieser Formalisierungen können bereits Daten im System bewegt werden. Im Unterschied zu Harvan und Pretschner²⁷ wird im Folgenden eine weitere Interpretation der Datenmenge D verwendet. Diese umfasst nicht nur Daten, die durch Policies der Nutzungskontrolle geschützt werden, sondern alle verfügbaren Daten. Dies ist erforderlich, um zwischen klassifizierten Daten, die das System weder verlassen noch vom Betreiber eingesehen werden dürfen, und unklassifizierten Daten unterscheiden zu können. $D = D_{cl} \cup D_{decl}$ sei daher aus zwei disjunkten Mengen der klassifizierten bzw. deklassifizierten Daten zusammengesetzt. Klassifizierte Daten können mittels der nicht-injektiven Deklassifikationsfunktion $decl : D_{cl} \rightarrow D_{decl}$ in unklassifizierte Daten konvertiert werden. Zusätzlich sei $D_{prot} \subseteq D$ die Menge der unklassifizierten Daten, die der Betreiber lesen, aber nicht modifizieren oder löschen darf (Metadaten wie Zeitstempel oder GPS-Daten). Da die Integrität der klassifizierten Daten sichergestellt werden muss, gilt $D_{cl} \subseteq D_{prot}$.

Im nächsten Schritt sind nun diejenigen Container im System zu definieren, die klassifizierte Daten enthalten dürfen. Entsprechend repräsentieren $c_{storage} \in C$ und $c_{decl} \in P$ den Speicher der Kamera (Speichermedium) bzw. den Systemprozess der Deklassifikationskomponente. Die Nutzungskontrollkomponente wird nicht als separater Container betrachtet, da sie als überwachende Instanz agiert und selbst keine Daten verarbeitet. Damit kann nun spezifiziert werden, wie die Deklassifikationsfunktion von der Deklassifikationskomponente benutzt wird. Dazu wird die Formalisierung der *read*-Operation so erweitert, dass klassifizierte Daten in die Deklassifikationskomponente hineinfließen können, wo dann eine deklassifizierte Repräsentation der Daten erzeugt wird.

$$\forall s \in [C \rightarrow 2^D], \forall l \in [C \rightarrow 2^C], \forall f \in [P \times F \rightarrow C], \forall p \in P, \forall e \in F_{dsc}, f(p, e) = c_{decl} : \quad (3)$$

$$((s, l, f), p, read(e), (s[t \leftarrow s(t) \cup s_{decl}(f(p, e))]_{t \in I^*(p)}, l, f)) \in R$$

Gl. 3 drückt aus, dass für jede *read*-Operation auf dem Deklassifikationscontainter c_{decl} jeweils auch die deklassifizierten Repräsentationen der Daten zurückgegeben werden. Dies

wird mittels einer modifizierten Speicherrelation $s_{decl} : C \rightarrow 2^{D_{decl}}$ erreicht, die automatisch die entsprechenden Daten deklassifiziert. Sie ist definiert als:

$$s_{decl}(c) = \{d \mid d \in (s(c) \setminus D_{cl})\} \cup \{decl(d) \mid d \in (s(c) \cap D_{cl})\} \quad (4)$$

Analog wird die *write*-Operation erweitert (vgl. Gl. 5), sodass diese ebenfalls deklassifizierte Daten schreibt, falls sie aus dem Deklassifikationscontainer stammen.

$$\begin{aligned} \forall s \in [C \rightarrow 2^D], \forall l \in [C \rightarrow 2^C], \forall f \in [P \times F \rightarrow C], \forall e \in F_{dsc} : \\ ((s, l, f), c_{decl}, write(e), (s[t \leftarrow s(t) \cup s_{decl}(c_{decl})]_{t \in t^*(f(c_{decl}, e))}, l, f)) \in R \end{aligned} \quad (5)$$

Mittels dieser erweiterten Definitionen der *read*- bzw. *write*-Operation können Daten nun in einer deklassifizierten Form angefordert werden. Noch immer können aber klassifizierte Daten die Deklassifikationskomponente verlassen. Um dies zu verhindern, werden im Folgenden die für die Sicherheit des Systems erforderlichen System-Policies spezifiziert. Davor sei allerdings noch der initiale Zustand $\sigma_i = (s_i, l_i, f_i) \in \Sigma$ in Gl. 6 angegeben.

$$\begin{aligned} s_i(c_{storage}) \cap D_{cl} &\neq \emptyset \\ s_i(C \setminus \{c_{storage}\}) \cap D_{prot} &= \emptyset \end{aligned} \quad (6)$$

Somit wird davon ausgegangen, dass klassifizierte Daten initial nur auf der Speicherkarte vorhanden sind, während noch keine schützenswerten Daten im System existieren.

6.2 System-Policies

Basierend auf dem angegebenen formalen Systemmodell können formale Nutzungskontroll-Policies spezifiziert werden, um die Nutzungen von klassifizierten Daten im System einzuschränken. Wie von Harvan und Pretschner beschrieben,²⁷ sind solche Policies zustandsbasiert. Sie spezifizieren nicht, welche Ereignisse unterbunden bzw. zugelassen werden sollen, sondern vielmehr welche unzulässigen Systemzustände zu vermeiden sind.

Die wichtigste Policy im betrachteten System verfolgt das Schutzziel der Vertraulichkeit und verlangt, dass klassifizierte Daten nicht durch den Betreiber betrachtet oder verarbeitet werden dürfen. Daher dürfen im formalen Modell ausschließlich die Container $c_{storage}$ als Quelle und c_{decl} als Deklassifizierungskomponente klassifizierte Daten enthalten.

$$\forall c \in C \setminus \{c_{storage}, c_{decl}\} : s(c) \cap D_{cl} = \emptyset \quad (7)$$

Die in Gl. 7 dargestellte Policy ist hinreichend, um unerwünschte Datennutzungen auszuschließen. Weiterhin sollen keine Aliasbeziehungen für diese beiden Container existieren, da klassifizierte Daten ansonsten entlang der Aliasbeziehungen abfließen könnten. Auch wenn Gl. 7 bereits implizite Flüsse klassifizierter Daten in Alias-Container verhindert, kann mit der Policy in Gl. 8 zusätzlich bereits das Entstehen entsprechender Aliasbeziehungen unterbunden werden.

$$l(c_{storage}) \cup l(c_{decl}) = \emptyset \quad (8)$$

Obleich die in Gl. 7 dargestellte Policy bereits alle Vertraulichkeitsanforderungen erfüllt, ist es ungünstig, über die gesamte Menge C der Container quantifizieren zu müssen. Da für den initialen Zustand angenommen wurde, dass nur $c_{storage}$ klassifizierte Daten enthält, kann die Policy in Gl. 7 auch als Restriktion für Datenflüsse zwischen $c_{storage}$ und c_{decl} ausgedrückt werden. Die Policy in Gl. 9 besagt, dass Daten aus dem Container $c_{storage}$ nur in den Container c_{decl} fließen dürfen. Die in Gl. 10 und 11 dargestellten Policies verlangen, dass nur deklassifizierte Daten aus c_{decl} heraus fließen können.

$$\forall((s, l, f), p, read(e), (\bar{s}, l, f)) \in R : f(p, e) = c_{storage} \implies l^*(p) = \{c_{decl}\} \quad (9)$$

$$\forall((s, l, f), p, read(e), (\bar{s}, l, f)) \in R : f(p, e) = c_{decl} \implies \bar{s}[l^*(p)] \subseteq D_{decl} \quad (10)$$

$$\forall((s, l, f), p, write(e), (\bar{s}, l, f)) \in R : p = c_{decl} \implies \bar{s}[l^*(f(p, e))] \subseteq D_{decl} \quad (11)$$

Im Unterschied zur Alias- und Namensrelation wird die Speicherrelation bei den betrachteten *read*- und *write*-Operationen zu \bar{s} modifiziert. Wenn wir o.B.d.A. annehmen, dass $p \notin \{c_{storage}, c_{decl}\}$ in der Policy in Gl. 10 und 11 gilt, so wird ersichtlich, dass ein System, das Gl. 9, 10 und 11 erfüllt, auch Gl. 7 erfüllt. Die Policies sind allerdings nicht äquivalent, da durch die Policies in Gl. 9, 10 und 11 ein Zugriff auf unklassifizierte Daten, die initial in $c_{storage}$ vorliegen, ebenfalls unterbunden wird. Wie im Folgenden gezeigt wird, eignet sich diese an Datenflüssen orientierte Spezifikation besser für die technische Umsetzung als die Policy in Gl. 7. Weiterhin kann auf die Nutzung von l^* in der Policy in Gl. 9 verzichtet werden, wenn gleichzeitig Gl. 8 gefordert wird.

Schließlich ist noch das Schutzziel der Integrität gegenüber einem Modifying-Attacker oder einem Destructive-Attacker (vgl. Abschnitt 4) zu behandeln. Diese sollen geschützte Daten weder verändern noch löschen können. Hierfür muss zu jedem Zeitpunkt gelten:

$$s_i(c_{storage}) \cap D_{prot} = s(c_{storage}) \cap D_{prot} \quad (12)$$

Die Policy in Gl. 12 verlangt, dass geschützte Daten, die initial auf dem Speichermedium vorliegen, für alle Zeit unverändert bleiben. Sie verhindert außerdem, dass neue geschützte Daten erzeugt werden, bspw. gefälschte GPS-Positionsdaten. Wegen $D_{cl} \subseteq D_{prot}$ stellt die Policy in Gl. 12 insbesondere sicher, dass auch klassifizierte Daten weder verändert noch vom ursprünglichen Speichermedium gelöscht werden können.

Der Akteur Strafverfolgungsbehörde taucht in der angegebenen Formalisierung des Systems nicht auf. Die zustandsbasierte Sicht auf das System eignet sich nur bedingt, um Benutzerautorisierungen auszudrücken. Es wird daher angenommen, dass eine Infrastruktur zur Verwaltung von Benutzerrechten auf dem formalisierten System aufsetzt und entscheiden kann, ob die angegebenen Policies durchzusetzen sind oder nicht.

7 Sicherheitsanalyse

Ausgehend von der Formalisierung eines Systems, das die Anwendung von Deklassifikationsfunktionen sicherstellt, können nun dessen Sicherheitseigenschaften in Bezug auf das Angreifermodell analysiert werden.

In Bezug auf den Datenschutz stellt sich zunächst die Frage, ob das System robust gegen einen Privacy-Attacker ist, d.h. ob das Schutzziel der Vertraulichkeit erfüllt ist. Gemäß der Policies in Gl. 9, 10 und 11 können klassifizierte Daten nur im Speicher der Kamera, sowie in der Deklassifikationskomponente existieren. Der Betreiber kann Daten nur einsehen, falls diese zuvor durch eine Deklassifikationskomponente anonymisiert wurden. Aufgrund dieser Eigenschaften ist sichergestellt, dass klassifizierte Daten niemals in un-anonymisierter Form das System verlassen. Dem Privacy-Attacker, und damit auch dem Betreiber der Dashcam, ist es nicht möglich klassifizierte Daten zu extrahieren, womit das Schutzziel der Vertraulichkeit erfüllt ist.

Für das Schutzziel der Integrität sind vor allem der Modifying-Attacker und der Destructive-Attacker relevant. Gemäß der Policy in Gl. 12 ist es nicht möglich, Daten auf dem Speicher der Kamera zu modifizieren oder zu löschen. Das System ist somit robust gegen Modifying-Attacker und ebenso gegen Destructive-Attacker, wobei hier vom mechanischen Zerstören der Speicherkarte abgesehen werden muss.

Damit die Analyse der Angreifer valide ist, müssen einige Annahmen an die Betriebsumgebung des implementierten Systems gemacht werden. Zunächst muss vorausgesetzt werden, dass das verwendete Betriebssystem die Integrität der einzelnen Komponenten sichert. Dies bedeutet insbesondere, dass der Betreiber der Dashcam auf dem Anonymisierungssystem keine Administratorrechte besitzt und der Zugriff auf die Programm- und Konfigurationsdateien der Systemkomponenten eingeschränkt ist. Auch das gemeinsame Geheimnis, welches auf dem Anonymisierungssystem gespeichert ist, muss vor dem Auslesen durch den Betreiber geschützt werden, bspw. durch Einsatz der Windows Data Protection API. Neben den Softwarekomponenten muss auch das physische System gesichert werden. Der Betreiber der Dashcam darf keinen direkten Zugriff auf die Festplatte des Anonymisierungssystems erhalten, ansonsten könnte er sich selbst Administratorrechte verschaffen. Diese Art des Umgehens von Sicherheitsmechanismen kann besonders wirksam durch den Einsatz eines sog. Trusted-Platform-Modules (TPM) verhindert werden.

8 Fazit

In diesem Beitrag wurde ein System entwickelt, das die Vertraulichkeit und die Integrität datenschutzrelevanter Bildbereiche gewährleistet. Mechanismen der Datenflusskontrolle lassen klassifizierte Daten von einem sicheren Speichermedium ausschließlich in die Deklassifikationskomponente fließen. Dort werden sie vor einem Zugriff durch den Betreiber anonymisiert, sodass dieser zu keinem Zeitpunkt die Möglichkeit hat, die Klavideodaten einzusehen. Durch die Auslagerung der Anonymisierung auf einen separaten Rechner können leistungsfähige Bildauswertelgorithmen eingesetzt werden, sodass (i) personenbezogene Daten nach dem Stand der Technik minimiert werden, und (ii) jede handelsübliche Dashcam um Datenschutzmechanismen erweitert werden kann. Ein datenschutzgerechter Betrieb von Dashcams erscheint somit unter den in Abschnitt 7 genannten Annahmen möglich.

Datenschutz im Fahrzeug der Zukunft: Vernetzt, Autonom, Elektrisch

Meiko Jensen¹, Nils Gruschka² und Jens Lüssem³

Abstract: Das Fahrzeug der Zukunft wird durch die Trends Vernetzung, autonomes Fahren und Elektromobilität bestimmt. Damit verbunden sind große Hoffnungen für die Aspekte Reisekomfort, Sicherheit im Straßenverkehr und Umweltfreundlichkeit.

Allerdings werden diese Fortschritte auch z. T. durch erhebliche Eingriffe in die Privatsphäre und in das Recht auf informationelle Selbstbestimmung erkauft. In diesem Beitrag beleuchten wir, welche Datenschutzprobleme sich durch technische Neuerungen im Fahrzeug der Zukunft ergeben können, wie diese rechtlich einzuschätzen sind, und welche Lösungswege für die zugrundeliegenden Fragen denkbar sind.

1 Überblick

Bereits heute bauen die großen Automobilhersteller massiv Vernetzungstechnologien in ihre aktuellen Modelle ein. Vom klassischen Automobil über Busse und Lastkraftwagen bis hin zu Einsatzfahrzeugen von Rettungsdienst, Feuerwehr und Polizei wird überall digitale Technologie eingesetzt, um das Fahren in Zukunft zu erleichtern. Dieser Trend der massiven Digitalisierung und Vernetzung wird in den nächsten Jahren noch deutlich verstärkt werden und neben den offensichtlichen Vorteilen zeichnen sich auch schon Probleme dieser Entwicklung ab. Bisher wird dabei primär die Sicherheit von Leben und Besitz betrachtet. Berichte über geknackte [ADA16] oder ferngesteuerte Autos [Gre] erlangen dabei immer wieder große Aufmerksamkeit und Experten gehen davon aus, dass diese Bedrohungen in Zukunft noch deutlich gravierender werden [BFN].

Wenig Beachtung haben bisher hingegen die Nebeneffekte in Bezug auf Datenschutz und informationeller Selbstbestimmung erfahren. So weiß kaum ein Autokäufer heutzutage, ob, wann, welche und wie viele Daten sein Fahrzeug an wen übermittelt. Gleichzeitig sind viele der gesammelten Daten in modernen Fahrzeugen teilweise äußerst sensibel, erlauben sie doch brisante Rückschlüsse auf Fahrverhalten, Bewegungsmuster, Angewohnheiten und sogar medizinisch relevante Umstände von Fahrern und Insassen solcher Automobile.

In diesem Beitrag geben wir einen Überblick über die bekannten und zukünftig zu erwartenden Daten, die im Fahrzeug gesammelt, ausgewertet und gegebenenfalls an Dritte

¹ FH Kiel, meiko.jensen@fh-kiel.de. Die Arbeit von M.J. wurde teilweise am Unabhängigen Landeszentrum für Datenschutz Schleswig-Holstein durchgeführt und dort im Rahmen des Projektes iKoPA durch das Bundesministerium für Bildung und Forschung (BMBF) gefördert.

² FH Kiel, nils.gruschka@fh-kiel.de

³ FH Kiel, jens.luessem@fh-kiel.de

übertragen werden, und beleuchten anschließend die in diesem Zusammenhang auftretenden Probleme hinsichtlich des Datenschutzes anhand von zwei ausgewählten Beispielen. Diesen folgt eine kurze datenschutzrechtliche Perspektive, und abschließend zeigen wir Wege auf, wie sich die Datenhaltung im Fahrzeug der Zukunft datenschutzfreundlicher gestalten ließe.

2 Stand der Technik

In der Diskussion der Zukunft des Automobils werden fast schon gebetsmühlenartig drei große Paradigmen wieder und wieder diskutiert. Zunächst ist als sicher anzunehmen, dass das Automobil der Zukunft *vernetzt* sein wird. Dies impliziert eine permanente, mobile Anbindung des Fahrzeugs an das Internet, über die eine bidirektionale Kommunikation zwischen Fahrzeug und entsprechenden Internetdiensten realisiert wird. Schon heute übermitteln moderne Fahrzeuge ihre Sensordaten per Mobilschnittstelle an ihre Hersteller, um etwa den Wartungs- und Inspektions-Prozess in Vertragswerkstätten zu optimieren.

Ein zweites großes Thema in Hinblick auf das Fahrzeug der Zukunft besteht in seiner Fähigkeit, zumindest teilweise *autonom fahren* zu können. Erste Ergebnisse lassen hier einen großen technischen Fortschritt erkennen (vgl. [Gui11, Gom16]), der mit großer Wahrscheinlichkeit in naher Zukunft auch in Deutschland umgesetzt werden dürfte. Zwar ist es noch ein weiter Weg bis die Technik das vollständig autonom fahrende Automobil zuverlässig implementiert hat, teilautonome Fahrassistenzsysteme – etwa beim Einparken oder zur Spurführung auf Autobahnen – sind jedoch bereits heute Teil modern ausgestatteter Fahrzeugmodelle.

Schließlich zeichnet sich ein deutlicher Trend weg von Verbrennungsmotoren hin zur Elektromobilität ab, bei dem Fahrzeuge *elektrisch* angetrieben werden. Diese Entwicklung geht jedoch mit veränderten Anforderungen an Verkehrsinfrastruktur und Fahrverhalten einher, da sich Fahrzeugbatterien nicht wie Verbrennungsmotoren binnen kurzer Zeit an beliebigen Tankstellen neu befüllen lassen. Auch die aktuellen Reichweitenbegrenzungen von Elektrofahrzeugen, vorgegeben durch die Maximalkapazität der Bordbatterie, dürften in naher Zukunft massiven Einfluss auf die Art und Weise haben, wie Elektrofahrzeuge in Deutschland genutzt werden.

Alle diese drei großen Neuerungen haben ihre jeweils eigenen, speziellen Anforderungen an Automobil, Fahrer, Hersteller und Verkehrsinfrastruktur, aber allen gemeinsam ist zu eigen, dass sie ihre Vorzüge nur dann ausspielen können, wenn die hierfür notwendigen technischen und infrastrukturellen Neuerungen geeignet realisiert werden. Ein Kernmerkmal ist hierbei die Nutzung gesammelter Sensordaten aus dem Fahrzeug zur Optimierung der jeweils benötigten Infrastrukturen. So steht beispielsweise zu erwarten, dass Navigationssysteme direkt mit Batteriesensoren und internetbasierten Batterie-Lade-Diensten gekoppelt werden, um die optimale Kombination von Fahrt, Batterieladung und ggfs. Batterietausch zu errechnen. Hierfür werden auch das Fahrverhalten des Fahrers, das Fahrzeuggewicht, der Grad der Abnutzung von Batterie und anderen Fahrzeugteilen, sowie weitere Sensordaten mit einbezogen.

3 Daten im Fahrzeug der Zukunft

Bereits heute fallen in gängigen Fahrzeugmodellen umfassende Messdaten verschiedenster Sensoren an, die oft fortlaufend erhoben, an entsprechende Steuergeräte im Fahrzeug übermittelt, dort aggregiert und ausgewertet und zur Regelung bzw. Steuerung der einzelnen Fahrzeugsysteme verwendet werden. Eine umfassende Auflistung verschiedener gängiger Sensoriksysteme in Fahrzeugen findet sich etwa bei Konrad Reif [Rei11]; eine (unvollständige) Übersicht relevanter Sensoren wird im Folgenden gegeben:

Sensorik zu Antrieb, Motorik und Fahrzeugzustand

- Raddrehzahlsensoren
- Beschleunigungssensoren
- Sensorik im Antrieb (Drucksensoren, Klopfensensoren, Drehzahlsensoren etc.)
- Lambdasonde (zur Steuerung der Kraftstoff-Luft-Mischung)
- Reifendrucksensoren
- Tankstands- und Tankdrucksensoren, bzw. Batterie-Ladestandssensorik
- Scheinwerfer-Neigungssensoren
- Fahrzeugneigungssensoren
- Überrollsensoren

Sensorik zur Fahrzeugsteuerung

- Lenkradwinkelsensoren
- Lenkrad-Drehmomentsensoren (Servolenkung)
- Pedalwegsensorik

Sensorik zur Überwachung von Innenraum und direkter Fahrzeugumgebung

- Sitzbelegungssensoren
- Sensorik der Zentralverriegelung und der Türschlösser
- Luftgütesensoren
- Feuchte- und Temperatursensoren
- Regensensor
- Abstandssensoren (Ultraschall, Radar, Lidar etc.)
- Kameras (Rückfahrkamera, Dashcam etc.)

Sensorik zur Navigation

- GPS-Sensoren
- Radio Data System (RDS-Sensor)
- Car2X-Sensoren (z.B. IEEE802.11p)
- Mobilfunkkommunikation

Wie sich unschwer erkennen lässt, erfasst die Sensorik im Fahrzeug der Zukunft umfassende Messdaten über das Fahrzeug und seine Umgebung. Insbesondere werden dabei vielfach auch personenbezogene Daten gesammelt. So hängen Daten zu Antrieb, Motorik und Fahrzeugzustand direkt vom Fahrer und seinem Fahrverhalten ab, ebenso wie die Sensorik zur Fahrzeugsteuerung. Die Navigationssensorik verarbeitet Daten über die Bewegungsmuster des Fahrzeugs – und damit die seiner Insassen. Schließlich verarbeiten vielfältige Überwachungssensoren sowohl personenbezogene Daten zu Fahrzeuginsassen als auch zu anderen Verkehrsteilnehmern im Umfeld des Fahrzeugs. Somit besteht ein klares datenschutztechnisches Interesse daran, die Art der jeweiligen Messdaten bezüglich ihres Verwendungszweckes und ihres Personenbezuges näher zu analysieren.

4 Datenschutzprobleme im Fahrzeug der Zukunft

Ein kurzer Blick auf die Liste der Sensordaten, die in modernen Fahrzeugen verarbeitet werden, offenbart bereits eine Fülle potentiell datenschutzrechtlich bedenklicher Datenverarbeitungsprozesse. So sind etwa GPS-Position und Car2X-Sensoren offensichtlich zur Erstellung von Bewegungsprofilen des Fahrzeugs – und damit des Fahrers – nutzbar, und folglich datenschutztechnisch hochgradig sensibel. Auch die Verbrauchsdaten zu Tankfüllung bzw. Batterieladestand bergen das Potential, gefahrene Strecken über die Änderungen des Füll- bzw. Ladestands zu erkennen.

Darüber hinaus gibt es in modernen Fahrzeugen aber auch Sensorik, deren Relevanz in Hinblick auf den Datenschutz nicht sofort offensichtlich wird. Im Folgenden werden hierfür zwei Beispiele gegeben, die aufzeigen, wie sich aus vermeintlich harmlosen Sensoren hochgradig sensitive Informationen ablesen lassen. Daran anschließend wird die aktuelle rechtliche Lage rund um Datenschutz in Fahrzeugen skizziert.

4.1 Erstes Szenario: Reifendruck

Neben der potentiellen Überwachung der Position und Bewegungsmuster eines Fahrzeugs – und damit einhergehend seines Fahrers bzw. der Mitfahrer – gibt es noch andere Arten personenbezogener Informationen, die sich mehr oder weniger direkt aus den Sensordaten des Fahrzeugs der Zukunft ableiten lassen. In diesem Szenario betrachten wir eine zunächst relativ harmlos wirkende Datenquelle: den Reifendrucksensor.

In modernen Fahrzeugen muss jeder Reifen gemäß EU-Verordnung 661/2009 Art. 9 Abs. 2 [EU609] mit einem entsprechenden Reifendrucksensor ausgestattet sein. Technischer Hintergrund hier ist die frühzeitige Erkennung von Über- oder Unterdruck während der Fahrt. Dadurch können beispielsweise Unfälle aufgrund akuten Druckverlustes in voller Fahrt vermieden werden, wenn der Fahrer sofort eine Warnmeldung erhält und zeitgleich das Lenksystem des Fahrzeugs auf den reduzierten Luftdruck in einem der Fahrzeugreifen reagieren kann. Gleichermaßen können dem Fahrer Warnmeldungen angezeigt werden, wenn der Luftdruck in den Reifen insgesamt zu hoch oder zu niedrig wird, oder wenn signifikante Druckunterschiede zwischen den verschiedenen Reifen des Fahrzeugs bestehen. Auch hängt der Verbrauch eines Fahrzeugs stark vom Reifendruck ab, so dass ein korrekter Reifendruck im Fahrbetrieb hilft, Kraftstoff bzw. Batteriekapazität zu sparen.

Aufgrund der Natur eines Reifendruckensors wird die Datenübermittlung an das Fahrzeug über eine kabellose Kommunikationsschnittstelle realisiert. Gängige Verfahren nutzen hier Techniken auf Basis von Funkwellen (z.B. Bluetooth), um die Reifendruckdaten des (rotierenden) Reifens an einen entsprechenden Empfänger am Fahrzeuggestell zu übermitteln. Moderne Reifendrucksensoren beziehen zudem bei ihren Messungen auch Temperaturunterschiede der Außenwelt über einen eingebauten Temperatursensor direkt mit ein. Ein weiterer, optionaler Bewegungssensor kann zudem ermitteln, ob ein Reifen sich gerade in Bewegung befindet oder nicht. Dies dient zum einen der präziseren Kalkulation des Reifendrucks im Stand bzw. in Fahrt, zum anderen kann der Reifendrucksensor aus Energieeffizienzgründen auf eine reduzierte Abtastrate schalten, wenn das Fahrzeug steht. Hierdurch verlängert sich die Funktionszeit der (häufig batteriebetriebenen) Reifendrucksensoren erheblich.

Die Historie der Reifendrucksensordaten wird üblicherweise in den Steuereinheiten des Fahrzeugs gespeichert, und lässt sich dort relativ leicht auslesen (vgl. etwa [Spa16]). Spätestens bei einer Übermittlung der Reifendrucksensordaten an Hersteller oder Werkstatt ist davon auszugehen, dass externe Organisationen relativ leicht direkten Zugriff auf die Reifendruckhistorie eines Fahrzeugs erlangen können.

4.1.1 Problemstellungen

Ein naheliegender Schwachpunkt in der Reifendruck-Messarchitektur besteht in der Verwendung einer drahtlosen Schnittstelle zur Datenübertragung vom Reifen an das Fahrzeug. Wie von Rouf et al. [RMM⁺10] nachgewiesen wurde, lässt sich eine solche Datenkommunikation im Fahrzeugbetrieb (d.h. auch während der Fahrt) erfolgreich angreifen, um etwa einen platten Reifen vorzugaukeln. Entsprechende Folgen eines solchen Angriffs reichen von einer akuten Warnmeldung an den Fahrer bis hin zu automatischen Ausweich- und Bremsmanövern, die jedoch von falschen Reifendruckwerten ausgehen und daher inkorrekte Fahrmanöver ausführen. Je größer hierbei die Autonomie des Fahrzeugs in Bezug auf sein Fahrverhalten ist, desto gravierender können die Folgen eines solchen Angriffs sein.

Ein anderer, datenschutzrechtlich wesentlich relevanterer Aspekt besteht in der Auswertung der Reifendruckhistorie. Abbildung 1 zeigt ein (konstruiertes) Beispiel für einen möglichen Druckverlauf in den Reifendrucksensoren eines Fahrzeugs über Zeit. Es ist deutlich zu erkennen, dass sich der Reifendruck mit der Zeit signifikant verändert. Stets folgt auf eine akute Erhöhung des Reifendrucks (Druckspitze) eine lange Phase geringfügig, aber signifikant höheren Luftdrucks. Dieser Unterschied lässt sich wie folgt interpretieren. Geht man von einem gewöhnlichen Fahrzeug mit vier Rädern aus, ergibt sich der tatsächliche Reifendruck hauptsächlich aus zwei Faktoren: der Luftmenge im Reifen selbst und dem Gewicht des Fahrzeugs, das darauf lastet. Bei einer angenommenen Gleichverteilung des Gewichtes auf vier Räder trägt folglich jedes Rad ein Viertel des Gesamtgewichtes des Fahrzeugs. Erhöht sich nun schlagartig der Reifendruck aller vier Reifen, so wurde entweder allen vier Reifen gleichzeitig mehr Luft zugeführt, oder das Gewicht des Fahrzeugs hat sich erhöht. Ein solches Muster ist also typischerweise stets dann zu beobachten, wenn das Fahrzeug beladen wird, bzw. wenn Personen in das Fahrzeug einsteigen – wobei ihr Körpergewicht das Gesamtgewicht des Fahrzeugs erhöht. Die Druckspitzen entstehen durch den akuten Beschleunigungsfaktor, der entsteht, wenn sich eine Person auf einen Fahrzeugsitz setzt. Anzumerken ist, dass die Messung von Druckspitzen nur in Abhängigkeit von der Abtastrate der Drucksensoren erfolgt, daher wird nicht jede Druckspitze auch aufgezeichnet. Der Übergang von einem Drucklevel zum nächsten ist aber stets ablesbar. Analog lassen sich entsprechende Reduktionen des Reifendrucks auf das Aussteigen von Personen zurückführen.

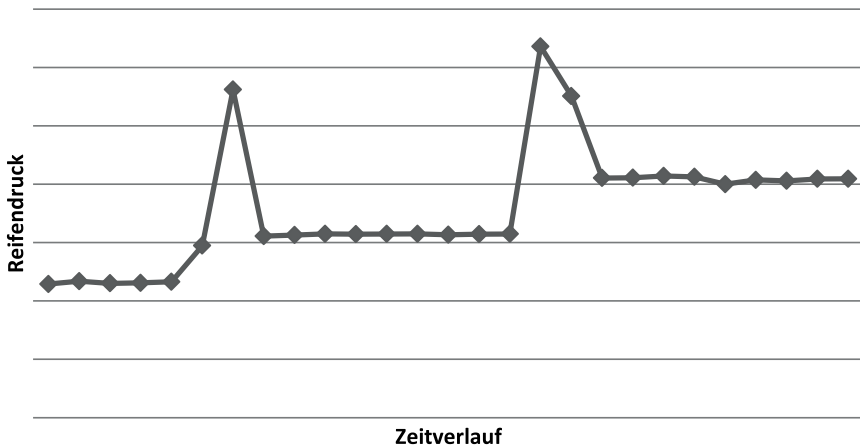


Abbildung 1: Beispiel eines Reifendruckverlaufs beim Einsteigen

Wie sich unschwer erkennen lässt, gibt die Reifendruckhistorie direkten Aufschluss darüber, zu welchen Zeitpunkten Personen das Fahrzeug bestiegen oder verlassen haben. Auch die Anzahl der Personen lässt sich leicht ableiten, entweder anhand der Zahl der Druckspitzen oder aus der Anzahl verschiedener Drucklevel im Verlauf. Rechnet man nun noch die Skaleneinheiten der Drucksensoren hinzu, lässt sich sogar das Gewicht der einzelnen Personen errechnen, indem die Differenz zwischen dem Druck vor Besteigen des Fahrzeugs und dem Druck nach Besteigen des Fahrzeugs gebildet wird. Sogar die Sitzplatzverteilung der Personen im Fahrzeug kann theoretisch ermittelt werden, wenn man annimmt, dass die Druckspitzen üblicherweise an der Fahrzeugecke am höchsten ausfallen, an der die Person

beim Einsteigen die meiste beschleunigte Masseeinwirkung ausgeübt hat. Folglich lassen sich aus den Reifendruckdaten mit relativ geringem Aufwand Profile erstellen, welche Person(en), identifiziert durch ihr Gewicht, zu welchen Zeitpunkten das Fahrzeug bestiegen oder verlassen haben.

Je nach Präzision der Gewichtsbestimmung der einzelnen Be- und Entladungen im Fahrzeug lassen sich sogar noch weitere Informationen ableiten, unter anderem zu folgenden Aspekten:

- Dauer der Fahrten (über Zeiträume zwischen Ein- und Ausstieg),
- Essgewohnheiten der wiederkehrenden Fahrzeuginsassen (über Gewichtsveränderungen),
- Schwangerschaften (über 9 Monate langsame Gewichtszunahme, gefolgt von akuter Gewichtsabnahme)
- Wachstumsverlauf und Alter von Kindern (über Gewichtsverlauf, Wechsel auf Vordersitz/Fahrersitz),
- Fahrverhalten (über Differenzen des Reifendrucks verschiedener Räder während der Fahrt, etwa bei Kurven und Brems- bzw. Beschleunigungsvorgängen), sowie
- Fahrziele (über Fahrdauer insgesamt und zeitliche Abstände von Kurven und Brems-/Beschleunigungsmanövern).

Wie unschwer zu erkennen ist, geben Reifendrucksensordaten eine Fülle zusätzlicher Informationen preis, wenn sie längerfristig aufgezeichnet werden. Einige dieser Informationen sind dabei sogar besonders sensibler Natur (siehe Schwangerschaften), sodass hier ein hoher Schutzbedarf für derartige Daten abgeleitet werden muss. Die Übertragung dieser Daten im Rohzustand an die Steuerungseinheiten des Fahrzeugs oder gar an externe Parteien (über Mobilschnittstelle oder Auslesen in der Werkstatt) ist somit als hochgradig bedenklich einzustufen.

Es bleibt anzumerken, dass viele der genannten Informationen sich in ähnlicher Form auch aus anderen Sensordaten im Fahrzeug der Zukunft ermitteln lassen. So sind Autositze bereits heute mit Gewichtssensoren ausgestattet, um ggfs. zu warnen, wenn ein Insasse des Fahrzeugs nicht angeschnallt ist. Auch die Auszugweite eines Anschnallgurtes oder die Einstellungen von Sitz- und Lehnenpositionen erlauben Rückschlüsse auf die Insassen.

4.1.2 Lösungswege

Für die genannten Einsatzzwecke des Reifendruckensors wäre es technisch hinreichend, die Einhaltung der relevanten Grenzen für den Reifendruck direkt im Sensor abzuprüfen, und nur bei Überschreiten einer solchen Grenze eine Kommunikation nach außen (d.h. zum Fahrzeug, zum Fahrer oder zu anderen Sensoren bzw. Steuergeräten) aufzubauen. Bei einer solchen Architektur würden die Informationen, die den einzelnen Reifendrucksensor

verlassen, kaum noch in realistischem Maße für die oben beschriebenen datenschutztechnisch bedenklichen Auswertungen nutzbar sein. Die relevanten Nutzinformationen (Reifendruck zu hoch oder zu niedrig) blieben erhalten, das Signal selbst würde jedoch verborgen bleiben. Entsprechend würde dem Datenschutzprinzip der Datensparsamkeit durch eine solche Architektur am ehesten entsprochen werden, ohne die legitimen Anwendungszwecke für Reifendrucksensoren zu verbauen.

Im Zuge der Vollvernetzung des Fahrzeugs findet sich heutzutage allerdings die gegenteilige Entwicklung: Reifendruckdaten werden im Original an die Steuereinheiten des Fahrzeugs übermittelt, dort aufgezeichnet und ausgewertet, und ggfs. an Hersteller, Werkstätten und andere Organisationen weitergegeben. Die Annahme hierbei ist, dass sich Reifendruckdaten auch für andere nützliche Anwendungszwecke einbeziehen lassen (ABS-Steuerung, Klimaanlagesteuerung, Überladungswarnungen etc.). Hier muss folglich eine umfassende Abwägung zwischen den verschiedenen legitimen und illegitimen Anwendungszwecken der Reifendruckdaten erfolgen. Zum gegenwärtigen Zeitpunkt ist hier aber nicht von einer datenschutzseitig vorteilhaften Ausgestaltung der Informationsflüsse im Fahrzeug der Zukunft auszugehen.

4.2 Zweites Szenario: Bewegungsprofile anderer Verkehrsteilnehmer

Damit sich Fahrzeuge autonom im Straßenverkehr bewegen können, müssen fortlaufend Informationen über die jeweilige Umgebung und damit auch über die eigene Position gesammelt werden. Diese Informationen über die Umgebung werden in der Regel in sogenannten SLAM-Algorithmen (Simultaneous Localization and Mapping) verarbeitet, um eine genaue Umgebungskarte zu generieren und gleichzeitig eine möglichst exakte Bestimmung der eigenen Lokation vorzunehmen.

Abbildung 2 zeigt die Einbindung der Kartographie in die Gesamtarchitektur eines autonomen Fahrzeugs. Hindernisse werden dabei mit Hilfe von Sensoren (Ultraschall, Radar, Laser, ...) detektiert [TMD⁺06]. Die Existenz oder die Nicht-Existenz von Hindernissen kann allerdings nur mit einer gewissen Wahrscheinlichkeit angegeben werden. Dies liegt beispielsweise an Artefakten wie Schattenwurf oder auch an Fehlfunktionen von Sensoren. Um diese Wahrscheinlichkeiten zu erhöhen, werden fortlaufend Umgebungskarten erstellt, mit bereits generierten Karten abgeglichen und auf diese Weise aktualisiert wie auch detailliert. De facto gelingt dies durch Verwendung bedingter Wahrscheinlichkeiten. So wird die Wahrscheinlichkeit der Existenz eines Hindernisses erhöht, wenn die nachfolgende Auswertung der Sensoren wiederum ergibt, dass sich an der bezeichneten Stelle ein Hindernis befindet.

Das oben beschriebene Vorgehen funktioniert zunächst nur für unbewegliche Hindernisse, kann aber auf bewegliche Objekte ausgeweitet werden. Hierzu müssen weitere Informationen wie beispielsweise Bewegungsrichtung und Objektgeschwindigkeit erfasst werden. Dennoch ist eine Detektion beweglicher Objekte mit größeren Unsicherheiten verbunden. Eine mangelhafte bzw. zu späte Detektion beweglicher Objekte (z.B. Ball, Mensch) kann problematisch werden, da hier – in Abhängigkeit von der Geschwindigkeit der bewegli-

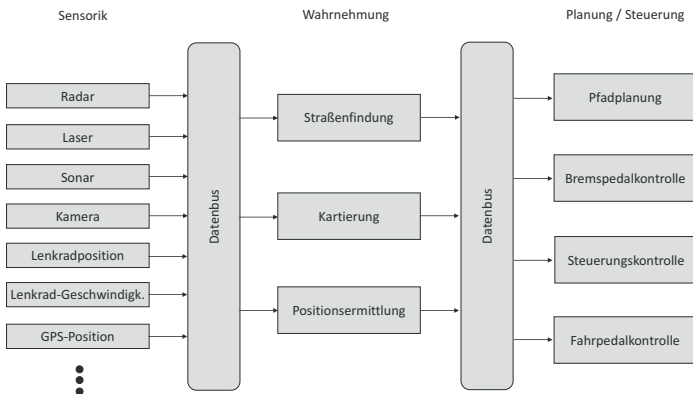


Abbildung 2: IT-Architektur eines autonomen Fahrzeugs (Ausschnitt, vereinfacht)

chen Objekte – ggf. zu hohe Anforderungen an die Reaktionszeiten von autonomen Fahrzeugen gestellt werden könnten.

In Abbildung 3 ist eine derartige Situation dargestellt. Die unterschiedlichen Grautöne geben das Maß an Sicherheit wider, mit der ein autonomes Fahrzeug ein Hindernis erkannt hat (hellgrau: Hindernis mit noch recht geringer Gewissheit erkannt, schwarz: Hindernis mit sehr hoher Gewissheit erkannt). In dieser Darstellung ist ebenfalls angedeutet, dass eine sichere Detektion beweglicher Objekte in der Regel mehr Sensordaten erfordert und damit zeitaufwändiger ist.

Um die Geschwindigkeit der oben skizzierten Kartierung zu erhöhen, könnten bereits existierende – aktuelle – Karten genutzt werden. Als Quellen derartiger Umgebungskarten kommen stationäre Objekte (z.B. Straßenlaternen) wie auch bewegliche Objekte (z.B. andere autonome Fahrzeuge) in Frage. So trägt beispielsweise eine Vernetzung von autonomen Fahrzeugen mit dem Ziel der Weitergabe bereits erstellter Karten dazu bei, dass sich die Güte der von den autonomen Fahrzeugen verwendeten Karten erhöht und damit auch die Grundlage für Entscheidungen verbessert.

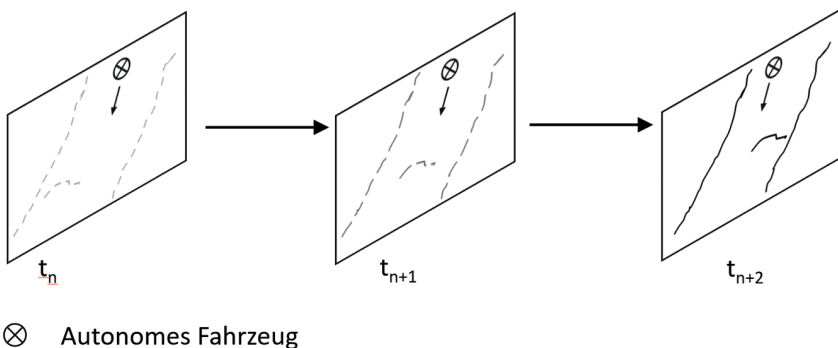


Abbildung 3: Kartierung mittels SLAM-Algorithmen (Prinzipskizze)

4.2.1 Problemstellungen

Die sich aus diesem Szenario ergebenden datenschutzrelevanten Problemstellungen sind mannigfaltig. Eine detaillierte Beschreibung sprengt den gesetzten Rahmen. An dieser Stelle soll daher nur auf die Möglichkeit der Generierung von Bewegungsprofilen anderer Verkehrsteilnehmer eingegangen werden.

Es werden in diesem Szenario Daten zwischen autonomen Fahrzeugen drahtlos übermittelt. Diese Daten sind keine Sensor-Rohdaten, sondern – in Form von Karten – bereits vorverarbeitete Daten. Die übermittelten Karten enthalten neben Informationen zu stationären Hindernissen oder Straßen auch Informationen über bewegliche Objekte, sofern diese eine Relevanz für die Exploration der Umgebung eines autonomen Fahrzeugs besitzen. Aus diesen – ggf. von mehreren autonomen Fahrzeugen – übermittelten Karten lassen sich zumindest über begrenzte Zeiträume Bewegungsprofile von Fußgängern erstellen, aus denen sich gegebenenfalls sogar die Identität dieser Verkehrsteilnehmer ableiten lässt.

Nicht nur autonome Fahrzeuge können solche Karten generieren. Auch stationäre Objekte wie z.B. mit entsprechenden Sensoren ausgerüsteten Straßenlaternen sind in der Lage eine Kartierung ihrer jeweiligen Umgebung vorzunehmen, um diese Informationen an autonome Fahrzeuge drahtlos zu übermitteln. Hier ist die Gefahr der Erstellung von Bewegungsprofilen anderer Verkehrsteilnehmer (z.B. Fußgänger) ungleich größer.

4.2.2 Lösungsszenarien

Da das Erkennen von Hindernissen (inkl. Fußgänger) und das Erstellen von (möglichst genauen) Karten ein integraler Bestandteil des autonomen Fahrens ist, lässt sich die dabei automatisch abfallende Erstellung von Bewegungsprofilen nicht einfach verhindern. Eine Möglichkeit zur Minderung dieser Gefahr wäre eine rein lokal Generierung von Karten ohne die Weitergabe an andere Teilnehmer. Allerdings verbessert die Verwendung von Karten anderer Provenienz die Genauigkeit der von einem autonomen Fahrzeug verwendeten Kartierung seiner Umgebung und erhöht damit auch die Sicherheit aller Verkehrsteilnehmer.

So verbleibt die Frage, ob insbesondere zu der Ausrüstung stationärer Objekte mit Sensoren Alternativen existieren, mit denen es nicht so einfach möglich ist, Bewegungsprofile von Verkehrsteilnehmern aufzuzeichnen.

4.3 Rechtliche Lage

Unabhängig von den konkreten einzelnen Datentypen lässt sich festhalten, dass das Fahrzeug der Zukunft umfassende personenbeziehbare und teils hochsensible Daten und Informationen verarbeitet. Entsprechend besteht nach BDSG §4 Abs. 1 (Verbot mit Erlaubnisvorbehalt) die Notwendigkeit einer geeigneten Rechtsgrundlage für die Erhebung und Verarbeitung dieser Daten. Zwar existieren hier teilweise entsprechende Regularien

(vgl. [EU609]), diese sind allerdings stets zweckgebunden und daher nicht ohne Weiteres auf die hier geschilderten Verarbeitungsvorgänge übertragbar. Insoweit bleibt für eine Rechtsgrundlage meist nur der populäre Weg der Einwilligung der Betroffenen (vgl. BDSG §4 Abs. 1 bzw. BDSG §4a). Allerdings ergeben sich im Kontext vernetzter Fahrzeuge hier große Herausforderungen: Wie sieht eine informierte, freiwillige Einwilligung in die Datenverarbeitung im Fahrzeug aus? Den Fahrer könnte man über die (nicht sehr ausdrucksstarken) Kommunikationsmöglichkeiten von Display und Bedienelementen im Armaturenbrett des Fahrzeugs erreichen, um eine Information und eine Einwilligung zu erhalten. Der Fahrzeughalter könnte möglicherweise bereits beim Kauf des Fahrzeugs über die Datenerhebung und -verarbeitung informiert und um Einwilligung gebeten werden.

Schwieriger wird die Einwilligung aber schon in Bezug auf Beifahrer, die ggf. nicht über ein eigenes Fahrzeug-Bedienelement verfügen. Wie in Abschnitt 4.1 ausgeführt werden auch personenbeziehbare Daten dieser Beifahrer umfassend Teil der fahrzeuginternen Datenerfassung, -verarbeitung und -weitergabe. Für häufig wiederkehrende Mitfahrer, etwa Familienmitglieder, ließe sich hier möglicherweise ein Konstrukt ähnlich der Haltereinwilligung zum Zeitpunkt des Fahrzeugkaufes realisieren. Dies dürfte jedoch für seltenere oder gar spontan aufgenommene Beifahrer nicht praktikabel sein.

Gänzlich unbrauchbar wird das Instrument der Einwilligung in Bezug auf Daten über andere Verkehrsteilnehmer. So kann ein Fußgänger an einem Zebrastreifen wohl kaum vorab um Erlaubnis gebeten werden, wenn er von den Kameras bzw. den hochauflösenden Laser- oder Radar-basierten Umgebungssensoren eines (teil-)autonomen Fahrzeugs erfasst und gescannt wird. Nichtsdestotrotz werden Daten über solche Verkehrsteilnehmer zweifelsohne erhoben und verarbeitet, ohne Einwilligung, und häufig ohne Kenntnis der Betroffenen.

Dass eine derartige Datenerhebung unbeteiligter Dritter rechtlich vermutlich unzulässig ist, lässt sich bereits aus der Urteilsbegründung des berühmten Dashcam-Urteils herauslesen (vgl. VG Ansbach in seinem Urteil vom 12.08.2014, Az: AN 4 K 13.01634), bei dem die Rechtmäßigkeit von Videoaufzeichnungen der direkten Umgebung eines Fahrzeugs zum Zwecke der Beweisdokumentation im Unfallfall deutlich angezweifelt wurde. Diese Einschätzung dürfte auf bildgebende Laser- oder Radarabtastungen ähnlich übertragbar sein, die zwar weniger Bildinformation liefern als eine vollwertige Kamera, aber dennoch hinreichend präzise Daten für eine Profilbildung und damit eine potentielle spätere Identifizierung von Personen erheben.

Es bedarf hier folglich einer Abwägung durch den Gesetzgeber: ist der Zweck der (teil-)autonomen Fahrzeugführung mit all ihren Vor- und Nachteilen ein überwiegendes Gut gegenüber den Beeinträchtigungen der Betroffenen? Ist hier ggfs. eine entsprechende Rechtsgrundlage zu schaffen? Zur Beantwortung dieser juristischen Kernfrage des autonomen Fahrzeugs der Zukunft bedarf es sicherlich zunächst weiterer technischer Forschungsdaten, um die tatsächlichen Vor- und Nachteile des autonomen Fahrens gegeneinander abwägen zu können. Rein juristisch ist vermutlich ein Großteil der vernetzt und (teil-)autonom fahrenden Fahrzeuge nach heutigem Rechtsstand nicht völlig rechtskonform im deutschen Straßenverkehr einsetzbar.

5 Lösungswege

5.1 Technische Lösungen

Zur Verbesserung der Privatsphäre können zum großen Teil bekannte Techniken des Datenschutzes verwendet werden. Allerdings können diese auch nur wirksam eingesetzt werden, wenn das komplexe Ökosystem rund um das Fahrzeug der Zukunft in Gänze analysiert wird und alle Bestandteile aufeinander abgestimmt sind. Da dies noch nicht der Fall ist, wird im Folgenden nur angedeutet, wie die jeweiligen Techniken zu einer Gesamtlösung beitragen könnten.

5.1.1 Zugriffskontrolle

Ein wichtiges Werkzeug zum Schutz von Daten (vor Auslesen aber auch vor Veränderung) ist die Zugriffskontrolle. Dabei wird an den Schnittstellen bzw. Diensten, über die auf die Daten zugegriffen wird, überprüft, ob das Subjekt die erforderlichen Privilegien besitzt. Dies erfordert detaillierte Richtlinien (engl. *Policies*), in denen die Zugriffsrechte festgelegt sind.

Während bei netzwerkbasierten Softwaresystemen (beispielsweise bei Web-Diensten) diese Technik standardmäßig verwendet wird, so ist dies bei intelligenten Fahrzeuge offensichtlich nicht der Fall. Dies zeigt auch der Fall des „gekaperten“ Chryslers [Gre].

Ein elaboriertes Zugriffskontroll-System könnte hier sowohl die Sicherheit als auch den Datenschutz verbessern. Man könnte damit den Zugriffe für Aussenstehende unterbinden und für Berechtigte zumindest einschränken. So ist es beispielsweise nicht notwendig, dass ein Mechaniker in der Werkstatt vollen Zugriff auf alle Daten erhält, sondern nur auf diejenigen, die für die momentane Tätigkeit notwendig ist. Weiterhin könnten auch bestimmte Datenzugriffe von außen die (temporäre) Freischaltung durch den Fahrer bzw. Halter erfordern.

5.1.2 Datensparsamkeit

Datensparsamkeit wird explizit im BDSG als Mittel zur Verbesserung des Datenschutzes erwähnt. Es sollten nur so wenig personenbezogene Daten wie möglich erhoben werden. Wie in diesem Artikel dargelegt, sind viele der im Fahrzeug der Zukunft gespeicherten oder kommunizierten Daten personenbezogen oder zumindest personenbeziehbar. Bei vielen der erhobenen Daten stellt sich auch die Frage nach dem Zweck der Speicherung (siehe beispielsweise [Spa16]). Hier sollte eine strenge Auswahl der relevanten Daten erfolgen.

5.1.3 Anonymisierung

Auch Anonymisierung wird bereits im BDSG aufgegriffen. Werden Daten so verändert, dass der Personenbezug nicht mehr enthalten ist, so sind sie für die Einzelperson unkritisch und fallen auch nicht mehr unter das BDSG. Bei vielen Daten rund um das Fahrzeug

der Zukunft ist der Bezug zu einer Person (oder einem Automobil) gar nicht relevant. Beispielsweise ist es zur Verkehrssteuerung typischerweise unwichtig, welche Fahrzeuge an einem Ort sind, nur die Anzahl der Fahrzeuge (und weitere Eigenschaften wie Geschwindigkeit, die keine Zuordnung erlauben) ist nötig.

Allerdings lässt sich Anonymisierung bei ungeschickter Umsetzung oder durch Erlangungen von zusätzlichen Informationen auch wieder aufheben. So kann beispielsweise die Information „silbergrauer Golf“ als anonym angesehen werden, während die Information „pinkfarbener Ferrari“ höchstwahrscheinlich genau ein Fahrzeug beschreibt (zumindest in einer Stadt). Ein anderes Beispiel wurde bereits im Abschnitt 4.2 beschrieben: das Bewegungsprofil einer (anonymen) Person ist unkritisch. Allerdings kann durch genaue Analyse und weitere Informationen (in welchen Hauseingang geht diese Person und wer wohnt in diesem Haus) auch hier die Anonymität wieder aufgehoben werden.

Anonymisierung ist also ein sehr mächtiges Werkzeug, welches die Weiterverarbeitung von Daten unter Beibehaltung der Privatsphäre ermöglicht. Allerdings müssen auch die Randbedingungen genau beachtet werden.

5.2 Rechtliche Lösungen

Wie bereits oben erwähnt, widersprechen wahrscheinlich ein Großteil der aktuellen oder geplanten Datenverwendungen den gesetzlichen Bestimmungen. Hier sollte die Einhaltung der Normen, z.B. von Datenschutzbeauftragten, verstärkt überprüft werden.

Weiterhin stellt sich aber auch die Frage, ob neuartige technische Systeme wie das Fahrzeug der Zukunft noch hinreichend vom BDSG abgedeckt werden können und nicht neue Datenschutzbestimmungen erforderlich sind. Hier wird interessant zu analysieren sein, wie sich die neue Datenschutz-Grundverordnung (vgl. [DSG16]) auf die Datenerhebung und -verarbeitung im Fahrzeug der Zukunft auswirken wird.

6 Zusammenfassung und Ausblick

Der vorliegende Artikel hat dargelegt, welche Datenschutzprobleme beim Fahrzeug der Zukunft auftreten können. Kritisch ist dabei die Kombination aus einer Vielzahl von Sensoren, exzessiver Speicherung von Sensordaten und der Vernetzung des Fahrzeugs mit vielen anderen Instanzen. Betrachtet man die aktuell bereits existierenden Systeme, so scheint zusätzlich beim Entwurf sowohl Sicherheit als auch Datenschutz nicht besonders beachtet worden zu sein.

Hier lassen sich in Zukunft durch verbessertes Systemdesign und Berücksichtigung von IT-Sicherheits- und Datenschutz-Mechanismen noch deutliche Fortschritte erzielen. Zusätzlich muss verstärkt auf die Einhaltung der gesetzlichen Bestimmungen geachtet werden, und diese sind gegebenenfalls an die neuartigen Anforderungen anzupassen.

Literatur

- [ADA16] ADAC. Autos mit Keyless leichter zu klauen. <https://www.adac.de/keyless>, 2016.
- [BFN] Manuel Bewarder, Florian Flade und Lars-Marten Nagel. BSI-Chef warnt vor Hackerangriffen auf Autos und Flugzeugen. *DIE WELT*, 26.04.2016.
- [DSG16] Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung), April 2016.
- [EU609] Verordnung (EG) Nr. 661/2009 des Europäischen Parlaments und des Rates über die Typgenehmigung von Kraftfahrzeugen, Kraftfahrzeuganhängern und von Systemen, Bauteilen und selbstständigen technischen Einheiten für diese Fahrzeuge hinsichtlich ihrer allgemeinen Sicherheit, 2009.
- [Gom16] Lee Gomes. When will Google's self-driving car really be ready? It depends on where you live and what you mean by "ready". *IEEE Spectrum*, 53(5):13–14, 2016.
- [Gre] Andy Greenberg. Hackers Remotely Kill a Jeep on the Highway—With Me in It. *WIRED*, 21.07.2015.
- [Gui11] Erico Guizzo. How google's self-driving car works. *IEEE Spectrum Online*, October, 18, 2011.
- [Rei11] Konrad Reif. *Bosch Autoelektrik und Autoelektronik*. Springer, ISBN: 978-3-8348-1274-2 (Print) 978-3-8348-9902-6 (Online), 2011.
- [RMM⁺10] Ishtiaq Rouf, Rob Miller, Hossen Mustafa, Travis Taylor, Sangho Oh, Wenyuan Xu, Marco Gruteser, Wade Trappe und Ivan Seskar. Security and Privacy Vulnerabilities of In-car Wireless Networks: A Tire Pressure Monitoring System Case Study. In *Proceedings of the 19th USENIX Conference on Security*, USENIX Security'10, Seite 21 ff., Berkeley, CA, USA, 2010. USENIX Association.
- [Spa16] Dieter Spaar. Daten auf Rädern: Was moderne Autos speichern und wie man an die Informationen herankommt. In *c't Magazin 09/2016*, Heise-Verlag, ISSN 0724-8679, Seite 170 ff., 2016.
- [TMD⁺06] Sebastian Thrun, Mike Montemerlo, Hendrik Dahlkamp, David Stavens, Andrei Aron, James Diebel, Philip Fong, John Gale, Morgan Halpenny, Gabriel Hoffmann, Kenny Lau, Celia Oakley, Mark Palatucci, Vaughan Pratt, Pascal Stang, Sven Strohband, Cedric Dupont, Lars-Erik Jendrossek, Christian Koelen, Charles Markey, Carlo Rummel, Joe van Niekerk, Eric Jensen, Philippe Alessandrini, Gary Bradski, Bob Davies, Scott Ettinger, Adrian Kaehler, Ara Nefian und Pamela Mahoney. Stanley: The robot that won the DARPA Grand Challenge. *Journal of Field Robotics*, 23(9):661–692, September 2006.

Datenschutzziele im Konflikt: Eine Metrik für Unverkettbarkeit als Hilfestellung für den Betroffenen

Christoph Bier¹

Abstract: Die 6 Datenschutz-Schutzziele befinden sich in einem dauerhaften Spannungsverhältnis. Insbesondere Transparenz und Unverkettbarkeit kollidieren. Das Recht auf Auskunft ist ein prominentes Beispiel für den Konflikt zwischen den beiden Schutzzielen. Es erfordert eine umfangreiche Rückverfolgung und mithin eine stärkere Verkettung personenbezogener Daten um den Betroffenen über den Umgang mit denselben in Kenntnis zu setzen. Im Sinne eines mündigen Bürgers sollte der Betroffene entscheiden, was ihm wichtiger ist: Die Unverkettbarkeit oder eine möglichst umfangreiche Auskunft. Dies bedarf jedoch einer klaren Kommunikation der Konsequenzen.

In diesem Beitrag wird auf Grundlage rechtlicher Anforderungen eine informationstheoretische Metrik für Unverkettbarkeit vorgestellt, modelliert und für ein automatisiertes Datenschutzauskunftssystem instantiiert. Die Metrik bildet die notwendige Informationsgrundlage, um es Betroffenen zu erlauben, selbst über den Trade-off zwischen Transparenz und Unverkettbarkeit zu entscheiden. Eine prototypische Implementierung zeigt die Berechenbarkeit der Metrik während das Ergebnis einer Nutzerstudie mit 31 Teilnehmern eine gute Akzeptanz der Metrik nahelegt.

Keywords: Datenschutz, Unverkettbarkeit, Transparenz, Auskunft, Metrik, Entropie

1 Einleitung

Das Datenschutzrecht versucht mit einer Vielzahl von Regelungen das informationelle Selbstbestimmungsrecht der Bürger und deren Handlungs- und Mitwirkungsfähigkeit im freiheitlich-demokratischen Gemeinwesen der Bundesrepublik Deutschland sicherzustellen.² Das Bundesdatenschutzgesetz (BDSG) und in die EU-Datenschutzgrundverordnung (EU-DSGVO) enthalten unterschiedlichste Anforderungen an die Verarbeitung personenbezogener Daten. Die in den letzten Jahren entwickelten Datenschutz-Schutzziele [RB11] geben dieser Vielfalt Orientierung. Sie konkretisieren das Recht auf informationelle Selbstbestimmung ohne der Gestaltung des Datenschutzrechts im Einzelnen vorzugreifen. Im Einzelnen lauten die Datenschutzziele *Vertraulichkeit*, *Integrität*, *Verfügbarkeit*, *Transparenz*, *Unverkettbarkeit* und *Intervenierbarkeit*. Diese Ziele sind allerdings nicht konfliktfrei. Insbesondere Transparenz und Unverkettbarkeit stehen im Widerspruch zueinander [HJR15].

Das Ziel der *Transparenz* fußt auf dem Recht des Einzelnen, beurteilen zu können, wer wann was über ihn weiß.³ Das Auskunftsrecht als Teil der Transparenz ist für den Betroffenen das wichtigste Datenschutzrecht [Si11]. Es ist Voraussetzung zur Wahrnehmung

¹ Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung IOSB, Fraunhoferstr. 1, 76131 Karlsruhe, christoph.bier@iosb.fraunhofer.de

² BVerfGE 65, 1 (43).

³ BVerfGE 65, 1 (43); BVerfGE 125, 260 (334).

der übrigen Betroffenenrechte auf Löschung, Sperrung und Berichtigung. Trotz seiner enormen Bedeutung für einen effektiven Datenschutz wird das Recht auf Auskunft durch die Praxis vernachlässigt. Auskünfte werden zwar erteilt, jedoch nur in Form von statischen Datenbankauszügen. An dieser Stelle setzt Personal-Data-Provenance an. Personal-Data-Provenance ist die dokumentierte Historie eines personenbezogenen Datums. Eine Provenance-Tracking-Infrastruktur verfolgt demnach den Lebenszyklus eines personenbezogenen Datums ausgehend von der Erhebung beim Betroffenen oder einem Dritten, über einzelne Verarbeitungsschritte bis hin zur Übermittlung. Alle Schritte werden mit dem Zweck der Erhebung und Verarbeitung des personenbezogenen Datums in Bezug gesetzt. Letztendlich soll der Betroffene die Möglichkeit bekommen, über eine Datenschutzauskunftsplattform jederzeit Einblick in den Umgang mit seinen personenbezogenen Daten zu nehmen [An15, BK16].

Unverkettbarkeit soll verhindern, dass es staatlichen Behörden und privaten Organisationen möglich ist, ein umfangreiches Persönlichkeitsprofil über jeden Einzelnen zu erstellen.⁴ Unverkettbarkeit fordert, dass personenbezogene Daten, Betroffene, Verarbeitungsprozesse und Nutzungsdomänen nicht miteinander verknüpft werden können. Aus datenschutzrechtlicher Sicht sind der Unverkettbarkeit insbesondere der Zweckbindungsgrundsatz⁵ sowie die informationelle und organisatorische Gewaltenteilung inhärent.

Transparenz erfordert eine ergänzende personenbezogene Sammlung von Protokolldaten und ihre Verknüpfung mit Bezug auf den Betroffenen. Jedes Mehr an Daten erhöht jedoch die Gefahr der Verkettbarkeit. Transparenz setzt die vollständige Verknüpfung von Daten und Verarbeitungsprozessen mit dem Betroffenen voraus, während Unverkettbarkeit des genauen Gegenteils bedarf. Das Ziel des in diesem Beitrag vorgestellten Ansatzes ist es, diesen Widerspruch sichtbar zu machen, um ihn, wenn möglich, aufzulösen. Entlang des Anwendungsfalls „Datenschutzauskunftssystem“ wird ein allgemeingültiges Modell für eine informationstheoretische Metrik für Unverkettbarkeit entworfen. Sie wird für ein Beispiel instantiiert und erläutert.

Verwandte Arbeiten. Ein Standardwerk zur Terminologie von Unverkettbarkeit in der Informatik ist bis heute die bis 2010 aktualisierte Veröffentlichung von Pfitzmann und Hansen [PH10]. Bohli und Pashalidis [BP11] formalisieren unterschiedliche Abstufungen der klassischen Unverkettbarkeit basierend auf der Unterscheidbarkeit von Systemrealisationen durch Angreifer mit unterschiedlichem, fest definierten Hintergrundwissen im Stil des IND-CPA-Modells zur Sicherheit kryptographischer Verfahren.

Die Idee, Anonymität informationstheoretisch zu beschreiben, wird bereits von Serjantov und Danezis ins Spiel gebracht [SD03]. Sie überführen das klassische „Anonymity Set“ auf ein nach den Wahrscheinlichkeiten der einzelnen Elemente der Menge gewichtetes Maß. Die Arbeiten von Steinbrecher und Köpsel [SK03] übertragen den informationstheoretischen Ansatz auf Unlinkability. Der Ansatz wird von Pashalidis [Pa08] von Äquivalenzrelationen auf alle zweistelligen Relationen verallgemeinert.

⁴ BVerfGE 65, 1 (42).

⁵ BVerfGE 65, 1 (43).

Gliederung. Der verbleibende Beitrag ist wie folgt gegliedert: Im Anschluss wird das Minimalbeispiel, anhand dessen die Berechnung der Metrik für Unverkettbarkeit verdeutlicht wird, geschildert. Anschließend werden im Abschnitt 1.1 die rechtlichen Anforderungen an die Ausgestaltung einer Unverkettbarkeitsmetrik erörtert. Das Systemmodell, die betrachteten Entitäten, das Angreifermodell und die informationstheoretische Metrik für Unverkettbarkeit werden in Abschnitt 2 eingeführt und im Abschnitt 2.3 für die unterschiedlichen Instanzierungen der Metrik konkret beschrieben. Ihre Implementierung wird im Abschnitt 2.4 erläutert. Im letzten Abschnitt wird der Nutzen der beschriebenen Metrik bewertet und ein Ausblick auf mögliche Erweiterungen gegeben.

Fortlaufendes Beispiel. Das fortlaufende Beispiel in diesem Beitrag nimmt ein Unternehmen namens AdBokis Buchclub GmbH, einen fiktiven Online-Händler für Bücher und Software, in den Fokus. Alice Fox ist Kundin dieses Händlers und möchte nach erfolgtem Einkauf ihr Auskunftsrecht wahrnehmen. Sie hat nach § 34 BDSG Anspruch auf Auskunft über die zu ihrer Person gespeicherten personenbezogenen Daten, deren Herkunft, Empfänger und den Zweck der Speicherung. Empfänger können der Betroffene, Dritte, Auftragsdatenverarbeiter und Stellen innerhalb der verantwortlichen Stelle sein. Der Stellenbegriff ist funktional und organisatorisch definiert [Si11].

In unserem Minimalbeispiel hat AdBokis zwei Kunden (Betroffene $b \in \mathcal{B}$): Alice Fox (b_1, s_1) und Peter Trollig (b_2, s_2). Die CloudyCloud GmbH (s_3) ist als Auftragsdatenverarbeiter für AdBokis tätig. Außerdem übermittelt AdBokis im Rahmen ihrer Geschäftsprozesse personenbezogene Daten an die PayPortal Inc. (s_4) und die Bonus Card GmbH (s_5). Intern spielen bei der Datenverarbeitung die Abteilungen Kundenbetreuung (s_6), Vertrieb, IT und Infrastruktur (s_{10}) und Recht (s_{11}) eine Rolle. In der Abteilung Vertrieb wird neben dem System für den Onlineverkauf (s_7) auch ein Archivserver (s_8) betrieben. Zudem gibt es Arbeitsplatzsysteme, die im Vertrieb normalerweise nicht für die Verarbeitung personenbezogener Daten vorgesehen sind. Exemplarisch ist deshalb im Minimalbeispiel der Workspace23 (s_9) enthalten. Alle diese Entitäten werden als Systeme $s \in \mathcal{S}$ bezeichnet. Die Verwendung des Begriffs *System* darf nicht verwirren. Ein System ist eine Ansammlung technischer (z.B. ein Cluster) oder organisatorischer (z.B. eine Abteilung) Entitäten, denen ein gemeinsames Wissen unterstellt wird.

Insgesamt verarbeitet AdBokis 30 personenbezogene Daten $d \in \mathcal{D}$ ihrer beiden Kunden in 17 Datenkategorien $\theta \in \Theta$. In Tabelle 1 sind exemplarisch die personenbezogenen Daten von Alice Fox aufgelistet.

Die Verarbeitung personenbezogener Daten findet entlang etablierter Verarbeitungsprozesse statt. Diese sind von der verantwortlichen Stelle gemäß § 4g Abs. 2 S. 1 BDSG i.V.m § 4e Satz 1 BDSG in einem internen Verzeichnisse zu dokumentieren. Teil dieser Dokumentation sind die verarbeiteten Datenkategorien, eine Beschreibung des Verfahrens sowie die möglichen Empfänger der Daten. Aus letzterer Information ergibt sich die Verknüpfung einzelner Verarbeitungsprozesse.

Auf eine Beschreibung des Datenschutzauskunftssystems wird aus Platzgründen an dieser Stelle verzichtet. Die wesentlichen Eigenschaften werden im Rahmen des fortlaufenden Beispiels erläutert.

d_x	θ_x	Datenkategorie	Inhalt
1	1	Vorname	Alice
2	2	Name	Fox
3	3	e-Mail	alice.fox@honigmail.de
		...	
16	14	Profilbild	[nicht darstellbar]
17	15	IP-Adresse	217.146.191.19
18	15	IP-Adresse	31.130.202.80
19	16	Empfehlung	Inges Braustubenführer
20	17	Rechnung	[nicht darstellbar]

Tab. 1: Auszug der Verarbeiteten Daten mit Personenbezug zu Alice

1.1 Aus der Unverkettbarkeit resultierende Anforderungen an ein Datenschutzauskunftssystem

Wie bereits im Abschnitt 1 erwähnt folgen aus der Unverkettbarkeit die Anforderungen der Zweckbindung und Zwecktrennung sowie die Organisatorische und technische Gewaltenteilung.

Zweckbindung und Zwecktrennung. Die Zweckbindung fordert, dass personenbezogene Daten nur zu dem Zweck verarbeitet und genutzt werden dürfen, zu dem sie auch erhoben wurden und der für sie dokumentiert wurde. Unter vielen findet sich diese Festlegung beispielsweise in § 28 Abs. 3 S. 7 BDSG. Die Zwecktrennung ist Ausfluss eines wesentlichen Aspekts der Unverkettbarkeit. Personenbezogene Daten, die zu unterschiedlichen Zwecken verarbeitet, insbesondere gespeichert, werden, dürfen nicht zusammengeführt werden. Mit technischem Bezug ist dies in der Anlage zu § 9 S. 1 BDSG Nr. 8 festgelegt.

Organisatorische und technische Gewaltenteilung. Die organisatorische und technische Gewaltenteilung im Rahmen der informationellen Gewaltenteilung folgt aus dem Gebot der Zweckbindung und Zwecktrennung. Ihr liegt das verwaltungsrechtliche Abschottungsprinzip zugrunde. Die informationelle Gewaltenteilung ist nicht im BDSG festgelegt, sondern ergibt sich aus der Rechtsprechung des Bundesverfassungsgerichts,⁶ in dessen Lichte das BDSG auszulegen ist. Während sich Zweckbindung und Zwecktrennung auf die personenbezogenen Daten selbst beziehen, ist die informationelle Gewaltenteilung eine Forderung, die direkt an die organisatorischen und technischen Einrichtungen gestellt wird. Die Zwecktrennung untersagt die Zusammenführung personenbezogener Daten. Die Gewaltenteilung verpflichtet dazu, dass Zugriffsrechte, Rollen sowie physische und logische Speicherorte nicht beliebig festgelegt werden. Sie sind entsprechend dem Zweck und nach dem Prinzip der Machtdistribution zu bestimmen.

Die organisatorische und technische Gewaltenteilung findet sich bisher noch kaum in expliziten Regelungsvorgaben des Bundesdatenschutzgesetzes. Sie ergibt sich jedoch indirekt aus anderen Anforderungen und Kriterien. Zunächst darf durch ein Datenschutzaus-

⁶ BVerfGE 65, 1 (69); BVerfG, NJW 1988, 959 (961).

kunftssystem keine Verknüpfung unterschiedlicher Daten möglich sein, soweit dies für die Auskunft nicht zwingend erforderlich ist. Außerdem dürfen die Bearbeitungswege und Weitergaben eines personenbezogenen Datums für eine Stelle, auch wenn sie an der Verarbeitung beteiligt ist, nicht durch das Datenschutzauskunftssystem nachvollziehbar gemacht werden. Daraus ergeben sich direkt die in Abschnitt 2.2 beschriebenen vier Unverkettbarkeitsrelationen.

2 Eine Metrik für Unverkettbarkeit

Unabdingbare Voraussetzung um über Unverkettbarkeit sprechen zu können, ist ein Modell des Ausschnitts der Realität zu definieren, in dem Unverkettbarkeit gemessen werden soll. Insbesondere sind die betrachteten Entitäten (\mathcal{E} - *items of interest*) und ihre Beziehungen zueinander (Verkettungsrelationen R) festzulegen. Außerdem ist ein Angreifer \mathcal{A} zu definieren, aus dessen Perspektive die Unverkettbarkeit bestimmt wird [PH10]. Eine Verkettungsrelation R ist eine Teilmenge des kartesischen Produkts von $n \geq 2$ Teilmengen $E_1, \dots, E_n \subseteq \mathcal{E}$ der Menge \mathcal{E} aller Entitäten: $R \subseteq E_1 \times \dots \times E_n$. (im Regelfall die Entitätsmengen der Entitätsklassen).

Unverkettbarkeit kann absolut oder relativ definiert werden. Absolut gesehen sind zwei oder mehrere Entitäten aus Sicht eines Angreifers dann unverkettbar, wenn der Angreifer nicht feststellen kann, ob die Entitäten innerhalb des definierten Modells in einem bestimmten Verhältnis zueinander stehen, oder nicht [PH10]. In einer Situation, in der ein gewisser Wissenszuwachs unabdingbar ist und in Kauf genommen wird, wie bei Auskunftssystemen, ist solch eine Metrik nicht hilfreich. Sie würde jederzeit trivial messen, dass die Unverkettbarkeit nicht gewahrt bleibt. Relative Unverkettbarkeit vergleicht die Unsicherheit des Angreifers \mathcal{A} bezüglich der wahren Verkettungsrelation R_τ nach Interaktion mit dem Gesamtsystem $\Sigma^{\mathcal{A}}$ mit der Unsicherheit, die bereits vor der Interaktion mit dem modellierten System bestand. Die Unsicherheit vor Interaktion ist vom Hintergrundwissen (A-priori-Wissen) des Angreifers abhängig. Die Interaktion mit $\Sigma^{\mathcal{A}}$ lässt den Angreifer die Beobachtungen I machen. Das kombinierte Wissen des Angreifers aus Hintergrundwissen und Beobachtungen wird auch als A-posteriori-Wissen bezeichnet. In der Literatur hat sich die informationstheoretische Bestimmung relativer Unverkettbarkeit etabliert (siehe Abschnitt 1). Im gesamten Beitrag werden Wahrscheinlichkeiten gemäß des Bayesschen Wahrscheinlichkeitsbegriffes als „Grad (vernünftiger) Glaubwürdigkeit/persönlicher Überzeugung“ (*degree of belief*) verwendet.

Sei X eine Zufallsvariable über der endlichen Menge der Kandidatenrelationen \mathcal{R} . Sowohl vor als auch nach Interaktion mit dem Gesamtsystem $\Sigma^{\mathcal{A}}$ weist der Angreifer \mathcal{A} allen Kandidatenrelationen $R \in \mathcal{R}$ einen Wahrscheinlichkeitswert $\mathbb{P}(X = R)$ zu. $\mathbb{P}(X = R)$ ist die angenommene Wahrscheinlichkeit, dass R die tatsächliche Relation R_τ zwischen den Entitäten aus E_1, \dots, E_n ist. Dann ergibt sich die Entropie des A-priori- bzw. A-posteriori-Wissens des Angreifers als:

$$H(X) = - \sum_{R \in \mathcal{R}} \mathbb{P}(X = R) \log_2 \mathbb{P}(X = R) \quad [\text{bit}]$$

Wobei $\mathbb{P}(X = R) \log_2 \mathbb{P}(X = R) = 0$ für $\mathbb{P}(X = R) = 0$ angenommen wird. Die Entropie misst die Informationsmenge, die \mathcal{A} noch braucht, um R_τ vollständig zu identifizieren.

Der Grad der Unverkettbarkeit ergibt sich als Verhältnis zwischen A-priori- und A-posteriori-Entropie (mit dem Beobachtungsereignis I):

$$\Delta(X, I) = \frac{H(X | I)}{H(X)}$$

Der Grad der Unverkettbarkeit beschreibt das Verhältnis zwischen der Situation nach und der Situation vor der Interaktion des Angreifers \mathcal{A} mit dem System $\Sigma^{\mathcal{A}}$ bezüglich des noch benötigten Wissens zur vollständigen Aufdeckung der Relation.

In bisherigen Arbeiten werden meist die A-priori-Situation und die maximale Unverkettbarkeit gleichgesetzt (Maximum-Entropie-Prior, $H(X) = H_{\max}(X)$). Die maximale Unverkettbarkeit ist $H_{\max}(X) = \log_2(|\mathcal{R}|)$. Ein Maximum-Entropie-Prior macht im diskutierten Szenario indes keinen Sinn. A priori sind damit Beobachtungen aus Datenverarbeitungsvorgängen ohne den Einsatz eines Datenschutzauskunftssystems vorausgesetzt. A posteriori werden die Beobachtungen aus den selben Datenverarbeitungsvorgängen unter Berücksichtigung des Einsatzes eines Datenschutzauskunftssystems ins Angreiferwissen mit aufgenommen. Statt eines Maximum-Entropie-Priors ist der Vergleichszustand (subjektiver Prior) schon ein Zustand mit partiellem Wissen.

Sei $H(X) \neq H_{\max}(X)$, dann ist bei Beobachtungen, die der A-priori-Annahme entgegengesetzt sind, hypothetisch ein $\Delta(X, I) > 1$ möglich. Allerdings ist als Maß der Unverkettbarkeit nicht die Grad der Unverkettbarkeit bezüglich eines bestimmten Angreifers von Interesse, sondern der niedrigste Grad über alle Angreifer. Trivial lässt sich immer ein Angreifer konstruieren, der kein Hintergrundwissen hat ($H(X) = H_{\max}(X)$) und durch seine Beobachtungen nichts dazulernen kann ($H(X | I) = H(X)$). Dessen Grad der Unverkettbarkeit ist immer $\Delta(X, I) = 1$. Somit ist der normierte globale Grad der Unverkettbarkeit $\|\Delta\| = \min_{\mathcal{A}} (\{\Delta(X_{\mathcal{A}}, I_{\mathcal{A}})\}) \in [0; 1]$.

2.1 Der Angreifer \mathcal{A}

Das Angreifermodell gibt die Leitlinien vor, an denen sich die Analyse der A-priori- und der A-posteriori-Situationen orientieren kann. Der Angreifer kann Teil der datenverarbeitenden Organisation (verantwortliche Stelle) sein oder außerhalb der Organisation zu finden sein. Im Folgenden werden exemplarisch nur zwei interne Angreifer betrachtet.

Der *Systemangreifer* \mathcal{A}^s verarbeitet möglicherweise selbst personenbezogene Daten. Er möchte aber Wissen über weitere Verarbeitungsvorgänge gewinnen. Vorstellbar ist beispielsweise eine Marketingabteilung, die wissen möchte, in welchem Maße und unter Preisgabe welcher Informationen ein Kunde bisher den Kundenservice angefragt hat. Der *zentrale Angreifer* \mathcal{A}^c entsteht erst durch das Datenschutzauskunftssystem. Die Datenschutzauskunft erfordert eine Aggregation der Personal-Data-Provenance vor der Weitergabe an den Betroffenen. Sie setzt voraus, dass ein Einstiegspunkt für den Abruf der

gesamten Provenance-Kette bekannt ist. Nur so kann die Vollständigkeit der Provenance gewährleistet werden. \mathcal{A}^s und \mathcal{A}^c werden als passive Angreifer angenommen. Sie halten die festgelegten Kommunikationsprotokolle des Datenschutzauskunftssystems vollständig ein. Eine Missachtung der Kommunikationsprotokolle kann von den Kommunikationspartnern festgestellt und organisatorisch verfolgt werden.

Das a-priori Hintergrundwissen der Angreifer umfasst das interne Verzeichnis und allgemeine Unternehmensstatistiken. Das Verzeichnis beinhaltet Informationen zu den Verarbeitungsprozessen, den beteiligten Systemen und den verwendeten Datenkategorien (siehe Abschnitt 1).

Annahme 1. *Dem Angreifer ist die Art und die Anzahl aller Systeme $s \in \mathcal{S}$ a-priori bekannt.*

Annahme 2. *Dem Angreifer ist die Anzahl der von der Datenverarbeitung betroffenen Kunden $|\mathcal{B}|$ a-priori bekannt. Dem Angreifer ist die Anzahl der verarbeiteten personenbezogenen Daten $|\mathcal{D}|$ a-priori bekannt.*

Personenbezogene Daten eines Betroffenen werden unabhängig davon erhoben und verarbeitet, ob personenbezogene Daten anderer Betroffener erhoben oder verarbeitet werden. Ob ein Kunde beispielsweise die Bezahlung mit Kreditkarte wählt hat keinen Einfluss darauf, ob dies ein anderer Kunde auch tut.

Annahme 3. *Die Zugehörigkeit eines personenbezogenen Datums zu einem Betroffenen (und umgekehrt) ist unabhängig von der Zugehörigkeit eines anderen personenbezogener Datums zu einem anderen Betroffenen (und umgekehrt).*

Ein Datum kann potentiell personenbezogenes Datum mehrerer Betroffener sein. Um die folgenden Überlegungen zu vereinfachen, wird dennoch angenommen, dass ein Datum nur einen Personenbezug zu einem Betroffenen haben kann.

Annahme 4. *Das Verhältnis von personenbezogenen Daten und Betroffenen ist eine $n:1$ -Beziehung.*

Die beiden letzten Annahmen 5 und 6, sind wichtige Annahmen zur Unabhängigkeit von Datenflüssen. Sie sind eine entscheidende Voraussetzung für die Berechenbarkeit der Unverkettbarkeitsmetriken. Beide Annahmen werden in der Realität nicht unbedingt vollständig eingehalten. Die durch sie induzierte Ungenauigkeit kann jedoch nur zu einem Unterschätzen des A-priori-Wissens des Angreifers führen. Das Delta zur A-posteriori-Situation wird dann größer. Die Gefährdung für den Datenschutz wird überschätzt. Deshalb sind die Annahmen vom Ergebnis her gedacht sinnvoller, als unbelegte Annahmen über die Abhängigkeit von Datenflüssen zu treffen, welche zu einem Unterschätzen des Datenschutzrisikos führen könnten.

Annahme 5. *Das a-priori Wissen zu Datenflüssen (Verarbeitungsprozessen) ist nur von der Kategorie der Daten, nicht von den Daten selbst abhängig.*

Annahme 6. *Die Flüsse zweier Daten sind unabhängig voneinander.*

2.2 Instanziierung der Unverkettbarkeit als Gegenspieler der Transparenz

Die relevanten Entitäten ergeben sich aus den Teilinformationen der Datenschutzauskunft. Es sind die Systeme $s \in \mathcal{S}$, die personenbezogenen Daten $d \in \mathcal{D}$ und die Betroffenen $b \in \mathcal{B}$. Gleiches gilt für die Verkettungsrelationen, die über diesen Entitäten definiert sind. Sie bilden das Interesse des Angreifers an den zu einem Betroffenen gespeicherten personenbezogenen Daten ($R^<$), an der Herkunft und den Empfängern personenbezogener Daten ($R^>$) und am zweckbestimmten Verarbeitungsort personenbezogener Daten (R^∇) ab. $R^=$ ergibt sich aus dem Gebot der Zwecktrennung. Wird die Zwecktrennung überwunden, kann ein Persönlichkeitsprofil des Betroffenen hergestellt werden, unabhängig davon, ob schon klar ist, wer er ist. Die vier genannten Relationen sind wie folgt definiert: (1) Die *Identifikationsrelation* $R^< \subseteq \mathcal{D} \times \mathcal{B}$ gibt an, ob das Datum $d \in \mathcal{D}$ einen Personenbezug auf den Betroffenen $b \in \mathcal{B}$ besitzt. (2) Die *Verknüpfungsrelation* $R^= \subseteq \mathcal{D} \times \mathcal{D}$ gibt an, ob zwei Daten $d_1, d_2 \in \mathcal{D}$ einen Personenbezug auf denselben (aber unbekanntem) Betroffenen besitzen. (3) Die *Speicher- und Verarbeitungsrelation* $R^\nabla \subseteq \mathcal{S} \times \mathcal{D}$ gibt für alle Systeme $s \in \mathcal{S}$ an, ob sie das Datum $d \in \mathcal{D}$ verarbeitet und/oder gespeichert haben. (4) Die *Datenflussrelation* $R^> \subseteq \mathcal{S} \times \mathcal{S} \times \mathcal{D}$ gibt für zwei Systeme $s_1, s_2 \in \mathcal{S}$ an, ob sie für ein bestimmtes personenbezogenes Datum $d \in \mathcal{D}$ in einer direkten Vorgänger-Nachfolger-Beziehung stehen.

$\Delta(X^<, I)$ und $\Delta(X^=, I)$ sind globale Metriken. Bei der Bestimmung des Grads der Unverkettbarkeit sind alle Betroffenen \mathcal{B} und alle personenbezogenen Daten \mathcal{D} mit einzubeziehen. Anders stellt sich die Situation bei $\mathcal{R}^>$ und \mathcal{R}^∇ dar. Der Grad der Unverkettbarkeit bezüglich dieser Mengen ist global und lokal bestimmbar. Lokal meint die Fokussierung auf bestimmte Systeme oder Betroffenen. Im Kontext der Datenschutzauskunft ist für einen Betroffenen nur relevant, wie sich die Unverkettbarkeit der Flüsse seiner personenbezogenen Daten entwickelt. Deshalb werden im Abschnitt 2.3.2 nur die Daten in der Teilmenge $\mathcal{D}_{\mathcal{B}} \subseteq \mathcal{D}$. Im Text wird dennoch im Sinne einer allgemeingültigen Darstellung von \mathcal{D} gesprochen. Gleichzeitig wird im Abschnitt 2.3.2 angenommen, dass dem Angreifer a priori bekannt ist, welche personenbezogenen Daten (aber nicht welcher Kategorie) zu welchem Betroffenen gehören. Die Unsicherheit über dieses Faktum wird bereits durch den Grad der Unverkettbarkeit von $\mathcal{R}^<$ gemessen.

2.3 Modellierung des A-priori- und A-posteriori-Wissens der Angreifer

Um den Grad der Unverkettbarkeit bezüglich der vier genannten Relationen zu bestimmen, ist es erforderlich, das Hintergrundwissen der Angreifer und den Wissenszuwachs durch die Einführung der Datenschutzauskunft messbar zu machen. Das Hintergrundwissen der Angreifer geht in die A-priori-Wahrscheinlichkeiten $\mathbb{P}(X = R)$ mit ein. Der Wissenszuwachs der Angreifer wird durch das Beobachtungsereignis I und die daraus resultierenden A-posteriori-Wahrscheinlichkeiten $\mathbb{P}(X = R | I)$ erklärt. A-priori- und A-posteriori-Wahrscheinlichkeitsverteilungen aller vier Relationen werden in diesem Abschnitt erläutert und anhand des Minimalbeispiels aus Kapitel 1 bestimmt.

2.3.1 Bestimmung der Wahrscheinlichkeitsverteilungen von $X^<$ und $X^=$

Die Mächtigkeit der Menge der Kandidatenrelationen ist unter Berücksichtigung von Annahme 4 für die Identifikationsrelation durch $|\mathcal{R}^<| = |\mathcal{B}|^{|\mathcal{D}|}$ gegeben.

Beispiel. Die Anzahl der Kandidatenrelationen für $|\mathcal{D}| = 30$ und $|\mathcal{B}| = 2$ ist 2^{30} .

$\mathcal{R}^=$ ist eine Äquivalenzrelation. Bei Äquivalenzrelation ist die Anzahl möglicher Relationen durch die Bellsche Zahl $B_{|\mathcal{D}|}$ gegeben. Die Bellsche Zahl lässt sich mit Hilfe der Stirling-Zahl zweiter Art bestimmen. Da $\mathcal{R}^=$ auf $\mathcal{R}^<$ zurückzuführen ist, sind die tatsächlichen Kandidatenrelationen durch die Anzahl der Betroffenen $|\mathcal{B}|$ beschränkt. Nur k -Partitionen mit $k \leq |\mathcal{B}|$ sind möglich. Die Formel ist deshalb in korrigierter Form anzuwenden: $|\mathcal{R}^=| = \sum_{k=0}^{\min(|\mathcal{B}|, |\mathcal{D}|)} S_{|\mathcal{D}|, k}$

Beispiel. Die Anzahl der Kandidatenrelationen für $|\mathcal{D}| = 30$ und $|\mathcal{B}| = 2$ ist $|\mathcal{R}^=| = \sum_{k=0}^2 S(30, k) = 536870912$.

Für beide Relationen gilt, dass es, mit Ausnahme der Mächtigkeit der Mengen \mathcal{B} und \mathcal{D} , kein globales Hintergrundwissen gibt.

Für den Systemangreifer \mathcal{A}^s unterscheidet sich der Unverkettbarkeitsprior nicht vom Posterior. Durch das Datenschutzauskunftssystem werden auf den Systemen nur solche Teile der Provenance vorgehalten, die auf Ereignisse im jeweilige System zurückzuführen sind. Daraus folgt für den Systemangreifer $\Delta(X^<, I) = \Delta(X^=, I) = 1$. Der zentrale Angreifer verarbeitet selbst keine personenbezogenen Daten. Für ihn sind $H(X^<) = H_{\max}(X^<) = \log_2 |\mathcal{R}^<|$ und $H(X^=) = H_{\max}(X^=) = \log_2 |\mathcal{R}^=|$. A posteriori, also unter Einsatz des Datenschutzauskunftssystems, erhält der zentrale Angreifer weitere Informationen I . Bei jeder Erhebung eines personenbezogenen Datums wird ihm ein pseudonymer Identifikator für das erhobene Datum zusammen mit Informationen zum Betroffenen übermittelt. Daraus kann der Angreifer zwar nicht schließen, welches personenbezogene Datum oder welche Kategorie personenbezogener Daten erhoben wurde. Allerdings kann er bestimmen, wie viele personenbezogenen Daten für jeden einzelnen Betroffenen erhoben wurden. Kandidatenrelationen, die keine entsprechende Struktur aufweisen, kann er ausschließen. Seien die dem Angreifer bekannt gewordenen k -Partitionen für die Menge der Daten \mathcal{D} von der Größe l_1, l_2, \dots, l_k . Dann sind

$$|(R^= \in \mathcal{R}^= | \mathbb{P}(X^= = R^= | I) \neq 0)| = \binom{|\mathcal{D}|}{l_1} \binom{|\mathcal{D}| - l_1}{l_2} \dots \binom{|\mathcal{D}| - (l_1 + l_2 + \dots + l_{k-1})}{l_k}$$

und

$$|(R^< \in \mathcal{R}^< | \mathbb{P}(X^< = R^< | I) \neq 0)| = k! \cdot |(R^= \in \mathcal{R}^= | \mathbb{P}(X^= = R^= | I) \neq 0)|.$$

Unter Weiterbestehen der Gleichverteilungsannahme ergibt sich die A-posteriori-Entropie direkt aus der Mächtigkeit der obigen beiden Mengen.

Beispiel. Für Alice wurden 20 personenbezogene Daten erhoben, für Peter 10. Damit sinkt die Anzahl der möglichen Relationen $\mathcal{R}^=$ auf $|(R^= \in \mathcal{R}^= | \mathbb{P}(X^= = R^= | I) \neq 0)| = \binom{30}{20} =$

$\binom{30}{10} = 30045015$. Folglich ist der resultierende Grad der Unverkettbarkeit $\Delta(X^{\equiv}, I) = \frac{\log_2 30045015}{\log_2 2^{29}} \approx \frac{24,8406}{29} \approx 0,8566$.

Die Anzahl der verbleibenden Identifikationsrelation ist $|(R^{\leq} \in \mathcal{R}^{\leq} \mid \mathbb{P}(X^{\leq} = R^{\leq} \mid I) \neq 0)| = 21 \cdot 30045015 = 60090030$. Entsprechend ist der resultierende Grad der Unverkettbarkeit $\Delta(X^{\leq}, I) = \frac{\log_2 60090030}{\log_2 2^{30}} \approx \frac{25,8406}{30} \approx 0,8614$.

2.3.2 Bestimmung der Wahrscheinlichkeitsverteilungen von X^{\triangleright} und X^{∇}

Unter der Annahme, dass die Weitergaben unterschiedlicher personenbezogener Daten voneinander unabhängig sind (Annahme 6), kann die Wahrscheinlichkeit $\mathbb{P}(X^{\triangleright} = R^{\triangleright})$ für eine Kandidatenrelation $R^{\triangleright} \in \mathcal{R}^{\triangleright}$ aus den Wahrscheinlichkeiten für die Teilrelationen je Datum $\mathbb{P}(X_d^{\triangleright} = R_d^{\triangleright})$ mit $R_d^{\triangleright} \subseteq \mathcal{S} \times \mathcal{S}$ berechnet werden. Es gilt $\mathbb{P}(X^{\triangleright} = R^{\triangleright}) = \prod_{d \in \mathcal{D}} \mathbb{P}(X_d^{\triangleright} = R_d^{\triangleright})$.

Das Wissen eines Angreifers wird zunächst dadurch charakterisiert, inwiefern ihm die Kategorie des personenbezogenen Datums bekannt ist. Die Kategorie des Datums bestimmt dessen Herkunft und Verarbeitung gemäß Verfahrensverzeichnis. Jedem Datum ist seine Datenkategorie über die Funktion $\vartheta : \mathcal{D} \rightarrow \Theta$ zugewiesen.

$$\mathbb{P}(X_d^{\triangleright} = R_d^{\triangleright}) = \sum_{\theta \in \Theta} \mathbb{P}(X_d^{\triangleright} = R_d^{\triangleright} \mid X_\theta = \theta) \mathbb{P}(X_\theta = \theta)$$

Beispiel. Die Zuordnung zwischen Daten und Datenkategorien ist einem Systemangreifer für diejenigen Daten bekannt, die er selbst verarbeitet. So ist \mathcal{A}_{s_3} bekannt, dass $\vartheta(d_{16}) = \theta_{14}$ ist. Für alle anderen Daten sowie grds. für den zentralen Angreifer muss entsprechend der Maximum-Likelihood-Methode die Gleichverteilung angenommen werden.

Das Wissen eines Angreifers wird außerdem dadurch charakterisiert, inwiefern ihm die Herkunft der personenbezogenen Daten bekannt ist. Die Herkunft der personenbezogenen Daten ist von der Kategorie der Daten abhängig. Die Wahrscheinlichkeiten der Zufallsvariable X_σ für bestimmte Startsysteme abhängig von der Datenkategorie lautet $\mathbb{P}(X_\sigma = s \mid X_\theta = \theta)$. Und es ergibt sich:

$$\mathbb{P}(X_d^{\triangleright} = R_d^{\triangleright} \mid X_\theta = \theta) = \sum_{s \in \mathcal{S}} \mathbb{P}(X_d^{\triangleright} = R_d^{\triangleright} \mid X_\sigma = s, X_\theta = \theta) \mathbb{P}(X_\sigma = s \mid X_\theta = \theta)$$

Beispiel. Alle personenbezogenen Daten bis auf jene der Kategorien Rechnung und Empfehlung werden ausschließlich beim Betroffenen selbst erhoben. Für diese ist das Herkunftssystem s_1 bekannt. Eine Rechnung wird immer im Onlineshopsystem des Vertriebs erzeugt. Das Herkunftssystem für die Rechnung ist damit aus Sicht der Angreifer eindeutig s_7 . Über eine Empfehlung ist hingegen nur bekannt, dass sie von einem Kunden kommen muss. Der Startvektor für die Empfehlung ist mit der Wahrscheinlichkeit $\frac{1}{2} s_1$ und mit der selben Wahrscheinlichkeit s_2 .

Anmerkung Jede zweistellige⁷ Relation R über endlichen Mengen kann als binäre bzw. boolesche Matrix $R^{\square} = (r_{ij})$, $r_{ij} \in \{0, 1\}$ dargestellt werden. Die Einträge der Matrix r_{ij} stehen für die Realisationen der wie folgt definierten Zufallsvariablen X_{ij} :

$$X_{ij} : \mathcal{R} \rightarrow \{0, 1\}$$

$$R \mapsto \begin{cases} 1 & (e_i, e_j) \in R \\ 0 & \text{sonst.} \end{cases}$$

Der Eintrag bzw. das Ereignis 1 bedeutet, dass die die durch den Index gegebenen Elemente in Relation zueinander stehen, der Eintrag 0, dass keine Beziehung vorliegt. Im Folgenden wird deshalb zur Vereinfachung nicht zwischen der Relation R und ihrer Matrixdarstellung R^{\square} differenziert.

$$r_{ij} = 1 \Leftrightarrow X_{ij} = 1 \Leftrightarrow (e_i, e_j) \in R$$

Daraus abgeleitet wird folgende Kurzschreibweise verwendet:

$$R_{r_{i_1, j_1}, r_{i_2, j_2}, \dots, r_{i_k, j_k}} : \Leftrightarrow R = \{e_{i_1}, e_{j_1}\} \cup \{e_{i_2}, e_{j_2}\} \cup \dots \cup \{e_{i_k}, e_{j_k}\}$$

Bezüglich der konkreten Datenflüsse stützten sich die Angreifer auf die Angaben des Verfahrensverzeichnis. Dieses hinterlegt für alle Daten die vorgesehenen Verarbeitungsprozesse. Das Wissen der Angreifer wird als Matrix der bedingten Flusswahrscheinlichkeiten W modelliert. Der Eintrag w_{ij} gibt die Wahrscheinlichkeit an, mit der ein Fluss von s_i nach s_j , angenommen wird, unter der Bedingung, dass das Datum bereits in s_i verarbeitet wurde:

$$W_{\theta, s} : \{1, \dots, m\} \times \{1, \dots, m\} \rightarrow [0, 1]$$

$$(i, j) \mapsto w_{ij} = \mathbb{P}(X_{dij}^{\triangleright} = 1 \mid X_{dii}^{\triangleright} = 1, X_{\sigma} = s, X_{\theta} = \theta)$$

mit $m = |\mathcal{S}|$. Implizit ist $w_{ii} = 1$. Der reflexive Fluss, gleichbedeutend mit der Speicherung und Verarbeitung im System ($\forall d, i : \mathbb{P}(r_{dii}^{\triangleright}) = \mathbb{P}(r_{dii}^{\nabla})$), ist vollständig durch die eingehenden Flüsse erklärt:

$$\mathbb{P}(X_{dii}^{\triangleright} = 1 \mid \exists j \in \{1, \dots, i-1, i+1, \dots, m\} : X_{dji}^{\triangleright} = 1, X_{\sigma} = s, X_{\theta} = \theta) = 1$$

$$\mathbb{P}(X_{dii}^{\triangleright} = 1 \mid \forall j \in \{1, \dots, i-1, i+1, \dots, m\} : X_{dji}^{\triangleright} = 0, X_{\sigma} = s, X_{\theta} = \theta) = 0$$

Für alle Datenflussrelationen gilt $\mathbb{P}(X_{dij}^{\triangleright} = 0 \mid X_{dii}^{\triangleright} = 0, X_{\sigma} = s, X_{\theta} = \theta) = 1$ und $\mathbb{P}(X_{dij}^{\triangleright} = 1 \mid X_{dii}^{\triangleright} = 0, X_{\sigma} = s, X_{\theta} = \theta) = 0$. Es kann keine ausgehenden Flüsse geben, falls es keinen eingehenden Fluss gibt. Damit ist der Wahrscheinlichkeitsbaum für die Datenflussrelation vollständig erklärt.

Die in den bedingten Flusswahrscheinlichkeiten zum Ausdruck kommenden Pfade ergeben sich aus dem Verfahrensverzeichnis. Das Erfahrungswissen des Angreifers lässt sich in zwei zentralen Parametern ausdrücken. Zunächst hängt die Wahrscheinlichkeit von Flüssen in linearen Verfahren maßgeblich von der *Fortschrittsquote* $\varpi \in (0, 5; 1]$ eines

⁷ Gilt grundsätzlich auch für mehrstellige Relationen.

Prozesses ab. Dieser Parameter wird überall dort in der Flussmatrix eingesetzt, wo ein Fluss einem Prozessschritt entspricht. Unvorhergesehene Abweichungen vom Verfahren werden durch eine Fehlerwahrscheinlichkeit $\varepsilon \in [0; 0,5)$ beschrieben. Mit dieser Fehlerwahrscheinlichkeit finden Flüsse zu und zwischen Systemen außerhalb des vorgesehenen Prozessablaufs statt.

Beispiel. Die AdBokis Buchclub GmbH hat folgende Verfahren etabliert: Registrierung, Bestellung, Zahlungsabwicklung, Kundendatenarchivierung, Missbrauchsbekämpfung und Kundenservice. Die Fortschrittsquote wird mit 90% ($\varpi = 0,9$) und die Fehlerwahrscheinlichkeit mit $\varepsilon = 0,02$ angenommen. Auf die Darstellung der einzelnen Prozesse muss aus Platzgründen verzichtet werden.

Auf dieser Grundlage kann die A-priori-Wahrscheinlichkeit für die einzelnen Kandidatenrelationen iterativ berechnet werden. Die Komplexität der vollständigen Berechnung der Wahrscheinlichkeiten aller möglichen Kandidatenrelationen ist allerdings in $O(2^{|\mathcal{S}| \cdot |\mathcal{S}|})$. Selbst bei wenigen Systemen ist somit die Berechenbarkeit der Wahrscheinlichkeitsverteilung nicht mehr gegeben. Deshalb ist nur eine heuristische Lösung entsprechend dem in Abschnitt 2.4 beschriebenen Verfahren möglich.

Beispiel. Für obige Flussmatrix ergibt sich die Wahrscheinlichkeit, dass das Datum nur im Herkunftssystem s_1 verarbeitet wird, sofern es von der Datenkategorie θ_{14} ist als

$$\mathbb{P}(X_d^\triangleright = R_{d,r_{d11}^\triangleright=1}^\triangleright \mid X_\sigma = s_1, X_\theta = \theta_{14}) = 1 \cdot 0,1 \cdot 0,98^9 \approx 0,0834$$

Würde man kombinatorisch aus den Wahrscheinlichkeiten $\mathbb{P}(X_d^\triangleright = R_d^\triangleright)$ die Gesamtwahrscheinlichkeit $\mathbb{P}(X^\triangleright = R^\triangleright) = \prod_{d \in \mathcal{D}} \mathbb{P}(X_d^\triangleright = R_d^\triangleright)$ berechnen, hätte dies eine Komplexität in $O(|\mathcal{D}|^{|\mathcal{S}| \cdot |\mathcal{S}|})$. Erfreulicherweise gilt für unabhängige Teilsysteme (Teilrelationen), dass die Entropie eine additive Größe ist ($H(X^\triangleright) = H(X_{d_1}^\triangleright) + \dots + H(X_{d_n}^\triangleright)$ mit $n = |\mathcal{D}|$). Somit lässt sich die Gesamtwahrscheinlichkeit aus den approximierten Teilwahrscheinlichkeiten bestimmen.

Die Wahrscheinlichkeitsverteilung für X^∇ lässt sich auf Grundlage der Wahrscheinlichkeitsverteilung von X^\triangleright ermitteln:

$$\mathbb{P}(X^\nabla = R^\nabla) = \sum_{R^\triangleright \in \mathcal{R}^\triangleright \mid R^\triangleright \equiv_\nabla R^\nabla} \mathbb{P}(X^\triangleright = R^\triangleright)$$

mit $R^\triangleright \equiv_\nabla R^\nabla \Leftrightarrow$

$$\forall d \in \mathcal{D}, s \in \mathcal{S} : ((d, s, s) \in R^\triangleright \wedge (d, s) \in R^\nabla) \vee ((d, s, s) \notin R^\triangleright \wedge (d, s) \notin R^\nabla)$$

Zu diesem allgemeinen Hintergrundwissen kommen noch die jeweiligen Beobachtungen I der Angreifer hinzu. Ein Systemangreifer kann die Datenflüsse durch sein System überwachen. Die Likelihood $\mathbb{P}(I \mid R^\triangleright)$ ist für solche Beobachtungen sicher 1 oder 0. Die A-posteriori-Wahrscheinlichkeit beträgt

$$\mathbb{P}(R^\triangleright \mid I) = \frac{\mathbb{P}(I \mid R^\triangleright) \mathbb{P}(R^\triangleright)}{\mathbb{P}(I)} = \frac{\mathbb{P}(I \mid R^\triangleright) \mathbb{P}(R^\triangleright)}{\sum_{R^{\triangleright'} \in \mathcal{R}^\triangleright} \mathbb{P}(I \mid R^{\triangleright'}) \mathbb{P}(R^{\triangleright'})}$$

und damit entweder 0 oder $\frac{\mathbb{P}(R^{\triangleright})}{\sum_{R^{\triangleright} \in \mathcal{R}^{\triangleright}} \mathbb{P}(I, R^{\triangleright} I)}$. Es findet also eine Normierung auf die Summe der Wahrscheinlichkeiten der Relationen, die die Beobachtung des Angreifers zulassen, statt.

Der zentrale Angreifer kann ohne das Datenschutzauskunftssystem keine Beobachtungen machen, sondern muss sich vollständig auf das Hintergrundwissen auf Grundlage des Verfahrensverzeichnis verlassen. Er ist jedoch der einzige Angreifer, der mit Hilfe des Datenschutzauskunftssystems weitere Beobachtungen I' machen kann. Für \mathcal{A}^S ist $\Delta(X^{\nabla}, I') = \Delta(X^{\triangleright}, I') = 1$. Bei der Registrierung neu erhobener personenbezogener Daten im zentralen Verzeichnis lernt der zentrale Angreifer die Quelle der personenbezogenen Daten und den Ort der ersten Verarbeitung im Unternehmen kennen (Likelihood von 1 oder 0). Als Ergebnis lassen sich jeweils die A-posteriori-Wahrscheinlichkeiten und der abgeleitete Grad der Unverkettbarkeit für X^{\triangleright} und X^{∇} nach dem im Abschnitt 2.4 beschriebenen Verfahren bestimmen.

2.4 Implementierung

Wie bereits im vorherigen Abschnitt erwähnt, ist die Wahrscheinlichkeitsverteilung für die Relation R_d^{\triangleright} auch schon bei wenigen Systemen, Daten und Datentypen nicht mit akzeptablem Aufwand an Zeit und Speicher vollständig berechenbar. Allerdings ist eine approximative Lösung möglich. Sind die Wahrscheinlichkeitsmatrizen nur spärlich mit Fortschrittswahrscheinlichkeiten belegt, ballt sich die Wahrscheinlichkeitsmasse bei denjenigen Kandidatenrelationen, die einen Fluss entlang des Verarbeitungsprozesses vorsehen. Kandidatenrelationen, die kaum Flüsse im Verarbeitungsprozess vorsehen, bilden den „Long Tail“ der Verteilung. Ihr Gewicht bei der Berechnung der Entropie ist gering. Diesen Umstand kann man sich bei der Berechnung des *belief* aus den Wahrscheinlichkeitsmatrizen zunutze machen, indem man systematisch zuerst die wahrscheinlicheren Kandidaten in die Berechnung aufnimmt und den „Long Tail“ nur bis zu einem gegebenen *Schwellwert* erschließt. Die Wahrscheinlichkeiten ergeben sich aus dem Wahrscheinlichkeitsbaum. Durch Tiefensuche in diesem Baum kann die Entropie, ausgehend vom Startsystem, approximativ erschlossen werden.

Beispiel. *Tabelle 2 enthält die berechneten Ergebnisse für unterschiedliche Schwellwerte. Dem Betroffenen könnte ein Mindestgrad an Unverkettbarkeit von 0,94 bzw. 0,90 garantiert werden.*

3 Zusammenfassung und Ausblick

In diesem Beitrag wurde eine informationstheoretische Metrik für Unverkettbarkeit entworfen und für ein Datenschutzauskunftssystem instanziiert. Die Modellierung des Angreiferwissens wurde anhand eines Beispiels eingeführt. Die Metrik beruht auf definierten rechtlichen Anforderungen und passt sich in die Datenschutz-Schutzziele ein. Sie erlaubt, den Trade-off zwischen Transparenz und Unverkettbarkeit deutlich zu machen.

Schwellwert	$H(X^{\triangleright})$	$H(X^{\triangleright}, I')$	$\Delta(X^{\triangleright}, I')$	$H(X^{\nabla})$	$H(X^{\nabla}, I')$	$\Delta(X^{\nabla}, I')$
10^{-1}	125,8201	114,9513	0,8927	96,2104	85,8917	0,9136
10^{-2}	139,5569	128,8865	0,9235	103,6033	92,4655	0,8925
10^{-3}	149,4219	139,3821	0,9328	107,1627	96,1635	0,8974
10^{-4}	156,1207	147,2097	0,9429	109,1815	98,5819	0,9029
10^{-5}	157,3259	148,4618	0,9436	109,4951	98,8992	0,9032

Tab. 2: Approximierte Werte für die Entropie und den Grad der Unverkettbarkeit der Relationen R^{\triangleright} und R^{∇} für unterschiedliche Schwellwerte

Im Rahmen einer Nutzerevaluation zu unserem Datenschutzauskunftssystem [BK16] haben wir auch die Einstellung von Betroffenen zur Metrik abgefragt, die in Form von Zustandsbalken in das Frontend des Systems eingebunden war. Von unseren 31 Teilnehmern hatten 21 das Konzept verstanden, 7 wahren sich aufgrund der gegebenen Kurzbeschreibung nicht sicher. Von diesen 21 hielten nur 6 die Metrik nicht für Hilfreich, für 9 war es eine nützliche Entscheidungsgrundlage für ihre Opt-out-Möglichkeit. Insgesamt ein zufriedenstellendes Ergebnis für eine bis dato unbekannte Entscheidungshilfe.

Literaturverzeichnis

- [An15] Angulo, Julio; Fischer-Hübner, Simone; Pulls, Tobias; Wästlund, Erik: Usable Transparency with the Data Track: A Tool for Visualizing Data Disclosures. In: Proc. of the 33rd ACM Conference on Human Factors in Computing Systems. S. 1803–1808, 2015.
- [BK16] Bier, Christoph; Kühne, Kay: PrivacyInsight: The Next Generation Privacy Dashboard. In: Proc. of the Annual Privacy Forum (APF 2016). 2016.
- [BP11] Bohli, Jens-Matthias; Pashalidis, Andreas: Relations among privacy notions. ACM Transactions on Information and System Security, 14(1):1–24, 2011.
- [HJR15] Hansen, Marit; Jensen, Meiko; Rost, Martin: Protection Goals for Privacy Engineering. In: Proc. of the 1st Int. Workshop on Privacy Engineering. S. 159–166, 2015.
- [Pa08] Pashalidis, Andreas: Measuring the Effectiveness and the Fairness of Relation Hiding Systems. In: Proc. of the Asia-Pacific Services Computing Conf. S. 1387–1394, 2008.
- [PH10] A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management. http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf, v0.34.
- [RB11] Rost, Martin; Bock, Kirsten: Privacy By Design und die Neuen Schutzziele. Datenschutz und Datensicherheit, 35(1):30–35, 2011.
- [SD03] Serjantov, Andrei; Danezis, George: Towards an Information Theoretic Metric for Anonymity. In: Proc. 2nd Int. Conf. on Privacy Enhancing Technologies. S. 41–53, 2003.
- [Si11] Simitis, Spiros, Hrsg. BDSG. Nomos, Baden-Baden, 7. Auflage, 2011.
- [SK03] Steinbrecher, Sandra; Köpsell, Stefan: Modelling Unlinkability. In: Proc. of the 3rd Int. Workshop on Privacy Enhancing Technologies (PET 2003). Springer, S. 32–47, 2003.

Eingeschränkte Selbstbestimmung im Onlineverkehr: Stärkung der Einwilligungserklärung durch Einführung vorformulierter Datenschutzbestimmungen

Maren Pollmann¹, Dennis-Kenji Kipker²

Abstract: Die informationelle Selbstbestimmung wird im Online-Zeitalter zunehmend herausgefordert, indem eine immer weitergehende Verlagerung der Datenverarbeitung auf private Akteure stattfindet. Insbesondere die datenschutzrechtliche Einwilligung droht an Effektivität zu verlieren, da Datenschutzbestimmungen nicht nur hochkomplex ausgestaltet, sondern länger und umfassender denn je sind. Unter diesen Gesichtspunkten werden zunächst verschiedene aktuell diskutierte rechtliche Ansätze vorgestellt, die der Autonomie über die eigenen Daten Rechnung tragen sollen. Im Anschluss an eine Bewertung der unterschiedlichen Vorschläge folgt die Feststellung, dass keine dieser Maßnahmen für sich genommen ausreicht, um den Persönlichkeitsschutz im Internet zu gewährleisten. Vorgeschlagen wird deshalb ein System vorformulierter Datenschutzbestimmungen, welches über die bloße datenschutzrechtliche Klauselkontrolle in AGB hinausgeht.

Keywords: Einwilligung, informationelle Selbstbestimmung, AGB-Kontrolle, Datenschutzbestimmungen

1 Einführung und Problemaufriss

Die hohe Innovationsgeschwindigkeit im Onlinebereich sowie neue Möglichkeiten zur Auswertung von Datensätzen, die in Sekundenschnelle Verknüpfungen zwischen unterschiedlichen Informationen herstellen können, stellen das geltende Datenschutzrecht und damit auch das Recht auf informationelle Selbstbestimmung mehr denn je auf die Probe. Insbesondere Privatunternehmen sammeln und verarbeiten personenbezogene Daten aus jeglichen Lebensbereichen mithilfe einer Vielzahl von internetfähigen Alltagsgeräten. Zudem wird schon lange über Mängel im Vollzug des Datenschutzrechts geklagt.³ Umso dringender ist es daher, den Datenschutz an die neuen Gegebenheiten anzupassen und ihn derart zu stärken, dass eine hinreichend effektive Wahrnehmung der informationellen Selbstbestimmung erreicht werden kann. Dieser Beitrag stellt einige der neuen Problemlagen heraus und präsentiert anschließend weitere

¹ Universität Bremen, Institut für Informations-, Gesundheits- und Medizinrecht (IGMR), Universitätsallee GW1, 28359 Bremen, m.pollmann@uni-bremen.de.

² Universität Bremen, Institut für Informations-, Gesundheits- und Medizinrecht (IGMR), Universitätsallee GW1, 28359 Bremen, kipker@uni-bremen.de.

³ Siehe dazu Ritter/Schwichtenberg, VuR 2016, 95 (96); Spiecker gen. Döhmann, in: Leible/Kutschke, Der Schutz der Persönlichkeit, S. 33.

Vorschläge und Ideen zum Schutz der informationellen Selbstbestimmung, speziell bezogen auf die datenschutzrechtliche Einwilligung.

2 Herausforderungen der informationellen Selbstbestimmung im Online-Zeitalter

2.1 Verlagerung auf private Akteure

Das Grundrecht auf informationelle Selbstbestimmung blickt bereits auf eine über 30-jährige Geschichte zurück. Als ein Recht, dessen Existenz und Schutzgewährleistungen im Wesentlichen auf dem technologischen Fortschritt im Bereich der automatisierten Datenverarbeitung begründet sind, unterliegt es einem steten Wandel. Insbesondere die modernen Online-Technologien fordern die individuelle Verwirklichung informationeller Selbstbestimmung stetig heraus, da die Geschäftsmodelle zahlreicher im Internet agierender Unternehmen auf der wirtschaftlichen Nutzung personenbezogener Daten basieren, so beispielsweise im Bereich kontextgesteuerter Werbeanbieter. Die Gefährdungslage für personenbezogene Daten, die sich aus dem Handeln solcher nicht-staatlichen Akteure ergibt, ist nicht mit der Entstehungsgeschichte und ursprünglichen Schutzrichtung des Grundrechts auf informationelle Selbstbestimmung identisch. Während es 1983 für den Fall des Volkszählungsgesetzes noch allein um den Schutz vor einer die Persönlichkeit über die Maße beeinträchtigenden staatlichen Datenverarbeitung ging⁴ und diese durch das klassische Subordinationsverhältnis geprägt war, bei dem die Freiheit und die Rechte des Bürgers den Bindungen und rechtfertigungsbedürftigen Befugnissen des Staates gegenüberstehen,⁵ spielt sich das Verhältnis zwischen Bürger und Unternehmen vielmehr auf einer Ebene der Gleichordnung ab. Trotz dieser zunächst formalen Gleichheit kann sich ein in der Intensität vergleichbares Gefährdungspotenzial für die informationelle Selbstbestimmung ergeben, das auf der „manchmal bestehenden Asymmetrie der Kräfteverhältnisse“⁶ beruht. Es besteht also ein Machtgefälle zwischen den beiden Akteuren, das zuvorderst nicht durch das Einräumen von Abwehrrechten, sondern im Wesentlichen nur durch die Erfüllung staatlicher Schutzpflichten ausgeglichen werden kann.⁷ Obwohl der Schutz der informationellen Selbstbestimmung ein hochrangiges Gut darstellt, ist er aber nicht frei von Beschränkungen. Jeder Einzelne ist als Bestandteil der Gesellschaft im Regelfall in diese eingebunden – und das auch in Bezug auf die Verarbeitung seiner personenbezogenen Daten.⁸ Dies ist ebenso zu berücksichtigen, wenn es um die Informationsnutzung durch Unternehmen geht, sodass

⁴ Gesetz über eine Volkszählung, Berufszählung, Wohnungszählung und Arbeitsstättenzählung (Volkszählungsgesetz 1983) vom 25. März 1982 (BGBl. I S. 369) - VZG 1983, BVerfGE 65, 1.

⁵ Grimm, JZ 2013, 585 (587).

⁶ Grimm, JZ 2013, 585 (587).

⁷ So auch Grimm, JZ 2013, 585 (587).

⁸ Vgl. BVerfGE 65, 1 (40).

im Rahmen der hier stattfindenden Interessenabwägung nicht nur das Grundrecht auf informationelle Selbstbestimmung, sondern auch die Entscheidung zu wirtschaftlicher Aktivität und damit die Unternehmensfreiheit sowie die Allgemeine Handlungsfreiheit (Artt. 12 Abs. 1, 2 Abs. 1 GG) als gegenläufige Interessen einbezogen werden müssen und nicht von vornherein geringer als die informationelle Selbstbestimmung bewertet werden dürfen.⁹ Soweit deshalb Lösungsansätze zum Schutz einer zukünftigen, zeitgemäßen informationellen Selbstbestimmung entwickelt werden, müssen sich diese auch an den technologisch möglichen und wirtschaftlich aufgegriffenen Nutzungsmöglichkeiten personenbezogener Daten orientieren.

2.2 Begrenzte Effektivität der datenschutzrechtlichen Einwilligung

Die Zulässigkeit einer Datenverarbeitung sowohl im öffentlichen als auch im nicht-öffentlichen Bereich wird durch das Verbotsprinzip mit seinen gesetzlichen Erlaubnistatbeständen bestimmt. Die Verarbeitung personenbezogener Daten ist danach grundsätzlich verboten. Lediglich eine individuelle Einwilligung des Betroffenen oder ein gesetzlicher Erlaubnistatbestand können dem Datenverarbeiter als Legitimationsgrundlage dienen.¹⁰ Auf diese Weise setzt das Verbotsprinzip die bereits vom Bundesverfassungsgericht geforderte Entscheidungsfreiheit des Einzelnen, über die Preisgabe und Verwendung seiner persönlichen Daten zu bestimmen, um.¹¹

Die Erteilung einer Einwilligung durch den von der Datenverarbeitung Betroffenen stellt als stärkster Ausdruck der informationellen Selbstbestimmung und der individuellen Entfaltungsfreiheit hierbei den zentralen Legitimationstatbestand dar. Dennoch unterliegt die Anwendung des Verbotsprinzips insbesondere im Verhältnis zwischen Privaten häufiger Kritik.¹² Infolge dieser wird zum Teil eine striktere Trennung der Regelungen der öffentlichen Datenverarbeitung durch den Staat als unmittelbaren Grundrechtsadressaten und der Datenverarbeitung zwischen Privaten, die nur mittelbar an die Grundrechte gebunden sind, gefordert.¹³ Aufgrund der hohen Datennutzungserfordernisse vieler Dienstleistungsunternehmen ist das Institut der datenschutzrechtlichen Einwilligung für den Online-Verkehr aber auch gleichsam essenziell, denn sie ermöglicht für die verarbeitende Stelle eine flexible Alternative zu den engen gesetzlichen Ausnahmetatbeständen. Ihre Wirksamkeit hängt jedoch von der Einhaltung gesetzlicher Voraussetzungen ab. Insbesondere die notwendige Informiertheit der Einwilligung sowie das Erfordernis einer bewussten und eindeutigen Erklärungshandlung stellen dieses Instrument regelmäßig auf die Probe. Dies verdeutlicht unter anderem ein Blick auf das vielzitierte Beispiel der Datenschutzbestimmungen des sozialen Netzwerks Facebook: Durch die darin enthaltenen Verweise auf die Klauseln integrierter Dienste und Anwendungen steigt die

⁹ Bull, NJW 2006, 1617 (1622).

¹⁰ Scholz/Sokol, in: Simitis, BDSG, § 4, Rn. 2.

¹¹ BVerfGE 65, 1 (42).

¹¹ S. unter 3.3.

¹³ Masing, NJW 2012, 2305 (2307).

Menge an Information für den Nutzer erheblich an.¹⁴ Dies hat zur Folge, dass der Einzelne angesichts der Vielzahl an datenschutzrechtlich relevanten Vorgängen faktisch nicht in der Lage ist, sich umfassend über die Verarbeitung seiner personenbezogenen Daten zu informieren. Der ohnehin schon häufige Fall, dass ein Betroffener für die Leistung eines bestimmten Onlinedienstes im Gegenzug seine personenbezogenen Daten in einem für ihn nicht nachvollziehbaren Umfang preisgeben muss, ohne die Reichweite seiner Einwilligung absehen zu können, darf durch derlei Anbieter nicht zum Regelfall werden. Ein solches Ergebnis widerspräche nicht nur dem Grundgedanken der informationellen Selbstbestimmung, sondern würde das Instrument der Einwilligung letzten Endes ad absurdum führen.

Des Weiteren bewirkt Big Data und die damit verbundene unbegrenzte Vernetzung von personenbezogenen Informationen, dass eine einmal gegebene Einwilligung für eine Vielzahl zum Teil unüberschaubarer Verarbeitungsvorgänge als Legitimationsgrundlage herangezogen werden kann. Das Verbotsprinzip kann dieser Entwicklung teilweise entgegentreten und den Einzelnen zumindest vor einem totalen Kontrollverlust über die Verwendung seiner persönlichen Daten schützen. Ein Festhalten an diesem Grundprinzip ist daher mehr als geboten. Doch angesichts der Defizite, die vor allem der Prozess der Einwilligungserteilung in der Praxis mit sich bringt, bedarf es neuer Regelungsinstrumente, welche die veränderten Rahmenbedingungen berücksichtigen und somit die Einwilligung zu einem wirksamen Instrument des Betroffenen machen, das seine schwächere Position gegenüber dem Online-Dienstleister auszugleichen vermag.

3 Bisher diskutierter Erkenntnisstand zum Schutz der informationellen Selbstbestimmung

Um dem Recht auf informationelle Selbstbestimmung auch im Zeitalter des unbegrenzten globalen Datenverkehrs noch zu hinreichender Wirksamkeit zu verhelfen, werden bereits seit mehreren Jahren unterschiedliche rechtliche Lösungsansätze diskutiert, die für sich betrachtet jedoch selten zu vollends befriedigenden Ergebnissen gelangen.

3.1 Eigentumsähnliche Ausgestaltung

Einige Vorschläge zur Verbesserung der informationellen Selbstbestimmung zielen darauf ab, das Recht am eigenen Datum zu einem eigentumsähnlichen Recht umzufunktionieren. Hierdurch soll Immaterialgüterschutz gewährt¹⁵ oder aber ein eigenes Datenwirtschaftsrecht¹⁶ geschaffen werden. Auch wenn es zunächst sinnvoll

¹⁴ Buchner, DuD 2015, 402 (404).

¹⁵ Kilian, in: Garstka/Coy, Wovon - für wen - wozu, S. 195 (211).

¹⁶ Reiners, ZD 2015, 51; Seidel, ZG 2014, 153 (158)

scheinen mag, das Recht an der Nutzung personenbezogener Daten vorrangig zu monetarisieren, so besteht das Risiko, dass durch eine solche Verwirtschaftlichung des Persönlichkeitsschutzes dem informationellen Selbstbestimmungsrecht seine ideelle Basis entzogen wird. Obgleich personenbezogene Daten als Grundlage für die Geschäftsmodelle einiger großer Internet-Unternehmen wie beispielsweise Google oder Facebook dienen und ihnen die Eigenschaft als Ware somit nicht vollständig aberkannt werden kann, ist dies nicht ihr einziges oder gar ausschlaggebendes Merkmal. Vor allem die ursprüngliche Herleitung des Rechts auf informationelle Selbstbestimmung aus dem Menschenwürdegrundsatz darf nicht völlig außer Acht gelassen werden. Eine Ablehnung der eigentumsähnlichen Ansätze schließt jedoch nicht die Möglichkeit aus, dem Betroffenen durch zukünftige Anpassungen seines Rechts auf informationelle Selbstbestimmung auch eine gewisse wirtschaftliche Teilhabe an dem Umgang mit den personenbezogenen Daten einzuräumen.

3.2 Ausgestaltung als objektiv-rechtliche Teilhabeordnung

Andere Ansätze sehen vor, das Recht auf informationelle Selbstbestimmung in Abkehr vom subjektiv-rechtlichen Gehalt als objektiv-rechtliche Informationsordnung auszugestalten. Der einem solchen Verständnis entsprechende Grundrechtsgehalt setzt sich aus objektiven Wertentscheidungen zusammen und tritt in der Dogmatik neben die subjektiv-rechtliche Abwehrfunktion der Grundrechte.¹⁷ Gemäß der objektiv-rechtlichen Teilhabeordnung muss der Staat die Grundrechte in seinen Entscheidungen berücksichtigen, sodass unterschiedlich ausgeprägte Pflichten zur Ausgestaltung der Rechtsordnung hergeleitet werden können.¹⁸ Ihn kann beispielsweise eine staatliche Schutzpflicht treffen, der zufolge er der informationellen Selbstbestimmung auch zwischen Privaten zur Geltung verhelfen muss.¹⁹ Eine weitere Ausprägung stellt der Grundrechtsschutz durch Organisation und Verfahren dar, welcher bereits im Volkszählungsurteil durch das Bundesverfassungsgericht als flankierende Maßnahme gefordert wurde.²⁰ Auf ebendiese objektiv-rechtliche Funktion der informationellen Selbstbestimmung stützen sich eine Reihe unterschiedlicher Ansätze zur Modernisierung des Grundrechts. Es soll dabei als ein Teilhaberecht ausgestaltet werden, das sich auf die Funktion des Datenschutzrechts als Kommunikationsordnung konzentriert und so dem Einzelnen ein Recht auf kommunikative Selbstbestimmung zuspricht.²¹ Gerade für den elektronischen Privatrechtsverkehr geben die Ansätze zur objektiv-rechtlichen Gestaltung des informationellen Selbstbestimmungsrechts wichtige Impulse: So kann die Stärkung der Teilhabe des Einzelnen an datenschutzrechtlich relevanten Vorgängen einen entscheidenden Beitrag leisten, um dem Machtungleichgewicht zwischen global

¹⁷ Bechler, Informationseingriffe durch intransparenten Umgang mit personenbezogenen Daten, S. 76.

¹⁸ Bechler, Informationseingriffe durch intransparenten Umgang mit personenbezogenen Daten, S. 77.

¹⁹ Britz, in: Hoffmann-Riem, Offene Rechtswissenschaft, S. 561 (585 f.).

²⁰ BVerfGE 65, 1 (58 f.).

²¹ Schoch, in: Sachs/Siekmann, Der grundrechtsgeprägte Verfassungsstaat, S. 1491 (1499); Simitis, in: Simitis, BDSG, § 1, Rn. 36 f.

agierenden Unternehmen und Individuen entgegenzutreten. Zudem sind die von Hoffmann-Riem herausgestellten Möglichkeiten für einen effektiven Selbstschutz durch die Nutzer von großer Bedeutung. Hierzu ist neben der Förderung von sozialen und technischen Kompetenzen auch die Aufklärung durch Verbraucherverbände vorgesehen.²² Die steigende Komplexität informationstechnischer Systeme erhöht jedoch auch die Herausforderungen, die mit der Realisierung des individuellen Selbstschutzes verbunden sind. So wird es auch für den technisch sachverständigen Bürger zunehmend schwieriger, die Datenverarbeitungsvorgänge zu überblicken und sein Handeln auf die Maximierung des Persönlichkeitsschutzes hin auszurichten.²³ Aus diesem Grund ist das Abstellen auf den individuellen Selbstschutz zwar notwendig, kann jedoch nicht die subjektive Dimension des informationellen Selbstbestimmungsrechts entbehrlich machen. Die Vorschläge, allein auf den objektiv-rechtlichen Gehalt des Grundrechts abzustellen, haben eine einseitige Konzeptionierung des Selbstbestimmungsrechts zur Folge, welche die Grundrechtsdogmatik nicht in ausreichendem Maße berücksichtigt.

3.3 Abkehr vom Verbotsprinzip

Weiters wird zur Novellierung des Rechts auf informationelle Selbstbestimmung die Abkehr vom bisher maßgebenden Verbotsprinzip der Datenverarbeitung diskutiert. Hierin wird teilweise dennoch eine Einschränkung der ebenfalls grundrechtlich geschützten Kommunikationsfreiheit gesehen, deren verfassungskonforme Wahrnehmung nur bei einer weiten Auslegung seiner Ausnahmen noch möglich sei.²⁴ Ebenso bewirke die weite Auslegung der Personenbezogenheit von Daten, dass ein Großteil der Kommunikation im Internet unter das Datenschutzrecht falle, was ein Durchsetzungshemmnis für die informationelle Selbstbestimmung zur Folge hat.²⁵ Mit dem Verbotsprinzip sei zudem die staatliche Schutzpflicht nicht erfüllt, da sie neben dem Persönlichkeitsschutz auch die Sicherstellung der größtmöglichen Entfaltungsfreiheit des Einzelnen umfasse.²⁶ Aller Kritik am Verbotsprinzip zum Trotz hätte jedoch der Einzelne ohne eine solche Lösung kaum mehr Möglichkeiten zu wissen, wer wann welche Daten von ihm erhebt und auf welche Weise verarbeitet. Um einem zunehmenden Kontrollverlust über die Verwertung der eigenen Daten entgegenzutreten, kann das Verbotsprinzip deshalb auch in Zukunft nicht entfallen. Dass ein Festhalten an diesem Prinzip geboten ist, zeigt neben der ausdrücklichen Normierung in Art. 8 Abs. 2 GR-Charta auch der europäische Reformprozess zur Datenschutzgrundverordnung, bei dem das Verbotsprinzip weder in einem der Entwürfe noch im vorausgehenden parlamentarischen Diskurs ausdrücklich infrage gestellt wurde.²⁷

²² Hoffmann-Riem, AöR 1998, 513 (536 f.).

²³ Bäcker, Der Staat 2012, 91 (112).

²⁴ Schneider, AnwBl. 2011, 233 (234); Härting, NJW 2013, 2065 (2067).

²⁵ Härting, in: Leible/Kutschke, Der Schutz der Persönlichkeit, S. 55 (58); Schneider, AnwBl. 2011, 233 (233).

²⁶ Giesen, CR 2014, 550 (552).

²⁷ Vgl. dazu jeweils Art. 6 der Entwürfe zur Datenschutzgrundverordnung von der Europäischen Kommission vom 25. Januar 2012 (KOM(2012) 11 endg., des Europäischen Parlaments vom 12. März 2014 und des Rats

3.4 Verknüpfung von Datenschutz und Wettbewerbsrecht

Einen entscheidenden Schritt zur Stärkung der informationellen Selbstbestimmung könnte eine engere Verzahnung zwischen dem Datenschutz- und dem Wettbewerbsrecht darstellen. Hier gilt es, bestehende Regelungen des UWG ausdrücklich für das Datenschutzrecht zu öffnen. Ein Rückgriff auf das Wettbewerbsrecht kann zum Persönlichkeitsschutz im Rahmen der Datenverarbeitung beitragen, indem es den Betroffenen eine Marktteilnehmerposition einräumt und hierdurch der kaum zu vermeidenden Entwicklung hin zu einer Kommerzialisierung personenbezogener Daten Rechnung trägt. Das Wettbewerbsrecht knüpft somit an den zunehmenden Vermögensaspekt von Informationen über Einzelpersonen an. Eine Verbindung der beiden Rechtsgebiete ermöglicht zudem, dass sich neben den Betroffenen auch Mitbewerber gegen unlautere Verhaltensweisen anderer datenverarbeitender Marktteilnehmer zur Wehr zu setzen können. Diese Option bietet das geltende Datenschutzrecht nicht, da es sich in der individuellen und behördlichen Rechtsdurchsetzung erschöpft.²⁸

3.5 Neuausrichtung einzelner Elemente des Datenschutzrechts

Neben Ansätzen, die eine fundamentale Umstrukturierung des Datenschutzrechts fordern, wird auch vorgeschlagen, mithilfe einzelner Anpassungen der bestehenden Grundsätze eine Stärkung der informationellen Selbstbestimmung zu erreichen. Insbesondere Big Data und die damit verbundene unbegrenzte Vernetzung von personenbezogenen Informationen stellen das Institut der datenschutzrechtlichen Einwilligung vor eine Herausforderung, da diese grundsätzlich zeitlich nicht begrenzt ist und somit für eine Vielzahl zum Teil unüberschaubarer Verarbeitungsvorgänge als Legitimationsgrundlage herangezogen werden kann. Daher wird teilweise gefordert, die Einwilligung zeitlich zu befristen.²⁹ Eine solche zeitliche Begrenzung der Einwilligung scheint jedoch wenig praktikabel, da es nur schwer möglich ist zu kontrollieren, ob das jeweilige Unternehmen nach Ablauf der Zeit eine neue Einwilligung von dem Betroffenen als Grundlage für die weitere Datenverarbeitung einholt.

Nichtsdestotrotz stellen aber über die Limitierung der Einwilligung hinaus Anpassungen im Hinblick auf die Klagebefugnis von Verbänden sowie der Transparenz der Datenverarbeitung vorteilhafte Ergänzungen im Rahmen einer Neuausrichtung des Datenschutzrechts dar.³⁰ Jedoch muss in diesem Zusammenhang auch bedacht werden, dass ein Ausbau an Informationspflichten nicht generell zu einer besser informierten

der Europäischen Union vom 15. Juni 2015, 9565/15; so auch Albrecht, CR 2016, 88 (91).

²⁸ Zech, WRP 2013, 1434 (1434).

²⁹ Spindler, GRUR-Beilage 2014, 101 (103).

³⁰ Zur Neuregelung des Verbandsklagerechts bei Datenschutzverstößen: Spindler, ZD 2016, 144; Ritter/Schwichtenberg, VuR 2016, 95; zur Forderung von mehr Transparenz unter anderem: Bechler, Informationseingriffe durch intransparenten Umgang mit personenbezogenen Daten, S. 194.

Einwilligung durch den Betroffenen führt, sondern auch in einen alltagsuntauglichen Formalismus münden kann. In den nächsten Jahren wird ferner davon auszugehen sein, dass das Urteil des EuGH zu Safe Harbor insbesondere die transnationale datenschutzrechtliche Debatte um die Informationsverarbeitung bei Privaten weiter prägen wird.³¹

3.6 Zwischenergebnis

Die verschiedenen Reformansätze zur Modernisierung des Rechts auf informationelle Selbstbestimmung im Internetzeitalter verdeutlichen die allgemeine Auffassung, dass der Schutz personenbezogener Daten in seiner derzeitigen rechtlichen Ausgestaltung in nicht mehr ausreichendem Maße den aktuellen technischen Anforderungen genügt. Insbesondere die veränderten Gegebenheiten für die automatisierte Datenverarbeitung im Internet und die dort angebotenen Dienste lassen sich nur schwer mit den geltenden nationalen und europäischen Vorschriften in einer Art und Weise vereinbaren, welche die Interessen aller Beteiligten in einen angemessenen Ausgleich bringt.

Bei den zur Lösung dieser Konfliktlage vorgestellten Ansätzen zeigen sich im Einzelnen Praktikabilitätsschwächen. Im Volkszählungsurteil leitete das Bundesverfassungsgericht das Recht des Einzelnen, „grundsätzlich selbst zu entscheiden, wann und innerhalb welcher Grenzen persönliche Lebenssachverhalte offenbart werden“,³² aus dem Allgemeinen Persönlichkeitsrecht ab. Dies geschah zu einer Zeit, in der im Regelfall weder zur Erledigung von Einkäufen noch zum Pflegen sozialer Kontakte, geschweige denn zum Auffinden von Informationen, personenbezogene Daten an Dritte preisgegeben werden mussten. Mittlerweile jedoch hat ein beträchtlicher Teil der persönlichen Lebensgestaltung eine Ausdehnung in den digitalen Raum erfahren. Diese Entwicklung hat es mit sich gebracht, dass sich personenbezogene Daten auch zum Wirtschaftsgut entwickelten und nunmehr neben ihrem ideellen persönlichkeitsrechtlichen Ursprung auch eine kommerzielle Komponente besitzen. Ein Datenschutzrecht, das zukunftsorientiert ausgestaltet ist, muss diesen beiden – nicht selten gegenläufigen – Interessen zwischen Persönlichkeitsschutz und Wirtschaftlichkeit in der Informationsverarbeitung gerecht werden. Es darf sich einerseits in seinem Wertgehalt nicht ausschließlich durch seine persönlichkeitsrechtliche Herkunft definieren, diese andererseits aber auch nicht über die Maße negieren, indem es das Recht auf informationelle Selbstbestimmung zu einem reinen Wirtschaftsrecht reduziert. Deshalb ist es zur Entwicklung eines praktikablen Datenschutzmodells notwendig, die verschiedenen Lösungsvorschläge, die für den zukünftigen Schutz der informationellen Selbstbestimmung diskutiert werden, miteinander zu kombinieren und hierdurch die unterschiedlichen Interessenlagen in Ausgleich zu bringen.

³¹ Siehe EuGH Urt. v. 6.10.2015 – C-362/14, EuZW 2015, 881 – Schrems.

³² BVerfGE 65, 1 (43); BVerfGE 63, 131 (142 f.).

4 Vorschlag zur Stärkung der informationellen Selbstbestimmung bei der Online-Datenverarbeitung: die vorformulierten Datenschutzbestimmungen

4.1 Unzulänglichkeit der AGB-Kontrolle von Datenschutzbestimmungen

Aufgrund der fehlenden Gestaltungsmöglichkeit durch den Betroffenen unterstellt die Rechtsprechung datenschutzrechtliche Einwilligungserklärungen regelmäßig der AGB-Kontrolle nach §§ 305 ff. BGB.³³ Dadurch wird der schwächeren Vertragspartei ermöglicht, auch bei umfangreichen AGB darauf zu vertrauen, dass diese keine überraschenden oder in besonderem Maße benachteiligenden, versteckten Klauseln enthalten. Dennoch trifft den Betroffenen bei einer Klage gegen die verarbeitende Stelle insoweit ein hohes Prozessrisiko, als dass, insbesondere bei den relativen Klauselverboten und der Generalklausel, das Ergebnis der Inhaltskontrolle nur schwer vorauszusehen ist. Zudem entfaltet ein etwaiges Urteil keine Wirkung gegenüber Dritten, sodass in einem zweiten Verfahren in Bezug auf dieselben AGB der Ausgang grundsätzlich offen ist. Nicht zuletzt auch angesichts der ohnehin schon sehr geringen Anreize für den Betroffenen, gegen Datenschutzverstöße vorzugehen,³⁴ stellt die AGB-Kontrolle im Bereich der Datenschutzbestimmungen kein zufriedenstellendes Instrument dar. Es bedarf somit spezieller Regularien, die eine effektive Kontrolle durch den Betroffenen schon im Vorfeld, das heißt bei Erteilung seiner Einwilligung, ermöglichen. Im Bereich der Datenschutzbestimmungen sollte aufgrund der vorgestellten Spezifika sowie des persönlichkeitsrechtlichen Schutzzinhalts die AGB-Kontrolle deshalb einem System vorformulierter Datenschutzbestimmungen weichen. Hierdurch kann eine transparentere Einwilligungssituation geschaffen und eine schnellere sowie praktikablere Kontrolle der Datenschutzbedingungen durch den Betroffenen selbst ermöglicht werden, ohne dass erst die Gerichte im Nachgang an die Einwilligungserteilung der informationellen Selbstbestimmung zu ihrer Wirksamkeit verhelfen müssen

4.2 Die Verpflichtung zur Nutzung vorformulierter Datenschutzbestimmungen

Die Schnellebigkeit in der Onlinewelt hat zur Folge, dass regelmäßig eine Vielzahl von Verträgen im Internet abgeschlossen wird, die in der Regel mit einer Preisgabe personenbezogener Daten einhergehen. Da es für den Einzelnen zeitlich oft unmöglich ist, alle Datenschutzbestimmungen zu lesen und zu bewerten,³⁵ bedarf es Instrumente, die es dem Betroffenen erleichtern, ihre Rechtskonformität selbst zu überprüfen und auf

³³ Siehe u.a. BGH GRUR 2008, 1010 (1011); LG Berlin; NJW 2013, 2605 (2606).

³⁴ Spiecker gen. Döhmman, in: Hain/Pfeifer (Hrsg.), Datenschutz im digitalen Zeitalter, S. 61 (76).

³⁵ Siehe hierzu auch Bolsinger, DuD 2016, 382 (385).

diese Weise die Ausübung seines Rechts auf informationelle Selbstbestimmung unterstützen.³⁶

Die Verwendung vorformulierter Datenschutzklauseln setzt zunächst voraus, dass auf gesetzgeberischer Seite ein Katalog von Klauseln erstellt wird, die im Einklang mit nationalen und europäischen Datenschutzvorgaben stehen und die von einer verarbeitenden Stelle zur Erstellung ihrer Datenschutzbestimmungen verwendet werden. Bei den vorformulierten Klauseln handelt es sich folglich um hinreichend detaillierte positivrechtliche Vorgaben schon bei Einwilligungserteilung, die als Substitut einer nachträglichen Verbotskontrolle durch die AGB-rechtlichen Vorschriften fungieren sollen.

Aufgrund der positiven Formulierung solcher Klauseln wird es dem Betroffenen erleichtert, die Schwere eines Eingriffs in sein Recht auf informationelle Selbstbestimmung eigenständig zu überprüfen und abzuschätzen. Selbst wenn die datenverarbeitende Stelle sich bei Abruf der Einwilligung nicht an die Klauselvorgaben halten sollte, ist das Prozessrisiko des Betroffenen im Vergleich zu einer gerichtlichen Überprüfung relativer Klauselverbote oder gar nur einer Generalklausel im Rahmen von AGB deutlich geringer, da nur eine formelle Überprüfung auf die Einhaltung des Klauselkataloges hin stattfinden muss. Die Schwelle für den Einzelnen, im Zweifelsfall eine gerichtliche Überprüfung von Datenschutzbestimmungen anzustreben, wird auf diese Weise deutlich gesenkt, sodass der Vorschlag ebenfalls einen Schritt zur Bekämpfung des Vollzugsdefizits darstellt. Die Verpflichtung zur Verwendung vorformulierter Datenschutzbestimmungen hat für den Betroffenen zudem den Vorteil, dass er sich mit den Datenschutzbestimmungen für einzelne Geschäftsbereiche nur einmal auseinandersetzen muss, da für alle Dienste dieselben Klauseln gelten.

Vorformulierte Klauseln bieten zudem die Möglichkeit, Datenschutzbestimmungen übersichtlich zu gestalten, indem sie ausschweifende Formulierungen der datenverarbeitenden Stelle verhindern. Deshalb sind sie ein wichtiger Beitrag zu dem Ziel, die Verarbeitung personenbezogener Daten für den Betroffenen transparenter zu gestalten. Insoweit stellen die vorformulierten Datenschutzbestimmungen auch eine Konkretisierung des allgemeinen Transparenzgebotes nach § 4a BDSG dar, an dessen Einhaltung es in der Praxis häufig mangelt, weil die Datenschutzbestimmungen letztlich frei vom Datenverarbeiter festgelegt werden.³⁷ Ebenso sind die fest vorgegebenen Klauseln dazu geeignet, weitere datenschutzrechtliche Grundsätze wie das Zweckbindungsprinzip oder den Bestimmtheitsgrundsatz konkret und für den Einwilligenden sowohl verständlich wie auch kontextbezogen wiederzugeben.

Das soziale Netzwerk Instagram hält sich beispielsweise in seinen Datenschutzrichtlinien die Möglichkeit vor, nach der Schließung oder Deaktivierung des

³⁶ Neben Ansätzen, die – wie in diesem Beitrag vorgeschlagen – den Inhalt der Einwilligungserklärung betreffen, sind auch Vorschläge, die dem Betroffenen eine visuelle Unterstützung bieten, vorstellbar, siehe dazu Pollmann/Kipker, DuD 2016, 378.

³⁷ Gola/Klug/Körffer, in: Gola/Schomerus, BDSG, § 4a Einwilligung, Rn. 23.

Kontos Informationen und Nutzerinhalte „für einen kaufmännisch angemessenen Zeitraum für Backup-, Archivierungs- bzw. Prüfzwecke“ aufzubewahren.³⁸ Für einen Privatnutzer ist aus dieser Klausel nicht erkennbar, mit welcher auch nur ungefähren Speicherdauer er zu rechnen hat und was konkret unter Archivierungs- und Prüfzwecken verstanden werden kann. Verstöße gegen das Transparenzgebot und den Bestimmtheitsgrundsatz liegen deshalb nahe. Letzterer verlangt, dass sich die Einwilligung auf einen genau umschriebenen Verwendungsvorgang bezieht, pauschal gehaltene Erklärungen bzw. nur allgemeine Angaben können deshalb nicht ausreichend sein.³⁹ Ein Verstoß gegen das Transparenzgebot ergibt sich ferner aus Art. 13 Abs. 2 der neuen EU-Datenschutzgrundverordnung, wonach die Angabe der Speicherdauer oder Kriterien für deren Festlegung eine notwendige Information darstellt, um eine transparente Datenverarbeitung zu gewährleisten. Eine vorformulierte Datenschutzklausel könnte unter diesen Gesichtspunkten eine maximale Speicherdauer festlegen oder eine Auswahl konkreter Kriterien zur Festlegung der Dauer bestimmen. Daneben ist es möglich, die Verwendungszwecke von Anfang an hinreichend konkret zu fassen, indem die für die Archivierung und Prüfung notwendigen Verarbeitungsvorgänge weiter aufgeschlüsselt werden.

Die vorformulierten Datenschutzbestimmungen bieten ferner den Vorteil einer leichten Überprüfbarkeit der Rechtslage, denn durch den Ausschluss individueller Klauseln des einzelnen Datenverarbeiters wird dem Betroffenen eine Sicherheit dahingehend gegeben, dass er von der Unwirksamkeit solcher Datenschutzbestimmungen ausgehen kann, die nicht mit den vorformulierten Klauseln konform sind. In diesem Zusammenhang bietet sich eine technisch verhältnismäßig leicht implementierbare Überprüfungsmöglichkeit über einen automatisierten Textabgleich der verwendeten Klauseln mit den offiziellen, von staatlicher Seite aus bestimmten Vorgaben für ein bestimmtes Geschäftsmodell an, die beispielsweise auf behördlichen Websites zum Download zur Verfügung stehen. Eine Aufspaltung der Klauselvorgaben nach unterschiedlichen Geschäftszwecken ist auch deshalb sinnvoll, weil in jeder datenverarbeitenden Branche mit verschiedenen Verarbeitungsszenarien zu rechnen ist. So ist es vorstellbar, unterschiedliche Bestimmungen zum Beispiel für soziale Netzwerke, E-Commerce-Unternehmen, Informationsdienste und Online Games zu schaffen. Insbesondere wird auch davon auszugehen sein, dass sich die spezifischen Anforderungen an eine Datenerhebung und -verarbeitung innerhalb einer bestimmten Branche bzw. eines Geschäftsmodells nur in geringfügigem Maße unterscheiden, sodass kein Bedarf besteht, die vorformulierten Datenschutzbestimmungen um individuelle Klauseln zu ergänzen.⁴⁰ Systematisch sind die Klauselvorgaben in der Form eines Baukastens zu fassen, aus welchem sich der Datenverwender entsprechend seinen Anforderungen die jeweiligen Bestimmungen

³⁸ <https://de-de.facebook.com/help/instagram/155833707900388/>.

³⁹ Simitis, in: Simitis, BDSG, § 4a, Rn. 77 ff.

⁴⁰ Beispielsweise zeigt ein Vergleich der Online-Games-Anbieter Europe Entertainment Ltd. und upjers GmbH, dass die Datenverwendungszwecke beider Anbieter grundsätzlich übereinstimmen; sie bestehen im Wesentlichen in der Bereitstellung und Gewährleistung des Spielbetriebs, der Zahlungsabwicklung, von Kundenserviceleistungen und in der Datenverarbeitung zu Marketingzwecken; siehe <https://www.stargames.net/de/allgemeine-geschaeftsbedingungen#privacy>, <https://de.upjers.com/privacy>.

auswählen und zu seinen individuellen Datenschutzbestimmungen zusammenstellen kann. Dadurch wird insbesondere auch solchen Anbietern Rechnung getragen, die eine Kombination verschiedener Dienste anbieten und sich keinem Geschäftsmodell alleinig zuordnen lassen. Freilich muss der Anbieter in überprüfbarer Weise kenntlich machen, innerhalb welcher Branchen er tätig ist, um eine Verwendung der vorformulierten Klauseln über das benötigte Maß hinaus zu vermeiden. Denkbar wäre hier eine Registrierung der Geschäftskategorien bei den Datenschutzbehörden. Diesen ohnehin schon überlasteten Einrichtungen wird es durch systematisch kategorisierte wie vorformulierte Klauseln zudem erleichtert, Bestimmungen einzelner Anbieter im Zweifelsfall – beispielsweise nach Anrufung durch einen Nutzer – zu überprüfen und zu bewerten.

Sämtliche Einwilligungserteilungen im Rahmen der vorformulierten Datenschutzbestimmungen haben dennotwendigerweise als Opt-in zu erfolgen. Dies allein schon deshalb, weil die Europäische Datenschutzgrundverordnung fußend auf Art. 4 Nr. 11 dem Opt-out-Modell jegliche gesetzliche Grundlage entzogen hat.

5 Fazit und Ausblick

Die informationelle Selbstbestimmung steht im Onlinezeitalter vor erheblichen Herausforderungen. Damit sie im Informationsalltag nicht leerläuft, bedarf es neuer Regelungsmechanismen. Dazu werden verschiedene Konzepte vertreten. Die Einführung vorformulierter Datenschutzbestimmungen stellt einen praktisch ausgerichteten Lösungsansatz dar. Indem durch ihre Verwendung die maßgeblichen, aber für den täglichen Gebrauch zu abstrakt gefassten Grundsätze des Datenschutzes in eine konkrete Form gegossen werden, können die Klauseln ein wesentliches Instrument sein, um die datenschutzrechtliche Einwilligung zukunftsfest zu machen.

Literaturverzeichnis

- [Bä12] Bäcker, Matthias: Grundrechtlicher Informationsschutz gegen Private. *Der Staat* 51/12, S. 91-116, 2012.
- [Be10] Bechler, Lars: Informationseingriffe durch intransparenten Umgang mit personenbezogenen Daten. Universitätsverlag Halle-Wittenberg, Halle (Saale) 2010.
- [Bo16] Bolsinger, Harald: Wo bleibt die digitale Dividende für Europas Konsumenten?. *Datenschutz und Datensicherheit* 06/16, S. 382-385.
- [Br10] Britz, Gabriele: Informationelle Selbstbestimmung zwischen rechtswissenschaftlicher Grundsatzkritik und Beharren des Bundesverfassungsgerichts. In (Hoffmann-Riem): *Offene Rechtswissenschaft*, S. 561-596, Mohr Siebeck Verlag, Tübingen 2010.
- [Bu15] Buchner, Benedikt: Message to Facebook. *Datenschutz und Datensicherheit* 06/15, S. 402-405, 2015.

- [Bu14] Buchner, Benedikt: Facebook zwischen BDSG und UWG. In (Alexander/Bornkamm/Buchner/Fritz, Hrsg.): Festschrift für Helmut Köhler zum 70. Geburtstag, S. 51-62, C.H. Beck Verlag, München 2014.
- [Bu06] Bull, Hans Peter: Zweifelsfragen um den informationellen Selbstbestimmung-Datenschutz als Datenaskese?. Neue Juristische Wochenschrift 23/06, S. 1617-1624, 2006.
- [GKK15] Gola, Peter; Schomerus, Rudolf: Bundesdatenschutzgesetz, Kommentar, 12. Auflage, C.H. Beck Verlag, München 2015.
- [Gi14] Giesen, Thomas: Für ein verfassungsgemäßes Datenschutzrecht in Europa. Computer und Recht 08/14, S. 550-556, 2014.
- [Gr13] Grimm, Dieter: Der Datenschutz vor einer Neuorientierung. JuristenZeitung 12/13, S. 585-592, 2013.
- [Hä12] Härting, Niko: Datenschutz und Persönlichkeitsrechte: Verbotsprinzip und offener Tatbestand. In (Leible/Kutschke, Hrsg.): Der Schutz der Persönlichkeit im Internet, S. 55-64, Boorberg Verlag, Stuttgart 2013.
- [Hä13] Härting, Niko: Anonymität und Pseudonymität im Datenschutzrecht. Neue Juristische Wochenschrift 29/13, S. 2065-2071, 2013.
- [Ho98] Hoffmann-Riem, Wolfgang: Informationelle Selbstbestimmung in der Informationsgesellschaft. Archiv des öffentlichen Rechts Band 123/98, S. 513-540, 1998.
- [Ki14] Kilian, Wolfgang: Strukturwandel der Privatheit. In (Garstka/Coy, Hrsg.): Wovon-für-wen-wozu-Systemdenken wider die Diktatur der Daten, S. 195-224, Berlin, 2014.
- [Ma12] Masing, Johannes: Herausforderungen des Datenschutzes. Neue Juristische Wochenschrift 32/2012, S. 2305-2311, 2012.
- [Pl13] Plath, Kai-Uwe: Kommentar zum BDSG sowie den datenschutzrechtlichen Regelungen des TMG und des TKG. Otto Schmidt Verlag, Köln 2013.
- [PK16] Pollmann, Maren; Kipker, Dennis-Kenji: Informierte Einwilligung in der Online-Welt. Datenschutz und Datensicherheit 06/16, S. 378-381, 2016.
- [Re15] Reiners, Wilfried: Datenschutz in der Personal Data Economy-Eine Chance für Europa. Zeitschrift für Datenschutz 02/15, S. 51-55, 2015.
- [RS16] Ritter, Franziska; Schwichtenberg, Simon: Die Reform des UKlaG zur Eliminierung des datenschutzrechtlichen Vollzugsdefizits – neuer Weg, neue Chancen?. Verbraucher und Recht 03/16, S. 95-102, 2016.
- [Sc11] Schneider, Jochen: Hemmnis für einen modernen Datenschutz: Das Verbotsprinzip. AnwaltsBlatt 04/11, S. 233-239, 2011.
- [Sc12] Schoch, Friedrich: Das Recht auf informationelle Selbstbestimmung in der Informationsgesellschaft. In (Sachs/Siekmann, Hrsg.): Der grundrechtsgeprägte Verfassungsstaat – Festschrift für Klaus Stern zum 80. Geburtstag, S. 1491-1512, Duncker & Humblot, Berlin 2012.
- [Se14] Seidel, Ulrich: Das Grundrecht auf Datensouveränität. Zeitschrift für Gesetzgebung 29/14, S. 153-165, 2014.

- [Si14] Simitis, Spiros: Kommentar zum Bundesdatenschutzgesetz. 8. Auflage, Nomos-Verlag, Frankfurt a.M. 2014.
- [Sp13] Spiecker gen. Döhmann: Die Durchsetzung datenschutzrechtlicher Mindestanforderungen bei Facebook und anderen Sozialen Netzwerken – Überlegungen zu Vollzugsdefiziten im Datenschutzrecht. In (Leible/Kutschke, Hrsg.): Der Schutz der Persönlichkeit im Internet, S. 33-54, Boorberg, Stuttgart 2013.
- [Sp15] Spiecker gen. Döhmann: Zur Architektur des europäischen und deutschen Datenschutzes im Zeitalter von Vorratsdatenspeicherung, Big Data und IT-Enhancement im Lichte der *Google Spain*-Entscheidung des *Europäischen Gerichtshofs*. In (Hain, K.-E./Pfeifer, N. Hrsg.): Datenschutz im digitalen Zeitalter – global, europäisch, national, S. 61-91, C.H. Beck Verlag, München 2015.
- [Sp14] Spindler, Gerald: Datenschutz und Persönlichkeitsrechte im Internet. Gewerblicher Rechtsschutz und Urheberrecht – Beilage 01/14, S. 101-108, 2014.
- [Sp16] Spindler, Gerald: Verbandsklagen und Datenschutz – das neue Verbandsklagerecht. Zeitschrift für Datenschutz 03/16, S. 114-119, 2016.
- [Ze13] Zech, Herbert: Durchsetzung von Datenschutz mittels Wettbewerbsrecht?. Wettbewerb in Recht und Praxis 11/13, 1434-1436, 2013.

Nutzerzentrierte Sicherheit – NzS 2016

Workshop: Nutzerzentrierte Sicherheit – NzS 2016

Andreas Heinemann¹ Steffen Lange² Marian Margraf³ Jörg Willomitzer⁴

Bei der Entwicklung von Sicherheitssoftware und -technologien steht primär die Umsetzung der klassischen Schutzziele Vertraulichkeit, Integrität und Verfügbarkeit im Vordergrund. Offenkundig hat die Benutzbarkeit und Verständlichkeit der zur Verfügung gestellten Sicherheitssoftware einen erheblichen Einfluss auf die Akzeptanz beim Anwender, damit auf deren fehlerfreie Benutzung und folglich auf das letztlich erreichte Sicherheitsniveau.

Im Entwurfsprozess werden die beiden Aspekte Benutzbarkeit und Verständlichkeit – wenn überhaupt – üblicherweise erst nach der Präzisierung des Sicherheitsmodells und der zu realisierenden Sicherheitsmechanismen in Betracht gezogen. Da insbesondere die zu realisierenden Sicherheitsmechanismen weitestgehend festlegen, auf welche Art und Weise der Anwender mit der Sicherheitssoftware interagiert, führt diese Vorgehensweise häufig zu einer geringen Akzeptanz beim Anwender.

Seit dem Siegeszug der Smartphones, Tablets und Wearables mit ihren intuitiven, oft touch-basierten Benutzerschnittstellen fordern Anwender mehr und mehr eine hohe Benutzbarkeit der angebotenen Softwareprodukte. Den Bedürfnissen der Anwender müssen deshalb beim Entwurf und der Implementierung von Softwareprodukten, insbesondere also auch bei der Entwicklung von Sicherheitssoftware immer stärker Rechnung getragen werden.

Der Workshop hat das Ziel, neue Ansätze zur Entwicklung von benutzbarer Sicherheitssoftware sowie zur Evaluation der beiden Aspekte Benutzbarkeit und Verständlichkeit zu thematisieren und zu diskutieren.

Nach eingehender Begutachtung durch jeweils drei Mitglieder des Programmkomitees wurden unter allen Einreichungen vier Beiträge ausgewählt. Wir danken den folgenden Mitgliedern des Programmkomitees für Ihre Mitwirkung im Begutachtungsverfahren:

- Jana Dittmann (Universität Magdeburg)
- Paul Gerber (TU Darmstadt)
- Norman Hänsch (Universität Erlangen-Nürnberg)
- Stephan Klein (Governikus GmbH & Co. KG)

¹ Hochschule Darmstadt, Schöfferstr. 10, 64295 Darmstadt, andreas.heinemann@h-da.de

² Hochschule Darmstadt, Schöfferstr. 10, 64295 Darmstadt, steffen.lange@h-da.de

³ Freie Universität Berlin, Takustr. 9, 14195 Berlin, marian.margraf@fu-berlin.de

⁴ Freie Universität Berlin, Takustr. 9, 14195 Berlin, joerg.willomitzer@fu-berlin.de

- Oksana Kulyk (TU Darmstadt)
- Luigi Lo Iacono (TH Köln)
- Stephan Neumann (TU Darmstadt)
- Lena Reinfelder (Universität Erlangen-Nürnberg)
- Maxim Schnjakin (Bundesdruckerei GmbH)

Darmstadt und Berlin, im Juni 2016

Andreas Heinemann & Steffen Lange (Hochschule Darmstadt)

Marian Margraf & Jörg Willomitzer (FU Berlin)

(Workshop-Organisation)

Usability der Identity- und Accessmanagement- komponenten von Endkundenportalen

Peter Weierich¹, Tobias Schmidt² und Sebastian Abeck³

Abstract: Für Kundenportale vieler Branchen hat sich das Identitäts- und Zugangsmanagement (IAM) zu einer Schlüsseltechnologie entwickelt, die gleichzeitig Geschäftsprozesse absichern muss und die Nutzerakzeptanz des Gesamtportals maximieren soll, indem sie möglichst niedrige Einstiegsbarrieren schafft. Die hier vorgestellte Studie evaluiert die Usability von 112 deutschen Unternehmensportalen verschiedener Branchen bezüglich der verwendeten IAM-Technologien und Prozesse. Erwartungsgemäß schneiden Einzelhandelsportale am besten ab, die niedrigste Marktdurchdringung von Personalisierungskomponenten weist die Versicherungswirtschaft auf.

Keywords: Usability, Personalisierung, Consumer Identity und Access Management, Consumer-IAM, Branchenvergleich

1 Einleitung

Digitalisierung und Personalisierung sind aktuelle Trends: Über alle Branchen hinweg versuchen Unternehmen damit, die Kundenbindung zu erhöhen und neue Geschäftsmodelle zu entwickeln [JP14]. Kunden begrüßen dies und nutzen digital verbundene Produkte und Services, um ihre Bedürfnisse zu erfüllen. Dieses Vorgehen ist in einigen Branchen zum Standard geworden [Fo14].

Das Identitäts- und Zugangsmanagement, also das Verwalten digitaler Identitäten und ihrer Berechtigungen ist die Grundvoraussetzung zur Umsetzung beider Trends [TB13], wurde in der Vergangenheit jedoch sehr stark branchenspezifisch entwickelt.

Grundsätzlich kann gelten, dass das traditionelle von der IT Security getriebene Unternehmens-IAM den dynamischen Anforderungen von Kunden nicht mehr gerecht wird [Fo14]. Daher entwickelt sich gerade eine neue Disziplin, das marketing-getriebene „Consumer-IAM“ (C-IAM), dessen größte Herausforderung darin besteht, Sicherheitsanforderungen und einfache Bedienbarkeit zu vereinen [Ma16].

¹ iC Consult GmbH, Keltlenring 14, 82041 Oberhaching, peter.weierich@ic-consult.com

² Karlsruhe Institut für Technologie (KIT), Forschungsgruppe Cooperation & Management (C&M), Zirkel 2, 76131 Karlsruhe, uhcrb@student.kit.edu

³ Karlsruhe Institut für Technologie (KIT), Forschungsgruppe Cooperation & Management (C&M), Zirkel 2, 76131 Karlsruhe, abeck@kit.edu

2 Vorgehen in der Studie

In einem ersten Schritt wurde erhoben, wie IAM-Systeme in Kundenportalen zur Personalisierung heute eingesetzt werden und welche Geschäftsprozesse sie absichern.

Bereits der Einstieg über einen Login oder Registrierung ist kritisch für die Akzeptanz und damit den Geschäftserfolg eines Kundenportals: Sicherheitssysteme müssen so gestaltet sein, dass sie gleichzeitig sicher und benutzbar sind [Ba04]. Beispielsweise muss die Registrierung eines Interessenten so einfach und effizient sein, dass die Einstiegsbarriere so niedrig wie möglich liegt. Anmeldeprozesse über Passworteingabe, soziale Medien oder andere Authentisierungswege müssen möglichst einfach und unempfindlich gegen das Vergessen von Zugangsdaten sein. Trotzdem verlangt der Kunde, dass sowohl Datenschutz und Sicherheit seiner Daten gewährleistet und sensible Geschäftsprozesse ausreichend abgesichert sind. Es gilt das Prinzip des "unmotivierten Benutzers" [WT99], wonach Sicherheitsmechanismen kein primäres Ziel für den Benutzer darstellen. Daher wirkt sich eine schlechte Usability deutlich negativ aus: Der Kunde bricht die Interaktion ab und wandert u.U. zu einem anderen Anbieter ab.

In einem zweiten Schritt wurde evaluiert, wie reif die eingesetzten IAM-Systeme hinsichtlich ihrer Benutzerfreundlichkeit sind. Ein Fokus wird dabei auf die Nutzung von "bring your own identity" (BYOID), insbesondere auf den "Social Login" gerichtet. Dabei können Kunden z.B. ihren Facebook- oder Google+ Account für die Anmeldung in anderen Portalen nutzen. Selbst in Umgebungen mit hohen Sicherheitsanforderungen kann dieses Vorgehen praktikabel und von den Kunden gewünscht sein [Ca14]. Das gilt umso mehr, als immer mehr kommerziell genutzte Anwendungen (LinkedIn, Salesforce dafür passende Federation Services anbieten.

2.1 Auswahl von Branchen: Vorreiter und Nachzügler

Um eine gute empirische Basis zu haben, sollten mehr als 100 Unternehmensportale evaluiert werden, die den folgenden Branchen zugeordnet werden können:

- Banken als die Vorreiter, die Online-Banking Systeme schon vor über 30 Jahren eingeführt haben [Wi14].
- Versicherungen, bei denen das Schlagwort „Digitalisierung“ derzeit besonders hoch im Kurs steht und die ebenso wie Banken stark reguliert sind.
- E-Commerce ist aus verschiedenen Gründen die „Kernbranche“ für C-IAM.
- Automobilhersteller, weil sie als early adopters neuer Technologien gelten.

Für die Messung der Benutzerfreundlichkeit wurde ein analytisches Verfahren gewählt, das ausschließlich die objektiven Eigenschaften der Systeme betrachtet und damit auf aufwändige Usability Tests mit vielen Testpersonen verzichtet. Die Bewertung orientiert

sich an dem Vorgehen der "Identification and Comparison Study" des Landesentrums für Datenschutz in Schleswig-Holstein [La03].

Bedienbarkeit	
Effektivität	Sind alle Benutzerziele bezüglich IAM online erreichbar?
Bedienbarkeitsbarrierefreiheit	Ist die Bedienung des IAM-Systems frei von Bedienbarkeitsbarrieren?
Fehlertoleranz und Benutzerführung	Wird der Benutzer während der Bedienung entlastet sowie vor Fehlern bewahrt?
Umsetzung von Self-Service	Ist es dem Benutzer möglich, Änderungen seiner digitalen Identität selbst durchzuführen?
Zugänglichkeit	
Verständlichkeit von Eingaben	Ist dem Benutzer ersichtlich und verständlich, was wo einzugeben ist?
Unterstützung des Benutzers	Werden zusätzliche Informationen zum Vollenden einer Aufgabe bereitgestellt?
Reduzierung kognitiver Barrieren	Ist die Bedienung des IAM-Systems frei von kognitiven Barrieren?
Darstellung	Ist das IAM-System sinnvoll gestaltet?

Tab. 1: Dimensionen der Checkliste zur Beurteilung von IAM-Systemen [WA15]

Um die Vergleichbarkeit zu gewährleisten, wurde eine Persona entwickelt. Diese stellt einen ca. 40jährigen männlichen Benutzer dar, der über gängige Anwenderkenntnisse in der Bedienung von Webseiten und Computern verfügt. Für die Analyse wurde damit eine einheitliche Beobachtungsposition geschaffen. Die Informationssammlung fand in zwei Iterationen statt: Zuerst wurden in einem Cognitive Walkthrough ausschließlich die Funktionen und Anwendungsbereiche der IAM-Systeme erhoben. Grundlage dafür waren branchenspezifische Szenarien, die die Persona zu durchlaufen hatte. Anschließend wurde die Usability anhand einer detaillierten Prüfliste erhoben, die sich an Best Practices der Usability-Forschung orientiert (z.B. die Grundlagen der Dialoggestaltung ISO 9241-110). Diese Prüfliste umfasst insgesamt 50 Einzelkriterien für die Themen Funktionsvollständigkeit, Verständlichkeit, Fehlertoleranz und Barrierefreiheit bezüglich Bedienung und Kognition (s. Tab. 1 und 2). Diese wurden gewichtet und den übergeordneten Attributen Bedienbarkeit und Zugänglichkeit zugeordnet. Die Bedienbarkeit bewertet ausschließlich auf die Interaktion mit dem IAM-System und die Zugänglichkeit adressiert die Kognition während der Interaktion. Die Bewertung der „Wichtigkeit“ erfolgte dabei über ein Konsensverfahren in der Gruppe der vier an der Analyse beteiligten Studierenden.

Bedienbarkeit	Einzelkriterium	Gewichtung
Effektivität	Registrierung möglich?	10
	Vollständige Online-Registrierung möglich?	7,5
	Anmeldung möglich?	10
	Datenänderung möglich?	6,5
	Löschung möglich?	5,5
	Rücksetzung des Kennwortes möglich?	9
Zugänglichkeit		
Reduzierung kognitiver Barrieren	Verzicht auf Policies (Passwortaufbau, PW-Erneuerung)	4,5
	Sinnvolle Nutzung von gestufter Authentifizierung, je nach Benutzerziel	6
	Werden nur Informationen abgefragt die benötigt werden?	6
	Wird eine kurze Erläuterung zur Datennutzung gegeben, nicht nur auf Richtlinien verwiesen?	4,5
	Vollständige Erläuterung?	4,5
	Daten nur für genutzte Dienste verwendet?	6,5
	Anzeige ob Daten sicher übertragen werden?	5
	Föderation umgesetzt?	3,5
	Social Login umgesetzt?	2

Tab. 2: Ausschnitt der Checkliste zur Bewertung der Usability

3 Ergebnisse

3.1 Branchenvergleich

In Abb. 1 sind die aggregierten Werte der einzelnen Branchen dargestellt: Die Zusammenführung erfolgte nach folgendem Schema. Wenn mehr als 50 % der Portale einer Branche eine bestimmte Eigenschaft erfüllen, wird diese der Branche positiv angerechnet. Die normierten Maximalwerte liegen sowohl für die Bedienbarkeit, als auch für die Zugänglichkeit bei 50 %. Durch die Breite der Säulen wird gezeigt, wie verbreitet IAM-Systeme in der jeweiligen Branchen sind.

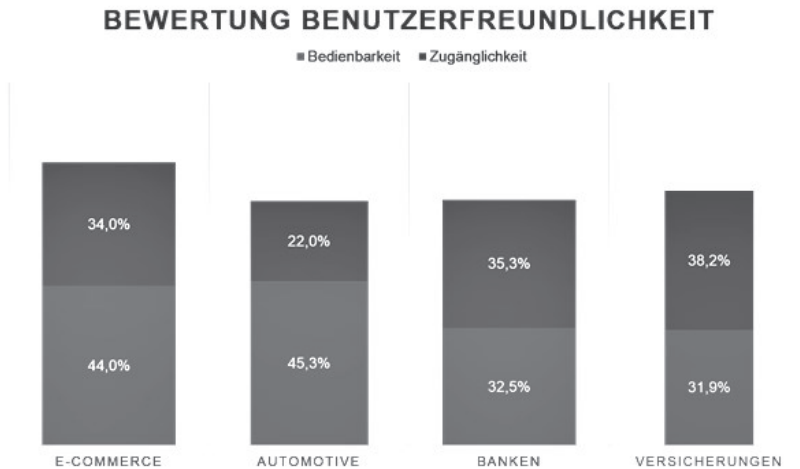


Abb. 1: Branchenvergleich der Benutzerfreundlichkeit: Erfüllungsgrade in Prozent; die Breite der Balken visualisiert die Marktdurchdringung von Personalisierungskomponenten.

3.2 Banken: Identitätsnachweis über das Internet ist die Ausnahme

Die untersuchten 20 Retail-Banken verfügen über mindestens ein IAM-System, das dem Kunden den Zugriff auf seine Konten erlaubt. Eine weitere Anwendung findet sich vor allem im Bereich der Direktbanken: Sie bietet Interessenten den Zugriff auf Tools wie Musterdepots oder Watchlists, die keine Kundenbeziehung voraussetzen und niedrigere Sicherheitsanforderungen aufweisen.

Für die Betrachtung der Funktionsvollständigkeit wurde angenommen, dass der Kunde selbst entscheiden möchte, "wann, wo und wie" er mit der Bank interagiert [Ba12a]. Das bedeutet, dass auf allen Kanälen der vollständige Funktionsumfang gewährleistet sein sollte (Registrierung, Anmeldung etc. im Self Service des Kunden).

Die Analyse hat ergeben, dass dies vor allem für Neukunden nicht zutrifft. Branchenweit war es bei der Ersterhebung nur bei der DKB möglich, Verträge – und damit einhergehende Registrierungen – vollständig über das Internet abzuschließen. Umgesetzt wird dies durch die Nutzung des elektronischen Bundespersonalausweises und über ein VideoIdent-Verfahren der Firma "WebID-Solutions GmbH". Dabei wird über eine web-basierte Videokonferenz das Personalausweisdokument geprüft. Das Angebot verspricht nach Aussage des Anbieters eine gesteigerte Konversionsrate, vor allem im Vergleich zum sonst gängigen PostIdent-Verfahren. Bestandskunden dagegen können neue Konten, Depots und andere Produkte buchen und in der Regel durch eine TAN legitimieren. Kritisch ist allerdings der reduzierte Funktionsumfang im Bereich des Mobile Banking: Dort können in der Regel nur Kontenübersichten erstellt und Standardtransaktionen ausgeführt werden.

Für die regelmäßige Benutzung sehen wir vor allem die kognitiven Barrieren als kritisch an: Der Kunde muss sich fast immer sowohl seine Zugangsnummer als auch eine PIN merken. Lediglich die Deutsche Bank bietet neuerdings im Mobile Banking die Möglichkeit, dass der Kunde sich über seinen Fingerabdruck authentisiert, was diesen sowohl kognitiv als auch in der Bedienung stark entlastet.

Ein Social Login könnte Benutzerfreundlichkeit verbessern. So führte die türkische Denizbank Social Login im Mobile Banking bereits 2012 ein. Zum Zeitpunkt der Durchführung der Studie war in Deutschland nur ein laufendes Projekt bekannt: Ein Anbieter mit einer jungen Zielgruppe, nutzte Social Login für ein Blog-Angebot.

3.3 Schlusslicht Versicherungen

Auf weniger als 50 % der untersuchten 47 Portale fand sich ein IAM-System für Kundenzugänge. Lediglich die Direktversicherer schneiden mit 72 % relativ gesehen besser ab, wodurch die klassischen Versicherer mit 40 % weit abgeschlagen sind. Das liegt im Wesentlichen daran, dass die klassischen Versicherungsvertreter weiterhin den „exklusiven“ Kundenzugang haben wollen und digitale Kanäle als Konkurrenz betrachten.

Kunden dagegen sehen das Internet als den wichtigsten Kanal zur Interaktion, für Vertragsabschlüsse, Vertragsverwaltungen und Schadensmeldungen [Ba12b]. Insofern bestehen in dieser Branche große Defizite zwischen Kundenerwartung und dem gelieferten Funktionsumfang: Nur wenige Anbieter offerieren die gewünschten Geschäftsprozesse, die zudem stark verbesserungswürdig umgesetzt sind. So müssen Online-Schadensmeldungen meist vollständig ausgefüllt werden und sind nicht über den Login an eine digitale Identität gekoppelt. Denn durch diese Verbindung würde der Kunde wesentlich entlastet werden, weil er weniger Daten eingeben müsste. Problematisch ist auch die Registrierung für den Online-Zugang. Dieser muss, wenn überhaupt vorhanden, in der Regel medienbruchbehaftet beantragt werden. Hier könnten die Bankenstandards genutzt werden, um Zugang mit Vertragsabschluss freizuschalten.

Für Interessenten bieten diese Portale fast keinen Mehrwert: Vereinzelt können diese einen Registrierungsprozess abschließen, um berechnete Angebote zu speichern. In vielen Fällen ist jedoch eine Berechnung von Versicherungskonditionen nicht möglich und es wird auf einen Kundenberater verwiesen.

Relative Vorreiter sind neben den Direktversicherern die Allianz und die Axa. Sowohl beim Funktionsumfang, als auch bei der Benutzerfreundlichkeit stehen sie an der Spitze. Trotzdem sind diese Portale nicht perfekt. Ähnlich der Bankenbranche existieren kognitive Barrieren durch die Nutzung des Standardlogins über Benutzer-Id und Passwort. Einige Anbieter nutzen den Bundespersonalausweis zur Authentifikation, welcher eine kognitive Entlastung darstellt – bzw. darstellen könnte, wenn er eine größere Verbreitung hätte.

Noch schlechter sieht es bei den mobilen Zugängen aus. Es werden zwar Apps ange-

boten, doch diesen fehlt häufig der Bezug zum Kerngeschäft: Ein Beispiel ist die weit verbreitete mobilen Applikation „Wettervorhersage“.

3.4 E-Commerce knapper Sieger

Für diese Studie wurden 25 E-Commerce Portale von Anbietern aus dem klassischen stationären Einzelhandel untersucht. Reine Online-Versandhändler wurden nicht betrachtet.

Eine Besonderheit hier ist, dass Einkäufe meistens ohne Benutzerregistrierung als „Gast“ abgeschlossen werden können. Es wird jedoch vom Kunden erwartet, dass er vollständige Angaben zu Anschrift und E-Mail-Adresse macht, die auch für eine „wiederverwendbare“ Registrierung erforderlich sind.

Der Mehrwert einer IAM-Nutzung besteht darin, dass der Kunde eine Übersicht über bisher getätigte Bestellungen bekommt und unterschiedliche Rechnungs- und Lieferadressen sowie Zahlungsmodalitäten für weitere Einkäufe gespeichert werden können. Zusätzlich können Wunschlisten u. ä. geführt werden.

Aber auch im E-Commerce werden nicht alle Standard-IAM-Geschäftsprozesse unterstützt. Beispielsweise ist es häufig nicht möglich, das Kundenkonto direkt im Portal zu löschen. In der Hälfte aller Fälle muss der Kunde eine Löschung via E-Mail oder Kontaktformular beantragen. Oft ist die Pflege von Stammdaten nur eingeschränkt möglich. Besonders auffällig sind Passworrichtlinien, die den Kunden zu stark einschränken und sogar behindern.

Generell ist die Benutzerfreundlichkeit von E-Commerce-Portalen unter den betrachteten Branchen am besten. Verbesserungspotenzial besteht, z.B. über Social Login. Dieser böte eine erhöhte Sicherheit bei Bestellungen von Gästen und die Komfortfunktion „automatisierte Datenübernahme“. Dafür kann der E-Commerce mit digitalen Gütern wie Spielen und Musik als Vorbild gelten: Dort ist Social Login weit verbreitet, um neue Kunden zu generieren und die Einstiegshürden abzusenken, ohne dass die Sicherheit darunter leidet. Als Positivbeispiel sind Mediamarkt und Saturn zu nennen, die ebenfalls mit digitalen Gütern handeln. Immerhin steigt die Nutzung des Social Logins spürbar: Während der Laufzeit der Studie (November 2014 – Januar 2015). In einigen Fällen wurde dieser während des Cognitive Walkthroughs noch nicht angeboten und stand bei der tieferen Analyse dann zur Verfügung.

3.5 Licht und Schatten in der Automobilindustrie

Von den 20 untersuchten Automobilherstellern haben zwar 19 einen geschützten Kundenbereich, dessen Nutzen häufig nicht ersichtlich ist: Kunden können zwar ihr Auto in einer virtuellen Garage parken, mehr als die Anzeige der passenden Betriebsanleitung oder der Speicherung des lokalen Händlers bekommen sie in der Regel jedoch nicht geboten. Auch die mobilen Apps zielen primär darauf ab, die Qualität der Produkte zu

erhöhen – beispielsweise durch die Integration des Smartphones in das Fahrerlebnis – als die Kundenbeziehung zu vertiefen.

Interessenten bietet der personalisierte Bereich ebenfalls begrenzten Nutzen. Über eine Speicherung von konfigurierten Autos geht der Funktionsumfang meist nicht heraus.

Dem vergleichsweise geringen inhaltlichen Nutzen stehen oft hohe Einstiegsbarrieren gegenüber: Es werden viele Benutzerdaten abgefragt, ohne zu erklären wofür diese benötigt werden. Ein solches Vorgehen kann dem Vertrauen, das der Benutzer in den Anbieter entgegenbringen muss, schaden [DD08]. Ein typisches Problem sind komplexe Passwortrichtlinien, um Kundenkonten zu schützen, die allerdings kaum schützenswert sind. Auch sollten Hürden wie Captchas oder Double-Opt-In entfernt werden, die in dieser Branche, im Gegensatz zum E-Commerce, exzessiv genutzt werden. Ein weiteres Verbesserungspotenzial besteht bei Unterstützung von Eingaben. So sind Fehlermeldungen häufig sehr allgemein gefasst und zeigen nur, dass ein Fehler aufgetreten ist, aber nicht die Art des Fehlers, manchmal sogar in einer Fremdsprache (z.B. französisch).

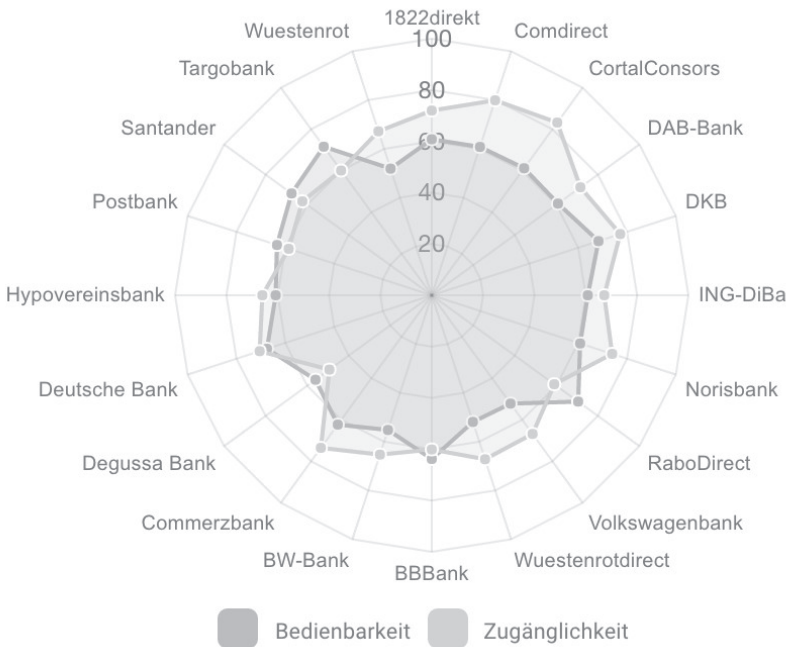


Abb. 2: Visualisierung der Detailergebnisse am Beispiel der Banken.

4 Fazit und Ausblick

Beim Funktionsumfang und bei der Benutzerfreundlichkeit von Personalisierungskomponenten in Webportalen hat die Versicherungsbranche den mit Abstand höchsten Nachholbedarf. Für alle Branchen gilt: Hauptsächlich werden IAM Technologien genutzt, um die Kundenbindung zu erhöhen. Das Potenzial, neue Kunden über personalisierte Services oder Communities zu gewinnen, bleibt weitgehend ungenutzt, obwohl Studien fokussierter Technologieanbieter (z.B. Giga, [Ma16]) belegen, dass die Konversionsrate von Webseiten-Besuchern damit massiv erhöht werden kann. Auch werden die Basis-Anforderungen der Kunden noch in keiner Branche vollständig befriedigt. So sind in vielen Portalen nicht alle Standardgeschäftsprozesse für das Identity- und Accessmanagement umgesetzt – insbesondere die Möglichkeit der Kontolöschung.

Alle Branchen haben erhebliches Potenzial die Benutzerfreundlichkeit zu verbessern. Die meisten Portale machen hierbei dieselben Fehler, unabhängig von der Branche. Eine Lösung böten insbesondere Social Login Mechanismen. Hierdurch besteht die Chance, Einstiegsbarrieren spürbar abzusenken. Für sensitive Geschäftsprozesse oder bei erhöhtem Risikokontext kann dieser Social Login durch zusätzliche Mechanismen – wie Mehrfaktorauthentifizierung oder die risikobasierte Authentifizierung [Fo12] – ergänzt werden, um die geforderten Sicherheitsniveaus zu erreichen. Idealerweise sollte dem Kunden gleichzeitig die Möglichkeit gegeben werden, die Verknüpfung seiner digitalen Identität mit einem sozialen Netzwerk aufzulösen und neue Verknüpfungen zu erstellen.

Die Studie ist als Longitudinalstudie angelegt: Daher erfolgt aktuell eine zweite Erhebung, die erfassen soll welche Veränderungen sich innerhalb der vergangenen 18 Monate ergeben haben. Dazu wurden die bisherigen Erhebungsdaten in eine Datenbank transferiert um beliebige Auswertungen online durchführen zu können (s. Abb. 2).

Literaturverzeichnis

- [Ba12a] Bain & Company, Retail-Banking: Die digitale Herausforderung, 2012
- [Ba12b] Bain & Company, Versicherungen: Die digitale Herausforderung, 2012
- [Ba04] Balfanz, D, et al., In search of usable security: five lessons from the field. IEEE Security & Privacy, vol. 2, no. 5, pp. 19-24, 2004
- [Ca14] Capgemini, The World Retail Banking Report 2014
- [DD08] Dhamlia, R.; Fousseault, L.: The Seven Flaws of Identity Management – Usability and Security Challenges, IEEE Security & Privacy, 2008.
- [Fo14] Forrester Consulting, Customer-Focused Organizations Must Take A Strategic Approach To “Identity Relationship Management”, 2014
- [Fo12] Forrester Consulting, The Forrester Wave™: Risk-Based Authentication, 2012
- [JP14] Jahn, B./Pfeiffer, M., Die digitale Revolution – Neue Geschäftsmodelle statt (nur) neue

- Kommunikation, In Marketing Review St. Gallen, vol.31, no.1, pp. 79-93, 2014
- [La03] Landeszentrum für Datenschutz Schleswig-Holstein, Identity Management Systems (IMS): Identification and Comparison Study, 2003
- [Ma16] Markowski, R.: CIAM als Basis für personalisiertes digitales Marketing, Vortrag bei der IAM Connect 2016, Berlin, 2016.
- [Ra14] Rademacher, P.: Welcher Kunde? – Wie die Digitalisierung den Autohandel überrollt, verfügbar unter: <http://www.automotiveit.eu/welcher-kunde-wie-die-digitalisierung-den-autohandel-ueberrollt/blickpunkt/id-0046378>, 2014
- [Ri04] Riehm, U.: E-Commerce: Begriff, Geschichte, aktueller Stand und Ausblick. in: Kübler, H.-D.; Elling, E. (Hrsg.): Wissensgesellschaft. Neue Medien und ihre Konsequenzen. Bonn: Bundeszentrale für politische Bildung (bpb), 2004
- [TB13] Tarzey, B. / Bamforth, R., Digital identities and the open business - Identity and access management as a driver for business growth, Quocirca, 2013
- [We15a] Weich, D.: Untersuchung der Identitäts- und Zugriffsmanagementsysteme von Endkundenportalen in ausgesuchten Branchen. Masterarbeit, KIT, Cooperation and Management, 2015.
- [We15b] Weierich, P., Abeck, S., Weich, D., Identitäts- und Zugangsmanagement für Kundenportale – Eine Bestandsaufnahme, in Alfred Zimmermann, Alexander Rossmann (Eds.), Digital Enterprise Computing (DEC 2015), Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn, S. 15 - 18, 2015
- [Wi14] Wirtschaft und Finanzen, Die Geschichte des online Bankings, verfügbar unter: <http://www.wirtschaft-und-finanzen.net/finanzen/die-geschichte-des-online-bankings.html>, 2014
- [WT99] Whitten, A.; Tygar, J. D.: Why Johnny Can't Encrypt – A Usability Evaluation of PGP 5.0, Proceedings of the 8th USENIX Security Symposium, 1999.

Beherrschbarkeit komplexer Firmeninfrastrukturen durch eigenschaftsbasierte Sicherheit ¹

Helmar Hutschenreuter² und Dieter Hutter³

Abstract: Mit der zunehmenden Vernetzung von IT-Infrastrukturen hat auch die Frage der Sicherheit solcher Infrastrukturen eine zentrale Bedeutung erlangt. Vielfach sind die Betreiber dieser Infrastrukturen bei der Planung und Implementierung ihrer Sicherheitskonzepte auf sich allein gestellt. Umso wichtiger ist es, dass sie geeignete Hilfsmittel an die Hand bekommen, um in dem Spannungsfeld zwischen Compliance- und Sicherheitsanforderungen des Betriebs einerseits und der notwendigen Verfügbarkeit diverser Software andererseits geeignete globale und langfristige Sicherheitspolitiken entwerfen und umsetzen zu können. Dabei ist von zentraler Bedeutung, dass die IT-Verantwortlichen die Konsequenzen ihrer Entscheidungen für die Sicherheit ihrer Infrastruktur jederzeit überschauen und gegebenenfalls adjustieren können. Dieser Beitrag diskutiert mögliche Lösungen insbesondere unter Usability-Gesichtspunkten anhand eines Projekts SaferApps für eine sichere Ausführung von Fremdapplikationen in einer Unternehmensinfrastruktur.

Keywords: Nutzerzentrierte Sicherheit, Identity Management, Eigenschaftszentrierte Sicherheit

1 Einleitung

Es ist inzwischen ein tradierter Gemeinplatz, dass der Mensch das “schwächste Glied” einer Sicherheitskette bildet [Sc00]. Die Gründe hierfür sind vielfältig. Die Absicherung eines Systems ist komplex und verlangt das Zusammenspiel vielfältiger Sicherheitsmechanismen. Die an das System zu stellenden Sicherheitsanforderungen sind von den Bedürfnissen der Anwender abhängig und können damit nicht fest vorgegeben werden. Die Erfassung seiner Bedürfnisse erfolgt aber in der Regel nicht in Termini seiner Applikationswelt, sondern orientiert sich an den zur Verfügung stehenden einzelnen (technischen) Sicherheitsmechanismen. Die Verantwortung für das geeignete Zusammenspiel der zum Einsatz kommenden Sicherheitsmechanismen wird dem Benutzer aufgebürdet, der mit dieser Aufgabe, die selbst für IT-Sicherheitsspezialisten schwer zu überschauen ist, komplett überfordert ist. Die Folge ist, dass der Benutzer die globalen oder langfristigen Konsequenzen seiner in Termini der technischen Sicherheitsmechanismen vorgenommenen Einstellungen nicht nachvollziehen kann. Die Situation gleicht einem Stellwerksbediensteten bei der Bahn, der für jede Weiche einzeln ihre Stellung festlegt, anstatt komplette kreuzungsfreie Streckenabschnitte für die Züge zu reservieren. Die Folgen sind hinreichend bekannt: Da die Konsequenzen für die Sicherheit eines Systems nicht abgeschätzt werden können, treten ritualisierte Entscheidungsmuster in den Vordergrund. Je nützlicher

¹ Diese Arbeit wurde vom deutschen Bundesministerium für Wirtschaft und Energie im Rahmen des Projektes SaferApps (FKZ KF2013021MS4) gefördert.

² DFKI GmbH, Cyber-Physical-Systems, Bibliothekstr.1, 28359 Bremen, helmar.hutschenreuter@dfki.de

³ DFKI GmbH, Cyber-Physical-Systems, Bibliothekstr.1, 28359 Bremen, hutter@dfki.de

ein System vom Benutzer eingeschätzt wird, desto mehr wird er bereit sein, dem System implizit zu vertrauen und die von ihm angefragten Rechte einzuräumen, da die sicherheitstechnischen Konsequenzen seiner Entscheidung - wenn überhaupt - nur sehr diffus und damit zweitrangig bleiben (siehe z.B. [AC04]). Das ist z.B. auch bei Sozialen Netzwerken der Fall, bei denen Sicherheits- und Privatsphäreneinstellungen für Nutzer oft nur schwer verständlich sind [FNS15]. Der Versuch, den Benutzer lediglich über die Bedeutung einzelner Sicherheitsmaßnahmen aufzuklären, greift zu kurz, da die Konsequenzen aus dem Zusammenspiel der verwendeten Mechanismen für den Benutzer trotzdem nicht begreifbar werden. Bezeichnenderweise ist selbst eine Vielzahl von Systementwicklern oder Administratoren ratlos, wenn sie danach gefragt werden, welche zusätzliche Sicherheitsmaßnahme für ihr System den besten Zugewinn an Sicherheit leisten könnte.

Eine benutzerzentrierte Sicherheit muss daher von den gewünschten Sicherheitsanforderungen, d.h. von den Sicherheitseigenschaften eines Systems, ausgehen und aus denen die zu ihrer Garantie notwendigen Sicherheitsmechanismen (bevorzugt automatisch) ableiten. Dies umfasst insbesondere eine dem Verständnis des Benutzers angepasste Formulierung bzw. Formalisierung seiner (langfristigen) Sicherheitsbedürfnisse. Deren Umsetzung in technische Sicherheitsanforderungen sowie auch deren Dekomposition in sie garantierende Sicherheitsmechanismen sollte im Idealfall im System selbst erfolgen. Betrachtet man beispielsweise die Konfiguration der Sicherheitseinstellungen eines Smartphones (z.B. Android), so muss der Benutzer seine globalen Sicherheitsbedürfnisse auf entsprechende Zugangsberechtigungen einer App auf die verschiedenen Ressourcen (GPS, Kamera, Netz etc.) herunter brechen. Die Lösung dieser Aufgabe erfordert dabei beim Benutzer ein semantisches Modell sowohl über mögliche Verbindungen zwischen den einzelnen Ressourcen als auch über das Zusammenspiel der verschiedenen Applikationen.

Dieses Papier diskutiert Technologien für eine solche benutzerzentrierte Sicherheit anhand einer ähnlich gelagerten Problematik, in der es um die Einbindung von Drittapplikationen in eine Firmeninfrastruktur geht. Basierend auf einem integrierten Identity- und Infrastrukturmanagementsystem sollen Applikationen von Fremdanbieter sicher in die Firmeninfrastruktur integriert werden.

2 Eigenschaftsbasierte Sicherheit

Die Sicherheit eines Systems definiert sich in vielen Fällen durch die in dem System zum Einsatz kommenden Sicherheitsmechanismen. Beispielsweise macht sich vielfach in der Industrie die (IT-)Sicherheit eines Programms an dem Einsatz einer Firewall und eines Virenerkennungsprogramms fest. Eine Garantie für die Sicherheit des Systems können solche Mechanismen nicht bieten, da sie nur bestimmte Angriffstypen unterbinden können. Analog bedeutet beispielsweise die Garantie der Abwesenheit von Speicherüberläufen in einem Programm nicht, dass das Programm eine von dem Mechanismen unabhängige Sicherheitseigenschaft besitzt. Andere Programmierfehler, die von den eingesetzten Sicherheitsmechanismen nicht erkannt werden, könnten zum Beispiel ähnlich fatale Folgen für die Sicherheit des Systems haben wie die oben erwähnten Speicherüberläufe. Damit ist der Einfluss der Sicherheitsmaßnahmen auf eine Garantie bzw. einen Nachweis der vom

Benutzer geforderten Sicherheitsanforderungen nur indirekt gegeben und der Benutzer mit der Frage, welche Mechanismen wann und wie zum Einsatz kommen sollen, entsprechend überfordert.

Dementsprechend besteht ein großes Interesse an einem Mechanismen-unabhängigen Sicherheitsbegriff, der sich über garantierte (positive) Sicherheitseigenschaften eines Systems definiert. Idealerweise sind solche Sicherheitseigenschaften gesucht, die bezüglich der Komposition sicherer Einzelsysteme oder über verschiedene Verfeinerungsebenen skalieren. Ein klassisches Beispiel eines solchen eigenschaftsbasierten Sicherheitsbegriffs ist die Abwesenheit von Informationsflüssen zwischen bestimmten Ein- und Ausgabegrößen und damit die Unabhängigkeit eines sichtbaren Verhaltens eines Systems von etwaigen vertraulichen Daten in dem System. Speziell im Bereich der Formalen Methoden (und in Deutschland im Rahmen des DFG Schwerpunkts SPP 1496 "Reliably Secure Software Systems) wird an der Entwicklung solcher eigenschaftsbasierten Sicherheitsbegriffen gearbeitet, die mit Hilfe des Einsatzes korrespondierender Sicherheitsmechanismen garantiert werden können.

Eine eigenschaftsbasierte Sicherheit als eine deklarative Spezifikation einer Sicherheitspolitik lässt sich besser einem Benutzer vermitteln als deren Umsetzung im Sinne einer Kollektion von implementierten Sicherheitsmechanismen, da diese Eigenschaften in der Regel nicht an eine spezifische Implementierung gebunden sind sondern auch auf abstrakten Ebenen formulierbar sind. Abhängigkeiten zwischen Ein- und Ausgangsgrößen eines Systems können auch ohne Kenntnis einer Implementierung dem Benutzer vermittelt werden. Eine eigenschaftsbasierte Sicherheit ist nicht an spezielle Mechanismen und damit auf spezielle Abstraktionsebenen, auf denen die Mechanismen definiert sind, gebunden. Sie erlaubt damit die Modifikation der Techniken, die zur Garantie der Eigenschaften beitragen. Will man beispielsweise im Bereich der verschiedenen Applikationen für die Firmeninfrastruktur formalisieren, dass sensitive Firmendaten nicht zu unbefugten Dritten gelangen können, sind die Mechanismen, mit denen dies später erreicht werden kann, erst mal zweitrangig. Damit bilden diese formalen, mathematisch verifizierbaren Sicherheitsbegriffe eine Brücke zwischen den technischen Sicherheitsmechanismen eines Systems und dem intuitiven Sicherheitsverständnis des Benutzers.

3 Fallbeispiel SaferApps

Das Projekt SaferApps beschäftigt sich mit der sicheren Integration von Unternehmensanwendungen in bestehende IT-Infrastrukturen. Die in SaferApps entwickelten Lösungsansätze werden am Beispiel des Serverbetriebssystems *Univention Corporate Server (UCS)*[Un] erprobt, welches von dem Bremer Unternehmen *Univention* entwickelt wird. UCS bietet die Möglichkeit, Unternehmensanwendungen einfach über einen AppStore zu installieren. AppStores sind bereits aus dem Smartphone-Bereich bekannt und erlauben die einfache Installation von Applikationen auf einem System. So können sich Anwendungen (Apps) einfach auf einer UCS-Installation im Unternehmensnetzwerk bereitgestellt werden. Da eine ungeprüfte Integration von Anwendungen aus verschiedenen Quellen aber die Sicherheit des Unternehmensnetzwerkes korrumpieren kann, ist es ein Hauptziel von Sa-

ferApps einen Mechanismus für eine sichere Integration bereitzustellen. Dabei sind die Hauptherausforderungen aus der Sicht der Usability, dass diese Integrationstechnik für die verschiedensten Apps anwendbar sein muss, ohne dass dabei signifikante Mehraufwände für die einzelnen Akteure, d.h. den Hersteller der App, den Betreiber des AppStore bzw. den Administrator der Firmeninfrastruktur anfallen. Dabei treffen die unterschiedlichen Interessen der beteiligten Akteure aufeinander. Der Betreiber des AppStore möchte zur Steigerung seiner Attraktivität eine Vielzahl verschiedener Apps in seinem Store anbieten gleichzeitig aber vermeiden, dass mögliche in einer App verborgene Schadware die Infrastruktur seiner Kunden korrumpieren kann und somit die Vertrauenswürdigkeit des AppStore selbst in Mitleidenschaft zieht. Andererseits möchte er den Aufwand, eine App in seinem Store anbieten zu können, auf ein Minimum begrenzen. Analog zu Smartphones besteht die Lösung dieses Problems in der Bereitstellung von Sandboxes in UCS, in denen einzelne Apps, abgeschottet von anderen Apps und kritischen Basisdiensten des Betriebssystems, ausgeführt werden.

Abbildung 1 gibt einen Überblick über den Lösungsansatz von SaferApps. Jede App ist in einer eigenen Sandbox gekapselt. Die Grenzen der Sandboxes sind durch eine doppelte Umrandung gekennzeichnet. Das UCS-System stellt Apps verschiedene Betriebsressourcen, insbesondere ein Dateisystem, ein LDAP-Verzeichnis sowie Zugriff auf das Netzwerk, zur Verfügung. Apps nutzen diese Ressourcen zur Kommunikation mit UCS und anderen Apps sowie zur Speicherung von Daten, wie beispielsweise zum persistenten Ablegen von Konfigurationen. Zur Installation neuer Apps verfügt der AppStore über ein SaferApps-spezifisches Installationsprogramm. Dieses setzt eine Zugriffskontrolle an den Grenzen der Sandbox auf, bevor es die neu installierte App innerhalb der Sandbox startet.

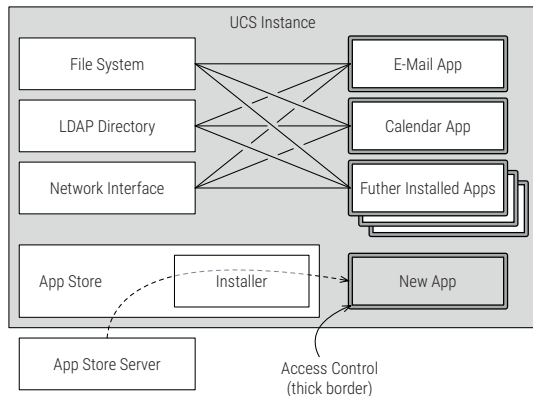


Abb. 1: SaferApps-Ansatz

Zur Installation neuer Apps verfügt der AppStore über ein SaferApps-spezifisches Installationsprogramm. Dieses setzt eine Zugriffskontrolle an den Grenzen der Sandbox auf, bevor es die neu installierte App innerhalb der Sandbox startet.

Diese Sandbox wird mit Hilfe der Linux-Container-Technologie Docker [Do] realisiert. Dafür wird eine Anwendung zunächst als sogenanntes Docker Image bereitgestellt und anschließend auf den UCS-Server heruntergeladen. Ein beliebiger Prozess (bspw. ein Serverdienst wie ein Webserver) kann auf Basis dieses Image als sogenannter Docker Container gestartet werden und getrennt von anderen Prozessen des Servers laufen. Der Vorteil von Docker ist, dass ein Container nativ im Linux-Kernel des Wirtssystems ausgeführt wird, keine Hardware emuliert werden muss und somit im Vergleich zu bestehenden Virtualisierungslösungen besonders ressourcensparend arbeitet. Docker erlaubt es, externe Ressourcen in den Container einzubinden. So lassen sich Dateien und Verzeichnisse des Hostsystems in das Dateisystem des Gastsystems im Container einbinden oder Netzwerkzugriffe

auf die Netzwerkschnittstelle des Hostsystems an die des Gastsystems weiterleiten. Am Ende entscheidet die Konfiguration des Containers, ob die App einen Zugriff innerhalb oder außerhalb der Sandbox ausführt, sodass die App nicht zwischen Ressourcen des Wirtbetriebssystems und denen ihrer Sandbox unterscheiden muss. Jede Kommunikation eines Docker-Containers nach Außen wird über den typbasierten Zugriffskontrollmechanismus AppArmor [Ap] überwacht. Für die Installation und den Betrieb von Apps, die beispielsweise Groupware-Funktionalitäten anbieten (E-Mail, Kalender, etc.), benötigen diese Apps partiellen Zugriff zu der im LDAP des Hostsystems realisierten Benutzerverwaltung. LDAP-Verzeichnisse sind Datenbanken, die bereits ihre eigene Zugriffskontrolle mitbringen, die sich über *Access Control Lists* (ACLs) konfigurieren lässt. Einträge sind in einem LDAP-Verzeichnis in einer Baumstruktur angeordnet und bestehen aus einer Liste von Key-Value-Paaren, den Attributen. Einträge, für die eine ACL gelten soll, lassen sich über ihre Position im Baum oder über Eigenschaften ihrer Attribute adressieren. Da bekannt ist, wie das Dateisystem oder LDAP-Verzeichnisse Daten intern verarbeiten, kann durch die Analyse von Zugriffen auf diese Ressourcen auf konkrete Informationsflüsse geschlossen werden. Bei Netzwerkdiensten ist nicht bekannt, wie diese Daten intern verarbeiten. Meist befinden sich diese auch außerhalb des Einflussbereichs des UCS-Systems, da sie nicht lokal ausgeführt werden. Daher ist hier eine Abschätzung des Informationsflusses durch die Ressource nicht möglich.

4 Usability in der Praxis

4.1 Ausgangslage

Wie bereits erwähnt, wird zur Kontrolle der Zugriffe einer Sandbox die Zugriffskontrollmechanismen AppArmor und LDAP-ACLs verwendet. Die Konfiguration dieser Zugriffskontrollen wird automatisch aus den Regeln einer Sicherheitspolitik generiert, die gemeinsam mit der App im AppStore hinterlegt ist und bei einem Installationswunsch mit ausgeliefert wird. Im Unterschied zu Smartphones definiert nicht der Endbenutzer sondern der Systemadministrator die (globalen) Sicherheitsbedürfnisse. Damit fallen Ansätze, die Usable Security über die Handlungen des Endnutzers definierten wie z.B. [Ye04, Ro12], aus. Das Konzept von SaferApps wird maßgeblich durch die Sicherheitsanforderungen und Anforderungen der Nutzer beeinflusst. Die zentrale Sicherheitsanforderung ist, dass die automatisiert konfigurierte Zugriffskontrolle zuverlässig arbeitet und ausschließlich der App die Zugriffe genehmigt, die durch ihre Sicherheitspolitik erlaubt sind. Die Anforderungen der Nutzer unterscheiden sich, orientieren sich aber gemeinschaftlich an einer einfachen Bedienung dem intuitiven Verständnis der Funktionsweise von SaferApps.

Um einen AppStore mit einer großen Bandbreite von Anwendungen füllen zu können, ist es wichtig, dass die Barrieren für App-Hersteller möglichst klein gehalten werden. Daher sollte ein Mechanismus, wie SaferApps, keinen signifikanten Mehraufwand für den Hersteller verursachen. SaferApps sieht vor, dass zu jeder App eine Sicherheitspolitik spezifiziert werden muss, die beschreibt, welche Zugriffsrechte die App zur Ausübung ihrer Funktionalität benötigt. Denkbar ist auch eine strukturierte Politik, die in Abhängigkeit

der benötigten Funktionalitäten die angeforderten Rechte reduziert. Inwieweit dies von den Herstellern allerdings unterstützt werden wird, ist fraglich.

Für den Betreiber des AppStore steht die Wartbarkeit des Store im Vordergrund. AppStores, wie sie von Mobilplattformen bekannt sind, lassen neue Anwendungen einen Zulassungsprozess durchlaufen. Hierbei wird geprüft, ob die Anwendung den Anforderungen des AppStore genügt. Dabei steht nicht zwingend die Sicherheit im Vordergrund, sondern auch ihre Inhalte oder ob sie rechtlichen Rahmenbedingungen entsprechen, die vom jeweiligen Betreiber des Stores vorgegeben werden. Unser Ziel war es, dass SaferApps keine weiteren Überprüfungen des AppStore Betreibers erfordert. Insofern wird die vom App-Hersteller formulierte Sicherheitspolitik nicht vom AppStore-Betreiber überprüft, sondern der Administrator einer UCS-Installation soll über eine Benutzeroberfläche die Zugriffsmöglichkeiten in Bezug auf seine individuelle IT-Infrastruktur regeln, denn er muss letztlich die Verantwortung für eine in seinem Unternehmensnetzwerk installierte App tragen und trifft damit die abschließende Installationsentscheidung.

Der Administrator ist verpflichtet, die Compliance- und Sicherheitsrichtlinien seines Unternehmens umzusetzen. Er benötigt deshalb ein Werkzeug, welches ihm ermöglicht zu prüfen, ob eine App diesen Richtlinien entspricht. SaferApps soll diese Prüfung insofern unterstützen als es den Administrator informiert, falls eine App den technischen Sicherheitsrichtlinien widerspricht. Das wird dadurch realisiert, indem die Richtlinien auf eine globale Sicherheitspolitik abgebildet werden, die auf der jeweiligen UCS-Installation hinterlegt wird. Die vom Hersteller vorgegebenen Politiken der Apps werden bei der Installation mit dieser Anwenderpolitik abgeglichen. Ein Schwerpunkt bei SaferApps ist umfassende und verständliche Information des Administrators. SaferApps stellt ein Werkzeug zur Verfügung, welches Kommunikationswege mit bereits installierten Anwendungen und dem UCS-System aufzeigt und diese (sofern vorhanden) mit Begründungen des App-Hersteller versieht. Das gibt dem Administrator eine Hilfestellung beim Verständnis von Seiteneffekten und Auswirkungen auf die Vertraulichkeit bestimmter Unternehmensdaten, die durch die Installation einer neuen App entstehen. Mit der zusätzlichen Möglichkeit bestimmte Funktionen von Anwendungen zu deaktivieren (sofern der App-Hersteller dieses unterstützt), kann der Administrator die vorgegebene Sicherheitspolitik einer App nachträglich gemäß seinen Anforderungen adjustieren.

4.2 Sicherheitspolitik

Analog zur Smartphone-Welt liefern App-Hersteller ihre Apps mit einer korrespondierenden Sicherheitspolitik aus, die die benötigten Zugriffsrechte der App definiert. Diese Politik kann strukturiert sein, indem sie Rechte einzelnen (optionalen) Funktionalitäten der App zuordnet, und damit es ermöglicht, die App auch mit eingeschränkten Rechten und damit eingeschränkter Funktionalität zu verwenden. Die von einem App-Hersteller bereitgestellte Sicherheitspolitik ist eine sortierte Liste von Zugriffsregeln. Bei der Auswertung der Politik ist die Reihenfolge der Regeln maßgeblich. Analog zu LDAP-ACLs gilt im Allgemeinen, dass für einen Zugriff immer die erste Regel gilt, deren Vorbedingung erfüllt ist. Dieses Konzept regelt damit einerseits den Umgang mit gegenseitig konkurrie-

renden Regeln und erleichtert andererseits die spätere Übersetzung der Zugriffsregeln für die LDAP in die standardmäßige ACL-Syntax.

SaferApps-Sicherheitspolitiken sind Metapolitiken, die nicht zur Laufzeit der App ausgewertet sondern zum Installationszeitpunkt einmalig ausgewertet werden, um die eigentliche Zugriffskontrolle (automatisiert) zu konfigurieren. Dabei wird die Sicherheitspolitik in die benötigten Konfigurationsdateien der verwendeten Zugriffskontrollmechanismen übersetzt. Bei dieser Übersetzung werden insbesondere auch Optimierungen, wie beispielsweise das Entfernen überflüssiger Regeln, vorgenommen.

Während im Falle des Dateisystems und des LDAP-Verzeichnis die Strukturen und der Zugriff der Apps auf die verschiedenen Speicherbereiche offensichtlich ist und dementsprechend disjunkte Bereiche den einzelnen Apps zugeordnet werden können, ist für einen Netzwerkdienst oft keine Struktur (bzw. Abgrenzung von Speicherbereichen) bekannt. Daher wird bei Daten, die von einer App über das Netzwerk versandt werden, immer davon ausgegangen, dass diese von einer anderen App über eine beliebige Netzwerkverbindung wieder empfangen werden können. Somit wird das Netzwerk in seiner Gesamtheit als einzelner (nicht weiter unterteilbarer) Speicherbereich gesehen, und Netzwerkzugriffsregeln dienen dazu die Konfiguration einer Firewall abzubilden.

Per Vergleich der Zugriffsregeln zweier Apps lässt sich herausfinden, ob ein Informationsfluss zwischen den Apps stattfinden kann. Informationsflüsse werden aus sogenannten Regelkollisionen impliziert. Eine Regelkollision entsteht, wenn sich Zugriffsrechte und Speicherbereiche zweier Regeln überschneiden. Die im Folgenden vorgestellte Regelsyntax zeigt, dass Regeln ähnlich wie postalische Adressen aufgebaut sind. Diese geben das Ziel einer Sendung über die Angabe von mehreren Attributen an. Wie die Attribute einer Postadresse den Weg einer Sendung vorgeben, so geben auch die Attribute einer Zugriffsregel die Route zu einem bestimmten Speicherbereich vor. Über einen Vergleich der Attribute zweier Regeln lässt sich feststellen, ob sich ihre Speicherbereiche überschneiden oder nicht. Allerdings ist anzumerken, dass eine Regelkollision lediglich auf einen potentiellen Informationsfluss hindeutet. Ob ein Informationsfluss tatsächlich stattfindet, hängt von den tatsächlichen Zugriffen der Apps ab. Diese werden jedoch von SaferApps nicht berücksichtigt. So ist es möglich, dass Informationsflüsse statisch erkannt werden, die dynamisch nicht auftreten. SaferApps geht von einem Informationsfluss von App1 nach App2 aus, wenn App1 schreibenden und App2 lesenden Zugriff auf einen gemeinsamen Speicherbereich (Dateisystem oder LDAP) hat.

Um den Beteiligten die Formulierung einer Sicherheitspolitik zu erleichtern, ist die Syntax der Regeln sehr einfach gehalten. Eine Regel ist eine Menge von Attribut-Werte-Paaren, welche leicht in gebräuchlichen Datenformaten, wie *XML*, *JSON* oder *YAML* abgebildet werden können. Jede Zugriffsregel muss die Standardattribute *type* und *permission* beinhalten. *type* gibt an, ob die Regel Rechte für Datei-, Netzwerk- oder LDAP-Zugriffe erteilt. *permission* gibt an, welche Zugriffsrechte erteilt werden. Gültige Werte des Attributs *permission* sind *ro* (read only) und *rw* (read/write). Um Netzwerkregeln intuitiver formulieren zu können, existieren zudem die Aliasnamen *listen* (Alias für *ro*) und *connect* (Alias für *rw*). Zudem existiert das optionale Attribut *features*. Dieses gibt an für welche Features der App die Regel benötigt wird. Fehlt dieses Attribut, wird ange-

nommen, dass die Regel essentiell für die Funktion der App ist und nicht an bestimmte Features gebunden ist. Neben den Standardattributen enthält jede Zugriffsregel abhängig von ihrem Typ weitere Attribute (vgl. Tabelle 1).

Attribut	Beschreibung
<i>Standardattribute</i>	
<code>type</code>	Typ der Zugriffsregel (Datei, Netzwerk, LDAP)
<code>permission</code>	Rechte, die diese Regel erteilt
<code>features</code>	Features, für die diese Regel erforderlich ist
<i>Attribute für Regeln zum Dateizugriff</i>	
<code>path</code>	Datei- oder Verzeichnispfad
<i>Attribute für Regeln zum Netzwerkzugriff</i>	
<code>host</code>	Hostname, IP-Adresse
<code>port</code>	Portnummer
<code>protocol</code>	Verwendung von TCP oder UDP
<i>Attribute für Regeln zum LDAP-Zugriff</i>	
<code>area</code>	LDAP-DN und ggf. DN-Selektor
<code>filters</code>	LDAP-Filter
<code>permitted_attrs</code>	Eingrenzung auf Attribute

Tab. 1: Attribute der verschiedenen Regeltypen

Falls eine App Funktionen besitzt, die optional sind und somit nicht zwingend für ihre Hauptfunktion erforderlich sind, kann der App-Hersteller diese ebenfalls in der Sicherheitspolitik definieren. Eine solche Feature-Definition ist, wie Regeln auch, ebenfalls eine Menge von Attribut-Werte-Paaren. Sie besitzt die Standard-Attribute `id`, `name` und `description`. `id` ist ein eindeutiger Identifier des Feature, der frei gewählt werden kann. Dieser Identifier wird verwendet, um im Regelattribut `features` auf ein Feature zu verweisen. Über die Attribute `name` und `description` soll der App-Hersteller durch die Wahl eines prägnanten Names und über einen Beschreibungstext das Feature für den Nutzer bzw. den Administrator einer UCS-Installation der App beschreiben.

4.3 User Interface

In SaferApps ist der Installationsprozess einer neuen App zentral. Hier wird die Zugriffskontrolle konfiguriert, wobei sich ihre Konfiguration aus der Sicherheitspolitik des App-Herstellers und den Konfigurationswünschen des Administrators der UCS-Installation, auf der die neue App installiert wird, ergibt. Um die Informationen aus diesen beiden Quellen zusammenzuführen, verwendet SaferApps eine graphische Benutzeroberfläche (GUI), die dem Administrator Informationen über die Konsequenzen der Installation gibt und ihm auch die Möglichkeit gibt, die Konfiguration der Zugriffskontrolle zu beeinflussen und auf diese Weise die vom App-Hersteller vorgegebene Sicherheitspolitik zu adjustieren.

GUIs vieler Installationsprogramme folgen einem Wizard-Konzept, um mit dem Benutzer zu interagieren. Nach diesem klickt sich der Nutzer durch eine Folge von Dialogfenstern, die aufeinander aufbauen. SaferApps verwendet ebenfalls einen Wizard. Im ersten Schritt

(siehe Abbildung 2) fragt der Wizard vom Administrator ab, welche Funktionalitäten der neuen App tatsächlich genutzt werden sollen. So können die Zugriffe bereits im Vorfeld eingeschränkt werden, indem Zugriffe für nicht benötigte Funktionalitäten automatisch ausgeschlossen werden. Durch die Entscheidung für oder gegen einzelne Funktionalitäten trifft der Administrator indirekt auch die Entscheidung für oder gegen einzelne Zugriffsregeln der Sicherheitspolitik der App. Durch die Verknüpfung mit den Funktionalitäten fällt ihm dieses einfacher, da sich aus den Zugriffsregeln nicht logischerweise bestimmte Funktionalitäten ableiten lassen, während umgekehrt für einzelne Funktionalitäten korrespondierende Zugriffe erforderlich sind.

Im zweiten Schritt (siehe Abbildung 3) wird der Administrator über Informationsflüsse informiert, die zwischen der neuen App und bestehenden Apps auftreten. Diese werden in verschiedenen Kategorien gruppiert. So gibt es Senden, Empfangen und bidirektionale Zugriffe sowie Netzwerkzugriffe. Falls in einer Kategorie keine Zugriffe vorhanden sind, wird die Kategorie komplett ausgeblendet. Während man auf Smartphone-Betriebssystem Apps Zugriffe auf bestimmte Arten von Daten (Kontakte, Kalender, etc.) erlauben kann, lässt sich aufgrund der individuellen Beschaffenheit der Installationen von Unternehmensservern die dort verarbeiteten Daten nicht eindeutig bestimmten Kategorien zuordnen. Daher werden die Daten nach den Apps gruppiert, die sie verarbeiten, um so dem Administrator eine intuitiv erfassbare Kategorisierung anzubieten.

Es ist anzumerken, dass die Menge der Informationen in diesem Schritt von der Anzahl und Art der bereits auf dem System installierten Apps abhängt. Treten nur wenige Regelkollisionen auf, so ist die Anzahl der Informationsflüsse gering. Somit erhält der Administrator gezielt die Zugriffe präsentiert, die tatsächlich einen Einfluss auf den Informationsfluss haben. Alle andere Zugriffe der Sicherheitspolitik muss er nicht prüfen.

Der Administrator ist an vorhandene Compliance- und Sicherheitsrichtlinien seines Unternehmens gebunden und muss diese durchsetzen. Deshalb ist es wichtig, dass er im Fall eines sich ergebenden Informationsflusses, der diesen Richtlinien widersprechen würde, die Sicherheitspolitik korrigieren kann. Eine Adjustierung ist allerdings nur bei Informationsflüssen möglich, die an Funktionalitäten der App gebunden sind, die im ersten

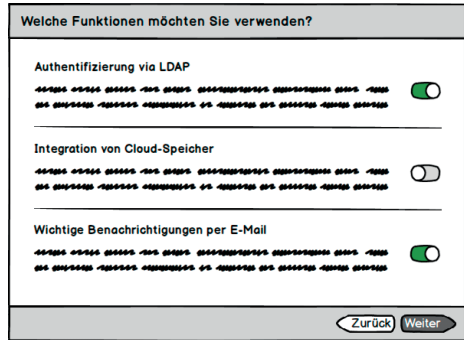


Abb. 2: Einschränkung der App-Funktionalität

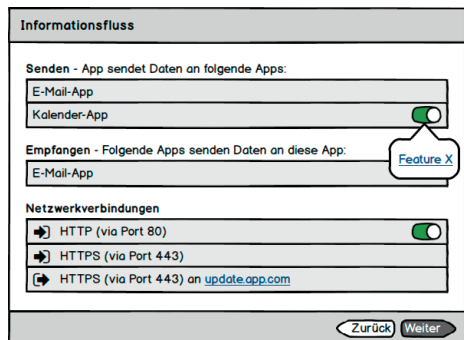


Abb. 3: Information den Informationsfluss

Schritt des Wizard ausgewählt werden konnten (also optional sind). Der Administrator soll so nicht in eine Situation geraten, in der eine installierte App nicht korrekt funktioniert, weil ihre Zugriffe von ihm zu rigide eingeschränkt wurden. Informationsflüsse, die sich verhindern lassen, sind mit einem On-Off-Switch versehen. Bei einem Klick auf diesen wird in einem Pop-Over angezeigt welche Funktionen der App von diesem Informationsfluss abhängen. Mit einen zusätzlichen Klick werden diese Funktionen nachträglich abgewählt und die aus ihnen resultierenden Zugriffe gesperrt.

5 Fazit

SaferApps zeigt, dass die Benutzerfreundlichkeit des Sicherheitskonzeptes wesentlich davon abhängt, dass die Konfiguration der Sicherheitsmechanismen in ihrer Konsequenz auf die Sicherheitseigenschaften des Gesamtsystems illustriert werden. Umgekehrt wäre es auch wünschenswert, dass aufgezeigt würde, welche Modifikation an den Sicherheitsmechanismen welche Effekte auf die Sicherheitseigenschaften haben. SaferApps ist ein erster Schritt in diese Richtung, wobei die angezeigten Informationsflüsse zwischen den Apps bisher die internen Flüsse einer App nicht berücksichtigen. Für (wesentlich kleinere) Smartphone-Apps ist eine solche feinere Informationsflussanalyse bereits heute automatisch machbar (e.g [Ar14]). Eine Analyse der Methodik hinsichtlich eines Security-Usability-Threat-Modell (z.B. [KFR10]) könnte helfen, um einen Kompromiss für die Granularität einer solchen Informationsflussanalyse zu ermitteln.

Literaturverzeichnis

- [AC04] Aytes, K.; Connolly, T.: Computer Security and Risky Computing Practices: A Rational Choice Perspective. *J. of Organizational and End User Computing*, 16(3):22–40, 2004.
- [Ap] AppAmor: , Linux application security system. <http://wiki.apparmor.net>.
- [Ar14] Arzt, S.; Rasthofer, S.; Fritz, C. et al.: FlowDroid: Precise Context, Flow, Field, Object-sensitive and Lifecycle-aware Taint Analysis for Android Apps. In: *Proceedings of the 35th ACM SIGPLAN Conference on Programming Language Design and Implementation. PLDI '14*, ACM, New York, NY, USA, 2014.
- [Do] Docker: , An open platform for distributed applications. www.docker.com.
- [FNS15] Foltz, C.B.; Newkirk, H.E.; Schwager, P.H.: An Empirical Investigation of Factors that Influence Individual Behavior toward Changing Social Networking Security Settings. *J. of Theoretical and Applied Electronic Commerce Research*, 11(2):1–15, 2015.
- [KFR10] Kainda, R.; Flechais, I.; Roscoe, A. W.: Security and Usability: Analysis and Evaluation. In: *10th Int. Conference on Availability, Reliability, and Security (ARES 2010)*. 2010.
- [Ro12] Roesner, F.; Kohno, T.; Moshchuk, A.; Parno, B.; Wang, H. J.; Cowan, C.: User-Driven Access Control: Rethinking Permission Granting in Modern Operating Systems. In: *2012 IEEE Symposium on Security and Privacy*. 2012.
- [Sc00] Schneier, Bruce: *Secrets & Lies - Digital Security in a Networked World*. John Wiley & Sons, 2000.

[Un] Univention: , Univention Corporate Server UCS. www.univention.de/produkte/ucs/.

[Ye04] Yee, Ka-Ping: Aligning security and usability. IEEE Security Privacy, 2(5):48–55, 2004.

NoPhish: Evaluation of a web application that teaches people being aware of phishing attacks

Alexandra Kunz¹, Melanie Volkamer¹, Simon Stockhardt¹, Sven Palberg², Tessa Lottermann² und Eric Piegert²

Abstract: Phishing has evolved to a serious cause of risk in our daily contact with the World Wide Web. Therefore, different extensions and plugins for web browsers were developed to detect phishing websites. To furthermore minimize the risk of falling for a phishing attack, the users themselves have to be educated. Therefore, the online game “NoPhish” has been developed, which explains the basics of phishing attacks and how to detect them efficiently. In the following study, the success rate of this online tool was measured. The goal was to determine which phishing strategies are effective in fooling users, which strategies can be practised well and which strategies are still effective in fooling users after having been taught by some educational material. The effectiveness of “NoPhish” in increasing users’ security awareness and the ability of detecting phishing URLs could be proven. Furthermore, it could be determined which types of phishing should be drawn special attention to in future development of phishing education material.

Keywords: Usable security, security awareness, phishing education

1 Introduction

Phishing represents an attempt from so-called phishers to elicit confidential information of users by using faked websites. These phishers want to get access to private account information and passwords which are used for e.g. e-banking, social networks or online shops. A successful phishing attack can have disastrous consequences for the victims leading to financial losses and identity theft. Usually these phishers send fraudulent e-mails or chat messages with a link and the order to click on it. There is a multitude of different phishing attacks like spear phishing where phishers want to increase their success rate by sending e-mails to specific companies with individual matched content. Another type of phishing is called clone phishing where phishers clone a previously sent message and replace the legit content with malicious information like links or formulas. That phishing plays a major role in our daily life shows the statistic of the Anti-Phishing Working Group (APWG), identifying around 50,000 new phishing websites every month, with retail being the most targeted industry sector at the moment and payment services close behind, but also in other sectors such as social networking. More and more companies fear that users will lose confidence in electronic commerce. Therefore, an

¹ Technische Universität Darmstadt, Fachbereich Informatik, SECUSO, 64289 Darmstadt, kontakt@secuso.org

² Technische Universität Darmstadt, Fachbereich Informatik, SECUSO, 64289 Darmstadt, kontakt@tu-darmstadt.de

efficient protection against phishing is needed.

Ludl, McAllister, Kirda and Kruegel (2007) engage in the effectiveness of techniques to detect phishing sites in their research. Their results show a 90% rate of detection of phishing attempts of blacklist-based solutions. The method identifies phishing sites after they are launched and reported as suspicious. These solutions represent an efficient way of discovery. However, their research covers only extensions and plugins for web browsers. But to furthermore minimize the risk, the users themselves have to be educated.

Dhamija, Tygar and Hearst (2006) studied which phishing attacks succeed in fooling users and why. They tested 9 different types of phishing attacks like different types of spoofing or different types of requested information. 22 participants were shown 20 web sites and asked afterwards to determine which ones were fraudulent. Additionally, the users had to evaluate their confidence about their decisions. Generally, the users were quite confident about their answers. However, 7 participants had never heard of phishing before they participated the study. Furthermore, 13 participants never paid attention to “HTTPS” and further 5 participants mentioned that they even never take a look at the address bar at all. To sum up, the results of the study lead to the conclusion that educating users about phishing has to be taken very seriously. These alarmingly high values show even more the importance of user education such as in the research of Sheng, Magnien, Kumaraguru, Acquisti, Cranor, Hong and Nunge (2007) with their online-game „Anti-Phishing Phil“. Their results demonstrate that games represent a highly efficient way of information transfer. Using a story-based approach with a phish character guiding through the game they provided a challenging, contextual and interactive game experience. Their results show the improved skills of the game users in detecting phishing-URLs compared to groups, which were not allowed to use the interactive game but instead read existing training material or had been tutored.

Canova, Volkamer, Bergmann and Reinheimer (2014) had a similar approach for educating users by an interactive game. They developed the Android app “NoPhish” which tutores the users in detecting phishing-URLs. Due to the fact that people are not regularly confronted with phishing attacks, the authors analysed its effectiveness on users’ knowledge retention. The results of their studies show that users of this app are more successful in detecting phishing-URLs, particular over a longer period.

Sheng, Holbrook, Kumaraguru, Cranor and Downs (2010) study the effectiveness of different educational materials to identify phishing webpages in their research. Participants of their study had to complete a role play to measure their susceptibility to fall for phishing attacks. On the basis of these results the participants were assigned to four different experimental groups with different educational material: a PhishGuru cartoon, Anti-Phishing Phil, popular web-based training materials and a combination of Anti-Phishing Phil plus a PhishGuru cartoon. The results of the authors’ study show that educational material reduces the end-users’ risk of supplying private information on phishing webpages by 40%.

A further work of Kumaraguru, Rhee, Acquisti, Cranor, Hong and Nunge (2007) is about

the comparison of three methods for improving the users' skills of detecting phishing attacks – two embedded training designs and another method consisting of simple email security notices. The embedded training methods consist of regularly sent phishing emails. If a user does not assess the mail as a phishing attack and clicks on a link, the user receives a warning. The first method presents information using text and graphics, the second method uses a comic strip format. The results of the authors' study demonstrate that the number of identified phishing emails decreased using the embedded training designs. The authors found out that the comic strip intervention was most effective while security notices were rather ineffective in teaching people about phishing attacks.

In the present study, the challenging and interactive web application of “NoPhish” was developed (based on Volkamer's “NoPhish” Android app) to educate users and increase their awareness and ability in detecting phishing URLs. The recruited participants had to do a pretest for checking their ability in detecting phishing attacks. Afterwards they ran through a training phase where they learnt the basic attacks of phishing and in a final step they had to test their learnt knowledge in a concluding posttest. The goal of the present study is to investigate 7 different types of phishing, more precisely to analyse which of them are effective. Another interesting fact is, which attacks can be practised well so that after some training time there is a high success rate in detecting phishing attacks. The last aspect is to investigate which phishing strategies are still effective in fooling users after they were educated by teaching materials.

2 Method

2.1 Participants

Participants were recruited from December 2015 until January 2016 via social networking sites like Facebook and via sending personal messages to circles of acquaintances. 65 users registered at the web application but 9 of them answered neither pretest nor posttest and 18 of them did not answer the posttest. Finally, 32 of 65 participants were left for analysing their data. The behaviour of 14 female participants and 18 male participants aged between 19 and 56 ($M = 28.54$, $SD = 10.79$) was analysed. They got no reward for participating the study. There were no special selection criteria or limitations to participate the study and the participants were not told about the purpose of the study.

2.2 Stimuli

The users were presented a total of 68 different stimuli, 28 items of the pretest and 40 stimuli of the education part. The same 28 items of the pretest were also presented in the posttest. The stimuli design and selection happened by the investigators. The stimuli were images of websites with corresponding web addresses of different types of phishing. Knowledge of the structure of web addresses could help identify phishing messages

(shown in Figure 1). The domain, the here called „who-section” (Wer-Bereich) is the most important part for detecting phishing URLs.

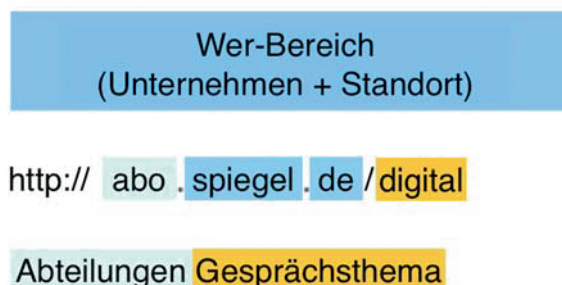


Fig. 1: Structure of a web address in “NoPhish”

In the study the following 7 types of phishing attacks were investigated:

- phishing type 1: URLs based on IP addresses (e.g. <https://87.122.24.91/ws>)
- phishing type 2: URLs where the who-section doesn't contain the company name (e.g. <https://www.hfkjt.com>)
- phishing type 3: URLs which contain the company name at the place of the department (Abteilung) (e.g. <https://www.instagram.account.com>)
- phishing type 4: URLs which contain the company name at the place of the topic of conversation (Gesprächsthema) (e.g. <https://www.account.com/t-online.de/settings>)
- phishing type 5: URLs where the who-section seems similar to the real URL but use an additional term (e.g. <https://www.bahn-support.de>)
- phishing type 6: URLs which contain typos (<https://www.facebok.com>)
- phishing type 7: URLs which contain similar looking letters and numbers like the real ones (e.g. <https://www.clropbox.com>).

The pretest has 14 faked items (2 items per phishing type) and 14 legitimate items. The same 28 items of the pretest are used for the posttest afterwards. The education part covers 8 different levels. Level 1 imparts knowledge of fundamental concepts to detect phishing attacks. Therefore, Level 1 doesn't contain any specific phishing addresses. Level 2 to Level 8 cover the above-named phishing types (1) to (7). Thereby each level highlights

one specific type of phishing, each with 6 corresponding URLs. The education part used only URLs without images of webpages. Every user got the same URLs in the same order.

2.3 Procedure

The experiment took about 45 minutes per person and was performed in 4 phases: (1) registration and personal data, (2) pretest, (3) education phase and (4) posttest.

Phase 1 – Registration and personal data. A first introduction about the experiment was shown to the participants on the welcome page of the web application. Furthermore, a definition of phishing and the following steps of the task were given. At first a pretest had to be done to classify the knowledge of every user, afterwards several levels to develop or strengthen the knowledge about phishing and finally a posttest for testing the learned skills. Furthermore, there were buttons to register or to login after a break. By clicking on the register-button the participant had to submit personal data like login credentials (e-mail-address and password), year of birth and gender. To provide a realistic experience the participants had to choose their frequently-used internet browser (Apple Safari, Google Chrome, Internet Explorer, Mozilla Firefox). According to the user's choice the images and address bars were personalised. In Mozilla Firefox, Internet Explorer and Google Chrome the important part of the web address is highlighted. This is not the case in Apple Safari. After submitting their data, the participants got to the pretest.

Phase 2 – Pretest. At the pretest every user was presented 28 images of a webpage with corresponding URLs. The participant's task was to identify if the shown webpage is a legit webpage or a phishing attack. Afterwards they had to state their confidence in their evaluation (on a scale of "very unsure", "unsure", "medium", "sure", "very sure"). After finishing the pretest the result was shown to the user (how many right answers they got).

Phase 3 – Education phase. After the pretest the training phase started. The participant got to an overview page of all levels, starting with a short introduction to the topic. After finishing the introduction, the task was to complete all levels. All levels are built up in the same way and cover one specific type of phishing. Every level consists of two parts, a theoretical part and a following practical task. The theoretical part explains the current phishing attack and how to detect it. After finishing the theoretical part, the user had to use his new learned skills in the practical task. An URL was presented and the participant had to identify if the URL is legit or not. After submitting the decision, the result was shown to the user. If it was a phishing URL, the following task was to highlight the who-section. Level 1 starts with the basics of a web address, the here called "who-section", the domain. Level 2 to Level 8 cover the different phishing attacks 1 – 7. At the beginning of every level the user got an overview of the so far learned phishing attacks. After finishing the practical task, the participants were shown how many questions they answered right and the following level was unlocked. The finished level was now locked so that the participants didn't have another chance to answer the questions again to get better results. After completing all levels there was an additional part, the concluding remarks which gave a short summary about phishing tricks which were not given attention to in the levels

of the web application.

Phase 4 – Posttest. After completing the education phase the participant finally had to pass the posttest. The proceeding was analogue to that in the pretest, as well the 28 items. Having finished the posttest, the participant was shown the performance for a last time.

3 Results

The data was analysed to determine if there are any differences in the test scores between the two times of pretest and posttest caused by the treatment. Because there are two times of measurements, multivariate techniques were used which allow multiple dependent variables. A within-subject multivariate analysis of variance (MANOVA) was done with 7 independent variables and 2 dependent variables. Pretest and posttest score were measured as dependent variables and phishing types 1 – 7 were measured as independent variables. Comparing the test score means in Figure 2, phishing type 5 and 6 got the worst results in the pretest and also in the posttest but with the fact that phishing type 6 documents a major increase than phishing type 5.

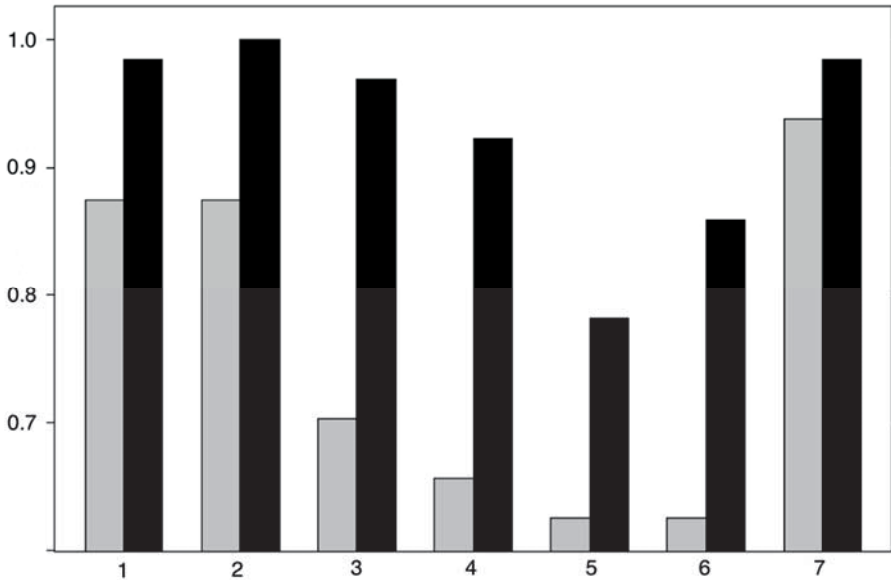


Fig. 2: Test score means of pretest (grey) and posttest (black), split up in the 7 types of phishing

Phishing type 7 reached the highest result in the pretest while phishing type 2 got the highest result in the posttest closely followed by phishing types 1 and 7. A total mean of all pretest scores was reached by a value of .74 (only test scores in a range between 0 and

1 are possible) and a total mean of all posttest scores was reached by a value of .93. Within the MANOVA the 4 following multivariate tests were done: (1) Wilks's lambda distribution, (2) Pillai's Trace Test, (3) the Lawley-Hotelling trace and (4) Roy's largest root. All of the 4 tests are based on the two matrices of "sum of squares" and "cross product". The 4 tests got significant with $F(7, 56) = 3.47$ and $p = .0037$ on a 99% level of significance. Conducting a series of follow-up ANOVAs of the dependent variables, split up in the 7 categories of phishing attacks, the values showed 6 of 7 significant results (Table 1). All of the results were significant except from phishing type 7.

phishing type	p-value	significance
1	.042	significant
2	.0071	significant
3	.000027	significant
4	.0029	significant
5	.0055	significant
6	.012	significant
7	.25	not significant

Tab. 1: Results of significance of ANOVAs

Following the principal analysis, a Welch two sample t-test was performed for testing the mean scores of pretest and posttest (of the particular phishing types) for significance. All tests were done on a 95% level of significance.

phishing type	p-value	significance	pretest score	posttest score
1	.04	significant	.88	.98
2	.009	significant	.88	1
3	.00005	significant	.70	.97
4	.003	significant	.66	.92
5	.006	significant	.50	.78
6	.01	significant	.62	.86
7	.30	not significant	.94	.98

Tab. 2: Results of Welch two sample t-test

In summary, 6 significant results were found (phishing types 1 – 6) and 1 non-significant result (phishing type 7). Furthermore, the correlations between the dependent variables were investigated, split up in the 7 categories of phishing attacks. The lowest correlation was shown between type 1 and type 6 with $r = .017$ and between type 1 and type 7 with $r = .034$. On the other hand, the highest correlation was shown between type 2 and type 3 with $r = .66$, between type 3 and type 4 with $r = .65$ and between type 4 and type 5 with $r = .52$. In general, only positive correlations appeared, varying around a mean of $r = .42$.

4 Discussion

The task of this research was to determine the efficiency of the online tool developed before. The question is whether the measurements of the test scores change over time. Non-constant measurements of the test scores would indicate influence of the training material. A MANOVA analysis of the data showed that there is a significant difference between the multivariate measurements of test scores at the 99% level of significance. Looking at the differences of individual measurements between the different types of phishing, type 1 – 6 got significant and type 7 got not significant. An overall analysis of combined pretest and posttest scores of each phishing attack showed that phishing types 1, 2 and 7 were the most often correctly detected attacks. Phishing types 5 and 6 were the phishing attacks where the most mistakes were done. Comparing the test score means, phishing types 5 and 6 got the worst results in the pretest and posttest but with the fact that phishing type 6 documents a major increase in the posttest than phishing type 5. Phishing type 7 reached the highest result in the pretest while phishing type 2 got the highest result in the posttest closely followed by phishing types 1 and 7.

The above listed results show significantly changes in the different test scores which indicate influence of the training material. Posttest scores and the results of the belonging t-tests proof that users make significantly less mistakes when they have to decide whether a website is faked or not than before using the online tool. Furthermore, it was the goal to analyse which phishing strategies are effective in fooling users and which are still effective after educating users with teaching material. Results show that especially phishing types 5 (URLs where the who-section seems similar to the real URL but use an additional term e.g. <https://www.bahn-support.de>) and 6 (URLs which contain typos <https://www.facebok.com>) were difficult to identify as phishing attacks. Regarding the collected data in total, the participants performed worst at phishing types 5 and 6, even in the posttest. This indicates that these types have to be considered seriously because they present the weakest points in users' attention. With regard to the research of Dhamija et. al (2006), nearly the same outcomes for phishing type 5 were found. About half of the presented URLs could not be identified as phishing attempts. A possible explanation for the bad performance regarding type 5 could be that users do not really know the exact URLs of their visited websites because they probably do not pay much attention to the important parts like the address bar, demonstrated in the study of Dhamija et. al (2006). A possible explanation for the bad performance at type 6 could be that URLs with typos look too similar to the real ones so that detecting the difference is rather unlikely by only having a short look at the URL. So, in future development of phishing education material special attention has to be drawn to these types of phishing. On the other hand, the posttest score means show high detecting results in phishing attacks 1, 2 and 7 which indicates that these types can be practised very well. Due to the fact that phishing type 7 got not significant in the univariate analysis and the t-test, it can be argued that type 7 (URLs which contain similar looking letters and numbers like the real ones e.g. <https://www.clropbox.com>) has a high success rate in detecting phishing URLs but additional training has no effect on the result. Because of the high scores in the pretest and also in the posttest, there is evidence

that this phishing attack is easy to discover and that this one is no effective technique to fool users. For now, there is no explanation why people are able to detect URLs with similar looking letters but on the other hand are not able to detect typos effectively. Interestingly, the study of Dhamija et. al (2006) shows completely opposed results in detecting phishing URLs with similar looking letters. With a success rate of about 9%, their results are in high contrast to this research with a success rate of about 94%. A possible explanation could be the different presentations of the URLs and corresponding websites. In this research only static images of websites were used, where Dhamija et. al (2006) presented websites with full functionality. This functionality, which made it possible for the participants to explore the website, its corresponding subpages and especially the changing might have influenced the participants' focus with regard to the important part (the address bar) for detecting phishing URLs. So, in summary only phishing types 1 and 2 can be practised well for detecting phishing URLs. The goal of this online tool is to educate users' security awareness and ability in detecting phishing URLs. Based on the total pretest score mean of all questions with .74 and posttest score mean of .93 a significant improvement can be proven. According to the research of Sheng et. al (2007) it could be confirmed that games represent a highly efficient way of information transfer.

A fact that also has to be mentioned is the decrease of participating users. 65 users signed up at our online tool but only 32 of them performed the training and the tests until the end. A possible explanation for this huge decrease is probably that people only deal with something when they are really interested in or when they are forced to. Due to the fact that one run through the tool including pretest, training phase and posttest took about 45 minutes probably a lot of a participants quit the game because they did not want to put so much time in it and unfortunately neither in further education. Another aspect which has to be considered are the correlations between different types of phishing. There were strong correlations between phishing types 2 and 3, 3 and 4, 4 and 5 which indicates that these types of phishing could influence each other, so that maybe people could be able to identify phishing URLs without the need of additional different training material.

Due to the fact that people are not regularly confronted with phishing attacks, Canova et. al (2014) analysed the effectiveness on users' knowledge retention. The results of their studies show that users of the "NoPhish" Android app are more successful in detecting phishing-URLs, particular over a longer period. Another interesting aspect for further research could be analysing the tool's effectiveness on user's knowledge retention in particular for every of the above named types of phishing. For future work, it could be also interesting to what extent the conceptual design turns out to be suitable for different types of user groups. Within the study, only the age of the participants was collected. Improved results in the posttest could be found in the age group between 50 and 60 (5 usable results) like in the age group between 20 and 40 (27 usable results). Due to insufficient available data, there cannot be provided any information regarding how far the training presents an adequate method for specific age groups.

An aspect of this study which wasn't considered are the different types of internet browsers. According to the user's choice the images and address bars were personalised. In Mozilla Firefox, Internet Explorer and Google Chrome the important part of the web address is highlighted. This is not the case in Apple Safari. It was not considered which consequence the user's choice had. By using Apple Safari, the important part is not highlighted therefore identifying the who-section was an additional challenge for these users. These difficulties could have influenced the results and could have produced more low values in the graphics. For further research, this has to be taken into account and possibly researched in an additional study. Another aspect, which was not considered, was the existing knowledge of the users. Beforehand, the participants were not asked about their knowledge of phishing. Due to this fact, test scores could have turned out better than they really are. For now, it is not possible to act on the assumption that some kind of phishing attacks is easy to detect because the high results could be also explained by the users' knowledge. But the most important point, regarding the alarmingly high statistics of APWG, is still minimizing the risk of falling for phishing attacks. Therefore, the users themselves have to be educated. The goal of this web application was to increase users' security awareness and the ability of detecting phishing URLs. This could be significantly achieved.

References

- [Ant15] Anti-Phishing Working Group (2015). Phishing Activity Trends Report. 4th Quarter 2014. Access at https://docs.apwg.org/reports/apwg_trends_report_q4_2014.pdf, 25.03.2016
- [Ca14] Canova, G., Volkamer, M., Bergmann, C., & Borza, R. (2014). NoPhish: an anti-phishing education app. In *Security and Trust Management* (pp. 188-192). Springer International Publishing.
- [DTH06] Dhamija, R., Tygar, J. D., & Hearst, M. (2006, April). Why phishing works. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 581-590). ACM.
- [DCF07] Dodge, R. C., Carver, C., & Ferguson, A. J. (2007). Phishing for user security awareness. *Computers & Security*, 26(1), 73-80.
- [ECH08] Egelman, S., Cranor, L. F., & Hong, J. (2008, April). You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1065-1074). ACM.
- [Ku09] Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L., Hong, J., Blair, M. A., & Pham, T. (2009, July). School of phish: a real-world evaluation of anti-phishing training. In *Proceedings of the 5th Symposium on Usable Privacy and Security* (p. 3). ACM.
- [Sh07] Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007, July). Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In *Proceedings of the 3rd symposium on Usable privacy and security* (pp. 88-99). ACM.

Usability-Untersuchung eines Papierprototypen für eine mobile Online-Ausweisfunktion des Personalausweises

Sandra Kostic¹ Andreas Heinemann² Marian Margraf³

Abstract: Die Online-Ausweisfunktion des Personalausweises in Deutschland besitzt eine Reihe von Usability-Schwierigkeiten und somit eine geringe Akzeptanz bei den Bürgerinnen und Bürgern. Die Umsetzung des Personalausweises in Form einer App auf einem Smartphone könnte hier Abhilfe schaffen. Mithilfe eines Papierprototypen wurden erste Usability-Untersuchungen durchgeführt, die zeigen, dass zwar die Benutzbarkeit der App gegeben ist, das Vertrauen in die Sicherheitsfunktionen der App jedoch nicht.

Keywords: Personalausweis, Mobile Online-Ausweisfunktion, Usability, Smartphone

1 Einleitung

Reisepässe und Personalausweise werden nicht nur im hoheitlichen Bereich, z.B. im Rahmen einer Grenz- oder polizeilichen Kontrolle zur Identifikation des Ausweisinhabers eingesetzt, sondern regelmäßig auch im privatwirtschaftlichen Umfeld. Die Einsatzmöglichkeiten erstrecken sich hier von stark regulierten Bereichen in denen eine gesetzlich vorgeschriebene Identifikationspflicht besteht, wie z.B. dem Banken- und Versicherungswesen, über kommerzielle und behördliche Anwendungen, z.B. Bürgerdienste, bis hin zu Geschäften zwischen Privatpersonen, z.B. beim Kauf eines Gebrauchtwagens. Mit der in den seit 1.11.2010 ausgegebenen deutschen Personalausweisen enthaltenen Online-Ausweisfunktion kann die Authentifizierung des Ausweisinhabers nunmehr auch Online erfolgen.

So ist es schon heute möglich, ein Konto bei der Bank für Investments und Wertpapiere AG oder der Deutschen Kreditbank AG mit Hilfe der Online-Ausweisfunktion zu eröffnen. Zahlreiche Bundes-, Landes- und Kommunalbehörden bieten darüber hinaus ihre Bürgerdienste (Beantragung von Geburtsurkunden, Feinstaubplaketten, Führungszeugnis etc.) ebenfalls über diese Funktion an, siehe [BMI15] für eine Übersicht.

Die heute zur Verfügung gestellte Infrastruktur für die Online-Ausweisfunktion adressiert vornehmlich eine Nutzung mittels PC. So steht die von der Bundesregierung bereitgestellte notwendige Software nur für die Betriebssysteme Mac OS X und Windows zur Verfügung ([Gov15]), Open-Source-Projekte bieten Lösungen für Linux-Varianten an. Ein weiteres Problem für die Nutzung der Online-Ausweisfunktion mit mobilen Endgeräten ist die Notwendigkeit eines Kartenlesers, mit dem die kontaktlose Schnittstelle des Personalauswei-

¹ Freie Universität Berlin, Takustr. 9, 14195 Berlin, sandra.kostic@fu-berlin.de

² Hochschule Darmstadt, Haardtring 100, 64295 Darmstadt, andreas.heinemann@h-da.de

³ Freie Universität Berlin, Takustr. 9, 14195 Berlin, marian.margraf@fu-berlin.de

ses kommuniziert. Zwar stehen Lesegeräte, die z.B. über Bluetooth mit Smartphones oder Tablets verbunden werden können, zur Verfügung ([BMI15]), allerdings ist aus Sicht der Nutzerinnen und Nutzer die Verwendung von Kartenlesern für Smartphone- und Tablet-Anwendungen noch unüblicher als mit PCs.

Ziel der vorliegenden Arbeit ist es zu evaluieren, ob ein auf einem Smartphone oder Tablet virtualisierter Ausweis (d.h. die Umsetzung der Online-Ausweisfunktion über eine App) für nicht-hoheitliche Anwendungen von Nutzerinnen und Nutzern akzeptiert wird. Zwar zeigen die Nutzungszahlen von z.B. Apple Pay (siehe [BoA16]), dass virtualisierte Sicherheitschips (hier Kredit- und Debitkarten) grundsätzlich angenommen werden. Die Frage, ob ähnliche Umsetzungen auch für von staatlichen Stellen verantwortete Lösungen ähnliche Akzeptanzraten zeigen, wird dadurch aber nicht beantwortet (siehe auch die Diskussion in Abschnitt 6).

Für die Untersuchung der vorliegenden Fragestellung wurde ein Prototyp entworfen (siehe Abschnitt 3) und einer Usability-Untersuchung unterzogen (Abschnitt 4). Eine sicherheitstechnische Evaluierung solch einer Umsetzung ist nicht Gegenstand dieser Arbeit. Eine mögliche Lösung wurde bereits in [Ot16] hinsichtlich einer sicheren Umsetzung untersucht.

Im Ergebnis (Abschnitt 5) wurde die App hinsichtlich ihrer Benutzbarkeit überwiegend positiv bewertet, es verbleiben aber Probleme, wie z.B. der aus Sicht der Nutzerinnen und Nutzer nicht bestehende Bedarf für die Nutzung und gefühlte Sicherheitsbedenken (siehe die Diskussion in Abschnitt 6).

2 Verwandte Arbeiten

Die Online-Ausweisfunktion des Personalausweises wird von Bürgerinnen und Bürgern nur sehr zögerlich eingesetzt [ID15]. So aktivieren lediglich ca. 30% der Besitzer des Personalausweises die Online-Ausweisfunktion. Die Gründe hierfür liegen nicht in sicherheits- oder datenschutzrechtlichen Problemen. Zahlreiche Untersuchungen haben gezeigt, dass die im Personalausweis zum Einsatz kommenden Verfahren sicher sind, vgl. [Be10a, Be10b, Be12]. Vielmehr zeigen Studien zur Online-Ausweisfunktion des deutschen Personalausweises, dass eine wesentliche Hürde zur Nutzung der Online-Ausweisfunktion die mangelhafte Verbreitung geeigneter Kartenleser in Haushalten ist, vgl. [As12]. Eine entwicklungsbegleitende Usability-Untersuchung der im Jahr 2014 veröffentlichten neuen Version der Software zur Nutzung der Online-Ausweisfunktion (AusweisApp2, [Gov15]) kam darüber hinaus zu dem Ergebnis, dass sich schwerwiegende Usability-Probleme aus der Gesamtinfrastruktur ergeben, die nicht allein durch die Anwendungssoftware gelöst werden können, [Wi16], siehe auch [As12]. So zeigt sich z.B., dass gerade die Verwendung von Kartenlesern zu einer sehr hohen Abbrecherquote bei bestimmten Nutzergruppen führt.

Ein weiteres Usability-Problem, das in [Wi16] aufgeführt wird, ergibt sich aus dem PIN-Management. Der Ausweis wird mit einer fünfstelligen Transport-PIN ausgeliefert. Vor der ersten Nutzung der Online-Ausweisfunktion muss diese in eine sechsstellige PIN um-

gewandelt werden. Da der Personalausweis keine Funktion vorsieht zu erfragen, ob die sechsstellige PIN bereits gesetzt wurde, kann die Software keine geeignete Nutzerführung umsetzen.

Die oben aufgeführten Usability-Probleme werden auch nicht über die Vorteile bei der Nutzung der Online-Ausweisfunktion kompensiert. Eine Vielzahl von Nutzerinnen und Nutzern sehen keinen Bedarf für diese Funktion, da ihnen geeignete Anwendungen nicht bekannt sind, vgl. [ZE16].

Eine wesentliche Voraussetzung für die Nutzung von Sicherheitsfunktionen ist die Benutzbarkeit, siehe z.B. [Wh99]. Benutzbarkeit kann aber nicht allein dadurch erreicht werden, dass Sicherheitsmechanismen ohne Nutzerinteraktion umgesetzt werden, da sonst das Vertrauen in die Sicherheitsfunktionen beim Nutzer nicht vorhanden ist, vgl. [Ru13].

3 Prototyp

Für die Untersuchung wurde ein Papierprototyp entwickelt, welcher jede einzelne Interaktion zwischen Benutzer und Software aufzeigt. Bei der Entwicklung des Prototypen wurden die Ergebnisse der in Abschnitt 2 erwähnten Studie [Wi16] berücksichtigt. Insbesondere konnten die dort aufgeführten Usability-Probleme durch das Konzept, die Online-Ausweisfunktion über eine App auf mobilen Endgeräten zu emulieren, gelöst werden (Wegfall eines Kartenlesers, bessere Umsetzung des PIN-Managements).

Um das Vertrauen in die Lösung zu erhöhen, müssen Nutzer bewusst Sicherheitsfunktionen anstoßen (Eingabe einer PIN), sind also Teil des Sicherheitsprozesses (siehe Abschnitt 2).

Die Teilnehmer des Tests haben positiv auf den Papierprototypen reagiert. Durch die Möglichkeit, direkt in den einzelnen Seiten des Prototypen Anmerkungen und Verbesserungen zu notieren, haben sich die Teilnehmer als Teil des Prozesses empfunden, ein noch konzeptionelles Produkt zu verbessern. Darüber hinaus zeigen z.B. [Sn03] und [Ri13], dass viele Nutzerinnen und Nutzer bei Einsatz von professionell wirkenden Prototypen (z.B. Prototypen, die mittels Mock-up-Tools entwickelt wurden), eine große Hemmschwelle haben, Kritik zu äußern, da gerade für technisch nicht versierte Nutzerinnen und Nutzer diese Lösungen bereits vollständig aussehen.

Gleichwohl haben die Teilnehmer den Prototypen als eine komplexe, funktionierende App wahrgenommen. In Situationen in denen der Papierprototyp eine Tätigkeit von ihnen erwartete, haben sie intuitiv richtig interagiert. Die Teilnehmer hielten zum Beispiel den Prototypen wie ein funktionierendes mobiles Gerät, um ein Foto zu machen oder verwendeten den Ziffernblock, um die für die Online-Ausweisfunktion notwendige PIN einzugeben.

So bietet er zum Beispiel ein Navigationsmenü (siehe Abb. 1) oder auch Untermenüs (siehe Abb. 2), um hier in einer Funktion *Verlauf* die ausgegebene Liste zu sortieren. Beispielhafte Anmerkungen der Testpersonen waren zu ergänzende Bedienelemente (siehe Abb. 3) oder erweiterte Funktionalitäten (siehe Abb. 4)

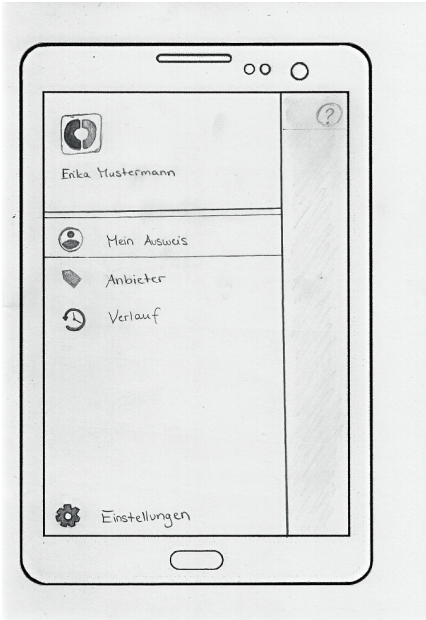


Abb. 1: Hauptmenü

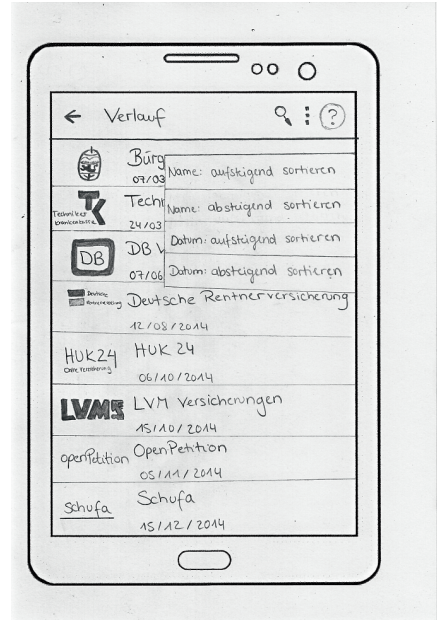


Abb. 2: Untermenü innerhalb der Funktion *Verlauf*

Die hier vorgestellten Oberflächen des entwickelten Prototypen sind ausgewählte Beispiele.⁴

4 Durchführung der Usability-Studie

An der Usability-Untersuchung haben insgesamt fünf Testpersonen teilgenommen. Nach [Ri13] ist diese Anzahl ausreichend, um im Rahmen von Usability-Tests schwerwiegende Usability-Probleme einer zu untersuchenden Lösung zu erkennen.

Voraussetzung für die Auswahl der Testpersonen war, dass sie ein Smartphone besitzen und dieses nicht nur für Telefonate nutzen, sondern mit der Nutzung von Apps auf Smartphones vertraut sind (siehe Zeile vier in Tabelle 1). Weiter wurden Testpersonen aus unterschiedlichen Altersgruppen, Geschlechtern, verschiedenem technischen Vorwissen (allgemein und Kenntnisse über die Online-Ausweisfunktion) ausgewählt. Die Online-Ausweisfunktion selbst hatte keine der Testpersonen zuvor benutzt. Die Informationen wurden im Rahmen von Interviews erhoben. Tabelle 1 fasst die Angaben zu den ausgewählten Testpersonen zusammen.

Der Usability-Untersuchung gliederte sich in zwei Teile:

⁴ Der vollständige Prototyp ist unter <https://ucs.fbi.h-da.de/wordpress/wp-content/uploads/2016/05/AusweisApp-Prototyp.pdf> einsehbar.

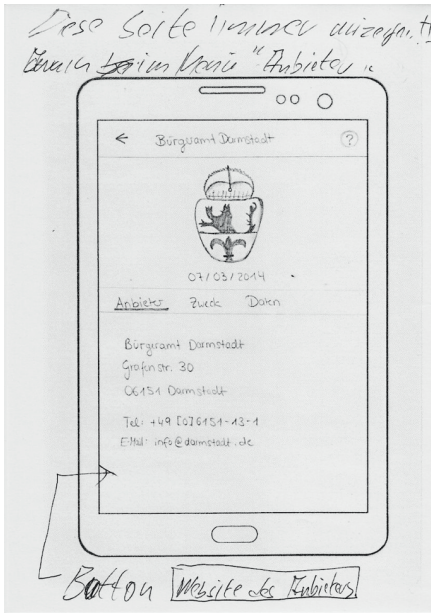


Abb. 3: Anmerkung eines Teilnehmers zur Ergänzung zum Reiter *Anbieter*

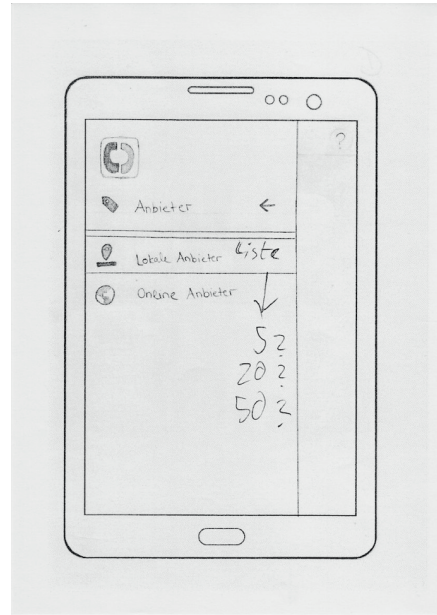


Abb. 4: Anmerkung eines Teilnehmers für die Karte der lokalen Anbieter

Geschlecht	P2, P3, P4 männlich, P1, P5 weiblich
Alter	P1: 33J., P2: 55J., P3: 56J., P4: 47J., P5: 45J., Ø 47 Jahre
technisches Wissen	stark: P2, P3, mittel: P4, gering: P1, P5
Nutzung von Smartphone-Apps	häufig: P1, P3, P4, mittel: P2, P5, gering: 0
Wissen über Online-Ausweisfunktion	stark: 0, mittel: P3, gering: P4, P5, kein: P1, P2
Nutzung der Online-Ausweisfunktion	keiner der Teilnehmer

Tab. 1: Zusammenfassung der Angaben zu den Testpersonen

- Der erste Teil* bestand aus der Durchführung von Thinking Aloud Tests an Hand des vorliegenden Papierprototypen. Dabei wurde die Installation der App auf einem Smartphone, die Hauptfunktion der App über einen beispielhaften Authentisierungsvorgang und die Deinstallation vorgestellt. Durch Nachfragen wurden die Testpersonen zusätzlich dazu angeregt, Verbesserungsvorschläge zu unterbreiten, aber auch Sicherheitsbedenken zu äußern. Anmerkungen und Verbesserungsvorschläge zum Prototypen wurden vom Teilnehmer selbst auf dem Prototypen notiert oder skizziert und in einem parallel geführten Protokoll festgehalten.
- Der zweite Teil* bestand aus der Beantwortung eines Fragebogens basierend auf der ISO-Norm 9241/110 (Grundsätze der Dialoggestaltung), [Pr09]. Auch wenn die meisten Fragen schon im Rahmen des zweiten Teils erhoben wurden, diente der Fragebogen dazu, die Ergebnisse der Usability-Untersuchung standardisiert zu präsentieren (siehe Abbildung 5).

5 Ergebnisse

Im Folgenden werden die Ergebnisse der beiden durchgeführten Usability-Tests vorgestellt. Eine Diskussion der Ergebnisse findet sich in Abschnitt 6.

5.1 Ergebnisse der Thinking Aloud Tests

Die Usability wurde durchweg positiv bewertet, siehe auch Abschnitt 5.2. Die Testteilnehmer haben insgesamt 54 Verbesserungsvorschläge unterbreitet. Diese betreffen insb. sicherheitsrelevante Fragestellungen (34 von 54). So wurde z.B. gefordert, dass die PIN nicht nur beim eigentlichen Authentisierungsvorgang eingegeben werden soll, sondern schon beim Öffnen der App (P2, P4). P2 sagte dazu:

„In der App ist mein Ausweis enthalten und damit meine Daten. Ohne die PIN hier eingeben zu müssen können ja Angreifer meine Daten sehen.“

Bemerkenswert ist die Tatsache, dass zwei der fünf Teilnehmer immer wieder von der eigentlichen Usability-Untersuchung abwichen und sicherheitsrelevante Fragen stellten (P2, P3).

Beispielhaft sei hier die Anmerkung von P4 wiedergegeben:

„Wenn aus Gründen der Bequemlichkeit bei der App darauf verzichtet wird, die Deinstallation vom Smartphone komplexer als üblich zu gestalten, besteht das Risiko, dass eines meiner Kinder diese App einfach so in den Papierkorb verschiebt! Was mache ich dann? Denn dann sind ja alle meine Daten weg.“

Oder die Anmerkung von P2:

„Der Benutzer-PIN genügt doch nicht als Sicherheitskriterium, genauso wie biometrische Daten. Das Medium Internet ist nicht dafür gedacht die Realität und Personen zu ersetzen.“

5.2 Auswertung des Fragebogens

Der im zweiten Teil genutzte Fragebogen basierend auf der ISO-Norm 9241/110 (Grundsätze der Dialoggestaltung) ist in sieben Kategorien unterteilt:

- Aufgabenangemessenheit
- Selbstbeschreibungsfähigkeit
- Erwartungskonformität

- Lernförderlichkeit
- Steuerbarkeit
- Fehlertoleranz
- Individualisierbarkeit

Die Kategorien Fehlertoleranz kann mittels eines Papierprototypen nicht sinnvoll getestet werden und wurde daher aus der Untersuchung herausgenommen. Fragen aus der Kategorie Individualisierbarkeit, wie z.B. ob sich die App für unterschiedliche Aufgaben anpassen lässt, waren nicht relevant, da ja nur eine Aufgabe (sichere Authentisierung) umgesetzt wird. Hier wurde lediglich die Frage, ob sich die App sowohl für Anfänger und Experten gleich gut eignet, untersucht. Diese wurde durchgängig positiv beantwortet.

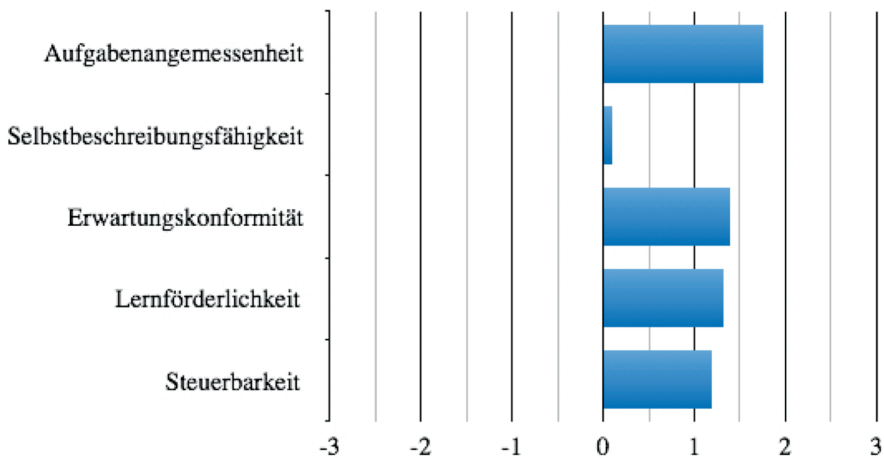


Abb. 5: Auswertung des Fragebogens (ISO 9241/110), Skala von -3 (sehr schlecht) bis 3 (sehr gut)

6 Fazit und Ausblick

Ziel dieser Arbeit war es festzustellen, ob ein auf dem Smartphone emulierter Ausweis von Nutzerinnen und Nutzern angenommen würde. Dabei wurde zunächst nur die Verwendung der Online-Ausweisfunktion für E-Business- und E-Government-Anwendungen untersucht. Die hoheitliche Ausweisfunktion, die z.B. im Rahmen einer polizeilichen Kontrolle verwendet wird, war nicht Gegenstand der Untersuchung.

Die Benutzbarkeit des untersuchten Prototypen wurde durchgängig positiv bewertet. Aus den Antworten der Testteilnehmer (siehe Abschnitt 5.1) lässt sich aber ableiten, dass das Vertrauen in die Anwendung eher gering ist. Dies hat mehrere Gründe:

- Selbst technisch nicht versierte Nutzer kennen die Gefahren, die von Schadsoftware auf Smartphones verursacht werden können. Auf der anderen Seite soll die

Online-Ausweisfunktion für Anwendungen eingesetzt werden, die ein hohes bis sehr hohes Sicherheitsniveau umsetzen müssen (z.B. Eröffnung eines Bankkontos, Abschluss eines Versicherungsvertrages, Abgabe der Steuererklärung). Dies empfanden die Testteilnehmer als Widerspruch. Hier müssen in weiteren Arbeiten Methoden (z.B. geeignete Visualisierung von Sicherheitsmechanismen) gefunden und untersucht werden, die diesen gefühlten Widerspruch auflösen.

- Die App wurde von Seiten der Testteilnehmer als Teil des Personalausweises verstanden. Insbesondere nahmen sie an, dass die App vom Staat bereitgestellt wird. An staatliche Einrichtungen wird aber mehr hinsichtlich der Umsetzung von Schutzmaßnahmen für IT-Sicherheit und Datenschutz erwartet. Zum einen, weil er als deutlich mächtiger empfunden wird als private Institutionen, zum anderen, weil staatlichen Stellen im Allgemeinen IT-Kompetenz nicht zugetraut wird. Auch hier müssen geeignete Methoden erarbeitet und untersucht werden, die Vertrauen in staatliche IT-Lösungen erzeugen.

Ein Problem verbleibt: Die Testpersonen sahen keinen Bedarf in der Verwendung der Online-Ausweisfunktion. Auch hierfür konnten mehrere Gründe aus den Antworten der geführten Tiefeninterviews abgeleitet werden:

Das Ergebnis aus Tabelle 1 zeigt, dass die Testpersonen sehr geringes Wissen bezüglich der Online-Ausweisfunktion haben. Dies liegt u.a. daran, dass das zuständige Ministerium die Online-Ausweisfunktion nur sehr wenig bewirbt (derzeit lediglich über die Internetplattform <http://www.personalausweisportal.de> und <http://www.ausweisapp.bund.de>).

Zwar gibt es bereits eine Vielzahl von Anwendungsfällen für die Online-Ausweisfunktion (siehe [BMI15]), für den durchschnittlichen Nutzer ergeben sich daraus aber lediglich ca. zwei Anwendungen jährlich (siehe [Ko16]). Umso wichtiger ist es, die Lösung so zu gestalten, dass sie benutzbar ist, also insbesondere ohne komplizierte Initialisierung auskommt und die Sicherheitsmechanismen, die eine Nutzerinteraktion benötigen, ohne Vorwissen genutzt werden können (bspw. Nutzung von biometrischen Merkmalen anstelle einer PIN).

Literaturverzeichnis

- [As12] S. Asheuer, J. Belgassem, W. Eichhorn, R. Leipold, L. Licht, Ch. Meinel, A. Schanz und M. Schnjakin. Akzeptanz und Nutzerfreundlichkeit der AusweisApp: Eine qualitative Untersuchung. Technische Berichte Nr. 69 des Hasso-Plattner-Instituts für Softwaresystemtechnik an der Universität Potsdam, 2012.
- [BoA16] Bank of America Newsroom. Bank of America Reports Fourth-quarter 2014 Net Income of \$3.1 Billion, or \$0.25 per Diluted Share. <http://newsroom.bankofamerica.com/press-releases/corporate-and-financial-news/bank-america-reports-fourth-quarter-2014-net-income-31-b>, Zugriff: 23.05.2016.

- [Be10a] J. Bender, D. Kügler, M. Margraf und I. Naumann. Privacy-Friendly Revocation Management without unique Chip Identifiers for the German National ID Card. . In: Computer Fraud & Security. Vol. 2010, No. 9. (September 2010), pp. 14-17.
- [Be10b] J. Bender, D. Kügler, M. Margraf und I. Naumann. Das Sperrmanagement im neuen deutschen Personalausweis. : Erschienen in: DuD-Datenschutz und Datensicherheit, Mai 2010.
- [Be12] J. Bender, Ö. Dagdelen, M. Fischlin und D. Kügler: The PACE|AA Protocol for Machine Readable Travel Documents, and its Security. Financial Cryptography 2012: 344-358 (2012).
- [BMI15] Bundesministerium des Innern (BMI). http://www.personalausweisportal.de/DE/Buergerinnen-und-Buerger/Online-Ausweisen/Online-Ausweisen_node.html, Zugriff: 25.05.2016.
- [Gov15] Governikus GmbH & Co. KG. <http://www.ausweisapp.bund.de>, Zugriff: 5.12.2015.
- [ID15] Initiative D21. 2015. eGovernment Monitor 2015: Nutzung und Akzeptanz von elektronischen Bürgerdiensten im internationalen Vergleich. Initiative D21 e.V., 2015, pp. 19-20.
- [Ko16] S. Kostic, A. Heinemann, M. Margraf. Nutzungspotential der Online-Ausweisfunktion aus Sicht der Bürgerinnen und Bürger. Preprint 2016.
- [Ot16] F. Otterbein, T. Ohlendorf und M. Margraf. Mobile Authentication with German eID. Extended Abstract for Presentation at the 2016 IFIP Summer School on Privacy and Identity Management, 2016.
- [Pr09] J. Prümper und M. Anft. Beurteilung von Software auf Grundlage der Internationalen Ergonomie-Norm DIN EN ISO 9241-110 (Langfassung) <http://people.f3.htw-berlin.de/Professoren/Pruemper/instrumente/ISONORM%209241-110-L.pdf>, Zugriff: 25.05.2016.
- [Ri13] M. Richter and M.D. Flückiger. Usability Engineering kompakt – Benutzbare Produkte gezielt entwickeln. Springer-Verlag, Berlin Heidelberg New York, 3. Auflage 2013.
- [Ru13] S. Ruoti, N. Kim, B. Burgon, T. van der Horst und K. Seamons. 2013. Confused Johnny: When Automatic Encryption Leads to Confusion and Mistakes. In: Proceedings of the Ninth Symposium on Usable Privacy and Security (SOUPS 13). ACM, New York.
- [Sn03] C. Snyder. Paper Prototyping - The Fast and Easy Way to Design and Refine User Interfaces. : Morgan Kaufmann, San Francisco, California, 2003.
- [Wh99] A. Whitten and J. D. Tygar. 1999. Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0. In: Proceedings of the 8th conference on USENIX Security Symposium - Volume 8 (SSYM'99), Vol. 8. USENIX Association.
- [Wi16] J. Willomitzer, A. Heinemann und M. Margraf. Zur Benutzbarkeit der AusweisApp2. In: Mensch und Computer 2016. Workshop Usable Security and Privacy: Ansätze und Lösungen zur nutzerzentrierten Entwicklung und Ausgestaltung von digitalen Schutzmechanismen. 2016.
- [ZE16] DIE ZEIT. Wo Deutschland bei der Digitalisierung lahmte. <http://www.zeit.de/digital/internet/2016-05/e-government-digitalisierung-deutschland-fortschrittsbericht>, Zugriff: 25.05.2016.

Workshop on New Security
Standards for IACS/SCADA
Industrial Systems

Workshop on New Security Standards for IACS/SCADA Industrial Systems

Jan deMeer¹, Karl Waedt²

Abstract: The IACS/SCADA-Security WS aims at Security Standards and Practice for Industrial Systems integrated by a sort of Distributed Middleware I4.0. A short tutorial into Security Standards is given by the WS Co-Chairs. In-depth aspects of this issue is discussed and presented by the invited authors from China, UK and Germany presenting: IEC 62443 Security Standards - Humans, the strongest and weakest link - Integrity Monitoring - Policy-based Monitoring - 3D-Modelling - Graded Security Forensics etc.

Keywords: Reference Architecture Model for Industrie4.0 (RAMI), Middleware, Industrial Automation and Control Systems (IACS), Supervisory Control and Data Acquisition (SCADA) Systems, Security Standards and Techniques.

1 WS General Objectives

The expected multipart standard *IEC 62443-g-p*, or *ISA99* for 'Industrial Process Management and Control' comprises $g=4$ groups with $p \leq 4$ parts each group.

Group no.1 'General' contains the parts of the terminology used, glossaries, security compliance metrics and a part with use cases; Group no.2 'policy and procedures' contains parts of security management requirements, implementation guidance, patch management etc; group no.3 'system' contains parts of security technologies, security levels for zones and conduits, security level requirements; group no.4 'component' contains the parts of product development requirements, technical security requirements; the latter part **IEC62443-4-2** currently is under discussion by ISO/IEC experts and most probably will be published during 2016 which completes the IACS series.

Those standards find their considerations by '*Industrie4.0*' but also by SCADA system developments and security evaluations. That's the main objective of this workshop to discuss the relationships between

¹ smartspacelab.eu GmbH, Berner Str. 21B,12205 Berlin, demeer@smartspacelab.de

² AREVA GmbH, PEAS-G, Henri-Dunant-Str. 50, 91058 Erlangen Karl.Waedt@areva.com

Industrial Standards addressing *IoT* vs. Industrial Systems implementing *IoT*³:



2 WS Participants Objectives

- to address the **Current Practice** of structuring, taking measures, evaluating benchmarking Industrie4.0 platforms and industrial automated control systems (IACS);
- to address Security Techniques, Architectures, Services, Features and Human-Machine-Interfaces in **Standardization** of Industrie4.0 platforms such as:
 - IEC TC65 Industrial Process Measurement, Control, Automation (IEC62443-p)
 - IEC TC57 Power System Management (IEC62351-p)
 - ISO JTC1/SC27 IT Security Techniques ISM, Process Control (ISO270 01/02/19)
 - BSI Protection Profile for Smart Grid GW, Energy Industry Act (TR03109)
 - ETSI CEN/CENELEC Smart Grid Coordination Group
 - NIST Smart Grid Interoperability Panel (NIST IR7628) etc.
- to address **Innovations** derived from features of industrial security & privacy standards and their impacts on industrial Control and Automation Systems IACS/SCADA/CRITIS;
- to address new Evaluation and Test Standards, i.e. '*Prijfnormen*', necessary to check correct implementations and impacts of **security & privacy measures in real and (ultra) large-scaled systems (ULS resp. CRITIS)**;
- to address **Laws and EU Regulations** that achieve Man-Machine Communication in the realm of '*Industrie4.0*'.

³ Copyright of left picture is reserved by IACS/SCADA WS Programme Committee Members Rainer Falk, Steffen Fries Siemens München, Copyright of right picture is reserved by Jan deMeer, ssl.eu GmbH Berlin;

3 WS General Co-Chairs

Jan deMeer⁽¹⁾, Karl Waedt⁽²⁾

- 1) smartspacelab GmbH, AIT, Berner Str.21B, 12205 Berlin,
demeer@smartspacelab.de
- 2) AREVA GmbH, PEAS-G, Henri-Dunat-Str. 50, 91058 Erlangen,
Karl.Waedt@areva.com

4 WS Programme Committee Members

Scott Cadzow, C2 Ltd. UK, ETSI TC Cyber;

Rainer Falk, Siemens AG München, Corporate Technology;

Steffen Fries, Siemens AG München, Corporate Technology;

Gerard Gaudin, EU Club R2GS France;

Hans-Joachim Hof, MuSe Munich IT Security Research Group, Munich University of Applied Sciences;

Peter Schaar, EU Academy for Freedom of Information and Data Protection Berlin;

Maik Seewald, CISCO Systems München;

Ulrich Seldeslachts, EU Club R2GS Belgium;

Kristina Unverricht, DIN Consumer Council Berlin;

5 WS Programme Structure

The IACS/SCADA Industrial Security Workshop is organized as a **half-day workshop, 2016, Sept. 27, 9h00-12h30**, with 2 main sessions, each 90 minutes and a coffee break of 30 minutes; thus giving room to 3-4 workshop presentations per session, ca.20 minutes each and sufficient time for discussions even during coffee break.

The WS raises following questions and discusses answers:

1. Do we have good IEC Industrie4.0 Security Standard?

Jan deMeer, ssl.eu GmbH et al 'New Security Standards for Automation and Control Systems, based on IEC 62443-4-2 (IACS/SCADA)';

2. How can IAC/SCADA Systems be secured by ICT?

Scott & Alexander Cadzow, C2 Ltd. UK 'Humans - the strongest and weakest link

in Securing Systems';

Mithil Parekh, OvG University Magdeburg et al.: OPANSec - Security Integrity Monitoring for Controllers;

3. How can IACS/SCADA System Security be analyzed by Formal Models?

Yan Gao, OvG University Magdeburg et al. 'SIEM - Policy-based Monitoring of SCADA Systems';

Simon Seibt, TH Nuremberg Institute of Technology et al. '3D Modelling of Selected Assets, Security Zones and Conduits';

Jianghai Li, Tsinghua University Beijing, China: 'Graded Security Forensics Readiness of SCADA Systems';

An Overview about the WS Programme can be gained from the following outline

09:00	6 - Registration & Opening Session: What does Security mean to Industries? <i>Waedt, Karl, de Meer, Jan</i> 09:00 - 09:50	1 - New Industrial Security Standards for Industrial Automation and Control Systems, ... <i>Mr. Jan DE MEER et al.</i>
10:00	7 - How can IACS be secured by ICT? <i>Waedt, Karl</i> 09:50 - 10:45	0 - Humans - the weakest and strongest link in securing systems <i>Mr. Scott CADZOW et al.</i> 5 - OPANSec – Security Integrity Monitoring for Controllers <i>Mr. Mithil PAREKH</i>
4 - Round Table Discussion with Hot Coffee		
11:00	2 - SIEM Framework for Policy-based Monitoring of SCADA Systems <i>Yuan GAO</i>	8 - How can IACS Security be analyzed by Models? <i>de Meer, Jan</i> 10:55 - 12:20
	4 - 3D and 2D modeling of selected assets, security zones and security conduits <i>Mr. Simon SEIBT</i>	
12:00	3 - Graded Security Forensics Readiness of SCADA Systems <i>Dr. Jianghai LI</i>	
5 - Wrap-up on Industrial Security WS		

6 WS Attendees Invited

- the IACS Workshop aims at practioners and engineers from Management, Administration, Security Operation, Security Incident Response Teams of SMEs and Providers of Industrial Infrastructures or Automated Control Systems;

- the IACS Workshop aims at experts from National, European and International Standardization & Regulation Alliances and Organizations such as DIN, ETSI, ISO/IEC, ANSSI, BSI, BNA, ENISA, CSA, ...
- the IACS Workshop aims at all Interested Parties, i.e. Students, Lecturers, Citizens of the Digital Society, who want to actively take part on the overwhelming industrial & societal revolution denominated as 'Industrie4.0' - Part-taking means to be a stakeholder (*Teilhaber*) and think about Regulations, Standards and IT-Laws on Privacy, Trustworthiness in Products, Built-in Security, Cyber Space Laws, Regulations and Measures to defend Cyber Crime

7 WS Background Information, Supporters and Links

<http://www.informatik2016.de/1127.html>

<http://germany.acm.org/aktivitaeten.html>

<http://www.school-of-technology.de/Club-R2GS-SoSo-english.html>



Club R2GS

New Security Standards for Industrial Automation and Control Systems, based on IEC 62443-4-2 (IACS/SCADA) - Extended Abstract

Jan de Meer¹, Karl Waedt²

Abstract: Der von Industrie und Infrastrukturbetreibern erwartete vierteilige Standard IEC 62443, bzw. ISA99 [IEC TC57/WG15, IEC SC45A/WG9, ISO/IEC JTC1/SC27/WG1] für Industrielles Prozess-Management und Prozesskontrolle (IACS) besteht aus 4 Gruppen, mit bis zu 4 Teilen je Gruppe. Nur der vorläufig letzte Teil *IEC 62443-4-2*, der von den Normungsorganisationen Anfang 2016 zur Kommentierung ausgeschrieben worden ist, spielt im Rahmen der Diskussion um das Schlagwort '*Sicherheit für Industrie4.0*' [Aca13, BSI12c, PA15, BWE16] eine wichtige Rolle für Industrieanlagen und ist daher Hauptgegenstand der Betrachtung dieses Aufsatzes. Die anderen Teile der Industrienorm IEC 62443, bzw. ISA99, befinden sich in unterschiedlichen Entwicklungsstadien. Während die älteren, bereits publizierten Teile *g-p*:= 1-1, 2-1, 2-3, 2-4, 3-1, 3-3, bzgl. ihrer Aktualität in den Gremien, begutachtet werden, sind andere Teile in Planung begriffen, wie z.B. die Teile 1-4: *IACS security life cycle use cases*, 2-2: *IACS Security Management System*, alle anderen Teile sind in Arbeit.

Keywords: Sicherheits-Standard, IACS, SCADA, RAMI, Industrie4.0, Security Operation Center, Cyber-Angriffe

1 Einführung in Industrie-System-Referenzmodelle

Die Digitalisierung hat uns die 4. industrielle Revolution *Industrie4.0* [Aca13, BSI12c, MH15] und damit eine neue Betrachtungsweise von der kontrollierten industriellen Produktion, beschert. Die Organisationen BITKOM, VDMA, ZVEI u.a. haben für das Konzept *Industrie4.0* eine 'Umsetzungsstrategie' im sog. Referenz-Architekturmodell Industrie4.0 'RAMI' [Aca13, PA15, MH15] beschrieben, bzw. formalisiert.

Das RAMI beschreibt die Elemente von industriellen Produktionsanlagen (IACS) in 3 Dimensionen:

1. Das IT-Schichtenmodell (*Layers*) mit den, von unten nach oben 6 geschichteten IT-relevanten Aspekten: *Assets, Integration, Communication, Information, Functional*, bis einschließlich *Business*;

¹ smartspacelab.eu GmbH Abt. AIT, Berner Str. 21B, 12205 Berlin, demeer@smartspacelab.de

² AREVA GmbH, PEAS-G, Henri-Dunat-Str.50, 91058 Erlangen, karl.waedt@areva.com

2. Das zweistufige *Life-Cycle und Value-Stream* Produkt-Modell, wie es in der Empfehlung IEC 62890 steht, bestehend aus den aufeinanderfolgenden Phasen: Produkttypenentwicklung, die in Produktionsprozesse von Produkttyp-Instanzen überführt werden;
3. Das 7-Hierarchien-Anlagen-Modell (*Plant Hierarchy Levels*), nach IEC 62264 oder IEC 61512, mit den technisch-organisatorischen Kontrollstrukturen von Produktionsanlagen: *Product, Field Device, Control Device, Station, Work Centers, Enterprise, Connected World*;

Das *Bundesamt für Sicherheit in der Informationstechnik* (BSI) spezifiziert in [BSI12c], welchen höchsten Kritikalitäten [JdM14], d.h. Top-10 der kritischen Bedrohungen, *Industrial Automation and Control Systems (IACS)*, ausgesetzt sein können: unberechtigte Nutzung von Fernwartungszugängen; online-Angriffen über die RAMI-Anlagen-Hierarchien; Angriffe auf Commercial *off-the-shelf (COTS)* IT-Komponenten; *DDoS*-Angriffe zur Störung von hierarchischen *IACS*-Kontrollstrukturen; Fehlverhalten und Sabotage von innen und außen einer Industrie-Anlage; Schadkode-Angriffen, mittels mobiler *IT*-Komponenten; *Man-in-the-middle (MiM) Attack*; nicht-autorisierter Zugriff auf *IACS*-Ressourcen; Angriffe auf allen IT-Schichten; Technisches Fehlverhalten und höhere Gewalt im gesamten *RAMI*-Raum.

SCADA - Supervisory Control and Data Acquisition [BWE16, TC11³] - wird seit Jahren weltweit in fast allen Arten Industrieller Produktionsprozesse verwendet. Eine sog. *SCADA*-Anwendung mißt Indikatordaten einer Anlage, um das System zu steuern bzw. zu kontrollieren. Sie besteht also aus den Teilen, (a) den überwachten Komponenten (sog. *Controls*), z.B. Verkehrsampel oder einer Energieversorgungsanlage, etc. (b) den Komponenten zur Überwachung (sog. *Intelligent Devices*), z.B. Sensoren und Aktoren (sog. *Remote Telemetry Units*) und einer Leitwarte (sog. *Security Operation Center, SCADA Master Units, ...*) und (c) einem Kommunikationsnetzwerk, das *SMUs (SOCs)* und *RTUs* im Feld miteinander verbindet.

2 Der IACS Sicherheitsstandard IEC 62443-4-2

Der vorläufig letzte Teil *IEC 62443-4-2* [IEC15a]¹ des vielteiligen 'Industrie-Standards',

³ In [16] stellt Siemens sein eigenes *SIMATIC SCADA System* für das Engineering großer Produktionsanlagen, vor. Es eignet sich als Entscheidungswerkzeug für Systemoptimierungen, es ist skalierbar zur Integration bestehender Anlagen, es kann mit der vorhandenen IT integriert werden, es bietet jederzeit Zugriff auf alle Anlagedaten. Für die Sicherheit, Zuverlässigkeit und Verfügbarkeit der Anlagedaten stehen geeignete Funktionen zur Verfügung. *SIMATIC SCADA* ist in vielen Branchen universell einsetzbar, u.a. Automobilindustrie, Stahlindustrie, Druck, Logistik, Transport, Verkehr, Öl-, Gasversorgung, Gebäudetechnik, Chemie, Pharmazeutik etc.

¹ Sowohl, im, von den Autoren benutzten Referenzdokument .../SC27/WG3 N1178, als auch im Dokument *DKE 716.0.1*, wird der Titel 'Security for Industrial Automation and Control Systems (IACS)' verwendet; Teil 4-2, worauf sich die Autoren in diesem Artikel beziehen, ist mit 'Technical Security Requirements for IACS Components' betitelt.

wird in erster Linie nicht mit Schnittstellen zwischen Mensch und Maschine (*Human-Machine-Interaction*) in Verbindung gebracht; jedoch spielt der *HMI*-Faktor dennoch eine gewichtige Rolle, nämlich dadurch, daß das Potenzial eines möglichen Angreifers auf ein reales Industrielles Automatisiertes, Kontrolliertes Produktionssystem (*IACS/SCADA*), in 4 Stufen des Aufwandes, den sog. *Security Levels* im Konzept der Cyber-Abwehr, eingeteilt wird.

Die Evaluierung Industrieller Automatisierung und Überwachungssysteme (*IACS/SCADA*), gestützt auf *benchmarking*, zielt ab, auf den Nachweis der operativen Nachhaltigkeit, bzw. des Erhalts der Funktionsfähigkeit in einer rauen Umgebung, die sich auf 4 verschiedenen Intensitätsstufen, die hier nicht weiter betrachtet werden können, stützt.

Der betrachtete, spezifische Standard IEC 62443-4-2 für die Sicherheit von *IACS* Komponenten, muss die Definition von Sicherheits-Anforderungen auf den gesamten *RAMI*-Raum (s. Kapitel 1) von Entwicklung und Produktion, plus die gesamte *RAMI*-Produktionshierarchie, vernetzt durch die IT-Schichten, erfüllen.

3 Sicherheits-Indikatorclassen für Produktionsanlagen

Der Standard-Entwurfsteil *IEC 62443-4-2* [IEC15a] spezifiziert die erforderlichen Sicherheits-Anforderungen für identifizierte *IAC* -System -Komponenten, *component requirements (CR)*, die zusätzliche Vorschläge enthalten können, wie man Sicherheitsanforderungen verbessern kann, sog. *requirement enhancements (RE)*. Die Sicherheitsanforderungen für Komponenten werden in 4 Sicherheits-Levels (*SL*) kategorisiert, welche die Anforderungen von 7 sog. grundsätzlichen Anforderungen, *foundational requirements (FR)*, erfüllen. Die *FR* stehen im Einklang mit 51 System-Anforderungsgruppen, *system requirements (SR)*.

IEC 62443-4-2 enthält Vorschläge für 7 Maßnahmefelder, *foundational requirements (FR)* genannt, die bei branchen-übergreifenden Sicherheits-, bzw. *hardening* Massnahmen oder Tests, in Betracht zu ziehen sind.

In einer angenommenen verletzlichen Umgebung ist jedes dieser 7 *foundation requirements (FR)* Angriffen unterschiedlicher Intensität, wie in den *security levels (SL)* beschrieben, ausgesetzt. *SL 1 bis 4* geben den Aufwand an, den man treiben muss, um, innerhalb einer spezifischen Sicherheitsmassnahme, *FR 1 bis 7*, alle Angriffe der Stärke *SL 1 bis SL 4* abzuwehren.

4 Ergebnisse und Ausblicke

Zusammenfassend, in Bezug auf das obige Beispiel $FR_{(j=6)}$, lässt sich sagen, dass für Tests des Maßnahmenbereichs $FR_6=TRE$, d.h. 'zeitgerechte Antworten auf

sicherheitsrelevante Ereignisse', z.B. auf der Sicherheitsstufe *SL2*:= 'geringe Motivation der Angreifer', *benchmarks*, bzw. *Audits* durchgeführt werden, die die Ausführung von *IACS* Operationen über einen bestimmten Zeitraum überwachen (*monitoring*) und über alle beobachteten Verstöße (*incidents*) gegen die Sicherheitsanforderungen Meldung geben, als auch Beweise zur Durchführung von nachfolgenden forensischen Untersuchungen sicherstellen.

In einem implementierten *benchmark* können die tatsächlichen Sicherheitsverstöße (*incidents*), z.B. für die Sicherheitsstufe *SL2*, die darin bestehen, dass nicht-autorisierte Personen oder Prozesse versuchen, sich Zugang oder Zutritt zu geschützten Daten oder Räumen zu verschaffen versuchen.

In Ergänzung zur sog. *SIEM-Landschaft* und um Ausdrucksstärke für die Definition von Sicherheitsereignissen und das Ereignis-*SIEM*-Management herzustellen, gibt es Bedarf für eine eigene '*SIEM-Sprache*', in dem besprochenen Kontext '*Common SIEM language*', kurz '*C-Slang*' (nicht Teil dieser Betrachtungen, ist aber geplanter *New Work Item (NWI)*, der *ETSI Industrial Specification Group for Information Security Indicators ISG ISI* bis Ende 2017).

In der Tat, spricht u.a. der Technische Report *WD 19608:2015-07-15* [ISO15] von einer *Common language*, die es für die Kommunikation zwischen den Systemteilhabern Konsument, Entwickler und Evaluatoren, sowohl für *security*, als auch für *privacy*, zu realisieren gilt.

Die vollständige Spezifikation der 7 *IACS*-Operationsfelder mit jeweils k_j Komponenten (s. 7 lists of *CR-benchmarks*), welche die Basis der *SIEM-Landschaft* bildet, werden in nachfolgenden normungsnahen Vorhaben behandelt und publiziert. Diese *benchmarks* sind jedoch für alle *KMUs* wichtig, weil sie die Grundlage für den sicheren und zuverlässigen Betrieb von industriellen Produktionssystemen und -plattformen (*IACS*), bilden.

Literatur

- [ISI13a] ETSI GS ISI 001-1/2 (2013) Information Security Indicators (ISI) – Indicators (INC) Part 1 „A full set of Operational Indicators for Organizations to use to benchmark their security posture“ & ISI INC Part 2 „Guide to select Operational Indicators based on the full set given in INC part 1“
- [ISI13b] ETSI GS ISI 002 (2013) Information Security Indicators (ISI) – Security Event Model (SEM) „A Security Event Classification Model and Taxonomy“
- [IEC15a] IEC 62443-4-2 Ed.1 Security for Industrial Automation and Control Systems (IACS) part 4-2: Technical Requirements for IACS components (IEC/TC57/WG15, SCA45A/WGA9, ISO/IEC JTC1/SC27/WG3 N1178 (2015-07) IT ST - Security Evaluation, Testing and Specification), https://webstore.iec.ch/preview/info_iec62443-2-4%7Bed1.0%7Db.pdf

-
- [ISO15] ISO/IEC WD19608:2015-07-08] ISO/IEC JTC1 SC27 WG3 N1193: IT ST Guidance for Developing Security and Privacy Functional Requirements based on ISO/IEC 15408
- [JdM15] Jan deMeer, ssl.eu GmbH Berlin: "Mehr Datenschutz und Betriebssicherheit durch Cyber Security Testing", ASQF SQ Magazin Ausgabe 34, März 2015, S.28-31
- [WDG15] K. Waedt, Y. Ding, Y. Gao, X. Xie: I&C Modeling for Cybersecurity Analyses, 1st TÜV Rheinland China Symposium - Functional Safety in Nuclear and Industrial Applications, Shanghai, 2015-10
- [BSI12a] Bundesamt für Sicherheit in der Informationstechnik (BSI): ICS Security top-10 Self-check - Machen Sie den Test, wie sicher ist Ihre Industrial IT?
- [BSI12b] Bundesamt für Sicherheit in der Informationstechnik (BSI): Empfehlung: IT in der Produktion - Industrial Control System Security, top-10 Bedrohungen und Gegenmassnahmen, BSI-CS100|v1.00, 29.5.2012
- [IEC15b] IEC 2015 NP 62443-3-2 - 65/611/NP - Security for Industrial Automation and Control Systems - Par 3-2: Security Risk Assessment and System Design
- [Aca13] Acatech, Forschungsunion: Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0 - Abschlußbericht des AK Industrie4.0, BMBF April 2013
- [BSI12c] BSI Empfehlung: IT in der Produktion - Industrial Control System Security - Top 10 Bedrohungen und Gegenmaßnahmen; BSI-CS10 | Version 1.00, 29.5.2012
- [PA15] Peter Adolphs: RAMI4.0 - An Architectural Model for Industrie4.0, DIN Berlin, 18.6.2015
- [MH15] Michael Hoffmeister, Festo AG&Co.KG: ZVEI The Industrie4.0 Component, Version 1.0 April 2015
- [JdM14] Jan deMeer, smartspacelab.eu GmbH Berlin: Dynamische Kritikalitätsbewertung von Prozessen in Kritischen Infrastrukturen, GI/ACM RG BB White Paper 3.2.2014, linkedin;
- [BWE16] BM f. Wirtschaft und Energie: IT-Sicherheit für die Industrie4.0; Abschlußbericht zur Studie im Auftrag des BMW, Januar 2016;
- [TC11] DPS Telecom, Fresno CA 93727-1523 USA: SCADA Tutorial Version 2.0, released August 8, 2011 - White Paper, www.dpstelecom.com;
- [Sim14] Siemens.de/scada: SIMATIC SCADA Systeme, Artikel Nr. E20001-A830-P810, Siemens AG 2014

Humans - the weakest and strongest link in securing systems

Scott Cadzow¹, Alexander Cadzow²

Abstract: Humans design, operate and are the net beneficiaries of most systems. However humans are fallible and make mistakes. At the same time humans are adaptable and resourceful in both designing systems and correcting them when they go wrong. These characteristics mean that humans can be both the strongest and the weakest link in system security. The aim of this paper is to look at how industrial control systems can use their human actors to build secure systems even noting the fallibility of the underlying machine.

Keywords: Security, Privacy, Standardisation, Human Factors.

1 Introduction

Humans design, operate and are the net beneficiaries of most systems. However humans are fallible and make mistakes. At the same time humans are adaptable and resourceful in both designing systems and correcting them when they go wrong. These characteristics mean that humans can be both the strongest and the weakest link in system security.

The set of Critical Security Controls (CSC) published by the SANS [SANS] Institute (see list below) are proposed as key to understanding the provision of security to systems. Misapplication of the controls by human error, malicious or accidental, will lead to system vulnerabilities. The importance of such controls has been widely recognised and they can be found, either duplicated or adopted and adapted for sector specific spaces, in ETSI, ISO and in a number of industry best practice guides.

- CSC 1: Inventory of Authorized and Unauthorized Devices
- CSC 2: Inventory of Authorized and Unauthorized Software
- CSC 3: Secure Configurations for Hardware and Software on Mobile Device Laptops, Workstations, and Servers
- CSC 4: Continuous Vulnerability Assessment and Remediation
- CSC 5: Controlled Use of Administrative Privileges

¹ Cadzow Communications Consulting Ltd (C3L), 10, CM21 9NP UK, scott@cadzow.com

² Bournemouth University and C3L, 10, CM21 9NP UK, alex@cadzow.com

- CSC 6: Maintenance, Monitoring, and Analysis of Audit Logs
- CSC 7: Email and Web Browser Protections
- CSC 8: Malware Defenses
- CSC 9: Limitation and Control of Network Ports, Protocols, and Services
- CSC 10: Data Recovery Capability
- CSC 11: Secure Configurations for Network Devices such as Firewall Routers, and Switches
- CSC 12: Boundary Defense
- CSC 13: Data Protection
- CSC 14: Controlled Access Based on the Need to Know
- CSC 15: Wireless Access Control
- CSC 16: Account Monitoring and Control
- CSC 17: Security Skills Assessment and Appropriate Training to Fill Gaps
- CSC 18: Application Software Security
- CSC 19: Incident Response and Management
- CSC 20: Penetration Tests and Red Team Exercises

In recognising that it is the human factor that generally identifies risk and maps out the functionality of a system - its goal in other words - it is clear that this strength can be undermined by fallibility. The question we need to ask is how do we optimise the strengths of the human element and minimise the risk they present to the system?

2 Discovery protocols in Industrial IoT for IACS

When a device is introduced to a system is a factor in the risk it introduces to the system. Specifying components, both software and hardware, during the design and initial deployment is critical in determining the security of the final deployment and the first few moments of runtime.

What needs to be asked of a system that has been designed with attributes of "Secure by default" and "Private by default" and with due attention to give assurance of the security attributes, is "is the system operating as expected?" This is where the capabilities of the controls under CSC-4, CSC-6 and CSC-19 in particular apply. However how does this

work in practice? It is essential to be able to know that what is in the system has a right to be in the system and here is where discovery protocols start to come to the fore. All that a discovery protocol does is allow a device to identify itself and its capability to other devices. This means that a well constructed discovery protocol can build an accurate contextual and system map. Furthermore when a discovery protocol is tied to a means of identifying and reporting system errors or faults the human operator is able to react. A final (perhaps), attribute of a discovery protocol is that misbehaviour, or unexpected behaviour of a discoverable element can be isolated and treated.

Discovery protocols are essential for semi- and fully-autonomous networks and systems. It is not proposed that IACS and I-IoT move towards full autonomy but it is proposed that a secure discovery protocol tied into the concepts underlying the CSCs of knowledge of what a system does and is doing through design, implementation, operation and disposal, will lend a system the ability to give benefit to the users and allow the human element to be supported in managing the system.

3 Summary and Recommendations

On the basis that humans are the net beneficiaries of IACS, and I-IoT, and that they are infallible every effort has to be expended in making it simpler to design systems that are secure by default, and that furthermore give assurance of the security offered. Assurance is needed at design time and the various assurance programs of Common Criteria and similar offer this, but we also need to give assurance of security at runtime (operational assurance) and on shutdown through disposal - lifetime assurance.

Thus the strong recommendation of this paper is to drive industrial technology towards life time assurance that supports the human user to both minimise the vulnerabilities introduced by human fallibility and to maximise the information given to the human user to resolve incidents.

References

[SANS] SANS institute, Critical Security Controls, <https://www.sans.org/critical-security-controls/>

OPANSec – Security Integrity Monitoring for Controllers

Mithil Parekh¹, Yuan Gao², Deeksha Gupta³, Christian Luschmann⁴

Abstract: Industrial automation and control systems (IACS) are more and more a combination of standardized hardware and software components that are respectively linked. A continuous increase in digital Instrumentation and Control (I&C) in production lines and critical infrastructures lead to a remarkable increase in computer-based digital security risk [IE10]. The Internet of Things, services, data and people also opens up new avenues for data theft, industrial espionage and attacks by hackers [BM16]. Currently many solutions are available for industrial network and software but more focus is still required for controller programming integrity checks.

Keywords: Integrity monitoring, programming mode authentication, whitelisting, virus protection, intrusion detection, hardening of PLC.

1 Introduction

The standard IEC 62443 allows validating potential weaknesses of automation and control technology and developing effective protection measures [IE15b]. Though this standard focuses on IT security of IACS, which are necessary for reliable and secure operation of automated systems and infrastructures [IE15c], continuous integrity checks for program and hardware configurations of controllers is somewhere neglected [TS12][LR12].

The network-based cyber-attacks on IACS got worldwide attention after Stuxnet incident [LR12]. During the life cycle management, it has been emphasized that within the plant network, controllers must be isolated from the office network [IE13]. Therefore, till now, more efforts have been put on plant networks hardening. Even after adequately securing plant networks, there could be possible weaknesses, like Stuxnet, to breach security barriers in automated plants [DREP14]. Therefore, a new approach has already been developed by AREVA GmbH which protects integrity for controllers by monitoring.

¹ Otto-von-Guericke University Magdeburg, Research Group Multimedia and Security, Universitätsplatz 2, 39106 Magdeburg, mithil.parekh@ovgu.de

² Otto-von-Guericke University Magdeburg, Research Group Multimedia and Security, Universitätsplatz 2, 39106 Magdeburg, yuan.gao@ovgu.de

³ Areva GmbH, Henri-Dunant-str. 50, 91058 Erlangen, deeksha.gupta@areva.com

⁴ Areva GmbH, Henri-Dunant-str. 50, 91058 Erlangen, christian.luschmann@areva.com

2 Literature review

There are various approaches which have been developed for security integrity monitoring for Programmable Logic Controllers (PLCs). A verification method has been developed to determine the safety and operability of PLC-based systems [MM94]. The method automatically checks sequential logic embedded in PLCs and provides counterexamples if errors are found. However, this modeling technique has been developed to verify only relay ladder logic. Another approach in [MS13] has a PLC backplane analysis system which connects directly to the PLC backplane to capture backplane communications between modules. WeaselBoard-backplane forwards inter-module traffic to an external analysis system that detects changes to process control settings, sensor values, module configuration information, firmware updates, and process control program (logic) updates [MS13]. This approach is based on real time monitoring but it is deployed on an additional workstation instead of on-device implementation. However, the approach is not practical but is recommended for understanding the properties of occurred vulnerabilities. PLC vendors normally do not guarantee for security if a backplane is used for capturing communication. But, OPANASec, as real-time and on-device security solution for PLC, takes just one cycle time for annunciation when the PLC integrity is being compromised.

3 Security integrity monitoring scope

With the advance of Industrial Ethernet solutions, increased networking with the office world and a large number of unprotected interfaces at the field level, security is of greatest importance. Therefore, knowledge of permitted communication relationships is a fundamental prerequisite for secure networks [IE10][SN13].

3.1 State-of the art for connecting office floors to automation facilities

In Fig. 1, a fundamental architecture is used for security solutions, especially to isolate industrial networks from office networks. The communication between two security zones (a concept introduced by IEC 62443) can be controlled by configuring a firewall and providing an additional zone which could be a Demilitarized Zone DMZ, as also described in IEC 62443. As indicated in Fig. 1, different security solutions like whitelisting, virus protection, and patch management are available for general IT application software. Some of these, like whitelisting, are also applicable for digital automation systems containing PLCs.

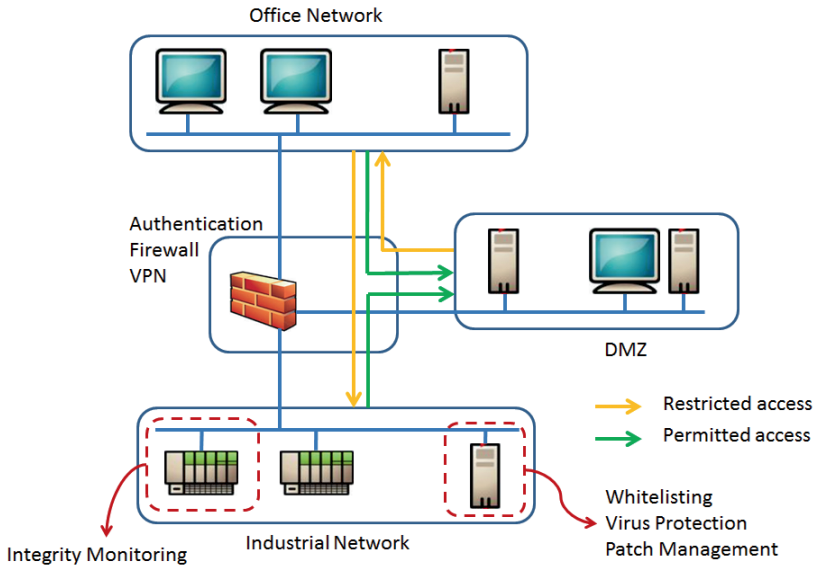


Fig. 1: Logically isolating plant controller by Firewall

3.2 Graded connections for high level Security Zones

Fig. 2 illustrates different graded security solutions for securely interconnecting industrial networks that are assigned to different logical security zones. Fig. 2(a) shows an elementary scenario without any type of control over the communication. This may be extremely dangerous and should usually be avoided for communication between different security zones. In the example in Fig. 2(b), the communication goes through a firewall, a scenario that was already elaborated in Fig.1. Particularly, SIEMENS recommends their own line of “Scalance S” security modules and “communication processor” for its platform. Difference between these modules and office-type devices is that these modules are hardened for use in industrial environments (IP30) and they are optimized for communication of process control information [SP12].

If two zones are physically separated without any hardwired interconnections or spanning networks, (non real-time) communication is still possible via an “air-gap connection”. In this case, which is frequently encountered at legacy systems, a portable storage device (e.g. a USB key) can be used for information exchange. The portable storage device may feature support for authentication, see Fig. 2(c).

A further, often considered more expensive solution, is the deployment of physically unidirectional security gateways, also known as data diodes. These enforce the unidirectional communication at the hardware level.

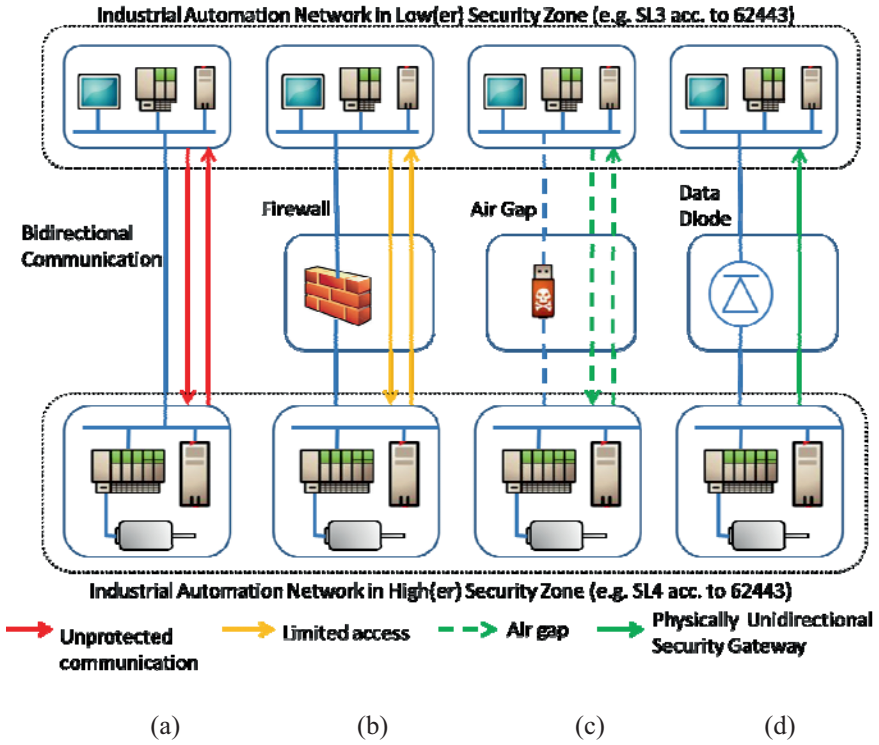


Fig. 2: Graded enforcement of communication restrictions between security zones

3.3 Secure connections originating from lower level Security Zones

Beyond the primary concern to isolate high security level zones, there are still many use cases where communication is required from site local intranets (lower security level) towards automation networks. Fig. 3 indicates a scenario in which a data diode is used to transfer set-points configuration files from a site-local intranet towards an automation network located in a higher security level zone.

This approach is resilient against many attack scenarios. While it provides a continuous communication, control and command loops that would allow a potential control from an outside location are not permitted.

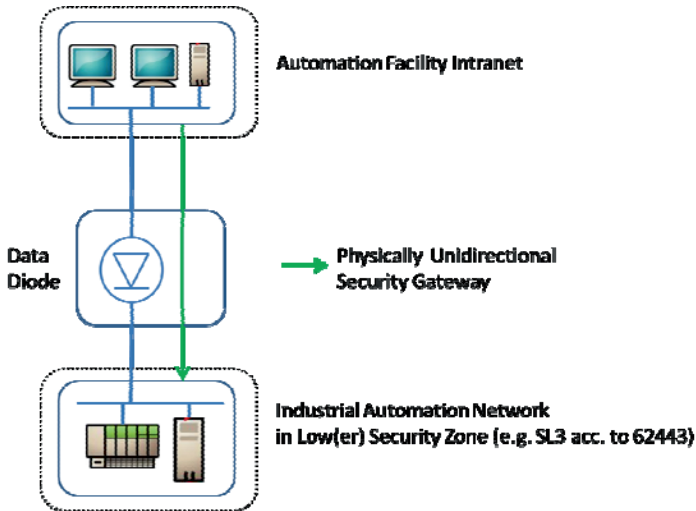


Fig. 3: Communication between Automation facility intranet to lower security zone

Additional verifications of the data received via the data diode can be performed based on the formal description of the content (e.g. simple syntax and semantics of a set-points file) using e.g. FPGA-based solutions, like SCOOP from seclab [SCO1].

3.4 Enforcing security integrity of automation systems

3.4.1 Motivation for the deployment of OPANASec

In an automated plant, the process control logic is mission- and, sometimes, life-critical, and is typically rigorously tested prior to deployment. It's remarkable that integrity checks are not performed on controller logic [WSGND]. Even if the PLC designers have not considered a malicious code threat, unintended changes to logic through memory manipulation and human error can have similar consequences.

OPANASec checks integrity of the application software and hardware configurations (including function block libraries) that run on PLCs. Therefore, it is suitable solution against types of attacks such as Stuxnet.

3.4.2 Existing security integrity monitoring solutions

Some security monitoring solutions already exist in current plant networks like whitelisting, anti-virus and patch management.

Application and file operations can also be considered by using whitelisting for controlling malicious activities. In this case, only specified, trusted applications can be run or only specified file operations can be performed. However, controlling a

manipulation of the application programs and hardware configurations of PLCs is not implemented today as a standard. Therefore, this can be an ideal time to introduce OPANASec as a state-of-the-art security solution for PLCs.

Patch management is used as a security measure to protect applications of all security techniques described in this security concept. For example, a potential attacker must first overcome multiple security barriers before weak points from a lacking security update can be exploited [IE15].

However, unauthorized changes to PLC logic are not controlled here by assuming inherent trust in those possibly compromised systems.

4 The OPANASec approach

The OPANASec solution is the implementation of an access control and integrity monitoring solution for SIMATIC S7 Programmable Logic Controllers.

It consists of two function blocks. They are added to the existing PLC program and can be used in CFC (Continuous Function Chart), as well as FBD (Function Block Diagram), COP (Contact Plan) or STL (Statement List). The function blocks continuously monitor the user program of the PLC, the hardware configuration of the PLC and also the configuration of network connections for changes. The changes are detected by checksums over the relevant data. If a change is recognized, it is stored to the non-volatile diagnostics buffer and alerted at the output. The output can be used for alerting, signaling, shut down, etc. To avoid deletion of the integrity monitoring solution by an attacker (virus, malware, human), the two function blocks of the solution monitor each other, and if any block fails, it is recognized and an alert is created as presented in Tab. 1.

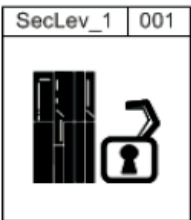
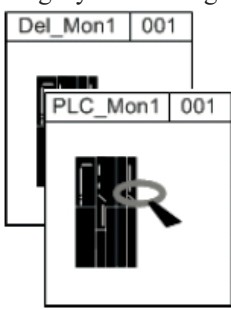
4.1 Key features

Some of the OPANASec key features are:

- Easy integration into existing PLC programs in STL/FBD/COP [IE15a]
- Free usage of output signals (HMI/SCADA, Annunciation Lamps, SMS, etc.)
- Part of AREVA's NPLib (Nuclear Power Library) for S7, IEC62138 certified by TÜV Süd
- AREVA standard solution for S7-300/S7-400 including F/H-Systems.

4.2 Limitations

Currently this is only implemented for SIMATIC S7 PLCs.

<p>Access control</p>  <p>SecLev_1 001</p>	<ul style="list-style-type: none"> • Key-Switch for blocking of program changes • Effect on the whole program (operational & failsafe) • Two factor identification in combination with engineering environment
<p>Integrity Monitoring</p>  <p>Del_Mon1 001</p> <p>PLC_Mon1 001</p>	<ul style="list-style-type: none"> • Online-Recognition of program changes on the PLC (including operational program code) • Online-Recognition of configuration changes (Parameterization of modules, network configuration) • Annunciation by SCADA and/or binary outputs • Logging of security events in diagnostic buffer and instance data block of PLC • Reciprocal self-monitoring of PLCMon and DelMon including annunciation of manipulation- and deletion attempts

Tab. 1: Two aspects of OPANASec

5 The OPANASec implementation

The integrity monitoring is used together with another function block of the OPANASec solution, which implements a control of the programming mode (Access control). The “SecLev” function can lock the PLC, so programming is prohibited. If programming mode is needed, the PLC has to be unlocked. This task can be completed only by the “SecLev” function block itself. Thus, a signal that is given to the function block e.g. by a key-switch in the cabinet or remote signal from a control room is used. Only if this particular signal that allows programming is present, changes on the program, HW configuration or network connection configuration of the PLC can be performed. Together with a password in the engineering environment (engineering software), that means a two factor identification ensures secured implementation of programming.

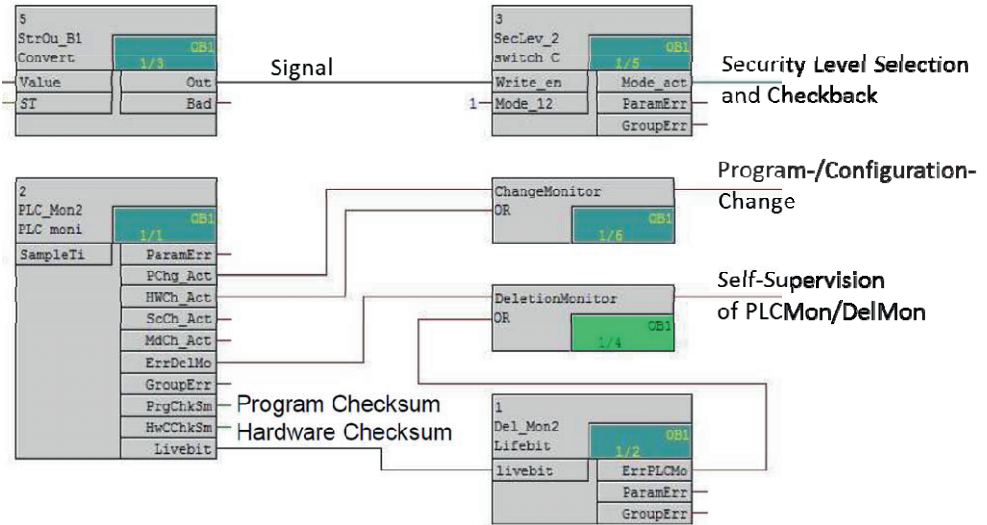


Fig. 3: Implementation of OPANASec

Only after access control through the “SecLev” function block is granted, a change/deletion in the PLC logic is possible. At each modification, an output signal from “ChangeMonitor” is generated and with this signal, it can be confirmed that an appropriate authentication is given to particular personnel, see in Fig. 3. Furthermore, using “Livebit”, both function blocks monitor each other against deletion. An annunciation is generated from “DeletionMonitor” if any of them is attempted for deletion.

6 Enforcing consistent security solutions for the process industry

The OPANASec approach described in the previous sections is effective in a framework of function diagrams and function blocks, as it is typically used in the process industry and for power plant automation. The difference, as compared to completely manual software development, is in the use of a comprehensive set of tools that generate most of the source code out of semi-formal graphical specifications.

The OPANASec function blocks make use of this framework to assure the software code integrity by a specific set of function blocks. Beyond this security integrity enforcement, further guidance on the design and interconnection of function blocks can be used similarly to security policies for secure code development. This guidance is at a higher level, as compared to source code level development details, like potential buffer overflows, handling of array indices or unsigned integer numbers arithmetic. An example of such guidance could be the recommendation (for lower security levels) or

requirement (for higher security levels) to always accompany a signal value with a signal status. This obviously has a benefit for safety, as the signal status may be considered (as part of the “active status processing”) in decisions on whether to rely on the current signal value. Additionally, there are benefits for security, in case the status processing is performed at all stages starting at the smart signals that collect and distribute the initial values, over the network devices that receive and forward the values within application level data messages. Thus e.g. a denial of service attack at a network (between the smart sensor and a processing unit) would potentially result in obsolete data. However, the signal status can be used to indicate this (invalid / obsolete data), so that the obsolete signal value can be excluded from further computations that would result in the actuation of physical aggregates, see Fig. 4.

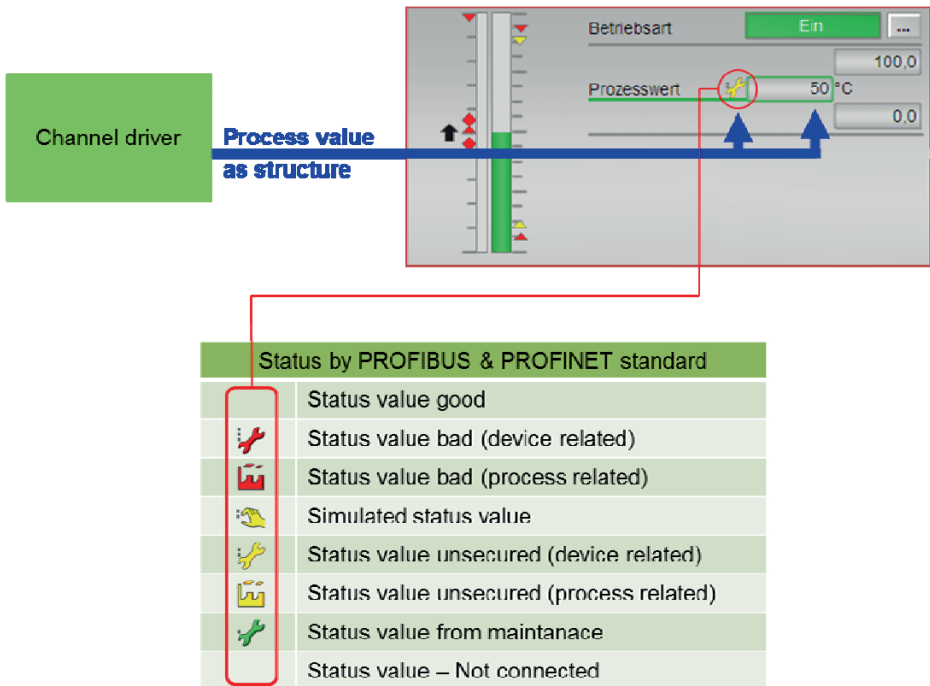


Fig. 4: NPLIB Signal status

Such a library, called NPLIB (for Nuclear Power Library) is mandated by AREVA GmbH for selected applications, even if the implementation is performed by a subcontractor. This assures that, at a higher level of the data flow, appropriate security policies can still be implemented, based on assumptions that are reliably enforced. Otherwise, the different subcontractors may not intend or may not even be able to provide the needed security assurance. This is a supporting step for implementing the

IEC 62443-2-4 [IE15a] requirements for service providers and also in line with [II14].

7 Conclusion

Continuous Security Integrity Monitoring can be performed with efficient and effective technical security solutions like OPANAsec. In order to enforce security in line with the Security Levels (SL1 ... SL4) of IEC 62443, additional solutions were addressed, like DMZ and Physically Unidirectional Security Gateways. While the use of physical data diodes may be more expensive than “air gap” connections, they may be mandatory for security conduits leading from high security level zones or in cases where real-time requirements cannot be met via manual “air gap” data exchange.

The function blocks that implement the OPANAsec solution are integrated into a certified function blocks library, called NPLIB. Enforcing the use of this library by internal engineering staff and by external engineering service providers assures a consistent level of error handling, e.g. signal status based, together with the provision of non-circumventable security controls.

References

- [BM16] Studie im Auftrag des Bundesministeriums für Wirtschaft und Energie - IT-Sicherheit für die Industrie 4.0 - Produktion, Produkte, Dienste von morgen im Zeichen globalisierter Wertschöpfungsketten, 2016.
- [IE15] IEC 62443-2-3: Security for industrial automation and control systems – Patch management in the IACS environment, Edition 1.0, 2015.
- [IE15a] IEC 62443-2-4: Security for industrial automation and control systems – Security program requirements for IACS service providers, Edition 1.0, 2015.
- [IE10] IEC 62443-2-1: Security for industrial automation and control systems – Establishing an industrial automation and control system security program, Edition 1.0 2010.
- [IE13] IEC 62443-3-3: Security for industrial automation and control systems – System security requirements and security levels, Edition 1.0, 2013.
- [IE15b] IEC 62443-3-2: [Draft] Security for industrial automation and control systems – Security risk assessment and system design, 2015.
- [IE15c] IEC 62443-4-2: [Draft] Security for industrial automation and control systems – Technical security requirements for IACS components, 2015.
- [IE13] IEC 62443-1-3: Industrial communication networks - Network and system security - System security compliance metrics, 2013.
- [WSG10] Wang, J; Stavrou, A; Ghosh, A.: HyperCheck: A Hardware-Assisted Integrity Monitor, 2010.

- [ACM15] Sadeghi, A.; Wachsmann; C.; Waidner, M.: Security and Privacy Challenges in Industrial Internet of Things, 2015.
- [DREP14] Hadžiosmanovic, D.; Sommer; R.; Zambon, E.; H.Hartel, P.: Through the Eye of the PLC: Semantic Security Monitoring for Industrial Processes, 2014.
- [SN13] SIMATIC NET - DMZ with the SCALANCE S623, 2013.
- [SCO1] SCOOP, <http://www.seclab-solutions.com/use-cases/prevent-data-leakage>, 2016.
- [II14] ISO/IEC 27036-1: Information technology — Security techniques — Information security for supplier relationships — Part 1: Overview and concepts, 2014.
- [MM94] I.Moon: Modeling programmable logic controllers for logic verification, 1994
- [MS13] Mulder, J.; Schwartz, M.; Berg, M.; Van Houten, J.; Urrea, J.; King, M.; Clements, A.; Jacob; J.: WeaselBoard: Zero-Day Exploit Detection for Programmable Logic Controllers, SANDIA REPORT, 2013
- [SP12] SIMATIC, Process Control System PCS 7, Security concept PCS 7 & WinCC (Basic), Function Manual, 2012.
- [TS12] Tyson, M; Singer, B; Cybersecurity for Industrial Control Systems SCADA, DCS, PLC, HMI, and SIS, CRC Press, USA, 2012.
- [LR12] Langner, R; Robust Control System Networks: How to Achieve Reliable Control After Stuxnet, Momentum Press, 2012.

SIEM: Policy-based Monitoring of SCADA Systems

Yuan Gao¹, Xin Xie², Mithil Parekh¹, Edita Bajramovic³

Abstract: Security Information and Event Management (SIEM) systems work on SCADA systems by observing and reacting to the dynamic security-related events of the target automation system. These events are created by collecting/filtering raw logs maintained by its sub-components. Preferably, logging items are attached with synchronized timestamps. Specific data of security-related event can be correlated and analyzed as security measures of the SCADA system. Possible correlation rules represent the power of SIEM system for handling security dynamics. A SIEM system can help to recognize security breach in a short time and optionally can react to the breach automatically. SIEM systems are able to monitor the system's residual risks, while continuously track the deployed security controls and measure their effectiveness. In this paper, we proposed the common requirements of a SIEM system and discussed its important enhancements within the context of SCADA systems. The SIEM system can be supported by the overall security model and designed in a model-driven manner. At last, this paper proposes a preliminary model of correlation rules.

Keywords: cybersecurity, SIEM, SCADA, log management, real-time system, network security, security modelling, security testing

1 Introduction

An information security management system (ISMS) [II13] is required by an industrial automation and control system (IACS) [Ie15]. Especially, the supervisory control and data acquisition (SCADA) system, which acts as the major type of an IACS, needs to be protected by ISMS too. For identifying as well as mitigating vulnerabilities of a SCADA system, security experts perform risk assessments according to the system's architecture and configurations. Vice versa, forensics investigation could be performed when a security breach in the daily operation is discovered, to determine how the breach happened as well as to collect the digital evidence. Both the risk assessment and the forensics investigation approach work either on a static system specification in its design phase or on a static snapshot of the system. Compared to this, the security information and event management (SIEM) system takes the responsibilities to handle the dynamic security statuses of a SCADA system. Within the run-time environment of the system, the SIEM system is able to monitor its residual risks and to track the triggered security

¹ Otto-von-Guericke University Magdeburg, Research Group Multimedia and Security, Universitätsplatz 2, 39106 Magdeburg, yuan.gao@ovgu.de

² Siemens AG, Digital Factory, Siemensallee 84, 76187 Karlsruhe, xin.xie@siemens.com

³ Friedrich-Alexander University Erlangen-Nürnberg, Informatics 1 Department - IT Security Infrastructure, Martensstr. 3, 91058 Erlangen, edita.bajramovic@gmail.com

events on-the-fly. Besides [Ie15], SIEM is also addressed in the security context of Industry 4.0 [Bu16].

Run-time events of an automation system carry the system’s statuses as well as their transitions in between. The logging functionality of the system will continuously record these events, typically with timestamps. One major purpose of recording these events is the diagnostic of the system. Furthermore, security-related events are part of these run-time events and are especially associated to the system’s security statuses, e.g. a locked door is opened with an employee card. Either a normal operation or a security breach can trigger multiple security-related events. Furthermore, among them, the event that indicates a possible security breach is named as security event (SE) [II11]. On one side, the SIEM system collects and analyzes them (in real-time) to create security warnings/reports which support the daily work of the incident response team (IRT). On the other side, with the record of system’s executions, the SIEM system can also assist security experts to perform risk assessments as well as support forensic investigations with strict constraints.

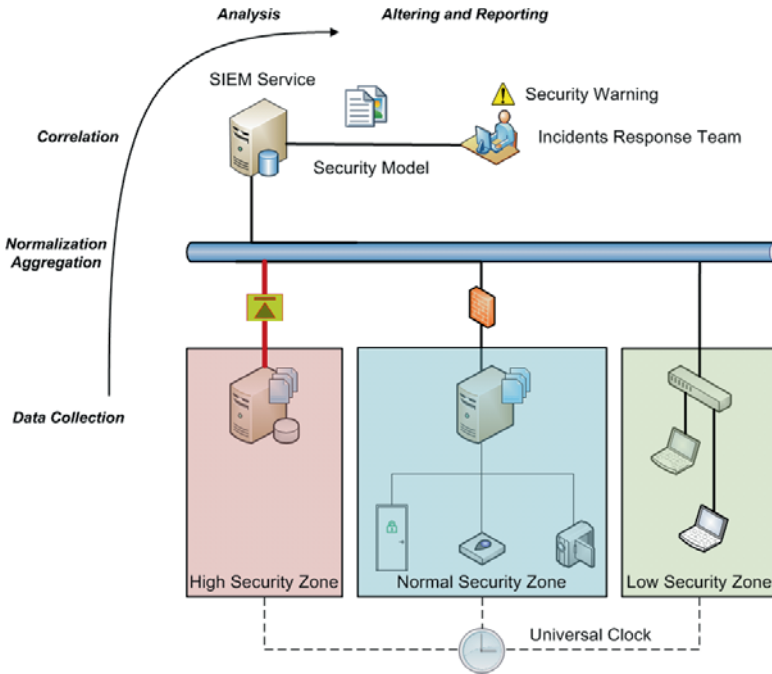


Fig. 1: A SIEM prototype and the associated data flow.

In this paper, we stick to the context of a SCADA system. However, with proper adaptations, the same design concepts of SIEM systems can be applied on a general IACS or its other sub-types, like distributed control systems (DCS) as well as compositions of Programmable Logic Controllers (PLCs) [Ni11].

2 Common Features of SIEM systems

Fig. 1 (partially based on preliminary work at ISO/IEC JTC1/SC27 WG4) illustrates a SIEM prototype and its associated data flow. Security-related events are created from logging items belong to separate components of a SCADA system. In this paper, we assumed that all logging items are already timestamped by its logging system. These event specific data will be collected to a central SIEM service for further processing. The curve in the left part of the picture shows the major steps of SIEM processes which are further explained in Section 2.4.

2.1 Timestamp of Security-related Events

For meeting the real-time impacts of SCADA systems, the timestamps attached to security-related events have to share a predefined accuracy across different sub-components. As depicted in Fig. 1, a universal clock is ideal for the synchronization task. However, the universal clock might be unavailable since it is not mandatory for a SCADA system. Normally each sub-component has its own clock. In this case, proper protocols need to be considered for the reliable time synchronization between different logging systems [St10]. Beside the time differences among multiple logging systems, the time accuracy also needs to be considered. When several events happen in the same minimum time unit (e.g. in one second), their sequence cannot be determined.

Within the same logging system, the timestamps bring the most important information to indicate the chronological sequence of events. Based on concrete implementations of logging systems, it might happen that events happened later were recorded in the log file priority to the events happened earlier. However, their sequence can still be determined according to the associated timestamps. When the execution time between two relevant events are not so important as well as these events are recorded in the log file strictly chronically, these events can be processed without timestamps.

A SIEM system can be created by introducing accurate time synchronization between sub-components in the design phase. Or with existing logging systems, the SIEM can be designed to be able handle timestamps in a fuzzy way which means inaccurate timestamping of events is allowed.

2.2 Collection of Event Specific Data

For correlating security-related events originating from different components, the relevant SIEM data, either events or raw logs, will be collected to a central service for further processing and analyzing. On one hand, the created security-related events need to retain the references to their raw log items, which is important for fulfilling the requirement of forensics investigations. On the other hand, the transfer of SIEM relevant data should neither overload the SCADA system nor breach network security controls.

More considerations about the transfer security are discussed in Section 4.2.

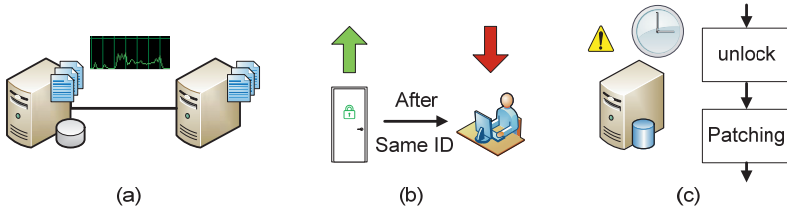


Fig. 2: Examples for correlation rules:

(a) abnormal network traffic (b) logon after leave out (c) configurations off working hour

2.3 Correlation of Security-related Events

Correlation is a key functionality of SIEM systems. Security-related events or SEs from the same component or across the considered SCADA system will be combined following predefined correlation rules. The historical correlations can back-track available security-related events in a long time to detect abnormal behaviors. For example as Fig. 2a, too much network traffic is observed or the statistic distribution of message types changes dramatically. Both of cases might indicate an attacker joined the network and manipulated the delivered messages. Furthermore, the example introduced in the preliminary work at ISO/IEC JTC1/SC27 WG4 (depicted in Fig.2b) also shows the potential of correlating events come from different components in real-time. In the example, two independent SEs of different components are taken into account:

1. The ID card of an employee was recorded by the entrance access control system that the person left the company.
2. After 30 minutes of the leaving event, a legal system logon action with the user/password belong to the same employee in the server room was logged.

Each of the two SEs complied with its relevant security controls. However, combination of the two events within a short period indicates a high probability of a security breach. An additional correlation example as illustrated in Fig. 2c could be the maintenance engineer patching the current system: when cabinet lock-monitoring system is in place, according to the flow chart, the uploading action should happen after the unlocking event of the cabinet. A reverse order or a missing of the unlocking event might indicate a compromise through network. This example can also be linked to a working schedule of the facility. When the uploading action happens after the regular working time, it is very likely an attack is ongoing.

In the correlation examples, different security systems and their associated controls as well as relevant security objects are involved. By utilizing the overall security model [Wa16] which has them already in place, correlations rules can be described by modelling the relations among them.

2.4 SIEM Process Steps

The curve in the left part of Fig. 1 shows the major steps of the SIEM processes. In this section, individual steps are introduced according to the sequence indicated by the arrow. Meanwhile, the requirements and potential risks of the SIEM system are further discussed. Especially, in parallel to these process steps, the raw logs need to be archived for later on tracing back or for possible forensic investigations.

Data Collection: The system's security-related events are collected from the raw logs of its sub-systems. These system logs and their associated logging functions are designed for either diagnosis purpose or compliance with security controls [Ie15]. Thus, the considered system logs firstly need to be examined whether they contain sufficient information as described in Section 4.1. Furthermore, during run-time of the SIEM system, log filtering is required to collect only security-relevant logging items.

Normalization and Aggregation: In this step, logging items from different sub-systems thus using various formats will be unified. The extra or duplicated information which is not required for the future steps will be reduced. Especially, in the normalization step, the associated timestamps need to be transformed into the same time system. At last in this step, security-related events will be created from the unified logging items.

Correlation: The correlation of security-related events is introduced in Section 0.

Analysis: In this step, correlation results can be linked to relevant security objects of the overall security model described in [Wa16]. With the model's support, potential security breaches (e.g. the abnormal logon action) could be discovered. Furthermore, the SIEM system will monitor the SEs associated to known vulnerabilities, thus to alert the incident response teams in advance. Meanwhile, the invoked SEs will be linked to applied security controls through the model. When a specific security control is triggered too often, then smart testing can be deployed on the associated security object to assist improving the system security.

Alerting and Reporting: The analysis results and possible subsequent system changes will be ordered according their timestamps. Alerts will be automatically sent to the IRT to acquire attention while the attached analysis report will assist the security staff to figure out the system vulnerabilities under attack and to deploy countermeasures.

Raw Log Archiving: Additionally, the raw log needs to be archived. On one hand, since the logging system could be interrupted to cause damage to log files or missing of events, log files need to be archived regularly. On the other hand, considering some sub-components like PLCs have only limited resources, the size of current log file is limited thus its content needs to be archived to a permanent storage. The archived raw logs can be utilized by the tracing back functionality or for fulfilling the regulation requirements on critical infrastructures to keep raw information for forensics investigations.

3 Key Features of Commercial SIEM Systems

Considering the amount of logging data that may be generated in real-time by SCADA systems, a commercial SIEM system is expected to be able to filter and search security logging information with precisely defined log file formats. Besides, the data selection process needs to cooperate with real-time system events. SPLUNK [Sp16] is one of the popular commercial solutions for the logging management of SCADA systems. Besides the ability to handle big amount of log data, it also supports different storage types such as cloud-storage, which brings more flexibility to various SCADA systems.

Other solution, the IBM QRadar also demonstrates the idea to connect the SIEM system and the IT Infrastructure under monitoring [Ib16]. It focuses on the network security while enables real-time correlation to identify high risks as well as security breaches. Since the associated firewalls as well as network monitoring systems are already policy-based, QRadar can be configured / extended for automated incidents response and the vendor announced to comply with regulatory requirements on data collection and secure reporting.

Furthermore, combining different existing tools is also an option. For example, a SIEM system could use SPLUNK to create real-time security-related events from a great amount of logs while data mining tools can be employed to build correlations from the big data set. Even different SIEM systems can be combined: for example, office IT has already SIEM from Vendor A while the SCADA system is associated with another SIEM system from Vendor B. These two SIEM systems can be combined together and share selected data between them. Furthermore, all commercial SIEM systems are able to receive threat intelligence. With this information SIEM system can always indicate the new threat. More attractively, the correlation and associated security objects could base the unified security model [Wa16] thus to implement a model-based SIEM system.

4 SIEM Framework for use in Industrial Automation

4.1 Policy on generation of Security-related Events

The security-related events created from raw logs have to fulfill the data requirements of the SIEM system. Considering different logging systems which have various non-standard data format, this is not a trivial task. In this paper, we gave the most important required information of a security event for guiding the generation process. Tab.1 lists the required attributes: *ID* is the unique identifier of the security event. *Timestamp* Attribute records when the event happened. When no universal clock is used, its specified clock reference needs to be attached. *Type* indicates which status changes happened in the system. For example, a booting event or a shutdown event has clear semantics relevant to the system. In the next, *Source Information* maintained the link from the security event to its raw log item. Finally *Invalidation Conditions* defines when

the security event turns to be invalid. The condition could be the arrival of a relevant security event or an expiration time is reached in a real-time system. For example, the close event of a door will terminate the open event of the same door. An authentication event should be valid only within a predefined amount of time.

ID	Timestamp	Type	Source Information	Invalidation Condition(s)
----	-----------	------	--------------------	---------------------------

Tab. 1: Essential Attributes of a security-related event.

4.2 Policy for secure collection of SIEM relevant data

The collection of security-related event specific data should not bring additional risks to the SCADA system. On one hand, the traffic of SIEM data collection will consume extra network bandwidth thus it needs to be well designed to avoid impacts to the system's normal operation. On the other hand, the SIEM central service collects data from components by a read-only manner. According to [Ie13], the communications between different security zones have to be constrained regarding to their security levels. As shown in Fig. 1, communications from high and middle security zones need to be protected against leaking to lower security zones through the SIEM service. The high security zone can be protected by data-diode while the middle security zone can be protected by firewalls configured with access restrictions, which acts as a physically unidirectional security gateway.

Besides the impact on communications, the secure collection of data requires appropriate configuration management of sub-components. The agent-based solution needs to install specific software on sub-components to handle the collection. One obvious advantage is the agent can be configured to periodically push log data to the SIEM service. However, considering attaching a SIEM system to an existing SCADA system, the agentless solution requires no modification of sub-components thus more compatible. The drawback is that for the SIEM to collect log data from a sub-component, a temporary user/account for connection is required and might introduce extra risks.

For protecting the target SCADA system, the integrity of raw log needs to be protected from tampering by indicating devices [MT15]. This will also help to prevent the manipulation or misleading of the SIEM system. The tamper resistance measurement can already be involved in the logging system. For example, in a CCTV system with long-term storage, the id of records should contain a continuously increased number. More concretely, assuming the CCTV system will record the internal vision of Room 1 every minute and save the video with the name contains the recording time (see Table 2.). A possible tampering method could be achieved by the attacker: firstly delete the record file in a previous time which contains criminal evidence. Then the system clock will be modified back to that time and the new automatically generated record will have the same name as the deleted one. At last, the system clock will be recovered to the normal status. However, as the example shown in Tab. 2, the record file name started with a continuously increased number. The file name of the new generated record starting with

250 is different with the illegally deleted one (starting with 100). Thus by checking the prefix number, the SIEM system can discover whether some records are removed illegally.

Record 1 (deleted)	Record 2 (new generated)
100 Room1 2016-06-15 12:38	250 Room1 2016-06-15 12:38

Tab. 2: An Example of Naming Convention for CCTV Records [Mo96]

Additionally, the tampering resistance can be addressed by applying correspondent storage security controls. The international standard [II15] demonstrates security controls as well as guidelines for storage security. Since in a SCADA system, the SIEM relevant raw logs are normally implemented as file-based storage, this standard is applicable for the security considerations of protecting them against tampering. For example, if the raw log is stored within the sub-component, the security controls of Direct Attached Storage (DAS) can be applied. More effectively, considering the data collection scenarios in a SIEM system, the security controls of Storage networking could address the network issues within the context of this paper. More important according to the guidelines provided by the standard, the backup of logs enable SIEM systems to detect tampering behaviors by comparing the on-line log and its backups. Meanwhile, it is also possible to recover logs after deconstruction attacks thus keep the evidence for further forensic investigations too. The security controls of Storage security services can be also applied for the consideration discussed in the next section (4.3).

4.3 Policy on Secure Evaluation and Reporting of SIEM Data

In general, the alerting/reporting of security incidents should comply with national rules as well as business domain specific collection of reports and finally towards fulfilling the requirements of BSI in Germany. In the case of that the IRT is separated from the data owner, authority processes need to be designed and agreed. Besides, regular security analytic reporting could support the daily work of IRT and empower their capabilities.

For reducing the workload of IRT, the alerting/reporting should be graded into different criticalities thus with different priorities. Meanwhile, the IRT staffs have the option to access different level of incident details when required. In some cases, the tracing back function can even navigate to the relevant raw logs. At last, correspondent to real-time systems, the announcement has to be made in time. Similar to the divisions of criticalities/priorities, SEs need to be grouped and analyzed according to their timestamps as well as time relevant invalidation constraints. The analyzing time of a group of security-related events should be predictable and ideally can be executed in different granularities which achieve a balance between execution time and analyzing accuracy.

Considering the confidentiality of the collected data, similar to the privacy-aware consideration addressed in [LMW10], the central SIEM service should try to keep data

inside and apply strong controls for data going outside which means locating in the same facility of the managed SCADA system. When a remote collection is required, transferred data need to be encrypted or obfuscated. Besides, the created security-related events as well as correlations/reports have to be controlled by the associated knowledge management system and/or asset management system. Practically, the data retention time needs to be considered and should cooperate with the requirements of forensic investigations.

4.4 Policy on Real-time Monitoring with the SIEM System

According to [Ie15], on one hand, the SIEM system should be able to monitor residual risks identified in the risk assessment approach. On the other hand, the effectiveness of security controls could be measured by the SIEM system.

For mitigating risks identified in the risk assessment, security controls will be deployed. However, in some cases, the criticality of residual risks is reduced but not eliminated. When deployed security controls are compromised too, these risks might still cause critical damage to the SCADA system. Fortunately, the SIEM system can react to knowing risks. For example, within the system abnormally increased network traffic is detected (Fig. 2). With previous risk assessment, the traffic could be with high possibility caused by an attacker who joined the network. Thus reacting to this residual risk, a scanning of the network as well as re-authentications can be performed by the SIEM system automatically or by the IRT staff manually.

Besides the residual risks, the SIEM system can monitor the involved security controls too. Taking the same example of the increased network traffic, scanning network discovers no additional device thus the warning will be suppressed. However, the control of the network traffic is triggered too frequently, which might indicate an unknown type of security breaches or a system specific false positive constraint is triggered. In the latter case, smart testing could be deployed to determine the possible reasons.

4.5 Example of SCADA network architecture with SIEM Evaluation

As part of the concept Security by Design, the SIEM system can be integrated into the SCADA system during the design phase. In this section, a conceptual design of the SIEM system is proposed. As different security zones are shown in Fig. 1: The high security zone on the left side is protected by a data-diode which is physical unidirectional. In this case, only the traffic from the high security zone towards the SIEM service is allowed. The data-diode blocks attacks that send manipulated messages from the compromised SIEM system to the critical control systems, like sensors and actuators. Meanwhile, the data-diode can only provide limited bandwidth for collecting SIEM specific data. The middle security zone in the center of Fig. 1 contains the access control equipment, like authentication devices and CCTV monitors. This security zone is separated from the SIEM service by a configured firewall. On one side, the firewall

protects the devices in the middle security zones from tampering actions through the network. On the other side, through the firewall, high-speed connections as well as reliable protocols, like TCP/IP, are still applicable. At last, on the right side, the low security zone is connected directly with the SIEM service which provides little protection while the maximum connections speed.

Through the heterogeneous kinds of connections, security-related events and data are collected to the central SIEM service. Here after normalizations/aggregations, uniformed events can be correlated to discover security breaches or to measure security controls. The correlation rules of events are defined based on the overall security model thus enable a model-driven SIEM system. Section 5 discusses several policy-based correlation rules and their associated constraints. These rules are policy-based thus can be defined without knowing implementation details of sub-components. They are defined and parameterized in the central SIEM service complying with correspondent security controls.

The correlation results of security-related events will be automatically examined by the SIEM service according to their associated constraints. Possible security breaches or statistic info of deployed security controls will be reported to the IRT. Especially, security warnings are pushed to the IRT staffs for further reactions. The IRT staffs can monitor the automatic processing of the SIEM service as well as directly take reactions to the system for ceasing ongoing attacks.

5 Attempts for Modelling Policy-based Correlation Rules

As the example mentioned in Section 4.5, the SIEM system needs to be able to reduce the workload of IRT thus automatic processing is required. The overall security model described in [Wa16] contains both security objects and associated security controls. By linking SIEM service to this model, it is possible to correlate/analyze security-related events automatically in a model-driven way.

For example, potential security breaches in the “logon after leaving” example (Fig. 2b) could be discovered by proper predefined configurations/rules: In the example, the associated security object is the employee appeared in the two events. The security object as an employee owns several attributes: e.g. title and location. A security event might *change* or *indicate* these attributes. On one hand, the “leaving” event changes the employee’s position outside the company. On the other hand, the “logon” event indicates her/his location (near the server) inside the company. Sorting the two events according to their timestamps, the conflict of the employee’s locations will reveal the possible security breach. Furthermore, when after the “leaving” event, an extra “entering” event occurred that changed the employee’s position again inside the company before the “logon” event. Therefore, there will be no conflict on the location attribute and no warning should be sent to IRT staffs.

Further extending the same example, what does that mean when we found 500 pairs of “entering” and “leaving” events between the first “leaving” event and the “logon” event (within 30 minutes)? Similar to the example of too many login attempts (e.g. 500 times) on a workstation in a short time (e.g. 10 minutes), the high frequency indicates possible security breaches. For discovering this kind of attacks, the resource concept needs to be introduced into the overall security model. A security object has always limited resources within a given time period. In this example, we can define one person can be associated up to 5 SEs during 30 minutes. Thus the real-time monitoring of the SIEM system will announce the security staff a warning since the resource (count of 5 SEs) was consumed up in a while. Similarly the resource (number of login attempts) could also be assigned to a server. Consuming up of the resource in a very short time might indicate a denial-of-service attack.

Considering the “configuration after working hour” example, after the regular working hours, the resource of allowed configuration changes will be shifted to 0. Thus any action could be counted as a configuration change, like setting parameters or patching software, will trigger a security warning.

6 Conclusions and Challenges

In this paper we discussed the common required features of a SIEM system and described its working approach by splitting it into different steps according to *the* preliminary work at ISO/IEC JTC1/SC27 WG4. Furthermore, a conceptual design of a SIEM system as well as associated network architecture is proposed. By linking the SIEM system to the overall security model of SCADA systems, a model-driven monitoring/reporting approach could be created to monitor system’s residual risks while measure effectiveness of deployed security controls. Considering the context of SCADA systems, the SIEM system needs to handle security-related events from different sources without accurate universal timestamps while needs to take archiving raw logs as well as generated events into account for later on tracing back or forensic investigations.

However, creating a SIEM system for a SCADA system still faces challenges. The raw logs of system’s sub-components are in different formats and probably do not share a universal clock. The correlation rules of security-related events can be complex while the system needs to react in real-time. At last, improper correlation rules and their constraints might lead to false positive alerts. These alerts will be generated especially when operation mode is changed (e.g. from normal mode to the maintenance mode) and associated rules as well as constraints are not adapted. This might result in that the IRT chooses to ignore or disable the feature.

According to the required security levels (SLs) it may be worth to spend effort in the standardization of logging format, especially for the sub-components with high SLs. This will help to avoid false positives as well as to support efficient evaluations for the most critical sub-components of a SCADA system.

Note: Some of the above described modelling-analyses are being elaborated as part of participation in the “SMARTTEST” Cybersecurity Testing R&D with three German University partners, partially funded by German Ministry BMWi.

References

- [Mo96] Mori, H.: The Perfect Insider, Kodansha, 1996
- [LMW10] Lisovich, MA.; Mulligan, DK.; Wicker, SB.: Inferring Personal Information from Demand-Response Systems, IEEE Security & Privacy, vol. 8, no. 1, pp. 11-20, 2010
- [MT15] Martyak, P.; Thow, M.: Enhancing Defense-in-Depth and Monitoring Programs to Protect Critical Digital Assets from Tampering, 9th Nuclear Plant Instrumentation and Control, Charlotte, 2015.
- [St10] Stocks, W.: When Anti-virus Doesn't Cut It: Catching Malware with SIEM, 24th Large Installation System Administration, San Joes, 2010.
- [Wa16] Waedt, K.; Parekh, M.; Tong, X; Gao, Y.; Ding, Y.; Xie, X.: Nuclear Safety and Risk based Cybersecurity Testing, 47th Annual Meeting on Nuclear Technology, Hamburg, 2016.
- [Bu16] Bundesministerium für Wirtschaft und Energie: IT-Sicherheit für die Industrie 4.0, 2016.
- [Ie13] IEC 62443-3-3: Industrial Communication Networks–Network and System Security – Part 3-3: System Security Requirements and Security Levels, 2013.
- [Ie15] IEC 62443-4-2 [Draft]: Security for industrial automation and control systems – Part 4- 2: Technical security requirements for IACS components, 2015.
- [III1] ISO/IEC 27035: Information technology — Security techniques — Information Security Incident management, 2015.
- [III3] ISO/IEC 27002: Information technology — Security techniques — Code of Practice for Information Security Controls, 2013.
- [III5] ISO/IEC 27040: Information technology — Security techniques — Storage security, 2015.
- [Ni11] NIST SP 800-82: Guide to Industrial Control Systems Security, 06.2011.
- [Ib16] IBM QRadar, www-03.ibm.com/software/products/en/qradar-siem, Accessed: 2016-05-30.
- [Sp16] SPLUNK, www.splunk.com, Stand: 30.05.2016.

3D Modeling of Selected Assets, Security Zones and Conduits

Simon Seibt¹, Karl Waedt², Hans Delfs³, Simon Odorfer⁴

Abstract: Current critical industrial infrastructure (CII) which uses industrial automation and control systems (IACS) can be a target of cyber-attacks or combined cyber-physical attacks. To ensure the security of the CII, international standards are gradually evolving. The IEC 62443-x-x series is a specialized multipart security standard for IACS. It introduces, among others, the concept of security zones and security conduits [IE09]. The security zones can be defined in a physical sense and for logical grouping. By applying the defense-in-depth concept, both definitions are relevant. The development of a realistic three-dimensional (3D) model of CII can support the subdivision for the zoning and further security analyses. The model includes the locations of security relevant physical assets, which can be grouped into security zones and linked with related security artefacts. By using a 3D model, all relevant two-dimensional views can be derived. This paper addresses the use of a 3D model to support the application of security controls and risk assessments in line with concepts elaborated by IEC 62443-4-2 [IE15].

Keywords: cybersecurity, industrial automation and control systems, physical assets, application security controls, security zones, security conduits

1 Introduction

Industrial automation platform developers and system integrators are gradually deploying commercial-off-the-shelf (COTS) technology for industrial automation and control systems (IACS) while still using legacy components that were initially designed for autarkic use only. This may cause increased vulnerabilities for cyber-attacks against the IACS equipment. The IT technology was initially developed for business systems for their daily processes and is usually not sufficiently robust or the legacy systems were designed for use in a different context. International organizations and committees have already responded to the increased threat level [St14]. Standards relating to the IT Security for control and safety systems are gradually released. Several parts of the multipart security standard IEC 62443 for IACS are already published. Key concepts of the IEC 62443 are Foundational Requirements (FR), System Requirements (SR),

¹ Nuremberg Institute of Technology, Department of Computer Science, Keßlerplatz 12, 90489 Nuremberg, seibtsi47063@th-nuernberg.de

² Areva GmbH, Overall & Safety I&C Engineering / Cybersecurity, Henri-Dunant-Str. 50, 91058 Erlangen, karl.waedt@areva.com

³ Nuremberg Institute of Technology, Department of Computer Science, Keßlerplatz 12, 90489 Nuremberg, hans.delfs@th-nuernberg.de

⁴ Nuremberg Institute of Technology, Department of Computer Science, Keßlerplatz 12, 90489 Nuremberg, odorfersi57887@th-nuernberg.de

Component Requirements (CR) and Requirement Enhancements (RE) [IE15]. As compared to generic information security standards, like ISO/IEC 27001/2 [III13a] [III13b], the concepts go along with a grading according to 4 Security Levels (SL) [IE13] and a comprehensive concept of security zones and security conduits.

The identification of physical security zones in a factory or plant or more generic for critical industrial infrastructures (CCI) may be difficult as the physical assets are often distributed over dozens to hundreds of rooms in different buildings on one or multiple sites. Three-dimensional (3D) models are increasingly used, for example to simulate facilities for test purposes or for virtual commissioning. The development of a 3D model can also support the definition of security zones, assignment of security controls as well as security risk analysis. The 3D model represents a hierarchical structure of the CII, which can be used to group the physical assets into correspondent security zones and to assign physical security controls and network interface security controls at the respective position in the model. These security controls, or more precisely Application Security Controls, as will be explained later, are linked to the refined security requirements and the corresponding reference of the security standard in order to assure the correct implementation. The links are bidirectional. Thus, a security auditor will also be able to navigate from the requirements to the security controls, which themselves can be associated to physical assets represented in a 3D model. Later on, these serve as one basis of the risk assessment, e.g. the extraction of attack trees and the prioritized analysis of paths of an attack tree [Wa16].

2 Principle and granularity of the 3D modeling

To model the physical properties of a CII together with the security relevant assets, a hierarchical structure for the 3D modeling is needed. The hierarchical structure is based on a graph of nodes with parent-child relationships. Each node can have an array of children but only one parent. A child of a node can be a single 3D object or another node, which includes further 3D objects or nodes. The structure is composed gradually and begins with the modeling of individual physical assets.

Note: Modeling of complex interrelationships is also an approach currently followed as part of Industry 4.0 in Germany [BM40] and similar approaches in other countries.

Note: The following figures show a part of a power plant as an example of a hierarchical structure. This example contains extreme simplifications and does not correspond to the reality.

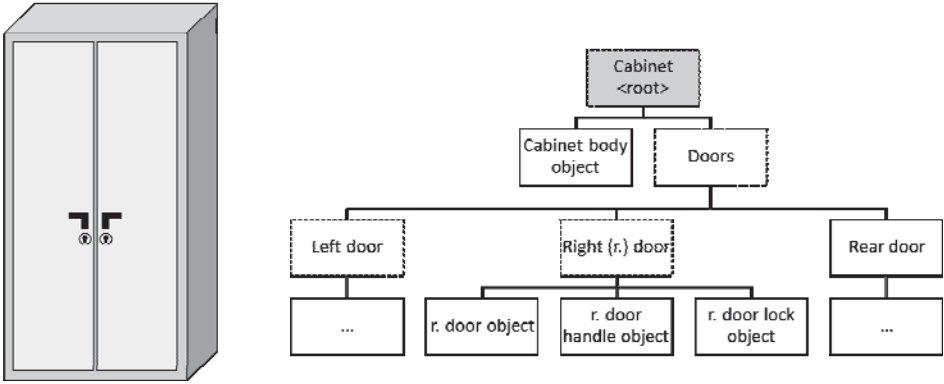


Fig. 1: Artefacts of an instrumentation and control cabinet

Fig. 1 shows an instrumentation and control (I&C) cabinet, which is an example of a security relevant asset. By modeling this cabinet, the root node is the “cabinet” element as indicated in Fig. 1. This node is linked to two children. One of them is a 3D object, which represent the body of the cabinet. The body is a modified standard cube mesh. The other child is another node with “Doors” as its name. This node groups the three door elements, two of which are positioned in the front of the cabinet and one at the rear of the cabinet (not visible in the 2D view at the left hand side of the figure). Each of the three door elements is a node that consists of a door object, a door handle object and a door lock object as part of the 3D mesh. These three downstream elements are simplified in this example because the elements could also be nodes, which further 3D objects. The door lock object is an example of security control (security counter measure) against unauthorized access to the cabinet.

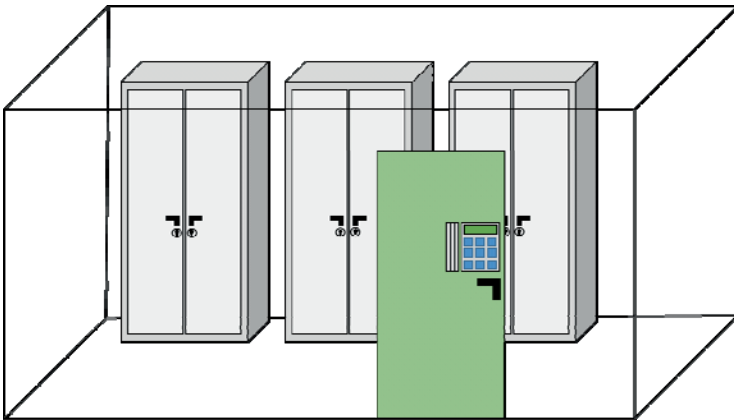


Fig. 2: Artefacts of an instrumentation and control room

In Fig. 2 the hierarchical structure is traced to the room level. The room is an example for an I&C room and consists of three cabinets that were modeled earlier. Furthermore,

the room has a door that also features an access control as a security counter measure. This access control could be a smart card reader with an additional PIN.

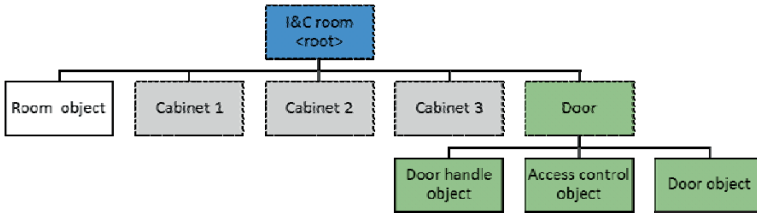


Fig. 3: Room level hierarchical representation of security relevant artefacts

By modeling this room, the root node is set to the “I&C room” as indicated in Fig. 3. This node consists of four other nodes and a 3D object, which is the “Room object”. The cabinet, which was grouped as a node, is tripled to new unique nodes in a child relation to the “I&C room”. The access-protected door is also another node with three child objects.

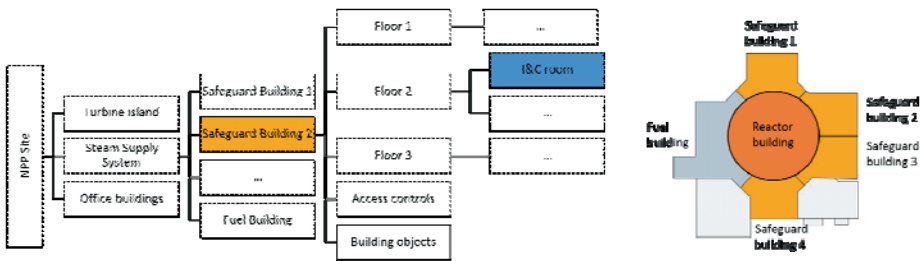


Fig. 4: Plant level hierarchical representation of security relevant artefacts

The next step is to expand the hierarchical structure to the buildings level or if there are more security relevant floors, each floor has to be modeled separately first. For this example, only floor 2 is considered as a relevant part. So the root node is set to “Safeguard Building 2”, as indicated in Fig. 4. To complete the safeguard building model, the procedure is the same as the modeling of the “I&C room” or the relevant assets.

By finishing the 3D model of a Power Plant with the introduced procedure, a scene graph with a hierarchical structure is created. This graph can be used to place security controls at their effective positions and to group the graph into security zones.

3 Application Security Controls

The security relevant assets should be protected by security controls such as physical entry controls or access controls. By modeling the assets, the security controls can be placed at their effective positions. The security controls describe measures to prevent an attack, detect an attack or initiate corrections/mitigations after an attack [WD15]. To assure that all security controls are correctly implemented, the security controls are linked to the description and implementation guidance from IEC 62443-x-x and ISO/IEC 27002:2013 [Wa16]. Another link can also be set to the initial requirements. Thus e.g. a RE according to IEC 62443-4-2 may be to have two authentication means. A further project-specific guidance may be to have iris scanners as additional access controls for each main security zone. From the Security Controls depicted in the 3D model it will be possible to navigate to the respective source requirements. However, this is beyond the focus of this paper. Furthermore, by using this linking method, weak points can be early identified on this level and potential physical intruder paths can be discovered by attack trees [WD15].

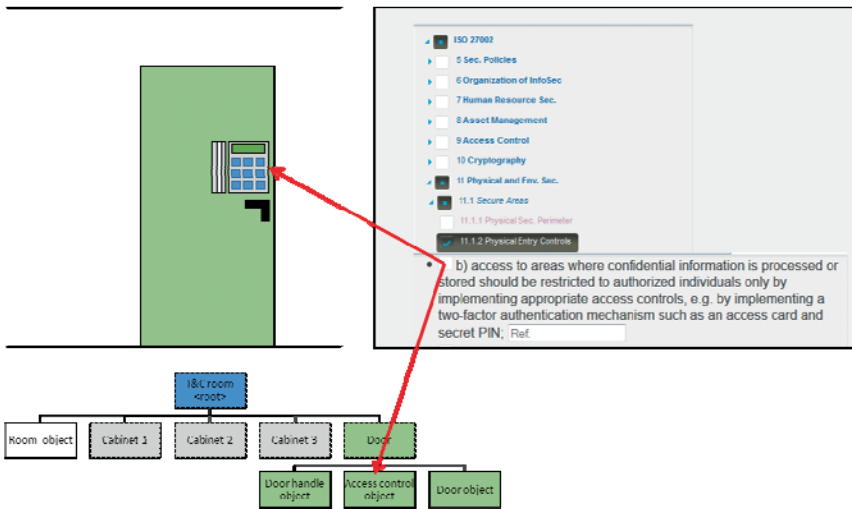


Fig. 5: Linking security controls

Fig. 5 shows an example for the security controls linking. The smart card reader with an additional PIN, which is a child object of the “Door” node, can be linked to the description and implementation guidance of the physical entry controls section of the ISO/IEC 27002:2013. The smart card reader can be used with cryptographic smart cards. The DES is a symmetric-key encryption and the RSA is a public-key cryptography [DK15]. Both methods are mostly used for the cryptographic smart cards. These smart cards can also be linked to cryptographic controls section of the ISO/IEC 27002:2013 to assure the correct implementation.

The general security standards like ISO/IEC 27002:2013 are structured hierarchically with different levels of detail. This can be modeled as XML tags or as JSON objects. Each section of the standard should get a unique ID for the linking. The file format for the 3D models is typically also based on XML or JSON. By modeling the 3D models with a scene graph, each object also get a unique ID for the identification.

To get a relation between the 3D object and the corresponding section of the standard, the Application Security Controls (ASCs) [II11] can be applied. The ASC is a data structure, which contain a precise enumeration and description of a security activity and its associated verification measurement to be performed at a specific point [II15]. An ASC data structure can also be created as XML tag or JSON object in a separate file or later in a database for each project or corresponding 3D model.

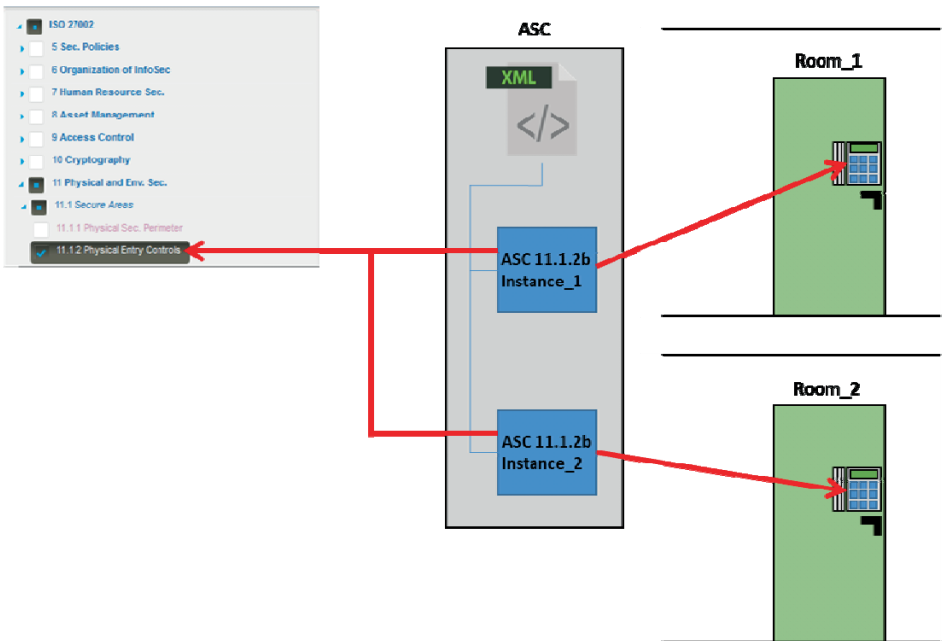


Fig. 6: Linking with ASC instances

Fig. 6 shows an example for a separate XML file with two ASC instances. For each security control linking, a new instance of the XML tag of the ASC is needed. In the example, two doors with their smart card reader as access control are indicated. The access control objects must be linked with the same corresponding section of the security controls standard (or guidance according to which the security controls are structured). Therefore, two instances of the ASC are needed. Each ASC also features additional information for the correct implementation of the security control. The listing example (Listing 1.) below shows a XML tag entry of an ASC.

```

<ASC ASC_ID="ASC_11_1_2_b_Instance_1">
  <ThreeDimensionalObject>
    <ID>AccessControl01_Door01_Room01</ID>
    <Filename>I C Room Scene</Filename>
  </ThreeDimensionalObject>
  <Standard>
    <ID>ISO_27002_11_1_2_b</ID>
    <Filename>ISO_27002</Filename>
  </Standard>
  <VerificationProcedure>...</VerificationProcedure>
  <SecurityTestCases>...</SecurityTestCases>
  <SecurityTestProcedure>...</SecurityTestProcedure>
  <Accountable>...</Accountable>
  <Responsible>...</Responsible>
  <Consulted>...</Consulted>
  <Informed>...</Informed>
  <ANF>...</ANF>
  <SecurityLevel>...</SecurityLevel>
  <SecurityZone>...</SecurityZone>
</ASC>

```

Listing 1: ASC instance as XML tag

Each ASC instance is also assigned a unique ID, which is a property of the start tag. The ID and Filename (or later file-path) of the 3D object and the corresponding section of the standard is integrated in a sub-tag of the ASC tag. After these entries, the additional ASC details are listed as own tags.

For example, the “SecurityTestCases” tag include different scenarios or use cases, which the security auditor wants to consider during the audit of the ASC. The tags “Responsible”, “Accountable”, “Consulted” and “Informed” base on the RACI model to store the role-based information of the involved persons. Furthermore, a link to the respective security zone can also be stored in the ASC instance.

While the concepts of Organization Normative Framework (ONF) and Application Normative Framework (ANF) [II15] will not be further evaluated here, the above example already indicates, that for complex industrial facilities, it is beneficial to use semi-formalized definitions of Application Security Controls (ASCs). Thus, interrelations between ASCs as well as links between ASCs and modeled objects become possible, together with commercial project relevant role assignments, activity planning and activity tracking, like the scheduling of security test activities and the currently achieved progress.

4 Security Zones and Security Conduits

Security Zones and Security Conduits are introduced as key concept in the IEC 62443-x-x. The zones can be defined in a physical sense and for logical grouping. By developing of a 3D model of a critical industrial infrastructure, multiple cybersecurity related activities are supported, starting from the Security by Design up to the virtual inspections. As part of the Security by Design, the security zones can be defined by navigating through the 3D model and by assigning the physical 3D container objects to Security Levels, e.g. SL1 to SL4 according to IEC 62443. Thus, the physical composition of Security Zones is defined.

Similarly, the logical Security Zones are defined starting from the business needs (business domains) and the network architecture. Note: This second part of logical definition of Security Zones as well as the combination of Physical Security Zones and Logical Security Zones is not addressed in this document.

Security Zones provide an effective means to design and provide Application Security Controls that are applicable to all assets contained in the respective Security Zone. At the modeling level and the (XML/ JSON) data representation level, this is similar to the inheritance of security controls to all assets in a Security Zone.

Beyond supporting an efficient representation (with less redundant repetitions), this also supports the effective demonstration of compliance with regulatory requirements, e.g. on how a requested Security Level is achieved by the contribution of multiple Application Security Controls, that are either inherited or explicitly assigned to the evaluated primary or supporting assets.

Security Zones can also support dynamic security analyses. As part of a security risk assessment scenario it can be assumed that one of the predefined threat agents is located in a specific Physical Security Zone. Based on the assumed threat agent characteristics, like general IT skills, general knowledge level of process engineering, knowledge level of automation platforms, knowledge of the specific configuration of the targeted industrial process, knowledge of security vulnerabilities etc., the threat agent may be able to traverse along different branches of an attack tree. The ability to traverse a path in an attack tree will not be only according to the threat agent capabilities but also according to the strength of the implemented application security controls.

In many cases, the strength of the application security controls can be assessed by security tests, like penetration and fuzz tests. Based on the detail of the 3D specification of cybersecurity relevant assets and artefacts, the benefit of cybersecurity tests expressed as expected effort to potential impact ratio can be prioritized. Note: “Smart” (as model based) cybersecurity testing is part of a joint project (SMARTTEST) by multiple German universities, partially sponsored by the German Ministry BMWi. The primary focus of the modeling for the cybersecurity testing is on the modeling of network protocols and state machines of the automation devices which allow more targeted attacks. However,

information from these behavioral models can be linked to the 3D and 2D models addressed in this document (e.g. link from an industrial network protocol model to a Security Conduit).

5 Conclusion

For complex industrial facilities, 3D modeling can effectively support multiple security activities, starting from the Security by Design up to the support of security audits. The 3D representation serves as unique source for deriving multiple 2D views. Traditionally, multiple 2D perspectives are used that may become inconsistent after manually modifying one 2D view. Nevertheless, objects in the 3D scenes can be linked to objects in unique (source) 2D representations, e.g. of network architectures. Similarly, objects in 3D scenes, can be linked to objects in unique 2D representations, e.g. a physical network interface card port in a 3D view of a network switch located in an electronics cabinet can be linked to the respective connection point in a 2D network architecture diagram.

Equipment from 3D representations of supporting automation and IT assets (e.g. a processing unit) can be linked to other 3D physical equipment, e.g. pumps, valves, smart sensors or actuation equipment which may be supporting or primary assets. While this is not elaborated in the current paper, it is a key step for evaluating the potential (graded) impact of a threat agent to the primary assets (e.g. a chemical process).

Summing up, this paper showed some modeling approaches that support the graded security levels, zones and conduits concepts of IEC 62443 in general and more specifically the risk assessment approaches being developed in IEC 62443-4-2. As indicated, several extensions have to be explored in order to make full benefit of the potential provided by an intuitive, semi-formal and, extensible modeling.

6 References

- [BM40] BMWi: IT-Sicherheit für die Industrie 4.0 - Produktion, Produkte, Dienste von morgen im Zeichen globalisierter Wertschöpfungsketten - Studie im Auftrag des Bundesministeriums für Wirtschaft und Energie, 2016.
- [DK15] Delfs, H.; Knebel H.: Introduction to Cryptography: Principles and Applications, 3rd ed., Springer, 2015.
- [IE09] IEC 62443-1-1:2009 Security for industrial automation and control systems – Part 1-1: Terminology, concepts and models, 2009.
- [IE15] IEC 62443-4-2:2015 [Draft] Security for industrial automation and control systems – Part 4-2: Technical security requirements for IACS components, 2015.
- [IE13] IEC 62443-3-3:2013, Industrial communication networks – Network and system security – Part 3-3: System security requirements and security levels, 2013.

- [III3a] ISO/IEC 27001:2013 Information technology – Security techniques – Information security management systems – Requirements, 2013.
- [III3b] ISO/IEC 27002:2013 Information technology – Security techniques – Code of practice for information security controls, 2013.
- [III11] ISO/IEC 27034-1:2011, Information technology - Security techniques - Application security - Overview and concepts, 2011
- [III15] ISO/IEC 27034-2:2015, Information technology - Security techniques - Application security - Organisation normative framework, 2015.
- [St14] Störkuhl, T.: IT-Sicherheit auf Basis IEC 62443 für elektrische Signalanlagen, Signal + Draht 10/2014, 10-12, 2014.
- [Wa16] Waedt, K.; Parekh, M; Tong X.; Gao, Y.; Ding, Y.; Xie, X.: Nuclear Safety and Risk-based Cybersecurity Testing, 47th Annual Meeting on Nuclear Technology, Hamburg, 2016.
- [WD15] Waedt, K.; Ding, Y.: IT Security / Interoperability, 1st Sino-German Intelligent Manufacturing/Industry 4.0 Standardization International Summit Forum (SGSF), Shanghai, 2015.

Graded Security Forensics Readiness of SCADA Systems

Jianghai Li¹, Edita Bajramovic², Yuan Gao³, Mithil Parekh³

Abstract: Security event logs are major indicators for the timely discovery of cyberattacks and during security incident examinations. Collection of sufficient logs of events associated with security incident time is critical for effective investigation. SCADA systems logging capabilities are intended for identifying and detecting process disruptions, not security incidents, and are frequently not suitable for digital forensic investigation [Ta13]. Nevertheless, logs provide tremendous support during digital forensics investigations as they consist of vast amounts of information, e.g. step-by-step events that occurred in a system in question, including time stamping [AIJ12]. In addition, logging is a major element of forensic readiness. Numerous tools and methods contribute to log monitoring, e.g. evaluating log records and correlating them through various systems. This can assist in incident handling, identifying policy violations, auditing, and other efforts. Within the general context described above and the more specific graded security approach of IEC 62443-x-x, this paper will identify cybersecurity specific SCADA component requirements, preconditions for subsequent forensic investigations, collecting potential digital evidence, graded forensic-related security controls, and forensic readiness during SCADA lifecycle phases.

Keywords: cybersecurity, forensic readiness, logging, time stamping, asset management, digital forensics investigation.

1 Introduction

As indicated in [IEC613] [IEC615], the application or device of an industrial automation control system (IACS) should have the ability to produce audit records related to security for the access control, request errors, control system events including activity and transactions logs, backup and restore event, configuration changes, and audit log events. However, many controllers deployed in SCADA systems do not have any capability to log security events. Many challenges with digital forensics investigation for SCADA systems exist, e.g. the field devices frequently do not have essential ability for comprehensive logging. Additionally, most of devices are not able to gather a sufficient amount of data [Hs08]. The problem studied in the paper is how to develop security logging capacity in SCADA system as SCADA system and network device logs are

¹ Tsinghua University, Institute of Nuclear and New Energy Technology, Beijing, China, lijianghai@mail.tsinghua.edu.cn

² Friedrich-Alexander University Erlangen-Nuremberg, Informatics 1 Department, Erlangen, Germany, edita.bajramovic@gmail.com

³ Otto-von-Guericke University Magdeburg, Group Multimedia and Security, Magdeburg, Germany, yuan.gao@ovgu.de

⁴ Otto-von-Guericke University Magdeburg, Group Multimedia and Security, Magdeburg, Germany, mithil.parekh@ovgu.de

crucial to detect cyber-attacks and perform forensics investigation. In addition, log integrity and reliability are equally important during the investigation; hence, logs must be regularly saved on a different system and constantly backed up. Also, logs should be cryptographically hashed to permit proper discovery of log modifications [Hs16]. Operators are also part of forensics investigation challenge. Usually they do not have the knowledge and expertise to gather, investigate, or evaluate SCADA system's traffic [Hs09]. As a substitute, operators are dependent on vendors for support during incident response. As a result, security incident response and resolution becomes very slow.

Within the context described above and as a preparation towards gradual compliance with IEC 62443-x-x, this paper is divided in six sections. The current section introduces general information regarding logs. The second section addresses Component Requirements and how the combination of Component Requirements (CRs) and Requirement Enhancements (REs) determines the Target Security Level that a component is capable of. The third section identifies preconditions for subsequent forensic investigations. These include the identification of assets, asset management, tracking, and time stamping of logging information. Precise asset identification is needed to address the security of SCADA systems. The fourth section refers to collecting evidence from servers and mobile equipment and issues related to proper collection of electronic evidence from the respective devices. Some of these electronic evidence collection topics are SCADA specific, and may involve e.g. Physical Unidirectional Security Gateways for conduits connecting to the highest security zones according to IEC 626443. The fifth section outlines graded forensic-related security controls. Lastly, the sixth section presents forensic readiness during all SCADA lifecycle phases. This is in line with newest Industry 4.0 BMWi IT Security standard [Bm16].

2 Component Requirements

Each system requirement (SR) has a basic requirement and zero or additional requirement enhancements (REs) to improve security [IEC615].

The control system should have ability to centrally maintain audit events and collect audit records from numerous devices through the control system into a one system (logical or physical). In addition, time-correlated audit track should be provided. Furthermore, the control system should have the ability to export these audit records in industry standard formats for enabling analysis by standard commercial log analysis tools, for example, security information and event management (SIEM) [IEC615].

Fig. 1 indicates how Component Requirements (CR) are derived from Foundational Requirements (FR) and System Requirements (SR) while taking into consideration a grading according to four Security Levels [IEC613] [WD14]. Standard format is important for higher security levels, lower probability of false negatives, and better basis for testing (different depth of testing of forensic readiness).

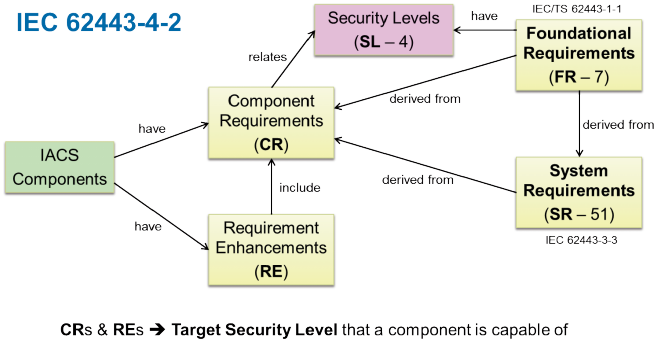


Fig. 1: Component Requirements for Security Graded SCADA Systems

3 Preconditions for Subsequent Forensic Investigations

Numerous international standards provide details regarding digital forensics investigation procedures and response to security incidents [EGI13] [EGI14] [ISIE11a] [ISIE14b] [ISI12b] [ISIE14c] [ISIE15c] [ISIE15d] [ISIE15e] [ISIE15f] [ISIE16a] [ISIE16b]. Nevertheless, many preconditions should be fulfilled for subsequent forensic investigation. The Asset Management addresses the systematic definition and maintenance of an up-to-date Asset Portfolio [ISO514a]. A first step towards an asset management is the asset identification [Wa16]. Asset identification is needed to address the security of SCADA systems [ISO514a] [ISO514b] [ISIE12a] [ISIE15a] [ISIE15b]. Automatic asset identification can increase the efficiency of security controls for communication networks, IT equipment and software assets. Non automatic asset identification could be implemented for some applications, e.g. safety-critical process automation, secure development environments, tamper indicating devices or legacy systems [Wa16]. The most critical assets are located in higher security zones.

3.1 Identification of Assets and Asset management

Understanding and monitoring SCADA system networks and asset management are essential elements in discovering cyberattacks and performing subsequent forensic analyses [LL15]. Asset identification serves numerous purposes inside a SCADA system. Four basic methods exist that support the identification of assets, as indicated in Fig. 2 [Le15].



Fig. 2: Asset Identification Basic Methods [Le15]

Additionally, many other methodologies in identifying assets exist; however, the risk that some assets will not be identified is present, e.g. certain legacy systems just do not use network for communication or they do rarely [NERC09a]. So in this case, physical identification of assets is needed [Le15]. Physical identification should be conducted occasionally to verify outcomes. Evaluation of configuration files on hardware, e.g. switches, discloses already identified and registered hardware. In addition, when hardware, e.g. a switch, is part of managed infrastructure, capturing network traffic necessary for passive scanning is possible. Evaluating data can be performed using open source tools, e.g. Wireshark, to accurately detect assets and communication patterns [Le15]. Passive scanning is good method for fast and competent identification of assets. On the other hand, active scanning on SCADA networks should not be always conducted as interrelating with sensitive hardware using unanticipated methods can cause interruption of normal operations or asset failure. Moreover, network hardware, e.g. proxies and firewalls often logically block involuntary communications [Le15]. As a result, partial network design records by active scanning are returned. Furthermore, sending communications through the network can alter the communication topologies causing difficulties for precise discovery and baseline.

3.2 Tracking for Correctly Associating to Sources

Asset tracking through SCADA system components provides benefits for the implementation of some security controls, but it also introduces the possibility of misuse. Locating assets is essential for safety and security reasons [NIST11]. For example, it should be possible to track the smart sensor that generated a specific event, even if the e.g. analog signal values provided by the sensor are received via intermediate (e.g. signal conditioning and preprocessing) automation equipment. Active log management functionalities may even identify cyberattack or security event in progress and provide information regarding place and traces leading to better response to the incident [NIST11]. Additionally, continuous tracking and monitoring of audit traces on critical zones of a SCADA system leads to discovery of malicious activities and thus permit the essential remedial actions. [Ba16b].

3.3 Time Stamping of Logging Information

During digital forensics investigation, proper time stamping and recording of events is achieved only if a high precision reference clock is available [NERC09b]. For example, timestamps might incorrectly specify that event A occurred 35 seconds before event B, when event A in reality occurred two minutes after event B [NIST06]. Thus, it is essential that reference system clock indicates accurate time. Inaccurate time leads to difficult log analysis. Furthermore, time stamps as outlined in IEC 62443 series, are required for successful audits and also for legal purposes. However, real-time clock may not be available in highest security zone or may not be supported by the embedded hardware or may deliver only a part of the time stamp (e.g. subdivision within one

second).

Nowadays, global positioning system (GPS) clocks are used to make sure universal time through the domain. However, network latency can cause substantial time dissimilarities between devices [NERC09b]. As a result, Network Time Protocol (NTP) can be utilized to approximate and calculate this latency. On the other hand, NTP can allow network spoofing attacks, initiating wrong time stamping [NERC09b]. Fig. 3 indicates real-time collection of the data and time stamping in a subsequent device (e.g. in a gateway). In a typical IT environment, time stamping is performed at the source, e.g. by the server or workstation, but a time stamp source may not be available for all SCADA components.

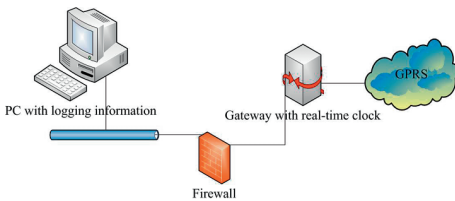


Fig. 3: Real-time Collection of the Data with Time Stamping in a Gateway

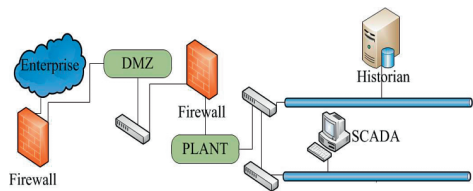


Fig. 4: Network Segmentation by Firewalls

4 Collecting Potential Digital Evidence

Collection of efficient and reliable data needed for forensic analysis cannot be fully extracted from numerous traditional device and control systems technologies due the nature of the technologies in question [Hs08] [Ed15]. Collecting evidence from servers and mobile equipment e.g. maintenance laptop, is not an issue as logging capability is available. But, collecting evidence from embedded devices is challenging as logging is not available and any transfer of evidence must be properly secured.

4.1 Regular Collection of Logging Information

Many challenges with regular collecting of logging information exist, e.g. restricted access to high security zones. Providing additional access, just to read the logging data brings in security risks. To overcome this issue additional firewalls can be implemented. Fig. 4 indicates an Industrial network that has been separated from an Enterprise network using Firewalls [Sc15].

Fig. 4 indicates Data Diode (DD) from High Security Zone to Low Security Zone, e.g. no manipulated command can be sent to SCADA system. Fig. 6 shows DD for transfer from an outside location of the utility e.g. intranet towards a Low Security Zone and no sensitive information will be stolen through Internet. Lower security zone, e.g. containing monitoring systems, needs connection only for time stamping.

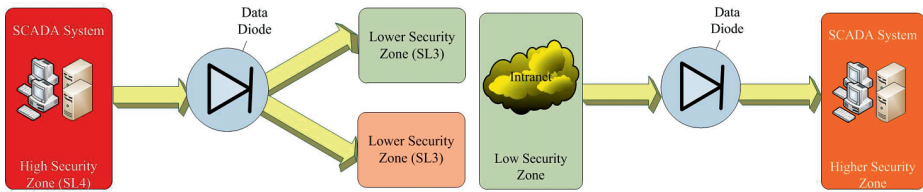


Fig. 4: Data Diode (High Sec. to Low Sec. Zone) Fig. 6: Data Diode (Low Sec. to Higher Sec.)

4.2 Identification of Users and Mapping to Roles

Identification of users is mandatory. No group logging should be allowed. However, multiple users may be assigned to the same role e.g. that allows them to change set-points of a SCADA component. This assures that they can perform their day-to-day work, assuming the respective work permits are issued (where appropriate), while a personal tracking of user activities related to the SCADA equipment is possible. When staff from multiple shifts is working with the same equipment, they need to login/logout. Therefore, efficient means for switching between users is needed.

4.3 Forensic Readiness for Maintenance Equipment

An essential aspect in preparing maintenance equipment for regular use during outages or for emergency situations is the incorporation of forensic readiness. The main idea is to assure that it was not manipulated, e.g. somebody replaces the software to be locally loaded. In this case, firewalls or a Physically Unidirectional Security Gateways [ISIE14a] are not applicable. Further, additional scanning before loading the new software to embedded devices is needed so that malicious software can be rejected or at least the loading behavior and authorized user will be recorded for forensic investigations.

4.4 Tamper Proof Logging

The following Building Technology and Data Diode example demonstrate how tamper proof logging is implemented and the purpose of a Physically Unidirectional Security Gateway (Data Diode).

4.4.1 Building Technology Example

Traditionally Building Technology (BT), often also called (Industrial) Facility Management, contained some security controls that still meet current cybersecurity recommendations [LW15]. For example, the commonly deployed logging of access of persons to different plant areas on a paper roll is definitely tamper-proof and can be used when addressing non-repudiation, as in Fig. 5, Current BT equipment, like heavy doors

or gates, includes automation equipment, usually with an electronics cabinet and a pneumatic cabinet per heavy gate. As an example, Fig. 5 also indicates how the networks connected to these gates have to be designed, in order to assure that an additional, completely independent monitoring and logging is in place [IEC611]. Tamper-proof logging on paper rolls is still in place. However, new cybersecurity technologies like Security Information and Event Management (SIEM) complement these [Ba16a].

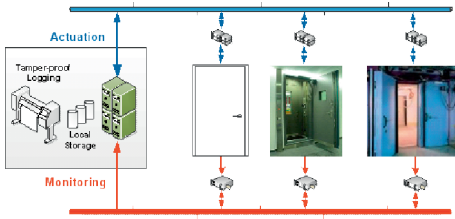


Fig. 5: Building Technology Example

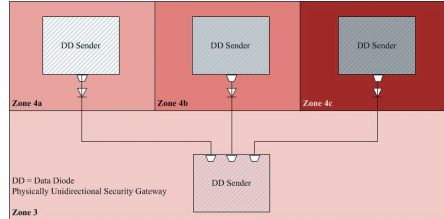


Fig. 8: Data Diode Example

They have to be considered in the Criticality Analysis with regard to correlating logging information of I&C and Electrical Systems (ES) with monitoring information provided by BT [WLZ15]. This logging information is also important with regard to Emergency Preparedness [Hs15] and protecting Critical Digital Assets from tampering [MT15].

4.4.2 Data Diode Example

According to the industrial facility specific security zone model, monitoring of systems through secure zone boundaries (security conduits) is required, in line with IEC 62443-3-3. Therefore, zone perimeter security policies must allow the transfer of security logs and events produced by monitoring devices to a central management system [WLZ15].

Data diodes can support the implementation of solutions that meet requirements of high security levels, as shown by the example in Fig. 8 [KL14]. As physical separations, they assure that the information flow is in one direction—away from the zones with higher security levels and in the direction of the central management system [Ba16a]. Additionally, data diodes make sure there is no capability for malicious traffic to penetrate the secure zone from the logging facility [KL14].

5 Graded Forensic-related Security Controls

For graded forensic-related security controls, a graded level of review stringency and testing depth is needed. Forensic-related security controls should be implemented, reviewed and tested according to security level presented in Fig. 6. Therefore, auditing on workstations, servers, and network devices must be permitted; audit records should be stored on centralized log servers. Critical applications should be configured towards auditing, including recording authentication attempts. Because logs manipulated, preparations, e.g. through their policies, guidelines, and procedures, should be done to

demonstrate the trustworthiness and integrity of such logs. Proactive approach should be considered in collecting valuable data. Configuring auditing on SCADA system, employing centralized logging, and implementing and using security monitoring controls in different security can all produce sources of data for further forensic efforts. In addition, Automated Asset Identification should be implemented.

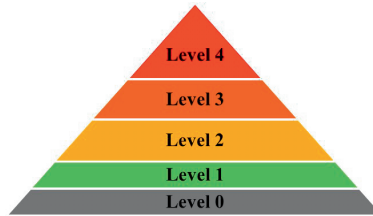


Fig. 6: Security Levels

Security levels are a concept that describes the degrees of security protection needed by different SCADA systems. Each level requires different sets of forensic-related security controls implemented in each security level satisfy the security requirements of that level [IEC614]. Multiple schemes and numbers of security levels or security grades are deployed by different industrial domains, e.g. four security levels according to IEC 62443 and as indicated in Fig. 6.

5.1 Automated Asset Identification

As mentioned in previous sections, Asset Management System needs to be implemented in order to assure an initial identification of assets at the right level of detail and for subsequent enforcement of identification updates after ongoing modifications. This can be supported by Asset Management Tools, including software modules for automatic asset identification. Automated asset identification may not be applicable for higher security zones either because it introduces an additional risk or because the SCADA components do not support automatic asset identification functionalities.

6 Forensic Readiness during all SCADA Lifecycle Phases

Incorporating forensic readiness into the SCADA system lifecycle phases, e.g. during development, leads to more successful control of security incidents. Examples include configuration management, change management, and validation and verification.

6.1 Configuration Management Processes and Assessment

Configuration management process is essential for the security of the SCADA hardware and software configurations. Each system alteration could have severe consequences on

its security [Pa13]. Therefore, system alteration must be properly assessed. Any modification could bring vulnerabilities that weaken security. Configuration management starts with properly tested and acknowledged security baselines for SCADA systems [Pa13]. Strong performance assessments are required in order to get successful forensic readiness during all SCADA lifecycle phases. Regular vulnerability assessment and automated auditing of the network and systems are essential part of the configuration management process [Pa13]. During development, it must be assured what is logged for changes of source code. In addition, reliance is on a configuration management system with logging enabled (assumptions, e.g. like ClearCase, that code is deleted, but all operations, including deletion, are recorded).

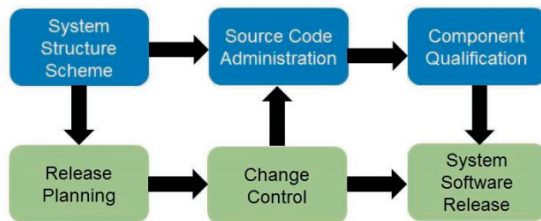


Fig. 7: Configuration Control

Configuration controls as outlined in Fig. 7 is also important against the insider threat as it ensure that I&C software employed by the customer is not maliciously manipulated during the software development lifecycle and during the software employment.

Each software component is uniquely identified according to a scheme of structured configuration IDs. Thus, e.g. a function block library (which may be mandatory for a high Safety Integrity Level and accordingly for e.g. SL3 or SL4 according to IEC 62443) is identified as one component and additionally each included function block is identified as an individual configuration management entity.

6.2 Change management

The purpose of change management is to manage the consequences of alterations in SCADA system configurations. In addition, it also prevents modifications that could have harmful impact to the security posture of a system. Therefore, modifications to an asset are only performed in a controlled manner. The process of managing changes reduces the risk of any changes made to a system, e.g. as insertions, installations, deletions and modifications that result in a compromise to system or data. To minimize risks of possible unfavorable consequences, the change management procedure requires verification of changes prior to implementation. In addition, risk assessment is conducted on all changes to the SCADA network that could have an impact on security, including configuration changes, the addition of network components, and installation of software [NIST11]. The current SCADA network configuration must always be well-known and documented [NIST11].

6.3 Verification and Validation (V&V)

V&V processes are associated to the analysis, evaluation, review, inspection, assessment, and testing of systems, software, hardware, and their interfaces in different lifecycle phases [ISII13]. These processes ensure that the final product meets the design anticipations. V&V is done using a graded approach according to IEC 62443.

7 Conclusion

Modern SCADA systems are dependent on information technology to operate efficiently. Understanding and monitoring SCADA networks and asset management are essential elements in discovering cyberattacks and performing subsequent forensic analyses [LL15]. During digital forensics investigation, proper time stamping and recording of events is required but it is achieved only if a high precision reference clock is available. In addition, many challenges with regular collecting of logging information exist, e.g. restricted access to high security zones. Providing additional access, just to read the logging data brings in security risks. Furthermore, an essential aspect in preparing maintenance equipment for regular use during outages or for emergency situations is the incorporation of forensic readiness. The main idea is to assure that it was not manipulated, e.g. somebody replaces the software to be locally loaded. In this case, firewalls or a Physically Unidirectional Security Gateways [ISIE14a] are not applicable. Further, additional scanning before loading the new software to embedded devices is needed. Forensic-related security controls should be also implemented, reviewed and tested according to security level. Auditing on workstations, servers, and network devices must be permitted; audit records should be stored on centralized log servers. Incorporating forensic readiness into the SCADA system lifecycle phases, e.g. during development, leads to more successful control of security incidents. Examples include configuration management processes and assessment, change management, and V&V.

References

- [AIJ12] Al-Nemrat, A.; Ibrahim, N.; Jahankhan, H.: Sufficiency of Windows event log as evidence in digital forensics. University of East London, London, 2012.
- [Ba16a] Bajramovic, E. et.al.: Cybersecurity Aspects in the I&C Design of Nuclear Power Plants. 3rd INPPS, Istanbul, 03.2016.
- [Ba16b] Bajramovic, E. et.al.: Forensic Readiness of Smart Buildings: Preconditions for Subsequent Cybersecurity Tests, Trento, 09.2016 (unpublished yet).
- [Bm16] Studie i.A. des BMWi IT-Sicherheit für die Industrie 4.0 Produktion, Produkte, Dienste von morgen im Zeichen globalisierter Wertschöpfungsketten, 2016.
- [Ed15] Eden, P. et.al.: A Forensic Taxonomy of SCADA Systems and Approach to Incident Response. 3rd International Symp. for ICS & SCADA Cyber Sec. Research, 2015.

-
- [EGI13] ETSI GS ISI 002: Information Security Indicators (ISI) — Event Model — A security event classification model and taxonomy, 2013.
- [EGI14] ETSI GS ISI 003: Information Security Indicators (ISI) — Key Performance Security Indicators (KPSI) to evaluate the maturity of sec. event detection, 2014.
- [Hs08] Recommended Practice: Creating Cyber Forensics Plans for Control Systems. Homeland Security, 08.2008.
- [Hs09] Recommended Practice: Developing an ICS CSIR Capability, 10.2009.
- [Hs16] Preparing for Cyber Incident Analysis. The National Cybersecurity and Communications Integration Center, Homeland Security, 2016.
- [HS15] Hollern, J.; Stringfellow, P.: Considerations for Integrating CS Reqs into the Nuclear Facility Emergency Preparedness Plan, 9th NPIC, Charlotte, 2015.
- [IEC611] IEC 61513: NPPs — I&C Sys. Important to Safety — General Req. for Systems, 2011.
- [IEC613] IEC 62443-3-3: Industrial communication networks – Network and system security – Part 3-3: System security requirements and security levels, 2013.
- [IEC614] IEC 62645: NPPs – I&C–Req. for sec. programmes for computer-based systems, 2014.
- [IEC615] IEC 62443-4-2: Security for IACS – Part 4-2: Technical security requirements for IACS components, 2015.
- [ISO514a] ISO 55001: Asset management — Overview, principles and terminology, 2014.
- [ISO514b] ISO 55001: Asset management — Management systems — Requirements, 2014.
- [ISIE11a] ISO/IEC 27035: IT — Security techniques — Inf. security incident management, 2011.
- [ISIE12a] ISO/IEC 19770-1: Software Asset Management — Processes and tiered assessment of conformance, 2012.
- [ISIE12b] ISO/IEC 27037: IT — Security techniques — Guidelines for identification, collection, acquisition and preservation of digital evidence, 2012.
- [ISIE14a] ISO/IEC 27033-4: IT — Security techniques — Network security — Part 4: Securing communications between networks using security gateways, 2014.
- [ISIE14b] ISO/IEC 27036-2: IT — Security techniques — Information security for supplier relationships — Part 2: Requirements, 2014.
- [ISIE14c] ISO/IEC 27038: IT — Sec. techniques — Specification for dig. redaction, 2014.
- [ISIE15a] ISO/IEC 19770-2: SW Asset Management — Software Identification Tag, 2015.
- [ISIE15b] ISO/IEC 19770-5: IT Asset Management — Overview and Vocabulary, 2015.
- [ISIE15c] ISO/IEC 27040: IT — Security techniques — Storage security, 2015.
- [ISIE15d] ISO/IEC 27041: IT — Security techniques — Guidance on assuring suitability and adequacy of incident investigative method, 2015.

- [ISIE15e] ISO/IEC 27042: IT — Security techniques — Guidelines for the analysis and interpretation of digital evidence, 2015.
- [ISIE15f] ISO/IEC 27043: IT — Security techniques — Incident investigation principles and processes, 2015.
- [ISIE16a] ISO/IEC DIS 27050-1[Draft]: IT — Security techniques — Electronic discovery — Part 1: Overview and concepts, 2016.
- [ISIE16b] ISO/IEC WD 27050-2 [Draft]: IT — Sec. techniques — Electronic discovery — Part 2: Guidance for governance and management of electronic discovery, 2016.
- [ISII13] ISO/IEC/IEEE 29119-1: Software and systems engineering — Software testing — Part 1: Concepts and definitions, 2013.
- [KL14] Knapp, E.; Langill, J.: Security Monitoring of Industrial Control Systems. In Industrial Network Security. 2nd Edition. Syngress Publishing, 29.12.2014.
- [Le15] Lee, R.: Active cyber defense cycle: Asset identification and network security monitoring. Control Engineering, 03.06.2015.
- [LL15] Lee, R.; Luallen, M.: Making digital forensics a critical part of your cyber security defenses, Control Engineering, 15.01.2015.
- [LW15] Lillo, E.; Waedt, K.: Challenges in Considering National and International CS Requirements and Performing a Criticality Analysis, IAEA Conf., Vienna, 06. 2015.
- [MT15] Martyak, P.; Thow, M.: Enhancing Defense-in-Depth and Monitoring Programs to Protect CDAs from Tampering, NPIC & HMIT Charlotte, 2015.
- [NERC09a] Security Guideline for the Electricity Sector: Identifying Critical Cyber Assets. NERC, 17.10.2009.
- [NERC09b] Security Guideline for the Electricity Sector: Time Stamping of Operational Data Logs. NERC, 2009.
- [NIST11] NIST SP 800-82: Guide to Industrial Control Systems Security, 06.2011.
- [NIST06] NIST SP 800-92: Guide to Computer Security Log Management, 09.2006.
- [Pa13] Paganini, P.: Improving SCADA System Security. INFOSEC Institute. 06.12.2013.
- [Sc15] Scott, A.: Tactical Data Diodes in IACS. SANS Institute, 18.05.2015.
- [Ta13] Taveras, P.: SCADA Live Forensics: Real Time Data Acquisition Process to Detect, Prevent or Evaluate Critical Situations. 1st Internat. Conf., Azores, 24.04.2013.
- [Wa16] Waedt, K. et.al.: Automatic Assets Identification for Smart Cities: Prerequisites for Cybersecurity Risk Assessments, Trento, 09.2016 (unpublished yet).
- [WD14] Waedt, K.; Ding, Y.: IT Security / Interoperability. 1st Sino-German Intelligent Manufacturing/Industry 4.0. Standardization Summit Forum, Shanghai, 16.12.2015.
- [WLZ15] Waedt, K.; Lillo, E.; Zavarsky, P.: Identification of the Critical Components of an ICS and Options to Protect Them, WINS. Workshop on Effective Integration of Physical Protection and Cyber Security, Vienna, 02. 2015.

Risikomodelle von komplexen Systemen

Risikomodelle von komplexen Systemen

Erich Schweighofer¹ und Felix Breitenecker²

Abstract: In diesem Workshop wird eine technische wie rechtliche Bewertung von Risikomodelle in komplexen Systemen vorgenommen. Die Risikoanalyse ist seit langem eine technische Methode und seit etwa 10 Jahren auch in der Rechtswissenschaft präsent. In sechs Beiträgen wird erstmals ein interdisziplinärer Überblick über den Forschungsstand mit einem Schwerpunkt in den Rechtswissenschaften gegeben.

Keywords: Risikoanalyse, Methodik, Nachvollziehbarkeit, Transparenz, Rechtsschutz

1 Darstellung des Workshops

Das Generalthema der Jahrestagung 2016 „Informatik: von Menschen für Menschen“ betont, dass Informatiksysteme der Zukunft den Menschen in den Mittelpunkt zu stellen haben: mit seinen Besonderheiten, Vorteilen, Nachteilen, in einer zunehmend komplexen Welt mit steigender Dominanz der IT.

In diesem Workshop wird eine technische wie rechtliche Bewertung von komplexen Systemen vorgenommen. Die Risikoanalyse ist seit langem eine technische Methode und seit etwa 10 Jahren auch in der Rechtswissenschaft präsent.

Für die Bewertung der Probleme der komplexen Gesellschaft und die Entwicklung von Lösungen – technischer wie juristischer Art - bedarf es eines interdisziplinären Dialogs. Dieser Workshop soll für Informatiker und Juristen diese wichtige Brückenfunktion erfüllen und unterschiedliche Themenbereiche umfassen.

2 Beiträge

Insgesamt wurden sechs der eingereichten Beiträge vom Programmkomitee angenommen. Die Themen sind sehr vielschichtig, aber durch die gemeinsame Sichtweise hinsichtlich des Risikos und einer Bewertungsmethodik verknüpft: Modellentwurf einer Risikoanalyse im Recht, Informationssicherheitsrisiken in komplexen Systemen bzw. in Klein- und Mittelbetrieben, Predictive Analytics,

¹Universität Wien, Arbeitsgruppe Rechtsinformatik, Schottenbastei 10-16/2/5, A-1010 Wien, erich.schweighofer@univie.ac.at.

²Technische Universität Wien, Institut für Analysis und Scientific Computing, Mathematische Modellierung und Simulation, Wiedner Hauptstraße 8-10, A-1040 Wien, felix.breitenecker@tuwien.ac.at.

Risikobewertung im Gasmarkt sowie Risikoanalyse durch agentenbasierte Simulation.

- Erich Schweighofer und Thomas Preiß, Risikoanalyse im Recht – eine neue juristische Methodik?
- Ingrid Schaumüller-Bichl und Andrea Kolberger, Information Security Risk Analysis in komplexen Systemen - neue Herausforderungen und Lösungsansätze
- Manfred Holzbach, Effiziente Risikoanalyse anhand praktischer Erfahrungsbeispiele
- Martin Setnicka, Predictive Analytics in der österreichischen Finanzverwaltung
- Maik Günther, Einsatz eines Gasmarktmodells zur Bewertung von Risiken
- Thomas Preiß, Konzept einer Risikoanalyse im Verwaltungsverfahren durch agentenbasierte Simulation

Die eingereichten vierseitigen Abstracts bzw. vollen Beiträge wurden durch das Programmkomitee begutachtet; die viele Kommentare sind in die finalen Beiträge eingeflossen.

3 Programmkomitee

Ao. Univ.-Prof. DI Dr. techn. Felix Breitenecker, TU Wien (Co-Tagungsleitung)
Ao. Univ.-Prof. Mag. DDr. Erich Schweighofer, Universität Wien (Co-Tagungsleitung)
Prof. Dr. Georg Borges, Universität des Saarlands
Dr. Gottfried Endel, Hauptverband der österreichischen Sozialversicherungen
Associate Professor Dr. Tobias Mahler, Universität Oslo
DI DI Dr. techn. Andreas Körner, Technische Universität Wien
DI Dr. Dr. Thomas Preiß, Amt der NÖ. Landesregierung/Universität Wien
DI. Dr. techn. Niki Popper, Technische Universität Wien
Univ.-Prof. Dr. iur. Dr. techn. Gerald Quirchmayr, Universität Wien
Assoz.-Prof. DI Mag. Dr. Michael Sonntag, Universität Linz
Prof. Dr. Peter Wahlgren, Universität Stockholm
Prof. Dr. Arthur Winter, Donau-Universität Krems

4 Dank

Den Autorinnen und Autoren gebührt unser herzlicher Dank für ihre Beiträge, die mit größtmöglicher Sorgfalt editiert wurden. Dem Programmkomitee ist es zu verdanken, dass möglichst viele Aspekte dieses Themas in den Beiträgen berücksichtigt werden konnten. Erstmals kann daher ein interdisziplinäres Workshop über Risikoanalyse mit besonderer Berücksichtigung der Rechtswissenschaften abgehalten werden.

Risikoanalyse im Recht – eine neue juristische Methodik?

Erich Schweighofer¹ und Thomas Preiß²

Abstract: Technische Methoden spielen in den Rechtswissenschaften schon lange eine wesentlich größere Rolle als üblicherweise angenommen. Es ist schon lange Standard in der Verwaltung, die Anbringen der Parteien zu speichern, zu sortieren, zu gruppieren, kurzum um anfallende Daten effizient und unter Beachtung gesetzlicher Vorgaben zulässig zu verarbeiten (seit etwa 20 Jahren unter dem Schlagwort E-Government). Nunmehr zielt der Einsatz wissensbasierender Systeme auf die Unterstützung rechtlicher Entscheidungsprozesse ab. Dies erfolgt entweder durch Analyse entsprechender Rechtsverfahren bzw. durch logische Verknüpfung rechtsrelevanter Texte mit der zugehörigen Judikatur. Diese Effizienzverbesserung muss aber von einer Risikoanalyse begleitet werden, damit die Grundwertungen der Rechtsordnung und die Menschenrechte, insbes. die Verfahrensgrundsätze, beachtet werden. Die bisherigen Ansätze liefern die Grundlagen für eine Systematik, um ein erstes Modell einer Risikoanalyse im Recht vorzulegen.

Keywords: Risikoanalyse, Methodik, Nachvollziehbarkeit, Transparenz, Rechtsschutz, Grundrechte

1 Einleitung

Ubi societas ibi ius. Wo es eine Gesellschaft gibt, gibt es ein Gesetz. Das Recht schafft die Rahmenbedingungen für unsere moderne Gesellschaft und den globalen Austausch von Waren- und Dienstleistungen. Ohne Recht wäre nicht das Vertrauen [Lu68] vorhanden, dass für die moderne Wissens- und Netzwerkgesellschaft [Sa15] unabdingbar ist. Vom Recht wird heute nicht mehr nur Streitbeilegung erwartet, sondern eine Vermeidung von rechtlichen Konflikten durch anerkannte und überall zugängliche Rechtsregeln. “While legal problem solving will not be eliminated in tomorrow’s legal paradigm, it will nonetheless diminish markedly in significance. The emphasis will shift towards legal risk management supported by proactive facilities, which will be available in the form of legal information services and procedures.” [Su96] Recht wird von der Wissenschaft der Streitbeilegung zur Wissenschaft der Streitvermeidung. Die Analyse der jeweiligen Risiken wird ein Kern der Methodik, um ein wesentliches Ziel der Rechtsordnung, die Bewahrung von Frieden und Ordnung, bestmöglich zu erreichen.

Risiko wird von der [ISO09] als Wirkung von Unsicherheit über Ziele verstanden [Ma10]. [Be86] weist in seinem populären Buch über die Risikogesellschaft auf die zentrale Bedeutung des Risikos hin – die Vielfältigkeit möglicher Schädigungsursachen unserer modernen Gesellschaft. Die Soziologie benutzt den Risikobegriff, um die Wirkung einer

¹ Universität Wien, Arbeitsgruppe Rechtsinformatik, Schottenbastei 10-16/2/5, A-1010 Wien, erich.schweighofer@univie.ac.at

² Amt der NÖ Landesregierung, Wissenschaftlicher Dienst / Universität Wien, thomas.preiss@aon.at

Entscheidung mit bzw. ohne Wissen zu beobachten und rückt Akteure und deren Entscheidungen und Folgenerwartungen ins Zentrum. Eine umfassende Risikodefinition bringen [KB14]: „Risiko ist die Entscheidung, einen Nutzen zu genießen und dabei einen zukünftigen Schaden mit einer mehr oder weniger gut bestimmbarer Eintrittswahrscheinlichkeit und einem ungewissen Ausmaß in Kauf zu nehmen“.

Ob, welche und wie viele Regeln in unserer Gesellschaft nötig sind, ist eine Abschätzung des Risikos möglicher Konflikte und deren sozialadäquater Beilegung. Statistisch bewertbare Faktoren spielen eine wesentliche Rolle: Akteure, Transaktionen, Konflikte, Schaden, Schadenseintrittswahrscheinlichkeit, Strafen etc. Moderne Rechtssysteme sind durch eine intensive Regulierung charakterisiert, deren Regeln über Rechtsinformationssysteme weitgehend kostenfrei zugänglich sind. Ob diese Regeln ihren Zweck erfüllen und Streit vermeiden und im Konfliktfall fair und gerecht lösen, ist eine andere Frage. Menschenrechtskataloge, insbes. die EU Grundrechtecharta, geben die wesentlichen Prinzipien des Rechtssystems vor.

Die Entscheidung über die Abschätzung des Risikos ist Sache der Gesetzgebung bzw. der Höchstgerichte. Diese stellen fest, ob es zu viele oder zu wenige Regeln gibt, wie diese zu interpretieren sind und verändern – wenn notwendig – das Rechtssystem und legen die Streitschlichtungsmechanismen fest. Das Rechtssystem als Wertungskatalog der Risiken kann grundrechtlich insbes. dann angefochten werden, wenn das Recht auf ein faires Verfahren verletzt wird.

Das Recht auf ein faires Verfahren (Art. 6 der Europäischen Menschenrechtskonvention, sowie Art. 47 EU Grundrechtecharta in Verbindung mit der Entscheidung U466/11 vom 14.3.2012 des Verfassungsgerichtshofs, die besagt, dass die „Grundrechte-Charta der Europäischen Union (...) wie die Verfassung zu sehen“ ist) bestimmt den Ablauf einer Risikoanalyse. Als wesentliche Norm der Risikoanalyse im Recht ist daher Art. 47 der EU Grundrechtecharta - Recht auf einen wirksamen Rechtsbehelf und ein unparteiisches Gericht anzusehen: „(1) Jede Person, deren durch das Recht der Union garantierte Rechte oder Freiheiten verletzt worden sind, hat das Recht, nach Maßgabe der in diesem Artikel vorgesehenen Bedingungen bei einem Gericht einen wirksamen Rechtsbehelf einzulegen. (2) Jede Person hat ein Recht darauf, dass ihre Sache von einem unabhängigen, unparteiischen und zuvor durch Gesetz errichteten Gericht in einem fairen Verfahren, öffentlich und innerhalb angemessener Frist verhandelt wird. Jede Person kann sich beraten, verteidigen und vertreten lassen. ...“ Es widerspricht dem Recht auf ein faires Verfahren, wenn die Bürger durch Regeln unsachlichen Risiken ausgesetzt sind. Wenn der Gesetzgeber nicht selbst die Regeln im Detail festlegt, tut er gut daran, Risikoanalysen als Vorgabe für ergänzende Regeln festzulegen. In jüngster Zeit ist dies bei der Datenschutz-Folgenabschätzung (Art. 35 DSGVO) erfolgt. Der österreichische Gesetzgeber hat im Gesundheits- und Ernährungssicherheitsgesetz (GESG, § 2) die „Risikoanalyse“ folgendermaßen definiert: als grundlegenden Prozess zur Gewinnung von Informationen und zur Entscheidung über Maßnahmen gesehen, der als geeignet gesehen wird, Risiken zu vermeiden. Das Ziel der Gefahrenforschung und -einschätzung ist die Auswirkungen, insbesondere der Schaden eines Gefahrenereignisses sollen dadurch

möglichst gering gehalten werden. Sinnvolle Regeln mit wirksamem Rechtsschutz sind das Hauptinstrument des Rechts zur Vermeidung von Gefahrenereignissen.

Der Gefahrbegriff ist dahingehend zu verstehen, dass ein mit staatlichen Zwangsmitteln herbeigeführter Zustand (als Ergebnis eines Gerichtsprozesses oder als Ergebnis eines Verwaltungsverfahrens) fehlerhaft ist. Aufgrund der Souveränität der Parlamente und dem Interesse an der Aufrechterhaltung einer grundsätzlichen Rechtssicherheit ist dies somit in erster Linie eine Aufgabe der Rechtspolitik. Nur wesentliche Fehler können im Wege des Rechtsschutzes bekämpft werden.

Diese Grundforderung an ein Rechtssystem kann durch eine Risikoanalyse wesentlich besser analysiert werden, als dies bisher der Fall war. Die Risikoanalyse liefert somit wissenschaftliche Grundlagen für die Rechtspolitik bzw. die grundrechtliche Bewertung des Rechtssystems.

Der weitere Teil des Beitrags ist folgendermaßen strukturiert: Initial werden methodologische Ansatzpunkte zur juristischen Risikoanalyse diskutiert, die gleichsam am Grad der gewählten Abstraktion differenziert werden. Danach folgt die dogmatische Konkretisierung eines Einsatzbereiches, der von einer bereits sehr in die Informationsverarbeitung verlagerte Vollziehung der betrachteten Rechtsmaterie ausgeht. Schließlich werden deren Applikationen und allfällige Aspekte weiterer Forschung beleuchtet.

2 Ansätze zur juristischen Risikoanalyse

Den Rechtsordnungen sind Risikoanalysen nicht fremd; diese wurden bisher jedoch nur in Randbereichen eingesetzt. Ausgangspunkt sind jeweils die Verfassung, die Grundrechte, Menschenrechtsverträge und darauf aufbauende Verfahrensrechte. Gelegentlich wird auch auf nicht positivierte Prinzipien wie die „zwingenden Denkgesetze“ Bezug genommen. Der Staat muss sich immer der Organe bedienen und diese haben ein Fehlerrisiko, das von der Rechtsordnung berücksichtigt werden muss. Der relevante Unterschied im Einsatz von menschlichen oder technischen „Beamten“ liegt in der unterschiedlichen Risikoanalyse, und zwar von Seiten der Behörde als auch insbes. des Bürgers. Hier ist das Recht auf ein faires Verfahren zu berücksichtigen. Auch elektronische IT Systeme, die diesem Zweck gewidmet sind und durch die Behörde entsprechend verwendet und parametrier³ werden, haben sicherzustellen, dass trotz der Verarbeitung großer Datenmengen dem Normunterworfenen ein gerechtes Verfahren zukommt. Es gibt bereits eine – wenn auch eher bescheidene – Judikatur, welche die Relevanz derartiger Risikofaktoren aufzeigt und unter dem Gesichtspunkt der

³ Hier kann es sich um eine Ermächtigung der Behörde zur Erlassung entsprechender Verordnungen (vgl. § 49a Abs. 1 VStG) handeln.

Verfahrensgerechtigkeit abhandelt. Der Blickpunkt der Rechtsphilosophie zur Gerechtigkeit an sich bietet regelmäßig eine wichtige Argumentationsgrundlage.

Auf Basis dieser Vorüberlegungen und der im Ansatz bestehenden Rechtsprechung kann aufgrund der Grundwertungen der Rechtsordnung und der Grundrechte, logischer Gesetzmäßigkeiten und rechtstheoretischen Überlegungen ein Modellentwurf zur Risikoanalyse vorgelegt werden. Dieser zeigt die wesentlichen Einflussfaktoren auf und ist einer statistischen Untersuchung zugänglich. Mit dieser Basis kann ein Risikomanagementprozess aufgebaut werden. Die wichtigen Vorarbeiten zu dieser Modellierung des Rechts aus der Sicht einer Risikoanalyse sind Graphentheorie und formale Sprachen, Statistik, Diskrete Simulation in ereignis- oder prozessorientierter Form sowie Hybride Modellbildung. In allen Modellen ist eine gewisse Formalisierung des Rechts jenseits der Sprache die Voraussetzung (vgl. zu sehr frühen Ansätzen [Re75]).

Graphentheorie und formale Sprachen: Die „nordische Schule“ ([Ma10] sowie [Wa13]) hat einen „dogmatischeren“ Ansatz gewählt. Es wird eine visualisierte „Sprache“, die statt Zeichen des Alphabets graphische Elemente⁴ verwendet, definiert, um den betrachteten juristischen Sachverhalt zu beschreiben. Diese Visualisierungen orientieren sich vermehrt an bestimmten Situationen, die bei Vertragsabschluss, bei der Erfüllung eines Vertrags sowie bei der Anrufung eines Gerichts eintreten können und eine Streitbeilegung erfordern. Es ist hier wesentlich, den betrachteten Prozess ausreichend gut zu kennen und darauf aufbauend die graphische Modellierungsmethode anzuwenden. Die „Einsicht“ über den Prozess stellt sicher, dass die Ergebnisse der Analyse valide sein werden; eine unmittelbare Verwertbarkeit des Modells „A“ auf andere Sachverhalte „B“ ist aber nur dann möglich, wenn deren Struktur vergleichbar ist: „ $A \sim B$ “. Die Ergebnisse, die mit diesem Ansatz erhalten werden, stellen sich als valide heraus, da eine beinahe unmittelbare Abbildung des realen Systems erfolgt. Diese Methodik ist sehr gut geeignet, die Begrifflichkeit sowie die jeweiligen Risiken aufzuzeigen bzw. auch Analysefaktoren und deren Gewichtung zu benennen.

Statistik: [SNM14] sowie [Wh08] wählen analytische Modelle, die der Ökonomie bzw. der Statistik entnommen sind und eine Schätzung von Eintrittswahrscheinlichkeiten des schädigenden Ereignisses ermöglichen. Die dem betrachteten System innewohnende Struktur wird nur im erforderlichen Maße zur Bildung des Modells herangezogen; weitere funktionale Abhängigkeiten werden durch (logistische) Regressionen erschlossen. Somit können Aussagen über Risikomanifestationen innerhalb bestimmter Konfidenzintervalle von Systemparametern erstellt werden. Es sind mittels geeigneter Software rasch Ergebnisse erzielbar, wobei die dem System innewohnende Struktur unerkannt bleibt („Black Box-Modellbildung“). Zusätzliche Modellparameter können auch in Hinblick auf Gestaltung von Strafen abgeleitet werden [Wh08].

⁴ Es ist zulässig, diesen Ansatz mit der Unified Modeling Language (UML) zu vergleichen. Website <http://www.omg.org/spec/UML/2.5/>, abgerufen am 21.6.2016. Darauf basierend wurde auch eine Sprache zur Systemdefinition „Systems Modeling Language (SysML)“ <http://www.omg.sysml.org/>, abgerufen am 21.6.2016, spezifiziert. Der Unterschied liegt darin, dass UML die Beschreibung beliebiger, SysML die Beschreibung physischer Systeme ermöglicht.

Diskrete Simulation in ereignis- oder prozessorientierter Form: [EPB13] verwenden für die Modellierung Methoden der diskreten Simulation, die nur mittelbar rechtliche Sachverhalte repräsentieren. Es werden die diskreten („konkreten“) Zeitpunkte beschrieben, an denen das System seinen Zustand ändert. Um dies konkret tun zu können, sind umfassende Systembeobachtungen und Datensammlungen erforderlich. Bei der ereignisorientierten Simulation werden aufgrund der Daten Wahrscheinlichkeitsverteilungen ermittelt, die dann eine große Menge von Datensätzen im Computer erzeugen, wodurch allfällig nicht beobachtbare Risikoeintrittswahrscheinlichkeiten und Schäden durch Risikomanifestationen beschrieben werden können.

Hybride Modellbildung: Bereits [TB90] haben erkannt, dass es Vorteile bringt, Methoden zu koppeln, um ein reproduzierbares, in geringer Zeit ermittelbares Simulationsergebnis zu erhalten. Graphische Methoden helfen, Prozesse effizient und valide zu beschreiben und unterstützen damit, dass eine große Menge vorliegender Daten wesentlich zu einer Ergebnisabsicherung beitragen kann. [EPB13] haben als hybriden Ansatz die Agentenbasierte Simulation gewählt. Unter der Voraussetzung, dass ein hierfür geeignetes Computerprogramm als Werkzeug gewählt wird⁵, wird pro im System vorkommender Entität („Agent“) das Systemverhalten mittels graphischer Beschreibungssprachen modelliert. Der Computer erzeugt im Rahmen eines Simulationsexperiments eine große Zahl von Agenten, deren Systemverhalten und dadurch ermittelte Parameter mittels statistischer Methoden analysiert werden können. Durch eine Vielzahl derartiger Experimente könnten wesentliche Grundlagen zur Modellvalidierung erstellt werden. Liegt ein überprüfbares Modell vor, ist eine Untersuchung möglicher Risikofaktoren beinahe unmittelbar möglich.

3 Modell einer juristischen Risikoanalyse

Ausgehend von rechtstheoretischen Überlegungen der Gerechtigkeit und des Rechtsschutzes in Hinblick auf die Verteilung des Risikos zwischen dem Staat und dem Bürger unterscheidet dieses Modell vier Analysephasen (basierend auf [Pr15]):

Anwendungsprozessanalyse: Als Grundlage bedarf es einer exakten Abbildung des Rechtsanwendungsprozesses: Prinzipien, Rechtsinstitute, Regeln, Verfahren und Rollen. Bei einer technologischen Unterstützung des Verwaltungshandels durch IT ist die Architektur des Softwaremodells zu beschreiben. Als Abbildung dieses Anwendungsprozesses eignen sich Modelle mit einer ontologischen und rechtslogischen Struktur, d.h. durch Verwendung einer informatorischen Beschreibungssprache (z.B. SysML). Hiermit werden die Entscheidungsfaktoren und Entscheidungsprozesse ausreichend präzisiert und vervollständigt.

⁵ Hier handelte es sich um AnyLogic, <http://www.anylogic.com>, abgerufen am 21.6.2016.

Risikofaktorenanalyse: Den Forschungen der Rechtstheorie zur Risikoanalyse ist zu entnehmen, dass aus ethischen Gründen das Risiko nicht ungleichgewichtig dem Bürger auferlegt werden darf. Entsprechend den erkannten Risiken muss der Rechtsschutz entsprechend gestaltet sein. Dies bedeutet, dass eine automatisierte Entscheidung ohne Kontrolle durch Organwalter und/oder entsprechenden Rechtsschutz unzulässig ist. Dieses Ergebnis ist auch in der EU Datenschutzrichtlinie sowie in der neuen EU Datenschutzgrundverordnung positiviert. Da in der österreichischen Verwaltung bei automatisierten Strafverfügungen nur die Anonymverfügung eingesetzt wird, gibt es keine praktischen Rechtsschutzprobleme. Im Bereich der FinanzOnline geht man etwas weiter und setzt Risikokontrollmethoden ein, weil die Kontrolle durch Organwalter wegen der Massenverfahren faktisch nur mehr bescheiden sein kann.

Faktenanalyse: Für eine Risikobewertung muss relevantes statisches Datenmaterial generiert werden. Die statistische Analyse ist insbesondere hinsichtlich des Rechtsschutzes notwendig, aber kann auch mögliches Versagen der IT – insbes. durch hohe Fehlerquoten - aufzeigen.

Wirksamkeitsanalyse: Durch festgelegte Parameter wird geprüft, ob das eingesetzte IT-gestützte Verwaltungsverfahren den Kriterien für die Rechtsstaatlichkeit entspricht, d.h. insbes., ob die Verfassungsmäßigkeit der Verfahren im Hinblick auf die Grundrechte sichergestellt ist. Fehleranfällige Verfahren ohne ausreichend Rechtsschutz sind nicht grundrechtskonform.

Abbildung 1 zeigt das Modell im Überblick:

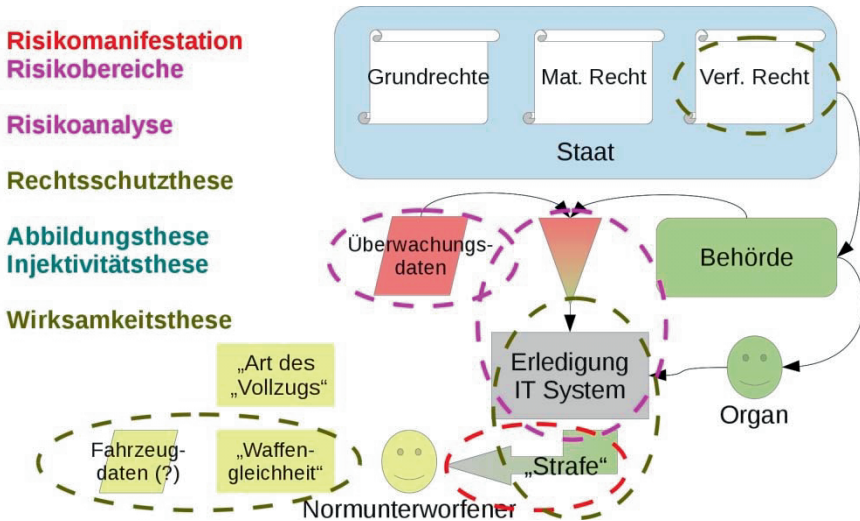


Abb. 1: "Vertikale" Struktur und Ansatzpunkte

Der Zusammenhang zwischen den etablierten Verwaltungsverfahren und den Ansatzpunkten für theoretische und methodische Überlegungen wird als „vertikale“

Struktur bezeichnet. Die vertikale Darstellung bezieht sich darauf, dass gleichsam von „oben“ - den Grundlagen des staatlichen Gemeinwesens - nach „unten“ - dem Aussprechen einer Strafe oder Erledigung im Verwaltungsverfahren - entsprechende Ansätze gefunden werden können. Ebenfalls mit der vertikalen Darstellung vergleichbar ist der Stufenbau der Rechtsordnung (Vgl. [WMK07], Rz 9), der sich an dem Wesen der Normerzeugung bzw. der derogatorischen Kraft von Normen orientiert.

Nochmals auf die Rechtsphilosophie zurückkommend gibt diese, auch in Hinblick auf die verfassungsrechtliche Ausgestaltung, den Rahmen vor, welcher Bereich der Gesetzgebung entsprechende Maßnahmen eines Risikomanagements vorzusehen hat. Hier ist wiederum die Berücksichtigung und Sicherstellung gerechter Verfahren gefragt.

Strukturell wird dieser dogmatische Prozess als „horizontal“ bezeichnet, was durch Abbildung 2 veranschaulicht wird. Der „horizontale“ Charakter findet sich dadurch veranschaulicht, dass gleichsam vom „Beginn“ zum „Ende“ oder von „links“ nach „rechts“ der wissenschaftliche Diskurs innerhalb verschiedener Disziplinen erfolgt. Betrachtungen der Grundlagen führen zur konkreten Ausgestaltung von Modellen, die dann auf die Anwendbarkeit im Rahmen von Verfahrensgerechtigkeit, Vorsorge und Transparenz untersucht werden. Dies fundiert auf wesentlichen Thesen, die diesbezüglich entwickelt wurden. Am anschaulichsten ist dieser Prozess nach der Besprechung der ausgewählten Judikate nachvollziehbar - die daraus entwickelten Thesen⁶ sichern die Zuordenbarkeit allfälliger Risikofaktoren zum Bereich der Verfahrensrealisierung an sich oder der Gewinnung von Überwachungsdaten sowie deren Würdigung.

Die konzeptionelle Neuerung besteht darin, dass ausgehend von rechtsphilosophischen Überlegungen und der Analyse bisheriger Judikatur Thesen zur Lösung der Forschungsfrage entwickelt werden. Weiters stellt die Berücksichtigung des „konträren“ Blickwinkels der Risikobetroffenheit im Zusammenhang mit (Verwaltungs)strafverfahren einen neuen Ansatz dar. Zweifellos sind strafbare Handlungen als Risikomanifestationen geeignet, Schäden zu verursachen. Gleichsam im Gegensatz dazu steht nun der Normunterworfenen als Beschuldigter im Mittelpunkt, der dem Agieren der Verwaltung entspringenden Risikomanifestationen „ausgeliefert“ ist.

Die auf den Thesen aufbauende Risikoanalyse (vgl. [Pr15], Anhang B, 205) verwendet den statistischen Ansatz. Dies ist darauf zurückzuführen, dass der zu analysierende Prozess bereits eine Modellierung und zumindest seitens der Behörde dadurch eine

⁶ Die „Beschlusswortung“ der Thesen erfolgt in Abbildung 1. Kurz zusammengefasst: Die Bereichsthese identifiziert Risikofaktoren innerhalb der betrachteten Rechtsmaterie vor allem im Zusammenhang mit „Bewertung“ und „Prozessen“. Risikoanalyse ist vergleichbar mit den Bestimmungen des GESG zu definieren. Effektives Risikomanagement sieht mit Reviews und Audits vergleichbares Vorgehen vor. Die Rechtsschutzthese fordert - vergleichbar mit dem Vorsorgeprinzip - vom Gesetzgeber ein Vorsehen von Rechtsmitteln gegen allfällige Risikomanifestationen. Die Abbildungsthese besagt, dass die Daten des IT Systems für eine Risikoanalyse geeignet sind. Die Injektivitätsthese sieht unter der Voraussetzung von rechtsökonomischen Grundsätzen (vgl. [No05], 74) ein gerechtes Rechtsergebnis - hier eine gerechte Strafe bei Verwaltungsübertretungen - geben. Die Konkretisierung effektiven Risikomanagements beschreibt die Wirksamkeitsthese (vgl. [Pr15], 215)

Validierung erfahren hat, indem ein entsprechendes IT System grundsätzlich effizient verwendet wird.

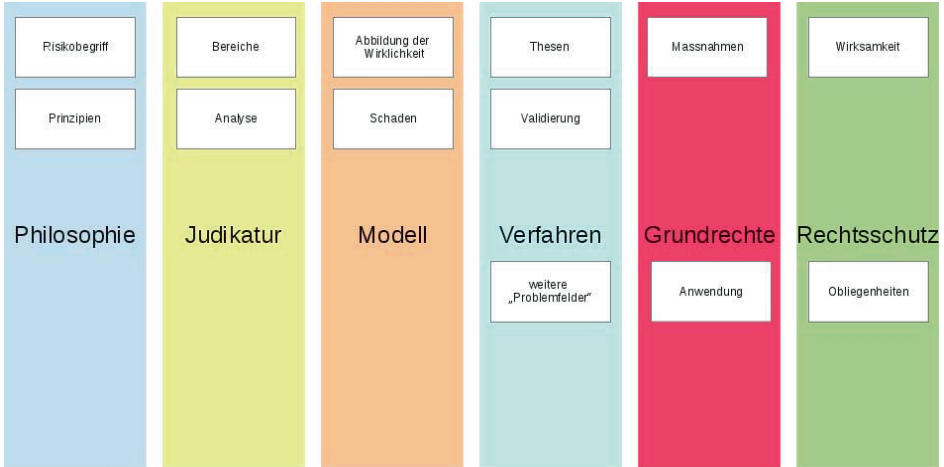


Abb. 2: "Horizontale" Struktur des Risikoanalyseprozesses

4 Einsatzmöglichkeiten der Risikoanalyse

Anhand eines Beispiels im Verwaltungsrecht werden die Einsatzmöglichkeiten für Risikoanalyse im Recht skizziert. Ausgehend von einem etablierten „Massenverfahren“, dem auch ein IT System zugrunde liegt, sollen durch technische, in diesem Fall auch mathematische, Methoden Risiko und Problemfelder - so diese existieren - aufgezeigt werden. Ein zu lösendes Problem stellt die Identifikation der genannten Risikofaktoren dar, die durch die Anwendung der vorgeschlagenen Methoden analysiert werden können.

Der Blick über den „Tellerrand“ des Verwaltungsstrafverfahrens kann weiters dadurch erfolgen, dass Überlegungen zur Gestaltung von diversionellen Maßnahmen gestellt werden. Die Praxis des Verwaltungsstrafverfahrens und dessen Vollzug zeigen, dass diese „Wiedergutmachung ohne Strafausspruch“ auch in diesem Bereich angedacht werden sollte. Diese „minder schwere“ Erledigungsart wäre analog zu den Grundsätzen des Vorsorgeprinzips dann zu überlegen, wenn einerseits dem Normunterworfenen zur objektiv einzusehenden „Unzeit“ eine Bestrafung widerfahren würde, andererseits wenn Einsicht besteht, aber eine Wiedergutmachungshandlung geboten erscheint.

Bei der Skizzierung der Thesen werden Audits und Reviews des Behördenhandelns angeführt. Dazu wird erforderlich sein, dass durch eine Aufsichtsstelle Daten eingesehen werden. Es wird unproblematisch sein, wenn nach Vorgaben der hierzu durch das Gesetz vorgesehenen Aufsichtsstelle anonymisierte bzw. aggregierte Daten seitens der Behörde, als datenschutzrechtlicher Auftraggeber übermittelt werden.

4.1 Beispiele

[Pr15, 10 ff., 30 ff.] sind Beispiele der österreichischen Judikatur im Zusammenhang mit automatisierten Verwaltungsstrafverfahren zu entnehmen, die schließlich dazu geführt haben, dass die dieser Rechtsmaterie zuzurechnenden Risikofaktoren zumeist in den Bereichen der „Bewertung“⁷ und des „Prozesses“⁸ zu finden sind. Wir stellen eine Auswahl an Sachverhalten dar, sowie deren Feststellbarkeit im Rahmen einer Risikoanalyse.

Messgeräte = Risikofaktor „Bewertung“: Es ist zu überprüfen, ob das „Eichalter“ (die Zeitspanne zwischen dem letzten behördlichen Eichungsvorgang und der tatsächlichen Messung) Rückschlüsse auf die Qualität des festgestellten Ergebnisses hat. Diese Untersuchung lässt sich durch Analyse der Varianz der Messungen bei gleichartigen Einsatzbedingungen feststellen.

Hierbei ist die Nullhypothese H_0 aufzustellen, dass die Varianz der Messungen nicht signifikant verschieden ist. Statistische Experimente mit der Programmiersprache SAS haben gezeigt, dass bei einem Signifikanzniveau von $\alpha = 0,05$ die Nullhypothese nicht zu verwerfen ist.

Ausfertigungsdauer = Risikofaktor „Prozess“: Ein wesentliches Maß, wie auch bei Überprüfungen durch den Rechnungshof angewandt, für ein ausreichend performantes Handeln der Behörde ist die Ausfertigungsdauer. Dass hier saisonale Effekte wesentliche Eingangsgrößen bilden, versteht sich von selbst. Wiederum kann durch eine Analyse durch geeignete inferente Statistiken, hier eine Zeitreihenanalyse, der saisonale Aspekt aufgezeigt werden und, vor allem im personellen Bereich, Vorkehrungen getroffen werden. Der Normunterworfene ist von diesem Risikofaktor in Sinne der „prozessualen Widerleglichkeit“ betroffen. Je länger die behauptete Verwaltungsübertretung zurückliegt, desto geringer ist die Wahrscheinlichkeit, dass wirksam ein zutreffendes Gegenteil bewiesen werden kann. Das Open Source Statistiksystem R liefert ein entsprechendes Ergebnis.

Höhe Strafbetrag nach rechtsökonomischen Grundsätzen = Risikofaktor „Bewertung“: Wir haben unter Bezugnahme auf [No05] und [Wh08] ausgeführt, dass unter rechtsökonomischen Aspekten optimale Strafen ermittelt werden können. Unter Anwendung des durch [Wh08] beschriebenen Modells werden die bisher durch die Behörde festgesetzten Strafbeträge analysiert und so gewissermaßen auf Gerechtigkeit überprüft. Folgendes SAS Programm berechnet dies:

```
%MACRO BV_WHITE_MODEL (LOK_VONDAT, LOK_BISDAT) ;
DATA ANZEIGE_AKT_1;
SET ANZEIGE_AKT;
WHERE DATUM_DER_UEBERTRETUNG
```

⁷ Darunter verstehen wir den Einsatz von Messsystemen, aber auch Techniken zur Strafbemessung.

⁸ Hier verstehen wir das Vorgehen der Behörde darunter, das zu einer Beschwerde des Normunterworfenen führt.

```

BETWEEN &LOK_VONDAT AND &LOK_BISDAT;
RUN;
TITLE "ADAPTIERTES MODELL WHITES - KOOPERATION EINNAHMEN
/ SICHERHEIT - ADAPTION VON 'P' - [&LOK_VONDAT -
&LOK_BISDAT]";
PROC NLIN
DATA=ANZEIGE_AKT_1 PLOTS=(FIT DIAGNOSTICS(STATS=NONE)) ;
PARMS
O = 1 A_1 = .5 A_2 = .5 B = 1 P_1 = 1 Q = 1 F = 1;
P = P_1 * (UEBERTREUNG_UM_KM_H**(-1));
MODEL STRAFE = ((O + A_2 * B * UEBERTREUNG_UM_KM_H**2)
/ (A_1 * P )) / UEBERTREUNG_UM_KM_H;
OUTPUT OUT=Y;
RUN;
%MEND;

```

Interessanterweise haben Simulationen ergeben, dass zu 95% eine Entsprechung realisiert wurde; außerhalb dieses Konfidenzbereich liegende Werte waren aber zum Teil signifikant abweichend. Das Vorliegen eines Risikofaktors ist somit nachgewiesen.

Einzahlungsbetrag = Risikofaktor "Bewertung": Obige Beispiele lassen unschwer nachvollziehen, dass Risikofaktoren durch Analyse von Daten elektronischer Systeme identifizierbar sind. Wesentlich schwerer zu fassen sind Risikofaktoren, die auf ein Handeln des Normunterworfenen zurückzuführen sind und keinen ausreichenden „Niederschlag“ in die Verwaltungsdaten finden. So verhält es sich, wenn Angaben zur korrekten Zuordenbarkeit des Strafbetrags (Identifikationsnummer, Betrag, Kontonummer der Behörde) fehlen. Für weitere Betrachtungen setzen wir voraus, dass der Bestrafte keine diesbezügliche Absichtlichkeit aufweist.

Der Gesetzgeber sieht eine schnelle, anonyme Erledigung einer Verwaltungsübertretung als aus einem Rechtsrahmen heraustretendes Entgegenkommen, sodass dem zu Bestrafenden in diesem Zustand des Verfahrens kaum Korrekturmöglichkeiten eingeräumt werden. Ist ein Fehler passiert, dann wird dessen Eintrittsrisiko wohl zur Gänze auf den Normunterworfenen übergewälzt.

Die aus den Prozessrechten bekannte „Waffengleichheit“ von Behörde und Beschuldigten ist nicht gegeben. Die Grundsätze der Transparenz und der Nachvollziehbarkeit des Behördenhandelns gebieten, so diese Risiken nicht abgebildet sind, ausreichende Dokumentation und Bereitstellen von Möglichkeiten und Verfahrensschritten, die ein Reagieren und Korrigieren des Beschuldigten vorsehen. Hier sind die Grundsätze des E-Governments im Sinne der Bürgernähe vermehrt zu beachten. Trotzdem sei, auch im Sinne der unten angegebenen Schlussfolgerungen, angemerkt, dass mathematische Modelle in der Wirtschaft auch „soft facts“ wie Kundenverhalten abbilden können. Diese mittels der Methode des „agent based modeling“ erstellten Systeme können Vorbilder für vergleichbare Anwendungen einer Risikoanalyse im Recht sein.

Die angeführten Beispiele zeigen, dass die großen „Kapitel“ der Modellbildungsmethoden auch bei juristischen Sachverhalten benützt werden können. Eine statistische Untersuchung kann bei Vorliegen entsprechend angelegter Verwaltungsdaten unmittelbar und bei aus heutiger Sicht überschaubarem Zusatzaufwand Nutzen generieren.

5 Schlussfolgerungen

Die Risikoanalyse im Recht ist ein Modell, das noch wenig in der Praxis umgesetzt wurde. Da ein wesentliches Element, das zu einem rechtlich bindenden Handeln des Staates führt, das Handeln des Normunterworfenen ist, bleibt hier noch ausreichend Raum für Untersuchungen von soziologischen Fragen. Dies ist darauf zurückzuführen, dass die bisherigen Modelle allfällige subjektive Momente nicht berücksichtigen. Hier sind die verschiedenen Gründe, die dafür ausschlaggebend sind, aufzuzeigen und in das Modell einzupflegen. Den dogmatischen Hintergrund dafür bildet die Entscheidungstheorie [Be86, 323].

Die Anwendung statistischer Methoden zur Wirksamkeitsanalyse stellt im Zusammenhang mit der „Qualitätssicherung“ IT gestützter Behördenverfahren ein zum Teil übliches Vorgehen dar. Dem Aspekt der Effektivität entsprechend, erscheint uns wichtig, dass vermehrt der Prozessmodellierung Aufmerksamkeit geschenkt wird. In der „horizontalen Struktur“ wird dadurch die Modellkomponente gleichsam gestärkt, da vor der Realisierung eines staatlichen Handelns dessen Auswirkungen wirksam und valide auf Risikofaktoren untersucht werden kann.

Literaturverzeichnis

- [Be86] Beck, U.: Risikogesellschaft. Auf dem Weg in eine andere Moderne. edition surkamp, Frankfurt am Main, 1986.
- [EPB13] Einzinger, P.; Popper, N.; Breiteneker, F.: The GAP-DRG Model Simulation of Outpatient Care for Comparison of Different Reimbursement Schemes. In (Pasupathy, R. et al., Hrsg.): Proceedings of the 2013 Winter Simulation Conference, 2013.
- [ISO09] ISO: IEC 31010: 2009. Risk management -- Risk assessment techniques, 2009.
- [KB14] Kristiansen, S.; Bonfadelli, H.: Risikoberichterstattung und Risikoperzeption. Reaktionen von Medien und Bevölkerung in der Schweiz auf den AKW-Unfall in Fukushima. In (Wolling, J.; Arlt, D., Hrsg.): Fukushima und die Folgen. Medienberichterstattung, Öffentliche Meinung, Politische Konsequenzen. Illmenau, S. 297-321, Open access unter: <http://neu-kommunikation.de>, 2014.
- [Lu68] Luhmann, N.: Vertrauen: Ein Mechanismus der Reduktion sozialer Komplexität. Enke, Stuttgart, 1968.
- [Ma10] Mahler, T.: Legal Risk Management, PhD Thesis, University of Oslo, 2010.

- [No05] Noll, J.: Rechtsökonomie. Verlag Österreich, 2005.
- [Pr15] Preiß, T.: Die Bedeutung der Risikoanalyse für den Rechtsschutz bei automatisierten Verwaltungsstrafverfahren, Dissertation, Universität Wien, 2015.
- [Re75] Reisinger, L.: Planspiel und Simulation im Recht. In: Winkler, G. (Hrsg.), Forschungen aus Staat und Recht 32, S. 148-166. Springer Verlag, Wien, New York, 1975.
- [Sa13] Saarenpää, A.: Legal Informatics Today, The View from The University of Lapland. In (Schweighofer, E.; Saarenpää, A; Böszörményi, J., Hrsg.): Proc. KnowRi\$ht 2012, S. 21-26, OCG, Wien, 2013.
- [SNM14] Surdeanu, M.; Nallapati, R.; Manning, C. D.: Risk Analysis for Intellectual Property Litigation, <http://www.surdeanu.info/mihai/papers/icail11.pdf>, Stand: 15.06.2016, 2014.
- [Su96] Susskind, R.: The Future of Law: Facing the Challenges of Information Technology. Clarendon Press, Oxford, 1996.
- [TB90] Troch I., Breitenecker F.: Simulation in der Regelungstechnik. Volume 12 of the series Fachberichte Simulation, S. 296-323, Springer, 1990.
- [Wa13] Wahlgren, P.: Legal Risk Analysis: A Proactive Legal Method. Juridiska fakulteten vid Stockholms universitet, Stockholm, 2013.
- [Wh08] White, M. D.: Time speeding behavior and optimal penalties. In: The Journal of socioeconomics, 384 - 399, Elsevier, 2008.
- [WMK07] Walter, R., Mayer, H.; Kucsko-Stadlmayer, G.: Grundriss des österreichischen Bundesverfassungsrecht, Manz, Wien, 2007.

Information Security Risk Analysis in komplexen Systemen - neue Herausforderungen und Lösungsansätze

Ingrid Schaumüller-Bichl¹, Andrea Kolberger¹

Abstract: Die Identifikation und Bewertung von Risiken, die die Informationssicherheit bedrohen (Information Security Risk Analysis, ISRA), ist in vielen Systemen von zentraler Bedeutung. Neue Technologien und Entwicklungen, wie etwa Industrie 4.0 oder das Internet der Dinge (Internet of Things, IoT) sowie generell die zunehmende Komplexität der IT-Systeme stellen neue Herausforderungen an die Risikoanalyse und das Risikomanagement. Der Artikel diskutiert die besonderen Anforderungen an ISRA in komplexen Systemen und geht der Frage nach, wie Risikoanalyse im IT- und Informationssicherheitsbereich in Zukunft aussehen kann. Welche Ansätze können übernommen und weiterentwickelt werden, wo braucht es neue Lösungen und wie können diese aussehen?

Keywords: Risikoanalyse, Risikomanagement, Informationssicherheit, IT-Sicherheit, komplexe Systeme

1 Einleitung

Ein systematischer Ansatz zur Identifikation und Bewertung von Risiken ist in der IT- und Informationssicherheit heute wichtiger denn je. Die Risikoanalyse ist Basis für die Auswahl und Implementierung adäquater Sicherheitsmaßnahmen. Zunehmend sehen auch Rechtsvorschriften, Normen, de-facto-Standards und branchenspezifische Vorgaben eine verpflichtende Auseinandersetzung mit Risiken vor. Aktuelles Beispiel dafür ist die neue EU Datenschutz-Grundverordnung. Sie sieht sowohl eine risikogerechte Auswahl von Sicherheitsmaßnahmen vor, als auch in potentiell kritischen Fällen die Durchführung einer DPIA (Data Protection Impact Analysis, Datenschutz-Folgenabschätzung), die als eine spezielle Form der Risikoanalyse im Datenschutzbereich betrachtet werden kann. Weitere Beispiele sind etwa die in ISO/IEC 27001 zentrale Informationssicherheits-Risikoanalyse sowie die Risikoanalyse im Rahmen des BCM Prozesses.

2 Risikoanalyseansätze in der Informationssicherheit – Stand der Technik und Grenzen

Die letzten Jahre haben im Bereich der Risikoanalyse für Informationssicherheit (ISRA)

¹ FH OÖ, Information Security Compliance Center (ISCC), Hafenstraße 47-51, 4020 Linz,
{ingrid.schaumueller-bichl, andrea.kolberger}@fh-ooe.at

einen Paradigmenwechsel gebracht. Während der Risikobegriff früher als möglicher Schaden aus der Ausnutzung von Schwachstellen durch eine Bedrohung definiert war (ISO/IEC 27005: "potential that a given threat will exploit vulnerabilities of an asset or group of assets and thereby cause harm to the organization" ([ISO11]), wird er nun, angelehnt an das generelle Risikomanagement, allgemeiner gefasst. So bezieht sich seit 2013 auch die ISO/IEC 27001, der weltweit verbreitete Standard zum Informationssicherheits-Management, auf ISO 31000 und damit auf folgenden Risikobegriff: "risk is the effect of uncertainty on objectives", definiert also Risiko als Abweichung von Zielen, egal ob positiv oder negativ ([ISO09], [ISO13]). Damit wurde der "klassische" Zugang zum Informationssicherheits-Risikomanagement, nämlich die Ermittlung von Bedrohungen, Schwachstellen und Schäden zur Risikoidentifikation und die Bewertung von Eintrittswahrscheinlichkeiten sowie der Schadenshöhe zur Abschätzung der Risiken erweitert und verallgemeinert.

Risikoanalyse unterscheidet typischerweise zwischen quantitativen und qualitativen Ansätzen. Die in der Literatur teilweise dargestellten semi-quantitativen Ansätze werden in der Folge in diesem Paper unter die qualitativen Ansätze gereiht.

Wie in anderen Bereichen auch haben quantitative Ansätze im IT- und Informationssicherheitsbereich vor allem einen entscheidenden Vorteil: sie erlauben eine monetäre Abschätzung der Schäden und Risiken, und bieten damit eine gute Argumentationsgrundlage für Investitionen in den Sicherheitsbereich. Arbeiten zur Berechnung des "Return on Security Investment" (ROSI) bzw. "Return on Information Security Investment" (ROISI) finden sich u. a. in [St08] und [Ko08].

Trotz wissenschaftlich interessanter quantitativer Verfahren werden in der Praxis im Informationssicherheitsbereich überwiegend qualitative Verfahren eingesetzt. Dies liegt in erster Linie darin begründet, dass eine genaue Abschätzung von Eintrittswahrscheinlichkeiten oder Häufigkeiten in diesem Bereich sehr schwer durchzuführen ist: es gibt wenige verlässliche historische Werte, anders als etwa bei der Häufigkeit von Blitzeinschlägen, Statistiken zur Lebenserwartungen und ähnlichen Ereignissen, wo man auf langjährige Beobachtungen und relativ konstante Entwicklungen zurückgreifen kann. Der IT-Bereich ist hingegen sehr raschen Änderungen unterworfen, die Bedrohungslandschaft ändert sich laufend und mit zunehmender Geschwindigkeit. Einen guten Überblick über die aktuelle Lage und künftig zu erwartende Risiken gibt etwa die Risikolandkarte der ENISA (ENISA threat landscape [EN15]), die auch auf künftig zu erwartende Risiken ("Emerging Risks") eingeht.

In der Praxis kommen in der IT- und Informationstechnologie daher vorwiegend quantitative Verfahren, häufig auf Basis von Risikomatrizen zum Einsatz. Oft wird dabei zur Unterstützung auf vorgegebene Bedrohungs- und Schwachstellenkataloge zurückgegriffen.

Resultat einer Risikoanalyse im Informationssicherheitsbereich ist damit in der Regel eine Fülle von Einzelrisiken, die im Anschluss gesondert und im Detail behandelt werden müssen. Maßnahmen zur Risikobehandlung setzen dort an, wo die Risiken

entstehen, d.h. sie können entweder die Eintrittswahrscheinlichkeit reduzieren oder die Höhe der Auswirkungen oder auf beide Faktoren Einfluss nehmen.

Risikoaggregation und damit die Abschätzung eines Gesamtrisikos für den IT-Bereich oder bestimmte IT-Anwendungen sind in der Praxis hingegen selten anzutreffen und auch schwer zu erreichen.

Zusammenfassend ist zu sagen, dass die heute üblichen Risikoanalyse-Methoden im IT-Bereich eine gute und unabdingbare Basis für die nachfolgende Auswahl von angemessenen, eben risikogerechten, Sicherheitsmaßnahmen darstellen. Eine Abschätzung des Gesamtrisikos für die Informationssicherheit und damit der Auswirkungen auf das gesamte Unternehmen oder die Gesellschaft, ist mit gängigen Verfahren aber nicht oder nur eingeschränkt zu erreichen.

3 Information Security Risk Analysis in komplexen Systemen – spezifische Anforderungen

In komplexen Systemen, wie z.B. industriellen Systemen oder dem Internet der Dinge (IoT), bestehen eine Reihe spezifischer Anforderungen an die Risikoanalyse bzgl. Informationssicherheit und die darauffolgende Risikobehandlung. Nachfolgend werden einige davon - ohne Anspruch auf Vollständigkeit - diskutiert.

Anzahl und Diversität der Komponenten: Komplexe Systeme sind typischerweise geprägt von einer Vielzahl unterschiedlicher Einzelkomponenten. Es müssen die Schwachstellen und potentiellen Angriffspunkte dieser Komponenten sowie die korrekte Konfiguration bekannt sein, was bei einer großen Zahl an unterschiedlichen Komponenten sehr aufwendig sein kann und jedenfalls gute Systemkenntnisse erfordert.

Häufige und rasche Änderungen des Systems: Die große Anzahl an Einzelkomponenten bringt es auch mit sich, dass sich das System meist sehr schnell ändert, neue Komponenten eingebunden werden oder die Komponenten anders vernetzt werden. Jede Änderung am System bringt potentiell neue Bedrohungen, Schwachstellen und Risiken mit sich. Ein Risikoanalyseverfahren sollte möglichst einfach und effizient auf solche Änderungen reagieren können, allerdings wird es in der Praxis nicht möglich sein, Risikoanalysen immer zeitnahe nachzuziehen. Hier sind auch entsprechende Maßnahmen im Systemdesign und im Risikomanagement erforderlich, wie etwa die Entwicklung resilienter Systeme sowie die Einbeziehung von Sensormessdaten und Inputs aus dem SIEM, wie in Kapitel 4 und 5 diskutiert.

Unterschiedliche Schutzziele: Informationssicherheit umfasst immer die Gewährleistung zumindest der Schutzziele Vertraulichkeit, Integrität und Verfügbarkeit, was generell erheblich zur Komplexität von ISRA beiträgt, da die Risiken in Bezug auf alle relevanten Schutzziele zu ermitteln sind. Je nach System können weitere Schutzziele von Bedeutung sein, wie etwa Authentizität, Gewährleistung der Privatsphäre (Privacy),

Zuverlässigkeit oder die Nachweisbarkeit von Aktionen (Non-Repudiation). Gerade in komplexen Systemen kann eine Vielzahl unterschiedlicher Schutzziele, die möglicherweise auch in Konflikt miteinander stehen, gegeben sein.

Konfigurationseinstellungen und deren Überwachung: Risiken entstehen in vielen Fällen nicht nur in Folge von systemimmanenten Schwachstellen, sondern auch durch unzureichende oder fehlerhafte Konfigurationen. Einstellungen müssen korrekt sein und laufend auditiert bzw. überwacht werden, Risiken aus Konfigurationsänderungen müssen in die Gesamtrisikobetrachtung einbezogen werden.

Einsatz von Low-Cost-Komponenten: In komplexen Systemen ist auch damit zu rechnen, dass aufgrund des Kostendrucks Komponenten zum Einsatz kommen, die keine oder schwache Sicherheitsfunktionen enthalten. Eine wichtige Frage für die Risiko- beurteilung ist auch, ob solche Komponenten ausreichende Update- und Steuerungsmöglichkeiten haben. Neue Technologien, wie etwa Physically Unclonable Functions (PUFs), sollen dem gegensteuern und als Kern für Sicherheitsmechanismen, etwa Authentisierungsmechanismen oder kryptographisches Key-Management, auch in solchen Anwendungen fungieren. In jedem Fall ist in der Risikoanalyse diesen Komponenten besondere Aufmerksamkeit zu widmen. Ein Risikoanalyseansatz für Physically Unclonable Functions ist in [KSD14] zu finden.

Abhängigkeiten: Eine Kompromittierung von für sich alleine genommen unkritischen und aus diesem Grund unzureichend geschützten Komponente kann unter Umständen einen erfolgreichen Angriff auf andere Komponenten nach sich ziehen. Insbesondere ist dabei zu bedenken, dass die Zusammenhänge und Abhängigkeiten und die Risiko- propagation in Abhängigkeit von den Schutzzielen (zumindest Vertraulichkeit, Integrität und Verfügbarkeit) unterschiedlich sein können. Während bei der Sicherstellung der Vertraulichkeit meist ein "Defense-in-Depth"-Konzept zum Tragen kommt, also unterschiedliche "Verteidigungslinien" aufgebaut werden, so dass bei Brechen einer dieser Linien noch nicht das Gesamtsystem korrumpiert ist, ist im Bereich der Verfügbarkeit eher die Gefahr gegeben, dass sich der Ausfall kleiner Komponenten kritisch auf wichtigere Komponenten und das Gesamtsystem auswirkt.

Risikoaggregation: Die heute üblichen Risikoanalyseansätze im Bereich Informationssicherheit bieten, wie in Kapitel 2 ausgeführt, eine gute Basis für die Auswahl von Sicherheitsmaßnahmen, aber nur bedingt eine Grundlage für übergreifende Geschäftsentscheidungen, da die Aggregation der Risiken zu einem Gesamtrisiko in den wenigsten Fällen zufriedenstellend gelöst ist. Gerade in komplexen Systemen ist es aber auch erforderlich, das Gesamtrisiko, das von solchen Systemen ausgeht, abzuschätzen, insbesondere Klarheit darüber zu bekommen, ob und in welchem Ausmaß dieses Risiko existenzbedrohend für ein Unternehmen sein kann oder gravierende Auswirkungen auf die Gesellschaft hat. Einen interessanten Ansatz zur Risikoaggregation auf Basis eines Graphenkonzeptes gibt [SC15].

4 Lösungsansätze

In diesem Abschnitt wird diskutiert, wie Information Security Risk Management für einen Einsatz in hochkomplexen Systemen aussehen kann. Welche Ansätze aus dem klassischen Risikomanagement können übernommen und weiterentwickelt werden, wo braucht es neue Lösungen und wie können diese aussehen?

Szenarienbasierte Impact Analyse:

Im Business Continuity Management (BCM) ist es bereits heute üblich, schon in einem frühen Stadium des BCM Prozesses eine Auswirkungsanalyse durchzuführen: in der BIA (Business Impact Analyse) wird zunächst festgestellt, welche Geschäftsprozesse die höchste Kritikalität haben, d.h. der Ausfall welcher Geschäftsprozesse den höchsten Schaden verursachen würde. Erst nachfolgend ist für diese Geschäftsprozesse eine "klassische" Risikoanalyse, allerdings konzentriert auf den Bereich Verfügbarkeit, durchzuführen. Einen ähnlichen Zugang finden wir in der neuen EU Datenschutz-Grundverordnung. Diese schreibt in Artikel 35(1) eine Datenschutz-Folgenabschätzung (Data Protection Impact Analysis, DPIA) vor, wenn "eine Form der Verarbeitung, insbesondere bei Verwendung neuer Technologien, aufgrund der Art, des Umfangs, der Umstände und der Zwecke der Verarbeitung voraussichtlich ein hohes Risiko für die Rechte und Freiheiten natürlicher Personen zur Folge (hat)". BIA und DPIA stellen jeweils auf einen spezifischen Aspekt der Informationssicherheit bzw. auf ein definiertes Schutzziel ab: die BIA auf Verfügbarkeit, die DPIA auf den Schutz personenbezogener Daten (Privacy).

Eine Impact Analyse im Bereich der Informationssicherheitsrisiken muss alle für das betrachtete System relevanten Schutzziele, also etwa Vertraulichkeit, Integrität, Verfügbarkeit, Zurechenbarkeit (Accountability), Verbindlichkeit (Non-Repudiation), oder Privacy, berücksichtigen. Als erster Schritt empfiehlt sich daher eine Feststellung der relevanten Schutzziele.

Zu diskutieren bleibt die Frage, auf welcher Ebene eine solche Impact Analyse ansetzen sollte. Die BIA im BCM setzt auf Geschäftsprozessebene an (s. etwa [BSI08c]), die DPIA bei der "Verarbeitung", also auf Ebene von Services. Auch im klassischen ISRA wird meist empfohlen, von den Geschäftsprozessen auszugehen, in der Praxis starten Risikoanalysen dennoch oft auch auf Ebene von Komponenten, wie beispielsweise Server oder Netzwerkkomponenten. Für einen durchgängigen Top-Down-Approach ist es sicherlich unumgänglich, auf einer hohen Ebene, also idealerweise auf Ebene der Geschäftsprozesse, und nicht auf Komponentenebene, anzusetzen.

Um der Komplexität der Fragestellungen gerecht zu werden, empfiehlt es sich, auf einen szenarienbasierten Ansatz zurückzugreifen. Dabei werden verschiedene Szenarien durchgedacht und ihre Auswirkungen auf die Schutzziele abgeschätzt. Bedrohungen, Eintrittswahrscheinlichkeiten und bestehende Sicherheitsmechanismen spielen in dieser Phase noch keine oder allenfalls eine untergeordnete Rolle. Die Szenarien müssen

jedoch grundsätzlich realistisch sein. Es ist zu ermitteln, welche Auswirkungen die betrachteten Sicherheitsvorfälle auf das Gesamtunternehmen und den Unternehmenserfolg haben. Die Szenarien mit dem größten Impact können dann näher untersucht werden, etwa in Form einer Top-Down-Analyse, vergleichbar mit einer Fehlerbaumanalyse (Fault Tree Analysis, FTA), wie sie in technischen Bereichen angewendet wird [DIN90]. In diesem Schritt sind dann auch detaillierte Risikoanalysen auf Basis von Komponenten durchzuführen, um konkrete Bedrohungen und Schwachstellen zu erkennen und auf deren Basis entsprechende Schutzmaßnahmen auszuwählen.

Toolunterstützung:

Eine der großen Herausforderungen an Risikoanalysen in IT-Systemen liegt in der raschen Veränderung der Parameter. Nicht nur die Bedrohungslage kann sich sehr rasch ändern, auch das System selbst ist meist vielfältigen und raschen Änderungen unterworfen: neue HW- und Softwarekomponenten, eine Änderung in baulichen Gegebenheiten oder in der Netzwerkstruktur sowie die Verarbeitung neuer Datenklassen können die Risikolage rasch verändern.

Die Erstmodellierung des Systems erfordert im Allgemeinen sehr viel Aufwand. Um zumindest rasch auf Änderungen reagieren zu können, ist es wichtig, die Modellierung toolunterstützt durchzuführen und Änderungen zeitnah einzupflegen. Eine weitgehende Automatisierung der Einbeziehung neuer Komponenten ist anzustreben.

SIEM-unterstützt:

Ein Security Information and Event Management System (SIEM) ermöglicht eine echtzeitnahe Analyse von Sicherheitswarnungen, die von Netzwerkkomponenten oder Applikationen generiert werden. SIEM-Lösungen bauen im Allgemeinen auf Log-Systemen auf und ermöglichen eine frühzeitige Erkennung von Angriffen und schnellere und bessere Reaktion auf Sicherheitsvorfälle (Security Incidents). Zunehmend werden SIEM-Lösungen heute auch zur Erkennung und Abwehr hochentwickelter zielgerichteter Angriffe (Advanced Persistent Threats, APTs) eingesetzt, da sie eine schnelle Reaktion auf die tatsächlichen Gegebenheiten in einem System ermöglichen.

Eine Einbindung von SIEM-Lösungen in das Risikomanagement ermöglicht zu einem gewissen Grad eine automatische Aktualisierung der Einschätzung der Risikolage und der Anpassung oder Einleitung von entsprechenden Sicherheitsmaßnahmen.

Eine automatisierte Einbeziehung von Messdaten und Sensoren in das Gesamtsystem erhöht die Aktualität der Risikobewertung weiter und unterstützt den Aufbau eines effizienten Risikomanagements. Ein gutes Beispiel für ein Risikoanalyse-System auf Basis von IT-Grundschutz und unter Einbeziehung von automatisierten Messdaten gibt etwa [Sc14].

5 Risikomanagementstrategien für komplexe Systeme

Wie oben ausgeführt ist eine Risikoanalyse wichtige Voraussetzung um Sicherheit in einem System zu gewährleisten. Für komplexe Systeme gilt dies in besonderem Maße. Dennoch sind gerade in komplexen Systemen auch andere Vorkehrungen, wie die Integration von Sicherheit bereits in der Entwicklung der Systeme, Absicherung durch Basisschutzmaßnahmen und die Entwicklung resilienter Systeme, unabdingbar. Einige davon werden in der Folge diskutiert.

Security by Design und Privacy-by-Design:

Im IT-Bereich werden Sicherheitsfunktionen heute oft erst im Nachhinein auf ein System aufgesetzt, was zu Fehlern führt und sehr kostenintensiv sein kann. Wünschenswert wäre eine integrierte Basissicherheit sowohl für die einzelnen Komponenten (s.u.) als auch für das Gesamtsystem ("security-by-design", "privacy-by-design"). Die Beachtung von Sicherheitsanforderungen bereits beim Systemdesign und in der Systementwicklung sowie die möglichst weitgehende Verwendung standardisierter Schnittstellen und Sicherheitsfunktionen können dazu beitragen, die Sicherheit der Systeme von vornherein entsprechend hoch zu gewährleisten.

Integrierte Sicherheit in allen Systemen und Komponenten:

In vielen Produkten des täglichen Lebens ist Sicherheit bereits ein integrierter Bestandteil. Elektroartikel, Kinderspielzeug und Autos unterliegen strengen Vorgaben, und ihre Qualität und Zuverlässigkeit wird streng geprüft. Produktrückrufe zeigen, dass zwar auch in diesen Bereichen keine 100%-ige Perfektion erreicht werden kann, dass jedoch ein gutes Prüf- und Kontrollsystem existiert und eine laufende Verbesserung vorgesehen ist. Angemessene Authentisierungsmechanismen, die Möglichkeit zu Updates und der Schutz von Vertraulichkeit und Integrität der Kommunikation sollten zum Standard-Repertoire aller Komponenten gehören. Mit Hilfe neuer Technologien, wie etwa PUFs, ist es heute möglich, auch in low-cost Komponenten, die für den Massenmarkt produziert werden, hohe Sicherheit zu gewährleisten. Dazu sind neben entsprechenden kryptographischen Verfahren (z.B. light-weight Kryptomechanismen) insbesondere auch spezifisch auf die verwendeten PUF-Technologien abgestimmte Fehlerkorrekturmechanismen erforderlich. [De15] gibt eine detaillierte Analyse und ein Design solcher Algorithmen anhand eines Prototyps zum Software-Schutz.

Basisschutz als Grundlage für das Risikomanagement:

In der Vergangenheit sehen wir unterschiedliche Zugänge zu Risikoanalyse und darauf aufbauender Risikobehandlung. Während in der 1970-er und 1980-er Jahren die "klassische" detaillierte Risikoanalyse für ein gesamtes System das Mittel der Wahl war, führte die zunehmende Komplexität der zu betrachtenden Systeme spätestens in den 1990-er Jahren zu einem Paradigmenwechsel. Mittels Basissicherheitsmaßnahmen sollten die Systeme gegen sogenannte pauschalisierte Gefährdungen geschützt werden, sehr rasch setzte sich ein kombinierter Ansatz durch, der eine Absicherung der

Systemteile und Komponenten mit normalem Schutzbedarf über generelle Sicherheitsmaßnahmen ("Grundschutzmaßnahmen") vorsieht, und für Bereiche mit sehr hohen Sicherheitsanforderungen eine detaillierte Risikoanalyse. Bekanntestes Beispiel dafür im deutschen Sprachraum ist der IT-Grundschutz des deutschen Bundesamtes für Informationssicherheit (BSI) ([BSI08a], [BSI08b]), auch das Österreichische Informationssicherheitshandbuch [BKA16] folgt dieser Vorgehensweise.

Während im kombinierten Risikoanalyseansatz meist bereits in einem frühen Stadium eine Trennung von normal- und hoch-schutzbedürftigen Systemteilen erfolgt, gehen neuere Ansätze davon aus, dass für alle Komponenten eine Basissicherung vorhanden sein sollte, auf der aufbauend weitere Standardschutzmaßnahmen sowie detaillierte Risikoanalysen für die besonders gefährdeten Szenarien / Komponenten erfolgen sollen. Der im Zuge der Modernisierung des IT-Grundschutzes des BSI gewählte Ansatz folgt dieser Vorgehensweise [BSI16]. Für komplexe Systeme ist es in besonderem Maße notwendig, dass die einzelnen Komponenten entsprechende Basisschutzmechanismen, wie etwa Authentisierungsmechanismen, haben.

Resilience:

Bereits beim Design komplexer Systeme sollte darauf geachtet werden, dass diese möglichst widerstandsfähig gegen Angriffe, aber auch gegen Fehlfunktion oder Ausfall einzelner Komponenten sind. Resiliente Systeme sind dadurch gekennzeichnet, dass sie auch im Angriffs-, Fehler- oder Katastrophenfall nicht vollständig ausfallen oder Informationen in großem Ausmaß preisgeben, und möglichst schnell und stabil wieder in ihren Ausgangszustand kommen. Sie ermöglichen per se Ausweichlösungen, wenn es zu Problemen kommt. Wichtige Mechanismen für Resilience im IT-Bereich sind etwa verteilte Verarbeitung, redundante Kommunikationsverbindungen oder Backup- und Recovery-Mechanismen.

6 Zusammenfassung und Ausblick

Risikoanalyse ist eine entscheidende, unverzichtbare Voraussetzung für die Abschätzung des Gesamtrisikos eines Systems und der Auswirkungen auf eine Organisation oder die Gesellschaft, sowie für die Auswahl von Sicherheitsmaßnahmen. Gerade im Bereich IT- und Informationssicherheit können Risikoanalysen allerdings auch für kleinere Systeme sehr aufwendig werden.

In komplexen Systemen ergibt sich eine Reihe von neuen Anforderungen an die Informationssicherheits-Risikoanalyse, die in Kapitel 3 diskutiert wurden. Es braucht neue Ansätze zur Risikoanalyse, insbesondere zur Risikoaggregation und zur Automatisierung der Risikoabschätzung (s. Kapitel 4), aber auch ein umfassendes Risikomanagement, wie in Kapitel 5 diskutiert, um die Sicherheit der Informationen und der sie verarbeitenden IT-Systeme auf Dauer zu gewährleisten.

Literaturverzeichnis

- [BKA16] Österreichisches Informationssicherheitshandbuch
- [BSI08a] Bundesamt für Sicherheit in der Informationstechnik (BSI): IT-Grundschutz-Vorgehensweise, BSI-Standard 100-2, Bonn, 2008
- [BSI08b] Bundesamt für Sicherheit in der Informationstechnik (BSI): Risikoanalyse auf der Basis von IT-Grundschutz, BSI-Standard 100-3, Bonn, 2008
- [BSI08c] Bundesamt für Sicherheit in der Informationstechnik (BSI): Notfallmanagement, BSI-Standard 100-4, Bonn, 2008
- [BSI16] Bundesamt für Sicherheit in der Informationstechnik (BSI): Modernisierung IT-Grundschutz, www.bsi.bund.de/IT-Grundschutz, Stand Juni 2016
- [De15] Deutschmann, M.: Mathematical Investigations on the Stability of PUF Responses Considering Different Error Correcting Mechanisms, Dissertation, Alpen Adria Universität Klagenfurt, 2015
- [DIN90] Deutsches Institut für Normung e.V. (DIN): DIN 25424 Fehlerbaumanalyse
- [EN15] European Union Agency for Network and Information Security (ENISA): ENISA Threat Landscape 2015, www.enisa.europa.eu/publications/etl2015, Stand Juni 2016
- [ISO09] International Organization for Standardization (ISO): ISO 31000 Risk management – Principles and guidelines, 2009
- [ISO11] International Organization for Standardization (ISO), International Electrotechnical Commission (IEC): ISO/IEC 27005 Information technology — Security techniques — Information security risk management, 2011
- [ISO13] International Organization for Standardization (ISO), International Electrotechnical Commission (IEC): ISO/IEC 27001 Information technology — Security techniques — Information security management systems — Requirements, 2013
- [Ko08] Kotsios, C. G.: ROISI – Investitionsentscheidungen im IKT-Sicherheitsbereich, VDM Verlag Saarbrücken, 2008
- [KSB14] Kolberger, A.; Schaumüller-Bichl, I.; Deutschmann, M.: Risk Analysis of Physically Unclonable Functions, Proceedings CMS2014, LNCS 8735, 2014
- [Sc14] Schiebeck, S.: An Approach to Continuous Information Security Risk Assessment focused on Security Measurements", Dissertation, Universität Wien, 2014
- [Sc15] Schiebeck, S.; Latzenhofer, M.; Palensky, B.; Schauer, S.; Quirchmayr, G.; Benesch, T.; Göllner, J.; Meurers, C.; Mayr, I.: Implementation of a Generic ICT Risk Model using Graph Databases, Proc. SECUWARE 2015, pp. 146-153, Venedig, 2015
- [St08] Stallinger, M.: IT-Governance im Kontext Risikomanagement, VDM Verlag Saarbrücken, 2008

Effiziente Risikoanalyse anhand praktischer Erfahrungsbeispiele

Manfred Holzbach¹

Abstract: Risikomanagement in der Praxis kann heute nicht mehr vollflächig erfolgen, sondern muss sich auf die lebenswichtigen Assets einer Organisation fokussieren und deren Bestand auch in Krisenfällen sichern können. Das gilt ganz besonders für KMUs, die immerhin das Rückgrat der Wirtschaft und wohl auch der Gesellschaft in Europa bilden. Anhand von 3 Beispielen unterschiedlicher Zielsetzung, Gefahrenstruktur und Schlussfolgerungen daraus wird gezeigt, wie mit Hilfe geeigneter Prioritätensetzung effiziente Risikoanalysen durchgeführt wurden resp. werden.

Keywords: Risikoanalyse, Risikomanagement, Cyber-Sicherheits-Check, sensible Daten, Attacken

1 Einleitung

Risikomanagement gemäß ISO/IEC 31000:2009 [ISO09] zur Identifikation, Analyse und Bewertung von Risiken für ein Unternehmen oder eine Organisation ist heute unbestrittener Weise unabdingbar. Nicht nur für große Einheiten, sondern auch für KMUs, Start-Ups oder ausgelagerte öffentliche Einrichtungen. Abhängig von ihrer Tätigkeit können diese im Zuge der Aufgabenteilung und Vernetzung volkswirtschaftlich kritische Points-of-Failure darstellen. So können Zulieferer für den Wirtschaftsstandort wertvolles Know-how besitzen oder Subauftragsnehmer sensible oder begehrte Informationen verarbeiten, welche für ihre Auftraggeber oder Dritte wesentliche Bedeutung haben. Somit können sie sich selbst, aber auch andere Einheiten in existenzbedrohende Lagen bringen – besonders in rechtlicher und finanzieller Hinsicht aber auch was Vertrauen und Reputation betrifft.

Risikomanagement kann überall angewendet werden, wo Risiken auftreten können. Dabei besteht die Zielsetzung im rechtzeitigen Erkennen von und dem Umgang mit solchen Risiken, welche die Existenz der Organisation gefährden, Leib und Leben anvertrauter Menschen bedrohen oder wo große Schadenspotenziale schlagend werden können. Es reicht dabei nicht, die identifizierten Einzelrisiken abzuarbeiten, sondern auch die Risikopotenziale aus deren Kombinationen zu erkennen und zu beherrschen.

Der Anforderung nach maximaler Breite und Tiefe an Analyse und Beherrschung stehen die endlichen verfügbaren Ressourcen, Zeit und Know-how, aber auch die Mobilität und

¹ Zentrum für sichere Informationstechnologie – Austria (A-SIT), Seidlgasse 22/9, 1030 Wien, manfred.holzbach@a-sit.at.

Kreativität von Angreifern entgegen. Es besteht auch immer die Gefahr, dass bei rein methodischem Ansatz etwas übersehen wird.

Mit der fortschreitenden Digitalisierung ist es aufgrund der immer stark ansteigenden Durchdringung, Vernetzung und Komplexität der sie ausmachenden IT-Systeme, Komponenten, Netzwerke und Applikationen nicht mehr möglich, alle potenziellen Gefahren und damit Risiken flächendeckend einzuschätzen. Nicht in der gegebenen Vielfalt, denn es gibt immer weniger industriell gefertigte Produkte, welche keine Software enthalten. So bringt es ein Herzschrittmacher auf 80.000 Codezeilen [JM13], das Elektrofahrzeug Chevrolet Volt auf 10 Millionen Codezeilen [LD10] und zum Vergleich die Flugsoftware der Boeing 787 auf 14 Millionen [DL11]). Und auch nicht im zeitlichen Verlauf, denn in jeder Sekunde werden täglich 390.000 neue Schadprogramme (also 4-5 pro Sekunde) registriert [SZ16].

Da sich dieser Beitrag an Risikomanagement-Experten wendet, wird darin Kenntnis der Risikomanagement- und Risikoanalysemethoden weitgehend vorausgesetzt.

2 Aufgabenstellung

Es handelt sich dabei um eine sich aus den praktischen Erfahrungen manifestierte Entwicklung.

Im Prinzip wird das gleiche Verfahren angewendet, unbeschadet davon ob es sich um ein großes Unternehmen oder ein Start-Up handelt. Der wesentliche Unterschied liegt in den verfügbaren Ressourcen, die für Detailtiefe, Einsatz von Experten, Tools, Ausführen von Pen-Tests zur Verfügung stehen.

Risikoanalysen, welche einen Bottom-Up Absatz verfolgen oder den flächendeckenden Anspruch stellen, alles abdecken zu können, werden etwa im Bereich der kritischen Infrastrukturen durchgeführt und stellen sich als Großprojekte über Monate oder Jahre mit Experten der unterschiedlichsten Disziplinen dar (Beispiel über eine umfassende Bewertung von Stromausfällen in [EC14]). In Bereichen wie exponierter Daten oder Software können sie nur bedingte Aussagen leisten bzw. sind sie mit vertretbarem Ressourcenaufwand nicht mehr umsetzbar, schon gar nicht für KMUs. Diese verarbeiten aber oft gesamt- oder volkswirtschaftlich äußerst wertvolle Assets, beispielsweise als Zulieferer oder Subauftragsnehmer.

Risikoanalyse ist heute noch viel wichtiger als je zuvor. Da sie jenseits kritischer Infrastrukturen nicht alles abdecken kann, ist es essentiell eine Methode zu haben die auf das jeweilige Problem, Budget und die Organisation angepasst ist.

Die in der Praxis zu lösende Aufgabe besteht darin, ein gezieltes, effizientes Vorgehen so zu gestalten, dass die Zielsetzung des Risikomanagements auch bei reduzierter Breite und Tiefe zuverlässig erreicht wird. Dies wird durch ein Schalenmodell, beginnend bei den Top-Assets (den „Kronjuwelen“) erreicht; dies schon aufgrund der Tatsache, dass

am Schluss immer ein Restrisiko in Kauf zu nehmen ist. Im Kern geht es darum, dort zu investieren, wo die Prioritäten sind und sich nicht in Details zu verlieren, welche keine lebenswichtigen Funktionen betreffen.

Das Schalenmodell geht vom Verstehen der Geschäftsprozesse aus und modelliert dann Assets nach Priorität aus deren Kritikalität:

- Welche Geschäftsprozesse sind für die Existenz des Unternehmens oder der Organisation unabdingbar?
- Was ist jeweils der maximal mögliche Schaden für die eigene aber auch davon mit betroffene Organisationen pro identifiziertem Bereich?

Die Assets sind exponiert gegen Gefährdungsbereiche, für welche Gefahrenkataloge und Kontrollziele aufgestellt werden – zunächst in groben Kategorien z.B. Gefahren im Zusammenhang mit Verfügbarkeit, Gefahren im Sinne des IT-Systems, dann je nach konkreter Situation und Bedarf weiter ins Detail gehend. Durch ein solches Top-Down Modell findet man die hohen Prioritäten vergleichsweise schnell.

Tiefer in Details wird dort gegangen, wo es sich dann als notwendig erweist oder wenn Problemzonen erkannt werden. Dabei wird zuerst hinterfragt, ob der betroffene Prozess oder sämtliche exponierten Informationen notwendig sind resp. ob und welche Alternativen es gibt. Somit werden sukzessive Schalen entfernt und zum Kern des Problems vorgedrungen. Jeder Ast des Risikobaumes, den man aufgrund geringer Priorität oder auch bereits festgestellter ausreichender Sicherheit abschneiden kann, bzw. jede eliminierte Schale spart Ressourcen in Form von Zeit und Geld.

Erst dann wird weiter nach unten in Richtung als kritisch ausgemachter Prozesse (insbesondere IT-Prozesse) weiter vorgegangen. Details werden erst nach deren Analyse gezielt untersucht (z.B. mittels Penetrationstests).

Zu betrachten sind unbedingt auch der Einfluss des Change Managements sowie von Notfallmaßnahmen und nicht standardisierbarer, aber im Krisenfall oft notwendiger Improvisation auf das Risiko.

Dies ist unabhängig davon, ob ein genormter Ansatz wie etwa ein Vorgehen gemäß BSI Standard 100-3 „Risikoanalyse auf Basis des IT Grundschutz“ [BSI08] gewählt wird oder nicht. Ein solcher erfüllt eine notwendige aber nicht hinreichende Bedingung für eine realistische Risikoeinschätzung. Denn es kann etwas übersehen werden oder eine aktuelle Schwachstelle entzieht sich den standardisierten Verfahren („Hintertüren“).

Im Vergleich mit dem heutigen Straßenverkehr könnte man sagen: „Genormter Ansatz bietet die STVO und einen Airbag, das notwendige Fahrkönnen muss als individuelle menschliche Fertigkeit eingebracht werden“.

„Kreativität ist gerade dort notwendig wo es gefährlich wird“.

„Analog zur Kriegsführung erweisen sich menschliche Qualitäten als unabdingbar“.

3 Praxisbeispiele

Vor diesem Hintergrund wird anhand von 3 praktischen Beispielen dargestellt, wie Risikoanalysen im Bereich von KMUs mit jeweils situativ unterschiedlichem Ziel, Zweck und unterschiedlicher Motivation und Methode, aber dennoch effizient und aussagekräftig durchgeführt wurden resp. werden. Die Beispiele handeln von Risikoanalysen, die A-SIT bei Organisationen tatsächlich umgesetzt hat resp. zur Zeit umsetzt. Die jeweiligen Unternehmen werden hier nicht genannt und Findings nicht veröffentlicht, wohl aber die unterschiedlichen Vorgehensweisen dargestellt, begründet und Erkenntnisse, welche die Methode betreffen, diskutiert.

3.1 Cyber-Sicherheits-Check bei einem öffentlichen Förderungsgeber

Die Servicestelle des Landes Steiermark für Unternehmen vergibt unter anderem Förderungsmittel des Landes und der Europäischen Union an Unternehmen, welche durch Innovationsmaßnahmen zur Stärkung der eigenen Wirtschaftskraft und des Wirtschaftsstandortes beitragen. Die Förderungsmittel werden beantragt, wobei die Innovation durch geeignete Detail-Unterlagen nachzuweisen ist, welche einen erheblichen wenn nicht unersetzlichen Wert und damit die Top-Assets darstellen. Die Schadenspotenziale reichen von erheblichen Ersatzforderungen über Compliance-Verletzungen bis zu massiver Image-Beschädigung nicht nur für die Servicestelle, sondern auch das Land.

Die Risikoanalyse erfolgte mittels Cyber-Sicherheits-Check, einer vom BSI² im Rahmen der Allianz für Cyber-Sicherheit vorgeschlagenen Vorgehensweise [BSI14], wobei die Risikoanalyse gemäß der „Cyber-Sicherheits-Exposition“ [BSI12] abgearbeitet wird. Als Grundlage für die Bewertung der implementierten oder geplanten Maßnahmen auf Notwendigkeit, Angemessenheit und Wirtschaftlichkeit werden die einzelnen Elemente der Infrastruktur, gespeicherten und übertragenen Daten sowie Verarbeitungsprozesse einer ganzheitlichen Cyber-Bedrohungsanalyse unterzogen und die Erkenntnisse kompakt und plakativ dargestellt. Damit erreicht man zügige und interaktive Kommunikation zwecks gemeinsamer Sichtweisen mit dem Management. Bemerkenswert dabei ist, dass man sich auf „Cyber“-Bedrohungen beschränkt, also Perimeter- und Safety-Aspekte nicht unmittelbar berücksichtigt.

Zunächst erfolgten interaktiv Bewertungen zur Cyber-Sicherheits-Exposition auf Basis standardisierter Fragestellungen zu Wert der Daten und Prozesse, Attraktivität für Angreifer, Art der Angreifer, zu erwartender Zielgerichtetheit von Angriffen, erfolgten Angriffen in der Vergangenheit, zu erwartendem maximalem Bedrohungsgrad und Transparenz für Angreifer – jeweils unterschieden zwischen Vertraulichkeit, Verfügbarkeit und Integrität. Die Einzelergebnisse werden nach einem vorgegebenen Gewichtungsschlüssel bewertet und eine Gesamt-Exposition pro Bereich ermittelt. Die

² Bundesamt für Sicherheit in der Informationstechnik, Bonn; Website www.bsi.bund.de.

„Schalen“ werden hier durch die Auswahl der Einzelfragen und Gewichtung der Antworten gebildet.

Aufgrund der wertvollen zu schützenden Daten ergeben sich entsprechend hohe Expositionen bei „Vertraulichkeit“, wobei zielgerichtete Angriffe denkbar aber bei wirksamen Maßnahmen unwahrscheinlicher sind als beim Unternehmen selbst, dem die Informationen gehören. Für die Exposition bei „Verfügbarkeit“ ist unter anderem die Möglichkeit von DoS Attacken zu bewerten.

Im zweiten Teil erfolgte die Bewertung des „Cyber-Sicherheits-Status“, d.h. der implementierten bzw. geplanten Maßnahmen. Dabei wird einem vorgegebenen Katalog von Maßnahenzielen gefolgt, allfällige Mängel aufgezeigt und die Wirksamkeit bewertet – sowohl tabellarisch mit Ampel-Indikatoren als auch verbal ausgeführt. In dieser Bewertung liegt die Priorisierung ausgehend von der Exposition und damit das Schalenmodell.

Parallel wurden auf Wunsch der Organisation zusätzliche Schwachstellen Scans mit einem Tool bei einem Outsourcing-Dienstleister durchgeführt. Das ist zwar kein zwingender Bestandteil des Cyber-Sicherheits-Checks, die Erkenntnisse sind allerdings in die Gesamtbewertung eingeflossen und erwiesen sich als durchaus wertvoll.

Der Cyber-Sicherheits-Check führt mit vergleichsweise kompaktem Aufwand zu einer sehr brauchbaren Bewertung der Risiken, allfälliger Schwachstellen und der Wirksamkeit von Maßnahmen. Das bewirkt Awareness und Mitwirkung des Managements, was gerade bei KMUs essentiell ist. Beim Scoring für die Exposition können sich bereits geringe Veränderungen in Einzelbereichen erheblich im Gesamtergebnis auswirken. Daher sollte die Methode von erfahrenen Auditoren durchgeführt werden.

3.2 Standardmethode nach Schalenmodell in einem Hochrisikobereich

Die primären Assets des Systems (die „Kronjuwelen“) sind Gesundheitsdaten, welche von unterschiedlichen Teilnehmern genutzt und verarbeitet werden, aber den Patienten gehören. Damit hat man mit der sensibelsten Art von Daten überhaupt, und für Angreifer äußerst wertvollen Informationen zu tun. Für Schutzmaßnahmen im Security-Bereich relevante Assets sind in erster Linie die Zugangsdaten, welche sehr heterogen strukturiert sind, was die Möglichkeiten und den Aktionsradius, somit das Gefahrenpotenzial, betrifft: Vom einzelnen Bürger über die Gesundheitsdienste bis zu zentralen Stellen. Unbeschadet genereller Leitlinien sind die von IT-Abteilungen betriebenen Systeme dezentral, unterschiedlich komplex und müssen Entscheidungsspielräume für die Akteure ermöglichen.

Das zentrale Risiko wäre ein Verlust der Vertraulichkeit, der Integrität aber auch der Verfügbarkeit von Gesundheitsdaten. Die Folgen könnten existenzbedrohende Ausmaße annehmen, sowohl in materieller Hinsicht (rechtliche Konsequenzen, Strafen,

Schadenersatz) aber auch was den Verlust an Reputation und Vertrauen betrifft. Eine besondere Problematik ergibt sich dabei aus der Struktur: Selbst wenn ein Vorfall an nur einer Stelle auftreten würde und nur ein begrenztes Ausmaß hätte, müsste mit heftigen Reaktionen der Öffentlichkeit, von Medien und der Politik gerechnet werden, welche ihren Fokus dann auf das Gesamtsystem richten und es in Frage stellen würden.

Die Risikoanalysemethode folgt einem Top-Down Ansatz: Zunächst werden Funktionen und Komponenten mit hoher Priorität identifiziert, welche sich in exponierten Bereichen befinden und Einfluss auf die Top-Assets haben. Die Fragestellung umfasst ihre jeweilige Risikoeinschätzung für sich, aber auch in welchen Bandbreiten sich ihre Restrisiken aufgrund ihrer Umgebung, verteilten Einsatzes, Eigenschaften und Zusammenwirken beeinflussen lassen. Niedrige Priorität haben dann Komponenten, bei denen man bereits ausreichende Sicherheit annehmen kann, etwa weil sie in geschützten Bereichen arbeiten resp. bereits begutachtet sind.

In der Folge werden Teilrisikoanalysen mit Hilfe eines Tools in mehreren Ebenen abgearbeitet, von den materiellen Assets über die Geschäftsprozesse, deren Relevanz auf das Gesamtsystem, die Komponenten und die Gefahren. Dies stellt den klassischen Teil der Risikoanalyse dar.

Die Ergebnisse werden für fiktive Angriffsszenarien herangezogen um die Robustheit als Grad, solche verhindern zu können, sowie die möglichen Folgen festzustellen, aber auch die Möglichkeiten sie zu erkennen und darauf effizient zu reagieren. Die gewonnenen Erkenntnisse fließen dann in Empfehlungen für die Implementierung präventiver, detektiver und reaktiver Maßnahmen ein.

Das Besondere an diesem Ansatz ist es (abgesehen vom Zusammenwirken mehrerer mit Security befasster Organisationen), einem heterogenen und aufgrund unterschiedlicher Go-Live Termine für geografische und organisatorische Einheiten dynamischen Gesamtsystem Rechnung zu tragen. Es gibt einerseits Standardverfahren, aber auch sehr spezifische resp. einzigartige Funktionen. Das erfordert menschliche Kreativität und Entscheidungsqualität und kann insgesamt nicht allein durch Befolgen von Standards hinreichend abgedeckt werden.

3.3 Risikoanalyse im Rahmen eines Qualitätsmanagements

Es handelt sich um eine sehr kleine Organisation, welche sich unter anderem mit Technologiebeobachtung befasst, Gutachten erstellt und sich dafür akkreditieren lässt. Der Zweck der Risikoanalyse ergibt sich daher zunächst aus der Notwendigkeit im Rahmen des Qualitätsmanagements und für die Pflichtversicherung. Da es sich um real vorhandene Risiken handelt, ist deren Einschätzung und Vermeidung aber auch für die Geschäftsgebarung unerlässlich.

Zu schützende Assets sind vertrauliche Unternehmens- und Produktdaten der zu prüfenden Organisationen, sowie Unterlagen und Erkenntnisse der

Technologiebeobachtung. Diese können immerhin Einfluss auf die Reputation von Systemen und deren Betreibern haben.

Das materielle Risiko liegt wie bei anderen Organisationen in deren Kompromittierung, Verlust oder Diebstahl. Die Folgen wären rechtliche Konsequenzen und Schadenersatzforderungen, welche auch hier existenzbedrohende Ausmaße annehmen könnten. Wegen der Rechtsform eines Vereins einer Teilorganisation birgt dies auch erhebliche Risiken für dessen Vorstände. Abdeckung durch Versicherungen ist nur begrenzt möglich, schon wegen der teuren Prämien in diesem Bereich.

Das Reputationsrisiko kann schon bei geringfügigen Vorfällen wie einem Einbruch mit Diebstahl der Handkassa existenzbedrohend werden, wenn damit das Vertrauen in die sichere Verwahrung von Dokumenten und das Image als Experte für Sicherheit damit in Frage gestellt würde.

Gefahren können weiters für die Integrität und Qualität der gemachten Aussagen auftreten, sowohl in inhaltlicher wie technisch-organisatorischer Hinsicht.

Bei der Risikoanalyse und dem daraus abgeleiteten Risikomanagement erfolgt die Überprüfung und schließlich der Nachweis, inwieweit die getroffenen Vorkehrungen und Maßnahmen wirksam sind. Ausgehend von den oben genannten Assets werden die maximal denkbaren Schadensszenarien abgeschätzt und welche Vorfälle sie verursachen könnten. In der nächsten Ebene (Schale) werden die kritischen Geschäftsprozesse zugeordnet und analysiert, welchen Einfluss handelnde Personen und örtliche Gegebenheiten darauf haben können. Dann wird die Stärke der Maßnahmen wie Perimeterschutz, Rollen- und Rechteverwaltung, verschlüsselte und gespiegelte Speicherung sowie für nachvollziehbare Richtigkeit der Aussagen die implementierten Review-, Freigabeprozesse und qualifizierte Signaturen bewertet und die Schlussfolgerungen gezogen.

3.4 Vergleich der Beispiele

Beispiel 3.1 hatte die Aufgabenstellung, in einem sehr kompakten Verfahren die bestehenden Sicherheitsmaßnahmen hinsichtlich allfälliger Schwachstellen zu bewerten und daraus Risiken und die Gesamtexposition zu erkennen sowie Empfehlungen zu generieren. Der gewählte Cyber-Sicherheits-Check kann dies fokussiert auf Security-Aspekte abdecken und ermöglicht auf Grund seiner Kompaktheit eine besonders aktive Mitwirkung des Managements. Erforderlich ist dazu allerdings viel Erfahrung und Übersicht bei den Auditoren und beim IT-Personal der untersuchten Organisation.

Beispiel 3.2 bearbeitet eine komplexe, heterogene und dynamische Systemlandschaft, in der es nicht möglich ist, sämtliche Einzelrisiken und ihr Zusammenwirken vollständig zu bewerten. Obwohl die Risikoanalyse sehr breit angelegt ist und mehrere Berater sowie Tools eingesetzt wurden, muss der Fokus dennoch auf den lebenswichtigen Assets und denjenigen Mechanismen liegen, die deren Funktion auch in Krisenfällen sichern. Der

Aufwand ist dementsprechend erheblich und den befassten Experten wird umfassendes Wissen, Erfahrung und Kreativität abverlangt.

Beispiel 3.3 wird für ein sehr kleines Unternehmen mit eng abgegrenzten, aber für die Organisation potenziell sehr gefährlichen Risiken verwendet. Der Vorgang ist individuell zugeschnitten, gestaltet sich allerdings auf Grund der Situation, dass die MitarbeiterInnen im Detail mit den Geschäfts- und IT-Prozessen sehr vertraut sind, als überschaubar und geradlinig. Es erfordert hohe Kompetenz der befassten Experten, aus dem begrenzten Input die Erkenntnisse nachvollziehbar zu dokumentieren und zu formalisieren, sodass sie für eine Akkreditierung brauchbar und hinreichend sind.

4 Weiteres Vorgehen

Das Österreichische Informationssicherheitshandbuch [BKA14] wird von A-SIT inhaltlich betreut und enthält ein umfassendes Kapitel über Risikoanalysestrategien und ihre Umsetzung. Dies wird nun anhand der gewonnenen Erkenntnisse und auch praktischen Erfahrungen auf den neuesten Stand gebracht. Ausgehend vom derzeit dargestellten „kombinierten Ansatz“ zur Risikoanalyse sollen das Konzept der Top-Down Vorgehensweise und das Schalenmodell konkret beschrieben und etablierte Methoden erläutert werden.

Herausfordernd dabei wird es sein, einen Leitfaden zu entwickeln, welcher einerseits generisch auf möglichst viele Fälle anwendbar ist, so umfassend als möglich ist und dennoch den Kern des Konzepts wiedergibt, wo es darum geht im konkreten Einzelfall die richtigen Prioritäten aber auch Zusammenhänge zu erkennen.

Zum einen wird man die allgemein gültigen Strukturelemente beginnend bei den lebenswichtigen Assets sowie häufige Einflussfaktoren aus der Umgebung und dem Zusammenwirken verschiedener Prozesse darstellen. Ähnlich wie im BSI Standard 100-3 [BSI08] macht es dabei auch Sinn, das Konzept und die Elemente anhand praxisnaher Beispiele zu erläutern.

5 Offene rechtliche Fragen aus der gewonnenen Erfahrung

- Da vollständige Sicherheit nicht gewährleistet werden kann, braucht es Richtwerte, was einer Organisation im Rahmen der Gesetze und Compliance an Aufwand und Zuwendung zumutbar ist. Wie weit reicht dann ihre rechtliche Verantwortung (Haftung) bzw. die ihrer Organe?
- Was sind dann die Grenzen der Prävention? Zum Beispiel können Firewalls und Virens Scanner längst nicht mehr vor allen Schadprogrammen schützen, weder von der Qualität noch der Quantität her.

- Inwieweit helfen Zertifizierungen, Haftungen zu reduzieren respektive abzuwälzen? Zertifizierungen oder ähnliche Begutachtungen dienen unter anderem als Beweismittel für korrekte Gebarung oder sichere Produkte – das gilt allerdings nur für den Zustand zum Zeitpunkt der Zertifizierung.
- Soll und darf man aktiv Gegenattacken unternehmen, um einen Angreifer auszuschalten?
- Welches Restrisiko darf man in Kauf nehmen? Ein Risiko, das einen Super-GAU (Existenzverlust des Unternehmens) in Kauf nimmt, wäre wohl unangemessen. Wie kann man es aber quantifizieren?
- Was kann rechtliche Beratung in diesem Bereich leisten? Ist man auf der sicheren Seite wenn man Beratung in Anspruch genommen hat?

6 Fazit

Absolute Sicherheit ist nicht erreichbar, das zeigte sich nicht nur in den durchgeführten Risikoanalysen, sondern ist durch die teilweise spektakulären Vorfälle und Leaks der letzten Jahre belegt. Während ein Verteidiger immer aktuell sämtliche Schwachstellen und ihr Potenzial kennen muss, genügt einem Angreifer eine einzige. Ziel ist es daher, existenzbedrohende Risiken zu kennen, zu beherrschen und glaubhafte Abwehr- und Gegenstrategien für den Krisenfall vorzuhalten und umsetzen zu können.

Kann man nachweisen, alle zumutbaren Überlegungen und Vorkehrungen getroffen zu haben, verringert sich dann auch das Schadenersatz-, Straf- und Reputationsrisiko, wiewohl diesbezügliche rechtliche Aspekte nicht von vornherein abgeklärt werden können.

Die Beispiele zeigen, dass sich Risikoanalysen nach einem an Organisation und Exposition angepassten Modell, innerhalb dessen dann Risikostränge mit als gering eingestufte Priorität wie Schalen abgelöst werden, effizient durchführen lassen und zu zuverlässigen, nachvollziehbaren und vor allem aktuellen Ergebnissen führen. Kritische Erfolgsfaktoren sind dabei Know-how, Erfahrung und Kreativität der Durchführenden.

Es ist allerdings darauf zu achten, dass die Bedeutung oder der Wert sensibler oder personenbezogener Informationen nicht unterschätzt wird.

Literaturverzeichnis

- [BKA14] Österreichisches Informationssicherheitshandbuch, Version 4.0.0, Bundeskanzleramt Österreich (BKA) und Zentrum für sichere Informationstechnologie – Austria (A-SIT), <https://www.sicherheitshandbuch.gv.at/2013/index.php>, Stand: 15.10.2014.

- [BSI08] BSI Standard 100-3 „Risikoanalyse auf Basis des IT Grundschutz, Bundesamt für Sicherheit in der Informationstechnik (BSI), https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzStandards/Standard03/ITGStandard03_node.html, Stand: 2008.
- [BSI12] BSI-Veröffentlichungen zur Cyber-Sicherheit, Empfehlung an Management, Cyber-Sicherheits-Exposition, BSI-CS 013| Version 1.00, https://www.allianz-fuer-cybersicherheit.de/ACS/DE/_/downloads/BSI-CS_013.pdf?__blob=publicationFile&v=2, Stand: 2012.
- [BSI14] Leitfaden Cyber-Sicherheits-Check, Bundesamt für Sicherheit in der Informationstechnik – BSI, <https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Broschueren/Leitfaden-Cyber-Sicherheits-Check.html>, Stand: 2014.
- [ISO09] ISO/IEC 31000:2009: Risk management — Principles and guidelines, 2009.
- [EC14] E-Control, Risikoanalyse für die Informationssysteme der Elektrizitätswirtschaft unter besonderer Berücksichtigung von Smart-Metern und des Datenschutzes, <https://www.e-control.at/documents/20903/-/-/3f89d470-7d5e-433c-b307-a6443692d8f7>, Stand: 2014.
- [LD10] Larry Dignan: GM's Volt: 10 million lines of code, <http://www.zdnet.com/article/gms-volt-10-million-lines-of-code/>, Stand: 1.11.2010.
- [DL11] David Lilienthal: NYC Aviation; Aviation News, Top 4 Fun Boeing 787 Technical Facts, 2011, <http://www.nycaviation.com/2011/09/fun-facts-revealed-at-boeings-787-technical-panel/#.V2Utv3qQp0>, Stand: 28.9.2011.
- [JM13] Zhihao Jiang und Rahul Mangharam: Entwicklung eines elektrophysiologischen Herzmodells für Regelkreistests von Herzschrittmachern in Echtzeit an der University of Pennsylvania, 2013.
- [SZ16] Jürgen Schmidt, Volker Zota: heise online, Zahlen, bitte! Täglich 390.000 neue Schadprogramme, <http://www.heise.de/newsticker/meldung/Zahlen-bitte-Taeglich-390-000-neue-Schadprogramme-3177141.html>, Stand: 19.4.2016.

Predictive Analytics in der österreichischen Finanzverwaltung

Martin Setnicka¹

Abstract:

Im Rahmen des Betrugsbekämpfungspakets zur Gegenfinanzierung der Steuerreform 2015 stellt ein Schwerpunkt das Beschreiten neuer Wege durch innovative Methoden in der Fallauswahl von Steuerprüfungsfällen dar. Mit Hilfe von Predictive Analytics wird versucht aus bereits festgestellten Nachforderungsfällen, inklusive Betrugsfällen, über deren Muster und Daten-Charakteristik auf ähnlich gelagerte Fälle zu schließen. Mit den Ergebnissen der Prüfungen dieser Fälle ist ein Rückschluss auf die Qualität der Fallauswahl, aber auch auf die Qualität der Prüfungen möglich. Daher kommt einer einheitlichen qualitativollen Prüfung dieser Fälle eine besonders hohe Bedeutung zu. Durch die Anwendung von interaktiven Entscheidungsbäumen in mehreren Modellen wurde versucht mit Hilfe einer Analysesoftware relevante Unterschiede zwischen den bereits abgeprüften Veranlagungsjahren mit einer bestimmten Abgabefestsetzung und jenen abgeprüften Veranlagungsjahren mit keiner Abgabefestsetzung in den Datenbasen zu finden. Da sich die Datenbasis, je nachdem welche Informationen (Daten) zu einem Steuerfall vorliegen (mit Beilageninformationen, mit Historie, etc.) relevant unterscheidet, wurden für bestimmte Segmente unterschiedliche Modelle entwickelt.

Keywords: Finanzverwaltung, Predictive Analytics, interaktive Entscheidungsbaum

1 Einleitung

Im Zusammenhang mit dem Betrugsbekämpfungspaket zur Gegenfinanzierung der Steuerreform 2015 stellt ein Schwerpunkt des Bundesministeriums für Finanzen das Beschreiten neuer Wege durch innovative Methoden im Bereich der Erhöhung der Kontroll- und Prüfungseffektivität in der österreichischen Finanzverwaltung dar. Mit Hilfe von Predictive Analytics wird versucht aus bereits festgestellten Nachforderungsfällen (auch Betrugsfällen) über deren Muster (Daten-Charakteristik) auf ähnlich gelagerte Fälle zu schließen, um die Fallauswahl zu optimieren. Das mit dieser Aufgabe beauftragte Predictive Analytics Competence Center, kurz PACC, wurde mit 1. Juni 2016 neu gegründet. Big Data und Data-Mining haben damit Einzug in die österreichische Finanzverwaltung gehalten.

Predictive Analytics hat daten-getriebene Entscheidungen als Ziel. Wenn die Kontroll- und Prüfungseffektivität stufenweise um bis zu 20 Prozent erhöht werden könnte, entspricht dies einer 20-prozentigen Mehrergebnissteigerung. Das wären jährlich rund 50

¹ Bundesministerium für Finanzen, Teamleiter und stv. Leiter Predictive Analytics Competence Center, Hintere Zollamtsstraße 2b, 1030 Wien, martin.setnicka@bmf.gv.at

Millionen Euro ab 2019 bei der Gemeinsame Prüfung lohnabhängiger Abgaben (GPLA) und Betriebsprüfung (BP) zusammengenommen. Ziel von dem Einsatz von Predictive Analytics in der Finanzverwaltung ist es Fälle zu finden, die mit einer bestimmten festgelegten Wahrscheinlichkeit im Zuge einer Prüfungsmaßnahme zu einer Nachforderung von mehr als 10.000 EUR je Jahr und Abgabenart zu führen.

2 Methodischer Ansatz

2.1 Predictive Analytics im Bundesministerium für Finanzen

Unter Predictive Analytics sind eine Reihe von unterschiedlichen Methoden zu verstehen, deren gemeinsames Ziel die Vorhersage von bestimmten Konsequenzen aus der Analyse vorliegender Informationen (Daten) ist. Eine Methode ist das Machine Learning, wo aus der Analyse bereits vorliegender Ergebnisse Erkenntnisse aus der Daten-Charakteristik Aussagen von wahrscheinlichen Ergebnissen bei gleichen bzw. ähnlichen Datenmustern möglich werden. In der österreichischen Finanzverwaltung kam diese Methode 2014 erstmals zum Einsatz. Mithilfe von bereits abgeprüften Nachforderungsfällen wurde ähnlich gelagerten Fällen eine Wahrscheinlichkeit über ebensolche Nachforderungen zugeordnet. Die Höhe der Wahrscheinlichkeit wurde von der Analysesoftware aus der Übereinstimmung mit bekannten Datenmustern zu bereits geprüften Fällen ermittelt.

Durch die Anwendung von interaktiven Entscheidungsbäumen in mehreren Modellen wurde versucht relevante Unterschiede in den Daten jener Fälle mit einer Abgaben-Festsetzung von mehr als 10.000 EUR (Risiko 1) und jenen Fällen, deren Prüfung ohne Abgaben-Festsetzung (Risiko 0) endete. In der nachfolgenden Tabelle werden die Prüfungsfälle der GPLA der letzten acht Jahre nach dem Risiko aufgeteilt.

Risiko	Anteil in %
0	97,75
1	2,25

Tab. 1: Darstellung Risiko bei Prüfungsfälle

Es bedarf keiner 100-prozentigen Wahrscheinlichkeit, um die Fallauswahl zu verbessern. Wie in Tabelle 1 abgebildet, lag die Quote der Nachforderungsfälle über 10.000 EUR bei 2,5 %, das sind 2 Nachforderungen über 10.000 EUR bei 100 geprüften Unternehmensjahren Mit dem Einsatz von Predictive Analytics soll diese Quote mittelfristig auf 25 Prozent gesteigert werden.

2.2 Segmentierung

Zuerst benötigt man jede Menge Informationen (Daten), die für eine solche Vorhersage

notwendig sind und daher verfügbar gemacht werden müssen. In einem ersten Schritt werden Fälle mit ähnlicher Datenstruktur und -umfang [Umsatzsteuer, Einkommensteuer, Körperschaftsteuer, Gewinnfeststellung sowie Lohnabgaben und zugehörige Gewinnermittlungs- Informationen (Betriebseinnahmen und - ausgaben)] in unterschiedliche Segmente eingeteilt. Für unterschiedliche Segmente werden in einem zweiten Schritt unterschiedliche mathematisch-statistische Algorithmen (Modelle) angewandt.

In Branchen, die über eine ausreichend große Anzahl von "Lernjahren" mit mehr als 10.000 EUR Abgaben-Nachforderungen verfügten, wurden eigene Branchenmodelle berechnet, da die Vorhersagen aus konsistenten Daten innerhalb einer Branche noch einmal präziser sind. Alle Fälle, für die bereits Ergebnisse aus vorangegangenen Prüfungen vorliegen, werden ebenso in einem Segment zusammengefasst und mit eigenen Modellen gesortiert (mit einer Wahrscheinlichkeit versehen).

2.3 Variable und Regeln

Durch die Feststellung, welche Variablen bzw. Kennzahlen (Beispiele: Umsatz, Einkünfte, Fremdleistungen, Personalausgaben, und viele andere mehr, aber auch Berechnungen wie Rohaufschlag, Schuldentilgungsdauer, Wertschöpfung und wiederum viele andere mehr) mit jenen Variablen und Kennzahlen korrelieren, wie sie die "Lernjahre mit Nachforderungen" im Gegensatz zu den "Lernjahren ohne Nachforderungen" aufweisen, können mathematische Wahrscheinlichkeiten errechnet werden, die einen noch ungeprüften Fall eher zu einem Fall mit Nachforderung von mehr als 10.000 EUR zuordnen als zu einem mit einer nicht zu erwartenden Nachforderung. Bei der Auswahl der GPLA-Fälle für 2016 wurden 227 unterschiedliche Variablen eingesetzt, darunter auch Verhältnisvariable wie „Fremdpersonal und -leistungen in Prozent des Personalaufwands“ oder „Personalaufwand in Prozent des Umsatzes“.

In der Fallauswahl der BP 2016 wurden 936 unterschiedliche Variablen zur Feststellung der Korrelationen mit den Lernfällen (0/1) eingesetzt.

Die Charakteristik einer solchen Korrelation findet sich in einem Knoten des Entscheidungsbaums. Mehrere solcher aneinandergereihter Charakteristiken ergeben die Regel in der Knotenbeschreibung.

Da eine Vielzahl an Segmenten mit einer Vielzahl von Knoten zu einer besseren Qualität der Wahrscheinlichkeitsvorhersage führen, ergeben sich eine Vielzahl unterschiedlicher Regeln.

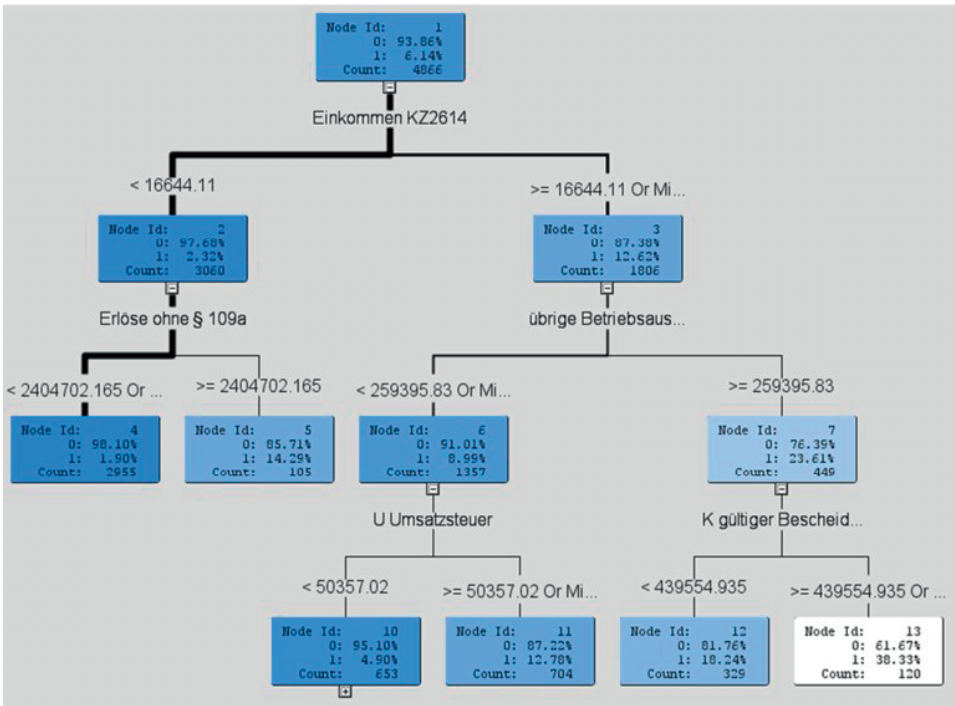


Abb. 1: Entscheidungsbaum BP 2016

2.4 Fallauswahl und Prüfungsmaßnahme

Die Korrelation bestimmter Kriterien zu einer Wahrscheinlichkeit besagt letztlich (nur), dass ein ungeprüfter Fall mit einer höheren Wahrscheinlichkeit in diesem "Risikojahr" den bereits abgeprüften Fällen des gleichen Segments, der gleichen Branche mit einer Nachforderung von mehr als 10.000 EUR ähnelt. Die dabei offen gelegten Korrelationsregeln sagen allerdings nicht unbedingt etwas über die Kausalität der Charakteristik in den Daten aus. Aus der Datencharakteristik der Fälle mit höherer Wahrscheinlichkeit können in aller Regel keine Schlüsse auf Prüfungstaktik und -methodik gezogen werden. Eine 40%ige Wahrscheinlichkeit besagt lediglich, dass Fälle mit dieser Datencharakteristik in der Vergangenheit bei 10 geprüften Veranlagungsjahren in 4 Fällen zu einer Nachforderung von mehr als 10.000 EUR führten.

Predictive Analytics wird 2016 in rund 40% aller zu prüfenden Fälle (GPLA und BP) als Fallauswahlmethode angewandt. Die Auswahl der restlichen Fälle obliegt den Prüfungsorganen nach Bedarf selbst. In der Pilotierung 2014 konnte gezeigt werden,

dass die Quote der Fälle mit einer Nachforderung über 10.000 EUR der "automatisierten" Auswahl höher ist als jene der "menschlichen" Auswahl.

2.5 Evaluierung

Für 2016 wurden 4.350 Fälle für die GPLA und 5.990 Fälle für die Betriebsprüfung mithilfe von Predictive Analytics ausgewählt. Diese Fälle stehen derzeit auf den Jahresprüfungsplänen der Finanzämter. Nach einiger Zeit wird das Ergebnis der durchgeführten Prüfungen evaluiert und Erkenntnisse daraus für die nächstjährige Fallauswahl gezogen.

Erst mit den Ergebnissen der Prüfungen dieser Fälle ist ein Rückschluss auf die Qualität der Fallauswahl, aber auch auf die Qualität der Prüfungen möglich. Daher kommt einer einheitlichen qualitätsvollen Prüfung dieser Fälle eine besonders hohe Bedeutung zu.

3 Zukünftiger Einsatz von Predictive Analytics

Predictive Analytics kann nicht nur in der GPLA und BP angewendet werden, sondern überall wo es bestimmte Kontrollmaßnahmen und Daten hierzu gibt. So könnte diese Methode sowohl im Zollbereich bei der Warenbeschau oder im Innendienst-Kontrollprozess der Steuerbescheide (Vor- oder Nachbescheidkontrolle) zum Einsatz kommen.

Einsatz eines Gasmarktmodells zur Bewertung von Risiken

Maik Günther¹

Abstract: Auf den wirtschaftlichen Erfolg der Stadtwerke München haben Energiepreise einen großen Einfluss. Ändert sich jedoch nur ein Parameter, beispielsweise der Ölpreis, kann dies Auswirkungen auf die Preise anderer Commodities haben. Die Stadtwerke München nutzen daher eine Vielzahl miteinander verketteter Marktmodelle. Mit dem weltweiten Gasmarktmodell WEGA können Gaspreise errechnet und beispielsweise der Effekt eines sinkenden Ölpreises auf den Gaspreis bestimmt werden. Aber auch Sensitivitäten zu veränderten Gasbedarfsszenarien, zu einem Überangebot an Flüssiggas oder zu einer Ukraine Krise können mit WEGA gerechnet werden. Aus den ermittelten Preiszeitreihen der relevanten Commodities werden anschließend die monetären Auswirkungen auf die Stadtwerke München abgeleitet. Diese Arbeit gibt einen Überblick über das Gasmarktmodell WEGA sowie über wesentliche Einflussgrößen auf den Gaspreis. Zudem werden die Zusammenhänge zwischen den einzelnen Commodities beleuchtet.

Keywords: Gasmarktmodell, WEGA, Commodities, NCG, Gasnachfrage, Ölpreis, Wechselkurs, LNG, Sensitivitäten, Risiken

1 Einleitung

Die Stadtwerke München (SWM) sind ein kommunales Versorgungs- und Dienstleistungsunternehmen. 2015 konnte mit 8.843 Mitarbeitern ein Umsatz von 6,6 Mrd. Euro erzielt werden [SWM16]. Beim Erdgas sind die SWM direkt und über Beteiligungen in allen Stufen der Wertschöpfungskette investiert. Hierzu gehören die Exploration und Förderung in der Nordsee, Gasnetze, Gasspeicherung, Gashandel sowie Vertriebseinheiten. Zudem haben die SWM ein Fernwärmenetz in München, welches u.a. mit gasbetriebenen KWK-Anlagen und Heizwerken versorgt wird.

Auf den wirtschaftlichen Erfolg der SWM haben Energiepreise einen großen Einfluss. Durch die Vielzahl an unterschiedlichen Assets der SWM und wegen der komplexen Wechselwirkungen zwischen den globalen Energie- und Finanzmärkten ist es nicht einfach, die Auswirkungen sich ändernder Rahmenbedingungen (z.B. Ölpreis, Gaspreis, Gasnachfrage, Wechselkurse) auf die SWM zu quantifizieren. Eine konkrete Fragestellung ist beispielsweise: Was bedeutet ein Ölpreis für die SWM, der dauerhaft 10 USD₂₀₁₅/bbl Brent unter den bisherigen Langfristprognosen bis 2040 liegt? Um derartige Fragestellungen zu beantworten, nutzen die SWM eine Vielzahl miteinander verketteter Modelle. So wirkt der Ölpreis auf den Gaspreis, der Gaspreis auf den Strompreis sowie auf den Kraftwerkspark der Zukunft. Der Einsatz dieses Kraftwerksparks hat wiederum Auswirkungen auf den Gasbedarf und somit auf den Gaspreis. Hinzu kommen Wechselwirkungen mit

¹ Stadtwerke München GmbH, Emmy-Noether-Str. 2, 80992 München, guenther.maik@swm.de

CO₂-Preisen, Kohlepreisen und Wechselkursen. Insgesamt gestalten sich diese Wechselwirkungen sehr komplex.

In dieser Arbeit wird daher auf das Zusammenspiel der Energiemärkte eingegangen, wobei dies aus Sicht der SWM mit ihren Assets aus den verschiedensten Bereichen geschieht. Dabei spielt die Modellwelt der SWM eine zentrale Rolle, weshalb die Wechselwirkungen der zahlreichen Markt- und Bewertungsmodelle ebenfalls erörtert werden. Vertiefend wird dabei der Gasmarkt beleuchtet, wobei das weltweite Gasmarktmodell WEGA² zum Einsatz kommt. U.a. wird in dieser Arbeit die Frage beantwortet, wie sich der Gaspreis bei einer Absenkung des langfristigen Ölpreisverlaufs verhält. Aber auch auf andere Treiber, die den Gaspreis beeinflussen, wird eingegangen. Im Rahmen von Sensitivitätsanalysen werden ihre Auswirkungen auf den Gaspreis quantifiziert.

Der Aufbau dieser Arbeit gestaltet sich wie folgt: Zunächst werden in Kapitel 2 die Wechselwirkungen zwischen den Energiemärkten aus Sicht der SWM erläutert. Anschließend wird in Kapitel 3 auf das Gasmarktmodell WEGA eingegangen. In Kapitel 4 werden Sensitivitätsanalysen mit WEGA durchgeführt. Diese Arbeit endet mit einer Zusammenfassung und einem Ausblick.

2 Zusammenspiel der Energiemärkte und Modellwelt der SWM

In der Einleitung wurde die Frage gestellt, welche Auswirkungen ein Ölpreis hat, der langfristig 10 USD₂₀₁₅/bbl Brent unter den bisherigen Preisannahmen liegt. Die Komplexität, die bei der Beantwortung einer derartigen Fragestellung berücksichtigt werden muss, soll nun aus Sicht der SWM verdeutlicht werden: Sinkt der Ölpreis, so sinken auch die Erträge aus der Ölförderung der SWM (bei konstanten Wechselkursen). Zudem sind Gaslieferverträge z.T. an den Ölpreis gekoppelt [IGU16], sodass Druck auf den Gaspreis ausgeübt wird, was wiederum auf die Aktivitäten der SWM in der Nordsee wirkt. Der Ölpreis hat nicht nur einen Einfluss auf den Gaspreis, sondern auch auf den Kohlepreis, da die kurzfristigen Grenzkosten für die Kohleförderung sowie die Transportkosten stark vom Ölpreis bzw. vom Preis von Ölprodukten (Diesel, Schweröl) abhängig sind. Sinkende Kohle- und Gaspreise sind zum einen positiv für den Kraftwerkspark der SWM, da Brennstoffe günstiger beschafft werden können. Jedoch werden durch sinkende Brennstoffpreise auch die zu erzielenden Strompreise gedrückt. Dies schmälert die Erlösseite der SWM-Kraftwerke. Zudem sinken die Erlöse aus der Fernwärme, da der Arbeitspreis für Fernwärme zu bestimmten Anteilen von den Preisen für Gas, Kohle und Heizöl extra leicht (HEL) abhängt. Hierbei ist zu beachten, dass der Ölpreis stark mit dem Preis für HEL korreliert. Auch die Erneuerbaren der SWM sind langfristig betroffen. Endet ihr Förderzeitraum, sind sie auf die Erlöse am Strommarkt angewiesen.

Es wird deutlich, dass die Änderung eines einzigen Parameters (beispielsweise der Öl-

² WEGA steht für Weltweites GAsmarktmodell.

preis) auch auf die Preise anderer Commodities wirken kann. Um die langfristigen monetäre Auswirkung einer Parameteränderung auf die SWM zu bewerten, müssen u.a. Preiszeitreihen für Öl, Gas, Strom, CO₂ und Kohle bis 2040 zur Verfügung stehen. Für die Öl- und Kohlepreise greifen die SWM auf kommerzielle Studien, Workshops und Experteninterviews zurück. Für die Gas-, Strom- und CO₂-Preise werden eigene Marktmodelle genutzt. Die Modellwelt der SWM ist in Abbildung 1 dargestellt. Hierbei sei erwähnt, dass die in Abbildung 1 dargestellten Modelle regelmäßig einer Validierung durch das Risikocontrolling unterzogen werden.

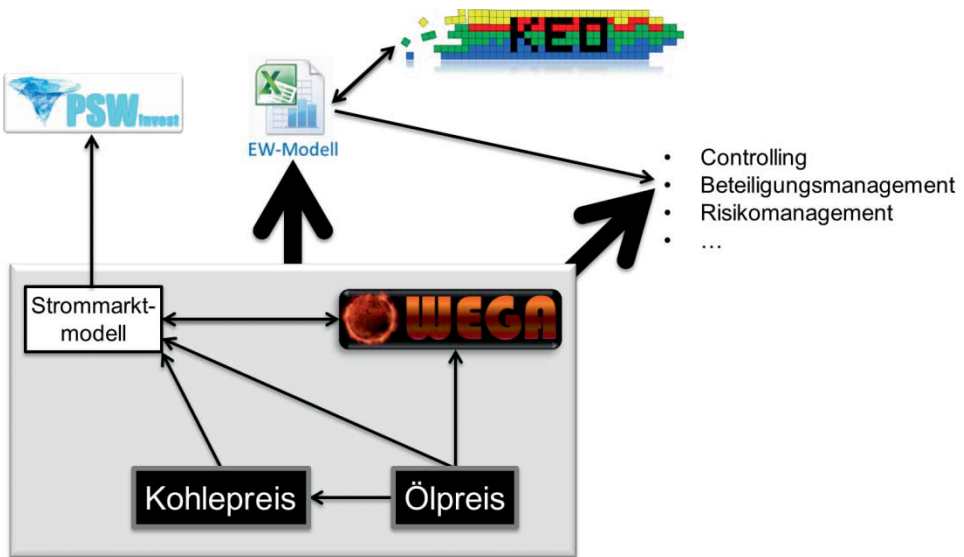


Abb. 1: Modellwelt der SWM

Ausgangspunkt in der SWM-Modellwelt ist der Ölpreis, der bei der Bestimmung des Kohlepreises Eingang findet. Zudem ist der Ölpreis wegen ölpreisindexierter Gaslieferverträge wichtig für den Gaspreis. Die SWM setzen das weltweite Gasmarktmodell WEGA ein, um Gaspreise und Gasflüsse zu errechnen [GS16], [Gü16b]. So kann mit dem Modell die Frage beantwortet werden, wie sich in Zukunft die Gaspreise bei einer Absenkung der bisherigen Ölpreisannahmen verhalten. Zusätzlich können mit WEGA auch die Auswirkungen von geänderten Eingangsparametern wie EUR/USD-Wechselkurs, Gasbedarf, Schiefergasförderung, Ukraine Krise, etc. untersucht werden. Des Weiteren wird von den SWM ein selbst entwickeltes fundamentales Strommarktmodell für Europa genutzt, um den Kraftwerkspark der Zukunft, den Wert der Erneuerbaren sowie die Strom- und CO₂-Preise zu bestimmen [SRF14]. In das Strommarktmodell gehen die Kohle-, Öl- und Gaspreise ein. Das Strommarktmodell liefert wiederum den Gasbedarf aus dem Kraftwerksektor an WEGA zurück. Die so erzeugten Commodity-Preise gehen zum einen in das EW-Modell ein, welches die Wertschöpfungsstufen der Strom- und Fernwärmeezeugung

der SWM abdeckt. Das EW-Modell selbst dient der Szenarienerstellung und Ergebnisanalyse. Konkrete Berechnungen des stundenscharfen Kraftwerkseinsatzes der SWM-Anlagen finden im Modell KEO statt [Gü13], [GGF15], [Gü16a]. Mit der Kombination aus EW-Modell und KEO werden die langfristigen Auswirkungen von geänderten Brennstoff-, CO₂- und Strompreisen auf den SWM-Kraftwerkspark sowie auf das Fernwärmenetz ermittelt. Ergebnisse des EW-Modells sowie alle Commodity-Preise werden außerdem zahlreichen Fachabteilungen zugänglich gemacht, wo u.a. weitere Berechnungen zu den monetären Auswirkungen stattfinden. In Abbildung 1 ist ebenfalls das Modell PSWinvest dargestellt. Mit ihm kann die Wirtschaftlichkeit von Pumpspeicherwerken am Spotmarkt berechnet werden [Gü15]. Die hierzu benötigte stündliche Zeitreihe der Strompreise stammt aus dem Strommarktmodell der SWM.

Neben Preiszeitreihen einzelner Commodities bis 2040 werden für gut abschätzbare Parameteränderungen zusätzlich Wahrscheinlichkeiten in einer groben Stufung angegeben. Mit Hilfe der erzeugten Preiszeitreihen kann anschließend die monetäre Auswirkung von geänderten Rahmenbedingungen auf die SWM berechnet werden.

3 Gasmarktmodell WEGA

Bei Gasmarktmodellen unterscheidet man Modelle, in denen von den Akteuren Marktmacht ausgeübt wird und Modelle, die mit vollständigem Wettbewerb arbeiten. Für einen Überblick über Gasmarktmodelle sei auf die Arbeiten von Chyong und Hobbs [CH14], Holz et al. [HHK08] sowie auf die Arbeit des Energy Modeling Forums [EM07] verwiesen.

WEGA arbeitet mit perfekter Vorausschau und vollständigem Wettbewerb. Marktmacht einzelner Akteure sowie Investitionsentscheidungen sind in Szenarien fest vorgegeben. Beispielsweise kann der Bau der Ostseepipeline Nord Stream 2 in WEGA hinterlegt werden. Hierdurch wird ein Szenario unterstellt, in dem Russland seine Marktmacht ausnutzt und bewusst Überkapazitäten erzeugt, um den Markteintritt anderer Teilnehmer zu erschweren. In beliebigen weiteren Szenarien können zusätzlich Preisvorgaben von Produzenten oder ein etwaiges Zurückhalten von Mengen unterstellt werden. Ein Szenario bis 2040 wird von WEGA in etwa 25 Minuten gerechnet.

Das Modell nutzt Linearer Programmierung (LP) und wurde in der Xpress Optimization Suite von FICO erstellt. Das Beratungsunternehmen Pöyry Management Consulting (UK) Ltd. hat das Modell entwickelt. Dort trägt es den Namen Pegasus [DaSa04], [Pö12]. 2013 haben die SWM Pegasus erworben und den Quellcode einer eingehenden Plausibilitätsprüfung unterzogen. Zudem wurde die zugehörige Datenbank von Pöyry beschafft und ebenfalls überprüft. Die Datenbank wurde von den SWM in Teilen geändert. So haben die SWM beispielsweise eigene Annahmen zum Ölpreis, zu Wechselkursen, zu den Grenzkosten von Speichern, zur Gasnachfrage oder zu Produktionsmengen einzelner Quellen. Diese Daten stammen aus eigenen Analysen oder aus dem fundamentalen Strommarktmodell der SWM. Zusätzlich kommen Daten von kommerziellen Anbietern wie PIRA

Energy Group, Wood Mackenzie, IHS und Bloomberg New Energy Finance, die z.T. eigene Gasmarktmodelle unterhalten.

WEGA ist ein weltweites Gasmarktmodell. Die Abbildung weltweiter Gasflüsse ist wichtig, da durch den zunehmenden Handel von verflüssigtem Erdgas (LNG) Märkte nicht mehr isoliert betrachtet werden können. Laut BP wird LNG bis 2035 einen größeren Anteil am weltweiten Gashandel haben als Pipelinegas [BP16]. Während der Transport mittels Pipeline an einen festen Anfangs- und Endpunkt geknüpft ist, kann LNG per Schiff von einem Verflüssigungsterminal aus zu jedem Regasifizierungsterminal der Welt transportiert werden. Gasmarktmodelle mit regionalem Fokus, die z.B. nur Nordamerika oder Europa betrachten und LNG-Flüsse fest vorgeben, bilden die Realität daher nicht ausreichend genau ab.

Mit Hilfe von unterschiedlichen Typen an Knoten und Kanten wird in WEGA der weltweite Gasmarkt abgebildet. Bei den Knoten unterscheidet man:

- **Bedarfszonen:** Dies können Länder, Regionen oder Handlungspunkte sein. Beispielsweise sind NetConnect Germany (NCG) oder Henry Hub (HH) in den USA als Handlungspunkte im Modell abgebildet. Jeder Bedarfszone ist ein tagesscharfer Gasbedarf bis 2040 zugewiesen.
- **Gasfelder innerhalb der Bedarfszonen:** Diese Gasfelder stellen die eigene Produktion innerhalb Bedarfszone dar. Die Produktionskapazität, die Produktionskosten und etwaige unterjährige Produktionsprofile (z.B. beim Gasfeld Groningen) sind bis 2040 hinterlegt.
- **Gasfelder mit Pipelineanbindung:** Im Modell werden Gasfelder zu größeren Gruppen zusammen gefasst, um die Rechenzeit zu reduzieren. Gasfelder haben eine Kapazität, Produktionskosten sowie ggf. ein unterjähriges Produktionsprofil und eine Mindestkapazität. Diese Daten werden bis 2040 vorgegeben. Ein Gasfeld in WEGA ist beispielsweise das Gasfeld Jamal in Russland.
- **Anlandepunkte für Pipelines:** Pipelines werden mit Hilfe eines Anlandepunktes an eine Bedarfszone geknüpft. Dieser Anlandepunkt hat eine maximale Kapazität. Als Beispiel kann die Anlandestation in Lubmin nahe Greifswald für die Ostseepipeline Nord Stream genannt werden.
- **Gasfelder mit Anbindung an LNG-Exportterminals:** Mit diesem Knoten werden LNG Verflüssigungsterminals modelliert. Sie haben eine Kapazität, Produktionskosten sowie ggf. ein unterjähriges Produktionsprofil und eine Mindestkapazität. Die Logik ist identisch zu Gasfeldern mit Pipelineanbindung. Es wird zwischen Gasfeldern mit Pipelineanbindung und mit Anbindung an LNG-Terminals unterschieden, um die Auswertung der Ergebnisse zu erleichtern. Auch hier wird ein Ausbaupfad für die Verflüssigungskapazität bis 2040 im Modell vorgegeben.
- **LNG-Importterminals:** Alle weltweiten Importterminals sind im Modell abgebildet. Auch ein etwaiger Ausbaupfad je Bedarfszone ist hinterlegt.

- Gasspeicher: Alle Gasspeicher Europas sind einzeln modelliert. Sie haben eine Ein- und Ausspeicherleistung sowie Kosten. Gasspeicher in Nordamerika oder China wurden zu Gruppen zusammengefasst. Etwaige Erweiterungen, Neubauten oder Stilllegungen sind bis 2040 abgebildet.
- Flexibilitätsoptionen innerhalb einer Bedarfszone: Hierzu gehören Preisgrenzen für den Coal-Gas-Switch sowie für abschaltbare Kunden. Zudem gibt es zwangsweise Lastabwürfe zur Vermeidung einer Unlösbarkeit des Optimierungsproblems. Falls diese Lastabwürfe auftreten, ist das Modell nicht ausreichend kalibriert und muss nachjustiert werden.

Für die Verbindung zwischen den Knoten gibt es eine Vielzahl an unterschiedlichen Kanten. Abbildung 2 zeigt das Beispiel zweier Bedarfszonen. Zwischen den Bedarfszonen gibt es beispielsweise eine Kante, die einen Interkonnektor darstellt. Sie erlaubt einen Gasfluss in eine vorgegebene Richtung. Für einen Interkonnektor, der Gasflüsse in beide Richtungen ermöglicht, werden daher zwei Kanten erzeugt. In Abbildung 2 ist ebenfalls zu erkennen, dass eine LNG Quelle (LNG Verflüssigungsterminal) mit mehreren Regasifizierungsterminals verbunden sein kann. Ebenso kann eine Quelle mehrere Pipelines haben und z.B. Gas aus Norwegen nach Deutschland und UK transportieren.

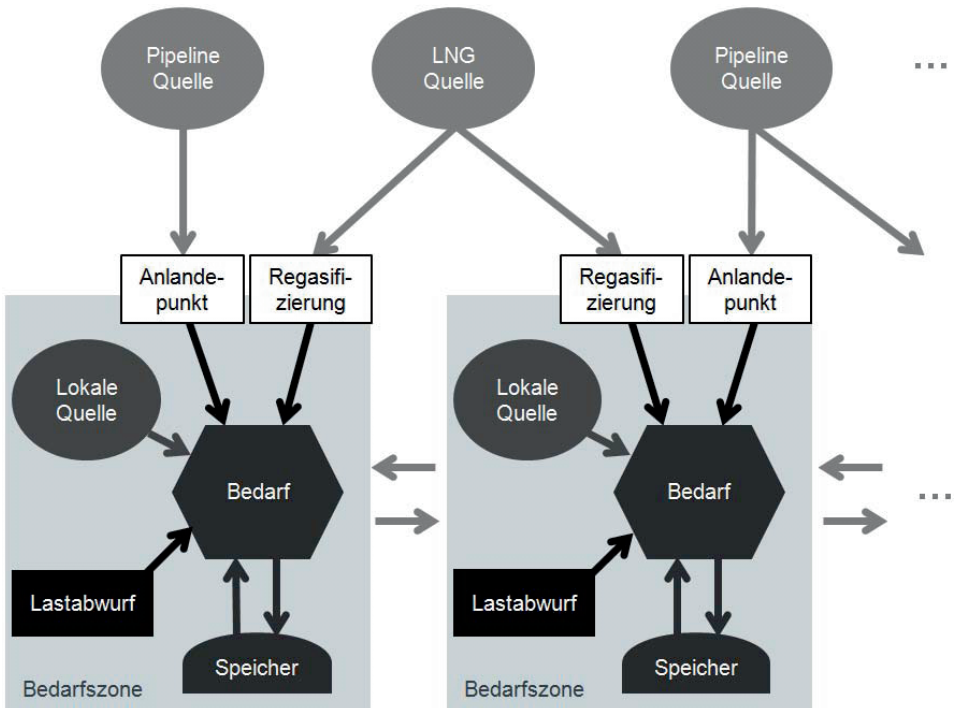


Abb. 2: Beispiel mit zwei Bedarfszonen

Das Modell enthält zudem eine Datenbank weltweiter Gaslieferverträge. Ein Gasliefervertrag hat ein Anfangs- und ein Enddatum. Er beinhaltet zudem Flexibilitätsoptionen, Take-or-Pay Klauseln sowie eine individuelle Preisformel. Diese Preisformel berücksichtigt eine Ölpreisindexierung, eine Hubindexierung oder eine Mischung aus Beidem. Etwaige Nachverhandlungen von Verträgen können ebenfalls hinterlegt werden. So wird im Modell von einer langsamen Abkehr von ölpreisindexierten Verträgen hin zu hubindexierten Verträgen oder hybriden Verträgen ausgegangen [IGU16].

WEGA optimiert jedes Gasjahr separat, wobei das Gasjahr in einzelne Tage ‚d‘ zerlegt wird. Das Ziel der Optimierung ist es, die Gesamtkosten zur Deckung des täglichen Gasbedarfs eines Gasjahres ‚K_{Gesamt}‘ zu minimieren (siehe Formel 1). Die Gesamtkosten setzen sich aus den tagesscharfen Produktionskosten ‚K_{Prod_d}‘, Transportkosten ‚K_{Transp_d}‘, Speicherkosten ‚K_{Spei_d}‘ sowie aus Flexibilitätskosten ‚K_{Flex_d}‘ zusammen.

$$\text{Min } K_{\text{Gesamt}} = \sum_{d=1}^D (K_{\text{Prod}_d} + K_{\text{Transp}_d} + K_{\text{Spei}_d} + K_{\text{Flex}_d}) \quad (1)$$

Bei der Minimierung der Gesamtkosten sind zahlreiche Nebenbedingungen einzuhalten. Eine harte Nebenbedingung ist die Deckung des Gasbedarfs jeder Bedarfszone ‚z‘ an jedem Tag ‚d‘ (siehe Formel 2). In Formel 2 meint ‚Ein‘ jeweils einen Zufluss zur Bedarfszone und ‚Aus‘ einen Abfluss. Der tägliche Gasbedarf einer Bedarfszone ‚z‘ wird durch Zu- und Abflüsse der zur Bedarfszone gehörenden Pipelines, LNG Regasifizierungsterminals (LNGEin_{z,d}) bzw. LNG Exportterminals (LNGAus_{z,d}), Interkonnektoren, Speicher und Flexibilitätsoptionen gedeckt. Weiterhin gibt es Zuflüssen durch Eigenproduktion innerhalb der Bedarfszone.

$$\begin{aligned} \text{Bedarf}_{z,d} = & \text{PipelineEin}_{z,d} - \text{PipelineAus}_{z,d} + \\ & \text{LNGEin}_{z,d} - \text{LNGAus}_{z,d} + \\ & \text{InterkonnektorEin}_{z,d} - \text{InterkonnektorAus}_{z,d} + \\ & \text{SpeicherEin}_{z,d} - \text{SpeicherAus}_{z,d} + \\ & \text{FlexibilitätsoptionEin}_{z,d} + \text{EigenproduktionEin}_{z,d} \quad \forall z \in Z, d \in D \end{aligned} \quad (2)$$

Weitere Nebenbedingungen beziehen sich auf die zuvor genannten Eigenschaften der Knoten, Kanten und Verträge. So darf beispielsweise die maximale Kapazität einer Pipeline oder eines LNG-Regasifizierungsterminals nicht überschritten werden. Auch vertraglichen Vereinbarungen aus den langfristigen Lieferverträgen sind einzuhalten.

Zur Visualisierung der Vielzahl an Eingangsdaten und Ergebnissen ist die Business Intelligence Software Tableau in das Gasmarktmodell integriert. Abbildung 3 zeigt einen Screenshot von WEGA. Mit Ziffer eins ist der Bereich markiert, in dem neue Szenarien erstellt oder bestehende Szenarien ausgewählt werden können. Darunter lassen sich im Bereich mit der Ziffer zwei vordefinierte Ansichten eines Szenarios auswählen, welche dann im Bereich mit der Ziffer drei angezeigt werden. Dort können sie auch manipuliert werden.

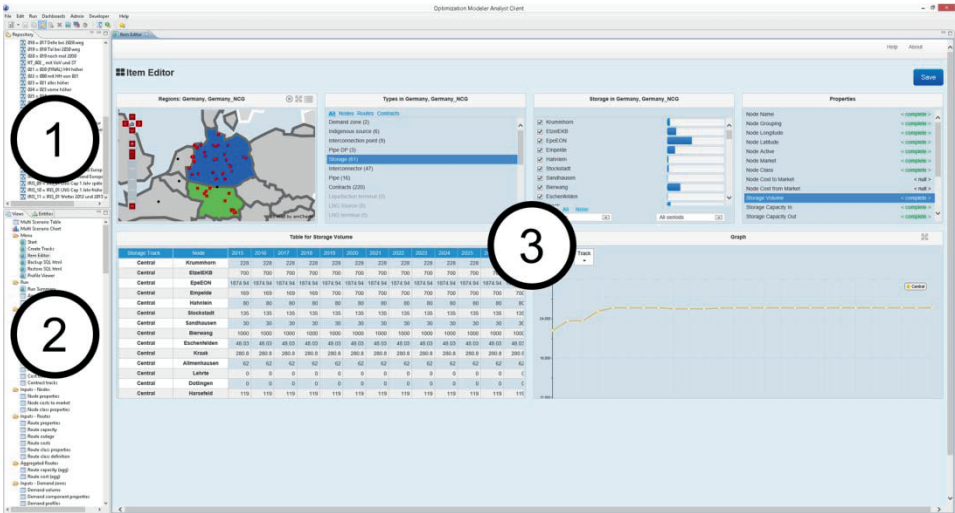


Abb. 3: Screenshot von WEGA mit Daten über Gasspeicher in Deutschland

4 Einfluss von Parametern auf den Gaspreis

In diesem Kapitel werden relevante Parameter vorgestellt, die einen Einfluss auf den Gaspreis haben. Anhand von Sensitivitätsanalysen mit Hilfe des Gasmarktmodells WEGA werden die Auswirkungen von Parameteränderungen quantifiziert. Soweit möglich, werden zudem grobe Aussagen getroffen, wie wahrscheinlich die untersuchten Parameteränderungen sind.

Ausgangspunkt für die Sensitivitätsbetrachtungen bildet ein Basis-Szenario, welches die bestmögliche Sicht auf die Zukunft darstellt. Das Basis-Szenario dieser Arbeit wurde im sinnvollen Rahmen modifiziert, um keine Rückschlüsse auf das SWM-Basis-Szenario zuzulassen (z.B. bei den Wechselkursen). Ausgehend vom Basis-Szenario dieser Arbeit werden Sensitivitäten abgeleitet, wobei immer nur ein Parameter geändert wird. Wegen interner Einschränkungen werden keine NCG-Preise angegeben, sondern prozentuale Preisabweichungen zum Basis-Szenario. An dieser Stelle soll der Hinweis gegeben werden, dass unter perfekter Vorausschau gerechnet wurde. Man wird bei den Sensitivitäten beispielsweise nicht von einem höheren Gasbedarf oder Bedarfsspitzen überrascht.

Geänderter Ölpreis: In Europa findet man eine Ölpreisindexierung vor allem noch bei Pipelineimporten aus Russland sowie bei LNG-Importen aus Algerien. Der Anteil der Ölpreisindexierung hat in der Vergangenheit zu Gunsten einer Indexierung auf Hub-Preise abgenommen. In Asien hingegen spielt die Ölpreisindexierung bei LNG-Importen noch eine relativ große Rolle [IGU16]. Wegen der derzeit niedrigen Ölpreise besteht aktuell wenig Druck, von der Ölpreisindexierung abzurücken.

In Abbildung 4 ist der Verlauf des NCG-Preises bei modifizierten Brent-Preisen dargestellt. Hier wurde der Brent-Preis in jedem Jahr um einen festen Betrag angehoben bzw. abgesenkt. Z.B. um +10 USD₂₀₁₅/bbl Brent als Realwert zum Basisjahr 2015. Man erkennt deutlich, dass der Ölpreis einen erheblichen Einfluss auf den NCG-Preis hat. Dieser nimmt jedoch in Zukunft ab, da in WEGA ein weiterer Rückgang des Anteils der Ölpreisindexierung unterstellt ist.

Auffällig ist, dass ein Preisanstieg bei Brent bis 2020 einen höheren Einfluss auf den NCG-Preis hat, als eine Preissenkung in den jeweiligen Jahren um den gleichen Betrag. Dies liegt daran, dass der Ölpreis momentan sehr niedrig ist und der Preis für ölpreisindexiertes Gas unter dem NCG-Preis liegt. Es macht somit Sinn, die komplette Volumenflexibilität ölpreisindexierter Verträge auszunutzen und möglichst viel Gas über diese Verträge zu beziehen. Erhöht man in dieser Situation den Ölpreis, so wird ölpreisindexiertes Gas zunehmend unattraktiv und man erreicht ein Niveau, in dem nur so viel ölpreisindexiertes Gas importiert wird, wie man laut Vertrag kaufen muss (Take-or-Pay Level). Das relativ günstige ölpreisindexierte Gas wird demnach nicht nur teurer, man bezieht von ihm auch weniger und muss andere, relativ teure Quellen nutzen. Bei einer Absenkung des Ölpreises von dem heute schon niedrigen Preisniveau ist der Effekt auf den NCG-Preis weniger stark ausgeprägt, da man bereits Mengen deutlich oberhalb des Take-or-Pay Levels bezieht und weniger Spielraum nach oben zur vertraglich maximal möglichen Menge hat.

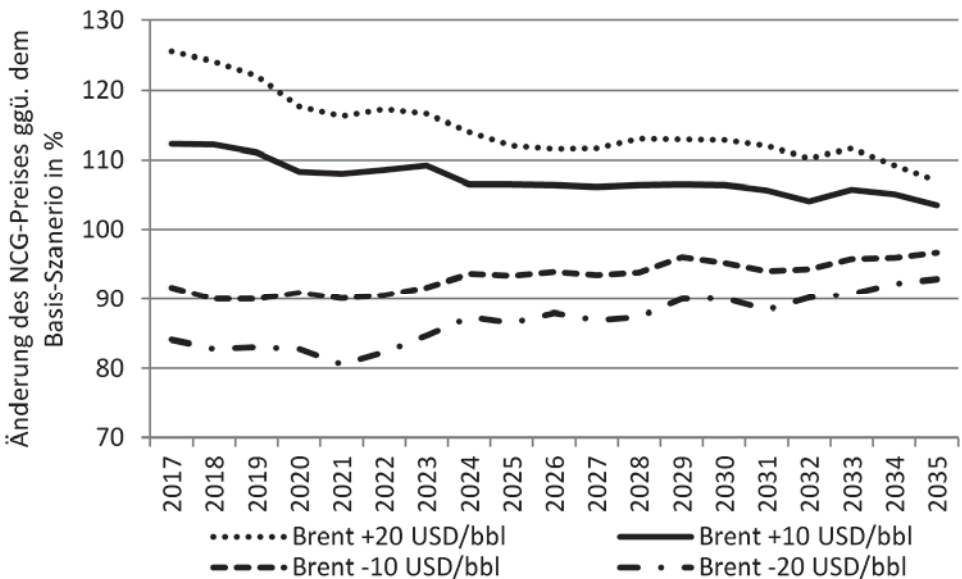


Abb. 4: Auswirkung geänderter Brent-Preise auf den NCG-Preis ggü. dem Basis-Szenario

Geänderter Wechselkurs: Eine Änderung des EUR/USD-Wechselkurses setzt ähnliche Mechanismen wie ein geänderter Ölpreis in Gang. Nordseeöl der Sorte Brent wird in USD/Barrel (bbl) gehandelt. Wird der Euro gegenüber dem Dollar stärker, so zahlt ein

Käufer in Europa, der in Euro bezahlt, weniger Euros pro bbl. Dies ist vor dem Hintergrund der zuvor dargestellten Ölpreisindexierung wichtig.

Die Auswirkungen eines EUR/USD-Wechselkurses von 1,0 sowie 1,1 und 1,2 sind in Abbildung 5 dargestellt. Je stärker der Euro gegenüber dem Dollar ist, umso niedriger ist der NCG-Preis. Den gleichen Preiseffekt würde man bei konstantem Wechselkurs und abgesehenem Ölpreis erhalten. Ein langfristiger EUR/USD-Kurs von 1,0 erscheint unrealistisch. Selbst bei den niedrigen Ölpreisen der letzten Monate ist dies kurzfristig nicht aufgetreten. Ein stärkerer Euro gegenüber dem Dollar ist eher wahrscheinlich.

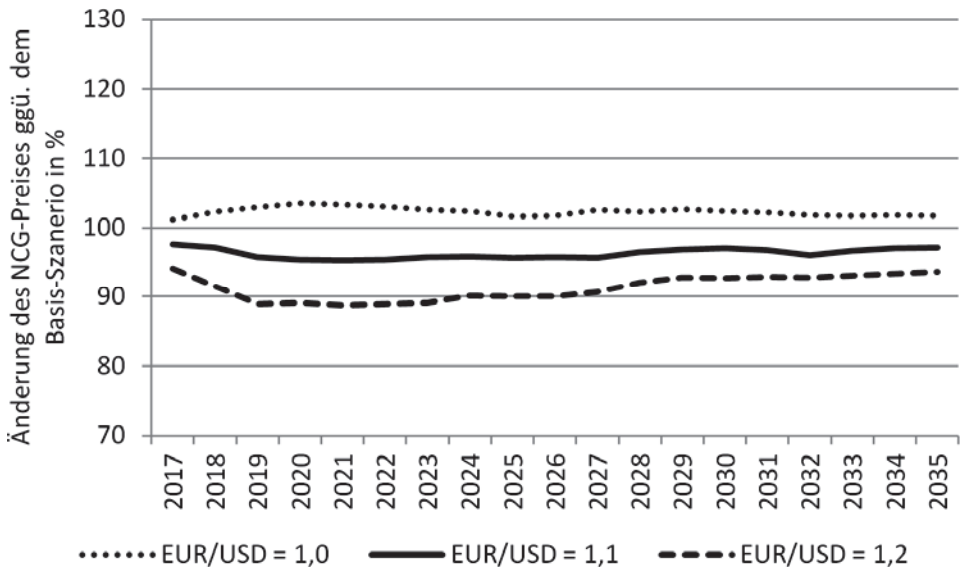


Abb. 5: Auswirkung geänderter EUR/USD-Kurse auf den NCG-Preis ggü. dem Basis-Szenario

Geänderter HH-Preis: Auch der Preis von HH hat einen Einfluss auf den Gaspreis in Europa. Denn die USA exportieren in Zukunft größere Mengen LNG, die meist auf den Preis am HH indexiert sein werden. Für die Berechnungen in Abbildung 6 wurden die Rahmenbedingungen im Modell so geändert, dass sich ein um 35% höherer bzw. niedrigerer Henry Hub-Preis ergibt.

Während die Änderung beim HH-Preis mit +35% relativ groß ausfällt, bewegt sich der NCG-Preis kaum. Dies liegt daran, dass nur der HH-Preis modifiziert wurde, nicht jedoch die LNG-Exportkapazitäten in den USA. Steigt der HH-Preis auf ein Niveau, bei dem ein Export zu langfristigen Grenzkosten unwirtschaftlich ist, wird das LNG-Exportterminal dennoch produzieren, solange es seine kurzfristigen Grenzkosten (z.B. ca. 3 EUR₂₀₁₅/MWh statt 15 EUR₂₀₁₅/MWh für Exporte von den USA nach Europa) verdient. Der Preisanstieg am HH wird damit gedämpft auf den NCG-Preis übertragen. Bei einem niedrigeren HH-Preis bewegt sich der NCG-Preis kaum. U.S.-LNG ist in der Merit-Order der Gasquellen im Basis-Szenario selten preissetzend, sodass ein Absenken beim Henry

Hub-Preis kaum auf den NCG-Preis wirkt. Sollte ein anderer HH-Preis als im Basis-Szenario eintreten, so ist ein niedrigerer HH-Preis eher wahrscheinlicher.

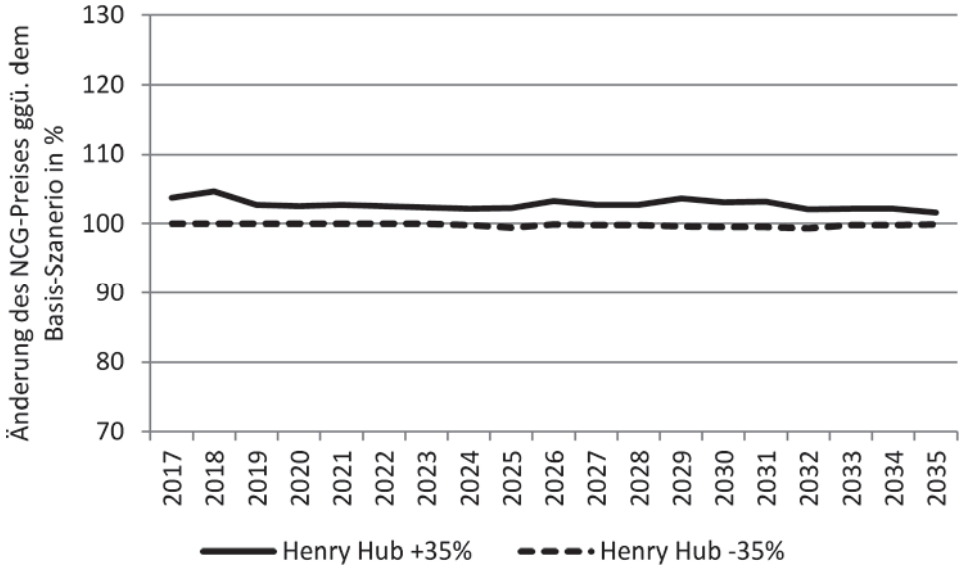


Abb. 6: Auswirkung geänderter Henry-Hub-Preise auf den NCG-Preis ggü. dem Basis-Szenario

Geänderter Gasbedarf Europas: Nachfolgend wird sich im Rahmen von Sensitivitäten einer veränderten Gasnachfrage gewidmet. Wiederum wird nur die Nachfrage geändert. Alle anderen Parameter bleiben unverändert. Zur Analyse des Einflusses des Gasbedarfs in Europa auf den Gaspreis für NCG wurde der Gasbedarf pauschal in jedem Jahr um 10% angehoben bzw. abgesenkt. Hierbei wurde die Vorgabe für den jährlichen Gasbedarf modifiziert. Über hinterlegte Profile wird aus ihm im Modell der tägliche Gasbedarf bestimmt. Die Profile selbst bleiben dabei unverändert.

In Abbildung 7 ist das Ergebnis der Berechnungen dargestellt. Bis zu den 20er Jahren bewirkt ein um 10% erhöhter bzw. abgesenkter Gasbedarf in Europa einen um mehr als 10% höheren bzw. niedrigeren NCG-Preis. Grundsätzlich ist die Tendenz zu erkennen, dass ein höherer Gasbedarf in Zukunft einen weniger hohen Preisanstieg nach sich zieht. Dies liegt daran, dass Europa bereits sehr gut mit Gas versorgt ist und Überkapazitäten an LNG bis Mitte der 20er Jahre nach Europa kommen werden. Auch zusätzliche Mengen an Pipelinegas aus dem Kaspischen Raum über die TAP/TANAP-Pipeline sowie russisches Pipelinegas aus Jamal gelangt in den nächsten Jahren nach Europa. Europa ist somit zunehmend mit Gas übertversorgt, was bis Ende der 20er Jahre anhält. Eine höhere Nachfrage kann daher zunehmend günstiger bedient werden.

Auffällig ist, dass eine Bedarfsabsenkung um 10% zunächst zu einer relativ starken Preissenkung bei NCG beiträgt. Da der Ölpreis momentan sehr niedrig ist, befinden sich ölpreisindexierten Verträge recht weit links in der Merit-Order der Gasquellen und sind dort

wegen der maximal genutzten Volumenflexibilität auch recht breit. Sinkt nun die Nachfrage, fallen teurere Quellen auf der rechten Seite aus der Merit-Order der Gasquellen. Im Modell ist ab Anfang der 20er Jahre ein Ölpreis auf einem höheren Niveau als heute hinterlegt. Damit besteht der zuvor dargestellte Effekt bzgl. der Merit-Order bereits Anfang der 20er Jahre nicht mehr. Die Mengen aus ölpreisindexierten Verträgen erreichen nicht mehr das maximal mögliche Volumen, die Merit-Order wird an dieser Stelle schmaler und teurere Quellen werden benötigt, um den Bedarf zu decken. Trotz eines bis Ende der 20er Jahre übertourten Marktes führt ein Bedarfsrückgang in Europa hierdurch nicht mehr zu den Preisabschlägen wie in den ersten Jahren.

Da der weltweite Gasbedarf und somit auch das weltweite Angebot wachsen, fällt eine Bedarfssteigerung bzw. ein Bedarfsrückgang in Europa von 10% in Zukunft anteilig immer kleiner aus. Auch dies führt zu dem Effekt, dass sich die beiden Kurven in Abbildung 7 der X-Achse annähern. Grundsätzlich verläuft der Gasbedarf in Europa bis 2035 verhältnismäßig konstant im Basis-Szenario. Sollte es zu einer Abweichung kommen, wird eine Abweichung nach unten als wahrscheinlicher angesehen als nach oben.

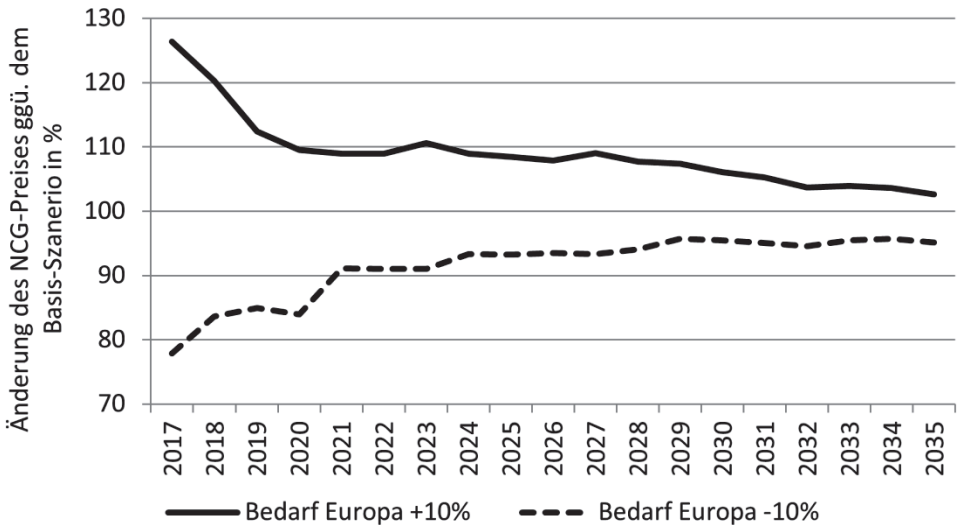


Abb. 7: Auswirkung geänderter Gasbedarfe auf den NCG-Preis ggü. dem Basis-Szenario

Geändertes Wetterprofil: Der Gasbedarf jeder Bedarfszone wird in WEGA als ein Wert pro Jahr vorgegeben. Mit Hilfe von Profilen werden daraus Tagesbedarfe generiert. Diese Profile wurden aus historischen Daten abgeleitet. Sie repräsentieren ein durchschnittlich warm-kaltes Jahr, ohne extreme Ereignisse. Die Ergebnisse in Abbildung 7 wurden beispielsweise mit diesen Durchschnittsprofilen gerechnet. WEGA bietet jedoch die Möglichkeit, die Profile historischer Jahre zu nutzen. So kann man beispielsweise berechnen, wie sich der Gaspreis ändert, wenn jedes Jahr in der Zukunft das Wetter vom Gasjahr 2012 hätte.

In Abbildung 8 sind die Ergebnisse für die historischen Profile von 2012 und 2013 dargestellt. Dabei handelt es sich immer um Profile eines Gasjahres. Während das Gasjahr 2012 relativ kalt war und Nachfragespitzen hatte, war das Gasjahr 2013 relativ mild. Dies erkennt man auch in den Ergebnissen. So ist der NCG-Preis mit dem 2012er-Profil höher als der NCG-Preis mit dem 2013er-Profil.

Außer in den Jahren 2017 und 2018 beim 2012er-Profil gibt es kaum Änderungen im Gaspreis für NCG. Der Gasmarkt ist gut versorgt und gerade Deutschland ist mehr als ausreichend mit Gasspeichern ausgestattet. Zwar geht die Produktion im Gasfeld Groningen in den Niederlanden zurück. Es kommt in den nächsten Jahren jedoch weitere Flexibilität durch zusätzliche weltweite LNG-Exportkapazitäten auf den Markt. Die Ergebnisse zeigen, dass der Wert von Flexibilität gering ist und weiter sinken wird.

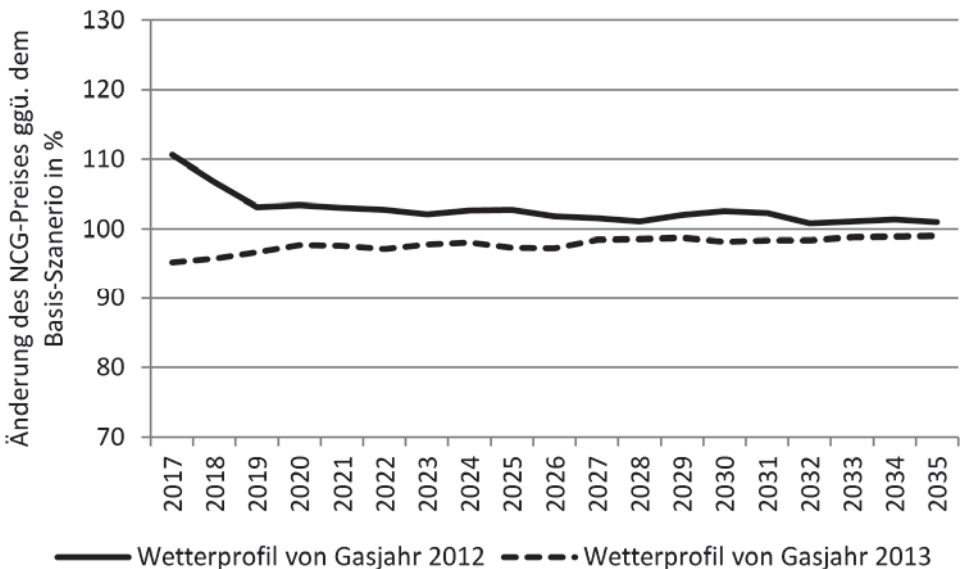


Abb. 8: Auswirkung historischer Bedarfsprofile auf den NCG-Preis ggü. dem Basis-Szenario

5 Zusammenfassung

In dieser Arbeit wurde zunächst das komplexe Zusammenspiel verschiedener Commodities aus Sicht der SWM erläutert. Dabei wurde auch auf die Modellwelt der SWM eingegangen. Beim Blick auf die Modelle lag ein Fokus auf dem Gasmarktmodell WEGA.

Konkrete Rechenergebnisse mit WEGA haben gezeigt, dass der Ölpreis auch in Zukunft einen signifikanten Einfluss auf den Gaspreis haben wird. Hierbei spielt auch der EUR/USD-Kurs eine wichtige Rolle. Weitere Berechnungen haben gezeigt, dass eine hö-

here europäische Nachfrage nach Erdgas in den 20er Jahren nur einen geringen Preisanstieg auslöst. Zudem wurde beim Einsatz verschiedener Wetterprofile deutlich, dass der Wert von Flexibilität eher gering ist und weiter abnehmen wird.

Zukünftig sollen deutlich mehr Sensitivitäten über alle Commodities hinweg gerechnet werden. Das Ziel ist die Erstellung eines Ergebnisraums, aus dem sich die Auswirkungen möglicher Parameteränderungen auf die SWM schnell ablesen lassen.

Literaturverzeichnis

- [CH14] Chyong, C. K., Hobbs, B. F.: Strategic Eurasian Natural Gas Market Model for Energy Security and Policy Analysis. *Energy Economics*, 44/09, S. 198-211, 2014.
- [DaSa04] Davies, G., Sarsfield-Hall, R.: Gas SCR – Cost Benefit Analysis for a Demand-Side Response Mechanism. A report to Ofgem, Pöyry, 2004.
- [EM07] EMF: Prices and Trade in a Globalizing Natural Gas Market, EMF Report 23, 2007.
- [GGF15] Günther, M.; Greller, M.; Fallahnejad, M.: Evaluation of Long-Term Scenarios for Power Generation and District Heating at Stadtwerke München. In: *Informatik Spektrum*, Vol. 38(2), S. 97-102, 2015.
- [GS16] Günther, M.; Schimpf, M.: Auswirkungen von Energieeinsparungen auf den Gasbedarf und den Gaspreis in Deutschland bis 2040, In (Mayr, H. C.; Pinzinger, M., Hrsg.): *Informatik 2016*, LNI, im Druck, 2016.
- [Gü13] Günther, M.: Power plant scheduling in long term scenarios with GAMS. In: 27th PuK-Workshop, KI 2013. 36th Annual Conference on Artificial Intelligence, S. 57-64, 2013.
- [Gü15] Günther, M.: Wirtschaftlichkeit von Pumpspeicherwerken am Spotmarkt, In (Heimerl, S., Hrsg.): *Wasserkraftprojekte. Ausgewählte Beiträge aus der Fachzeitschrift Wasserwirtschaft*, Band 2, Wiesbaden: Springer Vieweg, S. 341-346, 2015.
- [Gü16a] Günther, M.: Kraftwerkseinsatzoptimierung mit GAMS. Berechnung von Langfristszenarien für strategische Entscheidungen, In: *Einsatz von OR-Verfahren zur Analyse von Fragestellungen im Umweltbereich*, Heidelberg: Shaker Verlag, S. 20-33, 2016.
- [Gü16b] Günther, M.: Practical Application of a Worldwide Gasmarket Model at Stadtwerke München. In: (Dörner, K. u.a., Hrsg.): *OR Proc. (OR 2015)*, Springer, im Druck, 2016.
- [HHK08] Holz, F., von Hirschhausen, C., Kemfert, C.: A Strategic Model of European Gas Supply (GASMOD). *Energy Economics*, 30/08, S. 766-788, 2008.
- [IGU16] IGU, Wholesale Gas Price Survey - 2016 Edition. A global review of price formation mechanisms 2005-2015, 2016.
- [Pö12] Pöyry: How will intermittency change Europe's gas markets?, 2012.
- [SRF14] Schaber, K.; Roth, H.; Fallahnejad, M.: Can the Gas Sector Provide the Flexibility to the Power Sector for the Integration of Renewables? In: 13th WIW, S. 46-51, 2014.
- [SWM16] SWM: Stadtwerke München. Geschäftsbericht 2015, 2016.

Konzept einer Risikoanalyse im Verwaltungsverfahren durch agentenbasierte Simulation

Thomas Preiß¹

Abstract: Risikoanalyse vermag mögliche Schadensaspekte aufzeigen. Neben den Methoden der statistischen Modellbildung können durch Etablierung eines agentenbasierten Modells Auswirkungen eines konkreten Handelns von staatlichen Stellen aufgezeigt werden. Es werden die konzeptionellen Grundlagen des Modells, das sich hier exemplarisch am Verwaltungsstrafverfahren orientiert, sowie die Einbindung in prozessuale Vorgaben im Rahmen der Legislative dargestellt. Ein Ausblick zeigt die Ausweitung des in diesem Bereich neuen Modellierungsansatzes auf die durch den Gesetzgeber vorgesehenen Rechtsschutzmechanismen auf.

Keywords: Risikoanalyse, Verwaltungsverfahren, Verwaltungsstrafen, agentenbasierte Modellbildung und Simulation

1 Vorbemerkung

Gesetze und Normen regeln in einer Gesellschaft das Handeln der verschiedensten natürlichen und juristischen Personen zu dem Zweck, dass – wohl ganz allgemein gesagt – (sozialer) Friede und weitestgehender Wohlstand im Gemeinwesen vorherrscht. Die Gesetzgebung obliegt den durch die Verfassung eingerichteten Gremien. Dass das System aus Gesetzgebung, den an der Gesellschaft Beteiligten und den Organen des Staates, denen unmittelbar regelndes Eingreifen obliegt, ein komplexes ist, wissen wir aus der allgemeinen Lebenserfahrung.

2 Simulation im Recht und Risikoanalyse

Damit vor dem eigentlichen Wirkungseintritt einer Norm deren Folgen abgeschätzt werden können, schlägt [Re76] (S. 148, 162) den Einsatz von Simulation vor. Durch die Anwendung von mathematischen Modellen und deren Realisierung in Computersystemen können ressourcenschonend verschiedene Szenarien beobachtet und analysiert werden. [Ze76] (S. 27 f.) spricht hier von dem „experimental frame“, *Reisinger* von einem „Modellexperiment“. Gemeinsam ist diesen Begriffen, dass eine Wiederholbarkeit der Simulationsläufe gegeben ist, sodass die Übertragbarkeit der gewonnenen Ergebnisse auf die Wirklichkeit überprüft werden kann.

Dass der Gesetzgeber die Untersuchung von „was wäre wenn“ Szenarien angeordnet hat, ist naturgemäß – wohl auch wegen der relativ leichten Abbildbarkeit – im monetären

¹ Wissenschaftlicher Dienst, Amt der NÖ Landesregierung / Universität Wien, thomas.preiss@aon.at.

Bereich angesiedelt. Es ist im Rahmen der wirkungsorientierten Folgenabschätzung² die budgetäre Auswirkung von Regelungsvorhaben anzugeben. Die dem Utilitarismus nicht fernstehende Betrachtungsweise nach ökonomischen Gesichtspunkten lässt sich auch auf das Handeln des Staates in Hinblick auf die Bürger anwenden. So beschreibt [Wh08] (S. 386 f) eine „optimale“ Strafe im Verwaltungsstrafbereich, deren dogmatischer Hintergrund der Rechtsökonomie entnommen ist. Diese wurde bereits als Grundlage für Modellierungsthesen ([Pr15], S. 216: „Abbildungsthese“ und „Injektivitätsthese“) angewandt. Aus diesem dargestellten Modellierungsansatz folgt, dass nicht die Rechtsnormen als logisches Kalkül zur „Bewertung“ oder „Subsumption“ von Handlungen von Normunterworfenen modelliert werden, sondern die bereits unter Anwendung der durch den Gesetzgeber vorgegebenen Regeln „Durchführungsanweisungen“ (Algorithmen).

3 Agentenbasierte Modellbildung (ABM)

Agentenbasierte Modellbildung³ ist ein dezentralisierter, individualzentrierter (im Gegensatz zum systembasierten) Ansatz für den Modellentwurf. Bereits [Ki05] (S. 268, 291, 292) schlägt die Anwendung dieses zu diesem Zeitpunkt neuen Modellbildungsansatzes in der öffentlichen Verwaltung vor. So sei erst durch die Verknüpfung des Handelns verschiedener Agenten die Auswirkung von Regelungen in einem komplexen System beobachtbar. Hier handelt es unter anderem um Interaktionen zwischen der Verwaltung und den Bürgern, die trotz gegebener Heterogenität strukturelles Wirken an den Tag legen. Dadurch sieht [Ki05] auch Möglichkeiten zur Risikoanalyse und Risikovermeidung gegeben, da die Auswirkungen von (organisatorischen) Maßnahmen von einem ausreichend validen Modell abgelesen werden können. Anwendungen dieses Modellbildungsparadigmas im Zusammenhang mit rechtlichen oder soziologischen Fragestellungen wurden in den letzten Jahren entwickelt, sodass von einer Etablierung gesprochen werden kann (vgl. [Ai14],[BDP13],[LMT13], [RRS12]).

Ausgehend von bereits erfolgten strukturellen Überlegungen ([Pr15], S. 77)⁴ stellt Abbildung 1 die wesentlichen Elemente eines ABMs dar, die sich mit den Ausführungen von Fußnote 4 wiederfinden. Im Modellkonzept werden diese als Agenten oder einzelne Objekte vorgesehen und sind daher:

- das Verhalten des Bürgers, das zur Verwaltungsübertretung führt,

² Geregelt durch die „Verordnung der Bundesministerin für Finanzen über die Abschätzung der finanziellen Auswirkungen auf die öffentlichen Haushalte im Rahmen der wirkungsorientierten Folgenabschätzung bei Regelungsvorhaben und sonstigen Vorhaben (WFA-Finanzielle-Auswirkungen-Verordnung – WFA-FinAV).

³ Vgl. <http://www.anylogic.de/agent-based-modeling>, abgerufen am 04.04.2016.

⁴ Die Struktur wird als „Risikodreieck“ bezeichnet. Der Bürger („Ecke C“) begeht eine Verwaltungsübertretung, die seitens des Meldungslegers („Ecke A“) an die Behörde („Ecke B“) mittels geeigneter Beweismittel gemeldet wird. Nach einem geeigneten Verfahren zur Bewertung ergeht eine Verwaltungsstrafe als Erledigung wiederum an den Bürger als Normunterworfenen. Weitere strukturelle Beschreibungen orientieren sich an der von [EPB13], (S. 2300, 2301), gewählten Darstellungsform.

- das Verfahren der Meldungslegung (hier handelt es in erster Linie um automatisierte Messungen), das entsprechende Beweismittel bereitstellt,
- die von der Behörde etablierten Prozesse, die unter Anwendung von in IT Systemen festgeschriebenen Regeln zu einer Bewertung führen, die als
- geeignete Erledigung an den Bürger übermittelt wird.

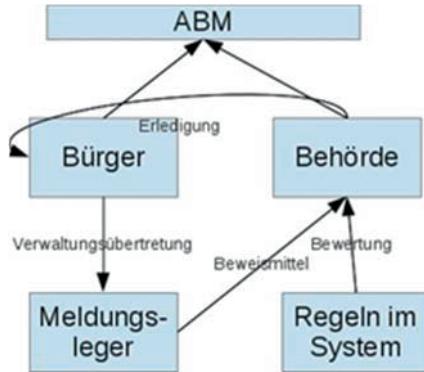


Abb. 1: : Struktur des Problems, angelehnt an das "Risikodreieck"

Der im Verfahren vorgesehene Rechtsschutz wird in dieser Sicht der Problemstruktur aus Systemvereinfachungsgründen vorerst nicht berücksichtigt.

3.1 Struktur des Modells

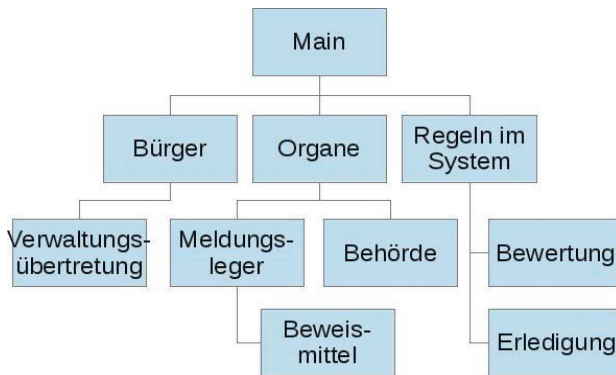


Abb. 2: : Hierarchische Struktur des ABM

Das Konzept des Modells orientiert sich an dem objektorientierten Ansatz, wie er von Systemen wie AnyLogic verwendet wird. Wesentliche Attribute und Eigenschaften der angeführten Objekte sind beim Bürger: übliche Stammdaten, bei der Verwaltungsübertretung: Ort, Zeit, Art, die Übertretung beschreibende Daten, beim

Meldungsleger: Ausbildungsgrad, Beweismittel: angewandte Ausrüstung, gesetzlich vorgeschriebene Nachweise, bei der Behörde: personelle Ausstattung, Ausbildungsgrad, bei den Regeln im System: „Entscheidungsontologien“, die zur Bewertung und Wahl der Erledigung verwendet werden.

3.2 Vereinfachende Annahmen

Bestimmte, einem Modell zur allfälligen Wahl einer „optimalen“ Strafe zuzuführende Parameter, werden durch vereinfachende Annahmen zu ersetzen sein. Der Grund dafür ist in erster Linie darin zu suchen, dass keine Daten den aktuell eingesetzten Datenbanken nicht entnommen werden können oder aus Datenschutzgründen nicht zur Verfügung stehen. Abbildung 3 zeigt ein vereinfachtes Modell der Behörde, das als Teil des Konzepts vorgesehen ist.

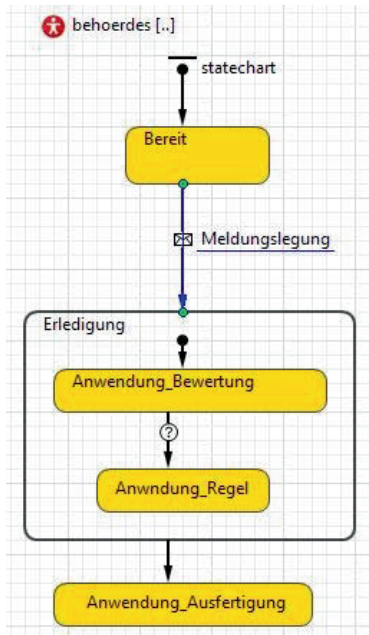


Abb. 3: Beispiel Behördenmodell

Es werden daher aus Informationen, die dem Bereich der „Open Government Data“ zuzuordnen sind, mittels Monte Carlo Methoden Basisdaten zur Simulation erzeugt. Das Modellkonzept sieht aber auch die Anwendung von Echtdate vor, so der Einsatz durch Auftraggeber zur jeweiligen Datenanwendung im Sinne des Datenschutzgesetzes (DSG 2000) erfolgt. Dadurch ist gewährleistet, dass durch Monte Carlo Verfahren nicht oder schwierig abbildbare Sonderfälle berücksichtigt werden können.

4 Prozessuale Anwendungen

Simulation führt zu Erkenntnissen, die auf die Wirklichkeit übertragbar sind. In unserem Fall ist dies das Erkennen von Auswirkungen von Prozessen innerhalb der Verwaltung, die in unterschiedlicher Qualität ablaufen. Zu deren Sicherstellung sind Institutionen wie die Rechnungshöfe, die Volksanwaltschaft und weitere gesetzlich vorgesehene Aufsichtsstellen eingerichtet.

Da hier stark formalisierte Verfahren, die einer effizienten Abbildbarkeit zugänglich sind, verwendet werden, können Änderungen des Prozesses und deren Auswirkungen, auch im Sinne der WFA, beobachtet werden. Geeignet sind hier die Ansatzpunkte der Etablierung von Verfahren zur Ermittlung einer optimalen Strafe nach *White*, der vermehrten Berücksichtigung der E-Government Strategie Österreichs und der Verbesserung der länderübergreifenden Behördenzusammenarbeit. Hier sind die genannten Aufsichtsstellen sowie der nationale und europäische Gesetzgeber befragt.

5 Ausblick

Dem Bürger sind Rechtsmittel zuzugestehen, die die Abwehr einer Rechtswirkung einer behördlichen Handlung ermöglichen. Diese finden dann Anwendung, wenn der Bürger das angewandte Verfahren oder Resultat als ungerecht empfindet. Der soziologischen, mathematischen Modellbildung sind Versuche zu entnehmen, wie diese wohl hoch emotional anzusetzenden Werte eines Rechtsempfindens abzubilden sind. Hier zeigt es sich, dass nicht auf Forschung im empirischen Bereich verzichtet werden kann. Ähnliche Problemstellungen wurden durch Auswertung von Fragebögen, die durch betroffene Bürger ergänzt wurden, gelöst.

Der Gesetzgeber hat Verfahren zur „Vorabkontrolle“ von Vorhaben, in erster Linie an Utilitaritätsgedanken orientiert, vorgesehen. Die hier zu einem Modellkonzept zusammengefassten Überlegungen scheinen geeignet zu sein, vermehrt einen agentenbasierten Modellierungsansatz vorzusehen.

Die Flexibilität dieses Ansatzes darf aber auch nicht dessen wesentliche Voraussetzungen bzw. Grenzen außer Acht lassen: Jeder Modellbildung hat eine geeignete Validierung zu folgen. Weiters sind Aspekte aufzuzeigen, die durch vereinfachende Modellannahmen zwar „ausgeblendet“ werden können aber für einen vertieften Erkenntnisgewinn wesentlich sind. Dies sind vor allem soziale Parameter, die einer „unmittelbaren“ Messbarkeit nicht zugänglich sind.

Schließlich ist „im Großen“ ein Verfahren vorzusehen, dass die Durchführung von logistischen Maßnahmen aufgrund von validen Simulationsergebnissen vorsieht bzw. vereinfacht.

Literaturverzeichnis

- [Ai14] Aizstrautsa, A. et al.: Architecture for Distributed Simulation Environment. ICTE in Regional Development. Valmiera, Latvia, 2014.
- [BDP13] Balke, T.; De Vos, M.; Padget, J.: I-ABM: combining institutional frameworks and agent-based modelling for the design of enforcement policies. In (Wyner, A.; Benn, N., (Hrsg.): Artificial Intelligence and Law 21, S. 371-398, 2013.
- [EPB13] Einzinger, P.; Popper, N.; Breitenecker, F.: The gap-drg model: simulation of outpatient care for comparison of different reimbursement schemes. In (Pasupathy, R et al., Hrsg.): Proceedings of the 2013 Winter Simulation Conference, 2013.
- [Ki05] Kiel, L. D.: A primer for agent-based modelling in public administration. In: Public Administration Quarterly Volume 29 No. 3/4, S. 268 - 296, 2005.
- [LMT13] Lotzmann, U.; Möhring, M.; Troitzsch, K. G.: Simulating the emergence of norms in different scenarios. In (Andrighetto, G. et al., Hrsg.): Artificial Intelligence and Law 21, S. 109-138, 2013.
- [Pr15] Preiß, T.: Die Bedeutung der Risikoanalyse für den Rechtsschutz bei automatisierten Verwaltungsstrafverfahren, Dissertation, Universität Wien, 2015.
- [Re75] Reisinger, L.: Planspiel und Simulation im Recht. In: Forschungen aus Staat und Recht 32, S. 148-166. Springer Verlag, Wien, New York, 1975.
- [RRS12] Riveret, R.; Rotolo, A.; Sartor, G.: Probabilistic rule-based argumentation for norm-governed learning agents. In (Andrighetto, G. et al., Hrsg.): Artificial Intelligence and Law 20: S. 383-420, 2012.
- [Wh08] White, M. D.: Time speeding behavior and optimal penalties. In: The Journal of socioeconomics, S. 384 - 399, Elsevier, 2008.
- [Ze76] Zeigler, B. P.: Theory of modelling and simulation, Wiley, New York, 1976.

Business Process Management
und Strategisches
Informationsmanagement
im Öffentlichen Sektor

Business Process Management und Strategisches Informationsmanagement im Öffentlichen Sektor

Dagmar Lück-Schneider¹, Konrad Walser²

Der öffentliche Sektor unterliegt anderen Rahmenbedingungen als Unternehmen. Auch ist das Aufgabenspektrum der Öffentlichen Verwaltung ausgesprochen groß und über unterschiedlichste Zuständigkeitsbereiche verteilt. Betrachtungen darüber, inwieweit bestimmte Vorgehensweisen auch im Öffentlichen Sektor zielführend sind, können somit bedeutsam sein. Das umfasst die Analyse durchgeführter Praxisaktivitäten genauso wie Transferüberlegungen zu aus der Wirtschaftsinformatik übertragenen Ansätzen.

Zu den ursprünglich für zwei getrennte Workshops aufgerufenen Call for Papers war die Einreichungslage so, dass sich die Organisatoren zu einer gemeinsamen Durchführung entschlossen. Andererseits konnten alle Einreichungen, die jeweils im Blind Review-Verfahren gesichtet wurden, angenommen werden. Die erste Einreichung, die dem Business Process Management zuzuordnen ist, befasst sich mit dem in der Wirtschaft in Prozessen mit Workflowmanagementsystem-Unterstützung bereits erprobtem Process Mining. Zwei weitere Beiträge sind im Bereich des strategischen Informationsmanagements angesiedelt. Im ersteren wird unter dem Ansatz der Transaktionskostentheorie ermittelt, unter welchen Bedingungen die Eigenerstellung oder eine Fremdvergabe (Make or Buy) von IT-Services für öffentliche Verwaltungen denkbar ist. Darüber hinaus wird untersucht, wie sich Gemeinden des Kantons Bern im Verhältnis zu diesen Bewertungen verhalten und es werden Empfehlungen für künftiges Agieren abgeleitet. Im zweiten Beitrag wird eine in mehrere Dimensionen zielende Befragung von über 130 CIOs aus deutschen Krankenhäusern zum aktuellen Stand von Informationsmanagement Systemen in diesem Segment vorgestellt. Last but not least befasst sich ein weiterer dem BPM zuordenbarer Beitrag mit der Branche, die ehemals staatlich war, die Telekommunikationsindustrie. In diesem Beitrag wird das subjektorientierte BPM auf das Prozessreferenz-Rahmenwerk enhanced Telecom Operations Map (eTOM) angewendet und die entsprechende Anwendung validiert.

¹ Hochschule für Wirtschaft und Recht Berlin, dagmar.lueck-schneider@hwr-berlin.de

² Berner Fachhochschule, konrad.walser@bfh.ch

Process-Mining in der öffentlichen Verwaltung

Dagmar Lück-Schneider¹

Abstract: In diesem Beitrag werden zunächst die für den Sektor der öffentlichen Verwaltung nicht verbreiteten Möglichkeiten des Process-Minings² vorgestellt und mit den für die öffentliche Verwaltung erhobenen Zielsetzungen des Prozessmanagements abgeglichen um dann Überlegungen zu Anwendungsfeldern und künftigen Entwicklungen dieser Methode für die öffentliche Verwaltung anzustellen. Grundlegende Aspekte und Voraussetzungen zur Anwendung der Methode werden genauso berücksichtigt wie die durch Nutzung der Methode resultierenden, über herkömmliches Geschäftsprozessmanagement hinausgehenden Potenziale. Darauf aufsetzend werden schließlich Betrachtungen angestellt, in welchen Bereichen der öffentlichen Verwaltung zum gegenwärtigen Zeitpunkt sowie künftig unter welchen Voraussetzungen Einsatzfelder gesehen werden können.

Keywords: Process-Mining, Geschäftsprozessmodellierung, Geschäftsprozessmanagement, öffentliche Verwaltung, Log-Dateien, E-Government

1 Einleitung

Im gleichnamigen Workshop der Informatik 2015 wurde das Forschungsgebiet Process-Mining in der Keynote vorgestellt. Es erschien spannend, sich damit zu befassen, welche Potenziale in dieser Methode für die öffentliche Verwaltung liegen könnten.

Ein solches Vorhaben umfasst, zunächst den gegenwärtigen Stand des Prozessmanagements in der öffentlichen Verwaltung darzustellen. Hierzu wird auf eine aus 2014 stammende Studie zurückgegriffen.

Außerdem erfolgt eine kurze Einführung in die Methode des Process-Minings, die im Umfeld der öffentlichen Verwaltung wenig bekannt ist – das bestätigte auch eine Nachfrage unter den Anwesenden auf dem Workshop BPMÖS der Informatik 2015 in Cottbus – und daher im Rahmen dieser Einreichung sinnvoll erscheint. Die hieraus resultierenden Möglichkeiten werden vorgestellt und darauf aufbauend werden schließlich Aussagen über die Anwendbarkeit der Methode auf Prozesse in der öffentlichen Verwaltung abgeleitet.

¹ Hochschule für Wirtschaft und Recht, Fachbereich Allgemeine Verwaltung 10315 Berlin, Alt Friedrichsfelde 60, dagmar.lueck-schneider@hwr-berlin.de

² Die Wahl für diese Schreibweise erfolgte in Anlehnung an die im Duden aufzufindende empfohlene Schreibweise für den Begriff „Data-Mining“, die auch die freie Enzyklopädie Wikipedia verwendet.

2 Stand Prozessmanagement im öffentlichen Sektor

2014 wurde durch PwC³ und IMTB⁴ eine Studie zur gegenwärtigen Situation des Geschäftsprozessmanagements in Deutschland auf den Verwaltungsebenen von Bund, Ländern und Kommunen durchgeführt. Ziel der Studie war es, „den aktuellen Umgang mit Prozessmanagement in der Verwaltung zu beleuchten und Empfehlungen zum Umgang mit diesem Thema abzuleiten“ ([PI14], S.11). Die Umfrage umfasste eine Erhebung quantitativer Daten über einen Online-Fragebogen sowie ergänzende qualitative Interviews ([PI14], S. 20). Insgesamt wurden 711 deutsche Verwaltungen um Teilnahme gebeten, von 223 gab es einen Rücklauf (32,1%) ([PI14], S. 20). Bezogen auf die verschiedenen Verwaltungsebenen verteilten sich die Rückläufe mit 41,5%, 28% und 31,6% auf Bund, Länder und Kommunen.

Die im Kontext dieses Beitrages wichtigsten Ergebnisse der Studie sind die folgenden:

- „Haupttreiber für den Einsatz von Prozessmanagement sind die organisatorische und die wirtschaftliche Notwendigkeit“ ([PI14], S. 8 sowie S. 27, Abb. 11 / die Abbildung ist in diesen Beitrag als Abb. 1 aufgenommen worden, vgl. folgende Seite).
- Herausragende Ziele liegen in der Optimierung des Personaleinsatzes sowie der Verkürzung der Bearbeitungszeiten; unter allen anderen wesentlich seltener genannten Zielen werden Qualitätsmanagement und Standardisierung am meisten genannt.

Gerade diese Notwendigkeiten und Ziele lassen sich mit der Methode des Process-Minings – über die traditionellen Ansätze hinaus – ebenfalls unterstützen, sofern die Datenlage der Prozesse dies ermöglicht.

Hinzu kommt, dass hierdurch nicht nur ein alternativer Zugang zu gleichen möglichen Erkenntnissen geschaffen wird, sondern Ergebnisse entstehen, die über die Aussagekraft herkömmlicher Methoden des Geschäftsprozessmanagements hinausgehen, weil sie auf real mit Zeitstempel mitgezeichneten Aktivitäten beruhen und dadurch weitere Steuerungsoptionen liefern. Durch Vergleiche von Ergebnissen beider Zugänge ergeben sich weitere Folgerungen. Auf diese Aspekte wird im Abschnitt 3 noch eingegangen.

Das vom Deutschen Institut für Normung herausgegebene Vorgehensmodell für Geschäftsprozessmanagement in der öffentlichen Verwaltung gibt als eine Möglichkeit zur Ermittlung des Sollprozesses die Simulation (Wenn-Dann-Analysen) an. Werden entsprechende Kennzahlen (Wahrscheinlichkeiten für Ablaufvarianten, durchschnittliche Bearbeitungsdauer, Liegezeiten etc.) zu Prozessschritten hinterlegt, so lassen sich „Auswirkungen von Optimierungs- und Gestaltungsmaßnahmen vor ihrer Umsetzung abschätzen und vergleichen“ ([DI09], S. 31). Geschäftsprozessmanagement-Tools weisen hierfür eigene Analyse- bzw. Simulationskomponenten auf. Derartige Möglichkeiten werden als **Ex-ante**-Techniken bezeichnet. „Diese kommen für die Bewertung von Ist-

³ Pricewaterhouse Coopers Aktiengesellschaft Wirtschaftsprüfungsgesellschaft

⁴ IMTB Group GmbH

Prozessen und vor der Umsetzung möglicher Soll-Alternativen zum Einsatz. Dabei wird jeweils davon ausgegangen dass die zugrunde liegende Prozessstruktur bekannt ist“ ([BK13], S. 147). Die hinterlegten Daten können für mathematische Auswertungen oder aber auch für Simulationen genutzt werden. In letzterem Falle bestimmen die hinterlegten Wahrscheinlichkeiten die Festlegung von an Verzweigungen aktiv werdenden Zufallsgeneratoren. In jedem Falle hängen die Ergebnisse von den zuvor ermittelten, zumeist durch Selbstaufzeichnung, Beobachtung oder Schätzung ermittelten Kennzahlen ab.

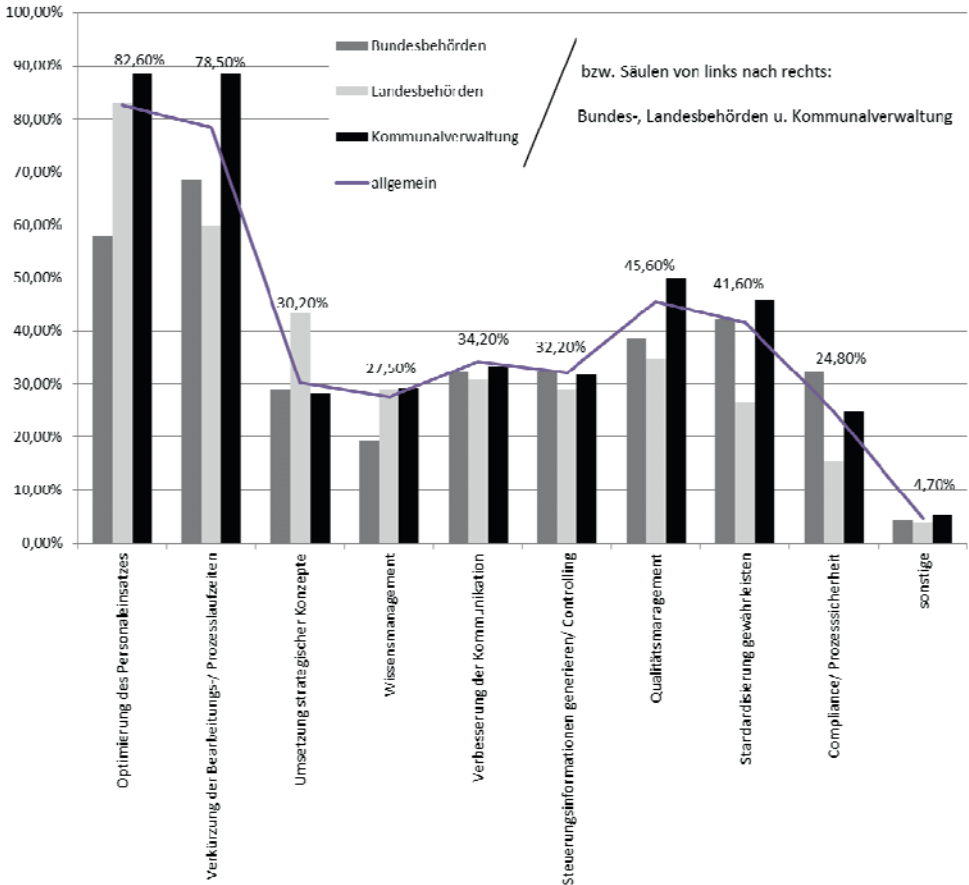


Abb. 1: Ziele des Prozessmanagements nach [PI14], S. 28

3 Möglichkeiten von Process-Mining

Process-Mining zählt zu den **Ex-post**-Analysen ([BK13], S. 147) und kann auch mit Echtzeit-Analysen gekoppelt werden. Basis für Process-Mining sind durch IT-Systeme generierte Ereignis-Daten (event data). Diese müssen bestimmte Voraussetzungen erfüllen. So müssen sie

- eindeutig bestimmten Fällen (Cases) und (*)
- Aktivitäten zugeordnet werden können und darüber hinaus
- mit einem Zeitstempel versehen sein.

Aus diesen Mindestangaben lassen sich dann für real vorgekommene Fälle Verläufe (Folgen von Aktivitäten) generieren. Existieren hinreichend viele Fälle, so können basierend auf diesen offensichtlich möglichen Verläufen Modelle gewonnen werden, ganz ohne herkömmliche Prozessmodellierung.⁵ Die Modelle dürften umso genauer sein, je mehr Fälle für die Generierung des Modells herangezogen wurden.

Das Beispiel in Tabelle 1 zeigt ein fiktives Beispiel zu einer Datenlage, wie sie durch ein Ticketsystem zur Bearbeitung von IT-Fehlermeldungen in Behörden oder Unternehmen entstanden sein könnte.

Für Process-Mining-Aktivitäten ist die Ticket-Nummer von wesentlicher Bedeutung, sie hält einen Fall zusammen und bildet die „Case ID“, die Fallidentifikationsnummer. Die Person, die den Fall auslöst (Beschäftigte(r) kann mehrfach in verschiedenen Fällen vorkommen, da ja durchaus mehrere Probleme mit der IT auftreten können und es nicht realistisch wäre anzunehmen, dass stets ein bereits gemeldetes Problem vor dem Auftauchen eines neuen behoben wäre. Sie wäre daher für eine Fallidentifizierung alleine nicht ausreichend. Die Aktivitäten werden für die Modellbildung benötigt.

Zusätzliche Spalten mit ihren Inhalten – in Tabelle 1 ist nur eine davon angedeutet – können ebenso für eine Modellbildung und für Analysen wesentlich sein. Gilt beispielsweise, dass Fehlermeldungen per Telefon, E-Mail oder online über das System gemeldet werden können, so führt dies im Beispiel zu unterschiedlichen Verläufen im Modell. Bei einer Online-Meldung entfällt das Aufzeichnen des Problems im Ticketsystem durch die IT-Beschäftigten, da dies dann der Beschäftigte selbst erledigt. Auch seine Daten (wer hat das Ticket ausgelöst?) werden automatisch erfasst. Alternativ denkbar wäre natürlich auch, dass das Ticketsystem hierfür von vornherein unterschiedliche Aktivitäten vorsieht. Eine Spalte mit dem bearbeitenden Sachbearbeiter des IT-Services ist ebenfalls wichtig. Und selbstverständlich wird in solchen System auch erfasst, wie der Fehler behoben wurde. Dies dient dem sukzessiven Aufbau einer Wissensdatenbank.

⁵ Eine solche Nutzung von Daten wird von van der Aalst als „play in“ bezeichnet, während er das Auslesen möglicher Verläufe als „play out“ und das Einspielen von Ereignisdaten in ein Modell als „replay“ bezeichnet ([Aa11], S. 18ff.).

Ticket-nr.	Aktivität	Zeitstempel	Beschäftigte(r)	Problem	...
1	Meldung Fehler	12.04.16-9:01:00	Arnold	Tastatur defekt	
2	Meldung Fehler	12.04.16-9:01:16	Broch	Word Problem	
1	Fehlerabschluss	12.04.16-9:15:00	Arnold	Tastatur defekt	
3	Meldung Fehler	12.04.16-9:20:25	Crems	Maus defekt	
2	Überarbeitung Fehlerbeschreibung	12.04.16-10:10:25	Broch	Bildausrichtung klappt nicht	
2	Recherche	12.04.16-10:13:09	Broch	Bildausrichtung klappt nicht	
...					

Tab. 1: Mögliche Ereignisdaten eines Ticketsystems

Gerade in Verbindung mit zusätzlichen Spalten können durch **Data-Mining** weitere Analysemöglichkeiten erschlossen werden. So kann man entlang der Spalteninhalte natürlich auch nach Zusammenhängen zwischen Spalten suchen, etwa ob sich bestimmte Probleme in bestimmten Abteilungen häufen oder ob es bestimmte Zeiten gibt, in denen Fehler öfter auftreten. Auch funktionale Abhängigkeiten sind denkbar, wenn etwa bestimmte Fehlermeldungen grundsätzlich mit anderen zusammentreffen. Dies könnte zu vorausschauendem Agieren des IT-Service führen. Entsprechende Auswertungen zählen zum Data-Mining. Hierbei werden keine Zusammenhänge zwischen Inhalten und Prozessverlauf hergestellt. Data-Mining-Möglichkeiten sind nicht Gegenstand dieses Beitrags, wenngleich sie ebenfalls genutzt werden können, um Prozesse z. B. nach Kundensegmenten orientiert zu optimieren.

Die Güte der durch Falldaten entwickelten Modelle hängt von der Anzahl der vorliegenden Fälle und deren Repräsentativität ab. Das generierte Modell kann zu allgemein oder auch zu speziell geraten. Ersteres gilt etwa, wenn es in unserem Beispiel nur aus den Aktivitäten Fehlermeldung – Fehlerbearbeitung – Fehlerdokumentation besteht, vielleicht, weil bis dahin noch keine anderen Situationen aufgetaucht sind (vgl. Abb. 2⁶). Letzteres ist möglich, wenn gleichartige Verläufe nicht zusammengefasst werden und für jeden vorliegenden Fall ein eigener Pfad in das Modell aufgenommen wird (vgl. Abb. 3), der jeweils exklusiv (Raute mit X) je nach Fall beschriftet wird.



Abb. 2: Prozessmodell in BPMN-Notation, wenig aussagekräftig, weil sehr generell gehalten

⁶ Alle Prozessmodelle dieses Beitrags sind in BPMN-Notation dargestellt.

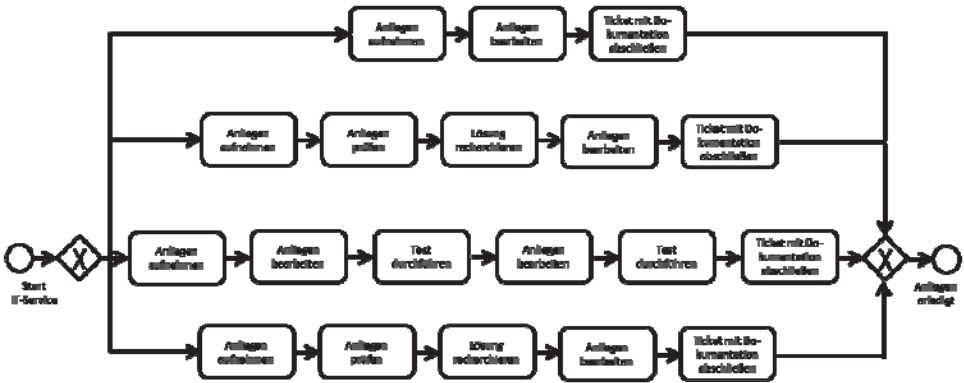


Abb. 3: Prozessmodell in BPMN-Notation, zu speziell, viele Aktivitäten doppelt enthalten

Dies zeigt, dass es keineswegs selbstverständlich ist, von vorliegenden Falldaten zu einem guten Modell zu kommen. Ein Modell, das den in Abb. 3 enthaltenen Fällen besser gerecht wird, ist in Abb. 4⁷ zu sehen. Hier ist keine Aktivität im Modell mehrfach vorhanden. Dazu trägt u. a. auch die vorkommende Schleife bei, die ein wiederholtes Testen abbilden kann.⁸

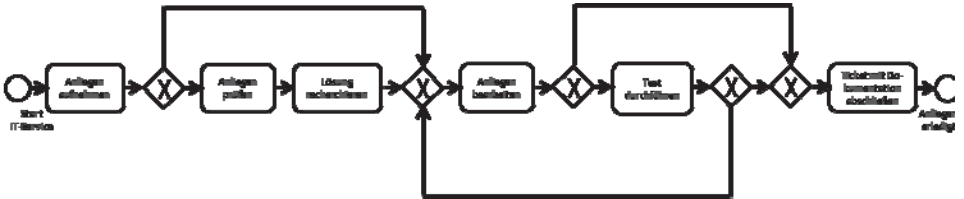


Abb. 4: nach aufgetretenen Fällen angemessenes Prozessmodell

Diese Betrachtungen verdeutlichen zudem, dass man keineswegs davon ausgehen kann, dass nur ein einziges Modell als Ergebnis entstehen kann. Unabhängig von den schon aus dem herkömmlichen Prozessmanagement bekannten vielen Prozessmodellnotationen ist die Auswahl des Algorithmus, mit dem aus den Falldaten ein Modell erzeugt wird, entscheidend für die Gestalt des Modells.

Und ohne weitere Fachbegriffe an dieser Stelle einzuführen ist besondere Aufmerksam-

⁷ Aus Platzgründen wurden in dieser Abbildung die Symbole für die Aktivitäten in der Darstellungsbreite an die Textinhalte angepasst.

⁸ Zusätzliche Fälle würden das Diagramm möglicherweise weiter verändern, so etwa, wenn festgestellt wird, dass die Prüfung des Anliegens nicht zwingend auch eine Lösungsrecherche nach sich zieht.

keit darauf zu richten,

- dass die Modelle eine Start- und Ende-Marke besitzen,
- dass alle möglichen Fälle stets bis zu einer Ende-Marke gelangen,
- dass es keine Modellabschnitte gibt, die gar nicht benötigt werden.
- dass der weitere Verlauf ab einer Marke stets klar definiert ist.

Besonders heikel ist das automatische Herausfinden von Schleifen in den Algorithmen. Die vorhandenen Programme bieten etliche verschiedene Möglichkeiten zur Erstellung von Modellen auf Basis von Ereignisdaten an. Am vorgenannten Extremfall, dass für jeden Fall im automatisch generierten Modell ein eigener Pfad mit eigenen Aktivitäten angelegt wird, kann man gut deutlich machen, dass auf diese Weise auf Schleifen komplett verzichtet werden kann. Ein solches Modell ist natürlich bei dem Vorliegen von Daten zu vielen Fällen völlig unbrauchbar und an einen Vergleich mit einem herkömmlich erzeugten Modell braucht man nicht einmal im Entferntesten zu denken.

Ärgerlich ist es aber auch, dass man nicht einfach nach einem „besten Algorithmus“ suchen kann. Je nach Datenlage kann mal der eine, mal der andere zu einem nach vorgegebenen Kriterien besten Ergebnis führen. Um die erzeugten Modelle fundiert beurteilen zu können, ist es erforderlich, sich sowohl mit möglichen Gütekriterien für Modelle als auch mit den Stärken und Schwächen der die Modelle generierenden Algorithmen zu befassen. Das ist ein umfangreiches und sehr anspruchsvolles Unterfangen (***) und kann innerhalb dieses Beitrags nicht geleistet werden. Hier muss auf die angegebene und weitere Fachliteratur verwiesen werden.

Im Folgenden wird davon ausgegangen, dass mit einem geeignet gewählten Algorithmus ein der Situation angemessenes Modell entstanden ist. Die dargestellten Möglichkeiten gelten für alle Anwendungsbereiche des Process Minings, sie sind also nicht spezifisch für den öffentlichen Sektor.

Prozessoptimierung

Über die vorliegenden Zeitstempel lassen sich in Verbindung mit den Fallidentifikationszahlen die Verläufe genau analysieren. Bestehen zwischen zwei Aktivitäten stets größere Differenzen bzgl. der Zeitstempel, sind hier lange Übergangszeiten. Ebenso lassen sich Schleifen entdecken (so etwa im Beispiel beim Auftreten mehrfacher Fehlerbeschreibungseinträge zu ein und demselben Fall), die Hinweise auf Problemsituationen sein können. Damit kann Fragen nachgegangen werden, die auf die Optimierung der Prozesse ausgerichtet sind. Auch selten auftretende Pfade lassen sich ermitteln und zur Grundlage von Prüfungen heranziehen (lassen sich solche Situationen ggf. völlig vermeiden?).

Übereinstimmungsprüfungen

Vergleiche mit herkömmlich gestalteten Modellen können Aufschluss darüber geben, ob in den automatisch generierten Modellen Pfade übersehen wurden oder ob möglicher-

weise bestimmte Pfade im durch Process-Mining entstandenen Modell noch nicht aufgetreten sind (conformance checking). Ebenso lassen sich Vergleiche so erzeugter Modelle mit den auf herkömmlichen Wegen erstellten Modellen zur Qualitätssicherung nutzen (compliance checking): Werden tatsächlich alle gewünschten Vorgaben berücksichtigt? Werden beispielsweise vor der Stellenbesetzung in sicherheitsrelevanten Bereichen tatsächlich stets vorab die Ergebnisse der Sicherheitsüberprüfung abgewartet?

Simulation

Und schließlich können durch ein Einspielen von vorhandenen Daten in die erzeugten Modelle (replay) tatsächliche Simulationen erfolgen, die z. B. Bearbeitungsschritte verdeutlichen können, an denen es stets zu Wartezeiten kommt. Das wiederum kann Hinweise darauf geben, dass Personal anders zugeordnet werden muss.

4 Prozess-Mining in der öffentlichen Verwaltung

Der Beitrag geht nun der Frage nach, welche Möglichkeiten sich durch Process-Mining für die öffentliche Verwaltung ergeben. Dabei muss zunächst die Frage gestellt werden, für welche Art von Prozessen sich die Methode eignet. Aus diesem Grund finden Sie im Folgenden zwei Abschnitte vor, einen, der erst einmal die Voraussetzungen beschreibt, die zur Anwendung der Methode erforderlich sind und einen weiteren, in dem Möglichkeiten aufgezeigt werden, die sich für geeignete Prozesse ergeben.

4.1 Voraussetzungen und Rahmenbedingungen

Die wichtigste Begrenzung der Methode besteht darin, dass sie nur auf Prozesse anwendbar ist, die geeignete Daten produzieren. Doch die im Unternehmenskontext vielfach vorhandenen homogenen, das gesamte Geschäftsfeld abbildenden IT-Verfahren stellen in der Verwaltung eher die Ausnahme dar. Die Prozessvielfalt in der öffentlichen Verwaltung ist immens. Die Prozesse sind zudem oft in vielfacher Hinsicht komplex. So können verschiedene Verwaltungsbereiche und Verwaltungen an einem Prozess beteiligt sein, ohne dass dies geeignet über IT-Schnittstellen nachvollziehbar wäre. Die Prozesse umfassen zudem nicht selten Aktivitäten, die überhaupt nicht datengestützt stattfinden (z. B. Beratungstätigkeiten oder noch nicht digitalisierte Teilaktivitäten) und die IT-Landschaften sind oft sehr heterogen. Organisationsübergreifende, eine einheitliche IT-Systemunterstützung ist nicht zu erwarten. Auch langfristig dürften einzelne Fachverfahren aufgrund ihrer Spezifika von bestimmten, unterschiedlichen IT-Anbietern dominiert werden. Und schließlich sind auch Workflowmanagementsysteme nicht sehr verbreitet.

Damit ist die Anwendung der Methode in der öffentlichen Verwaltung erschwert. Viele Prozesse fallen weg, weil sie nur teilweise informationstechnologisch unterstützt werden. Da ließen sich, wenn überhaupt, nur bestimmte Prozessabschnitte entsprechend analysieren. Bei anderen Prozessen wäre der Aufwand hoch, Daten aus verschiedenen

Systemen für Process-Mining zunächst in eine einheitliche Form zusammenzuführen. Erst danach wäre ein Vorgehen wie folgend beschrieben denkbar.

Sofern die Situation gegeben ist, dass ein IT-Verfahren einen Verwaltungsprozess durchgängig unterstützt, ist zu prüfen, inwieweit die unter (*) angegebenen Voraussetzungen für die Datenlage zutreffen und ob es geeignete Schnittstellen zum Auslesen dieser Daten gibt.

Aufgabenfelder der öffentlichen Verwaltung, die IT-technologisch auf Datenbanksystemen abgebildet werden, sind grundsätzlich mögliche Kandidaten. Auf Datenbanksystemen erfolgen im Rahmen eines Verwaltungsvorganges Einträge oder Änderungen nach mehr oder weniger im Ablauf vorstrukturierten Arbeitsschritten. Solche Vorgänge sollten auf produktiven Datenbanksystemen schon aus Sicherheitsgründen grundsätzlich protokolliert werden, auch wenn diese Mitzeichnungen natürlich Performance-Einbußen mit sich bringen und Speicherplatz kosten. In Verbindung mit dem Transaktionsmanagement auf Datenbanksystemen besitzen diese Vorgänge immer einen Zeitstempel. Bei rechtsrelevanten Verfahren (z. B. beim Erstellen von Bescheiden zu einer Leistungsgewährung, einer Berechtigung oder einer Steuerfestsetzung) dürfte die Protokollierung so eingestellt sein, dass sich auch nachvollziehen lässt, wer für welches Ereignis verantwortlich war.

Allerdings werden aufgrund des Verzögerungseffektes vielfach ausschließlich die Datenbank verändernde Aktivitäten aufgezeichnet. Dies reicht, um bei einem Datenverlust aufbauend auf der letzten Sicherung den aktuellen Datenbestand wieder rekonstruieren zu können. Lesevorgänge, die bei manchen Prozessen durchaus auch von Bedeutung sein können (Stellensuche für einen Arbeitssuchenden während eines Gesprächs), wären dann „unsichtbar“. Dabei können auch solche Aktivitäten inhaltlich und zeitlich äußerst relevant sein, wie das Beispiel zeigt. Aber selbst das Protokollieren von Lesevorgängen ist aus Sicherheitsgründen nicht irrelevant. Personenbezogene oder andere besonders schützenswerte Daten (z. B. Patente, Ausschreibungsgebote etc.) könnten ohne Chance auf Entdeckung unbefugt abgerufen werden. Dabei ist es unerheblich, ob die Daten verschlüsselt abgelegt sind. Sind sie erst einmal in Besitz eines „Interessenten“, ist es lediglich eine Frage der Zeit, sie zu entschlüsseln.

Das Generieren einer Log-Datei, die alle Aktivitäten auf der Datenbank protokolliert, wird von leistungsfähigen Datenbankmanagementsystemen unterstützt. Sollte man die Daten also für Analysezwecke wünschen, so wäre z. B. auch eine auf einen Analysezeitraum befristete komplette Protokollierung denkbar. Die Trennung verschiedener Fälle ist zumindest dann gewährleistet, wenn zur Ablaufsteuerung der Prozessschritte Workflowmanagementsysteme zum Einsatz kommen. Diese steuern die Abarbeitung aller erforderlichen Schritte eines jeden Falls vom Anfang bis zum Ende, indem sie z. B. auf Antragsingänge und fertige Aktivitäten durch Weiterleitung und das Anstoßen weiterer Aktivitäten bzw. durch das Beenden eines Falls reagieren.

Ein Problem stellt sich hinsichtlich des Darstellungsformates der Daten. Die aufgezeichneten Daten sind nicht selten nur binär verfügbar (vgl. z. B. [Ko05], S. 525). Dann wie-

derum benötigt man Funktionen, die eine Umwandlung in Formate realisieren, die von Process-Mining Programmen gelesen werden können. Das sind beispielsweise in dem open source Programm ProM txt⁹-Dateien oder xes¹⁰-Dateien. Dateien im csv¹¹-Format sind ebenfalls geeignet, weil sie sich entsprechend transformieren lassen. Wenn die Daten in einer für diese Programme lesbaren Datei vorliegen, kann mit einer Datenanalyse und dem Generieren von Modellen (play in) sowie mit Replay-Szenarien begonnen werden. Dies wird durch solche Programme unterstützt.

Neben technischen Aspekten sind natürlich auch Ressourcen für die Bereitstellung der Daten erforderlich. Im Allgemeinen ist das IT-Service-Personal auf nicht tägliche Routinen auf den Fachverfahren nicht vorbereitet. Und zum gegenwärtigen Zeitpunkt ist der Betrieb von Process-Mining-Software ebenfalls in der Regel noch nicht gegeben. Ausnahmen können bestehen, wenn Geschäftsprozessmanagementtools entsprechende Komponenten aufweisen. Darüber hinaus kann nicht davon ausgegangen werden, dass die in Kapitel 3 unter (***) erläuterten benötigten Kenntnisse zur Beurteilung erzielter Modell-Ergebnisse bereits im öffentlichen Sektor vorliegen. Sollte die Methode relevant werden, so sicherlich zunächst mit externer Unterstützung.

Und schließlich soll noch darauf hingewiesen werden, dass Auswertungen von Nutzerverhalten, selbst wenn dies anonymisiert erfolgt, i. d. R. das Einschalten des Betriebsrates erfordert. Das gilt allerdings im Kontext von Prozessanalysen generell, nicht nur für die Methode des Process-Minings. Hier kommt hinzu, dass elektronisch Daten ausgelesen werden und sich diese „Kontrolle“ den Blicken der Beschäftigten entzieht.

4.2 Vorteile und Möglichkeiten durch die Auswertungen für die öffentliche Verwaltung

Bei Vorliegen der für Process-Mining geforderten Voraussetzungen gelten die in Kapitel 3 geschilderten Möglichkeiten der Methode generell für Arbeitsabläufe, sei es im Unternehmensumfeld oder in der öffentlichen Verwaltung.

Daher kann man mit Process Mining im Bereich der öffentlichen Verwaltung insbesondere

- ausgehend von realen Vorgängen entsprungene Prozessmodelle für Verwaltungsprozesse generieren,
- belastbares Zahlenmaterial zum Auftreten von Fällen und Fallvarianten erheben,
- belastbares Zahlenmaterial zum zeitlichen Ablauf der einzelnen Aktivitäten der vorkommenden Fälle erheben,
- eindeutige Hinweise auf Engpässe erhalten,

⁹ Dateierweiterung für eine *Textdatei*

¹⁰ Dateierweiterung für eine Datei des Typs *extensible event stream*

¹¹ Dateierweiterung für eine Datei des Typs *comma seperated values*

- eindeutige Hinweise auf Abweichungen vom gewünschten Vorgehen (z. B. Genehmigungen ohne vorherige Prüfungen) erzielen.

Die erhaltenen Daten lassen sich nutzen, um sinnvolle Prozess- oder Software-Veränderungen zu identifizieren und hierauf aufbauend Veränderungsvorhaben zu planen. Sie können auch genutzt werden, um eine bessere Allokation der personellen Ressourcen zu den Aktivitäten im Prozess vorzunehmen.

Im Kontext der zu realisierenden Sicherheitsmaßnahmen kann die Methode auch verwendet werden um Verstöße zu erkennen, insbesondere im Umgang mit Vorgängen, die besonders schützenswert sind, so etwa mit VS-NFD¹²-Vorgängen aber auch zur Erkennung von Datendiebstahl.

5 Ausblick

E-Government verfolgt u. a. das Ziel Verwaltungsvorgänge mit moderner Informationstechnik zu unterstützen und im Idealfall vollautomatisierte Services für Bürger und Unternehmen oder auch intern Verwaltungseinheiten zur Verfügung zu stellen. Die gesetzlichen Grundlagen wurden inzwischen derartig angepasst, dass herkömmliche Antragstellungen mit Unterschrift inzwischen in vielen Fällen durch gleichgestellte elektronische Formen ersetzt werden können. Es gibt allerdings immer noch eine Vielzahl von Hemmnissen, durchgängig automatisierte Lösungen zu offerieren. Ein Problem stellt etwa der Datenaustausch bei komplexeren Vorgängen zwischen diversen Behörden dar. Dennoch ist langfristig davon auszugehen, dass vollautomatische Prozesse eine zunehmende Rolle spielen werden, zumal Bürokratie von Unternehmen als wesentlicher Standortfaktor wahrgenommen wird (vgl. [Kf14], S. 6).

Es ist abzusehen, dass mit fortschreitendem E-Government Techniken des Process-Minings auch für die öffentliche Verwaltung von hoher Relevanz sein werden. Gerade für voll automatisierte Prozesse besitzt diese Methode enormes Potenzial. Unglücklich gestaltete Benutzerschnittstellen, die zu fehlerhaften Einreichungen oder Angaben führen, könnten schnell entdeckt werden. Anfangsschwierigkeiten, die im Abruf geeigneter Daten liegen dürften, sollten bei zunehmender Beliebtheit der Methode verschwinden. Ebenfalls können tatsächliche personelle Engpässe schnell aufgedeckt werden.

Von ergänzendem Interesse wären auch Auswertungen im Vorfeld der Auswahl eines E-Government-Prozesses auf Internetseiten (Tracking¹³, Suchworteingaben). Diese Analysen ähneln dem Process-Mining sehr, auch wenn bei den verschiedenen Suchaktivitäten keine Sachbearbeitung erfolgt. Sie können wichtige Hinweise für die Verbesserung der Auffindbarkeit gewünschter Anliegen geben.

¹² Verschlussache – nur für den Dienstgebrauch

¹³ hierunter soll eine Auswertung des Zugangs auf die Angebotsseite sowie des Bewegungsverhaltens auf den Angebotsseiten verstanden werden.

Literaturverzeichnis

- [Aa14] Aalst, W. M. P.: Geschäftsprozessmodellierung: Die „Killer-Applikation“ für Petrinetze. Informatik Spektrum 37, 3/14, S. 191-198, 2014.
- [Aa11] Aalst, W. M. P.: Process Mining. Discovery, Conformance and Enhancement of Business Processes. Springer, Heidelberg, 2011.
- [Bu15] Burattin, A.: Process Mining Techniques in Business Environments. Theoretical Aspects, Algorithms, Techniques and Open Challenges in Process Mining, Berlin: Springer, 2015.
- [BK13] Bayer, F.; Kühn, H. (Hrsg.): Prozessmanagement für Experten. Impulse für aktuelle und wiederkehrende Themen, Springer Gabler, Berlin.
- [DI09] DIN, Deutsches Institut für Normung e.V, Modell zum prozessorientierten Vorgehen in der öffentlichen Verwaltung. DIN, Berlin, 2009.
- [HK00] Herbst, J; Karagiannis, D.: Integrating Machine Learning and Workflow Management to Support Acquisition and Adaptation of Workflow Models. Intelligent Systems in Accounting, Finance and Management, Volume 9 (2), S. 67-92, 2000.
- [Kf14] KfW, KfW Bankengruppe: KfW- Wettbewerbsindikator 2014, Frankfurt a. M., 2014, Zugriff am 15.04. unter <https://www.kfw.de/PDF/Download-Center/Konzernthemen/Research/PDF-Dokumente-Wettbewerbsindikator/Wettbewerbsinsikator-2014.pdf>
- [Ko05] Kofler M.: MySQL 5. Einführung, Programmierung, Referenz. Addison-Wesley, München, 2005.
- [Lü12] Lück-Schneider, D.: Geschäftsprozessmanagement in der öffentlichen Verwaltung, Schriftenreihe der Forschungsstelle für Betriebsführung und Personalmanagement e. V. (9), Neuhofen/Pf., 2012.

IT-Sourcing in Schweizer Kommunen – Fallstudien aus dem Kanton Bern

Konrad Walser, Anna Barbara Brechbühl und Reto Zraggen¹

Abstract: Verwaltungstätigkeiten sind künftig noch effizienter, qualitativ hochwertiger und kundennäher zu gestalten. Hierfür ist die effiziente Organisation der Informatik in der öffentlichen Verwaltung ein zentraler Faktor. Die zunehmende IT-Komplexität auch ausgehend von E-Government und der Vernetzung legt die Frage nahe, ob IT-Services selbst erbracht oder ausgelagert werden sollen. Der vorliegende Beitrag zeigt den Stand der Ausprägungen des IT-Sourcings anhand von Berner Kommunen. Als Entscheidungsgrundlage, ob IT-Services auszulagern sind, empfiehlt es sich, eine entsprechende Methode zur Beurteilung auszuwählen. Im vorliegenden Fall ist dies der Ansatz der Transaktionskostentheorie. Dieser zeigt auf, unter welchen Bedingungen die Eigenerstellung oder eine Fremdvergabe (Make or Buy) von IT-Services denkbar ist. Daraus lassen sich sogenannte transaktionstheoretische Normstrategien ableiten. Die Entscheidungskriterien lauten wie folgt: Spezifität der Leistung, strategische Bedeutung, Unsicherheit und Häufigkeit der Aufgabenerfüllung. Für diese Kriterien wird eine Bewertung in „hoch“ und „niedrig“ vorgenommen, danach können die Normstrategien abgeleitet werden. Ausgehend davon kann entweder Eigenerstellung oder Outsourcing empfohlen werden. Aus Analyse und Auswertung geht hervor, dass zwei auf Basis der Transaktionskostentheorie aufgestellte Hypothesen nicht bestätigt werden können. Im Kanton Bern gibt es Gemeinden, welche ihre IT-Services aufgrund der niedrigen strategischen Bedeutung und/oder der niedrigen Spezifität auslagern. Es ist kein durchgängiges Muster für die untersuchten Gemeinden ersichtlich. Viele IT-Services werden in den Gemeinden trotz niedrig bewerteten Einflussgrößen selbst erbracht. Ausgehend davon lautet eine Empfehlung: Eine Ist-Analyse bestehender Leistungen durchzuführen und diese in einem weiteren Schritt zu strukturieren, um auslagerbare IT-Services zu identifizieren, etwa basierend auf einem Servicekatalog. Danach kann strukturiert über mögliche Eigenerstellung oder Auslagerung von IT-Services entschieden werden.

Keywords: Informationstechnologie, IT-Outsourcing, Kommunen, Make-or-Buy-Entscheidung, Transaktionskostentheorie, Normstrategien, IT-Servicekatalog

1 Einleitung

1.1 Problemstellung

Die Verwaltungstätigkeit ist künftig noch effizienter, qualitativ hochwertiger und kundennäher zu gestalten. Fakt ist, dass die IT in der öffentlichen Verwaltung ein zentraler Erfolgs- und Standortfaktor geworden ist. E-Government und vernetzte

¹ Berner Fachhochschule, E-Government-Institut, Brückenstrasse 73, CH-3005 Bern, konrad.walser.@bfh.ch; anna.brechbühl@post.ch; reto.zraggen@sbb.ch

Verwaltung stellen zudem wesentliche Herausforderungen dar [Br12]. Was macht eine solche aus? [LM12]. Komplexität und Management der IT erhöhen sich durch die fortschreitende Vernetzung. Aufgrund dessen, dass Erwartungen und Ansprüche von Leistungsbezügern stetig steigen, sind Vertreter der öffentlichen Verwaltung angehalten, Themen wie Zentralisierung und Standardisierung von IT-Leistungen sowie zunehmendes Eingehen von Kooperationen und IT-Sourcing verstärkt zu prüfen und anzugehen [LM12]. Hinsichtlich solcher Vorhaben ist zu prüfen, welche Einzellösungen durch standardisierte Anwendungen zu ersetzen sind und welche IT-Services ev. intern erbracht oder ausgelagert werden können. Um das Ziel zu erreichen, dass das Verwaltungshandeln effizienter und qualitativ hochwertiger gestaltet und umgesetzt wird, sind Führungspersonen angewiesen darauf, die richtigen Entscheidungen bezüglich IT-Sourcing-Vorhaben treffen zu können. Ein denkbarer Ansatz um IT-Sourcing-Vorhaben zu identifizieren und zu beurteilen stellt der Transaktionskostenansatz basierend auf der gleichnamigen Theorie dar, der im vorliegenden Beitrag zur Anwendung gelangt [Wi85].

1.2 Zielsetzung und Abgrenzung

Basierend auf dem aktuellen Stand der Themengebiete Outsourcing sowie Transaktionskostentheorie (TRXT) soll geprüft werden, was der Stand des IT-Sourcings in der Öffentlichen Verwaltung bei Berner Gemeinden ist. Zudem gilt es zu prüfen, ob sich die von den Verfassern aufgestellten Hypothesen bestätigen lassen oder zu falsifizieren sind. In der folgenden Arbeit wird für die Prüfung des IT-Sourcings ausschließlich auf die TRXT gesetzt, dies ausgehend von den Transaktionseigenschaften Spezifität, strategische Bedeutung, Unsicherheit und Häufigkeit. Der Fokus zielt auf größere Gemeinden des Kantons Bern.

1.3 Methodisches Vorgehen und Forschungsdesign

Das Forschungsdesign im vorliegenden Beitrag zur Untersuchung des IT-Sourcing Standes in der Öffentlichen Verwaltung auf kommunaler Ebene im Kanton Bern sieht wie folgt aus: 1. Durchführung Literaturrecherche sowie Präzisierung Forschungsfrage; 2. Definition Forschungsdesign; 3. Datenerhebung mittels strukturierter Interviews; 4. Auswertung der erhobenen Daten; 5. Interpretation und Darstellung daraus gewonnener Erkenntnisse. In einem ersten Schritt wird die Thematik Outsourcing und die TRXT anhand der bestehenden Literatur erläutert sowie die Forschungsfrage aufgrund der verwendeten Literatur präzisiert, aber auch zwei Hypothesen entwickelt. Darauf basierend wird ein adäquates Forschungsdesign gewählt und angewendet. Mittels der genannten zwei Schritte ist so die Frage zu klären, wie die TRXT auf der Gemeindeebene im Kanton Bern verwendbar ist und inwiefern mögliche IT-Sourcing-Vorhaben im Kontext von IT-Services beurteilt werden können. Zur Erhebung der Daten wird ein Interviewleitfaden basierend auf [Gi05] erstellt. Anschließend daran werden Daten mittels strukturierter Interviews erhoben. Die Auswahl der zu interviewenden

Personen erfolgt durch die Bestimmung von fünf der zehn größten Berner Gemeinden (gemäß [BFS13]) anhand der Einwohnerzahl. Im letzten Schritt werden die erhobenen Daten ausgewertet und den theoretischen Grundlagen gegenübergestellt.

2 Theoretische Grundlagen

Outsourcing wird seit den 80er-Jahren als Begriff für die Auslagerung von IT-Prozessen und -Services aus Unternehmen verwendet. Inzwischen, rund 30 Jahre später, gehört das IT-Outsourcing in das Repertoire strategischer und operativer Managementinstrumente aller Branchen, auch der Öffentlichen Verwaltung. Der Begriff setzt sich aus den englischen Begriffen **outside**, **resources** und **using** zusammen [LL11]. [Kn14] definiert IT-Outsourcing als die mittel- und langfristige Auslagerung bisher innerbetrieblich erfüllter IT-Aufgaben an ein oder mehrere rechtlich unabhängige Dienstleistungsunternehmen. Er unterscheidet die Formen Infrastruktur-Outsourcing, Outsourcing der Anwendungsentwicklung, Outsourcing des Systembetriebs und Outsourcing der Softwarewartung. Outsourcing kann verschiedene Formen annehmen. Es können aus Sicht des Unternehmensarchitekturmanagements Geschäfts- oder IT-Prozesse ausgelagert werden, Anwendungen, Services (aus IT-Servicemanagementsicht oder aus Sicht einer Service-orientierten Architektur), Infrastrukturen, die Wartung oder die Softwareentwicklung, etc. (vgl. u.a. [Kn14]). [Or15] zählt zu den Vorteilen von Outsourcing Kostenreduktionen und -optimierungen, Ressourceneinsparungen, die Skalierbarkeit von Leistungen, der Risikotransfer auf einen (externen) Leistungserbringer und der Fokus des Unternehmens auf die Kernkompetenzen. Die Nachteile sieht [Sc11] in der Abhängigkeit vom Leistungserbringer, im höheren Datenschutzrisiko und in den erhöhten Kosten während des Outsourcing Projekts. Weiter oben wurde gezeigt, was Outsourcing bedeutet und erklärt, welche Formen davon existieren. Um ein allfälliges Outsourcing oder eine Teilauslagerung von IT-Services durchzuführen, ist eine Entscheidungsmethodik erwünscht – dies ist im vorliegenden Falle die TRXT.² Die TRXT zeigt auf, unter welchen Bedingungen Eigenerstellung oder Fremdvergabe (Make or Buy) von IT-Services denkbar ist. Da die TRXT abstrakt ist, ist eine Abbildung aller möglichen Konstellationen unmöglich. Um auf dieser Grundlage eine begründete Entscheidung zu treffen, ist eine spezifische Analyse nötig [Ni04]. Die einzelnen Transaktionen bilden bei der Transaktionskostenanalyse die Basis für die grundlegende Untersuchung [Gi05]. Dabei wird unter einer Transaktion die Übertragung von Verfügungsrechten an Gütern und Dienstleistungen verstanden [Ni04]. Die für die Vertragserstellung für die Pflege der

² Es gibt auch andere theoretische Erklärungsmodelle für das IT-Outsourcing. Vgl. hierzu [Di04]. In [Di04] werden u.a. die folgenden Theorieansätze thematisiert, welche zur Erklärung des IT-Outsourcing beigezogen werden können: Agency Theory, Game Theory, Innovation Theories, Power and Politics Theories, Relationship Theories, Resource Theories, Social Exchange Theory, Strategic Management Theories sowie Transaction Cost Theory. Da die Forschung zum IT-Outsourcing in der öffentlichen Verwaltung noch weitgehend am Anfang steht, wurde zunächst mit einer Untersuchung auf Basis der Transaktionskostentheorie begonnen. In weiteren Forschungsschritten ist geplant, auch weitere theoretische Ansätze zur Erklärung des IT-Outsourcings in der öffentlichen Verwaltung zur Anwendung zu bringen.

vertraglichen Beziehung anfallenden Kosten werden als Transaktionskosten bezeichnet [Ni05]. Die Methode basiert auf den zwei Verhaltensannahmen: Begrenzte Rationalität und Potentieller Opportunismus [Di06]. Durch die TRXT soll die effizienteste Organisationsform einer gegebenen Transaktion gefunden werden. Sind die Transaktionskosten zwischen den Beteiligten in einem Outsourcing-Vertrag minimiert, ist die Gültigkeit der TRXT gegeben [Gi05]. An der TRXT wurde zwischenzeitlich auch schon vielfältig Kritik geübt, etwa aus dem Bereich der Wirtschaftswissenschaft, von Soziologen oder Organisationstheoretikern. Sie sehen diese als Konkurrenz der eigens in ihren Disziplinen angebotenen Erklärungsansätze. In der Literatur ist eine systematische Zusammenfassung der Kritik am Transaktionskostenansatz nicht auffindbar [Dö99]. Im Rahmen der TRXT werden die Kosten einer Transaktion vor dem Hintergrund der Verhaltensannahmen durch drei Faktoren beeinflusst [Ni04]: „The principal dimensions with respect to which transactions differ are asset specificity, uncertainty, and frequency. The first is the most important and most distinguishes transaction cost economics from other treatments of economic organizations, but the other two play significant roles“ ([Wi85], S. 52). Die wichtigsten Transaktionseigenschaften sind somit: **Spezifität:** Der Spezifität wird eine große Bedeutung beigemessen. Hierbei können im Laufe oder als Folge einer Transaktion transaktionsspezifische Investitionen auftreten. Investitionen in produktspezifische Fertigungsanlagen oder unternehmensspezifische Qualifikationen werden als Spezifität definiert (z.B. Spezialmaschinen, spezifisches Humankapital, geographische Spezifizierung). Die Beziehung der Akteure ändert sich durch die zunehmende Spezifität der Einsatzfaktoren [Ni04]. Der Wettbewerb wird durch die Spezifität ausgeschaltet und die Transaktionspartner entsprechend voneinander abhängig. Bei einer Beziehungsauflösung verlieren spezifische Investitionen an Wert [Di06]. Ein Faktor ist dann spezifisch, wenn der Ertrag, den er in der nächstbesten Verwendungsmöglichkeit einbringt, geringer ausfällt als in der jetzigen Verwendungsrichtung [An15]. **Unsicherheit:** Verträge bleiben unvollständig, es treten Fehlanpassungen auf [Di06]. **Häufigkeit:** Durch die ansteigende Anzahl identischer Transaktionen besteht die Möglichkeit zur Realisierung von (Fix-)Kostendegressions-, Skalen- und Synergieeffekten. Durch die Nutzung dieser Potenziale sinken die Produktions- und Transaktionskosten je Transaktion mit der Zunahme identischer Transaktionen [Ni04]. Tendenziell steigt mit der Häufigkeit der Trend zur Eigenfertigung, mit der Begründung, dass Kapazitäten zur Eigenfertigung meist erst ab einer bestimmten Größenordnung rentieren (Economies of Scale) [Gi05]. Die strategische Bedeutung, welche unter den drei Faktoren nicht explizit aufgeführt ist, steht in engem Zusammenhang mit der Spezifität einer Leistung (Vgl. Tab. 1 im Anhang).

Eine hohe Spezifität einer Leistung lässt auf eine hohe strategische Bedeutung rückschließen. Dabei handelt es sich meist um Kernkompetenzen. Solche Leistungen sind für die Differenzierung gegenüber der Konkurrenz von hoher Bedeutung und aus diesem Grund unternehmensintern zu organisieren und nicht auszulagern. Strategisch bedeutsame Leistungen sind zudem häufig mit schützenswertem Know-how gekoppelt [Gi05]. In der folgenden Tabelle wird die Wirkung der Ausprägungen einzelner Einflussgrößen gemäß [Pi91] präsentiert. Basierend auf den in der obigen aufgeführten Ausprägungen der Einflussgrößen auf die Transaktionskosten können sogenannte Norm-

strategien abgeleitet werden. Nach Empfehlung von Picot soll dazu die Portfolio-Technik angewendet werden, welche besagt, dass von den vier Einflussgrößen deren zwei auszuwählen sind. Eine weitere Empfehlung ist, dass dazu die zwei wesentlichen Einflussgrößen (strategische Bedeutung und Spezifität) auszuwählen sind [Gi05]. Die Normstrategie Outsourcing ergibt sich, wenn die Ausprägung der beiden Einflussgrößen niedrig ist. Sind die Ausprägungen beider Einflussgrößen hoch, resultiert die Normstrategie Eigenerstellung. Fallen die Ausprägungen unterschiedlich aus, ist die daraus resultierende Normstrategie unter Bezug auf die beiden unterstützenden Einflussgrößen (Unsicherheit und Häufigkeit) zu bestimmen [Gi05] (Vgl. Tab. 2 im Anhang). Die daraus resultierenden Normstrategien (N1 – N8) sind in der Tabelle 3 im Anhang ersichtlich [Gi05].

3 Forschungsfragen und Methodisches Vorgehen (Forschungsdesign)

Seit den 80er-Jahren werden, im Zusammenhang mit der IT, Funktionen von Unternehmen an externe Unternehmen im In- und Ausland ausgelagert [LL11]. Dabei kann Outsourcing verschiedene Formen annehmen [RK11]. Zu den Vorteilen von Outsourcing gehören gemäß [Or15]: Kostenreduktionen und -optimierungen, Ressourceneinsparungen, die Skalierbarkeit von Leistungen, der Risikotransfer auf einen externen Leistungserbringer und der Fokus des Unternehmens auf die Kernkompetenzen. Mögliche Nachteile sind die Abhängigkeit vom Leistungserbringer, das höhere Datenschutzrisiko und die erhöhten Kosten während des Outsourcing-Projekts [Sc11]. Es lässt sich festhalten, dass auf der kommunalen Ebene im Kanton Bern die Mehrheit der Gemeinden die Mehrheit ihrer IT-Services selbst erbringt. Lediglich ein paar wenige Gemeinden lagern (eine Mehrheit oder alle ihre) IT-Services aus [WB10]. Gemäß [Gi05] soll durch die TRXT die effizienteste Organisationsform einer gegebenen Transaktion gefunden werden. Basierend auf den Ausprägungen der thematisierten Einflussgrößen (Spezifität, strategische Bedeutung, Unsicherheit und Häufigkeit) auf die Transaktionskosten können transaktionskostentheoretische Normstrategien abgeleitet werden.

Im vorliegenden Beitrag wird mittels qualitativen Methoden aufgezeigt, ob diese theoretischen Grundlagen im Themenbereich Outsourcing und der TRXT des IT-Sourcings in der Öffentlichen Verwaltung auf der kommunalen Ebene im Kanton Bern zutreffen oder nicht. Die Forschungslücke liegt dabei darin, dass zwar bereits Untersuchungen durchgeführt wurden, für den spezifischen Fall der kommunalen Ebene im Kanton Bern aber keine expliziten Forschungsergebnisse zum IT-Outsourcing vorliegen. Der Beitrag soll zeigen, wie der aktuelle Stand bezüglich IT-Sourcing in der Öffentlichen Verwaltung auf der kommunalen Ebene im Kanton Bern ist, und ob die Feststellung, dass Gemeinden die IT-Services mehrheitlich selbst erbringen, zutrifft oder nicht.

Die folgende Forschungsfrage ist somit im vorliegenden Beitrag zu beantworten. Was ist der Stand des IT-Sourcings in der Öffentlichen Verwaltung auf der kommunalen Ebene

im Kanton Bern? Ausgehend davon werden basierend auf der TRXT zwei Hypothesen abgeleitet. H1: Da die IT für eine Gemeinde im Kanton Bern keine hohe strategische Bedeutung aufweist, werden IT-Services ausgelagert. H2: Da die IT für eine Gemeinde im Kanton Bern keine Spezifität aufweist, werden IT-Services ausgelagert.

4 Fallstudien, Analyse und Auswertung

Insgesamt wurden Interviews mit Mitarbeitern von vier unterschiedlichen Gemeinden durchgeführt, wovon ein IT-Verantwortlicher für zwei Gemeinden zuständig ist. Der Einfachheit halber wird diese in der Analyse und Auswertung als eine Gemeinde (Gemeinde D) aufgeführt.

4.1 Gemeinde A

Kurzbeschrieb der Gemeinde A: Die IT der Gemeinde A ist gemäß aktuellem Organigramm der Finanzdirektion unterstellt, welche durch die Stadtpräsidentin, resp. dem Gemeinderat geführt wird. Dabei war die IT der Gemeinde A vor ein paar Jahren noch dem Stadtpräsidenten direkt unterstellt. Die Stadtverwaltung der Gemeinde A beschäftigt derzeit zirka 200 bis 250 Personen. Dabei arbeitet ein großer Teil der Beschäftigten außerhalb eines Büros (z.B. in der Baudirektion) und ist entsprechend nicht direkt auf die Unterstützung der Informatik angewiesen. D.h., dass ungefähr 200 Personen Zugriff haben auf die Infrastruktur der Informatik. Das Team der Informatik besteht insgesamt aus drei Personen, wovon eine Person für die Führung sowie sämtliche Managementaufgaben verantwortlich ist. Die zwei weiteren Angestellten sind für die technischen Belange und den Support zuständig. Der Zuständigkeitsbereich der IT reicht von der Beschaffung bis hin zur Entsorgung der IT-Infrastruktur. **Analyse der Gemeinde A:** Aufgrund der zur Verfügung stehenden Ressourcen werden in der IT der Gemeinde A Leistungen von außen bezogen, da mit dem internen Know-how derzeit nicht alle Tätigkeiten abgedeckt werden können. Dabei werden Leistungen nicht unterschieden, d.h. in der IT der Gemeinde A sind alle Leistungen grundsätzlich IT-Services. Vier ausgelagerte IT-Services konnten bei der IT der Gemeinde A explizit identifiziert werden: „Aufsetzen der Server“, „Netzwerk“, „Beschaffung von Verbrauchsmaterial“ und „Website“, welche in der Tabelle 4 im Anhang aufgeführt sind. Der Grund dafür, dass das „Aufsetzen der Server“ (Provisioning Server, BackOffice Server sowie Terminal Server) ausgelagert wurde, ist, dass der Aufbau von spezifischem Know-how intern nicht nötig ist. Beim IT-Service „Netzwerk“ (WLAN auf Switches, Definition Routing etc.) sind Sicherheitsüberlegungen die treibende Kraft, den IT-Service auszulagern. Zudem wurde auch hier der interne Aufbau von spezifischem Know-how als nicht notwendig angesehen. Die „Beschaffung von Verbrauchsmaterial“ (Druckertoner) ist ein automatisierter Prozess, welcher bei einem bestimmten Tonerstand eine Bestellung beim Lieferanten auslöst. Der Lieferant ist entsprechend für den IT-Service zuständig. Als vierter und somit letzter IT-Service der Gemeinde A, wurde die

„Website“ identifiziert. Aus Gründen der erforderlichen 7x24h Abdeckung wurde auch dieser IT-Service ausgelagert. In der folgenden Tabelle ist die von der Autorschaft vorgenommene Wertung der Einflussgrößen (hoch, niedrig) bezüglich der IT-Services der Gemeinde A gemäß der TRXT abgebildet (Vgl. Tab. 4 im Anhang).

Restliche Leistungen oder IT-Services werden von der Gemeinde A selber erbracht. Weitere IT-Sourcing-Vorhaben sowie ein mögliches Insourcing von IT-Services sind derzeit nicht geplant. **Auswertung der Gemeinde A:** Der Vergleich mit den weiter oben beschriebenen Normstrategien zeigt, dass die Gemeinde A alle ihre IT-Services gemäß der TRXT abgeleiteten Normstrategien ausgelagert hat. Aufgrund der niedrigen Spezifität sowie der niedrigen strategischen Bedeutung ist der Entscheid, sämtliche IT-Services auszulagern, richtig und ein Outsourcing in diesem Fall korrekt (Vgl. Tabelle 5 im Anhang).

4.2 Gemeinde B

Kurzbeschreibung der Gemeinde B: Die IT der Gemeinde B war vor dem Jahr 2001 dem Departement Finanzen/Steuern unterstellt. Diese Organisationsstruktur wechselte etwas später und die IT wurde in der Präsidialabteilung angesiedelt. Ausgehend von einer personellen Änderung im Sommer 2015 wurde die IT wieder dem Departement Finanzen/Steuern unterstellt. Die Gemeinde B beschäftigt, mit steigender Tendenz, ca. 180 Mitarbeitende. Diese Arbeitsplätze werden durch die IT mit einem Team von drei Mitarbeitenden und einem Lernenden betreut. Die drei Personen und der Lernende betreuen die gesamte IT-Infrastruktur inklusive einer Hotline, welche während den Schalteröffnungszeiten erreichbar ist. Eine Person hat zusätzlich die Leitung der Informatik inne. Da die IT in der Gemeinde B keine Durchsetzungsgewalt hat, ist es oft schwierig Entscheide durchzusetzen. Dies führt oft zu Unstimmigkeiten, da die Mitarbeitenden die Entscheide nicht mittragen, welche die IT umsetzen muss. **Analyse der Gemeinde B:** Für die Gemeinde B ist ein IT-Service eine Dienstleistung, die von der IT den Benutzern zur Verfügung gestellt wird. In der Gemeinde B werden drei IT-Services unterschieden: „Telefonie“, „Druckerei“ und „Arbeitsplatz“. Der IT-Service „Telefonie“ beinhaltet alle analogen und digitalen Telefonapparate sowie das Telefonnetz. Die „Druckerei“ beinhaltet gemietete Multifunktionsgeräte, bei welchen der Provider das Papier und den Toner automatisch nachliefert. Der IT-Service „Arbeitsplatz“ beinhaltet alles, was benötigt wird, um einen Arbeitsplatz betreiben zu können. Dazu gehören der Server, das Netzwerk, die Hardware und die Hotline. Von den drei IT-Services hat die Gemeinde B nur die „Druckerei“ ausgelagert. Gründe für ein Outsourcing dieses IT-Services waren vor allem die tieferen Kosten. Ausgelagert werden bei der Gemeinde B ebenfalls diejenigen Arbeiten, welche spezifisches Know-how voraussetzen. Dies kommt oft bei Projekten vor, da das Wissen in der IT-Abteilung nicht vorhanden ist und aufgrund der Einmaligkeit der Projekte auch nicht aufgebaut werden soll. Auch bei Know-how Problemen (3rd Level Support) können Anfragen an die Provider oder Lieferanten abgegeben werden. Da es sich bei diesen beiden geschilderten Auslagerungen aber gemäß Definition der Gemeinde B nicht um IT-Services handelt,

wurden diese nicht in die Tabelle 6 im Anhang aufgenommen. In dieser ist die von den Verfassern vorgenommene Wertung der Einflussgrößen (hoch, niedrig) bezüglich der IT-Services der Gemeinde B, gemäß der TRXT ersichtlich (Vgl. Tabelle 6 im Anhang).

Der IT-Service „Arbeitsplatz“ wurde in der Gemeinde B mit einer hohen Spezifität bewertet, da der Service viele Fachanwendungen beinhaltet. Diese werden aufgrund der Eigenschaften einer Fachanwendung, insbesondere da diese ein spezifisches Know-how erfordern, als spezifisch eingestuft. Für die Gemeinde B ist momentan kein weiteres Outsourcing von bestehenden oder ein Insourcing von bereits ausgelagerten IT-Services geplant (Vgl. Tabelle 7 im Anhang).

Die **Auswertung der Gemeinde B** zeigt, dass diese Gemeinde nicht alle IT-Services gemäß der von der TRXT abgeleiteten Normstrategien ausgelagert hat. Richtig ist, gemäß den Normstrategien, dass der Service „Druckerei“ aufgrund der niedrigen Spezifität und der niedrigen strategischen Bedeutung ausgelagert wird. Die „Telefonie“ sollte aber aufgrund der niedrigen Spezifität und der geringen strategischen Relevanz ausgelagert werden. Auch den IT-Service „Arbeitsplatz“ sollte gemäß Normstrategie 4 ausgelagert werden. Die Gemeinde B lagert diese IT-Services nicht aus, da sie der Meinung ist, ein Outsourcing verursache höhere Kosten. Auch will man möglichst viel selbst betreiben um den Mitarbeitenden eine interessante und abwechslungsreiche Arbeit zu bieten. Zudem kann so bei Problemen schnell und direkt reagiert werden.

4.3 Gemeinde C

Kurzbeschreibung der Gemeinde C: Die IT der Gemeinde C ist in der Direktion für Finanzen/Personal/Informatik (FPI) angesiedelt und umfasst 66 Mitarbeitende, was 55 Vollzeitäquivalenten entspricht. Als zentraler Informatik-Dienstleister der Stadtverwaltung und der Volksschulen betreuen die IT-Mitarbeitenden rund 3'500 Mitarbeitende mit insgesamt 2'000 PC-Arbeitsplätzen in der Verwaltung und 2'100 PC-Geräte in den Schulen. Zu den Aufgaben der IT gehören zusammengefasst die Planung, Installation, der Betrieb, die Wartung und die Betreuung von IKT-Infrastrukturen. Dies umfasst den Betrieb der Rechenzentren, die gesamte Systemtechnik (Server, Netzwerk, Storage, Backup und Telefonie), die Betreuung von ca. 300 Fachanwendungen und das Service Desk, welches als Single Point of Contact (SPOC) für Kundenanfragen dient. Die IT deckt somit das gesamte Spektrum ab, das in der Verwaltung genutzt wird.

Analyse der Gemeinde C: Die Gemeinde C definiert ein IT-Service als eine Dienstleistung, welche sie weiter in Basis- und Zusatzdienstleistungen unterscheiden. Basisdienstleistungen sind betriebsnotwendig, Zusatzdienstleistungen hingegen sind je nach Bedarf unterschiedlich und nicht Teil der Grundversorgung. Diese werden entsprechend separat verrechnet. Die Gemeinde C beschreibt ihre Dienstleistungen in einem Servicekatalog, den die Leistungsbezüger erhalten. In diesem sind alle Dienstleistungen der IT, die Zusatzdienstleistungen, die Leistungsabgrenzungen, die Mitwirkungspflicht der Leistungsbezüger sowie die Verrechnung beschrieben. In der Gemeinde C werden insgesamt sieben IT-Services unterschieden, welche größtenteils selbst erbracht werden:

der „Basis-Arbeitsplatz“ beinhaltet einen standardisierten Arbeitsplatz, der mit dem städtischen Netz verbunden ist und alle Kommunikationsdienste und Print-Services nutzen kann. Der „File-Service“ enthält den Zugriff auf ein gesichertes persönliches Laufwerk und weitere Speichermöglichkeiten. Der „Datenbank-Service“ beschreibt den Aufbau, Betrieb und Unterhalt von Datenbank-Management-Systemen. Der „Application-Services“ umfasst die Betreuung von Fachanwendungen. Der „Remote Service“ beinhaltet den gesicherten Zugriff auf Ressourcen innerhalb des Stadtnetzes über das Internet. Der Service „Schulungen“ beinhaltet ein Standard- oder individuelles Coaching und der Service „Telefonie“ beinhaltet fixe Apparate und die Mobilkommunikation. Der IT-Service „Basis-Arbeitsplatz“ lässt sich in zwölf weitere Dienstleistungen unterteilen: Erstinstallation, IT-Servicezentrum, Asset Management (Inventar), Lizenz Management, Release Management (Client), Virusprotection-Services, IT-Netz, Mail-Services, Web-Services, Print-Services, Domain-Services sowie die Citrix-Farm. Von diesen zwölf Dienstleistungen wurden drei ausgelagert: die Wartung von Multifunktionsgeräten und Drucker (Print Services), die Bewirtschaftung des Schulnetzes und das Netz von/zu einzelnen Außenstationen (IT-Netz) und die Softwarepaketierung (Release Management). Zur einfacheren Darstellung und Bewertung wurden diese IT-Services in der Tabelle 8 im Anhang separat geführt. Auch die Wartung von Fachanwendungen wird zum Teil von den Lieferanten selbst übernommen. Teilweise werden auch Schulungen ausgelagert und der Internetauftritt der Gemeinde C wurde ebenfalls an einen externen Dienstleister übertragen. Beweggründe für das Outsourcing dieser Services sind einerseits die Wirtschaftlichkeit und interne Ressourcenengpässe, andererseits das fehlende interne Know-how zur Erbringung der Leistungen. Die folgende Tabelle visualisiert die von den Verfassern vorgenommene Wertung der Einflussgrößen (hoch, niedrig) bezüglich der IT-Services der Gemeinde C gemäß TRXT (vgl. Tab. 8 im Anhang).

Für die Gemeinde C wurden vor allem die IT-Services, welche in Zusammenhang mit den spezifischen Fachanwendungen stehen, mit einer hohen Spezifität bewertet. Dies, da es sich bei Fachanwendungen meist nicht um Standardsoftware, sondern um gemeindespezifische Software handelt. Für die Gemeinde C ist momentan kein weiteres Outsourcing von bestehenden oder ein Insourcing von bereits ausgelagerten IT-Services geplant.

Auswertung der Gemeinde C: Die Auswertung der Gemeinde C zeigt, dass die heutigen IT-Services, welche bereits durch einen Dritten erbracht werden, anhand der Bewertung der TRXT ausgelagert werden. Die Bewertung in der Tabelle 9 zeigt aber auch, dass generell alle Services der Gemeinde C aufgrund der TRXT ausgelagert werden sollten. Dies, da die Spezifität und die strategische Bedeutung der IT-Services durch die Verfasser meist als niedrig eingestuft wurde. Nur IT-Services im Zusammenhang mit spezifischen Fachanwendungen der Gemeinden wurde die Spezifität hoch zugeordnet (Vgl. Tabelle 9 im Anhang).

4.4 Gemeinde D

Kurzbeschreibung der Gemeinde D: Die IT der Gemeinde D ist dem Departement „Umwelt und Betriebe“ untergeordnet. Sie ist derzeit mit elf Personen und zwei Lernenden (Total: 950% Stellenprozent exkl. Lernende) besetzt. Das Team der IT ist in zwei Gruppen aufgeteilt (Systemadministration und Applikationsbetreuung). Dazu kommen administrative Tätigkeiten sowie die Leitung des Teams, welche durch eine Vollzeitstelle abgedeckt wird. Grundsätzlich ist die IT der Gemeinde D für sämtliche Belange verantwortlich, d.h. vom dezentralen Arbeitsplatz bei den Usern bis hin zu den Servern im Rechenzentrum. Dazu gehören auch ungefähr 120 Applikationen, welche betrieben werden müssen. Im Gegensatz zu den Gemeinden A bis C ist die Gemeinde D bezüglich IT gleichzeitig für zwei Gemeinden verantwortlich. **Analyse der Gemeinde D:** Die IT der Gemeinde D erbringt außer der Softwareentwicklung sämtliche Leistungen selber. Dabei sind alle Leistungen grundsätzlich als IT-Services zu betrachten. Dabei ist die IT der Gemeinde D nicht ISO/IEC 20000 zertifiziert und besitzt keinerlei Zertifikate. Sie arbeitet nach „best practice“-Ansätzen. Ausgerollte Hardware gehört der IT Gemeinde D. Leistungen werden zu einem definierten Preis erbracht und intern nicht 1:1 verrechnet. Ausnahmen sind die gebührenfinanzierten Abteilungen (Kehricht, Abwasser, Wasser, Feuerwehr), denen die IT-Arbeitsplätze zu Vollkosten verrechnet werden. Die IT der Gemeinde D lagert außer den von ihr definierten IT-Services „Multifunktionsgeräte“ und „Website Hosting“ derzeit keine weiteren IT-Services aus. Sie fungiert im Grunde genommen als Sourcingpartner für die Gemeinden D1 und D2. Aus dem Grund, dass die IT der Gemeinde D gleich zwei Gemeinden repräsentiert, sind in diesem Fall zwei Szenarien denkbar. Zum einen ist die IT der Gemeinde D als integrierter und zentraler Dienstleister zu analysieren (aktuell), zum andern könnte die IT der Gemeinde D als ausgelagerter Sourcingpartner für die Gemeinden D1 und D2 betrachtet und analysiert werden. **Szenario 1: IT als integrierter Dienstleister der Gemeinde D:** Es folgt die Analyse, bei welcher die IT der Gemeinde D als integrierter Dienstleister betrachtet wird (Vgl. Tabelle 10 im Anhang).

Restliche Leistungen resp. IT-Services werden von der IT der Gemeinde D selber erbracht. Weitere IT-Sourcing-Vorhaben sowie ein mögliches Insourcing von IT-Services sind derzeit nicht geplant. **Szenario 2: IT als ausgelagerter Sourcingpartner für die Gemeinden D1 und D2:** Nachfolgend wird die Analyse präsentiert, bei welcher die IT der Gemeinde D als ausgelagerter Sourcingpartner für die Gemeinden D1 und D2 betrachtet wird. Die in Tabelle 11 aufgeführten IT-Services wurden von den Verfassern basierend auf dem durchgeführten Interview zusammengestellt (Vgl. Tabelle 11 im Anhang).

Auswertung der Gemeinde D: Der Vergleich mit denen im Kapitel 2.2.2 beschriebenen Normstrategien zeigt, dass die IT Gemeinde D alle ihre IT-Services basierend auf der aus der TRXT abgeleiteten Normstrategien ausgelagert hat (Vgl. Tabelle 12 im Anhang).

Aufgrund der niedrigen Spezifität sowie der niedrigen strategischen Bedeutung ist der Entscheid, sämtliche IT-Services auszulagern, richtig und ein Outsourcing in der Betrachtung der IT der Gemeinde D als integrierter Dienstleister (Szenario 1) korrekt (Vgl. Tabelle 13 im Anhang).

Im Fall der zweiten Auswertung, in welcher die IT der Gemeinde D als ausgelagerter Sourcingpartner betrachtet wird (Szenario 2), fehlen den Verfassern detaillierte Angaben zu den jeweiligen Leistungen. Eine mögliche Zusammenstellung von IT-Services ist in der Tabelle 13 im Anhang enthalten. Der Vergleich mit den weiter oben definierten Normstrategien zeigt, dass nicht alle IT-Services gemäß der TRXT ausgelagert sind. Aufgrund der niedrigen Spezifität sowie der niedrigen strategischen Bedeutung sollten die IT-Services „Netzwerk (LAN, WAN, Servernetzwerk)“, „Infrastruktur“, „Storage“, „Datenbank Server“ und „File Server“ ausgelagert werden. Dasselbe gilt für den IT-Service „Basis-Arbeitsplatz (dezentrale Clients, Thin Clients, PC, Desktop, Laptops, Tablets)“, welcher trotz der hohen Spezifität aufgrund der TRXT, ausgelagert werden sollte.

4.5 Gesamtauswertung

Die nachfolgende Gesamtauswertung über den Ist-Zustand sämtlicher Gemeinden zeigt, dass von den total 21 erwähnten und identifizierten IT-Services lediglich rund ein Drittel in Eigenerstellung erbracht werden. Die restlichen rund zwei Drittel werden bereits heute durch einen externen Leistungserbringer erbracht. Dabei gilt es festzuhalten, dass die Vergleichbarkeit aufgrund der unterschiedlichen Auslegung und Identifikation von IT-Services seitens der Gemeinden grundsätzlich nicht gegeben ist. Zudem wurden Leistungen, welche nicht explizit als IT-Services erwähnt oder identifiziert wurden, nicht abgebildet. Aufgrund dessen, dass die IT der Gemeinde D heute als integrierter Dienstleister zu betrachten ist, wurde das Szenario 2 der Gemeinde D in der Gesamtauswertung nicht miteinbezogen. Basierend auf der TRXT und der danach abgeleiteten transaktionstheoretischen Normstrategien ist ein Outsourcing sämtlicher identifizierter IT-Services aller befragten Gemeinden im Kanton Bern anzustreben. Dies geht aus der in Tabelle 14 im Anhang ersichtlichen Zusammenstellung der gesamten identifizierten IT-Services der Gemeinden A bis D hervor.

5 Empfehlungen und Kritische Würdigung der Arbeiten

Fast schon im Sinne eines Ausblicks zur weitergehenden Forschung zu IT-Services in Kommunen empfehlen die Verfasser, basierend auf den gewonnenen Erkenntnissen aus dem vorherigen Kapitel (primär die ungleiche oder Nicht-Standardisierung und -Strukturierung von IT-Services, was deren Vergleichbarkeit massiv erschwert), eine Ist-Analyse der bestehenden IT-Services durchzuführen und diese in einem weiteren Schritt zu strukturieren. Hierbei stellt die Information Technology Infrastructure Library ITIL, welche eine Sammlung von Best Practices in einer Reihe von Publikationen zur

Umsetzung eines IT-Service-Managements (ITSM) bereitstellt, eine Hilfe dar dazu, wie IT-Services identifiziert und strukturiert werden können.³ Gemäß ITIL kann ein sogenannter Servicekatalog erstellt werden [Hu11], welcher diejenigen Services enthält, welche aktiv sind oder für bestehende oder zukünftige Kunden genehmigt wurden. Der Servicekatalog dient der Kanalisierung von Serviceaufträgen und der Definition von Service Level Requirements und definiert die Leit- und Richtlinien sowie die Verantwortlichkeiten und Ziele der Services. In diesem Zusammenhang kann dann auch strukturierter über die Auslagerung oder Eigenerstellung entschieden werden. Der Servicekatalog enthält auch Preisinformationen, Service Level Verpflichtungen und die Bedingungen für die Servicebereitstellung [OGC07]. Ist der Servicekatalog erstellt, kann in einem nächsten Schritt basierend auf den dannzumal bestehenden Fakten fundierter über Eigenerstellung oder ein Outsourcing von IT-Services entschieden werden, um gemäß [Or15] allfällige Vorteile des Outsourcings ausnutzen zu können. Um die „richtigen“ Entscheidung herbeizuführen, können die in der Tabelle 3 im Anhang enthaltenen transaktionstheoretischen Normstrategien (N1 – N8) gemäß [Gi05] herbeigezogen werden. Eine Festlegung einer IT-Strategie oder einer Sourcing-Strategie könnte ebenfalls helfen, Entscheidungen zu vereinfachen.

Aus der Analyse und Auswertung geht überdies hervor, dass die von den Verfassern formulierten Hypothesen beide nicht bestätigt werden können. Ausgehend von den fünf analysierten Gemeinden im Kanton Bern gibt es Gemeinden, welche ihre IT-Services aufgrund der niedrigen strategischen Bedeutung und/oder der niedrigen Spezifität auslagern, jedoch ist kein durchgängiges Muster für alle untersuchten Gemeinden ersichtlich. Viele IT-Services werden in den Gemeinden trotz niedrig bewerteten Einflussgrößen der TRXT noch selbst erbracht. Der Fokus beim Entscheid, ob ein IT-Service ausgelagert werden soll oder nicht, richtet sich meist weder auf die Strategie noch auf die Spezifität, sondern viel mehr auf die Kosten, das vorhandene Wissen und die personellen Ressourcen. Sofern internes Know-how sowie die dazugehörigen Ressourcen vorhanden sind, sehen sich die untersuchten Gemeinden im Kanton Bern nicht gezwungen, weitere IT-Services auszulagern. Laut der Aussage eines befragten Gemeindemitarbeiters sollen IT-Mitarbeitende, welche in einer Gemeinde angestellt sind, eine interessante und abwechslungsreiche Arbeit haben, was bei einem Outsourcing nicht mehr gegeben sei. Zudem können gemäß Aussagen aus den Interviews bei einem Outsourcing die Leistungen und der Support für die Kunden nicht mehr im gleichen Ausmaß gewährleistet werden. Als wie erwähnt problematischer Punkt an der Ist-Situation ist festzuhalten, dass nur wenige Gemeinden ihre IT-Services systematisch identifiziert und strukturiert haben und mittels Servicekatalog oder Ähnlichem vorzeigen können. Dies erschwert die grundlegende Entscheidung – dannzumal allenfalls auch fundiert über ein Benchmarking – ob es zweckmäßig ist, IT-Services auszulagern. Basierend auf der TRXT und der danach abgeleiteten transaktionstheoretischen Normstrategien könnten

³ Vgl. hierzu z.B. auch den Begriff Service Archetypes, welcher von [IN07] im ITIL-Band Service Strategy erwähnt wird. Eine weitere Möglichkeit wäre, vom Ansatz des Capabilities Management auszugehen und über Capabilities Services zu eruieren. Möglicherweise käme man dabei auf ähnliche Lösungen wie über den Ansatz des Service-Archetype-Managements.

die Gemeinden im Kanton Bern zukünftig mehr IT-Services auslagern und damit ggf. von den Vorteilen eines Outsourcings profitieren. Die Verfasser sind der Meinung, dass die Vergleichbarkeit aufgrund der unterschiedlichen Auslegung und Identifikation von IT-Services der Gemeinden grundsätzlich nur bedingt gegeben ist. Ein Raster mit standardisierten IT-Services hätte geholfen, die von den Gemeinden erwähnten Leistungen zu strukturieren, zu bewerten und anschließend zu vergleichen. Dies wird in einer weiteren Forschungsarbeit an die Hand genommen. Im Sinne einer Limitation des Forschungsvorhabens ist anzumerken, dass Aussagen auf die einbezogenen fünf Gemeinden aufgrund der kleinen Anzahl befragter Gemeinden nicht verallgemeinert werden können. Neben größeren Gemeinden müssten auch kleine und mittlere Gemeinden befragt werden, um systematischere und möglicherweise repräsentativere Aussagen machen zu können.

Literaturverzeichnis

- [An15] Anonymous (2015): Transaktionsdimensionen. Online: www.wiwi.uni-muenster.de/06/toplinks/glossar/glossar.php?begriff=66 (12.05.2015).
- [Br12] Bundesrat (2012): Cloud Computing Strategie der Schweizer Behörden 2012-2020. Online: www.egovernment.ch/umsetzung/00881/00882/index.html?lang=de (24.05.2015).
- [CKL09] Csoka, J.; Knolmayer, G.; Loosli, G. (2009): Informatik. Eigenleistung oder Fremd-bezug? Schweizer Gemeinde. 2/09, S. 16-17.
- [Di04] Dibbern, J.: Information Systems Outsourcing: A Survey and Analysis of Literature. In: The Data Base for Advances in Information Systems 35 (2004) 4, S. 6-102.
- [Di06] Dietl, H. (2006): Transaktionskostentheorie. Online: www.business.uzh.ch/professorships/som/stu/Teaching/fs13/BA/LM/Transaktionskostentheorie.pdf (28.04.2015).
- [Dö99] Döring, H. (1999): Verbreitete Kritik am Transaktionskostenansatz. Online: webdoc.sub.gwdg.de/diss/1999/doering/ (05.05.2015).
- [Gi05] Gimpert, M. (2005): Möglichkeiten und Grenzen des CRM- Prozessoutsourcings im schweizerischen Telekommunikationsmarkt. Lizentiatsarbeit, Bern: Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität Bern.
- [Hu11] Hunnebeck, L. (2011): Service Design, TSO, Norwich.
- [IN07] Iqbal, M.; Nieves, M. (2007): Service Strategy (ITIL V3), TSO, Norwich.
- [Kn14] Knolmayer, G. (2014): Outsourcing. Online: www.enzyklopaedie-der-wirtschaftsinformatik.de/wi-enzyklopaedie/lexikon/uebergreifendes/Globalisierung/Outsourcing/index.html?searchterm=it-outsourcing (14.04.2015).
- [LL11] Lewinski-Reuter, V.; Lüddemann, S. (2011): Glossar Kulturmanagement. Wiesbaden: VS Verlag für Sozialwissenschaften.
- [LM12] Loher, D.; Mégroz, M. (2012): Spotlight. IT-Trends in der öffentlichen Verwaltung. Die Volkswirtschaft Das Magazin Für Wirtschaftspolitik, 3-2012, S. 50–51. Online:

dievolkswirtschaft.ch/content/uploads/2012/03/13D_Loher.pdf (25.05.2015).

- [Ni04] Nienhüser, W.; Jans, M.; Köckeritz, M. (2004): Grundbegriffe und Grundideen der Transaktionskostentheorie. Online: www.uni-due.de/apo/GrundbegriffeTAKT.pdf (25.04.2015).
- [OGC07] Office of Government Commerce (2007): Service Strategy. TSO, Norwich.
- [Or15] Ortner, G. (2015): Projektmanagement-Outsourcing. Chancen und Grenzen erkennen, Heidelberg: Springer-Verlag.
- [Pi91] Picot, A. (1991): Ein neuer Ansatz zur Gestaltung der Leistungstiefe. Zeitschrift für betriebswirtschaftliche Forschung, 43, S. 336-357.
- [RK11] Reed, A. H.; Knight, L.V. (2011): Major Virtual Project Risk Factors. Online: <http://jitm.ubalt.edu/XXII-3/> (20.4.2015).
- [Sc11] Schäfer, S. (2011): IT Outsourcing Leitfaden. Von der Idee bis zur Umsetzung, Norderstedt: Books on Demand GmbH.
- [Si14] Siepmann, F. (2014): Managing Risk and Security in Outsourcing IT Services. Onshore, Offshore and the Cloud, Boca Raton: Taylor & Francis Group.
- [VBK10] Von Stetten, A.; Beimborn, D.; Kuznetsova, E.; Moos, B. (2010): Proceedings of the 43rd Hawaii International Conference on System Sciences: The Impact of Cultural Differences on IT Nearshoring – Risks from a German Perspective. Kauai, Hawaii, January 5 – January 8 2010. Danvers: IEEE.
- [WB10] Walser, K.; Breidung, M. (2010): Vergleichende Analyse des IT-Service-Providing in der Öffentlichen Verwaltung in Deutschland und der Schweiz. Beitrag, Bern: Berner Fachhochschule Kompetenzzentrum Public Management und E-Government
- [Wi85] Williamson, O. E. (1985): The Economic Institutions of Capitalism, New York.

Anhang

Ausgehend von den LNI-Längenbeschränkungen für diesen Beitrag mussten die umfangreichen aber wesentlichen Tabellen zu diesem Beitrag in ein externes Dokument ausgelagert werden, das auf einem öffentlichen Dropbox-Ordner abgelegt ist. Der Link hierzu lautet wie folgt:

https://www.dropbox.com/s/hoclx9g48jh9cxg/IT_Sourc_Komm_Vier_Cases_V%20Outsourcing_Tabellen%20-%20Kopie.pdf?dl=0

The Status Quo of Information Management in Hospitals - Results of an Online Survey

Christian Kücherer¹, Jan David Liebe², Michael Schaaf³, Johannes Thye², Barbara Paech¹, Alfred Winter³ and Franziska Jahn³

Abstract: **Context:** Information Management (IM) departments in hospitals provide IT services supporting patient care and administrative hospital functions. Complexity and impacts of IT failures are continuously increasing. Therefore, professional IM departments are necessary, even though their importance is commonly underestimated. **Objective:** Little is known about the organization and variety of IM departments in German hospitals. Therefore, we want to characterize their capabilities. In this paper, we present a study that analyzes the current status of IM in German hospitals in the dimensions (i) organization of IM departments, (ii) educational degree of Chief Information Officers (CIOs), (iii) communication of the CIO with the hospital management, (iv) usage of IT-process- frameworks, and (v) application system categories used for IM-related tasks. **Method:** The evaluation is based on an online survey of 134 CIOs. The survey questions were developed according to domain-specific literature, the SNIK-ontology, and interviews with domain experts. The survey questions were improved by a pretest. **Results:** The survey indicates that most of the CIOs are graduates in informatics with 13 years of experience, who are responsible for one hospital without being member of the hospital management. CIOs communicate in weekly formal meetings with the hospital management where they discuss projects, finance, security, and critical IT- and hospital services. Most IM departments do not use IT-process frameworks, but nearly all of the IM departments use ticketing systems, network management systems, project management tools, collaboration tools, BI tools, and ERP systems. The results show, that IM departments are generally well organized with the potential for improvement in IT-process-frameworks and application systems for IM functions.

Keywords: information management, hospital, online survey, application systems, CIO, IT department, ITIL, management communication.

1 Introduction

Information management (IM) aims at the systematic management of information systems. An information system can be defined as the socio-technical subsystem of an enterprise comprising processes, information technology, and humans in their information processing roles. In hospitals, the permanent responsibility for life and death

¹ Heidelberg University, Institute for Computer Science, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany, (kuecherer | paech)@informatik.uni-heidelberg.de

² University of Applied Science Osnabrück, Faculty of Business Management and Social Science, Health Informatics Research, Group PO Box 1940, 49009 Osnabrueck, Germany, (J.Liebe | Johannes.Thye)@hsosnabrueck.de

³ Leipzig University, Institute for Medical Informatics, Statistics and Epidemiology, Härtelstr. 16-18, 04107 Leipzig, Germany, (franziska.jahn | michael.schaaf | alfred.winter)@imise.uni-leipzig.de

of patients and effectiveness rather than efficiency determine the activities of the medical professionals [LR13], [MC09]. Therefore, IM in hospitals has to consider organizational and legal requirements. However, information systems in health care are often said to lag behind information systems in industry by a decade. Several German studies support the existence of this felt gap. Two studies published in 2008 revealed the low maturity of applications in German hospitals and low IT budgets in comparison to industry [LKH08], [MSM08]. So far little attention has been paid to the internal structures, IM functions, and application systems used in IM departments (synonyms ICT or IT department). According to Winter et al. [Wi11], IM in hospitals has to differentiate between of strategic, tactical, and operational IM. Strategic IM plans, monitors, and directs the information system's long-term development. Tactical IM updates certain parts of the information system through projects. Operational management has to ensure the information system's daily operation. For all these IM tasks, called the IM functions, clear procedures, applications, competent staff, and a highly skilled manager in the form of a CIO is needed, who is responsible for the functioning of the IM department. Several studies indicate that business success is correlated with IT competencies or the adoption of standardized processes [MSM08], [PSBQ10], [MK11]. Therefore, professional IM departments are an important success factor for information systems.

The overall goal of this paper is to present a first descriptive evaluation to characterize the capabilities of IM departments in German hospitals. The goal is structured by the application of the Goal Question Metric (GQM) approach [Va02] by Basili et al. and is formulated as: *Determine* the capabilities of IM departments in German hospitals *with respect to* (D1) the CIO's position in the hospital management hierarchy, (D2) his/her educational background and experience, (D3) his/her communication habits with the hospital management, (D4) the use of IT-process-frameworks, and (D5) IM-specific applications used *for the purpose of* a professional and reliable strategic, tactical, and operational management *in the context of* IM divisions of German hospitals *from the viewpoint of* researchers and CIOs. We address the aforementioned goal by investigating 9 research questions (RQs), as documented in Table 1. The column *research question* shows the RQs, the column *dimension* shows the mapping of RQ to the dimensions D1-D5 in the goal. We expect the chosen RQs to be of high interest in the IM community.

RQ.ID	Research Question	Dim
RQ.1	How is the CIO positioned in the hospital management hierarchy?	D1
RQ.2	Which educational background does the CIO have?	D2
RQ.3	Which work experience does the CIO have?	D2
RQ.4	How does the CIO communicate with the hospital management?	D3
RQ.5	How often does the CIO communicate with the hospital management?	D3
RQ.6	What are major issues in the communication with hospital management?	D3
RQ.7	Which IT-process-frameworks are used for strat/tact/operat.IM functions?	D4
RQ.8	What is the level of utilization for the ITIL-framework?	D4
RQ.9	Which IM-specific application systems are used to support IM functions?	D5

Tab. 1: Research questions and the dimensions (Dim.) of the goal

The position of the CIO in the management hierarchy is interesting since it shows the status of the CIO and his/her decision power (RQ.1). The educational background and the work experience (RQ.2) of CIOs influence the organization and standard-orientation of the IM department. The work experience (RQ.3) of CIOs does highly influence the organization of the IM department and decisions. Communication is a key management factor. Thus, we are interested in the CIO's communication habits (RQ.4). The form and frequency of communication (RQ.5) with the hospital management indicates the quality of relationship between these two. The issues the CIO discusses with the hospital management (RQ.6) are interesting since they show important aspects of a CIO's daily business. IT-process management frameworks (RQ.7), such as ITIL, COBIT, and PRINCE2, allow reliable and controlled operation processes to support IM functions. The level of utilization of IT process management frameworks (RQ.8) shows to which extent the IM departments act according to the frameworks. IM departments use a variety of application systems to support several IM functions. We want to understand which classes of application systems are used for which IM functions (RQ.9).

This paper is structured as follows: Section 2 explains the methodological background of the survey design and the development of the study questions. The evaluation of the online survey and the answers to the research questions are given in Section 3. Section 4 discusses the results and the insights gained in this study. Potential threats to validity are discussed in Section 5. The conclusion and ideas of future work is given in section 6.

2 Methodology

This section describes the study design process used to construct the online survey, and in brief the performed online survey.

2.1 Survey Design

In the development of the study, four roles were involved. First, the editor team consisted of two persons familiar with the hospital IM domain and basic questionnaire-design. Second, the survey team consisted of two persons who have significant experience with study- and questionnaire-design and are able to rate the defined questions from a survey point-of view. Third, the domain expert is a person, who is familiar with the hospital IM domain and all variations of different hospital IM departments. Fourth, the beta testers, a heterogeneous group of 10 persons who completed the questionnaire as CIO and gave detailed feedback. The survey team also contributed to the selection and review process.

Figure 1 shows the process of the survey design. In the selection process, the editor team formulated the goal and question proposals of the survey according to the GQM scheme, based on three sources. For the structuring of IM functions in hospitals, we adopted the strategic, tactical, and operational IM classification as proposed by Winter et al. [Wi11].

In order to extend the hospital-specific work of Winter et al., the domain-independent IM work of Heinrich/Stelzer [HRS14] has been used. For all topics that relate to tactical IM, i.e. project management, the work of Ammenwerth et al. [Am15] provided a good structure of tasks and best practices. These sources were used to create an ontology of IM in hospitals (SNIK- ontology) [Sc15] that contain IM-specific concepts. The SNIK ontology provided a glossary for the process of questionnaire construction and the formulation of the question and answer sets. The question proposals were refined, restructured, or rejected with the help of domain expert interviews. In the review process, the finalized set of questions was discussed with the survey team. Questions that did not match the survey criteria were redesigned or were refused. The survey criteria for valid questions are: (i) question must contribute to the goal of the survey (ii) question must be easily understandable (iii) question must be answerable by most of the hospitals and suit most of the IM settings (iv) evaluation must be reasonably practicable (i.e. limited number of free-text questions) (v) filter questions may be applicable to reduce the effort for participants. In the pretest phase, the survey team created a pretest version of the online survey based on the questions from step 2 in the tool Unipark⁴. This online survey was evaluated by the beta tester, domain experts, editor- and survey team. All involved roles gave sound feedback and proposed changes to the questions, which were incorporated in the next iteration of the selection and review process. The pretest was iterated twice. Finally, the online survey was executed in the phase online survey.



Fig. 1: The questionnaire development and study design process

2.2 Conducting the Online Survey

The study is designed as a cross-sectional study with possible repetitions. The online survey⁵ comprises 59 questions and contains also questions for other goals, not focused in this article. There are approx. 1980 hospitals in Germany [St15], some of them belonging to a group of hospitals. Thus there is a lower number of CIOs. We contacted N=1284 CIOs via e-mail with an invitation to participate in the online survey. The CIO's E-Mail addresses were available to the survey team from previous surveys.

⁴ <http://www.unipark.com/de/>

⁵ The final online survey questions are available for download

<http://www.snk.eu/de/Ergebnisse/fragebogen2016.pdf>, available in German only.

We collected the data from the online survey between February 12, 2016 and the beginning of April 2016. The survey resulted in 176 analyzable questionnaires, which are completed at least half. 134 of the 176 were completely filled. This results in a response rate of 13.7 %. From the participating hospitals were 11.8 % private hospitals, 37.2 % public hospitals, and 51.0 % in independent nonprofit organizations (n=176).

3 Status of Information Management in German Hospitals

This section describes the status of IM in German hospitals by investigating the research questions in Table 1. The number of answers considered is indicated as n and the standard deviation as σ for each evaluated question. The question number in the questionnaire is indicated by Q. We evaluated the stated questions, whereof the questions marked with /F include the evaluation of free text fields.

3.1 Findings for RQ.1: The CIO's Position in Hospital Management Hierarchy

The position of the CIO in the hospital management hierarchy, as requested in RQ.1, depends on the CIO's inclusion into the hospital management, the job description, and the number of subordinated hospitals. Slightly more than half of the CIOs (54.9 %, n=95) are responsible for a single hospital (Q3, total n=173), whereas all others are responsible for more than one hospital (45.1 %, n=78). Those CIOs who are responsible for more than one hospital take care of 3.97 hospitals in average ($\sigma = 6:394$, n=78). Only a minority of 4.7 % (n=8) are female CIOs, whereas 95.3 % of the CIOs are male (Q5, total n=171). All participants were requested to name their job description (Q6/F, total n=170), which show a great variety. However, 82.9 % of the job descriptions indicate an executive status of the participant by containing the keywords leading, leader, manager, head of, etc. Interestingly, only in 3.5 % of the job descriptions, the term CIO is contained explicitly. The majority of 94.4 % (n=152) of the CIOs are not members of the management (Q18, total n=161). **Summarizing RQ.1**, we can state in most cases the CIO is male and not a member of the hospital management, but in the majority of the cases the job description reveals the CIO's executive status. Slightly more than half of the CIOs are responsible for one single hospital and in all other cases they are responsible for up to four hospitals.

3.2 Findings for RQ.2: The Educational Background of CIOs

Information about a professional training or graduate occupation, and certificates characterizes the educational background of CIOs (RQ.2).

More than half of the participants (52.5 %, n=94) hold a graduate degree from a university or a university of applied sciences. Nearly half of the participants (40.2 %, n=72) completed a professional training, six participants (3.4 %) hold a PhD, and seven

participants (3.9 %) have earned a GMDS medical informatics certificate (Q10/F, total n=179). Those participants that completed a professional training gave information about their specialization of education (Q11, total n=96). 23 participants (24 %) have a professional training in business administration, four participants completed a training in nursing or medical care (4.2 %), two participants finished a medical or therapeutic training (2.0 %) and the vast majority of 49 % (n=47) are qualified IT specialists. Other trainings were completed by 20 participants (20.8 %) in the areas of electronics, chemistry, and mechanical engineering.

Degree	University (% , n)		University of applied sciences (% , n)	
Diploma	32.7 %	34	36.5 %	38
B.Sc.	1.9 %	2	8.7 %	9
M.Sc.	5.8 %	6	6.7 %	7
State examination	3.8 %	4	2.9 %	3
Magister (M.A.)	1.0 %	1		

Tab. 2: Academic degrees.

Regarding the academic degree (Q12, total n=104), Table 3 shows that a Diploma is most common, nearly equally distributed between universities and universities of applied sciences. In contrast, B.Sc. and M.Sc. degrees are still very uncommon. These statements were not introduced until several years after the Bologna Process in 1999 in Germany has begun. Due to the work experience (see findings for RQ.3 in subsection 3.3), only a minority of graduates with M.Sc. and B.Sc. degrees already are in a CIO position. Other degrees are state examination (n=7) and Master of Arts (n=1) that are held by a minority of 7.7 %. The majority of CIOs has graduated from a university of applied sciences. The specialization in academic studies (Q13, total n=117) is in 30.8 % (n=36) informatics and business informatics, followed by 16.2 % (n=19) in business administration and 14.5 % (n=17) in engineering sciences. Only 12 participants (10.3 %) specialized in medical informatics. Medicine, natural sciences, and other specializations make up 28.3 % (n=33). Although medical informatics being a field of studies preparing for the job of hospital CIO, there are fewer graduates in medical informatics than in engineering. The subject of a PhD (Q14/F, total n=6) is in 2 out of 6 cases related to biology and in 1 out of 6 cases related to physics. One participant stated a topic related to computer tomography, which can be located in medical informatics or computer science. Two participants did not state their topic. **Summarizing RQ.2**, it can be said that slightly more than half of the CIOs have graduated with a diploma or a master's degree in a subject related to informatics or business administration, whereas a PhD is an exception. Less than half of the CIOs have completed a professional training.

3.3 Findings for RQ.3: The Work Experience of CIOs

The average work experience of a CIO in this position is 13.52 years, whereas the minimum was 0 and the maximum 35 years (Q7, n=170, $\sigma = 8$; 738). In average, a CIO is with his/her employer for 11.54 years (Q8, n=170, $\sigma = 8$:093). **Summarizing RQ.3**,

we can say that the typical CIO has more than 13 years of experience in his/her job, and stays more than 10 years with the same employer.

3.4 Findings for RQ.4: Communication of the CIOs with Hospital Management

The communication with the Hospital Management is supported by visual aids that contain or display information in the form of office documents, reports, or dashboards. Documents can be digital documents presented on displays or projectors, or they can be printouts of digital documents. Digital documents such as office documents and printouts are used frequently, as shown in Table 4a (Q22/F, total n=215). Information from reporting tools or dashboards is less frequently used. Interestingly, only a minority does not use visual aids at all. Other visual aids mentioned are flip-charts, whiteboards, and video conferencing systems. The most important contents of the visual aids (Q23, total n=149) are decision memos, recommendations, trends, and comparisons of departments or facilities, as depicted in Table 4b. Other contents mentioned were mind-maps, meeting minutes, or e-mails.

Visual aids	%	n	Content	%	n
digital documents	62.6	134	decision memos	85.2	127
printouts of documents	48.8	105	recommendations	73.2	109
reporting tools. dashboards	23.7	51	trends	72.0	108
no visual tools	3.7	8	comparisons of facilities	29.5	44
			comparisons of departm.	22.1	33

(a) Visual aids for communication

(b) Content of visual aids.

Tab. 3: Visual aids and their content.

The contents of digital documents (Q24, total n=147) in the form of tables (17.4 %, n=115) and text (17.7 %, n=117) are considered to be more important than diagrams (13.9 %, n=92). The same applies to printouts of documents: tables (12.1 %, n=80) and text (14.2 %, n=94) are considered important, whereas diagrams (9.7 %, n=64) are less important. Interestingly and intelligibly is that diagrams (6.0 %, n=64) and tables (5.9 %, n=39) are more important when dashboards are used, compared to text (3.2 %, n=21). **Summarizing RQ.4**, it can be said that CIOs communicate with the hospital management by mostly using electronic documents with text and tables. The electronic documents contain decision memos, recommendations, and trends of the IM department. Dashboards focus on diagrams and tables, but are surprisingly less frequently used.

3.5 Findings for RQ.5: Frequency of Communication with the Hospital Management

CIOs communicate with the hospital management in 42.3% (n=91) once a week. A minority of 16.7 % (n=36) of the participants communicate monthly. Only 10.7 % (n=23) communicate daily and only 1.4 % (n=3) communicate less frequently than yearly (Q19, total n=154). Formal meetings are used in 54.9 % (n=85) of the cases to

communicate with the hospital management. In 45.1 % (n=70) of the cases, CIOs rarely have a formal meeting (Q20, total n=155). Informal meetings with the hospital management (Q21, total n=155) are used rarely in 66.5 % (n=103), whereas in 33.5 % (n=52) of the cases they were used frequently. Therefore H.6 is true. **Summarizing RQ.5**, it can be said that most of the CIOs communicate weekly in a formal meeting. Informal meetings are used only rarely.

3.6 Findings for RQ.6: Major Communication Issues with the Hospital Management

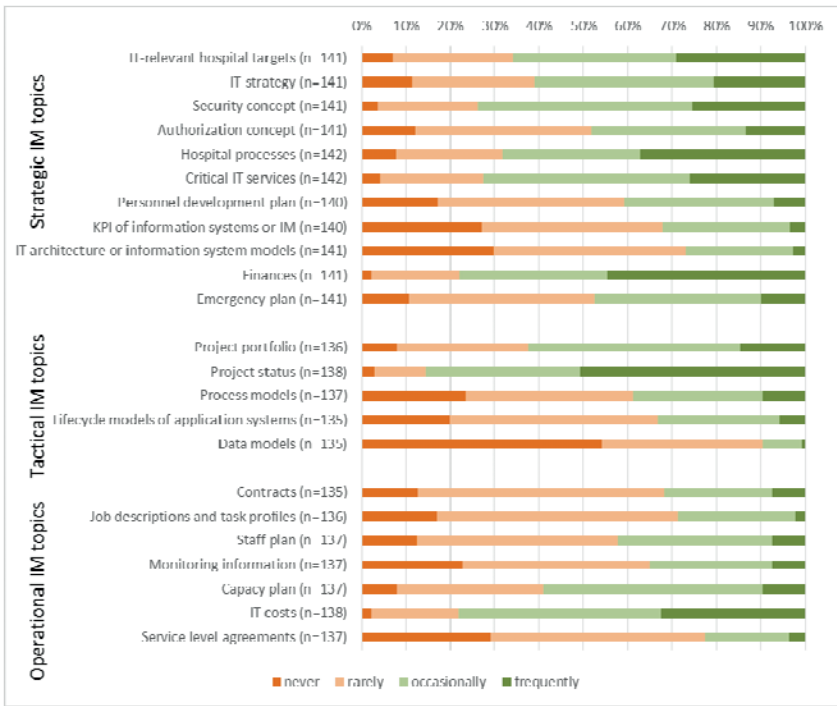


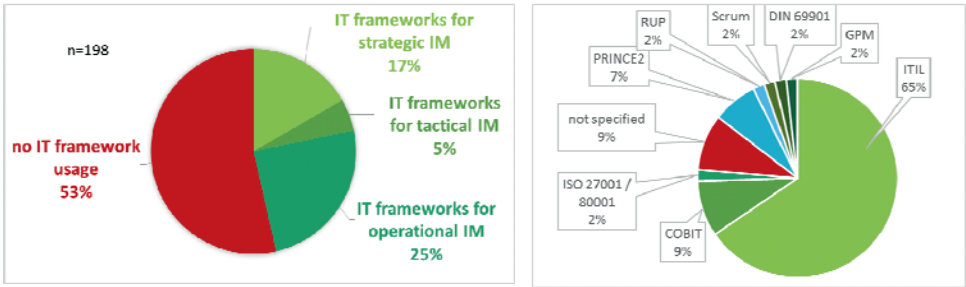
Fig. 2: Observed frequency of used strategic, tactical, and operational information

Nearly all information shown in Figure 2 with its frequency is used in the communication with the hospital management. When considering occasional and frequently used information together, the information most commonly used is project status, financial issues, IT costs, security concepts, critical IT services, hospital processes, and IT-relevant hospital targets. Other strategic information subject for communication (Q40/F, total n=141) that was stated by participants is: daily business, IT incidents, Hospital information systems, interoperability to medical technology, IT requirements from other departments, responsibilities. Other tactical information (Q49/F) is information supplied by users, cost plans, and time schedules.

Summarizing RQ.6, it can be said that major communication issues are related to projects, finance, security, critical IT- and hospital-services, and IT-relevant hospital targets. Monitoring information (status of helpdesk, system workload) is rarely issues.

3.7 Findings for RQ.7: IT-Process-Framework Usage

The survey reveals the IM categories, in which IT frameworks are used, as shown in Figure 3a, (Q16, n=198). A majority of 53 % does not use a framework at all. For operational IM functions, 25 % adopt an IT framework, followed by 17 % for strategic IM functions. For tactical IM functions, only a minority of 5 % adopt an IT framework.



(a) usage of IT frameworks in categories of IM

(b) IT frameworks frequency

Fig. 3: IT framework usage

The most frequently used IT framework (Q37/F, Q45/F) is ITIL (65 %), followed by COBIT (9 %), and PRINCE2 (7 %), as depicted in Figure 3b. Other IT frameworks such as ISO 27001/80001, RUP, Scrum, DIN 69901, and GPM (all 2 %) were mentioned each by one participant only, and therefore do not play a significant role. Interestingly, CMMI and PMBOK are not used at all. ITIL can be viewed as both, operational and strategic IM. The IT-governance framework COBIT covers strategic, tactical, and operational IM aspects. As a project management framework, PRINCE2 is a tactical IM framework. **Summarizing RQ.7**, it can be said that for strategic IM ITIL and COBIT are used, for tactical IM PRINCE2 is used and for operational IM ITIL and COBIT are used.

3.8 Findings for RQ.8: The Level of Utilization of the ITIL-Framework

ITIL Process	not adopted	ad. planned	partially ad.	fully ad.
Service strategy	34.8 % 16	19.6 % 9	43.5 % 20	2.2 % 1
Service design	28.3 % 13	28.3 % 13	39.1 % 18	4.3 % 2
Service transition	29.8 % 14	21.3 % 10	42.6 % 20	6.4 % 3
Service operation	10.6 % 5	14.9 % 7	53.2 % 25	21.3 % 10
Cont. Service improvement	32.6 % 15	32.6 % 15	26.1 % 12	8.7 % 4

Tab. 4: Degree of ITIL process adoption (ad.) for operational IM (rows present % and n).

Table 5 shows the degree of ITIL process adoption (Q57, n=47), whereas 36 participants adopt at least one of the ITIL processes partially or fully. Service operation is adopted by 35 participants partially or fully. Interestingly, more than 70 % use ITIL and most of the participants adopted it partially. It is surprising that strategy, design, and transition are adopted partially by many participants (43.5 %, resp. 39.1 %, resp. 42.6 %). **Summarizing RQ.8**, it can be said that ITIL service operation has the highest level of utilization, followed by service transition and service strategy.

3.9 Findings for RQ.9: Application System Categories for IM Functions

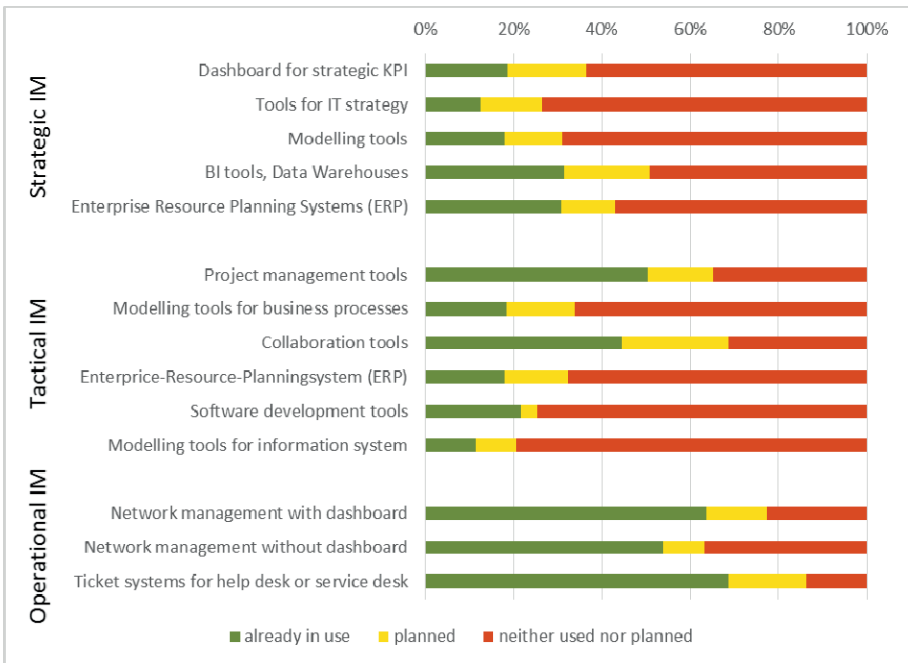


Fig. 4: Application systems used for IM functions

IM departments apply several tools in the form of application systems for IM functions (RQ.9). Since there is a great variety of different tools from different vendors, we examine the categories of application systems. For instance the category of *project management* application systems comprise the products Microsoft Project, inloox, FreeProject, and ProjectLibre. Figure 4 shows the distribution of application systems categories used for strategic, tactical, and operational IM (Q38, Q47, Q53). Interestingly, the vast majority of IM departments does not use application systems for various IM functions except for project management (n=68, total n=138), collaboration (n=61, total n=138), network management (n=79, total n=138), and ticketing (n=94, total n=138). In consequence, these IM functions seem to be important and complex and therefore

require tool usage. For strategic IM functions, mainly BI tools and Data Warehouses and ERP systems are used by 31 % of the IM departments. Only 30 % of IM departments use ERP systems for strategic IM. Project management tools and collaboration tools (SharePoint, Wikis etc.) have a high popularity for tactical IM functions. Network management systems with dashboards and ticketing systems for help- or service-desks have an extensive use of 57 % resp. 68 % in operational IM functions. Participants have additionally mentioned (Q39/F, Q48/F, Q54/F) that they use documentation tools and knowledge management tools (wikis and mind maps) for tactical and operational IM functions, as well as server-/client-management systems and software distribution systems for operational IM functions. **Summarizing RQ.9**, it can be said that most of the IM departments do not use application systems for strategic and tactical IM functions, but do almost always use application systems for operational IM functions. When application systems are used for IM functions, they are (in order of frequency) ticketing systems, network management systems, project management tools, collaboration tools, BI tools, Data Warehouses, and ERP Systems.

4 Discussion

In the following, we relate our study results to former studies on CIOs and IM in hospitals. Whereas our study aimed at the analysis of internal functions, application systems, certification, and staff-related issues of IM departments in hospitals, the study by Leimeister et al. [LKH08] from 2008 focused on strategic IT goals, IT cost, and the functionalities and user satisfaction in the context of hospital information systems. In our study, we could confirm some characteristics of hospital CIOs that did not change in the past 8 years. Regarding RQ.1, the CIO's position, there is still only a very small number of „real“ CIOs who are officially named „chief information officer“. The results of the survey indicate that the hospital management knows about the importance of the IM department, although they do not organize the IM department as an inherent part of the hospital management. Regarding RQ.2, the educational background, the number of CIOs holding an academic degree (52.5 %) was comparable to the corresponding value in [BMB06] (59.5 %). Interestingly, only few CIOs have graduated in medical informatics, although they are supposed to be domain experts. There might be two reasons: first, there are only few graduates or second, the position of a CIO combines management functions with domain knowledge acquired on the job. The work experience of CIOs of 13 years with the same employer for 11 years as investigated by RQ.3 indicates a strong relationship between hospital management and the CIO. The observed steadiness is of high importance for continuous and reliable management, and the strategic alignment of IM departments. Concerning the communication between the CIOs and the hospital board (RQ.4 and RQ.5), Burke et al. [BMB06] associated CIOs reporting to the CFO (Chief Financial Officer), the CEO (Chief Executive Officer), or the COO (Chief Operations Officer) and the revenues of the hospital. They found out that reporting to the CFO correlated with higher revenues in American hospitals. Our focus on the frequency of contact and the topics which are relevant in the communication between the CIO and

the hospital board add another, more qualitative view on the relationship between the CIO and the hospital board. The communication in weekly, formal meetings indicates an appropriate and purposeful cooperation of IM department and hospital management. Regarding RQ.6, the observed issues in communication were expected, especially finance and security topics. Interestingly, helpdesk and system workload rarely are issues. The adoption of ITIL in hospitals (RQ.7) of five European regions was analyzed by Hoerbst et al. [Ho11] in interviews conducted in 2008. At that time, only 5 out of 75 participating hospitals in Austria, Bavaria, Slovakia, South Tyrol, and Switzerland had already implemented parts of ITIL processes. Now we can see that the adoption of ITIL has increased over the past few years. However, 53 % of the hospitals still do not use a framework at all. In these cases the management of various IM functions has the opportunity to improve. The level of ITIL process utilization (RQ.8) shows the operational importance of the ITIL framework. However, it also shows the strategic utilization, which is surprising for us. Regarding RQ.9, the evaluation of the categories of application systems used in IM departments shows a clear deployment of tools for operational IM functions. Obviously, there is a lack of usage for strategic and tactical IM functions. Application systems for those IM tasks might just not be necessary or there might be a high potential for customized tools that support efficient IM functions.

5 Threats to Validity

The threats to validity of this study are structured according to Wohlin et al. [WHH03]. **Construct validity** considers whether the study measures what it claims [WHH03]. This study is a cross-sectional study that evaluates the capabilities of IM departments in German hospitals. Survey questions were designed by the use of literature and were subject of a review process and a pretest. Nevertheless, IM departments are different. Therefore, the questions allow to capture a wide range of possibilities. By comments in free text fields, we enabled the participants to submit additional information. Questions might be misunderstood or some participants might not be familiar with the IM classification. Therefore, we performed a review process and a pretest. We also gave examples and explanations in the questions that help to understand terms and grouping. **Internal validity** determines the extent of conclusions that can be drawn from a study, in particular by eliminating the bias of the study [WHH03]. Participants might bias this study, since only interested persons contribute to the survey. Therefore, the set-up of non-participating IM departments remains unknown. This problem is mitigated by motivating the participants with an incentive⁶, an invitation sent by a professor of medical informatics, and the collaboration with the health IT report [Hü14] research group. Another threat to validity is that non-CIOs could also have conducted the survey. We mitigated this by addressing the invitation to CIOs only, which is evident by the job descriptions in RQ.1. There were no non-CIO-like job descriptions given. **External Validity** describes the possible generalization or transfer of the study results to other

⁶ Participants could win a license for the enterprise modeling tool 3LGM2 , <http://www.3lgm2.de/>

situations [WHH03]. This study targets German hospitals and might not be transferable to other countries. As the focus is strictly on IM in hospitals, the results cannot be transferred to IM departments of other domains. The small number of 176 participants is mitigated by repeating the survey after one year.

6 Conclusions and Future Work

This article has investigated the status quo of IM in German hospitals in five dimensions. It shows that IM departments have a reliable foundation but have potential for improvement in IT-process-framework usage and a better utilization of application systems for IM functions. Future works include the analysis of more dimensions that characterize IM in a further study. Also the IM department's professionalism correlated to its outcome and success needs to be evaluated in a further study based on existing data. The IM department's success comprises support for processes, satisfaction of users and hospital management, and IT costs. In a combination with the health IT report, the success factors and dimensions D1-D5, which characterize the IM department, could be linked with each other. We also need to understand, how the IM changes over time and which future challenges in IM can emerge.

Acknowledgments

The authors would like to thank all participants of the online survey, Mrs. Hübner and her Health Informatics Research Group for their survey support, Thomas Wendt for his early feedback and all beta testers, in particular Jens Hüasers, Frank Meineke, Stefan Smers, Anatoli Zeiser, and others. This work was supported by the DFG (German Research Foundation) under the Project SNIK: Semantic Network of Information Management in Hospitals, Grant no. 1605/7-1 and 1387/8-1.

References

- [Am15] Ammenwerth, Elske; Haux, Reinhold; Knaup-Gregori, Petra; Winter, Alfred: IT Projektmanagement im Gesundheitswesen. Schattauer, Stuttgart, 2nd edition, 2015.
- [BMB06] Burke, Darrell; Menachemi, Nir; Brooks, Robert: Health care CIOs: assessing their fit in the organizational hierarchy and their influence on information technology capability. *The health care manager*, 25(2):167–172, 2006.
- [Ho11] Hoerbst, Alexander; Hackl, Werner O; Blomer, Roland; Ammenwerth, Elske: The status of IT service management in health care-ITIL in selected European countries. *BMC medical informatics and decision making*, 11(1):1, 2011.
- [HRS14] Heinrich, Lutz Jürgen; Riedl, René; Stelzer, Dirk: *Informationsmanagement*. De Gruyter, Berlin, 11th Edition, 2014.

- [Hü14] Hübner, Ursula Hertha; Liebe, Jan-David; Straede, Matthias-Christopher; Thyse, Johannes: IT-Report Gesundheitswesen-Schwerpunkt IT klinischer Prozesse. Schriftenreihe des Niedersächsischen Ministeriums für Wirtschaft, Arbeit und Verkehr, 2014.
- [Kü15] Kücherer, Christian; Jung, Manuel; Jahn, Franziska; Schaaf, Michael; Tahar, Kais; Paech, Barbara; Winter, Alfred: System Analysis of Information Management. In: INFORMATIK 2015, Volume LNI P-246, Cottbus, Germany, pp. 783–796, 2015.
- [LKH08] Leimeister, Jan Marco; Klapdor, Sebastian; Hörmann, Christian: IT-Management in deutschen Krankenhäusern: Eine empirische Studie unter IT-Entscheidungssträgern. BoD–Books on Demand, 2008.
- [LR13] Lorenzi, Nancy M; Riley, Robert T: Organizational aspects of health informatics: managing technological change. Springer Science & Business Media, 2013.
- [MC09] Meyer, Derek; Cox, Benita: Can signalling theory and the semaphoric nature of information systems explain clinicians' ambivalence to informatics? Studies in health technology and informatics, Volume 160, pp. 671–674, 2009
- [MK11] Marrone, Mauricio; Kolbe, Lutz M: Uncovering ITIL claims: IT executives' Perception on benefits and Business-IT alignment. Information Systems and E-Business Management, 9(3):363–380, 2011.
- [MSM08] Messerschmidt, Marcus; Schüle, Peter; Murnleitner, Martin: Der Wertbeitrag der IT zum Unternehmenserfolg. 2008.
- [PSBQ10] Papke-Shields, Karen E.; Beise, Catherine; Quan, Jing: Do project managers practice what they preach, and does it matter to project success? International Journal of Project Management, 28(7):650 – 662, 2010.
- [Sc15] Schaaf, Michael; Jahn, Franziska; Tahar, Kais; Kücherer, Christian; Winter, Alfred; Paech, Barbara: Entwicklung und Einsatz einer Domänenontologie des Informationsmanagements im Krankenhaus. In: INFORMATIK 2015, Volume LNI P-246, Cottbus, Germany, pp. 753–765, 2015.
- [St15] Statistisches Bundesamt: Gesundheit. Grunddaten der Krankenhäuser 2014. Fachserie 12 Reihe 6.1., 2015.
- [Va02] Van Solingen, Rini; Basili, Vic; Caldiera, Gianluigi; Rombach, H Dieter: Goal question metric (gqm) approach. Encyclopedia of software engineering, 2002.
- [WHH03] Wohlin, Claes; Höst, Martin; Henningsson, Kennet: Empirical Methods and Studies in Software Engineering. In: Empirical Methods and Studies in Software Engineering: Experiences from ESERNET, volume LNCS 2765 of Lecture Notes in Computer Science, pp. 7–23. Springer Berlin Heidelberg, 2003.
- [Wi11] Winter, Alfred: Health information systems. Springer, London, 2nd edition, 2011.

Development of a Subject-Oriented Reference Process Model for the Telecommunications Industry

André Nursinski-Stolberg¹, Kiritharan Gangatharan² und Christian Czarnecki³

Abstract: Generally the usage of reference models can be structured top-down or bottom-up. The practical need of agile change and flexible organizational implementation requires a consistent mapping to an operational level. In this context, well-established reference process models are typically structured top-down. The subject-oriented Business Process Management (sBPM) offers a modeling concept that is structured bottom-up and concentrates on the process actors on an operational level. This paper applies sBPM to the enhanced Telecom Operations Map (eTOM), a well-accepted reference process model in the telecommunications industry. The resulting design artifact is a concrete example for a combination of a bottom-up and top-down developed reference model. The results are evaluated and confirmed in practical context through the involvement of the industry body TM Forum.

Keywords: Subject-oriented Business Process Management, Reference Process Model, eTOM, Telecommunications Industry.

1 Introduction

The design and improvement of business processes supported by reference process models are intensively researched topics that are influenced by a broad variety of methodical approaches (e.g. Becker and Schütte 2004; Malinova et al. 2013; Houy et al. 2014). The emerging digital age has a significant impact on business processes in order to remain competitive with respect to fast changing market and innovation requirements (Schmiedel and vom Brocke 2015). Flexibility and agility (Richter and Esswein 2014), collaborative modeling concepts combined with increased involvement of stakeholders (Krumeich et al. 2013; vom Brocke 2015), and the direct execution of process models (Börger and Fleischmann 2015) are demanded.

A common way of structuring the work is a hierarchical decomposition from a high-level process framework to detailed operational processes (Rammner and Ramias 2010). According to this hierarchical process structure, two different starting points can be distinguished. Firstly, the performance of current processes is measured on an operational level to identify weaknesses that are improved in an incremental way, like e.g. proposed by Six Sigma (Harmon 2010). Secondly, future (target) processes are

¹ Detecon International GmbH, Sternengasse 14-16, 50676 Köln, andre.nursinski-stolberg@detecon.com

² Detecon International GmbH, Sternengasse 14-16, 50676 Köln, kiritharan.gangatharan@detecon.com

³ Hochschule für Telekommunikation Leipzig, Gustav-Freytag-Straße 43-45, 04277 Leipzig, czarnecki@hft-leipzig.de

completely re-designed starting with a high-level understanding of the value chain, like e.g. proposed by Business Process Reengineering (Hammer and Champy 1994, Harmon 2010). Hence, according to the starting point those approaches can be differentiated into top-down and bottom-up.

With respect to bottom-up design and implementation of business processes, there is currently a discussion about the subject-oriented Business Process Management (sBPM) (Fleischmann 2013). sBPM offers a process modeling notation that focuses on the process actors and their interactions combined with a process modeling tool that allows a direct execution (Fleischmann 2013). As sBPM starts with a bottom-up approach on an operational level with direct involvement of the process actors, an advantage is seen in a fast organizational implementation (Hübner 2015). The practical usage of sBPM is described in various case studies by different authors. The case studies underline the benefits of a bottom-up approach with respect to higher acceptance of the process design as well as to fast implementation in the organization (cf. Section 2.1). However, the sBPM usage described in these cases focuses on a small amount of operational processes that are improved through design workshops with process actors (e. g. Augl and Stary 2015; Lederer et al. 2015). The restriction to a small amount of operational processes can be seen as a limitation of the bottom-up approach proposed by sBPM.

Therefore our paper offers a combination of the well-accepted top-down reference process model *enhanced Telecom Operations Map* (eTOM) with a bottom-up reference process modeling approach, i.e. for both the development as well as the usage. In the telecommunications industry the reference process model eTOM is widely used and confirmed as a de facto standard (ITU 2007; TM Forum 2015) as well as discussed in scientific research (Kelly 2003; Holschke et al. 2009; Czarnecki et al. 2013). We have researched the question how a combination of the sBPM bottom-up approach with the top-down reference process model eTOM could be realized. We have structured this question as follows: (1) the development of a general method for combining sBPM with a reference process model and (2) the application of this method in order to design concrete sBPM artifacts.

The results presented in this paper are based on the work of the TM Forum eTOM working group³ that is an industry organization consisting of telecommunications companies, vendors, and consulting companies worldwide. We have developed the results in an iterative manner over two years considering practical requirements and feedback of the TM Forum. The results are presented here in a generalized and summarized form. The developed artifacts are an extension of the existing reference process model eTOM following the design science approach (Hevner et al. 2004). As an industry-specific reference model the results support telecommunications companies to address their transformational needs of the emerging digital age in an agile manner. From a general perspective, it contributes to the combination of inductive, bottom-up development with top-down reference models based on a practical example.

³ The authors are part of this working group and have developed the presented results.

The paper is organized in the following way. Section 2 provides a short overview of reference process modeling, the concrete reference process model eTOM, and the bottom-up process design approach sBPM. In section 3, the development of our subject-oriented reference model for eTOM is described. The results are illustrated with the example of the customer-centric *Request-to-Answer* process. Furthermore, the evaluation of the results is discussed. Conclusion and limitations are presented in section 4.

2 Related Work

2.1 Reference Process Modeling

The design and improvement of processes can be supported by reference process models (e.g. Becker and Schütte 2004; Malinova et al. 2013; Houy et al. 2014). The usage of reference process models is related to cost and quality benefits (Fettke and Loos 2007; Fettke et al. 2006), widely accepted in practice, and often organized in a top-down manner (Becker and Schütte 2004; vom Brocke 2015). As a reference process model should be applicable for a range of situations, the generalization from a concrete implementation is necessary. In this context, reference process modeling is structured in (1) the development of a reference process model in order to support its reuse in various similar problem domains, and (2) the usage of an existing reference process model in order to develop a so-called application model customized to the requirements of a specific situation (Becker and Schütte 2004; vom Brocke 2015). The development of reference process models can be either organized top-down or bottom-up. While deductive development of reference process models starts top-down based on general theories and concepts (Becker and Schütte 2004; vom Brocke 2015), inductive development of reference process models derives generalizable content bottom-up from individual process models (Ardalani et al. 2013, Martens et al. 2015, Yahya et al. 2012). There is an intense discussion about the inductive development of reference process models in the research community that is mainly related to heuristic methods (Martens et al. 2015). So far, the usage of existing reference process models is typically organized in a top-down manner starting with a process framework followed by its decomposition to an operational level combined with a customization according to company-specific requirements. (Rosemann 2003; Rummler and Ramias 2010; Malinova et al. 2013).

2.2 The Reference Process Model eTOM

In the telecommunications industry the reference process model *enhanced Telecom Operations Map* (eTOM) is widely used and confirmed as a de facto standard (ITU 2007; TM Forum 2015) as well as discussed in scientific research (Kelly 2003; Holschke et al. 2009; Czarnecki et al. 2013). eTOM was developed and is continuously updated by the non-profit industry organization TM Forum. eTOM supports telecommunications companies in their transformational needs, e. g. caused by major changes in the value

chain and harmonization efforts of processes and information systems (Bub et al. 2011, Czarnecki et al. 2013). It consists of a high-level reference structure for processes in the telecommunications industry and their hierarchical decomposition on different levels of detail (Kelly 2003). Hence, originally eTOM offers a hierarchical definition and structuring of processes without any logical process flow. This point was addressed by Czarnecki et al. (2013) by describing the extension of eTOM through reference process flows in three process domains, namely customer-centric domain, network domain and product domain. The eTOM sub-processes are connected and arranged in 18 process flows.

The customer-centric domain captures the following seven end-to-end reference process flows and act as a basis for the development of our subject-oriented reference process model: (1) Request-to-Answer, (2) Order-to-Payment, (3) Usage-to-Payment, (4) Request-to-Payment, (5) Termination-to-Confirmation, (6) Problem-to-Solution, (7) Complaint-to-Solution. eTOM is a valid example for a well-accepted top-down reference process model. The application of sBPM to eTOM was proposed by the authors as a new development to the TM Forum eTOM working group which has accepted this topic. The results of this extension are described in this paper.

2.3 Subject-Oriented Business Process Management

Subject-oriented Business Process Management (sBPM) is a process modeling approach based on a communication view between the involved actors (subjects) within a process (Schmidt et al. 2009). It is based on the idea that business process descriptions in natural language are complete sentences composed of subject, predicate and object (Schmidt et al. 2009). It is argued that in commonly applied approaches the focus is more on the predicate and object and less on the subjects. Therefore, sBPM focuses on the involved subjects within a process (Fleischmann et al., 2009).

It follows a short explanation of the subject-oriented approach according to Fleischmann et al. (2009): The focused subjects within a process can be concrete persons, roles, or machines that exchange messages among each other and thereby represent the process at hand. Within the first step of sBPM modeling the subjects and their interactions are covered, leading to a *Subject Interaction Diagram* (SID). The SID demonstrates the exchanged messages among the subjects. The second step, the behavior of the subjects is modeled by sequencing the activities of the respective subject using three states and transitions, namely *Send*, *Receive* and *Function*. The result of this step is captured in a *Subject Behavior Diagram* (SBD).

Since 2009 the usage of sBPM is discussed by practitioners and researchers on a yearly conducted dedicated sBPM conference⁴. Also various practical case studies of the sBPM usage were presented there, and published by Fleischmann et al. (2015). Please see Table 1 for a summary of relevant case studies.

⁴ Please see www.s-bpm-one.org for further information.

Source	Case Scope	sBPM Usage
Lederer et al. (2015)	optimization of customer support process, monitoring of product development process	workshops with process actors for as-is analysis and to-be design, detailed modeling of existing BPMN diagrams, implementation in monitoring tool
Augl and Stary (2015)	process analysis and development at a university clinic for radiotherapy-radiation oncology	workshops with process actors for modeling as-is situation and developing model variants
Sprogies and Schmidt (2015)	design of the software deployment process for an IT service provider	workshops with process actors for modeling the as-is situation, bottom-up validation and optimization
Hübner (2015)	standardization of an IT hardware procurement and allocation process across subsidiaries	workshops with subsidiaries for as-is modeling
Lorbacher (2015)	digitalization of contract performance process	Scrum-based process modeling workshops, implementation based on sBPM tool and existing ERP system

Table 1. Cases describing the sBPM usage

All these case studies are focused on a small amount of operational processes, e. g. Lederer et al. (2015) describe the detailing of selected existing process diagrams through sBPM. They all describe bottom-up workshops for as-is analysis and/or incremental process optimization, e. g. Hübner (2015) emphasizes the bottom-up approach of sBPM, and Augl and Stary (2015) provide a detailed illustration of the as-is design workshops. All cases describe advantages in the acceptance of the process design as well as the fast implementation which confirms the benefits of a bottom-up approach with respect to agile and flexible processes. However, none of the cases describe a company-wide process design or the usage of a reference model.

The need for using reference models with sBPM is discussed by some authors. Forbrig (2015) identifies the requirement to reuse certain modeling parts in sBPM and proposes the concept of generic components. Garon et al. (2014) propose a sBPM reference implementation of the ITIL change management process. Piller (2015) suggests a sBPM representation of a reference process for maintenance. Mesbahipour et al. (2014) discuss how architecture planning on business, application, and data is achieved using the concepts of sBPM in rapid architecture cycles, creating greater agility. Hence, in all these papers sBPM is used for the detailing of single operational processes. In contrast, our paper describes the application of sBPM to a whole reference process model by combing the bottom-up and top-down perspectives.

3 A Subject-oriented Reference Process Model for eTOM

In our research we adhere to the principles of the design science research methodology (Hevner et al. 2004): after the initial problem identification, we developed a method for combining the bottom-up approach of sBPM and the top-down perspective of eTOM (cf. Section 3.1), and we have designed concrete sBPM artifacts based on this method (cf. Section 3.2). We have conducted both in an iterative way with input from real-life implementation projects. A reference model is a point of reference for a whole range of situations (Fettke and Loos 2007). Therefore a major part of our work is a generalized structure of process subjects that is independent from a concrete organizational structure. We have based this structure on a generalized job profile database that was successfully implemented in various projects. Furthermore we have studied the applicability of our sBPM artifacts in expert discussions and workshops (cf. Section 3.3). We have also undergone the evaluation by the TM Forum that has agreed the publication of our work.

3.1 Development of the sBPM Reference Process Model for eTOM

Our work is structured into (1) the development of a general method for combining sBPM and eTOM and (2) the application of this method in order to design concrete sBPM artifacts as an extension of eTOM.

The method is organized in the following four steps (cf. left part of Fig. 1):

1. *Identification and detailing of relevant processes*: Due to the bottom-up approach of sBPM detailed process definitions are required.
2. *Definition of generalized subjects*: The reference model eTOM is independent from a concrete organizational structure. The definition of subjects as a general reference is required.
3. *Mapping of subjects to sub-processes and tasks*: sBPM starts in a bottom-up manner with the interaction between the subjects involved in a process (cf. Section 2.1). Therefore, a mapping between the reference model eTOM and the generic subjects is a prerequisite for the design of sBPM artifacts.
4. *Development of Subject Interaction and Subject Behavior Diagrams*: The sBPM diagrams provide a concrete reference for the interaction and behavior of subjects as a process definition according to the sBPM concept.

We have applied the above method for the design of concrete sBPM artifacts as an extension of the existing reference model eTOM (cf. right part of Fig. 1). We have conducted the whole development in an iterative manner including regular expert discussions in the TM Forum working group and real-life data from projects in the telecommunications industry. The scope for our development of a subject-oriented process model is based on the eTOM domains (cf. Section 2.2). Mainly due to higher use and application in practice as well as the availability of valuable input from prior transformation projects, we focus in the current work on the customer-centric domain,

covering all activities related to sales, production, logistics, and after-sales initiated by the customer (Czarnecki et al. 2013). According to our scope, the seven reference process flows of the customer-centric domain were used as input from the existing eTOM model. Those were analyzed with respect to their sub-processes and the involved tasks performed by the particular organizations and their representatives. Fig. 2 shows an example of the sub-processes and tasks of the reference process flow *Request-to-Answer* that were used as input for our development.

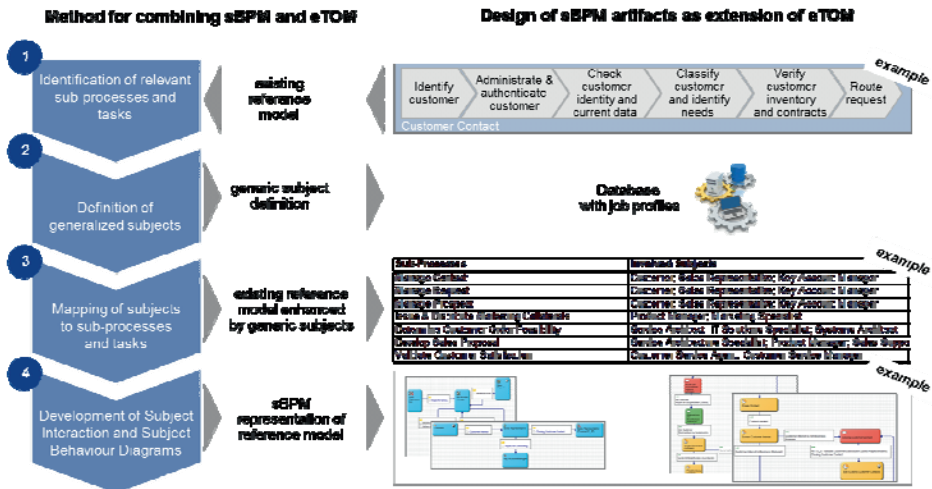


Fig. 1. Development of a subject-oriented reference process model for eTOM

According to our aim of a bottom-up development, in the next step insights from completed transformation projects in the telecommunications industry were used. On the one hand, real-life data of organizations that have successfully implemented those customer-centric processes were studied with the goal to identify the involved subjects that are part of these processes in their day-to-day business. On the other hand, job profiles and descriptions of the individual organizations and their subjects were screened with respect to the above mentioned sub-processes and their related tasks. The job titles of those representatives that were involved in the mentioned tasks to a strong extent were extracted in a bottom-up manner. However, in order to ensure that the retrieved job titles do not carry any specifics of the concrete company, they were generalized by comparing the individual job titles across organizations and reducing their nomenclature to their core functions based on the standardized definition of the eTOM model.

Afterwards the bottom-up collected and thereafter generalized subjects are mapped to the tasks and sub-processes of the seven reference process flows from the customer-centric domain of eTOM. The boundaries of the activities assigned to the individual subject are based on the same logic as they were extracted. The tasks are allocated to the subjects based on their job profiles and from observations how the responsibilities are

defined in practice within the analyzed organizations. The company-specific insights used in this work are derived from a long established and regularly updated internal database that consists of retained results from multiple, international transformation projects in the telecommunications industry. The database contains samples of successfully implemented customer-centric processes, organizational structures, and job profiles from the respective companies. Actually in this database more than 500 entries are listed categorized in 40 job families.

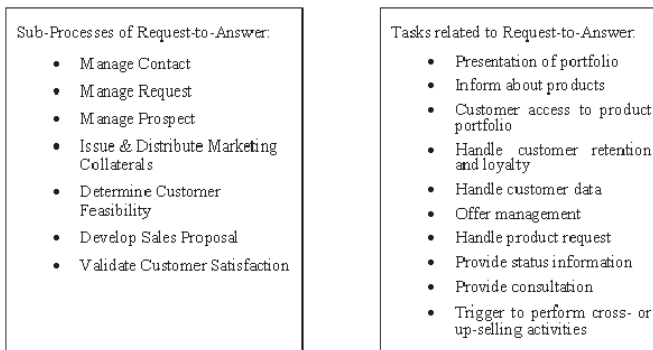


Fig. 2. Detailed sub-process and tasks of the reference process flow “Request-to-Answer”

With the aim to add the subject view to the eTOM model for each sub-process of the high-level process flow, detailed process models were designed based on the sBPM notation (Fleischmann et al., 2009). Firstly, the subject-oriented communication flows were modeled in Subject Interaction Diagrams. In combination with the generated generic subjects according to the above described approach, the process flows are the basis for the Subject Interaction Diagrams. The Subject Interaction Diagram shows the sent and received messages between the involved subjects. Secondly, for every subject of the Subject Interaction Diagram, the internal behavior was captured in Subject Behavior Diagrams that summarize the activities of each subject from its point of view within that particular process. Moreover, the messages that the subjects receive and send by interacting with other actors involved in the process are illustrated.

3.2 Illustration of the Results

The complete subject-oriented process reference model for the customer-centric processes of eTOM contains a Subject Interaction Diagram and a Subject Behavior Diagram for 97 individual subjects. Due to the limited space, we present within this paper an exemplary extract from the process flow *Request-to-Answer*.

Comparing the implemented *Request-to-Answer* process at various telecommunication companies, we identified the generic subject *Sales Representative* as the first subject involved in managing the customer contact and requests. He is the first touch point for the customer. The *Sales Representative* receives the request and either performs the tasks

related to the sub-process *Manage Contact* himself or involves the *Key Account Manager* who is another generic subject. This particular interaction among those subjects illustrates an exemplary extract from the whole *Request-to-Answer* process (cf. Fig. 3).

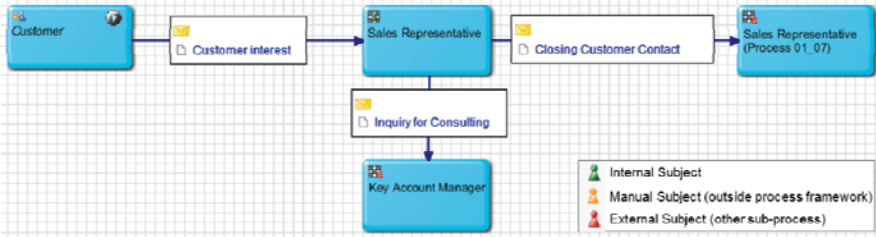


Fig. 3. Extract of Subject Interaction Diagram for “Request-to-Answer” process

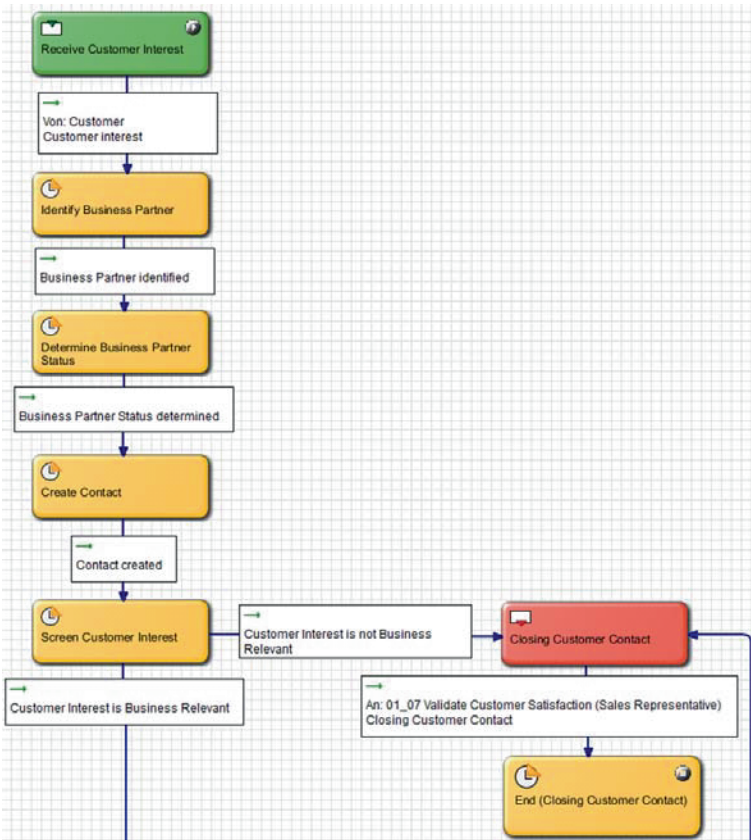


Fig. 4. Subject Behaviour Diagram for the generic subject “Sales Representative”

This extract also shows how the different sub-processes are connected to each other. In this example the *Sales Representative* forwards the task *Closing Customer Contact* to himself in the sub-process *Validate Customer Satisfaction*. Fig. 4 illustrates an example of the Subject Behavior Diagram for the subject *Sales Representative*. At first, the *Sales Representative* receives a request from the customer. Thereafter, he identifies the business partner and his status. Then he creates a contact after confirmation followed by screening the customer need. If it is not relevant for the concrete business, it is closed by the *Sales Representative*. Otherwise the *Sales Representative* verifies and updates the existing customer request and proceeds with the subsequent tasks. In this manner, we have designed the Subject Interaction Diagrams and the respective Subjective Behavior Diagrams for all seven reference process flows from the customer-centric domain.

3.3 Evaluation of the Results

In design science the development of an artifact is an iterative process that requires evaluation and refinement, e. g. based on case studies, analytical argumentations, or field studies (Hevner et al. 2004). A reference model is developed for various model users that could apply the reference model in a comparable problem domain (Fettke und Loos 2007). Both the applicability for different real-life scenarios as well as the anonymity of the possible model users makes the evaluation of a reference model a difficult task (Frank 2007).

In our context, the developed sBPM reference process flows are an extension of the existing reference model eTOM. Hence, the development is based on content that was already evaluated and confirmed by the industry organization TM Forum. The evaluation of our designed sBPM artifacts can be structured into the following two parts:

1. Evaluation of the generalized subject structure
2. Evaluation of Subject Interaction Diagrams and Subject Behavior Diagrams

For the generalized subject structure we have used an own organizational database that contains the organizational structure of various telecommunication companies worldwide. Those organizational structures were generalized in job profiles defining organizational entities that are commonly used in telecommunication companies. Those job profiles were again used as a reference in various reorganization projects (cf. Table 2) which confirm our generalized subject structure.

Our results are illustrated as detailed Subject Interaction Diagrams and Subject Behavior Diagrams. Those diagrams are the outcome of two years iterative development including workshops with worldwide experts and possible users in the telecommunications industry. In addition, we have presented and discussed our design artifacts in the eTOM working group as responsible body of the TM Forum. The TM Forum has agreed the official publication of our sBPM artifacts. This is an additional evaluation of our sBPM reference process model by an industry organization.

Project	Case Scope
Reorganization of North African telecommunication company	Development of a target organizational structure, including comprehensive job descriptions
Restructuring of South African telecommunication company	Development of new target structure and organizational detailing (business tasks and job descriptions)
Process design for Asian telecommunication company	Development of detailed target processes based on eTOM and mapping to organizational structure
Process management for Asian telecommunications company	Implementation of eTOM-based process management, mapping to target organization

Table 2. Real-life implementation projects of our generalized job profiles (extract)

4 Conclusion and Limitations

In this paper, we provide a concrete example for a combination of a top-down reference process model with a bottom-up process modeling approach. Our model is derived from real-life organizational data and applies the subject-oriented process notation (sBPM). Our development is related to the well-recognized reference model eTOM and the industry body TM Forum. Our results are based on real-life implementations and are confirmed by industry experts. Furthermore, the TM Forum has accepted its publication after presentations and discussion in the eTOM working group.

By working in accordance with the design science paradigm (Hevner et al. 2004), we have developed a subject-oriented reference model for the customer-centric domain of eTOM. Major challenge was the definition of a generic subject structure in order to define the subject behavior and interaction. In this paper, we first explain the general development process. Hereby, generic subjects are created based on a wide range of real-life transformation projects at international telecommunications companies. Based on our detailed job profile database, we have performed a mapping between the sub-processes and tasks to generic subjects. Furthermore we explain the concrete results of our sBPM artifacts based on an example of the customer-centric process *Request-to-Answer*. The Subject Interaction Diagram and the related Subject Behavior Diagram for the subject *Sales Representative* are illustrated in this paper. Our results are evaluated and confirmed by experts, possible users, and the TM Forum.

Our contribution to theory and practice is twofold. Firstly, we illustrate the integration of an inductive, bottom-up reference modeling approach with the top-down approach of an existing reference model. We contribute a concrete, practical example to the discussion of inductive reference modeling. Secondly, we have extended the well-accepted reference process model eTOM by developing a subject-oriented reference model with generic subjects and their interactions. With our sBPM extension we support

telecommunication companies to implement eTOM in a more efficient way. The sBPM extension identifies concrete subjects (functions and roles) with their respective duties and responsibilities in the selected process. However, the analysis of concrete benefits (e.g. less effort) would require further practical project implementations.

Although we have derived our reference model from real-life organizational data, the design process was a consensus-based, iterative development with the involvement of the TM Forum working group and various experts. The development could be further improved by applying heuristic methods of inductive reference modeling. As we have now convinced the TM Forum to integrate eTOM with a bottom-up approach, this could be the starting point for further research.

References

- Ardalani, P., Houy, C., Fettke, P., Loos, P.: Towards a minimal cost of change approach for inductive reference process model development. Proceedings of the 21st European Conference on Information Systems, ECIS-2013, Utrecht, Netherlands (2013).
- Augl, M., Stary, C.: Communication- and Value-Based Organizational Development at the University Clinic for Radiotherapy-Radiation Oncology. In: Fleischmann, A., Schmidt, W., Stary, C. (eds.) S-BPM in the Wild, pp. 35–53. Springer, Heidelberg (2015)
- Becker, J., Schütte, R.: *Handelsinformationssysteme*. Redline Wirtschaft, Frankfurt a. M. (2004)
- Börger, E., Fleischmann, A.: Abstract State Machine Nets: Closing the Gap Between Business Process Models and Their Implementation. In: Proceedings of the 7th International Conference on Subject-Oriented Business Process Management. ACM, New York (2015)
- Bruce, G., Naughton, B., Trew, D., Parsons, M., Robson, P.: Streamlining the telco production line. *Journal of Telecommunications Management*, 1:15–32 (2008)
- Czarnecki, C., Winkelmann, A., Spiliopoulou, M.: Reference Process Flows for Telecommunication Companies: An Extension of the eTOM Model. *Bus. Inf. Syst. Eng.* 5:83–96 (2013)
- Fettke, P., Loos, P.: Referenzmodellierungsforschung. *Wirtschaftsinf.* 46:331–340 (2004)
- Fettke, P., Loos, P.: Perspectives on Reference Modeling. In: Fettke, P., Loos, P. (eds.) *Reference Modeling for Business Systems Analysis*, pp. 1–20. IGI Global (2007)
- Fleischmann, A.: *S-BPM illustrated: a storybook about business process modeling and execution*. Springer, New York (2013)
- Fleischmann, A., Schmidt, W., Stary, C.: *S-BPM in the Wild*. Springer, Heidelberg (2015)
- Forbrig, P.: Reuse of models in S-BPM process specifications. In: Proceedings of the 7th International Conference on Subject-Oriented Business Process Management. ACM, New York (2015)
- Garon, P., Neumann, A., Bensberg, F.: Design of a Subject-Oriented Reference Model for Change

- Management. In: Nanopoulos, A., Schmidt, W. (eds) S-BPM ONE - Scientific Research. LNBIP, vol. 170, pp. 74–88. Springer, Heidelberg (2014)
- Grover, V., Saeed, K.: The telecommunication industry revisited. *Commun ACM*, 46:119–125 (2003)
- Hammer, M., Champy, J.: *Reengineering the corporation: a manifesto for business revolution*. HarperBusiness, New York (1994)
- Harmon, P.: The Scope and Evolution of Business Process Management. In: vom Brocke, J., Rosemann, M. (eds.) *Handbook on Business Process Management 1*, pp. 83–106. Springer, Heidelberg (2010)
- Hevner, A.R., March, S.T., Park, J., Ram, S.: Design Science in Information Systems Research. *MIS Quarterly*, 28:75–105 (2004)
- Holschke, O., Levina, O., Rake, J., Offermann, P.: Verbesserung der Wirksamkeit des SOA-Design durch Referenzmodelle. In: 9th International Conference on Wirtschaftsinformatik, pp. 233–242 (2009)
- Houy, C., Fettke, P., Loos, P.: Zur Evolution der Ereignisgesteuerten Prozesskette. In: Tagungsband Multikonferenz Wirtschaftsinformatik, pp. 1020–1033 (2014)
- Hübner, L.: A Service Hardware Application Case Fiducia. In: Fleischmann, A., Schmidt, W., Stary, C. (eds) *S-BPM in the Wild*, pp. 75–95. Springer, Heidelberg (2015)
- ITU: Recommendation M.3050.0: Enhanced Telecom Operations Map (eTOM). (2007)
- Kelly, M.B.: The TeleManagement Forum’s Enhanced Telecom Operations Map (eTOM). *Journal of Network and Systems Management*. 11:109–119 (2003)
- Krumeich, J., Werth, D., Loos, P.: Nutzung des Viewpoint-Konzepts zur Unterstützung kollaborativer Modellierung - Konzeption und prototypische Implementierung. In: 11th International Conference on Wirtschaftsinformatik, pp. 1261–1275 (2013)
- Lederer, M., Schott, P., Kurz, M.: Subject-Oriented Business Processes Meet Strategic Management: Two Case Studies from the Manufacturing Industry. In: Fleischmann, A., Schmidt, W., Stary, C. (eds) *S-BPM in the Wild*, pp. 13–34. Springer, Heidelberg (2015)
- Lorbacher, F.: Designing an Agile Process Layer for Competitive Differentiation. In: Fleischmann, A., Schmidt, W., Stary, C. (eds) *S-BPM in the Wild*, pp. 97–110. Springer, Heidelberg (2015)
- Malinova, M., Leopold, H., Mendling, J.: An Empirical Investigation on the Design of Process Architectures. In: 11th International Conference on Wirtschaftsinformatik, pp. 1197–1211 (2013)
- March, S.T., Smith, G.: Design and Natural Science Research on Information Technology, Decision Support Systems. (15:4), pp. 251–266 (1995)
- Markus, M.L., Majchrzak, A., Gasser, L.: A Design Theory for Systems that Support Emergent Knowledge Processes, *MIS Quarterly*. (26:3), pp. 179–212 (2002)
- Martens, A.; Fettke, P.; Loos, P. (2015): Inductive Development of Reference Process Models Based on Factor Analysis, in: Thomas. O.; Teuteberg, F. (Hrsg.): *Proceedings der 12.*

Internationalen Tagung Wirtschaftsinformatik (WI 2015), Osnabrück, S. 438-452

- Mesbahipour, R., Nursinski, A., Spiller, M.: Architecting the Enterprise along Communication Paradigm Using the TOGAF® Framework. In: Zehbold, C. (eds.) S-BPM ONE. CCIS, vol. 422, pp. 157-163. Springer, Heidelberg (2014)
- Ohlsson, J., Händel, P., Han, S., Welch, R.: Process Innovation with Disruptive Technology in Auto Insurance: Lessons Learned from a Smartphone-Based Insurance Telematics Initiative. In: vom Brocke, J., Schmiedel, T. (eds) BPM - Driving Innovation in a Digital World, pp. 85–101, Springer, Heidelberg (2015)
- Piller, C.: (2015) A Reference Model for Maintenance Processes. In: Fleischmann, A., Schmidt, W., Sary, C. (eds) S-BPM in the Wild, pp. 153–169. Springer, Heidelberg (2015)
- Pousttchi, K., Hufenbach, Y.: Value Creation in the Mobile Market: A Reference Model for the Role(s) of the Future Mobile Network Operator. *Bus. Inf. Syst. Eng.* 53:299–311 (2011)
- Richter, P., Esswein, W.: Betriebliche Prozesse und Projekte im Spannungsfeld zwischen Standardisierung und Agilität. In: Proceedings MKWI 2014, pp. 1075–1087 (2014)
- Rosemann, M.: Preparation of Process Modeling. In: Becker, J., Kugeler, M., Rosemann, M. (eds.) *Process Management*, pp. 41–78. Springer, Heidelberg (2003)
- Rummler, G.A., Ramias, A.J.: A Framework for Defining and Designing the Structure of Work. In: vom Brocke, J., Rosemann, M. (eds.) *Handbook on Business Process Management I*, pp. 83–106. Springer, Heidelberg (2010)
- Scheer, A.-W.: *Business Process Engineering: Reference Models for Industrial Enterprises*. Springer, Heidelberg (1998)
- Schmidt W., Fleischmann A., Gilbert O.: Subjektorientiertes Geschäftsprozessmanagement. *HMD – Praxis der Wirtschaftsinformation*, 52-62 (2009)
- Schmiedel, T., vom Brocke, J.: Business Process Management: Potentials and Challenges of Driving Innovation. In: vom Brocke, J., Schmiedel, T. (eds) BPM - Driving Innovation in a Digital World, pp. 3–15, Springer, Heidelberg (2015)
- Sprogies, M., Schmidt, W.: Introducing S-BPM at an IT Service Providers. In: Fleischmann, A., Schmidt, W., Sary, C. (eds) S-BPM in the Wild, pp. 55–74. Springer, Heidelberg (2015)
- TM Forum: GB921 Process Framework Concepts and Principles. (2015)
- Vom Brocke, J.: *Referenzmodellierung: Gestaltung und Verteilung von Konstruktionsprozessen*. 2. Aufl., Logos-Verlag, Berlin, Deutschland (2015).
- Yahya, B.N., Bae, H., Be, J., Kim, D.: Generating Valid Reference Business Model using Genetic Algorithm. *International Journal of Innovative Computing, Information and Control* 8, 1463-1477 (2012)

Workshop zum Stand, den
Herausforderungen und
Impulsen des Geschäfts-
prozessmanagements

Workshop zum Stand, den Herausforderungen und Impulsen des Geschäftsprozessmanagements

Michael Fellmann¹, Ralf Laue², Agnes Koschmider³, Andreas Schoknecht³,

Das Geschäftsprozessmanagement (GPM) ist mittlerweile eine etablierte Forschungsdisziplin. Um zukünftig die Relevanz der erreichten Forschungsergebnisse zu sichern und eine zielgerichtete Weiterentwicklung zu ermöglichen, müssen die Methoden und Instrumente des GPM an die sich im ständigen Wandel befindlichen Ansprüche aus Wissenschaft, Wirtschaft und Verwaltung angepasst werden. Darüber hinaus bedarf auch die Integration neuer Technologien (beispielsweise RFID, mobile Geräte, Wearable Devices, Spracherkennung und weitere Sensorik) in das GPM der weiteren Untersuchung. Mit dem Workshop soll daher eine Plattform geschaffen werden, um über die Ideen für die zukünftige Forschung im Kleinen, aber auch die Ausrichtung der GPM-Disziplin im Großen diskutieren zu können. Der Workshop möchte somit dazu beitragen, wichtige Impulse und Anregungen für die zukünftige Forschung zu liefern.

Die akzeptierten Beiträge spiegeln die große Bandbreite an möglichen Themen wider, indem sie Vorschläge zur Modellierung und Verbesserung von Geschäftsprozessen, dem Vergleich von Modellen, der Entwicklung von Referenzmodellen und schließlich auch der tatsächlichen Ausführung von Prozessen unter Nutzung neuartiger Hardware wie Datenbrillen unterbreiten.

Der Beitrag von Bergner et al. befasst sich mit der Analyse von Rückmeldungen aus sozialen Netzwerken zur Verbesserung von Geschäftsprozessen. Zwei Beiträge einer Forschungsgruppe der Universität Bayreuth (jeweils Baumann et al.) diskutieren den Begriff der Ähnlichkeit von Geschäftsprozessmodellen. Dieses Thema ist wichtig für die Wiederverwendung von Modellteilen und die Verwaltung von Modellsammlungen. Der Artikel von Rehse et al. befasst sich mit der induktiven Entwicklung von Referenzmodellen, während der Beitrag von Niemöller et al. wieder ein sehr neues Thema aufgreift: den Einsatz von Datenbrillen zur Unterstützung von Prozessabläufen.

Insgesamt kristallisieren sich trotz des großen inhaltlichen Spektrums zwei verbindende Themen über alle Beiträge heraus: Zum einen die Verwendung von Modellen und der damit implizierte, nach wie vor vorhandene hohe Stellenwert von Modellen für das Geschäftsprozessmanagement. Zum anderen die Verwendung neuer Ansätze zur Erstellung und zum Vergleich von (Referenz-)Modellen, die auf semantischen Technologien wie Sprachverarbeitung aufbauen oder neue algorithmische Verfahren einsetzen. Die Mo-

¹ Universität Rostock, Institut für Informatik

² Westsächsische Hochschule Zwickau, Fachgruppe Informatik

³ Karlsruher Institut für Technologie, Institut für Angewandte Informatik und Formale Beschreibungsverfahren

dellierungsforschung und die Verbesserung der Werkzeuge bleiben damit auch zukünftig ein dynamisches und fruchtbares Teilgebiet der Forschung zum Prozessmanagement.

Alle Einreichungen zu diesem Workshop sind durch mindestens zwei Experten begutachtet worden. Wir danken den Mitgliedern des Programmkomitees für ihre Bereitschaft zur Mitwirkung:

Peter Fettke, Deutsches Forschungszentrum für Künstliche Intelligenz Saarbrücken

Susanne Leist, Universität Regensburg

Ulrich Reimer, FHS St. Gallen

Frank J. Rump, Hochschule Emden/Leer

Kurt Sandkuhl, Universität Rostock

Werner Schmidt, Technische Hochschule Ingolstadt

Meike Ullrich, Karlsruhe Institute of Technology

Supporting Business Process Improvement with Natural Language Processing: A Model-based Approach

Martin Bergner¹ Hans-Georg Fill² Florian Johannsen³

Abstract: The research area of business process improvement (BPI) focuses on the development of methods and techniques for redesigning processes as well as improving the process performance accordingly. Thereby one particular challenge lies in the classification of customer feedback that is received via various online and offline channels. In the paper at hand we present an approach how customer feedback that is issued via social media can be classified using natural language processing. For this purpose, a domain-specific modeling method has been designed that permits to specify the processing steps in visual form and that can integrate the results in existing modeling methods for BPI. For a first evaluation, the approach has been prototypically implemented on the ADOxx-based SeMFIS platform and assessed using a SWOT analysis.

Keywords: Business Process Improvement, Natural Language Processing, Conceptual Modeling

1 Introduction

One of the current challenges in research on business process management is to provide manageable and practicable approaches that support the goal-oriented conduction of business process improvement (BPI) projects [Va16]. On the one hand, this is regarded as an academic need to close the gap between the areas of business process management and process-oriented quality management [St06]. On the other hand, corresponding support in terms of methods and tools is also strongly sought after by practitioners who are often overwhelmed by the large quantity and complexity of available techniques [JF14a, Ha15]. Besides traditional approaches such as the numerical analysis and simulation of business processes [Ba11], a wide range of methods and tools have been specifically developed for BPI (e.g., Six Sigma) [Ze11].

In a recent attempt to provide a manageable set of BPI techniques and establish a best-practice approach for the systematic conduction of BPI projects, a *roadmap for business process improvement (BPI)* has been developed [JF14a]. The roadmap guides the user during the systematic development of process improvement opportunities, starting with a first analysis of a business process in the form of Supplier-Input-Process-Output-Customer (SIPOC) diagrams, the investigation of customer feedback to a detailed analysis of the

¹ University of Vienna, Faculty of Computer Science, Waehringerstrasse 29, 1090 Vienna, martin.bergner@univie.ac.at

² University of Vienna, Faculty of Computer Science, Waehringerstrasse 29, 1090 Vienna, hans-georg.fill@univie.ac.at

³ University of Regensburg, Chair of Business Engineering, Universitätsstraße 31, 93053 Regensburg, florian.johannsen@wiwi.uni-regensburg.de

process performance by means of key performance indicators (KPIs). In the following, the emphasis is put on the investigation of customer feedback, which represents the base for the definition of project goals [Me13].

With the upcoming of social media platforms such as Twitter, Facebook, LinkedIn, or Xing [HKP12], organizations started to collect customer opinions via these channels complementing traditional sources such as CRM systems or online and offline surveys. In this regard, customer posts, e.g., on an enterprise's Facebook page, represent the so-called "Voice of the Customer (VOC)" [PNC00], capturing consumers' current attitude towards the product and service offerings or the company in general. This information provides valuable insights into customers' behavior and serves as a base for triggering marketing efforts or BPI projects [GR10]. The information in social media channels can be analyzed in real-time and is - compared to the collection of primary data - accessible at low costs. Further, customers' sentiment as reflected by the posts is highly up-to-date, eliminating the risk of receiving outdated data, which is given when analyzing secondary data exclusively (e.g., complaint reports) [Li12].

The challenge when retrieving customer feedback from social media platforms is that the provided data is largely unstructured. Especially customer utterances in the form of comments are expressed in natural language, which needs to be processed by humans before it can be used for management decisions. Therefore, the following two research questions can be derived:

1. How can support be provided and technically realized for BPI initiatives to classify natural language expressions?
2. How can this support be integrated with model-based approaches for BPI?

For providing support in classifying these natural language expressions and assigning them to BPI initiatives, a model-based approach for the machine-based processing of natural language has been designed and prototypically implemented [Be15]. With this approach it is aimed for a solution that can be adapted to multiple scenarios in BPI. Furthermore, the approach can be integrated with existing modeling methods for BPI. The addressees of the modeling method are BPI experts requiring support in managing BPI projects.

The remainder of the paper is organized as follows: In section 2 we will discuss foundations in the area of model-based BPI, social media platforms, and natural language processing. Then, in section 3 the modeling method will be described. This will include the description of the modeling language and the implementation. Subsequently, a first evaluation using a SWOT analysis is given in section 4. The paper will be concluded by an outlook on future work.

2 Foundations

Before we continue with the description of the modeling method, we first briefly outline some foundations that are necessary to understand the design of the approach.

This includes a short overview on the topic of model-based BPI, a characterization of social media platforms, and an outline of the core techniques that we re-used from the area of natural language processing.

2.1 Model-Based Business Process Improvement

The field of BPI is closely related to the process-oriented quality management discipline [St06]. Besides the domains of business management and information technology, also process-oriented quality management has for a long time been concerned with the simplification of business processes, the control of the quality of the outputs and the corresponding IT support [Ha15, St06]. Considering the multitude of BPI techniques existing [An99, Me13, HGJ06], the challenge lies in selecting the most suitable and effective techniques for supporting employees in the conduction of particular BPI projects [Ze11, JF14a]. This becomes even more important because practitioners increasingly refrain from using extensive methods for performing BPI projects (e.g., Six Sigma) and prefer a manageable set of BPI techniques instead [Da13]. Thereby, in the course of a project, it is not only essential to make use of adequate data sources for investigating the current functioning of a process but also to codify the BPI project participants' knowledge adequately [SM09].

In [JF14a] a proposal for such a set of techniques has been generated in the form of a roadmap – see figure 1. The roadmap contains 11 techniques for business process improvement that are ordered according to the phases "Define", "Measure", "Analyse", "Improve", and "Control" (DMAIC), which were derived from the Six Sigma cycle [PNC00]. For reasons of brevity we only discuss the first phase "Define" in the following and refer interested readers to the details in [JF14a].

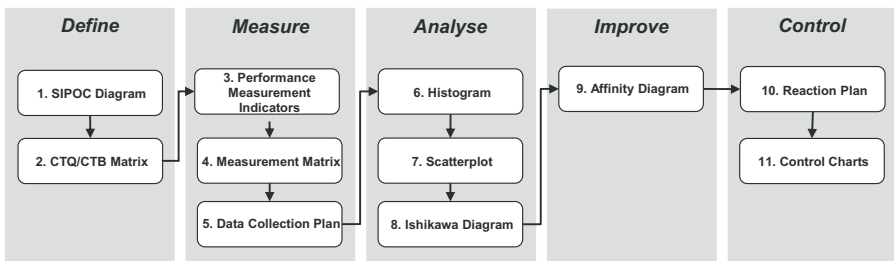


Figure 1: Roadmap for Business Process Improvement [JF14a]

The first BPI technique of the roadmap is the so-called SIPOC diagram. It gives a high-level overview of the inputs and outputs of a business process as well as the process steps and the corresponding process customers [Me13]. Subsequently, the Critical-to-Business (CTB) and Critical-to-Quality (CTQ) factors are elaborated by help of the CTQ/CTB Matrix. For this purpose, customer (VOCs) and employee requirements (Voice of the Business - VOBs), e.g., extracted from CRM systems or social media channels, are collected and condensed to core statements from which CTQ and CTB factors are derived that deter-

mine the goals of a BPI project. For instance the VOC statements "*the processing times are quite long*" and "*it took a long time until I received feedback*" can be consolidated to a core statement "*process cycle times*" from which the CTQ factor (project goal) "*reduction of process cycle times by 2 working days*" may be derived. So far, this condensation has usually been accomplished manually in BPI projects. However, considering the large amount of data processed (e.g., data retrieved from social media channels) and limited human resources in BPI projects [GR05] machine-based support is required. All BPI techniques of the roadmap, as shown, were designed as conceptual model types [JF14a] allowing to efficiently codify, communicate and process the results emerging in BPI projects. Accordingly, we speak of an approach for *model-based BPI* in that context [JF14a].

2.2 Social Media Platforms

Social media are today not only used for private communication purposes but are of increasing importance for the communication between enterprises and customers [KA14]. Via social media channels customer inquiries can be efficiently handled, marketing material widely shared or complaints solved quickly amongst others [PM12]. According to Berthon et al. social media thereby fulfill three functions [Be12]: a. the dissemination of information, b. the interaction with people, and c. the interpretation / sense-making of information. Regarding the collection of customer requirements the second and the third function are relevant. The interaction thereby includes both personal interaction with individual users, e.g., via personal messages, as well as in the form of posts and comments on social media content such as Twitter tweets or Facebook posts. The interpretation of information and sense-making refers to the discussion and reflection of phenomena by social media users.

Social media platforms support these functions by providing IT support for constructing a public or semi-public profile, connecting to other users, and exchanging information with them on various levels of interaction [BE08]. For most of these platforms it is possible to retrieve information programmatically via web interfaces. For example, the Facebook Graph API⁴ lists more than 50 root nodes as entry points for querying information. This includes aspects such as friend lists, notifications, messages, comments, or photos as well as system information such as copyright rules or debug information. Some of this information is structured, e.g., the number of likes received for an object, but most information is expressed in natural language.

2.3 Natural Language Processing (NLP)

The machine-based processing of natural language information or computational linguistics comprises a large number of different activities that are today supported by a wide

⁴ Facebook Graph API Version 2.6: <https://developers.facebook.com/docs/graph-api/reference/> accessed 04-05-2016

range of libraries [IMF13]. In the following we briefly discuss the most important concepts used in NLP. For more detailed information we refer to the descriptions in [JM00, IMF13].

One of the first steps is the *identification of the used language*. Thereby, the language is identified in the form of a score to determine the confidence of the identification algorithm.

Part-of-Speech Tagging (POS) assigns labels for grammatical classes of words, i.e., nouns, verbs, adjectives, etc. POS is the basis for resolving ambiguities, e.g., for homonyms, i.e., lexical equivalents with different meanings.

Named Entity Recognition (NER) classifies the mentions of people, organizations, locations, and other named entities. It is essential for aspects such as text summarization or the identification of contextual information.

Sentence Boundary Disambiguation stands for the splitting of sentences into smaller parts. Although this may seem trivial at first sight, the punctuation especially in German and such used in social media contexts requires specific procedures, e.g., to identify emoticons.

Lemmatization refers to the task of finding the base form of a word, i.e., the form contained in a dictionary. It is required for subsequent tasks such as query expansion and finding similar words. It shall not be mixed up with *stemming* which does not take the context of a word into account. *Word Splitting* stands for the splitting of compound words into their individual parts. This applies to both orthographic compounds (e.g., 'school-bus') as well as semantic compounds (e.g., 'bittersweet').

Query Expansion is a common technique in information retrieval for enhancing the query performance by adding additional words. A typical approach is to add synonyms and hypernyms by using a thesaurus to increase the probability of finding words with related meaning in other texts.

Term Weighting measures the occurrence of terms in a text and uses it to assign weights to the terms. The assumption is that more important terms occur more frequently. Based on these weights, comparisons between texts can be conducted to measure their similarity. This is accomplished by *Similarity and Dissimilarity Measures*, e.g., the often used cosine similarity between two vectors of term frequencies.

Another important technique that is often used in natural language processing is *Stop Word Removal*. Thereby, words that carry little semantic meaning are removed from a text before it is being processed. However, it has to be carefully chosen which words are to be removed as they are sometimes decisive for semantic meaning in combination with other words.

3 A Modeling Method for Natural Language Processing in Business Process Improvement

With the above described foundations we can now advance to the description of a modeling method for supporting natural language processing in the context of BPI. Thereby we follow a traditional design science research process in which we focus on the design and

development of the artifact in form of the modeling method [Pe07, KK02]. The goal of the modeling method is to close the gap between business requirements and technological opportunities by simplifying the processing of natural language information from social media [De08]. We therefore focused in particular on the second step of the BPI roadmap where the utterances of customers (VOCs) are classified and used for deriving project goals.

In this context, the major challenge is to assign customer utterances in social media channels to the affected business processes. For humans who are familiar with both, the business processes of a company and the way customers express feedback via social media, this can be achieved quite easily. However, for machines that have to process this information automatically this is a rather complex issue. One of the reasons for this complexity is the fact that the knowledge about the business processes is typically not accessible by machines. Although one might argue that it could be referred to business process models or internal process descriptions that document this knowledge [KJS96], the problem is the *semantic gap* between descriptions of business processes that are used by a company internally and such that customers perceive beyond company borders. This also manifests itself in analyses of the language used by customers in social media channels and that used in business process descriptions. We thus had to find a mediator between these two language worlds.

The way we addressed it was to revert to so-called 'FAQ' sections on company websites. These sections document questions that are frequently asked by customers including the answers returned by employees of that company. Although these sections do not directly refer to actual business processes of a company, the fact that they are often well maintained and structured permits to easily infer which business processes are affected by company professionals. In addition, they include enough quantity of textual information as it is required for successful natural language processing.

3.1 Design of the Modeling Method

When building NLP applications a large number of APIs, services, and platforms are available today [IMF13]. Some of these platforms include visual modeling languages for specifying the processing of natural language information by invoking distinct processing components. Two notable examples include the visual editor included in the current release of Microsoft's Azure Machine Learning Studio (AzureML)⁵ and Unitex [Pa16]. Besides several machine learning algorithms, AzureML permits to visually specify steps for natural language processing such as the pre-processing and filtering of datasets, format conversions as well named entity recognition or different methods for term frequency calculations. Unitex on the other hand is a collection of programs that have been developed for analyzing natural language texts. It is a low-level tool for processing natural language that follows a graph-based approach. Unitex graphs can handle the automatic inflection of dictionaries, the pre-processing of texts, the normalization of text automata, the search for patterns and several more [Pa16][p.119ff.].

⁵ See <https://studio.azureml.net/> last accessed 06-05-2016

For our approach we chose to take a high-level approach that can be integrated with existing enterprise modeling methods in the area of business process management and BPI. For this purpose we specified three model/diagram types as shown in the meta model in figure 2.

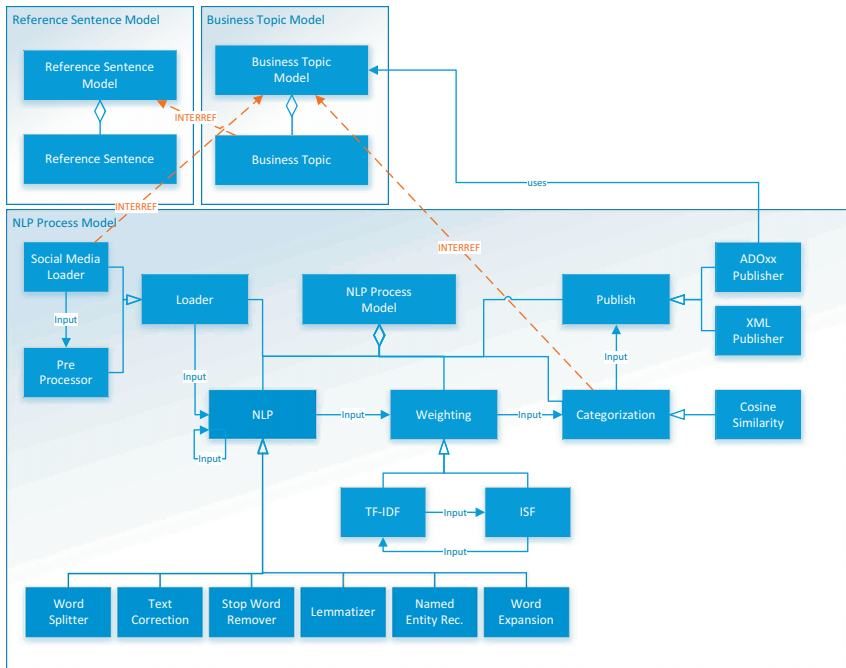


Figure 2: Meta Model Showing the Three Model Types: Reference Sentence Model, Business Topic Model, and NLP Process Model

The *business topic model* and the *reference sentence model* are used to represent categories of topics to which information from social media shall be assigned. This can for example be used to represent the headings of a company's FAQ sections together with the corresponding questions and answers given in that section. Another use case would be to define arbitrary business sections that correspond to a company's business areas and assign textual information, e.g., from the company website or other available documents.

The *NLP process model* is responsible for specifying the steps for processing natural language information. It is linked to the business topic model type to receive the information about the categories to which the natural language information shall be assigned. Furthermore, it contains the following modeling classes that a user can arrange to determine the processing steps. These classes generally follow the process of *loading* natural language data, *preparing* the data for categorization using NLP techniques, *categorizing* the data based on the user-defined categories in the business topic model, and *publishing* the results.

The *Social Media Loader* is responsible for the settings of the data loading process. It enables the user to choose between several social media platforms and to define the corresponding settings, e.g., to retrieve posts from a company's Facebook page. The *Pre Processor* class is responsible for the splitting of posts into sentences and several filterings prior to processing. Users are able to filter posts by length and language, replace abbreviations, hyperlinks, etc. The *Word Splitter* class splits compound words into individual words. The *Stop Word Remover* class handles the removing of stop words based on pre-defined stop word lists. The *Word Lemmatizer* is concerned with the the lemmatization and correction of the capitalization of words. The *Named-Entity Recognition* class is used to identify locations and persons using a corpus. The *Word Expansion* class handles the addition of synonyms to each word by using a thesaurus file in the sense of a query expansion. The *TF-IDF* weighs each term in a given document by using the TF-IDF algorithm. The *ISF* class weighs all sentences of a document, using the the inverse sentence frequency, a modified form of the Inverse Document Frequency (IDF), which is used to identify and remove unnecessary sentences in a document. The *Cosine Similarity* class categorizes the incoming natural language information depending on the prior weighting values. The *ADOxx Publisher* class is responsible for exporting an ADOxx compatible XML file, which contains the results of the categorization in a Business Topic and Reference Sentence Model. The *XML Publisher* class exports an XML file with the analysis results that can be processed by a specifically created web application called Social Media Categorization Viewer, which will be discussed in the next section.

3.2 Prototypical Implementation

The modeling method has been implemented as a prototype on the ADOxx meta modeling platform [FK13] – see the screenshot in figure 3. In order to enable later extensions for semantic processing with ontologies, it was chosen to extend the SeMFIS library [Fi16]. SeMFIS includes several model types for representing ontologies which will be used in the future to extend the descriptions used in the business topic and the reference sentence model.

In addition to the creation of visual model editors for the model types described in the previous section, several additional software components and services were implemented to realize the NLP steps. As the intended target application area of the system were German social media posts and tweets, the corresponding versions for German were selected.

For pre-processing, the Apache Commons Lang library⁶ was used, for the language detection the language detection library for Java⁷, for word splitting the jWordSplitter library by Daniel Naber⁸, for lemmatizing the mate-tools⁹, and for named entity recognition the Stanford CoreNLP toolkit [Ma14].

⁶ Apache Commons Lang: https://commons.apache.org/proper/commons-lang/download_lang.cgi last accessed 06-05-2016

⁷ <https://github.com/shuyo/language-detection/blob/wiki/ProjectHome.md> last accessed 06-05-2016

⁸ Decomposition of German compound words with jWordSplitter http://www.danielnaber.de/jwordsplitter/index_en.html last accessed 06-05-2016

⁹ <https://code.google.com/archive/p/mate-tools/> last accessed 06-05-2016

To establish the interfaces to social media platforms the RestFB library¹⁰ for Facebook and the Twitter4j library for twitter¹¹ were included.

For several of the natural language processing components external datasets and corpora are required. For the lemmatizer and the POS model we used the datasets of the mate-tools, for the stop word removal the stop word list provided on Google Code¹², for the text correction the 'incorrect word list' from the German Wikipedia¹³, for the detection of emoticons by the pre-processor the emoticon list from the English Wikipedia¹⁴, for the Named Entity Recognition the WaCky Corpus [Ba09], and for the Word Expander the thesaurus provided by Open Thesaurus¹⁵.

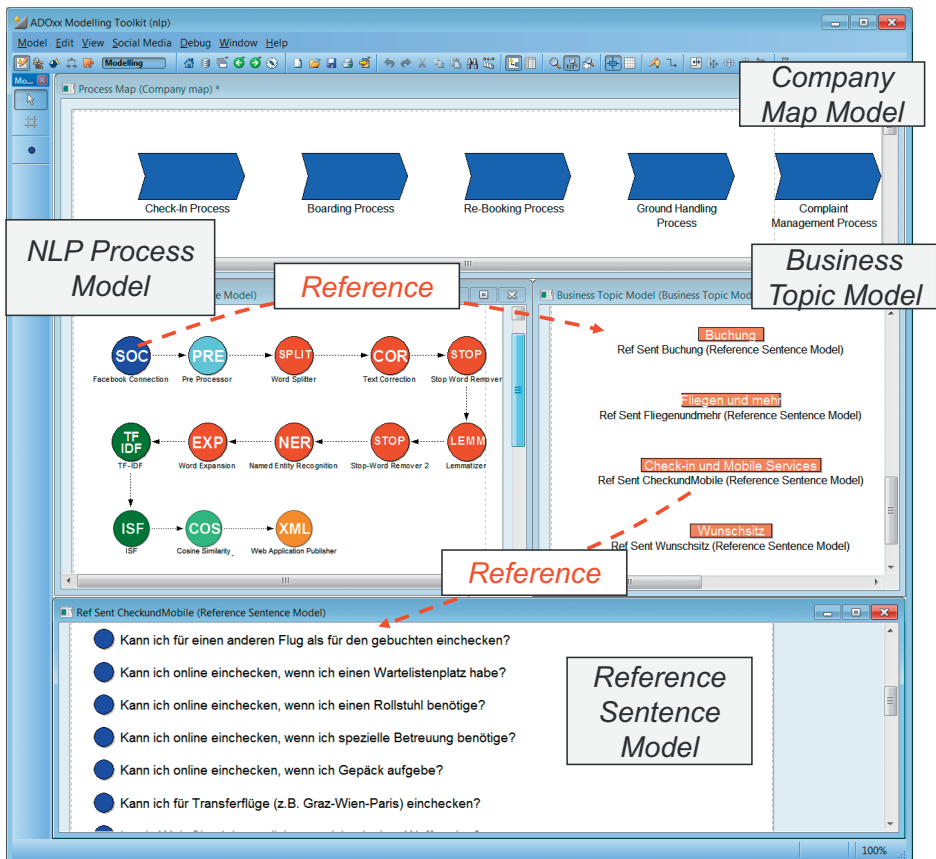


Figure 3: Screenshot of the Implementation of the Modeling Method on ADOxx

¹⁰ <http://restfb.com/> last accessed 06-05-2016

¹¹ <http://twitter4j.org/> last accessed 06-05-2016

¹² <https://code.google.com/archive/p/stop-words/> last accessed 06-05-2016

¹³ <https://de.wikipedia.org/wiki/Kategorie:Wikipedia:Falschschreibung> last accessed 06-05-2016

¹⁴ https://en.wikipedia.org/wiki/List_of_emoticons last accessed 06-05-2016

¹⁵ <https://www.openthesaurus.de/> last accessed 06-05-2016

In figure 3 at the top, a sample model of a process map is shown that is included in the standard distribution of SeMFIS and below the three new model types. The sample NLP process model shows a standard NLP process for processing posts that are retrieved via the Facebook connection. It includes steps such as pre-processing, stop word removal, lemmatization, named entity recognition, word expansion and cosine similarity calculation based on the TF-IDF and ISF measures. The final step in the process is the publication of the results via a web application. On the right hand side an example for a business topic model is shown. The model is referenced from the Social Media Loader class. The topics in the model contain references to several reference sentence models. An example is shown at the bottom. The information from these models is used in the NLP process model to assign the Facebook posts to the categories in the business topic model.

In figure 4 the web application 'Social Media Categorization Viewer' for analyzing the results of the execution of the NLP process are shown. It receives the data from the main application running on ADOxx via an XML interface. Via this application, users can inspect the assignment of the Facebook posts to the categories specified in the business topic model. The current version of the web application also includes the possibility for users to give information about the adequacy of the assignment of the posts. For this purpose, a mechanism has been added in the implementation where users can manually classify a number of posts to the defined categories. For assessing the accuracy of the automatic assignment, the standard measures for precision and recall can then be shown in addition to the classification for each category and for all posts in total.

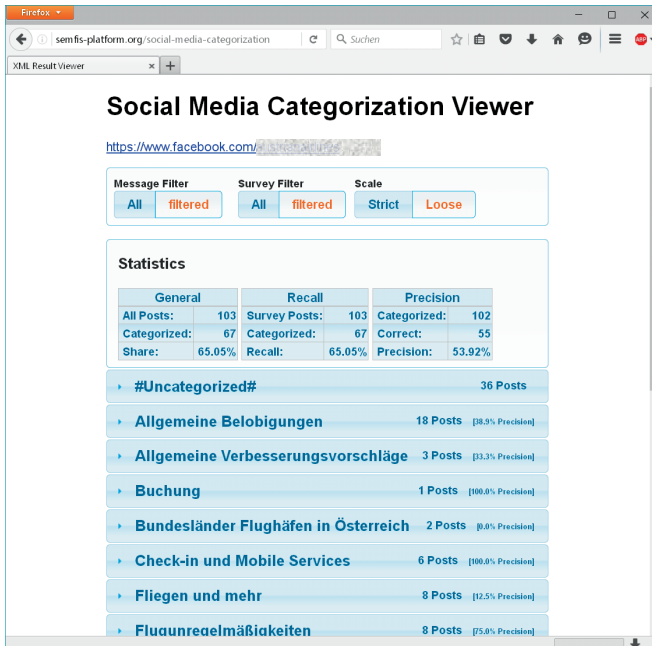


Figure 4: Screenshot of the Social Media Categorization Viewer Web Application

4 Evaluation Using a SWOT Analysis

The approach has undergone a first evaluation in the form of a SWOT analysis. Based on the application of the approach to social media data that has been retrieved from the Facebook presence of an Austrian airline company, the following strengths, weaknesses, opportunities, and threats could be assessed. More detailed evaluations including detailed user evaluations and empirical analyses are already planned and will be conducted in the near future.

The *strength* of the approach lies in the integration with existing enterprise modeling approaches. Especially for the support of BPI projects, the visual specification of topics and reference sentences is easy enough to be accomplished also by domain experts. As the technological details of the natural language processing steps are abstracted through the modeling language, users do not need in-depth technological know-how for applying the approach. The standard NLP process as shown in the sample above does not necessarily need to be modified and can thus be applied to various NLP classification tasks.

The major *weakness* of the approach is that it is currently tied to natural language processing of posts and comments in German. Although plenty of libraries exist for conducting a similar procedure in English or other languages, technical know-how about the NLP configuration process is required to achieve multi-language capabilities. However, the modeling language as implemented on ADOxx already offers extension points for this case so that for example English thesauri or stop-word lists can be assigned via changes in attribute values of the model.

An *opportunity* of the presented approach is its extensibility with further NLP processing features and its integration in further modeling methods. Besides the mentioned algorithms and NLP services, a multitude of other NLP algorithms and libraries exist. By extending the NLP process model, these can be integrated into the processing pipeline. In addition, the implementation of the modeling method based on the SeMFIS library offers the opportunity of using more formal corpora for the NLP tasks. For example, by reverting to the OWL ontology model type of SeMFIS, formal specifications between terms can be expressed that can be subsequently used by the NLP components. Regarding the integration with other modeling methods, not only the integration with methods directed towards BPI such as the RUPERT modeling tool [JF14b, JF15] become possible. Also other modeling approaches e.g., in the area of strategic or innovation management could profit from the integration with classification results from social media. Generally, regarding BPI, the automatized analysis of social media data enables to get detailed insights into those issues that negatively or positively shape customers' current sentiment (e.g., service quality, employee friendliness, etc.). This allows a firm to trigger corresponding BPI efforts in consequence. At present, the classification of customer posts mostly is done manually by the workforce (e.g., by reverting to the CTQ/CTB Matrix), which is a time-consuming and resource-intensive process. Thus, the solution introduced above represents a significant contribution for the goal-oriented conduction of BPI projects.

Although the proposed approach does not lead to any direct *threats* in our view, there are some indirect consequences that may result from it. For example, users of the NLP

processing may be tempted to rely too much on its results and may not inspect any more how these results have been achieved. Although first test runs with sample data have shown that well-configured NLP processes can lead to satisfying results, these are far from perfect - i.e., they never reach 100% accuracy in terms of precision and recall. Despite the great advances also in the area of NLP there will always be some uncertainty that users have to deal with.

5 Conclusion and Outlook

In this paper we have presented an approach for simplifying the use of machine-based natural language processing with a modeling method. The goal of the method is to support BPI by classifying in particular the utterances of customers (VOCs) in social media and by highlighting the affected business processes accordingly. The approach has undergone a first evaluation in the form of a SWOT analysis.

However, there are several aspects that need to be evaluated in more detail. These can be grouped into technical / architectural, NLP and BPI-specific aspects. Regarding technical aspects, the performance of the presented approach will have to be evaluated in detail. This affects how many posts, e.g., from Facebook or Twitter, can be retrieved and correctly classified in a given time frame. For any practical usage this is highly relevant as the amounts of information in social media channels increase dramatically every day. In this context also the architecture consisting of the meta modeling platform and the NLP components will have to be assessed. The current prototype may not yet fully exploit the optimal exchange of information between all components and will thus have to be optimized also in this aspect.

The use of the NLP components will have to undergo a separate evaluation. Here, it will need to be checked if the components that are currently used lead to the optimal performance of the overall system or whether some of them can be replaced with more powerful versions. For this purpose, a number of test runs with different test data will have to be performed. The results will then have to be compared with manually classified natural language data. This is already technically supported by the presented application so that detailed performance results on the classification will be delivered.

Finally, aspects that are specific to the domain of BPI will have to be evaluated. As BPI projects are conducted by practitioners it will need to be tested whether the approach meets their requirements or not. This concerns the usability of the tool, the embedding into BPI projects as well as the quality of the generated results. Therefore, in-depth investigations with real-life social media data from companies will have to be performed and evaluated together with practitioners. If necessary, this will lead to refinements of the modeling as well as the natural language processing approach.

The immediate next steps include the further development of the prototype to advance it to a level that can be tested with practitioners and academics alike. To permit the evaluation in the context of BPI it is also aimed for the integration of the modeling method in the RUPERT modeling toolkit that has been previously applied to BPI use cases [JF14a, JF14b].

References

- [An99] Andersen, B.: Business process improvement toolbox. ASQ Quality Press, 1999.
- [Ba09] Baroni, M.; Bernardini, S.; Ferraresi, A.; Zanchetta, E.: The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, 43(3):209–226, 2009.
- [Ba11] Barjis, J.: Enterprise Modeling and Simulation with Enterprise Engineering. *Journal of Enterprise Transformation*, 1(3):185–207, 2011.
- [BE08] Boyd, D.M.; Ellison, N.B.: Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13:210–230, 2008.
- [Be12] Berthon, P.R.; Pitt, L. F.; Plangger, K.; Shapiro, D.: Marketing meets Web 2.0, social media, and creative consumers: Implications for international marketing strategy. *Business Horizons*, 55:261–271, 2012.
- [Be15] Bergner, M.: Integrating Natural Language Processing with Semantic-based Modeling. Master's thesis, University of Vienna, 2015.
- [Da13] Davis, D.: 3rd Biennial PEX Network Report: State of the Industry - Trends and Success Factors in Business Process Excellence. Technical report, PEX Network, 2013.
- [De08] Demirkan, H.; Kauffman, R. J.; Vayghan, J. A.; Fill, H.-G.; Karagiannis, D.; Maglio, P.P.: Service-oriented technology and management: Perspectives on research and practice for the coming decade. *Electronic Commerce Research and Applications*, 7(4):356–376, 2008.
- [Fi16] Fill, H.-G.: SeMFIS: A Flexible Engineering Platform for Semantic Annotations of Conceptual Models. accepted for *Semantic Web (SWJ)*, 2016.
- [FK13] Fill, H.-G.; Karagiannis, D.: On the Conceptualisation of Modelling Methods Using the ADOxx Meta Modelling Platform. *Enterprise Modelling and Information Systems Architectures*, 8(1):4–25, 2013.
- [GR05] Gijo, E.V.; Rao, T. S.: Six Sigma Implementation - Hurdles and More Hurdles. *Total Quality Management & Business Excellence*, 16(6):721–725, 2005.
- [GR10] Gallagher, J.; Ransbotham, S.: Social Media and Customer Dialog Management at Starbucks. *MIS Quarterly Executive*, 9(4):197–212, 2010.
- [Ha15] Harmon, P.: The Scope and Evolution of Business Process Management. In (Vom Brocke, J.; Roseman, M., eds): *Handbook on Business Process Management 1*. Springer, pp. 37–80, 2015.
- [HGJ06] Hagemeyer, C.; Gershenson, J.K.; Johnson, D.M.: Classification and application of problem solving quality tools: A manufacturing case study. *The TQM Magazine*, 18(5):455–483, 2006.
- [HKP12] Heidemann, J.; Klier, M.; Probst, F.: Online social networks: A survey of a global phenomenon. *Computer Networks*, 56(18):3866–3878, 2012.
- [IMF13] Ingersoll, G.S.; Morton, T.S.; Farris, A.L.: *Taming Text - How to Find, Organize and Manipulate it*. Manning, 2013.
- [JF14a] Johannsen, F.; Fill, H.-G.: Codification of Knowledge in Business Process Improvement Projects. In: *European Conference on Information Systems (ECIS'2014)*. AIS, 2014.

- [JF14b] Johannsen, F.; Fill, H.-G.: RUPERT: A modelling tool for supporting business process improvement initiatives. In: *Proceedings of DESRIST'2014*. Springer, pp. 418–422, 2014.
- [JF15] Johannsen, F.; Fill, H.-G.: Supporting Knowledge Elicitation and Analysis for Business Process Improvement through a Modeling Tool. In: *International Conference on Business Informatics 2015*. AIS, 2015.
- [JM00] Jurafsky, D.; Martin, J.: *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Prentice Hall, 2000.
- [KA14] Kane, G.C.; Alavi, M.: What's Different About Social Media Networks? A Framework and Research Agenda. *MIS Quarterly*, 38(1):275–304, 2014.
- [KJS96] Karagiannis, D.; Junginger, S.; Strobl, R.: Introduction to Business Process Management Systems Concepts. In (Scholz-Reiter, B.; Stickel, E., eds): *Business Process Modelling*. Springer, Berlin et al., pp. 81–106, 1996.
- [KK02] Karagiannis, D.; Kuehn, H.: Metamodeling Platforms. In: *Third International Conference EC-Web 2002 â DEXA 2002*. Springer, Aix-en-Provence, France, p. 182, 2002.
- [Li12] Liu, B.: Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167, 2012.
- [Ma14] Manning, C.; Surdeanu, M.; Bauer, J.; Finkel, J.; Bethard, S.J.; McClosky, D.: The Stanford CoreNLP Natural Language Processing Toolkit. In: *52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. pp. 55–60, 2014.
- [Me13] Meran, R.; John, A.; Roenpage, O.; Staudter, C.: *Six Sigma+Lean Toolset*. Springer, 2013.
- [Pa16] Paumier, S.: *UNITEX User Manual 3.1 - March 27, 2016*. Universite Paris-Est Marne-la-Vallée, 2016.
- [Pe07] Peffers, K.; Tuunanen, T.; Rothenberger, M.A.; Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3):45–77, 2007.
- [PM12] Pinto, M. B.; Mansfield, P.: Facebook as a complaint mechanism: An investigation of millennials. *Journal of Behavioral Studies in Business*, 5, 2012.
- [PNC00] Pande, P.S.; Neuman, R.P.; Cavanagh, R.R.: *The Six Sigma Way: How GE, Motorola, and other top companies are honing their performance*. McGraw-Hill, 2000.
- [SM09] Seethamraju, R.; Marjanovic, O.: Role of process knowledge in business process improvement methodology: a case study. *Business Process Management Journal*, 15(6):920–936, 2009.
- [St06] Stracke, C.: Process-oriented quality management. In (Ehlers, U.D.; Pawlowski, J.M., eds): *Handbook on Quality and Standardisation in E-Learning*. Springer, pp. 79–96, 2006.
- [Va16] Vanwersch, Rob J. B.; Shahzad, K.; Vanderfeesten, I.; Vanhaecht, K.; Grefen, P.; Pintelon, L.; Mendling, J.; Merode, G.G.; Reijers, H.A.: A Critical Evaluation and Framework of Business Process Improvement Methods. *Business & Information Systems Engineering*, 58(1):43–53, 2016.
- [Ze11] Zellner, G.: A structured evaluation of business process improvement approaches. *Business Process Management Journal*, 17(2):203–237, 2011.

On Behavioral Process Model Similarity Matching: A Centroid-based Approach (Enlarged Abstract of [BBJ15])

Michaela Baumann¹, Michael Heinrich Baumann^{1,2}, Stefan Jablonski¹

Summary A great number of related work points out the need for similarity detection between process models, be it the management of process model repositories, compliance checking, the reuse of model parts, etc. [BL12]. Among the different aspects of process models used for measuring similarity, behavior is an important part. [BBJ15] introduces a behavioral similarity measure suitable for many-to-many correspondences (cf. [Ba14]) that is still easy to compute.

Methods The paper uses an abstract definition of process models which is transferable to many process modeling languages. In a first step, a map between the compared models is established, mapping sets of activities to sets of activities. An example mapping is given in Fig. 1 indicated with different patterns of the activities.

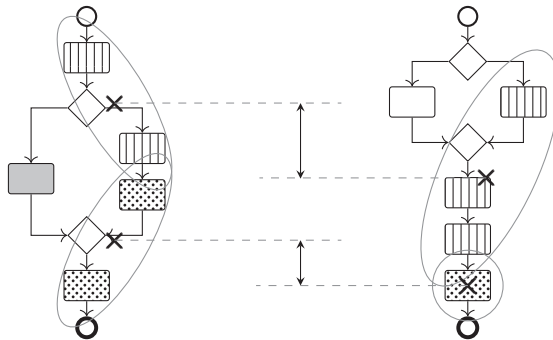


Fig. 1: Schematical representation of the comparison of (positional) centroids for mapped sets of process tasks.

In a second step, three behavioral features are determined for each activity: relative position, repeatability, and optionality. The relative position of activity a is calculated as the length of the shortest path from start to a divided by the length of the shortest path from start to end crossing a . For the grey activity in Fig. 1 this would be $3/6 = 0.5$. Repeatability and optionality are boolean. In the example, all activities are not repeatable but some are optional. For each mapped set and for each feature these values are averaged. In Fig. 1, the positions of the mapped sets, that are the averaged positions of the respective activ-

¹ University of Bayreuth, Faculty of Mathematics, Physics & Computer Science, {firstname.surname}@uni-bayreuth.de

² The work of M. H. Baumann is supported by a scholarship from Hanns-Seidel-Stiftung which is funded by Bundesministerium für Bildung und Forschung.

ities are indicated with an \times , called centroid, each. The differences of a centroid and its image then contribute to the similarity: the lower the differences in average, the more similar the process models. An extension is given to penalize inhomogeneous activity sets via variances.

To validate the approach, a comparison with causal footprints (CF) [Di11], a widely accepted technique for measuring behavioral similarity, and “smallest” CF [BL12] is applied. All approaches show the same trend but the centroid-based (CB) approach shows greater variations: “similar” models are rated more similar and “dissimilar” models are rated more dissimilar than by CF and smallest CF. The major distinction is the computation effort of the approaches: the CB approach scales linearly with the number of activities whereas the smallest CF method scales quadratically and the CF method exponentially.

Discussion In order to evaluate the unpenalized CB approach against common sense of (behavioral) similarity, we carried out another study.³ For five reference models we provided three alternative models, each, and asked modeling experts for the most similar and the most dissimilar one. The results are presented in Tab. 1.

		1-1	1-2	1-3	2-1	2-2	2-3	3-1	3-2	3-3	4-1	4-2	4-3	5-1	5-2	5-3
	CB	94	61	89	83	100	75	85	89	48	65	91	97	99	93	98
survey	sim	55	9	36	0	100	0	9	91	0	36	0	64	27	0	73
	dissim	0	82	18	9	0	91	18	9	73	64	36	0	27	46	27

Tab. 1: Unpenalized CB data and survey data rounded and in percent. CB values are all between 0 and 100 while the survey values sum up to 100.

For the first four reference models, the expert judgements support the CB approach. The fifth model, however, does not fit exactly. Nonetheless, the CB approach shows satisfying results in case studies. It does neither consider causal dependencies nor execution traces but position, repeatability, and optionality. The CB approach is a tradeoff between simplification and calculation effort.

References

- [Ba14] Baumann, Michael Heinrich; Baumann, Michaela; Schöning, Stefan; Jablonski, Stefan: Towards Multi-perspective Process Model Similarity Matching. In: EOMAS, volume 191 of LNBP, pp. 21–37. Springer Berlin Heidelberg, 2014.
- [BBJ15] Baumann, Michaela; Baumann, Michael Heinrich; Jablonski, Stefan: On Behavioral Process Model Similarity Matching: A Centroid-based Approach. In: ICCGI 2015, volume 5, pp. 125–131. thinkmind, 2015.
- [BL12] Becker, Michael; Laue, Ralf: A comparative survey of business process similarity measures. *Computers in Industry*, 63(2):148–67, 2012.
- [Di11] Dijkman, Remco; Dumas, Marlon; van Dongen, Boudewijn; Käärik, Reina; Mendling, Jan: Similarity of business process models: Metrics and evaluation. *Information Systems*, 36(2):498 – 516, 2011.

³ The corresponding material is available under

https://docs.google.com/forms/d/1VGwvJ7RrC08gQs9rvHEqsiBdBfhG5gt86-fwmqnUE1I/viewform?c=0&w=1&usp=mail_form_link [2016-04-11]

Ansätze zum Ähnlichkeitsabgleich von deklarativen Geschäftsprozessmodellen

Michaela Baumann¹, Michael Heinrich Baumann^{2,3}, Lars Ackermann¹, Stefan Schönig¹, Stefan Jablonski¹

Abstract: Prozesse in Organisationen lassen sich grundsätzlich in zwei Klassen unterteilen: Strikte Routineprozesse und flexible Prozesse mit einem sich dynamisch entwickelnden, variablen Kontrollfluss. Aufgrund der unterschiedlichen Natur strikter und flexibler Prozesse entstanden auch zwei verschiedene Klassen von Prozessmodellierungssprachen: imperative und deklarative. Während imperative Modelle die exakte, schrittweise Reihenfolge der Aufgaben in einem Prozess beschreiben, formulieren deklarative Modelle lediglich Regeln, welchen die Prozessausführung folgen muss. Aufgrund immer größer werdender Prozessmodellsammlungen ist es notwendig, unabhängig vom Modellierungsansatz ähnliche Modelle zu erkennen und zusammenzuführen, was mit dem Begriff Ähnlichkeitsabgleich bezeichnet wird. Für imperative Prozessmodelle existiert bereits eine Vielzahl von Methoden zur Lösung dieses Problems. Da es für deklarative Prozessmodelle jedoch keine allgemein gültigen Lösungen gibt, werden in der vorliegenden Arbeit grundlegende Möglichkeiten des Ähnlichkeitsabgleichs diskutiert.

Keywords: Geschäftsprozessmanagement, Prozessmodellähnlichkeit, Deklarative Prozessmodelle

1 Einleitung

Das Geschäftsprozessmanagement (GPM) stellt eine etablierte Methode zur Modellierung, Ausführung und Analyse von Prozessen dar [Du13]. Prozesse in Organisationen lassen sich grundsätzlich in zwei Klassen unterteilen [Ja94]: strikte Routineprozesse mit eindeutig vorgeschriebenem und vorab spezifizierbarem Kontrollfluss und flexible, entscheidungsintensive Prozesse mit einem sich dynamisch entwickelnden und von Fall zu Fall unterschiedlichen Kontrollfluss. Folgerichtig entstanden auch zwei Paradigmen von Prozessmodellierungssprachen: Imperative und deklarative. Erstere fokussieren das exakte Beschreiben der Reihenfolge der Arbeitsschritte in einem Prozess. Deklarative Modelle hingegen formulieren lediglich Regeln für die Prozessausführung, wobei grundsätzlich jede Ausführung, die keine Regel verletzt, valide ist. Derartige Regeln können die Reihenfolge der Prozessschritte, aber auch die Abhängigkeiten derselben von Daten, Personen und Werkzeugen repräsentieren. Während sich strikte Prozesse häufig besser mittels imperativer Modellierungssprachen (z.B. BPMN) repräsentieren lassen, eignen sich für flexible Prozesse meist eher deklarative Notationen (z.B. CMMN) [Pi12].

¹ Universität Bayreuth, Institut für Informatik, Universitätsstraße 30, D-95447 Bayreuth, {michaela.baumann, lars.ackermann, stefan.schoenig, stefan.jablonski}@uni-bayreuth.de

² Universität Bayreuth, Institut für Mathematik, Universitätsstraße 30, D-95447 Bayreuth, michael.baumann@uni-bayreuth.de

³ Die Arbeit von Michael Heinrich Baumann wird durch ein Begabtenstipendium der „Hanns-Seidel-Stiftung e.V. (HSS)“ aus Mitteln des „Bundesministeriums für Bildung und Forschung (BMBF)“ gefördert.

Aufgrund der wegen Mehrfachmodellierungen immer größer werdenden Sammlungen von Prozessmodellen wird es zunehmend schwieriger, diese zu verwalten. Mehrfachmodellierung kann dabei unter anderem durch verschiedene Versionen desselben Prozessmodells im Rahmen des Business Process Reengineering [TG12], zielgruppenspezifische Modellanpassungen im Zuge der Internationalisierung [Di09] oder Unternehmensfusion verursacht werden. Zur Verwaltung großer Prozesssammlungen ist es notwendig, ähnliche Modelle zu erkennen und zusammenzuführen, was mit dem Begriff Ähnlichkeitsabgleich bezeichnet wird [Di11]. Ein solcher Abgleich dient daneben auch der Wiederverwendbarkeit von Teilprozessen sowie dem Vergleich mit Referenzprozessen (Konformitätsprüfung) und Vorgaben (Complianceprüfung) [BL12]. Die größte Herausforderung für einen solchen Abgleich ist die Heterogenität in Konzeption (Granularität der Aktivitäten) und Realisation (Repräsentation des Verhaltens, Formulierung der Beschriftungen und verwendete Terminologie) der abzugleichenden Modelle [WDM10].

Im Gegensatz zu imperativen Modellen existieren für deklarative Prozessmodelle bislang jedoch kaum Lösungen für den Ähnlichkeitsabgleich, bzw. die vorhandenen können nur sehr fallspezifisch angewendet werden. Ausgehend von einer Einführung zu deklarativen Prozessmodellen stellt die vorliegende Arbeit einen Überblick über mögliche Verfahren des Ähnlichkeitsabgleichs für deklarative Prozessmodelle dar.

2 Deklarative Prozessmodelle

Fahland et al. [Fa09] sehen den wesentlichen Unterschied zwischen imperativen und deklarativen Prozessmodellen darin, wie ermittelt werden kann, ob ein gegebenes Verhalten konform bezüglich des Prozessmodells ist oder nicht. Im Falle der imperativen Modelle muss ein Verhalten mittels eines Pfades durch das Modell rekonstruiert werden können. Im Gegensatz dazu wird bei deklarativen Prozessmodellen geprüft, dass zu keinem Zeitpunkt eine der enthaltenen Regeln verletzt worden ist. Ein deklaratives Prozessmodell ist immer gegeben durch eine Menge an Aktivitäten und eine Menge an Regeln, die die Ausführung der Aktivitäten bedingen. Diese Regeln können beispielsweise in linearer temporaler Logik oder Prädikatenlogik formuliert sein [Fa09]. Wir verwenden im Folgenden keine festgelegte deklarative Modellierungssprache, um die Idee unseres Ansatzes so einfach wie möglich darzustellen. Ein Modell, bestehend aus den Aktivitäten A , B und C , könnte beispielsweise folgende Regeln umfassen: „ A muss mind. einmal und kann max. zweimal ausgeführt werden“, „ B muss genau einmal ausgeführt werden“, „ C darf max. einmal ausgeführt werden“, „ B kann nur dann ausgeführt werden, wenn A bereits ausgeführt wurde“, „Wenn C ausgeführt wird, muss auch B später noch ausgeführt werden“

3 Stand der Wissenschaft

Für imperative Prozessmodelle werden in der Literatur zahlreiche Möglichkeiten für einen Modellabgleich vorgeschlagen. Zusammenfassungen und Auswertungen verschiedener Ansätze finden sich beispielsweise in [Di11]. Die Methoden reichen dabei von strukturellen, graphbasierten Ansätzen über labelbasierte Ansätze bis hin zu verhaltensbasierten

Ansätzen [BBJ15, Di09, LR15]. Informationen, die verwendete Dokumente, zugewiesene Agenten und Rollen oder zu benutzende Werkzeuge betreffen, können ebenfalls für einen Ähnlichkeitsabgleich herangezogen werden [Ba14, WDM10].

Alle Perspektiven, die weder Struktur noch Verhalten betreffen, z. B. Label, können für deklarative wie für imperative Modelle gleich, oder zumindest ähnlich, behandelt werden. Folglich sind im Weiteren vor allem Verhalten und Struktur von Interesse. Für die Ähnlichkeitsbestimmung zweier deklarativer Prozessmodelle können zwei Ansätze der Literatur entnommen werden. Zum einen ist es in manchen Fällen möglich, deklarative Prozessmodelle in imperative zu überführen [PDCM14], für die dann die oben erläuterten Methoden angewendet werden können. Wie in [GH01] gezeigt, ist es auch möglich, bestimmte deklarative Prozessmodelle auf endliche Zustandsautomaten abzubilden und dann einen Editierabstand (Graph-Edit-Distance) zu berechnen [Wo06]. Da in einem Zustandsautomat jedoch Zustände, und nicht Aktivitäten, die Knoten darstellen, ist eine Abwandlung der Abbildung von Aktivitäten auf eine Abbildung von Zuständen erforderlich. Auch ist zu beachten, dass deklarative Modelle, die Variablen benutzen, nicht zwingend in einen endlichen Automaten überführt werden können [ZSJ14]. Es müssen also Ansätze auch für solche Prozessmodelle gefunden werden.

4 Ansätze zum Ähnlichkeitsabgleich

In deklarativen Modellen wird der Kontrollfluss implizit über Regeln eingeschränkt. Letztere bilden demnach in deklarativen Modellen die Grundlage für einen Ähnlichkeitsabgleich für die Verhaltensperspektive. Die Ähnlichkeitsmaße der anderen Perspektiven sollen analog zu denen imperativer Prozessmodelle übernommen werden. Wie für imperative Prozessmodelle wird auch für deklarative Prozessmodelle eine Abbildung M , die zwei Modelle G_1 und G_2 aufeinander abbildet, festgelegt. Für die im Folgenden aufgeführten Ideen zur Ähnlichkeitsberechnung wird zunächst eine 1:1-Abbildung angenommen. Das Finden der besten Abbildung M^* erfolgt dann über einen Optimierungsansatz und wird hier nicht weiter behandelt. Die Ansätze müssen dahingehend bewertet werden, als dass sie Ähnlichkeitswerte liefern, die mit dem „Common Sense“ übereinstimmen, aber auch in der Praxis berechenbar sind.

4.1 Ähnlichkeitsabgleich via Ausführungspfade und Konformitätsprüfung

Die erste vorgestellte Möglichkeit geht ähnlich vor, wie es bei Process-Mining-Verfahren üblich ist, um die Fitness von Logdaten bezüglich der erzeugten Modelle zu bestimmen [RvdA06]. Es werden simulierte, im besten Fall alle möglichen, Ausführungspfade für ein Modell verwendet und dann mit dem zweiten Modell abgeglichen, d. h. auf Ausführbarkeit im zweiten Modell getestet. Da in deklarativen Prozessmodellen jedoch prinzipiell auch unendlich lange Ausführungspfade möglich sind, kann über Simulation nur eine endliche Menge an möglichen, endlichen Ausführungspfaden erzeugt werden. Dieses Thema der Automated Sequence Generation findet sich beispielsweise in [Ha12]. Die Re-

präsentativität des Ergebnisses der Ähnlichkeitsprüfung hängt von der Menge an simulierten Pfaden ab. Im Gegenzug muss auch verglichen werden, ob Ausführungspfade, die in einem Modell nicht erlaubt sind, im anderen auch nicht erlaubt sind. Insgesamt ergäben sich also vier Gruppen von Pfaden (erlaubt-erlaubt, erlaubt-nicht erlaubt, nicht erlaubt-erlaubt, nicht erlaubt-nicht erlaubt). Als Ähnlichkeitsmaß bietet sich der Quotient aus gleichartigen und allen Pfaden an. Sind alle diese Zahlen endlich, kann das Maß einfach berechnet werden, sind die Anzahlen jedoch unendlich, existiert lediglich ein Grenzwert. Es kann dabei vorkommen, dass dieser Grenzwert immer aus $\{0,1\}$ ist.

4.2 Ähnlichkeitsabgleich via Verhaltensmuster und Wahrheitstabellen

Die zweite Möglichkeit ist eine Approximation der Methoden aus 4.1. Es werden zunächst für die abgebildeten Aktivitäten vorher festgelegte Verhaltensmuster bestimmt, die eine bestimmte Anzahl an Aktivitäten umfassen. Bei bis zu 3-stelligen Mustern für die Aktivitäten A und B ergeben sich beispielsweise folgende Muster, wobei nur positive Zusammenhänge berücksichtigt werden: $-, A, B, AA, AB, \dots, BBB$. Es wird dann mit Hilfe der Regelmengen beider Prozessmodelle jedem Muster ein Wahrheitswert zugewiesen, d. h., es wird geprüft, ob das Muster im jeweiligen Modell so ausführbar ist. Ein einzelnes A würde in diesem Fall bedeuten, dass A bzw. $M(A)$ irgendwann einmal ausgeführt wird und davor und danach weder A noch B (bzw. $M(A)$, $M(B)$) ausgeführt werden. Es wird hier zunächst nicht unterschieden, ob die Ausführung direkt oder irgendwann nacheinander erfolgen muss. Sind die Wahrheitswerte für die Muster in beiden Modellen bestimmt, werden sie abgeglichen: Ist ein Muster in beiden Modellen erlaubt oder in beiden Modellen nicht erlaubt, wird ihm der Wert 1, ist ein Muster in einem Modell erlaubt, im anderen nicht, der Wert 0 zugewiesen. Diese Werte werden addiert und durch die Anzahl der Muster geteilt, was einen Mittelwert im Intervall $[0,1]$ liefert. Erlauben beide Modelle die gleichen Muster, ist der Wert 1. Je weniger Muster übereinstimmen, desto mehr tendiert der Mittelwert zu 0. Solch ein Mittelwert wird für je zwei abgebildete Aktivitäten berechnet. Für eine Abbildung mit drei abgebildeten Aktivitäten werden also drei Mittelwerte gebildet, die noch einmal gemittelt werden. Die wichtigste Frage, die sich für diese Möglichkeit stellt, ist die, welche Verhaltensmuster für den Abgleich verwendet werden sollen, um ein möglichst objektives und verallgemeinerbares Ergebnis zu erhalten. Ein Vorteil dieser Methode ist, dass Transitivität in den Regelmengen berücksichtigt wird.

4.3 Ähnlichkeitsprüfung via Prädikatenlogik und Regelähnlichkeit

Eine weitere Möglichkeit, die sich im Zusammenhang mit einem Ähnlichkeitsabgleich von zwei Prozessmodellen ergibt ist die, die Regeln der Prozessmodelle auf Ebene der zugrunde liegenden Logik direkt miteinander zu vergleichen. Es stellt sich die Frage, ob aus der Logik bekannte Erfüllbarkeits- bzw. Entscheidbarkeitskriterien für eine Ähnlichkeitsbestimmung benutzt werden können, wie zum Beispiel in [Bi92] oder [SS94] für statistische Lernverfahren verwendet. Nach einer ersten Einschätzung ergibt sich für die Regelmengen deklarativer Prozessmodelle, zumindest dann, wenn sie Aussagen über komplexere

Zusammenhänge erlauben, dass sie unentscheidbar werden. Der Vergleich kann (gegeben entscheidbare Regelmengen) so erfolgen: Kann eine Regel aus G_1 aus den Regeln von G_2 abgeleitet werden, wird sie mit 1 bewertet, sonst mit 0, und andersherum. Dann werden diese Werte aufsummiert und durch die Anzahl der Regeln geteilt. Ein Problem stellt die Anzahl der Regeln dar, da Regelmengen, die zu einem gleichen Verhalten führen, nicht identisch, also vor allem nicht gleich mächtig, sein müssen. Das heißt, sowohl Dividend als auch Divisor in diesem Ähnlichkeitsmaß sind nicht eindeutig bestimmt.

5 Schluss und Ausblick

Die vorliegende Arbeit zeigt einige Ansätze und Schwierigkeiten beim Ähnlichkeitsabgleich deklarativer Prozessmodelle auf. Da die bisher in der Wissenschaft verfügbaren Techniken nicht für alle flexiblen Prozesse verallgemeinerbar sind, werden neue Ansätze vorgestellt, die jedoch noch weiter untersucht werden müssen. Nach dem derzeitigen Stand der Forschung scheint die Methode *Verhaltensmustervergleich* am praktikabelsten und vielversprechendsten, sowohl von der Konzeption als auch der technischen Umsetzbarkeit. Jedoch sind zunächst alle Ansätze auszuarbeiten und zu evaluieren. Aufgrund mangelnder Abgleichsansätze für deklarative Prozessmodelle, muss für einen Referenzwert zunächst auf das Urteil von Modellierungsexperten für eine Evaluation zurückgegriffen werden. Die vorgestellten Ansätze sollen dann im Weiteren mit Ähnlichkeitswerten anderer Perspektiven, beispielsweise den Aktivitätenbeschreibungen [LR15] oder den Agenten und Ressourcen [Ba15], kombiniert werden.

Literaturverzeichnis

- [Ba14] Baumann, M. H. et al.: Towards Multi-perspective Process Model Similarity Matching. In: Enterprise and Organizational Modeling and Simulation, Jgg. 191 in LNBIIP, S. 21–37. Springer Berlin Heidelberg, 2014.
- [Ba15] Baumann, M. et al.: Resource-Aware Process Model Similarity Matching. In (Toumani, Farouk; et al., Hrsg.): Service-Oriented Computing - ICSOC 2014 Workshops, Jgg. 8954 in Lecture Notes in Computer Science, S. 96–107. Springer International Publishing, 2015.
- [BBJ15] Baumann, M.; Baumann, M. H.; Jablonski, S.: On Behavioral Process Model Similarity Matching: A Centroid-based Approach. In: ICCGI 2015, S. 125–131. IARIA, 2015.
- [Bi92] Bisson, G.: Learning in FOL with a similarity measure. In: Proc. of the Nat. Conf. on Artificial Intelligence. S. 82–82, 1992.
- [BL12] Becker, M.; Laue, R.: A comparative survey of business process similarity measures. Computers in Industry, 63(2):148–67, 2012.
- [Di09] Dijkman, R. et al.: Aligning Business Process Models. In: Int. Enterprise Distributed Object Computing Conf., S. 45–53. IEEE, 2009.
- [Di11] Dijkman, R. et al.: Similarity of business process models: Metrics and evaluation. Information Systems, 36(2):498 – 516, 2011.

- [Du13] Duman, M. et al.: *Fundamentals of Business Process Management*. Springer, 2013.
- [Fa09] Fahland, D. et al.: Declarative versus Imperative Process Modeling Languages: The Issue of Understandability. In: *Enterprise, Business-Process and Information Systems Modeling*, Jgg. 29 in LNBIP, S. 353–366. Springer Berlin Heidelberg, 2009.
- [GH01] Giannakopoulou, D.; Havelund, K.: Automata-based verification of temporal properties on running programs. In: *16th Annual Int. Conf. on Automated Software Engineering (ASE)*. S. 412–416, 2001.
- [Ha12] Hallé, S. et al.: A Logical Approach to Data-Aware Automated Sequence Generation. In: *Transactions on Computational Science XV*. Springer Berlin Heidelberg, S. 192–216, 2012.
- [Ja94] Jablonski, S.: MOBILE: A modular workflow model and architecture. In: *Proc. of Int. Working Conf. on Dynamic Modelling and Information Systems*. Citeseer, 1994.
- [LR15] La Rosa, M. et al.: Detecting approximate clones in business process model repositories. *Information Systems*, 49:102 – 125, 2015.
- [PDCM14] Prescher, J.; Di Ciccio, C.; Mendling, J.: From Declarative Processes to Imperative Models. In: *SIMPDA*. S. 162–173, 2014.
- [Pi12] Pichler, P. et. al.: Imperative versus Declarative Process Modeling Languages: An Empirical Investigation. In: *Business Process Management Workshops*. Springer Berlin Heidelberg, S. 383–394, 2012.
- [RvdA06] Rozinat, A.; van der Aalst, W. M. P.: Conformance Testing: Measuring the Fit and Appropriateness of Event Logs and Process Models. In: *Business Process Management Workshops*. LNBIP, Springer Berlin Heidelberg, S. 163–176, 2006.
- [SS94] Sebag, M.; Schoenauer, M.: A rule-based similarity measure. In: *Topics in Case-Based Reasoning: First European Workshop, EWCBR-93*. Springer Berlin Heidelberg, S. 119–131, 1994.
- [TG12] Tka, M.; Ghannouchi, S. A.: Comparison of Business Process Models as Part of BPR Projects. *Procedia Technology*, 5:427 – 436, 2012.
- [WDM10] Weidlich, M.; Dijkman, R.; Mendling, J.: The ICoP Framework: Identification of Correspondences between Process Models. In: *Advanced Information Systems Engineering*, Jgg. 6051 in LNCS, S. 483–498. Springer Berlin Heidelberg, 2010.
- [Wo06] Wombacher, A.: Evaluation of Technical Measures for Workflow Similarity Based on a Pilot Study. In: *On the Move to Meaningful Internet Systems (OTM)*, Jgg. 4275 in LNCS, S. 255–272. Springer Berlin Heidelberg, 2006.
- [ZSJ14] Zeising, M.; Schönig, S.; Jablonski, S.: Towards a common platform for the support of routine and agile business processes. In: *Int. Conf. on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*. S. 94–103, 2014.

Inductive Reference Model Development: Recent Results and Current Challenges

Jana-Rebecca Rehse, Philip Hake, Peter Fettke, and Peter Loos¹

Abstract: Reference modeling offers attractive benefits for both research and practice. The inductive strategy for reference model development derives reference models by generalizing individual enterprise models. It has recently gained attention in research, however, its practical application still faces numerous challenges. The objective of the article at hand is to identify recent contributions to the field of inductive reference model development and use them to analyze the current challenges that impede the application of their results in practice. We identify a total of 18 contributions, either scientific articles describing inductive methods for reference model development, or practical reports describing the concrete development of a reference model for a certain domain. They are all analyzed by means of a six-stage-framework for reference model development. For each stage, we derive specific challenges and point out acknowledgments and potential solutions.

Keywords: Inductive Reference Model Development, Reference Modeling, Grand Challenges

1 Introduction

Reference modeling offers attractive benefits for both research and practice [FL07]. Following the epistemologically established differentiation between rationalism and empiricism as two fundamental ways of cognition, reference modeling differentiates a deductive and an inductive strategy for reference model development [BS97]. Model development according to the deductive strategy employs generally accepted theories and principles, while the inductive strategy is based on the generalization of individual enterprise models. It focuses on the commonalities of the individual models and abstracts from specific features. Hence, deductive development proceeds from the general towards the specific (“top-down”), whereas inductive development evolves from the specific into the general (“bottom-up”).

Although several concepts, methods, and tools for the support of reference modeling exist by now, practical development of reference models still faces a variety of challenges. The open questions range from project preparation, over individual steps of pre-processing and derivation, to maintenance and continuous improvement of a reference model. Hence, the development process of a reference model today is often barely structured, nontransparent, and only marginally justified and thus hardly communicable in terms of course and characteristics. In other words, the ideal of a generally repeatable, engineering-style approach is usually not yet reached. With that, a tool support is also only possible in a limited way.

¹Institute of Information Systems at the DFKI, Campus D3.2, 66123 Saarbruecken, Jana-Rebecca.Rehse@iwi.dfki.de, Philip.Hake@iwi.dfki.de, Peter.Fettke@iwi.dfki.de, Peter.Loos@iwi.dfki.de

The article at hand intends to pave the way for a structured, methodical, repeatable, and thus justifiable practical application of inductive reference modeling techniques. Therefore, we analyze recent contributions to the field of inductive reference model development and point out the current challenges that impede the application of their results in practice. We consider both scientific articles focusing on the underlying methodology and reports describing the concrete inductive development of a domain-specific reference model in practice, intending to identify as many practical challenges as possible.

The challenges as presented in the following are comparable to the currently discussed Grand Challenges of Mertens and Barbian [MB15]. While they discuss Grand Challenges of Business and Information Systems Engineering (BISE) or “Wirtschaftsinformatik” in general, the contribution at hand is limited to the specific challenges of inductive reference modeling. Although width and relevance of the challenges discussed here are much narrower in the first place, we also intend to fruitfully stimulate future research. Like Mertens and Barbian, we do not assume our list to be complete, but rather a first attempt towards a more structural examination of inductive reference model development.

The article is structured as follows. After the introductory section, important foundations of inductive reference modeling are described in Section 2. Section 3 outlines the identification of challenges by means of a literature review. The identified challenges are grouped and described in section 4. Section 5 critically assessed our findings, before concluding the article with an outlook on further research.

2 Foundations of Inductive Reference Model Development

The research agenda for reference modeling can be outlined by two main questions:

1. Questions regarding Reference Models: How to design a reference model for a certain application domain? A reference model can capture the current organizational context (descriptive interpretation) or make a suggestion for an innovative organizational context (prescriptive interpretation).
2. Questions regarding Reference Modeling Methods: How to develop and apply reference models? Methodological questions can examine characteristics and performance features of known methods, refine existing methods, or invent and evaluate completely new methods.

Both research questions can be viewed regarding the inductive development strategy. Answers concerning the current state of research are exemplified by the following aspects:

1. Reference Models: Several works report on the inductive development of a reference model ([AFF11, KPR08, DM05], q.v. the attribute “construction method” of the reference models listed in the catalog at <http://rmk.iwi.uni-sb.de>).

2. **Reference Modeling Methods:** Known procedure models for reference modeling indicate that existing individual enterprise models as well as further sources of knowledge should be identified and considered within the context of reference modeling [Be02, Th06]. In addition, there are several contributions that suggest methodological approaches for different aspects of inductive reference model development [Ar13, GvJV08, LRW09, RFL13].

Although the research agenda for inductive reference modeling has made a certain progress, there exists a variety of challenges, which have not yet been systematically addressed. In order to provide an extensive and systematic assessment of challenges, we follow the method for inductive reference model development, as defined by Fettke [Fe14]. It consists of six consecutive stages, each identifying and describing a central task necessary for the practical development of a reference model. The stages are shown in Fig. 1 and shortly outlined in the following.



Fig. 1: Stages of Inductive Reference Modeling [Fe14]

1. **Preparation of Reference Model Development:** Before a reference model can be developed, several preparatory decisions and assumptions have to be made regarding the context of the reference model development. Among other things, this stage includes identifying stakeholders' requirements and defining modeling conventions.
2. **Collection of Individual Models:** The inductive development strategy is based on individual enterprise models as obligatory input data. These models are collected in this stage. Depending on the size of the reference model domain, the collection may be split into several sub-stages, identifying classes of models and representative organizations before collecting the actual models.
3. **Preprocessing of Individual Models:** After the individual models have been collected, they have to be aligned and harmonized before they can be used for the derivation of a reference model. This stage includes checking and establishing necessary modeling conventions as well as identifying correspondences.
4. **Acquisition of the Reference Model:** This stage describes inductive reference model development in a narrower sense. The preprocessed individual models are used to derive a reference model, using an arbitrary similarity measure and construction approach. Depending on the number of input models, it may make sense to split them into several model clusters and develop an individual reference model for each cluster.
5. **Postprocessing of the Reference Model:** As the previous stage is often fully automated and, even if not, limited by the contents of the input model data, the acquired reference model should be manually postprocessed in this stage. This may include

connecting unconnected model parts, adding, deleting or renaming nodes, or complementing the reference model by deductively developed model parts.

6. Evaluation and Enhancement of the Reference Model: A reference model will be especially useful if it is regularly evaluated and replenished as necessary. New individual models have to be included, as they arise. If the process domain changes significantly, it might make sense to develop a completely new reference model as well.

3 Identification and Analysis of Relevant Literature

3.1 Literature Review

In order to identify practical challenges to the inductive development of reference models, the first step is to identify relevant contributions to the field, i.e. literature that possibly describes such challenges. Therefore, we conduct a literature review, following the methodology as proposed by Fettke [Fe06]. After defining the objective and research question of our literature review, we conduct the literature search by entering pre-defined keywords in several research databases. The obtained results are assessed regarding their relevance to the posed research question. Thereby, we focus on contributions that concretely describe an inductive method, as opposed to methodically open ones (e.g. [Th06]). While those contributions may generally apply to the inductive strategy, they do not report on any concrete challenges, hence not providing answers relevant to our research question. In this assessment, we follow the Model of System Design [FHL10]: The respective contribution has to describe a technique, i.e. a reliable means to realize a certain goal, and explicate its foundation on empirical knowledge in order to be classified as inductive.

For conducting a literature review, a set of keywords has to be defined. Since the term “Inductive Reference Model Development” is not consistently used, several different keyword combinations assure that as much literature as possible is covered. We combine the term “reference model” with different verbs describing the process of automatic model development. Each verb can appear in either its infinitive form (“develop”), as a gerund (“developing”), or as the corresponding noun (“development”). Overall, there are eight keyword combinations, as listed in Table 1. The search is limited to title, keywords, and abstract of the articles. If those indicate a relation to reference modeling the full text is assessed. We conduct the search in five databases, namely Academic Search Complete (ASC, via EBSCO)², AIS Electronic Library (AISeL)³, Business Source Premier (BSP, via EBSCO)⁴, io-port⁵, and Scopus⁶.

After accumulating all relevant literature, we conduct a backward search, analyzing the citations of the identified articles to account for prior contributions to consider. Finally,

² <http://search.ebscohost.com/>

³ <http://aisel.aisnet.org/>

⁴ <http://search.ebscohost.com/>

⁵ <http://www.io-port.net>

⁶ <http://www.scopus.com>

we use *Google Scholar* to conduct a forward search, finding articles that cite the articles identified in the previous steps. As reference modeling is traditionally a topic of the German-speaking research on information systems, we also conduct an equivalent literature search using German keyword combinations. As the chosen databases mainly contain English articles, this search does not yield any significant results. However, we include German-language contributions in our forwards and backwards search to yield a better coverage of the literature.

Search Terms	ASC		AISeL		BSP		io-port		Scopus	
	All	Rel.	All	Rel.	All	Rel.	All	Rel.	All	Rel.
“reference model” AND inductive	1	0	7	2	0	0	0	0	11	2
“reference model” AND mining	10	1	8	0	8	1	19	4	163	3
“reference model” AND automatic	25	0	3	0	8	0	0	0	391	0
“reference model” AND (generation OR generate OR generating)	118	1	25	0	45	1	43	3	920	5
“reference model” AND (construction OR construct OR constructing)	72	0	52	0	30	0	45	1	692	0
“reference model” AND (development OR develop OR developing)	278	0	188	2	194	0	137	0	2859	4
“reference model” AND (derivation OR derive OR deriving)	59	0	32	1	19	0	26	1	355	0
“reference model” AND (discovery OR discover OR discovering)	38	1	2	0	20	0	18	1	142	3

Tab. 1: Quantitative Results of the Literature Review

The quantitative results of the literature search are displayed in Table 1. For each keyword combination and each database, we list the overall number of search results (in column “All”) and the number of relevant search results (in column “Rel.”). The relevance of an article is assessed manually, based on its title, keywords, and abstract. While the number of search results differs considerably, the number of relevant results is consistently very small. The large number of insignificant results is caused by the fact that, although relevant to the field of reference modeling, many of the used search terms (e.g. “construction” or “generation”) can be used in completely different contexts. In addition, many research disciplines use reference models as a scientific tool or methodology.

Accumulating all relevant articles and eliminating duplicates results in a total of ten contributions that we consider relevant for our research question. Performing a forwards and backwards search yields an additional eight articles, resulting in a total of 18 contributions, as listed in table 2. Of these 18, three [AFF11, GS14, KPR08] have the character of a practical experience report on inductive reference modeling, whereas the remaining 15 take on a rather scientific point of view.

3.2 Analysis of Contributions

In order to identify challenges to inductive reference model development, we analyze each contribution for challenges that are either explicitly mentioned as such, or implicitly addressed by according measures. For example, a paper may require previously defined activity correspondences in form of identical descriptions (implicit), while others use dedicated techniques for correspondence identification (explicit). The results of this assessment are shown in table 2. Explicit references are symbolized by a full circle (●), implicit references by a half-full circle (◐). An empty circle (○) means that the respective challenge is not mentioned in this contribution. All challenge have to be explicitly mentioned at least once. In order to provide a better overview and to point out relations, each challenge is allocated to one of the six stages of reference modeling. An in-depth description of each challenge is contained in the following section 4.

Challenge		[AFF11]	[Ar13]	[GvJV08]	[GS14]	[KPR08]	[LRW09]	[LRW10]	[LRW11]	[MFL14]	[MFL15]	[MG13]	[PiŠK12]	[RFL13]	[RFL15]	[WFL13]	[YB11]	[Ya12]	[YWB12]
1	Indication Criteria	○	○	○	○	○	○	○	●	○	○	○	●	○	○	●	○	○	○
	Modeling Language	○	●	◐	●	●	●	●	●	●	●	◐	●	○	●	○	○	○	○
	Modeling Conventions	○	●	◐	●	●	●	●	●	○	◐	◐	○	○	○	●	●	●	●
2	Model Character	◐	○	○	◐	●	○	○	○	●	○	○	○	○	●	○	○	○	○
	Model Access	◐	○	●	◐	●	○	◐	◐	○	○	○	○	◐	○	●	○	○	○
	Collection Effort	◐	○	◐	◐	◐	○	◐	◐	○	○	○	○	◐	○	●	○	○	○
3	Modeling Language	◐	○	○	○	◐	◐	◐	◐	○	○	○	○	○	○	●	○	◐	◐
	Modeling Conventions	◐	○	○	○	◐	◐	◐	◐	○	○	○	○	○	○	○	●	◐	◐
	Correspondences	●	●	●	●	●	○	◐	●	●	●	○	●	○	○	●	◐	○	○
4	Abstraction	○	●	○	◐	●	○	○	○	◐	○	○	○	●	●	●	●	○	○
	Model Clustering	●	○	○	◐	○	○	○	○	○	●	○	○	○	○	●	●	○	○
	Decomposition	○	○	○	●	●	○	○	○	○	○	○	○	◐	○	○	○	○	○
	Model Variants	◐	○	●	○	●	◐	●	●	○	○	○	○	◐	○	●	○	○	○
	Contortion	○	○	○	○	○	○	○	○	○	●	○	○	○	○	●	○	○	○
	Inductive Fallacy	●	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	○	○
Complexity	○	●	◐	○	○	●	●	●	●	●	○	○	○	○	●	○	●	●	
5	Additions	○	○	○	◐	●	○	○	○	○	○	○	○	○	○	●	●	○	○
	Model Connection	○	○	○	◐	●	○	○	○	○	◐	○	○	○	○	●	●	○	○
6	Evaluation	●	○	○	◐	●	○	○	○	○	○	○	○	●	○	○	○	○	○
	New Models	○	○	○	◐	○	○	○	●	○	○	○	○	○	○	○	○	○	○

Tab. 2: Foundation of Identified Challenges in Literature

The results of our assessment allow for several interesting insights. First of all, none of the challenges is particular to one specific contribution. Each of them is mentioned at least twice, although sometimes implicitly, meaning that the respective contribution describes measures to address this challenge, without explicating why or how it influences

the reference model development. Most of these implicit references appear in case studies or empirical evaluations, such as [AFF11] or [GS14]. Rather epistemological challenges, such as the problem of input data contortion or inductive fallacies are rarely mentioned. Instead, most contributions focus on practical or technical challenges, such as reaching a feasible algorithmic complexity or ways to establish correspondences between individual models in terms of a matching. While these challenges are unquestionably important for the practical application of the inductive strategy, epistemological aspects should also be considered, as they play an important role for evaluating the resulting reference model.

Regarding the six stages of reference modeling, it is apparent that stage four, concerned with the derivation of the reference model, contains the most challenges and is addressed by every single contribution. As this stage contains the inductive reference model development in a narrower sense, it is the core of most inductive development approaches and hence poses more challenges that have to be addressed. The preparatory first stage is also well-covered, although this is mostly caused by assumptions and restrictions to the input data. Nevertheless, this critical assessment is important for developing new approaches to reference modeling, which can potentially overcome these restrictions. On the contrary, both the collection stage (two) and the postprocessing and evaluation stages (five and six) are only sparsely covered by the contributions. While it is possible to develop a reference model without postprocessing and evaluating it, the collection of the input models is central to the success of the model development itself, as the contents of the reference model is directly influenced by the contents and character of the input models. It is remarkable that the challenges of this stage are only addressed by about half the contributions and mostly in an implicit way. Many contributions assume to have access to a sufficiently large and representative collection of digitally represented individual models, which simplifies most research activities, but is a non-trivial problem in practical applications.

Regarding the contributions, it is apparent that some take on a much broader view on inductive reference model development than others. Especially the practical reports on inductive reference model development [AFF11, GS14, KPR08] cover many challenges that are not considered by other, rather methodical contributions. Among these, the degree of examining potential challenges differs considerably. While some contributions try to take a rather hollistic view on reference model development [RFL15, LRW11], others focus on one particular challenge, which is analyzed in depth (for example input model abstraction in [RFL13]). Other contributions have a rather technical view on reference model development, focusing on algorithmic aspects and technical details (e.g. [MG13, YB11]).

4 Current Challenges of Inductive Reference Model Development

4.1 Preparation of Reference Model Development

Before beginning to collect data, several preliminary questions have to be answered. Essential challenges are:

- **Indication Criteria:** It is indisputable that according individual models have to be available to enable inductive reference model development. If, during the preparatory stage, it is unclear whether an adequate amount of individual models exist or can be easily acquired, the entire project is at risk. Hence, we require criteria determining when to resort to an inductive development and when to avoid it. A well-founded empirical knowledge base is a crucial success factor for the inductive strategy [WFL13]. Moreover, for organizational reasons, the development of a new reference model might not always be supported [LRW11].
- **Choice of Modeling Language:** The choice of the reference modeling language in the context of the inductive strategy is important for two reasons. First of all, it makes sense to follow the individual models in order to avoid unnecessary conversions and transformations. Therefore, algorithmic approaches as described in Martens et al. [MFL14] or Yahya et al. [YWB12] require a formal mostly graph-based representation of the individual models and the resulting reference model. Second, due to the number and variety of the individual models to process, it is often necessary to represent the model variants in an appropriate way. Hence, the target language should provide concepts for variant management as described in [GvJV08].
- **Modeling Conventions:** In general, modeling conventions are regarded to be important. However, due to the plurality of modeling contexts, it cannot be assumed that modeling conventions are always factually enforced when creating the input models. Hence, according compromises have to be found. Especially the algorithmic solutions we identified address this challenge, since they rely on characteristics such as unique labels [LRW09] or block-structuredness of the input process models [Ar13].

4.2 Collection of Individual Models

Individual enterprise models are an obligatory input and thus a necessary requirement for applying the inductive development strategy. They can either be collected originary (primary collection)[PIŠK12] or existing information can be reused (secondary collection) [GvJV08, AFF11]. Essential challenges are:

- **Representative Character of the Individual Models:** If the representative character of the individual models is questionable, there is a risk of contortion. For instance, particular model variants may be overrated in the derivation of the reference model. Meanwhile, recognizing such contortions requires to know the entire model domain. In some cases, however, this domain can only reasonably be assessed after the reference model has been described. In other applications, it may make sense to purposefully avoid a representative character, for instance to examine contrasting individual models. In this case, the question arises how many contrasting cases are required. Hence, Karow et al. [KPR08] require a rich empirical database and an inductive validation of the reference model. Pajk et al. [PIŠK12] address this challenge by first developing the model on the empirical data available and then assessing it according to predefined requirements.

- **Access to Individual Models:** According to Karow et al. [KPR08], accessing enterprise models can be difficult to impossible. Individual models can be subject to confidentiality or organizations see the risk to loose competitive advantages by allowing them to circulate. Hence, cooperation cannot be expected when applying the inductive development strategy. Therefore, the proof of concept for the method proposed in [KPR08] is settled in the field of public administration and it is based on public domain models.
- **Collection Effort:** The effort necessary for collecting the individual models can overcompensate the benefits of reference modeling. This risk is especially apparent, if cost and benefits are very differently distributed among the stakeholders [RFL15].

4.3 Preprocessing of Individual Models

Before concrete methods for reference model acquisition can be applied, a general preprocessing of the individual models appears to be useful. Essential challenges are:

- **Transformation of the Modeling Language:** If the individual models are not represented in a uniform modeling language, it is suitable to transform and convert them into a uniform target language. This is crucial for the algorithmic approaches we obtained, since the processing relies on a uniform representation, for example as an Event-Driven Process Chain [RFL15]
- **Examination and Establishment of Modeling Conventions:** If certain conventions have been determined, it makes sense to examine the models for meeting them. If deviations exist, according corrections of the individual models are required. It may also make sense to reconstruct the factually used conventions [WFL13].
- **Identification of Correspondences:** Single constructs in different individual models can be equivalent, similar, or different. It is essential to know about such correspondences. The need for and the challenges of identifying these correspondences is addressed in most of the approaches that aim at an algorithmic solution [MFL15, Ar13, RFL15]. The correspondences can be derived from solely structural characteristics of the input models (for example similar graph-theoretic model features [RFL15]) or be related to semantic aspects (for example similar process traces or words in node labels [Ar13, MFL15, MFL14]). The fact that structural similarities are neither necessary nor sufficient for semantic similarities, is an additional challenge.

4.4 Acquisition of the Reference Model

The aim of the reference model acquisition stage is to derive a reference model from a set of individual models. Essential challenges are:

- **Abstraction and Generalization:** Methods for abstracting specific model features are required to generalize the contents of the individual models. If the individual models contain commonalities, they are to be represented in the reference model. This way, the reference model contains the typical structures of the application domain. Hence, the acquired reference model should abstract from specific features of individual models. In analogy to the identification of correspondences described above, not only the model structure, but also the model semantics should be considered. In [RFL13], a variety of abstraction approaches are evaluated in terms of applying them in the context of inductive reference model development.
- **Clustering of Individual Models:** Developing model clusters is important for two reasons. First, individual models can stem from completely different contexts, so a cluster analysis can be used to identify possible domains [WFL13](for instance for cross-sectional tasks like accounting or logistics). Second, Aier et al. [AFF11] postulate a cluster analysis in order to group a variety of model variants within the same domain.
- **Decomposition of Individual Models:** For the decomposition, two aspects should be considered. First, individual models often consist of numerous partial models, which should be analyzed separately [GS14]). Second, even a single partial model may be decomposed into fragments. This makes sense, because two models can be different, while individual model fragments may be similar or even identical.
- **Plurality of Model Variants:** When acquiring a reference model, two opposing design objectives have to be balanced. One extreme case includes all individual models as model variants in the input model, the other extreme case only captures the commonalities of the individual models in the reference model and abstracts from individual model features. Aier et al. [AFF11] describe this challenge as a fundamental conflict between standardization and individualization of the reference model.
- **Handling Contortions:** If the collected individual models do not constitute a representative sample of the domain, it is to be checked whether possible contortions exist and can be corrected by taking according measures. Contortions can be prevented by selecting a representative or sufficiently large set of organizations [RFL15].
- **Risk of Inductive Fallacy:** Philosophy in general and (inductive) logic in particular often refer to the weaknesses of inductive conclusions [AFF11]. This aspect should also be considered in the inductive development strategy. The fact that models may contain descriptive as well as prescriptive content also has to be taken into account [RFL13].
- **Algorithmic Complexity:** There exists a variety of algorithms for model analysis that include techniques for comparing process models. These techniques include graph-based structural comparisons [YB11, Ya12, YWB12] as well as comparisons that deal with the processing and comparison of natural language [Ar13, MFL15]. Due to their high runtime and memory consumption, such algorithms are unable to cope with larger model sets as they typically appear within the inductive development strategy.

4.5 Postprocessing of the Reference Model

Essential challenges when postprocessing a reference model are:

- **Additions:** The inductive approach of Grger and Schumann [GS14] includes an enhancement of the inductively derived reference model. Reference models can be expected to be complemented by additional contents. The question arises, how to combine inductive and deductive strategies into a hybrid development strategy.
- **Connection of Partial Models:** It should be considered to group separate reference models. Grouping makes sense, if the individual models were accordingly decomposed in a previous step [GS14, KPR08]. Adding such connections is also useful to point out similarities between reference models of different domains (for example warehousing processes in retail and industry).

4.6 Enhancement of the Reference Model

Reference modeling is especially useful if the reference model is constantly evaluated and enhanced. Otherwise, there is a risk of obsolescence. Essential challenges are:

- **Evaluation:** In this context, it is important to assess the generated reference model with regard to the predefined goals. Depending on the specific stakeholder or perspective, the verification and validation varies. While Aier et al. [AFF11] conduct interviews with process users, Pajk et al. [PIŠK12] propose a simulation or even an empirical evaluation by applying the reference model.
- **New Individual Models:** If new variants are derived from a reference model, they should be considered to be included in the reference model. Reference model stability should be aimed for, i.e. not every variant should be included. However, it is obviously sensible to include standard adaptations, which are regularly executed during an application, into the reference model. It might be desirable to constantly enhance the reference model to minimize its distance to all existing variants [LRW11].

5 Discussion and Conclusion

The objective of the article at hand was to identify a catalog of potential challenges to the practical application of the inductive development strategy for reference modeling. This catalog is meant to stimulate further research by providing a list of empirically grounded challenges, which can be addressed in a structured way to advance the inductive strategy both in research and practice. However, we do not claim our list of challenges to be complete or exhaustive. For example, none of the contributions discusses the possibility for a regular re-design of the reference model, which could become necessary, if the input models change considerably. In future research, we intend to validate and extend the identified

challenges by conducting expert interviews before dealing with each challenge individually. Interviewing reference modeling experts can offer a complementary perspective and lead to additional insights from a practitioner's point of view.

Our contribution in this article can only be a starting point for the in-depth analysis of the challenges to inductive reference modeling. The term "reference model" is mainly popular in German-speaking countries and Australia, while other communities prefer terms such as "reusable model". Extending our literature review by those terms can further advance our findings. Meanwhile, we cannot assure that additional challenges are mentioned in previous works on reference modeling, not focusing on the inductive strategy. However, it is almost impossible to assess their relevance without taking explicit features of the inductive strategy, such as the foundation on empirical knowledge, into account.

It should also be noted that none of the identified contributions contains a direct reference towards Business Process Management in particular, although it appears self-evident that inductive reference modeling offers interesting potentials for classical BPM topics, such as continuous process improvement, customization and implementation of standard software, or process documentation. Traditional applications of reference modeling, such as design, optimization, simulation, or certification of business processes may also be addressed by inductive methods. Simultaneously, many identified challenges are related to topics that are already addressed by BPM researchers. For example, advancements on naming conventions, model enhancement, or variant management could be built upon to deal with the according challenges.

Despite the numerous challenges presented in this article, the inductive development strategy has considerable advantages when compared to the deductive strategy. Analyzing individual models and their potentials for generalization is not only intellectually interesting, but appears almost obligatory if the models are easily available. Second, a direct comparison does not require general principles and theories to derive a reference model. Moreover, an inductively developed reference model can be expected to have a higher degree of detailing, maturity, and acceptance. It thus appears to be especially fruitful to advance the integration of inductive and deductive development into a hybrid strategy. In general, inductive reference modeling is still at a rather early stage of development and there is a lot of potential for further research.

In addition, methods of inductive reference modeling offer interesting potentials for related applications. For example, model comparison, model integration, or analysis of internal model variants come to mind. Besides these practical applications, there are also interesting potentials for an application in research. Inductive methods can for example be used for a scientifically well-grounded comparison of models. They are hence an important foundation for one of the Grand Challenges identified by Mertens and Barbian [MB15]: Following the Human Genome Project, the "Wirtschaftsinformatik" should chart existing IT systems in order to draw conclusions for the future employment of IT and the accuracy of its systems. Such activities may significantly benefit from inductive methods and thus discover new potentials for known works on organizational typology and the model of shell and nucleus.

References

- [AFF11] Aier, Stephan; Fichter, Michael; Fischer, Christian: Referenzprozesse empirisch bestimmen : von Vielfalt zu Standards. *Wirtschaftsinformatik & Management : WuM*, 3(3), 2011.
- [Ar13] Ardalani, Peyman; Houy, Constantin; Fettke, Peter; Loos, Peter: Towards a Minimal Cost of Change Approach for Inductive Reference Model Development. In: *Proceedings of the 21st European Conference on Information Systems. European Conference on Information Systems (ECIS-13)*. AIS, 2013.
- [Be02] Becker, Jörg; Delfmann, Patrick; Knackstedt, Ralf; Kuroпка, Dominik: Konfigurative Referenzmodellierung. In (Becker, J.; Knackstedt, R., eds): *Wissensmanagement mit Referenzmodellen: Konzepte für die Anwendungssystem- und Organisationsgestaltung*, pp. 25–144. Springer, 2002.
- [BS97] Becker, Jörg; Schütte, Reinhard: Reference Information Systems for Retail: Definition, Use and Recommendations for Design and company-specific Adaption of Reference Models. In: *Wirtschaftsinformatik*, pp. 427–448. Springer, 1997. In German.
- [DM05] Daun, Christine; Matheis, Thomas: Constructing a reference process model for E-government. In (Mosca, R., ed.): *Proceedings of the 7th International Conference on The Modern Information Technology in the Innovation Processes of the Industrial Enterprises (MITIP)*, Genua. pp. 10–14, 2005.
- [Fe06] Fettke, Peter: State-of-the-Art des State-of-the-Art. *Wirtschaftsinformatik*, 48(4):257–266, 2006.
- [Fe14] Fettke, Peter: Eine Methode zur induktiven Entwicklung von Referenzmodellen. In (Kundisch, Dennis; Suhl, Leena, eds): *Tagungsband Multikonferenz Wirtschaftsinformatik 2014*. Universität Paderborn, 2014.
- [FHL10] Fettke, Peter; Houy, Constantin; Loos, Peter: On the relevance of design knowledge for design-oriented business and information systems engineering. *Business & Information Systems Engineering*, 2(6):347–358, 2010.
- [FL07] Fettke, Peter; Loos, Peter: Perspectives on Reference Modeling. In (Fettke, Peter; Loos, Peter, eds): *Reference Modeling for Business Systems Analysis*, pp. 1–20. Idea Group Publishing, 2007.
- [GS14] Gröger, Stefan; Schumann, Matthias: Entwicklung eines Referenzmodells für die Gestaltung des Drittmittel-Prozesses einer Hochschule und Ableitung von Einsatzgebieten für Dokumenten- und Workflow-Management-Systeme. *Arbeitsbericht der Universität Göttingen*, 1, 2014.
- [GvJV08] Gottschalk, Florian; van der Aalst, Wil; Jansen-Vullers, Monique: Mining Reference Process Models and Their Configurations. In: *On the Move to Meaningful Internet Systems: OTM 2008 Workshops*. Springer, pp. 263–272, 2008.
- [KPR08] Karow, Milan; Pfeiffer, Daniel; Räckers, Michael: Empirical-Based Construction of Reference Models in Public Administrations. In: *Multikonferenz Wirtschaftsinformatik*. pp. 1613–1624, 2008.
- [LRW09] Li, Chen; Reichert, Manfred; Wombacher, Andreas: Discovering Reference Models by Mining Process Variants Using a Heuristic Approach. In (Dayal, Umeshwar; Eder, Johann; Koehler, Jana; Reijers, Hajo, eds): *Business Process Management*, volume 5701 of *Lecture Notes in Computer Science*, pp. 344–362. Springer Berlin Heidelberg, 2009.

- [LRW10] Li, Chen; Reichert, Manfred; Wombacher, Andreas: The MinAdept clustering approach for discovering reference process models out of process variants. *International Journal of Cooperative Information Systems*, 19(03n04):159–203, 2010.
- [LRW11] Li, Chen; Reichert, Manfred; Wombacher, Andreas: Mining business process variants: Challenges, scenarios, algorithms. *Data & Knowledge Engineering*, 70(5):409–434, 2011.
- [MB15] Mertens, Peter; Barbian, Dina: Researching “Grand Challenges”. *Business & Information Systems Engineering*, 57(6):391–403, 2015.
- [MFL14] Martens, Alexander; Fettke, Peter; Loos, Peter: A Genetic Algorithm for the Inductive Derivation of Reference Models Using Minimal Graph-Edit Distance Applied to Real-World Business Process Data. In (Kundisch, Dennis; Suhl, Leena; Beckmann, Lars, eds): *Tagungsband Multikonferenz Wirtschaftsinformatik 2014*. Universität Paderborn, 2014.
- [MFL15] Martens, Alexander; Fettke, Peter; Loos, Peter: Inductive Development of Reference Models Based on Factor Analysis. In (Thomas, Oliver; Teuteberg, Frank, eds): *Proceedings der 12. Internationalen Tagung Wirtschaftsinformatik (WI 2015)*. volume 12, Universität Osnabrück, Osnabrück, Germany, pp. 438 – 452, 2015.
- [MG13] Mejri, Asma; Ghannouchi, Sonia Ayachi: Discovering Reference Process Models in the Context of BPM Projects. *Procedia Technology*, 9:489–497, 2013.
- [PIŠK12] Pajk, Dejan; Indihar-Štemberger, Mojca; Kovačič, Andrej: Reference model design: An approach and its application. In: *Information Technology Interfaces (ITI)*, Proceedings of the ITI 2012 34th International Conference on. IEEE, pp. 455–460, 2012.
- [RFL13] Rehse, Jana-Rebecca; Fettke, Peter; Loos, Peter: Eine Untersuchung der Potentiale automatisierter Abstraktionsansätze für Geschäftsprozessmodelle im Hinblick auf die induktive Entwicklung von Referenzprozessmodellen. In (Alt, Rainer; Franczyk, Bogdan, eds): *Proceedings of the 11th International Conference on Wirtschaftsinformatik*, Leipzig, Germany. 2013. In German.
- [RFL15] Rehse, Jana-Rebecca; Fettke, Peter; Loos, Peter: A graph-theoretic method for the inductive development of reference process models. *Software & Systems Modeling*, tba:1–41, 2015.
- [Th06] Thomas, Oliver: *Management von Referenzmodellen: Entwurf und Realisierung eines Informationssystems zur Entwicklung und Anwendung von Referenzmodellen*. Logos-Verlag, 2006.
- [WFL13] Walter, Jürgen; Fettke, Peter; Loos, Peter: How to Identify and Design Successful Business Process Models: An Inductive Method. In (Becker, Jörg; Matzner, Martin, eds): *Promoting Business Process Management Excellence in Russia - Proceedings and Report of the PropelleR 2012 Workshop*. volume 15. European Research Center for Information Systems, Münster, pp. 89–96, 2013.
- [Ya12] Yahya, Bernardo; Bae, Hyerim; Bae, Joonsoo; Kim, Dongsoo: Generating valid reference business process model using genetic algorithm. *International Journal of Innovative Computing, Information and Control*, 8(2):1463–1477, 2012.
- [YB11] Yahya, Bernardo; Bae, Hyerim: Generating Reference Business Process Model Using Heuristic Approach Based on Activity Proximity. In: *Intelligent Decision Technologies*, pp. 469–478. Springer, 2011.
- [YWB12] Yahya, Bernardo; Wu, Jei-Zheng; Bae, Hyerim: Generation of Business Process Reference Model Considering Multiple Objectives. *Industrial Engineering & Management Systems*, 11(3):233–240, 2012.

Shaping the Future of Mobile Service Support Systems – Ex-Ante Evaluation of Smart Glasses in Technical Customer Service Processes

Christina Niemöller¹, Dirk Metzger¹, Michael Fellmann², Deniz Özcan¹ and Oliver Thomas¹

Abstract: The recent introduction of smart glasses through media spurred new concepts of service support systems. Especially in the domain of Technical Customer Services (TCS), the opportunity to access information hands-free provides additional benefits. However, due to the novelty of the technology besides various technical issues the question of usefulness is of vital importance. To date, little research provides guidance for researchers and practitioners. Goal of this contribution is to systematically elicit functionalities of smart glasses to offer an overview of the emerging technology and apply the functionalities to requirements of the TCS to investigate the benefits of smart glasses in this domain. Therefore, a multi-method approach for the elicitation of 20 features has been conducted followed by mapping the features to 36 previously derived requirements. In total, the contribution can serve as guidance for the strategic evaluation of smart glasses supporting service processes.

Keywords: Smart Glasses, Augmented Reality, Technical Customer Services, Taxonomy.

1 Introduction

Due to recent technological advancements in hardware and software, the character of IT transformed from an attached commodity to the center of new product and service ideas [BLM14]. Especially services whose work is performed in the field, such as Technical Customer Service (TCS) processes, may profit from mobile devices [AH10]. Mobile services enable technicians to directly capture information more quickly and more precisely, while they are performing maintenance, inspection or repair tasks [Le11]. So far, mobile solutions mostly have been implemented in the form of handheld devices, however, handheld devices are not the likely end of the mobile computing innovation vector [Lu13]. Emerging technologies such as smart glasses are discussed to support complex service processes [HE15].

Although the potential induced by traditional head-worn-displays has already been discussed in research in the field of maintenance [HM05, HF09, LP98, RBH03], still only

¹ Universität Osnabrück, Informationsmanagement und Wirtschaftsinformatik, Katharinenstr. 3, 49074 Osnabrück, vorname.nachname@uni-osnabrueck.de

² Universität Rostock, Wirtschaftsinformatik insbes. Betriebliche Informationssysteme, Albert-Einstein-Str. 22, 18057 Rostock, michael.fellmann@uni-rostock.de

few scientific papers exist that address smart glasses. Especially in the professional context, such wearable devices meet the need of workers that are not able to use handheld devices due to the nature of their work, e.g. when free or clean hands are mandatory [LH14]. There is a need of studying emerging technologies such as smart glasses in a business context, since manufacturers tend to focus predominantly on consumer markets and requirements. Therefore, in our paper, we study the impact of smart glasses on technical customer services. With this, we aim to lead the emerging technology to its full potential in business processes and as a consequence to foster service innovation.

The scope of our work is a structured identification and classification of functionalities of smart glasses as to date there is no typology known from literature. However, to conduct an ex-ante evaluation of smart glasses in the context of TCS which we envision in the paper at hand, such a typology is needed. The rationale for this ex-ante evaluation is that an early evaluation and documentation makes it easier for researchers to communicate intermediate products of a Design Science Research (DSR) process to the research community. With this, not only consensus on the relevance, novelty, and importance of a chosen problem domain can be build early in DSR processes [SB12]. Other researchers can build upon this rigorously developed knowledge right away, without waiting for the instantiation of the IS prototype itself. This is why we validated the potential usage of smart glasses for TCS processes through our approach.

As a result, we hope to both provide a solid foundation for further research activities as well as insights relevant for the practical application of smart glasses in enterprise settings. In more detail, our contribution to research is twofold. First, we develop a taxonomy of functionalities of recent smart glasses. Second, we map them to requirements in the TCS. This in turn leads to a strategic decision-making basis for the introduction of smart glasses in companies. With the help of this, functional management is equipped with an instrument to decide – based on their specific requirements – whether smart glasses create value and what functionality contributes to that value. With our research, we fill two research gaps. First, to the best of our knowledge, no taxonomy of smart glasses technology and their functionalities have been published so far. Second, also the actual impact on business cases in the TCS area has not been discussed in the Service Science literature until now which we do in our contribution. This provides a fertile ground for future research concerning the design of smart glasses systems in TCS carrying out the system design and ex-post evaluation.

Against this background, the questions that guide our research are: RQ1: What functionalities do the emerging technology smart glasses offer? and closely connected to the previous question: RQ2: Are the requirements of the TCS addressed by the capabilities of smart glasses? To address our research questions, we proceed as follows: In Section 2, we give an overview on our methodical approach which also introduces the structure of this paper. In Section 3, we give background information about TCS processes. In Section 4, functionalities of smart glasses are explored. Further in this section, we will examine whether smart glasses create value for the TCS. In Section 5, we discuss our findings, conclude our approach and point to future research.

2 Research Model

As our research focuses on the design of an artefact to support the service processes of the TCS we follow a DSR approach [cf. Pe07, SB12]. As SONNENBERG AND VOM BROCKE suggest, we conducted an ex-ante evaluation before constructing the artefact itself [SB12] to validate e.g. system designs, requirements or guidelines [Of10]. Thus, as shown in Figure 1, we follow a two-step approach: (1) We analyze the support technology smart glasses and their functionalities (technological perspective). After that, we conduct an ex-ante evaluation to validate whether smart glasses act as technology to support the TCS, (2) we establish the relations between the two aspects via a mapping that allows us to derive the impact of smart glasses on TCS. With the ex-ante evaluation, we validate on the one hand the applicability and the suitability of using smart glasses to support the identified requirements and, hence, the TCS process. On the other hand, while mapping the requirements with the smart glasses functionalities, we implicitly also evaluate completeness, correct level of detail, internal consistency and applicability of our derived smart glasses taxonomy [SB12].

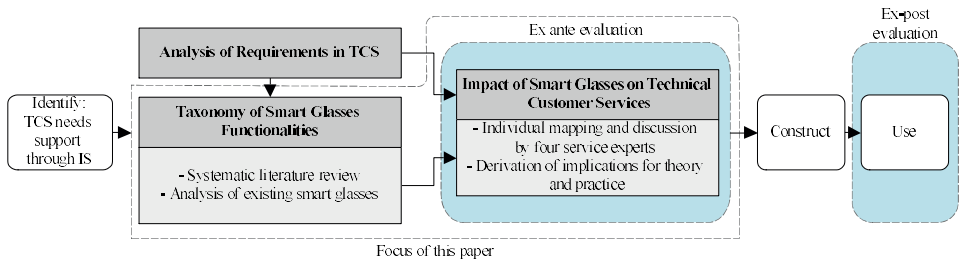


Figure 1. Research model referring to SONNENBERG & VOM BROCKE [SB12]

In previous work, requirements of TCS were elicited by observing 77 TCS processes (one process = one customer order) of a leading company in the intralogistics sector in Germany. The processes obtained maintenance, repair, and adjustment and consultation orders. A detailed description of the research methodologies and their findings has already been published in [Ma13].

Due to the novelty of the smart glasses topic, we had no access to practitioners with real-life implementation of a smart glasses system and just a few scientific papers have been published. For this reason, we decided to elicit the functionalities from a systematic literature review (cf. [BR09, WW02]) in the classical IS databases as well as investigate a market analysis of the smart glasses that are already available or announced to be available in 2015/2016 (cf. Section 4). The search term “smart glasses” was used. Querying the databases returned 211 hits, which were reviewed in respect to title, abstract and keywords (if available). We excluded articles if they do not focus on smart glasses defined in terms of augmented reality, mobility or head-worn displays. Thus, 169 papers were classified as out of scope and five duplicates were eliminated. This revealed a total of 37 relevant articles. This set of articles contains theory-oriented papers [RBI15, cf. Wa14a] as well as

application-oriented papers [Gi12, cf. He14b] also listed in the mentioned databases. The remaining articles were discussed by the research team. By the analysis of the papers for smart glasses functionalities another nine papers were excluded. We excluded papers if (1) they did not address functionalities in an IT-related sense or if (2) they provided mere hardware descriptions such as the weight or battery life that did not count as a feature in our analysis or if (3) they provide pure project descriptions or business-oriented use case descriptions lacking a sufficient level of detail were excluded. After implementing the exclusion criteria, 28 papers remained. Passing on to the analysis of our search results, in a first step, we assigned keywords to the articles in a process comparable to open coding [CS15]. In this process, we identified key features of smart glasses and assigned corresponding keywords (codes) to the articles such as e.g. “GPS Navigation”. Due to the reason that there are only a few scientific papers found and the publishing process sometimes has a longer time to succeed than the announcements of new technologies from the manufacturer of new information technologies, we decided to complete and double check the list of smart glasses features with a market analysis of the existing and announced smart glasses. Therefore, we analyzed the 10 most recent and most relevant smart glasses specifications. To include all current and future features, we used a mixture between already released smart glasses and announced ones that are likely to be release within the next 2 years. For the collection of the features official websites and specifications were used. After having identified the single features of smart glasses, we clustered them inductively into groups [Ma10].

To answer the question whether smart glasses have an impact on the TCS, we analyzed if the functionalities do meet the requirements for mobile support systems in TCS. Hence, in a team of four service experts, we matched the functionalities with the TCS requirements. All experts share a Service Science background.

3 Technical Customer Service Process

For the development of a service support system for the TCS, it is necessary to identify and structure underlying requirements based on service processes. For this purpose, a reference maintenance process is used as initial point for service activity description. Figure 2 represents a reference maintenance process in the field of the TCS.

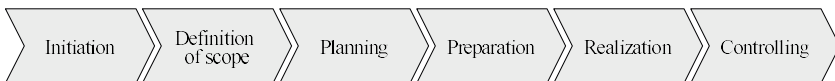


Figure 2. Reference maintenance process of the TCS [Da15]

The first phase of the service process starts with the *initiation* which is usually triggered by a service request induced by the customer. This phase can in turn roughly be distinguished into the sub-processes customer request management in which the TCS receives the service order in form of claims, repair or supply requests from customers and the equipment management composed of scheduled maintenance and on-request maintenance

[LX96]. However, since both activities are heavily intertwined, we do not further distinguish between them as separate steps. After the *initiation*, the *scope of the service order is defined* which builds the foundation for the service planning phase. The appropriate service technician is informed about the symptoms obtained from the customer himself or the service center as point of contact [LX96]. Aiming to fulfil the determined goals, required resources such as spare parts or tools need to be checked for service delivery. Following the *planning* of the forthcoming service tasks, additional control measures need to be considered and the ascertainment of the service delivery begins. For this purpose, dates, specific work orders and the required resources must be provided and *prepared* [Da15]. During *realization phase*, the service technician delivers his work order in physical appearance on the service object at the customer's site [MBM03]. The process ends with the *controlling phase* comprising the analysis and reflection of the service delivery, with respect to the quality and productivity of the service technician's activities [Da15]. In the latter, the *controlling phase* is dropped as it is a cross-departmental function that only indirectly benefits from smart glasses.

4 Functionalities of Smart Glasses

Within the 28 publications, a consolidated number of 20 features, consisting of a combination of hardware and software components, were mentioned. Except for a *Temperature Gauge*-feature, which was presented as special feature within one paper, all of them were mentioned in at least three publications. The most mentioned features include *Voice Recognition*, *Picture and Video taking* as well as *GPS Navigation* and *Information Searching*.

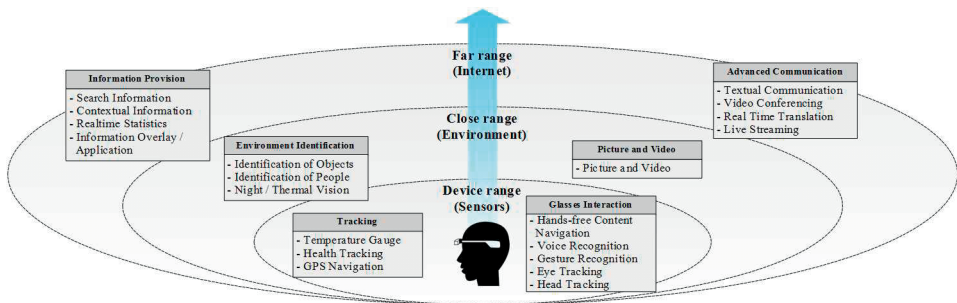


Figure 3. Categorization of functionalities

When it comes to categorizing the functionalities, we clustered them into six groups by their topical closeness: the identified groups (1) *Tracking* and (2) *Glasses Interaction* include features existing on the device itself working with sensors; (3) *Environment Identification* and (4) *Picture and Video* contain features operating at the environmental range close to the device; (5) *Information Provision* and (6) *Advanced Communication* comprise features working also in far distance e.g. via internet connection (cf. Figure 3).

	<i>Feature</i>	<i>Description</i>	<i>Literature</i>	
Device Range (Sensors)	(1) Tracking	Temperature Gauge	Measurement of the ambient temperature.	[Ka13a]
		Health Tracking	Tracking of health activities by measuring for example the heart rate using sensor components.	[Be14, He14a, Ov13, Wa14a, Wi13]
		GPS Navigation	A navigation system that provides location information.	[Go13, He14a, Vu13, Wa14b, Wa14a, Wi13, He14b, Ka13b, Ka13a, Mu12, OI13a, OI15, Ov13, RBI15]
	(2) Glasses Interaction	Hands-free Content Navigation	Navigation without using hands, in order to carry out commands or browse content.	[Ca13, Da14, OI13b, Ov13, Vu13, Wa14b, Go13, He14a, He14b, Ka13c, Ka13b, Ka13a, Mo15, OI13a]
		Voice Recognition	Receiving and interpreting dictation, or understanding and carrying out spoken commands.	[Be14, Ca13, RBI15, Si13, Vu13, Wa14b, Wa14a, Da14, Gi12, He14a, He14b, Jo14, Mo15, Na13, Ov13]
		Gesture Recognition	Interpretation of human gestures via mathematical algorithms in order to carry out commands.	[Be14, Da14, Fo13, He14a, Jo14, Mc15, Na13, OI15, Ov13, Wi13]
		Eye Tracking	Measurement of either the point of gaze or the motion of an eye relative to the head. It can be used for either analysis or navigation.	[Be14, Ov13, Wa14b, Wa14a, Wi13]
Head Tracking	Monitoring the head movement. This allows moving the camera in the same orientation as the user's head does.	[Ca13, Jo14, Ka13a, Mu12, OI13a, OI15, Vu13, Wa14b]		
Close Range (Environment)	(3) Environment Identification	Identification of Objects	Using camera function to identify, track and recognize objects.	[Be14, Ca13, He14a, Mu12, Na13, Ox14, RBI15, Si13, Wa14b, Wa14a]
		Identification of People	Using camera function to identify, track and recognize people.	[Be14, Go13, He14a, He14b, Ov13, Ox14, Si13, Wa14a]
	Night/Thermal Vision	Collecting the infrared radiation from objects in the scene to be able to see in low light conditions.	[Mu12, Ov13, Ox14]	
	(4) Picture and Video	Pictures and Video	Taking pictures and videos using the internal camera.	[Be14, Ca13, OI13a, Ox14, Vu13, Wa14b, Wa14a, Wi13, Gi12, Go13, He14b, Ka13c, Ka13a, Ma14, Mc15, Mo15]
Far Range (Internet)	(5) Information Provision	Search Information	Search for information based on search terms, pictures etc.	[Be14, Ca13, Ov13, RBI15, Wa14a, Da14, Gi12, Go13, He14a, He14b, Mo15, Mu12, Na13]
		Contextual Information	Collection and analysis of data about a device's surroundings in order to present relevant, actionable information.	[Be14, Ca13, Da14, Fo13, He14a, He14b, Jo14, Na13]
		Real-time Statistics	Compiling of real-time statistics about relevant information and displaying them.	[Be14, Ca13, Fo13, He14a, Ka13a, OI15, Wi13]
		Information Overlay/Application	Adding additional information by overlaying virtual objects upon the user view of the real world.	[Be14, Gi12, He14a, Mu12, RBI15, Wi13]
	(6) Advanced Communication	Textual Communication	Send and receive textual messages using communication channels such as e-mail in conjunction with input mechanisms such as voice recognition.	[Ca13, Gi12, Go13, Ka13b, Ka13a, Mo15, RBI15, Vu13, Wa14a]
		Video Conferencing	Communication of the user with two or more other users by real-time bidirectional video and audio transmissions.	[Ca13, He14b, Mo15, Wa14a]
		Real-time Translation	Translation of foreign languages on-the-fly and displaying subtitles.	[Be14, Go13, He14a]
		Live Streaming	Video content is sent in compressed form over the Internet and displayed by the viewer in real time.	[Da14, He14a, Ka13a, Mo15, Mu12, OI15, Ov13]

Table 1. Overview of smart glasses features from literature

Noteworthy about the categories is that although the *Glasses Interaction* is a support function and is not completely distinct from the other groups, it forms a separate group as the

interaction with the emerging technology is one of the key factors for its success. Furthermore, the *Picture and Video* functionality led to its own category as it was the most mentioned feature. Table 1 provides further information for the introduced features with a short description, the group they were assigned to and the supporting literature that initially mentioned them.

Within the 10 smart glasses that were analyzed, we found that most but not all of the features from literature are mentioned. Especially in *Tracking* the features *Temperature Gauge* and *Health Tracking* as well as *Identification of People* and *Night/Thermal Vision* are not found on the 10 smart glasses analyzed. We explain that mainly by the specialness of the first two and privacy concerns for the last.

Category	Feature	Google Glass	Vuzix M100	Glassup	Moverio BT-200	Golden-I	Motorola HCl	Atheer One	Meta Pro	Recon Jet	Microsoft Hololens	Total
(1) Tracking	Temperature Gauge											0
	Health Tracking											0
	GPS Navigation	x	x	x	x		x			x		6
(2) Glasses Interaction	Hands-free Content Navigation	x	x		x		x	x	x	x	x	8
	Voice Recognition	x	x			x	x				x	5
	Gesture Recognition		x			x	x	x	x		x	6
	Touch Recognition							x	x		x	3
	Eye Tracking										x	1
	Head Tracking	x	x	x	x	x	x	x	x	x	x	10
(3) Environment Identification	Identification of Objects	x	x					x	x		x	5
	Identification of People											0
	Night/Thermal Vision											0
(4) Picture and Video	Pictures and Video	x	x		x	x	x	x	x	x	x	9
(5) Information Provision	Search Information	x	x		x		x	x	x	x	x	8
	Contextual Information	x	x	x	x		x	x	x	x	x	9
	Real-time Statistics	x	x	x						x		4
	Information Overlay/Application				x			x	x		x	4
(6) Advanced Communication	Textual Communication	x		x			x			x	x	5
	Video Conferencing	x	x							x	x	4
	Real-time Translation	x		x	x							4
	Live Streaming	x	x		x						x	4
	Total	13	12	6	9	4	9	9	9	9	14	

Table 2: Overview of smart glasses and features from market data

An additional feature was found that was not mentioned in literature before which is *Touch Recognition*. This feature enables the smart glasses to track the contact to virtual object that are projected. Therefore, a depth camera is needed to enable the measurement of hands

distance. It can be used for virtual graphic user interfaces and interaction with virtual objects. Regarding groups, we included the new feature in *Glasses Interaction*. An overview of the found features and the smart glasses is given in Table 2. We put an x when the feature is mentioned for the smart glasses.

4.1 Using Smart Glasses to Support Technical Customer Service Processes

In the previous sections of this chapter, several functionalities of smart glasses have been identified and described. To investigate whether those functionalities support the service technician better, it has to be analyzed whether the functionalities do meet the requirements for mobile support system in TCS. We repeated the incremental approach of first working independently and then consolidating the results that we already used for Table 1. The results are shown in Table 3. Within the analysis we put an “x”, if the functionalities of smart glasses create added value. No “x” is stated, if the functionality does not offer a direct benefit. For example, *Glasses Interaction* – e.g. *Hands-free interaction* – can be helpful with almost every process step and requirement, but we just put an “x” when the requirement concerns a work step where the technician needs his hands for another activity as like repairing the service object or environmental circumstances such as soiled hands occur. For example, the *Filling-out assistant for forms and reports* could be used with handhelds, with just winking or voice control, to fill in the invoices, but there is no direct advantage compared to using a handheld keyboard. However, concerning the *Electronic checklist for customer service operations* or the *Interactive assistance for customer service operations*, hands-free interaction helps to communicate with the device while operating at the service object. Finally, we used an “(x)” if the benefit is not evident in the first place but under certain conditions. Regarding an overall of 36 requirements, we conscientiously evaluated and discussed every mark. In the following three examples are given.

First, regarding the requirement *Diagnostic function for maintenance objects*, we detected a beneficial influence by *Environment Identification*, *Information Provision* and a minor influence by *Tracking*. This is due to the fact that along the service process the technician, first of all, has to identify the service object by *Object Recognition*. Second, he has to investigate the object which can be assisted by *Information Provision* in form of an information overlay of non-visual parts of the machine or next service steps. Conditionally, the *Tracking* can be beneficial as for instance the room temperature can be tracked and might be an indicator for the systems malfunction.

Second, considering the requirement *Electronic checklist for customer service operations* an aid is created by *Glasses Interaction*, *Environment Identification*, *Information Provision* and *Tracking*. When the technician arrives at the customer-site the checklist is automatically shown based on *Tracking* (e.g. *GPS Navigation*) and leads him through the service process. The ability of *Glasses Interaction* with *hands-free content browsing* and *Voice Recognition* allows the hands-free navigation through the checklist.

Requirements	Functionalities						Total
	(1) Tracking	(2) Glasses Interaction	(3) Environment Identification	(4) Picture and Video	(5) Information Provision	(6) Advanced Communication	
Customer complaint management				(x)			1
Localized information				(x)			1
Order description				(x)			1
Returns management				(x)			1
Diagnostic function for maintenance objects	(x)		x		x		3
Electronic checklist for customer service operations	x	x	x		x		4
Interactive assistance for customer service operations		x	x		x	x	4
Interface for the parameterization of maintenance objects		x					1
Preview function of documents					x		1
Remote diagnosis functions for maintenance objects	x					x	2
Scan function for optical and electromagnetic codes	x	x					2
Updating information resources	x	x	x	x			4
Electronic spare parts procurement			x	x	x		3
Linked information objects			x		x		2
Plausibility check for data collection	x		x				2
Search and call of structured and unstructured data		x			x		2
Intelligent disposition	x				x		2
Filling-out assistant for forms and reports	x	x	x	(x)	x		5
News service					x		1
Proactive information provision			x		x		2
Proactive order provision (Management of orders)	x		x		x		3
Real-time transmission of order-related data					x		1
Real-time communication with int. and ext. Actors						x	1
Service-related Key Performance Indicator measurement	x		x				2
Updating of the knowledge database	x		x	x			3
Updating of the service history	x	(x)	x				3
Opportunity to influence the disposition	x						1
Support of weakly structured processes					x	x	2
Forecast function for resource and tool requirements	(x)		x		x		3
Local contact persons (contact info, kind of knowledge)			x		x		2
Reminder function for appointments	x				x		2
Report error				x		x	2
(Partially) automated document creation	x		x	x			3
Decision support for customer service					x		1
Management of suggestions for improvement				x			1
Preparation of cost estimates	(x)		x		x		3
Total	17	8	17	11	19	5	

Table 3. Coverage of the requirements by smart glasses functionalities

So, working with both hands and concurrent information supply via checklist enhance the technician. On top the *Information Provision* via *Contextual Information* offer additional information directly when needed. Finally, *Environment Identification* is useful in terms of *Object Recognition* as it can be used for identification of the serviced object and might automatically check items on the checklist when the completion is detected.

Third, one requirement is *Report Error* which deals with the demand of immediate help when something goes wrong. Advantages of smart glasses attached to this requirement are *Advanced Communication* as well as *Picture and Video*. In case of an error, the need for timely expertise is crucial to prevent further damage. Therefore, *Advanced Communication* with *Video Conferencing* is beneficial as a colleague or supervisor can be called and with *Live Streaming* the Point-of-view is transferred which enables the technician at site to show the current problem. On top he can document the problem by taking pictures or videos.

Interpreting Table 3, the greatest impact can be achieved by fulfilling requirements concerning the operations at the service object; for example *Electronic checklist for customer service operations* and *Interactive assistance for customer service operations*. In addition, requirements that concern the feedback linked to the service operations as for example *Updating information resources* or *Filling-out assistant for forms and reports* are as well considered positive. Expected as the most value-creating functionalities are *Information Provision*, *Environment Identification* and *Tracking*. This could be explained, analogous to the statement concerning the requirements that the biggest benefit occurs while operating at the service object. The *Advanced Communication* is not as value-creating as the others. This might be due to the fact, that the service technician is able and encouraged to perform independently. Additionally, the added functionalities as for example through *Information Provision* decrease the need for additional communication even further. The small benefit of *Advanced Communication* could be explained by the fact, that the interaction with the device is beneficial in situation where hands-free interaction is needed, but due to the other functionalities, such as *Environment Identification* or *Information Provision*, the communication happens partly automated. Finally, there are four requirements that do not directly benefit from the functionalities of smart glasses. This could be explained by the fact that those four requirements are solely linked to the first two phases *Initiation* and *Definition of Scope* where the focus of the technicians is on customer contact and *planning*, mostly operated at the service center and not at the point of service.

Although most requirements benefit from the support through smart glasses, there are process steps where smart glasses might have negative implications. The service process comprises a lot of documentation which at first do not seem more natural on smart glasses, depending on the model of the smart glasses and their display size. This documentation work is already implicitly included in requirements, for instance, *Updating of the service history* or *Automated document creation*. Using smart glasses, these requirements can directly be met (video, audio, environmental identification etc.). Hence, in future research we will analyze which influence the benefit of the features of smart glasses on the one hand and the disadvantages of their limited display size at the other will exert on the overall acceptance and satisfaction of the service technician. However, when it comes to documentation the use of smart glasses might also have negative impact as the different input methods do need some kind of adaption phase. Further, they might not satisfy the user when he is forced to insert information as text.

5 Conclusion and Outlook

Based on a literature review and smart glasses manufacturer's data, we derived and determined relevant features of smart glasses regarding the TCS. In order to check the applicability and usefulness of smart glasses and to make sure that smart glasses functionalities support the service technician better, we examined the capabilities of smart glasses by satisfying the existing requirements for mobile service support systems. According to our results, the greatest impact using smart glasses can be achieved by fulfilling requirements concerning the direct operations on the service object (e.g. a machine that is to be repaired). The service technician has the ability to process the service order while wearing the support system instantaneous in front of his eyes. So, he can receive information while working and without stopping the service delivery as it would require interacting with a traditional handheld device. The most value-creating functionalities of smart glasses are expected in *Information Provision*, *Environment Identification* and *Tracking*. These are features that are inevitable for the service technicians while operating at the service object.

A limitation of our research is that our insights are not based on the real use of smart glasses in the TCS since experiences in that direction that are backed by large datasets of real use cases are not yet available. Therefore, we go the first step into this direction in describing smart glasses functionalities and confronting them with requirements from the TCS. In doing so, we pave the way for further research and implementation activities with the ultimate goal of evaluating smart glasses in the TCS domain in real-world field studies. As part of our future research, we will also address the potentially negative effects that smart glasses may exert on work processes. For example, the customer may feel disturbed by a technician with a camera mounted on his head that may affect the customer loyalty. Some companies might even forbid devices with cameras due to confidentiality concerns. Also, regarding the ergonomics of smart glasses, some questions remain open. Since information and data are presented immediately in the service technicians' field of vision, a reduction of head and neck movements during repair can be achieved as there is no need to turn to a handheld display [HF11]. Further, the investigation of the actually perceived positive effects in terms of ergonomics would also require real-world experiments and the result may be highly dependent on the concrete tasks. However, our research results are still valuable since they form the basis to easily generate hypotheses regarding the value of smart glasses which then can subsequently be confirmed or rejected. This, in turn, may be highly dependent on the characteristics of the cases that are going to be analyzed. It, thus, requires a lot of future research effort to finally come up with a valid theory of smart glasses impact on TCS which finally may be formulated e.g. as an explanatory design theory [BP10]. Our results contribute towards the development of such a theory.

All in all, the main contribution we provide is twofold. First, the systematical analysis of smart glasses functionalities (RQ1) and creation of a taxonomy can be used for further research and deal as foundation of a state-of-the-art. Second, the application on requirements in TCS (RQ2) enables a more deeply understanding of why the emerging technology might change the field. Consequently, the impact on the theoretical work on TCS in

future certainly has to keep the technology in mind. The practical impact of the contribution is that it might support decisions about using functionalities of smart glasses in terms of the requirements that they help to address. Even if many questions about the acceptance of smart glasses and the design of a system are still unanswered, this contribution can provide answers to the question about the general usefulness of smart glasses in TCS. Consequently, a responsible manager could use the contribution to match their own requirements with the ones given and determine whether smart glasses are beneficial.

6 References

- [AH10] Andersson, B. ; Henningsson, S.: Developing mobile Information Systems: Managing additional Aspects. In: Alexander, T. ; Turpin, M. ; van Deventer, J. (Hrsg.): Proc. of the 18th Europ. Conf. on Information Systems. Pretoria, South Africa, 2010, S. 1–12
- [BP10] Baskerville, Richard ; Pries-Heje, Jan: Explanatory Design Theory. In: Business & Information Systems Engineering Bd. 2 (2010), Nr. 5, S. 271–282
- [Be14] Bendel, Oliver: Die Datenbrille aus Sicht der Informationsethik. In: Informatik-Spektrum, Springer Berlin Heidelberg (2014), S. 1–9
- [BLM14] Böhmann, Tilo ; Leimeister, Jan Marco ; Möslin, Kathrin: Service Systems Engineering. In: Business & Information Systems Engineering Bd. 6 (2014), Nr. 2, S. 73–79
- [BR09] vom Brocke, Jan ; Simons, Alexander ; Niehaves, Björn ; Riemer, Kai ; Plattfaut, Ralf ; Cleven, Anne ; Brocke, Jan Von ; Reimer, Kai ; u. a.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. In: 17th European Conference on Information Systems (2009), S. 2206–2217
- [Ca13] Canaday, Henry: Maintenance Mobility. In: Aviation Week & Space Technology Bd. 175 (2013), Nr. 45, S. 14
- [CS15] Corbin, J ; Strauss, A: Basics of qualitative research: Techniques and procedures for developing grounded theory. Thousand Oaks, California : Sage Publications Ltd., 2015
- [Da14] Davis, Brian: Ready to Wear. In: Professional Engineering Bd. 27 (2014), S. 31–34
- [Da15] Däuble, Gerald; Özcan, Deniz; Niemöller, Christina; Fellmann, Michael; Nüttgens, Markus; Thomas, Oliver: Information Needs of the Mobile Technical Customer Service - A Case Study in the Field of Machinery and Plant Engineering. Proceedings of the 48th Annual Hawaii International Conference on System Sciences (HICSS) (2015). Manoa, S. 1018–1027.
- [Fo13] Forbes.com: Startup Meta Challenges Google Glass With Sleek, \$3,000 Smart Glasses. In: Forbes.com (2013)
- [Gi12] Giles, Martin: Through a Glass digitally. In: Economist (2012)
- [Go13] Google: Capturing the Value of Smart Glasses. In: Trends Magazine Bd. 10 (2013), Nr. 10, S. 37–41
- [HM05] Haritos, T. ; Macchiarella, N.D.: A Mobile Application of Augmented Reality for

- Aerospace Maintenance Training. In: 24th Digital Avionics Systems Conference. Bd. 1 : IEEE, 2005, S. 5.B.3–1–5.B.3–9
- [HF09] Henderson, S ; Feiner, S: Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In: Proceeding ISMAR '09 Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality. Washington, DC, USA, 2009, S. 135–144
- [HF11] Henderson, S ; Feiner, S: Exploring the benefits of augmented reality documentation for maintenance and repair. In: Visualization and Computer Graphics, IEEE Transactions Bd. 17 (2011), Nr. 10, S. 1355–1368
- [He14a] Hegel, Theresa: Through the Looking Glass. In: Wearables Bd. 18 (2014), S. 38–39
- [He14b] Hemmadi, Murad: Wearable Tech Comes to the Office. In: Canadian Business Bd. 87 (2014), Nr. 13, S. 12–14
- [HE15] Herterich, Matthias M ; Peters, Christoph ; Uebernickel, Falk ; Brenner, Walter ; Neff, Alexander A: Mobile Work Support for Field Service : A Literature Review and Directions for Future Research. In: Internationale Tagung Wirtschaftsinformatik (2015), S. 134–148
- [Jo14] Jooyeon Ham, Jonggi Hong, Youngkyoon Jang, Seung Hwan Ko, Woontack Woo: Smart Wristband: Touch-and-Motion–Tracking Wearable 3D Input Device for Smart Glasses. In: Streitz, N. ; Markopoulos, P. (Hrsg.): Distributed, Ambient, and Pervasive Interactions, Lecture Notes in Computer Science. Bd. 8530. Cham : Springer International Publishing, 2014, S. 109–118
- [Ka13a] Karp, Andrew: Get ready for Eyewear-able Technology (2013), 20/20 Magazine
- [Ka13b] Karp, Andrew: Eye2 Explores Smart Glasses, Glasses Free 3D and More. In: Vision Monday Bd. 27 (2013), Nr. 1, S. 23
- [Ka13c] Karp, Andrew: Celebrating a Year of Eye² Innovation. In: Vision Monday Bd. 27 (2013), Nr. 16, S. 8
- [LP98] Lawson, Shaun W. ; Pretlove, John R. G.: Augmented reality for underground pipe inspection and maintenance. In: Stein, M. R. (Hrsg.): Photonics East (ISAM, VVDC, IEMB) : International Society for Optics and Photonics, 1998, S. 98–104
- [Le11] Legner, Christine ; Nolte, Christoph ; Nils, Urbach: Evaluating Mobile Business Applications in Service Maintenance Processes: Results of a Quantitative-Empirical Study. In: ECIS 2011 Proceedings., 2011, S. Paper 247
- [LX96] Lee, Jae Sik ; Xon, Yong Xune: A customer service process innovation using the integration of data base and case base. In: Expert Systems with Applications Bd. 11 (1996), Nr. 4, S. 543–552
- [LH14] Lindström, John ; Hanken, Claas: Wearable Computing: Security Challenges, BYOD, Privacy, and Legal Aspects. In: Nemati, H. R. (Hrsg.): Analyzing Security, Trust, and Crime in the Digital World : IGI Global, 2014, S. 96–120
- [Lu13] Luftman, Jerry ; Zadeh, Hossein S ; Derksen, Barry ; Santana, Martin ; Rigoni, Eduardo Henrique ; Huang, Zhengwei (David): Key information technology and management issues 2012–2013: an international study. In: Journal of Information Technology Bd. 28,

- Nature Publishing Group (2013), Nr. 4, S. 354–366
- [MBM03] Mayer, Karl J. ; Bowen, John T. ; Moulton, Margaret R.: A proposed model of the descriptors of service process. In: *Journal of Services Marketing* Bd. 17 (2003), Nr. 6, S. 621–639
- [Ma10] Mayring, Philipp: *Qualitative Inhaltsanalyse Grundlagen und Techniken*. Weinheim : Beltz, 2010
- [Ma13] Matijacic, Michel; Fellmann, Michael; Özcan, Deniz; Kammler, Friedemann; Nüttgens, Markus; Thomas, Oliver: Elicitation and Consolidation of Requirements for Mobile Technical Customer Services Support Systems – A Multi-Method Approach. In: *Proceedings of the 34th International Conference on Information Systems (ICIS) (2013)*. Mailand, Italien, S. 1–16.
- [Ma14] Martini, Paul: A secure approach to wearable technology. In: *Network Security* Bd. 2014 (2014), Nr. 10, S. 15–17
- [Mc15] McCandless, Karen: Seeing into the future. In: *Automotive Manufacturing Solutions* Bd. 16 (2015), Nr. 2, S. 26–28
- [Mo15] Moshtaghi, Omid ; Kelley, Kanwar S ; Armstrong, William B ; Ghavami, Yaser ; Gu, Jeffery ; Djalilian, Hamid R: Using google glass to solve communication and surgical education challenges in the operating room. In: *The Laryngoscope* (2015)
- [Mu12] Murphy, David: Vuzix Glasses Bring Augmented Reality to Your Games, Face. In: *PC Magazine* (2012)
- [Na13] Nadler, Daniel: Wearable FinTech Part I: Target, Ticker. In: *Institutional Investor-International Edition* (2013)
- [Of10] Offermann, Philipp ; Blom, Sören ; Schönherr, Marten ; Bub, Udo: Artifact Types in Information Systems Design Science – A Literature Review. In: *Global Perspectives on Design Science Research*. LNCS, vol. 6105 Bd. 6105 (2010), S. 77–92
- [Ol13a] Olson, Parmy: Cheaper Google Glass Competitor Vuzix Ships Developer Units. In: *Forbes.com* (2013)
- [Ol13b] Olson, Parmy: Epson Smart Glasses Browse YouTube With A Nod And Tilt Of The Head. In: *Forbes.com* (2013)
- [Ol15] Olson, Parmy: Smart Glasses Won't Hit The Mainstream Till Our Bosses Start Buying Them. In: *Forbes.com* (2015)
- [Ov13] Overton, Gail: Head-worn displays: Useful tool or niche novelty? In: *Laser Focus World* Bd. 49 (2013), Nr. 7, S. 33–37
- [Ox14] Oxford, University of: Smart glasses offer spatial awareness to the visually impaired. In: *Engineer (Online Edition)* (2014)
- [Pe07] Peffers, Ken ; Tuunanen, Tuure ; Rothenberger, Marcus a. ; Chatterjee, Samir: A Design Science Research Methodology for Information Systems Research. In: *Journal of Management Information Systems* Bd. 24 (2007), Nr. 3, S. 45–77
- [RB15] Rauschnabel, Philipp A. ; Brem, Alexander ; Ivens, Bjoern S.: Who will buy smart glasses? Empirical results of two pre-market-entry studies on the role of personality in

- individual awareness and intended adoption of Google Glass wearables. In: *Computers in Human Behavior* Bd. 49 (2015), S. 635–647
- [RBH03] Rügge, Ingrid ; Boronowsky, Michael ; Herzog, Otthein: Wearable Computing für die Industrie. In: *Industrie Management* Bd. 6 (2003)
- [Si13] SITA Lab: Google Glass Boarding. In: *Asian Aviation Magazine* Bd. 11 (2013), S. 31
- [SB12] Sonnenberg, Christian ; vom Brocke, Jan: Evaluations in the Science of the Artificial - Reconsidering the Build-Evaluate Pattern in Design Science Research. In: Peffers, K. ; Rothenberger, M. ; Kuechler, B. (Hrsg.) *DESRIST 2012, LNCS 7286*. Berlin Heidelberg, Springer-Verlag (2012), S. 381–397
- [Vu13] Vuzix: Futuristic Heads-up Glasses with GPS Go to App Developers. In: *GPS World* Bd. 24 (2013), Nr. 4, S. 18
- [Wa14a] Wang, Chih-Hsuan: A market-oriented approach to accomplish product positioning and product recommendation for smart phones and wearable devices. In: *International Journal of Production Research* (2014), S. 1–12
- [Wa14b] Wang, Chih-Hsuan: Integrating correspondence analysis with Grey relational model to implement a user-driven STP product strategy for smart glasses. In: *Journal of Intelligent Manufacturing* (2014)
- [WW02] Webster, Jane ; Watson, Richard T: Analyzing the Past to Prepare for the Future: Writing a Literature Review. In: *MIS Quarterly* Bd. 26 (2002), Nr. 2, S. xiii – xxiii
- [Wi13] Wiederhold, Brenda K: Time to port augmented reality health apps to smart glasses? In: *Cyberpsychology, behavior and social networking* Bd. 16 (2013), Nr. 3, S. 157–158

Enterprise Architecture Management

Enterprise Architecture Management – Aktueller Stand und Ausblick

Carsten Brockmann¹, Eldar Sultanow²

Auf der Informatik 2016 wurden im Workshop Enterprise Architecture Management aktuelle Stände der Forschung veröffentlicht und während der Konferenz diskutiert. Der Workshop bestand aus zwei Blöcken.

Den ersten Block eröffnete der Beitrag zu maritimen Referenzmodellen. Die Autoren haben eine standardisierte Methode vorgestellt, um Architekturen für die Domäne der Seeverkehrsindustrie zu erstellen. Der nächste Beitrag behandelt den Einsatz stilisierter Fakten zur Bewertung wandlungsfähiger Unternehmensarchitekturen. Die Autoren haben herausgearbeitet, dass Unternehmensarchitekturen über Indikatoren gemessen werden. Stilisierte Fakten zeigen auf, wie die Indikatoren zusammenwirken. Beim dritten Beitrag wird der Frage nachgegangen, wie eine leichtgewichtige IT-Modernisierung in der öffentlichen Verwaltung mit der Hilfe von EAM erfolgen kann. Ziel der Methode ist es, die besonders betroffenen Bereiche mittels einer Heatmap zu identifizieren und mit geringem Aufwand Abhilfe zu schaffen.

Der zweite Block wird mit dem Beitrag „Design und Nutzung einer industriespezifischen Referenzarchitektur für die Telekommunikationsindustrie“ eröffnet. Die vorgestellte Referenzarchitektur fand bei drei Projekten Anwendung. Der nächste Beitrag „Entwicklung eines Architekturkonzepts zum flexiblen Einsatz von Analytics“ beschäftigt sich mit der Frage, wie mit cyber-physischen Systemen (CPS) oder einer Cloud Problemstellungen des Business Analytics behandelt werden können. Der letzte Beitrag stellt ein Standardisierungsprojekt zur Entwicklung einer Referenzarchitektur für die Luftfahrt vor.

¹ Capgemini, Potsdamer Platz 5, 10785 Berlin, Germany, carsten.brockmann@capgemini.com

² Capgemini, Bahnhofstraße 11C, 90402 Nuremberg, Germany, eldar.sultanow@capgemini.com

A domain-specific architecture framework for the maritime domain

Benjamin Weinert¹, Axel Hahn² and Oliver Norkus³

Abstract: The maritime domain has the goal to harmonize heterogeneous systems and integrate new approaches into existing structures. In this paper we discuss the need for architectural methodology to overcome the challenge to coordinate the development of new systems considering technology issues, governance aspects and users between existing architectures in the maritime domain. We introduce already approved domain-specific architecture frameworks and present our approach for a suitable architecture framework which takes into account the appropriate structure of maritime domain.

Keywords: Maritime domain, system architecture, enterprise architecture framework, MAF, e-navigation

1 Introduction

Vessels and shore-based facilities are equipped with a variety of maritime systems to improve safe navigation and traffic management. These heterogeneous systems such as automated broadcasting of vessel information (AIS), Electronic Chart and Data Information Systems (ECDIS) or shore-based vessel traffic management differ in a high degree.

Stakeholders in the maritime domain face the challenge to harmonize existing systems and integrate new approaches and technologies into existing technical and organizational structures for sustainable, reliable and safe maritime transportation. This affects technical components, organizational structures and human users as well as the common interaction as elements of socio-technical systems [Se13].

Hence, the International Maritime Organization (IMO) – the organization for global maritime matters as part of the United Nations - established the e-Navigation strategy. This is initiated as “the harmonized collection, integration, exchange, presentation and analysis of marine information on-board and ashore by electronic means to enhance berth to berth navigation and related services for safety and security at sea and protection of the marine environment” [MS09], [NC14].

¹ OFFIS Institute for Information Technology, R&D Division Transportation, Escherweg 2, 26121 Oldenburg, Germany, benjamin.weinert@offis.de

² University of Oldenburg, Department of Computing Science, Ammerländer Heerstr. 114-118, 26129 Oldenburg, Germany, axel.hahn@uni-oldenburg.de

³ University of Oldenburg, Department for Business Computing, Escherweg 2, 26121 Oldenburg, Germany, oliver.norkus@uni-oldenburg.de

This strategy is for the maritime domain. It defines a set of goals for the evolution of the maritime domain. E-Navigation focuses on the users needs shipside and ashore and aims to reduce the risk of accidents on sea and improve safety and efficiency in the maritime processes. This strategy bases mainly on interoperable information exchange between services and systems [NC14].

Thus, it is a vision in which direction the further development of the maritime domain shall go. However, the contribution wherein the strategy is defined, lacks on tools and methodologies to accomplish the goals defined in the strategy [MS14], [MS09].

Therefore, the maritime stakeholders face the challenge to operationalize the e-Navigation strategy successfully. This includes the use of a domain-wide suitable methodology considering relevant characteristics of the maritime domain to enable a structured and consistent view on maritime systems from different perspectives. This methodology needs to support a common development process for upcoming e-Navigation systems.

Thus, this paper introduces the Maritime Architecture Framework (MAF) as a domain specific architectural methodology to overcome the challenge to coordinate the development of new systems between technology issues, governance aspects and users between existing architectures.

In the following sections, we discuss first the need for a maritime domain specific architecture framework. Second, we present existing approaches for a domain-wide architecture model before we introduce the MAF as maritime domain specific architectural methodology including its relationship to the structure of the maritime domain. Finally, this paper is concluded in section 5.

2 Challenge for a domain-specific architecture framework

The existing systems in the maritime domain are stand-alone solutions for each use case. These systems are barely integrated and combined with each other. Currently applied solutions also do not support a domain-wide information exchange. In addition, they do not follow the e-Navigation strategy and thus they are less future-oriented but well embedded in the maritime domain infrastructure [DN15].

The IMO aims with its e-Navigation strategy for a harmonized collection and integration of maritime information and systems. Upcoming solutions such as new maritime systems, for instance communication between ship and shore, should consider the high number of already existing heterogeneous systems [NC14]. This results in a high integration and coordination effort for managing the interactions with these various systems [MS09].

Therefore, it happens that existing socio-technical systems are based on different architectural designs and are embedded in various organizational and governance

structures in international, regional and national matters. Following IMO's e-Navigation vision for seamless integration between existing and upcoming systems, existing systems have to be analyzed in a structured way on their system architectures including the specific context and organizational structures in which a system is embedded to enable views from different operational or technical perspectives on the examined systems. This will enhance the understanding of the relationship between the maritime world and the examined systems and to consider this during the development of new e-Navigation systems (see figure 1).



Fig. 1: Challenge for a domain-specific high-level framework

The answer to this challenge should consider the overarching e-Navigation strategy and IMO's reflection of the maritime domain. Consequently, it shall provide a consistent methodology to structure the engineering process of socio-technical system concepts and to align technical systems. Hence, such a framework needs to frame the maritime domain including its stakeholders, the existing and upcoming technical system (architectures), related business processes and organizational structures including governance and regulation aspects to enable a complete view on the maritime infrastructure.

Summarized, there are challenges within the maritime domain, which could be solved by an integrated maritime framework. The requirements for such a maritime framework are:

- Support the development of maritime system architectures by offering a structured methodology, which enables to identify possible interoperability issues to ensure a communication and cooperation across the systems or to identify overlaps and gaps within their system architectures.
- Support the analysis of maritime systems regarding their technical architectures and organizational structures.
- Ensure a domain-wide consistent terminology for a common understanding using

this framework.

- Allow the design of organizational and technical harmonized and optimized architectures.

3 The Maritime Architecture Framework

To overcome the challenges and requirements described in the upper section, the authors have negotiated to develop the Maritime Architecture Framework (MAF) in a broad community process. MAF is derived from the successfully established architecture model in the energy domain named Smart Grid Architecture Model (SGAM). SGAM is shortly explained in section 3.1, where after the specific development of the MAF is discussed (section 3.2). Section 3.3. introduces the concept before a well suitable visualization and modelling approach for the MAF is shown in section 3.4. The methodology to close the gap between the generic e-Navigation strategy and the current architectural world in the maritime domain is presented in section 3.5. Next to this introduction of the MAF, the following chapter will discuss the application and evaluation.

3.1 SGAM - A domain specific framework from the energy domain

Smart Grid Architecture Model (SGAM) is a successfully established implementation of an Enterprise Architecture Framework to address domain specific issues in the electric utilities domain. SGAM was developed to handle the complexity of the Smart Grid system-of-systems approach with focus on interoperability and standardization aspects for business and governance as well as for technical issues [CE12].

SGAM structures the electric domain to ensure a consistent basis for discussions in heterogeneous groups. It takes established domain-independent frameworks (e.g., TOGAF [TO16]) as well as domain-specific models into account. SGAM is used to describe technical use cases as well as business cases [UE15].

SGAM structures the aspects of the domain into three dimensions. It covers an interoperability dimension with five layers of specification aspects ranging from business objectives down to physical components. Furthermore, it includes a hierarchical dimension (named *zones*), which structures the power system management as well as a domain related dimension (therefore named *domain*) to represent the energy production and consumption chain. Besides the multidimensional model, SGAM's methodology supports the specification of use cases including stakeholders, actors and technical aspects and allows the mapping to its model [UE15].

The SGAM approach is successfully applied in the energy domain and already adapted in other domains. The industry 4.0 initiative in the industry domain uses the design

principles of SGAM in the Reference Architectural Model Industry 4.0 (RAMI 4.0) to enable different user perspectives to achieve a common understanding of the industrial value chain including their components, standards and use cases [HR15].

3.2 Development

The MAF is adopted from the SGAM Model by a community process. The authors derived an initial version directly from SGAM using existing maritime architectures and reference architectures. This includes the Common Shore Based System Architecture [IA15a], [IA15b] and IMO's e-Navigation architecture as well as the maritime service portfolio, a collection of maritime services done by the IMO [NC14].

In a second step the MAF proposal was discussed with numerous stakeholders from industry (provider for navigation systems and communication technology, system integrators, ship-owners, pilots and others) as well as engineers working for governmental maritime agencies (e.g. Germany, Denmark, Sweden, Korea) and researchers (in Norway, Australia, Austria, Korea). This led to an updated version as introduced by this paper.

At present, the MAF is under discussion by the working group of the eNAV committee of the IALA, the International Agency for Aids to Navigation and Lighthouse Authorities. Additionally, there are the test cases for using the MAF to structure a base architecture for e-Navigation named the Maritime Cloud, the development of data exchange to portable pilot units (PPU) and synchronization of port services within the project CPSE-Labs (see chapter 4).

3.3 Concept

The MAF needs to establish clear relationships between technical systems, users and related governance aspects. This includes:

- existing business objectives, that explain the benefits of the systems,
- regulation and governance aspects, which regulates the maritime domain,
- technical functions, that are required to realize the business objectives,
- information exchange between those technical functions including the related information types and / or data models,
- communication protocols to allow the aspired information exchange and
- components, which are required to implement the technical hardware in the system.

It needs to cover the domain as stated by the IMO [MS09]. To enable the representation

of relationships between technical issues with organizational aspects such as business objectives or governance issues, the framework needs to include different interoperability views. Therefore, the MAF derives the interoperability layer used in the SGAM approach [CE12].

Similar to the user-needs driven e-Navigation strategy, the MAF focuses on a user-oriented analysis of specific use-cases with respect to the examined architectures. It adapts the Maritime Service Portfolios (MSP) as improved provision of services to vessels and users [MS09] to define the scope of the MAF and as a basis for user and use case specifications of examined architectures.

Referring to established approaches, the framework is divided into two parts:

- The multidimensional cube for a graphical representation of the underlying maritime domain and the examined system architecture,
- A methodology to structure the examined system including the system requirements and (possible) use cases in a consistent way.

3.4 The multidimensional cube

The multidimensional cube (MAF-Cube) offers a consistent view on architectures on different levels. As shown in figure 2, the requirements which were defined in section 3.2 are used to display (1) interoperability-, (2) hierarchical- and (3) topological aspects of the maritime domain. From now on, we use the terminology *axis* for the name of the dimension and *layer* for the surface with all aspects, which belong to a *category*. All aspects on a layer can be sub-structured by the categories of other dimensions.

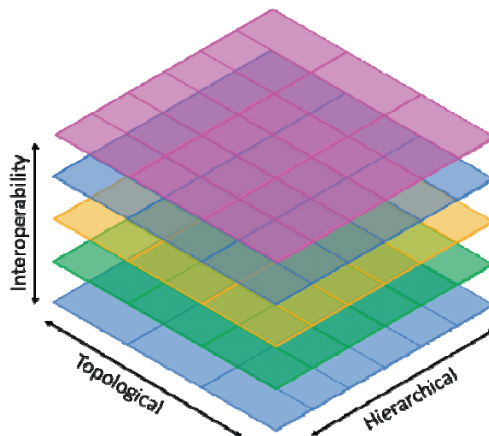


Fig. 2: The MAF-Cube

The topological axis represents the logical location where a technology component is located. The interoperability axis addresses communication, data and information, usage and context of a maritime system. The hierarchical axis substructures management and control systems of the maritime domain, for example for maritime transportation systems from the traffic management of a coastal area down to the radar echo of a vessel.

3.4.1 Topological axis

The layers are derived from IMO's breakdown of the maritime domain [NC14] and cover the structure of the maritime domain in a logical location. The axis contains the layers:

- Ships and other maritime traffic objects: Representing entities in the maritime domain (e.g., vessels). It covers the ship-side entities of the e-Navigation architecture.
- Link: Representing entities dedicated to physically interact between maritime traffic objects and shore, such as telecommunication methods and protocols. Represents the three levels of *Operational links*, *Functional links* and *Physical Links* between shipside and shore-side.
- Shore: Representing entities of the shore side infrastructure, activities and systems on shore including interfaces to logistical movements in/out of maritime domain.

3.4.2 Interoperability axis

The interoperability layers cover organizational, informational and technical aspects and include the different levels of interaction (*operational*, *functional*, *technical* and *physical*) as stated in IMO's e-Navigation vision [MS09] and are derived from SGAM [CE12].

- Regulation & Governance: Role and legal basis of international, regional or national (shipping) authorities.
- Function: Functions and (elemental) services including their relationships.
- Information: Data and information that is being used and exchanged between functions, services and components. It describes data and information objects including its semantic and data models.
- Communication: Protocols and mechanisms for the interoperable exchange of data between components.
- Component: Required components in engineering terms. This includes, amongst others: systems, actors, applications, services, network infrastructure.

3.4.3 Hierarchical axis

This axis covers economic, technical and physical issues of a maritime system. It starts with the classification of the examined system into its field of activity and continues with system-specific operations before the system will break down into technical services and their components as well as interfaced physical components.

- Fields of activity: Systems, which support or manage different markets or eco systems along the maritime domain.
- Operations: Global, regional, national and local operational perspectives used by companies or authorities (e.g. a traffic flow management).
- Systems: Technical systems, which integrate or use technical services for gaining a virtual representation and control of the transport processes.
- Technical Services: Single technical and logical services.
- Sensors & Actuators: Local infrastructure for detecting objects with physical means and receiving / processing the results with physical systems and hardware.
- Transport Objects: Entities of maritime transport processes such as vessels, floating objects and aircrafts operating in the maritime domain.

3.5 Methodology

The methodology is composed of three main steps leading to enable an easy mapping of system architectures to the MAF-Cube. The scope of this process is to structure the system engineering phases starting from planning over the identification of requirements to the use case development in a harmonized and formal way. This allows the user to map the results, to visualize in the MAF-Cube to explore interoperability issues and to identify spots which need to be standardized.

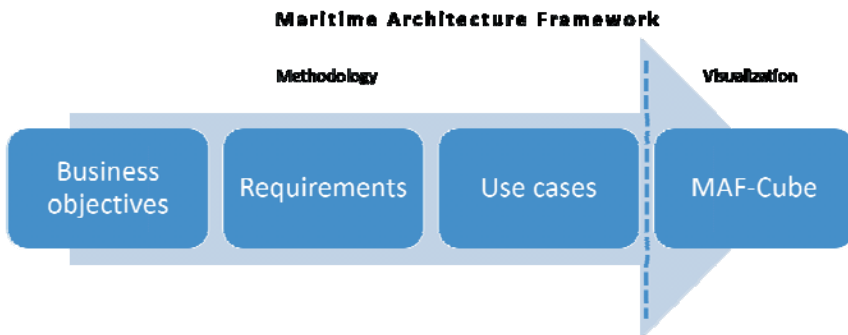


Fig. 3: The Maritime Architecture Framework

As shown in figure 3, the methodological steps are built upon each other. During this process, the MAF orients towards at least three different user-groups:

- (1) Business: Sorts all high-level decision makers governing the business objectives and the ambient business structures.
- (2) Engineer: Addresses the technical working group responsible for the realization of the objectives, given by the Business group.
- (3) Consumer: Groups users, who are not participating in the system development but have interests in the future use of the systems.

The framework provides this methodology, tools and templates for an alignment of characteristics of different system architectures. This framework includes a derivation of the Requirement Abstraction Model [GW06] which originally ensure interoperability between product requirements. This approach is used in the MAF to concretize the business objectives and requirements to the same abstraction level as basis for the development process of a system architecture or for an alignment between system architectures.

Furthermore, the MAF methodology includes a use case template, which follows the international standard for use case methodology in [IE15] and is extended considering the domain-specific characteristics. The use case template comes with an actor list, which is derived from the maritime users as defined by [MS09] and is enriched by facilitating the described information to UML diagrams. Therefore, it is supposed to support a consistent understanding of functionality, actors and process across different projects and between different organizational structures in the maritime domain.

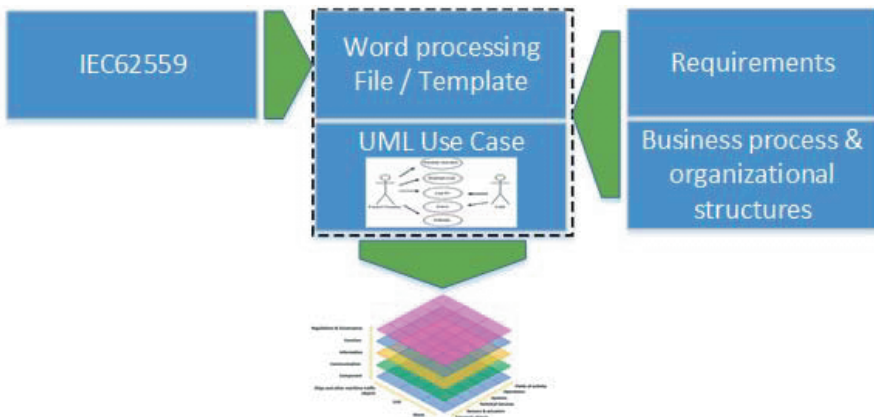


Fig. 4. The MAF process

Finally, the methodology includes a process to extract described information of a

systems architecture from the use cases to ensure a successive mapping to the MAF-Cube. This step identifies physical components and business relatives of the examined architectures and includes standards, rules and regulations regarding to the interoperability layers. The cube allows to locate the system architecture and related components in the maritime domain (shipside, shore-side or as a link between them). Furthermore, the cube enables to structure the system construction including its field of activity in a hierarchical order (see figure 4).

4 Application & Evaluation

The Maritime Architecture Framework supports different functions. First, it ensures a common basis including consistent terminology to map existing architectural approaches and interfere between them.

Second, the framework can be used to support the development of an enterprise architecture by representing the technical aspects of systems and the interrelationships to organizational structures of the targeted field of application. This includes to follow a top-down approach, developing a business architecture and in relation to that, the belonging IT-environment. Otherwise, the framework is also proposed to follow a bottom-up approach to support the adaption of existing heterogeneous architecture into an e-Navigation compliant architecture to ensure interoperability with other systems.

First evaluations of the Maritime Architecture Framework are done as part of internal projects and interviews with experts of the maritime domain as well as experts for IT- and enterprise architectures. Regarding to this, the MAF is work in progress and has reached a stable maturity. Therefore, the framework is part of the 2nd round of experiments of the CPSE Labs Design Centre Germany North [CP16a]. The experiments focus on the development of different maritime systems in context of e-Navigation. The MAF will be used for evaluation on a larger scale by supporting the development of such Cyber-physical systems in the experiments [CP16b]. This includes the support for interoperability and integration for cooperation between the systems.

5 Conclusion

This paper addresses the need of a common methodology to align and integrate existing system architectures in the maritime domain. According to this, we present the Maritime Architecture Framework as a standardized methodology to assemble existing architectures in a meaningful and unique way to identify interoperability issues, interfaces and links to other (upcoming) systems. Thus, as an impact of the MAF, the gap between the visionary IMO strategy and the current problems of maritime players is handled to make the maritime domain more sustainable and future-ready. As first applications and evaluation steps show, the MAF is an effective, well suitable tool for

that case. In subsequent stages of the MAF it is used in larger scenarios and further improved.

6 Acknowledgement

The work presented in this paper is part of the maritime research at the University of Oldenburg and OFFIS, Germany. The authors of this document explicitly invite maritime stakeholders to contribute to the further development of this framework. Upcoming extensions and changes will be discussed in an open governance group. We thank the Ministry of Science and Culture of Lower Saxony for supporting us with the graduate school Safe Automation of Maritime Systems (SAMS).

7 References

- [CE12] Smart Grid Reference Architecture, CEN-CENELEC-ETSI Smart Grid Coordination Group, 2012.
- [CP16a] Cyber-Physical Systems Engineering Labs: 2nd Open Call for innovation projects, http://www.cpse-labs.eu/downloads/CPSELabs-2nd-open-call_CLOSED.pdf, accessed 10.05.2016.
- [CP16b] CPSE-Labs, www.cpse-labs.eu, accessed 09.05.2016.
- [DN15] SHIP CONNECTIVITY, DNV GL Strategic, 2015.
- [GW06] Gorschek, Tony; Wohlin, Claes: Requirements Abstraction Model, 2006.
- [HR15] Hankel, Martin; Rexroth, Bosch: Industrie 4.0: The Reference Architectural Model Industry 4.0 (RAMI 4.0), ZVEI - German Electrical and Electronic Manufacturers' Association, 2015.
- [IA15a] IALA Guideline 1113 on Design and Implementation Principles for Harmonised System Architecture of Shore-based Infrastructure, IALA, 2015.
- [IA15b] IALA Guideline 1114 on A Technical Specification for the Common Shore-based System Architecture (CSSA), IALA, 2015.
- [IE15] IEC 62559-2 Use case methodology – Part 2: Definition of the templates for use cases, actor list and requirements list, International Electrotechnical Commission, 2015.
- [MS09] MSC 85/26/Add.1, Annex 20 STRATEGY FOR THE DEVELOPMENT AND IMPLEMENTATION OF E-NAVIGATION, International Maritime Organization, 2009
- [MS14] MSC 94/18/8 Development and implementation of e-navigation', International Maritime Organization, 2014.
- [NC14] NCSR 1-28 - Report to the Maritime Safety Committee, International Maritime

Organization, Sub-Committee on navigation communications and search and rescue, 2014.

- [Se13] Sekimizu, Koji: A Concept Of A Sustainable Maritime Transportation System, World Maritime Day, 2013
- [TO16] The Open Group Architecture Framework, <http://pubs.opengroup.org/architecture/togaf8-doc/arch/>, accessed 09.05.2016
- [UE15] Uslar, Mathias; Engel, Dominik: Towards Generic Domain Reference Designation: How to learn from Smart Grid Interoperability, 2015.

Einsatz stilisierter Fakten zur Bewertung wandlungsfähiger Unternehmensarchitekturen

André Ullrich¹, Edzard Weber²

Abstract: Die Bewertung von Unternehmensarchitekturen ist abhängig von den angestrebten Zielen der Bewertung sowie von denen des betrachteten Unternehmens und kann anhand von Architekturmerkmalen, Qualitätseigenschaften oder Zielgrößen durchgeführt werden. Eine strategische Zielgröße ist die Wandlungsfähigkeit eines Unternehmens. Zur Anzeige dieser Eigenschaften können Indikatoren verwendet werden. Dies ermöglicht auch die Erfassung latenter Merkmale. Jedoch ist nicht einheitlich geklärt, welche Indikatoren wandlungsfähige Unternehmensarchitekturen erklären. Das Ziel des vorliegenden Beitrages ist es, die Identifikation von allgemein belastbaren Indikatoren zur Bewertung wandlungsfähiger Unternehmensarchitekturen vorzustellen. Zu diesem Zweck wird die Methode der Stilisierten Fakten angewendet. Dieses beginnt mit einem Literaturreview zur Erstellung der notwendigen Datenbasis. Im Ergebnis liegen Hypothesen in Form von stilisierten Fakten über Indikatoren wandlungsfähiger Unternehmensarchitekturen vor.

Keywords: Indikatoren, Stilisierte Fakten, Unternehmensarchitekturen, Wandlungsfähigkeit

1 Einleitung

Unternehmensarchitekturen sind aus informationstechnischer Sicht das „Zusammenwirken technologischer, organisatorischer und psychosozialer Aspekte bei der Entwicklung und Nutzung von betrieblichen soziotechnischen Informationssystemen“ [Gr06]. Technische, organisatorische und soziale Erfolgsfaktoren bedingen sich gegenseitig und dürfen nicht losgelöst voneinander betrachtet werden. Die Bewertung von Unternehmensarchitekturen ist abhängig von den angestrebten Zielen der Bewertung sowie von denen des betrachteten Unternehmens [Sa10]. Diese kann anhand von Architekturmerkmalen wie Modularität [Ko99], Qualitätseigenschaften wie Wartbarkeit [Gu08] sowie anhand von Zielgrößen wie Wirtschaftlichkeit [MS04] oder Effizienz [Fi03] durchgeführt werden. Einige dieser Architekturmerkmale sind Befähiger des Wandels. Wandlungsbefähiger charakterisieren individuelle und ungerichtete, abrufbare Eigenschaften von Wandlungsobjekten zum Wandel und tragen maßgeblich zum Wandel bei [He03]. Dabei bezeichnet eine Eigenschaft das, was einem Objekt (Person, Gegenstand, Begriff) zu eigen ist. Merkmale sind erkennbare Eigenschaften von Objekten zur Unterscheidung von anderen Objekten. Kriterien sind

¹ Universität Potsdam, Lehrstuhl für Wirtschaftsinformatik, insb. Prozesse und Systeme, August-Bebel-Str. 89, 14482 Potsdam, aullrich@lswi.de

² Universität Potsdam, Lehrstuhl für Wirtschaftsinformatik, insb. Prozesse und Systeme, August-Bebel-Str. 89, 14482 Potsdam, eweber@lswi.de

Merkmale, die als Grundlage für eine Entscheidung herangezogen werden.

Eine (weitere) organisationale Zielgröße ist die Eigenschaft, die Anforderungen der Umwelt (Turbulenzen) schnell, effizient und selbst handhaben zu können [Gr06]. Diese wird als Wandlungsfähigkeit bezeichnet. Das Äquivalent im Kontext kritischer Infrastrukturen ist die Resilienz. Resilienz ist die Fähigkeit eines Systems, Störungen und Fehlordnungen zu widerstehen [GP12], indem schädliche Ereignisse antizipiert werden sowie diesen vorgebeugt wird und das System selbstheilend agiert [Ho06].

Ein praktikabler Ansatz zur Bewertung ist die Verwendung von Indikatoren als Anzeiger von Architekturmerkmalen [AG05]. Ein Indikator ist ein Operator, der ein Indiz erzeugt. Ein Indiz stellt keinen Beweis dar, ist aber belastbarer als eine reine Vermutung. Indikatoren werden eingesetzt, wenn das eigentliche Betrachtungsmerkmal nicht direkt zu messen ist [Kr91]. Sie stellen eine Hilfsvariable dar bzw. liefern eine Hilfskennzahl. Der Indikator schlägt an, wenn ein bestimmter Schwellwert überschritten wird. Das ist dann ein Indiz dafür, dass z.B. im Kontext von Umweltturbulenzen Wandlungsbedarf oder sogar akuter Wandlungsdruck besteht. Ein Indikator kann also unterschiedliche Indizien erzeugen, je nachdem wie stark die Kennzahl ausschlägt. Darüber hinaus ermöglichen Indikatoren auch die Messung latenter Eigenschaften und sind dementsprechend insbesondere für den Einsatzzweck der Bewertung von Unternehmensarchitekturen geeignet. Jedoch stellt sich bei Ausrichtung eines Unternehmens anhand der Zielgröße der Wandlungsfähigkeit die Frage: Mittels welcher Indikatoren können wandlungsfähige Unternehmensarchitekturen bewertet werden?

Das Ziel des vorliegenden Beitrages ist es, die Identifikation von Indikatoren zur Bewertung wandlungsfähiger Unternehmensarchitekturen vorzustellen. Zu diesem Zweck wird die Methode der Stilisierten Fakten angewendet (nach dem Vorgehen von [HFL09]). Dieses beginnt mit einem Literaturreview zur Erstellung der notwendigen Datenbasis. Im Ergebnis liegen Hypothesen in Form von stilisierten Fakten über Indikatoren wandlungsfähiger Unternehmensarchitekturen vor.

Der Beitrag gliedert sich wie folgt: Problemstellung, Motivation und Zielstellung wurden in Abschnitt 1 dargelegt. Abschnitt 2 stellt den verwendeten Forschungsansatz dar. Darin erfolgt zuerst eine allgemeine Beschreibung des Konzepts der stilisierten Fakten sowie anschließend dessen Anwendung zur Generierung von Indikatoren zur Bewertung wandlungsfähiger Unternehmensarchitekturen. Abschnitt 3 umfasst die Ableitung der stilisierten Fakten am Beispiel der „Selbstorganisation“ sowie die Darstellung der Untersuchungsergebnisse. In Abschnitt 4 erfolgt die Diskussion der Ergebnisse und Abschnitt 5 gibt eine Zusammenfassung sowie einen Ausblick.

2 Forschungsansatz - Konzept der stilisierten Fakten

2.1 Allgemeine Beschreibung des Konzeptes

Für die Darstellung von Wandlungstreibern und Wandlungsbefähigern, die in der Literatur diskutiert werden, eignet sich prinzipiell die Methode der Stilisierten Fakten. Die Methode der Stilisierten Fakten wird seit einiger Zeit auch in der Wirtschaftsinformatik diskutiert [HFL11, Ze11]. Durch diese Methode werden neue Erkenntnisse aus der Zusammenfassung abstrahierter Einzeluntersuchungen gewonnen [Ti94, HFL09]. Diese Erkenntnisse zu einem bestimmten Betrachtungsgegenstand sind zwar einerseits weniger spezifisch, gründen aber andererseits auf einer breiteren Expertenbasis. Stilisierte Fakten bezeichnen charakteristische Merkmale, Wissens- und Erfahrungselemente, die empirisch breit gestützte, essentielle und generalisierte Eigenschaften eines Phänomens darstellen [Lo11]. Diese Elemente sollen erklärungsrelevant, methoden- und theorieneutral sowie nichtstrittig sein. Die Fakten bestehen aus: (1) wissenschaftlich fundierten Aussagen, welche aus dem zu bearbeiteten Themengebiet stammen, (2) Implikationen sowie (3) dem abgeleiteten stilisierten Fakt. Der Prozess der Generierung von stilisierten Fakten bedient sich vornehmlich Methoden der qualitativen Inhaltsanalyse, wobei auch formale oder quantitative Methoden zur Ableitung angewendet werden können [HFL09]. Die erforderlichen Implikationen werden subjektiv von unabhängigen Individuen gebildet. Dies stellt wesentliche Anforderungen an diese Individuen. Damit die Implikationen als gültig anerkannt werden können, müssen möglichst logische Verknüpfungen zwischen den Aussagen und den Implikationen bestehen. Erforderlich sind dabei tiefe Kenntnisse des Untersuchungsgegenstandes sowie die Anforderung, dass der Bearbeiter die Aussagen der jeweils entsprechenden Personen bei möglicher Mehrdeutigkeit richtig interpretiert.

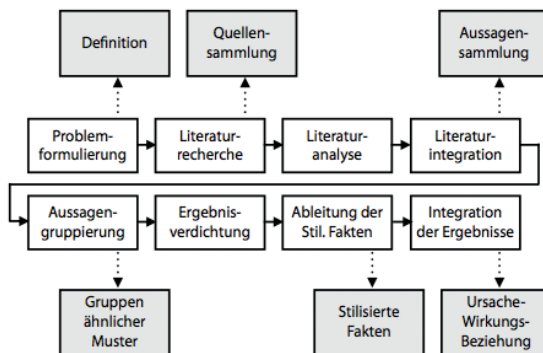


Abb. 1: Vorgehen zur Generierung stilisierter Fakten

Zur Bildung stilisierter Fakten sind mehrere Schritte notwendig (vgl. Abb. 1). Zuerst muss die Problemstellung präzise formuliert werden, um anschließend die Literatur nach relevanten Aussagen, welche sowohl im weiteren als auch im engeren Sinne in den

Problembereich fallen, zu sammeln und zu analysieren. Dabei werden einzelne im Detail auch unterschiedliche Aussagen zusammengefasst. Dies ermöglicht die Extraktion gebündelter und abstrahierter Eigenschaften. Wobei durch inhaltliche Übereinstimmung von unterschiedlichen Aussagen sowie durch ein hinlängliches Maß an Repräsentativität diese abstrahierten Eigenschaften rechtfertigbar sind [Ze11, HFL11]. Durch die Prozessschritte Literaturanalyse, Literaturintegration sowie Aussagengruppierung entsteht eine gruppierte Aussagensammlung zu einem Phänomen. Anhand derer wird eine Ergebnisverdichtung mittels Inhaltsanalyse durchgeführt. Darauf aufbauend kann die Ableitung der stilisierten Fakten erfolgen. Gemäß [HFL09] sind aus diesen auch Ursachen-Wirkungs-Zusammenhänge deutbar. Der Ableitungsprozess der Bildung der stilisierten Fakten ist durch entsprechenden Konsens von Fachexperten abzusichern.

An die Generierung von stilisierten Fakten werden bestimmte Grundanforderungen gestellt, die eine akzeptable und methodisch saubere Ableitung gewährleisten sollen [Sc01]: (1) Offenlegung des Generierungsprozesses, (2) methodenübergreifende Faktengenerierung, (3) Vollständigkeit der Faktenmenge, (4) hinreichender Konsens über jeden Fakt und (5) Offenheit der Faktenmenge. Die Einhaltung dieser Forderungen sowie der Schritte der Vorgehenssystematik sichern eine hohe intersubjektive Nachvollziehbarkeit der generierten Ergebnisse. Wesentlich ist, dass stilisierte Fakten durch die oben beschriebene Vorgehenssystematik erstellt und mittels Expertenkonsens legitimiert werden. Dabei beurteilen die Experten eines Forschungsfeldes die Qualität einer Implikation. Diese Fakten sind jedoch nicht als Gesetze oder Regeln zu verstehen, sondern vielmehr als schwerwiegend begründete und legitimierte Vermutungen. Dementsprechend ist der Anspruch kausale Zusammenhänge durch stilisierte Fakten aufzudecken, nicht zu halten. Dieser Zusammenhang ist nicht aus empirischen Daten herleitbar, da Kausalität bei Ursache und Wirkung nicht physikalischer Gesetzmäßigkeiten wiederum eine Implikation darstellen würde [Wi09]. So kann lediglich die Korrelation von empirischen Daten und der Generalisierung der Zusammenhänge aufgedeckt werden. Ein weiterer nicht unumstrittener Aspekt ist die Verwendung empirischer Daten aus heterogenen Quellen. Entsprechend des Konzepts ist es unerheblich, ob die Erkenntnisse aus Grundlagenforschung, anwendungsorientierter Forschung oder aus Lehrbuchwissen stammen. Die unterschiedlichen Quellen werden als etablierte und konsolidierte Expertenauffassung betrachtet. Ein weiteres Problem ist die Anzahl potenzieller Quellen stilisierter Fakten. In der Wirtschaftsinformatik existieren noch nicht ausreichend empirische Datensätze (auch Sekundärdaten), um daraus effektiv stilisierte Fakten bilden zu können [Lo11]. Darüber hinaus könnten die gebildeten stilisierten Fakten als vereinfachte Alternativen zu Theorien verstanden werden, was möglicherweise die Integration von verfügbarem theoretischem Wissen in gestaltungsorientierte Vorhaben erschwert.

Subjektive Einflüsse auf die Generierung stilisierter Fakten können durch strukturierte und offen dokumentierte Prozessschritte aufgefangen werden [Lo11]. Ebenfalls wichtig ist die Bemessung des Suchrasters der Aussagen. Ist dieses zu eng formuliert, bleiben relevante Erkenntnisse aus benachbarten Wissensbereichen unbeachtet. Ist es hingegen

zu weit formuliert, so ist die Extraktion aller relevanten Aussagen komplex und ressourcenaufwändig.

2.2 Vorgehen zur Identifikation von Indikatoren aus stilisierten Fakten

Das vorhandene Wissen zum vorliegenden Untersuchungsgegenstand - den Indikatoren der Wandlungsfähigkeit - liegt in großem Umfang in Form von (wissenschaftlichen) Texten vor. Zur Gewährleistung einer systematischen Vorgehensweise wird die Stichprobe der zu untersuchenden Beiträge in Anlehnung an das Vorgehensmodell zur Ableitung der stilisierten Fakten ermittelt sowie die Inhalte der Beiträge analysiert. Zur Durchführung qualitativer Inhaltsanalysen wird das vorliegende Material hinsichtlich einer vorher identifizierten Fragestellung untersucht [Di07]. Die entsprechende Fragestellung für die Problemstellung der Identifikation Stilisierter Fakten hinsichtlich von Indikatoren zur Bewertung wandlungsfähiger Unternehmensarchitekturen lautet: Welche Aussagen beinhalten potenzielle Informationen über Indikatoren zur Bewertung wandlungsfähiger Unternehmensarchitekturen? Vor dem Hintergrund dieser Fragestellung erfolgt eine explorative Sichtung des relevanten Materials. Dabei ist das Material zum Zweck der Identifikation möglichst aller untersuchungsrelevanten potenziellen Fakten jeweils mindestens zweimal gesichtet worden. Zur Faktenerhebung werden wissenschaftliche Beiträge zu den Themenbereichen Wandlungsfähigkeit und Resilienz mittels eines Reviews untersucht. Dabei werden diese Beiträge bezüglich der expliziten und impliziten Nennung von Indikatoren und Befähigern betrachtet.

Die Methode des Review untersucht von einer bestimmten Fragestellung ausgehend die relevanten Arbeiten sowie Erkenntnisse eines Themengebiets und ist durch zwei Merkmale charakterisiert: (1) „Ein Review basiert auf einer Menge von Primäruntersuchungen zu einer oder mehreren ähnlichen Forschungsfrage(n).“ (2) „Ein Review zielt darauf ab, die Ergebnisse ausgewählter Primäruntersuchungen zu beschreiben, zusammenzufassen, zu bewerten, zu klären oder zu integrieren“ [Fe06].

Forschungsgebiet	Format	<i>n</i> vor Relevanzprüfung	<i>n</i> nach Relevanzprüfung
Wandlungsfähigkeit	Journal	56	46
	Konferenz	20	16
	Buch(beitrag)	37	29
	Bericht o.a.	5	4
Resilienz	Journal	9	8
	Konferenz	2	2
	Buch(beitrag)	3	2
	Bericht o.a.	0	0
Summe		132	107

Tab. 1: Deskriptive Auswertung zur Literaturbasis der stilisierten Fakten

Die verwendete Datenbasis wurde systematisch in Literaturdatenbanken (econis, genios, scholar.google) mittels Schlagwortsuche (Wandlungsfähigkeit, Wandlungsbefähiger, wandlungsfähige Architekturmerkmale, wandlungsfördernde Eigenschaften, Resilienz, resilienzfördernde Eigenschaften; sowie deren englischsprachige Äquivalente) recherchiert und bezüglich ihrer Relevanz für die Ableitung stilisierter Fakten gemäß den oben skizzierten Anforderungen überprüft. Dies ist zulässig, da an ein Review nicht die Anforderung gestellt wird, sämtliche Arbeiten zu einem Themengebiet zu berücksichtigen [Fe06] und die dort erarbeitete Literatur eine sowohl breite als auch tiefe themenspezifische Datenbasis darstellt. Insgesamt wurden 132 Beiträge aus dem Zeitraum von 1993 bis 2014 gesichtet. Nach einer Relevanzprüfung konnten noch 107 Beiträge berücksichtigt werden (Tab. 1). Die Beschränkung des Zeitraums führt einerseits dazu, dass möglicherweise potenzielle Fakten nicht berücksichtigt werden, andererseits wird somit ein hoher Gegenwartsbezug gewährleistet. Die betrachteten Beiträge umfassen sowohl deutsch- als auch englischsprachige Journal-, Konferenz- und Buchbeiträge sowie Monographien (die letzten beiden sind zur Kategorie „Buch(beitrag)“ zusammengefasst).

Die Untersuchung und Bewertung des bestehenden Konsenses zu einem stilisierten Fakt (Konsensanalyse) erfordert ein Klassifikationskriterium. Dieses ist in der vorliegenden Untersuchung die Anzahl der Nennungen von Indikatoren oder Befähigern der Wandlungsfähigkeit in Beiträgen von Experten auf den Feldern der Wandlungsfähigkeits- und Resilienzforschung, die den jeweiligen Fakt stützen. Die Anzahl von übereinstimmenden Aussagen ermöglicht hinreichende Rückschlüsse auf die Qualität der stilisierten Fakten [WL07a]. Die jeweiligen identifizierten Fakten werden - in Analogie zu [HFL09] - als gut gestützt (A), durchschnittlich gestützt (B) sowie schwach gestützt (C) klassifiziert. Die stilisierten Fakten der Kategorien B und C sind im Vergleich zu den Fakten aus Kategorie A einer weiteren Überprüfung bedürftiger. Dies impliziert jedoch nicht, dass es sich dabei um unbegründete Zusammenhänge handeln muss. Vor allem die stilisierten Fakten der Kategorie C stellen angenommene Tendenzen dar, die es zukünftig anzunehmen oder abzulehnen gilt. Aus diesem Grund werden in der vorliegenden Untersuchung die Fakten der Kategorien A und B zur Weiterverarbeitung genutzt. Für die Einteilung der Fakten stehen zwei unterschiedlichen Kategorisierungsprinzipien zur Auswahl. Die absolute Kategorisierung fordert für jede Güte-Kategorie eine Mindestanzahl an Quellen ein, die in einem angemessenen Verhältnis zur Gesamtquellenanzahl steht. Eine relative Kategorisierung legt fest, wie viele Fakten (absolute oder prozentuale Anzahl) einer Kategorie zugehörig sein sollen. Die relative Kategorisierung kann dazu führen, dass ursprüngliche A-Fakten durch zusätzliche Quellenauswertung zu B-Fakten abgewertet werden. Es ist aber trotzdem gewährleistet, dass die besser belegten Fakten höher einkategorisiert werden. Die Grenzen der hier verwendeten Kategorien sind wie folgt festgelegt:

- Kategorie A: mindestens 9 Quellen, die eine Aussage stützen,
- Kategorie B: mindestens 3 Quellen, die eine Aussage stützen sowie

- Kategorie C: mindestens 1 Quelle, die eine Aussage stützt.

Experten greifen in der Regel ihre Erkenntnisse aus vorangegangenen Publikationen wieder auf. Dies ist im Kontext der Konsensanalyse zu berücksichtigen. Zur Vermeidung von Doppelzählungen werden aus diesem Grund unterschiedliche Beiträge desselben Autors oder derselben Autorengruppe zur gleichen Aussage lediglich einmal in der Zählung erfasst.

Die Gruppierung der identifizierten Aussagen erfolgt anhand von gemeinsamen Implikationen der jeweiligen Inhalte. Dabei wird eine Verdichtung der Ergebnisse erzielt, indem zu inhaltlichen Aussagen Implikationen gebildet werden. Aus diesen Implikationen werden die stilisierten Fakten abgeleitet. Eine explizite Integration der Ergebnisse in Form von Ursache-Wirkungsbeziehungen kann durch die Methode der Stilisierten Fakten nicht direkt geleistet werden. Um dazu belastbare Erkenntnisse zu generieren, sind eigenständige empirische Untersuchungen oder Simulation notwendig. Beispielsweise sind Wechselwirkungen zwischen den Indikatoren sowie unterschiedliche Gewichtungen dieser nicht ohne weiteres ableitbar.

Entsprechend des angewandten Vorgehens sowie den Ausführungen im vorliegenden und im folgenden Abschnitt werden die fünf Grundanforderungen an die Generierung von stilisierten Fakten (Offenlegung des Generierungsprozesse, methodenübergreifende Faktengenerierung, Vollständigkeit der Faktenmenge, hinreichender Konsens über jeden Fakt, Offenheit der Faktenmenge) adressiert. Einzig die Vollständigkeit der Faktenmenge wird durch die Eingrenzung des Betrachtungszeitraum bewusst beschränkt.

3 Stilisierte Fakten der Indikatoren zur Bewertung wandlungsfähiger Unternehmensarchitekturen

Gemäß der im vorhergehenden Abschnitt dargestellten Vorgehensweise wurden Indikatoren zur Bewertung wandlungsfähiger Unternehmensarchitekturen als stilisierte Fakten ermittelt. Zur besseren Nachvollziehbarkeit wird das Vorgehen am Beispiel des stilisierten Faktus „*Selbstorganisation ist ein Indikator der Wandlungsfähigkeit*“ verdeutlicht (Abb. 2). Die Sichtung der Literatur fokussierte insbesondere auf Indikatoren, Befähigern und fördernde Eigenschaften der Wandlungsfähigkeit. Aussagen wie „*Aus der Fabrikplanung stammen die Indikatoren Skalierbarkeit, Modularität, Mobilität und Interoperabilität. Die Betrachtung autopoietischer Systeme fügt die Indikatoren der Selbstorganisation, Selbstähnlichkeit und Redundanz hinzu.*“ [AGS05] wurden gesammelt, daraus allgemeine Implikationen („*Neben anderen ist Selbstorganisation ein Indikator der Wandlungsfähigkeit.*“) abgeleitet, sodass diese zu einem stilisierten Fakt gebündelt werden können.

Aussage	Implikation	Stilisierte Fakt
<p>„Aus der Fabrikplanung stammen die Indikatoren Skalierbarkeit, Modularität, Mobilität und Interoperabilität. Die Betrachtung autopoietischer Systeme legt die Indikatoren der Selbstorganisation, Selbstähnlichkeit und Redundanz hinzu.“ [AGS05]</p> <p>„That autonomy in living systems is a feature of self-production (autopoiesis), and that a living system is properly characterized only as a network of processes of production of components that is continuously, and recursively, generated and realized as a concrete entity (unity) in the physical space, by the interactions of the same components that it produces as such a network.“ [Ma79]</p> <p>„In den unterschiedlichen Unternehmensbereichen wie Einkauf, Produktion oder Logistik wirken sich diese Umweltturbulenzen unterschiedlich stark aus. Sind Bereiche nicht in der Lage, diese Turbulenzen zu beheben, werden sie sich verändern müssen, wenn das Unternehmen auch zukünftig erfolgreich produzieren will.“ [WJ07b]</p> <p>„Selbststeuerung kann als Organisationsprinzip begriffen werden und könnte aus Sicht der Fabrikplanung als Konzept zur operativen Realisierung von Wandlungsfähigkeit eingesetzt werden. Beide Konzepte soll ein Produktionssystem zur Bewältigung der gestiegenen Anforderungen befähigen.“ [SS11]</p> <p>„Im Konzept des wandlungsfähigen Unternehmens dominieren die Prinzipien der geplanten Evolution und der gelenkten Selbstorganisation. Dies impliziert, dass neben der Selbstorganisation -als notwendige Voraussetzung und Garant der Wandlungsfähigkeit- ein gewisses Minimum an Fremdorganisation unabhängig ist.“ [ER07]</p> <p>„Durch den Transfer der Idee der Selbstorganisation auf betriebswirtschaftliche Systeme erfordern sich die Anwendung von verbesserten Anpassungs- und Reaktionsfähigkeit auf die sie umgebenden komplexen dynamischen Umweltbedingungen und erfordern so eine höhere Robustheit des Systems“ [HW06].</p> <p>Die Eigencharakter der Wandlungsfähigkeit der Wandlungsfähigkeit von Schutzsystemen im Bereich Kritischer Infrastrukturen wird anhand folgender Indikatoren gemessen: ... Selbstorganisation ... [MSW10]</p>	<p>Lebende Systeme organisieren sich und ihr Fortbestehen selbst.</p> <p>Dies unternehmerische Subsysteme unabhängig von anderen Subsystemen agieren können, ist eine notwendige Bedingung dafür, dass sie mit ihren betriebswirtschaftlichen Turbulenzen umgehen können.</p> <p>Selbststeuerung kann als ein Konzept zur Realisierung operativer Wandlungsfähigkeit genutzt werden.</p> <p>Selbstorganisation ist ein dominierendes Prinzip eines wandlungsfähigen Unternehmens</p> <p>Selbstorganisation wirkt positiv auf Anpassungs- und Reaktionsfähigkeit eines betriebswirtschaftlichen Systems</p> <p>Selbstorganisation ist ein Indikator der Wandlungsfähigkeit</p>	<p>Selbstorganisation ist ein Indikator für Wandlungsfähigkeit</p>

Abb. 2: Stilisierter Fakt: Selbstorganisation ist ein Indikator der Wandlungsfähigkeit

Entsprechend der Anzahl vorhandener Nennungen (Klassifikationskriterium), die zu einem Stilisierten Fakt gebündelt werden können, erfolgt die Kategorisierung der Fakten (Tab. 2). Interessierte Leser können sich hinsichtlich einer nachvollziehbaren Darstellung der verwendeten Originalzitate für die Ableitung der stilisierten Fakten an den Erstautor dieses Beitrags wenden.

Stilisierte Fakten der WF-Indikatoren	Kategorie	Belegende Quellen
Modularität	A	[EI05, KS11, HW05, AGS05, Sp07, KN09, Du01, SS07, NK02, Ne01, BK04, Wi06, DEK10, KG13, Wu08, RSW10]
Skalierbarkeit	A	[Wi07, Du01, AGS05, KS11, KN09, DEK10, RSW10, EI05, HW05, Wu08, KG13, SS07,

Mobilität	A	He07] [AGS05, EI05, KG13, KN09, Wu08, SS07, HW05, Sp07, He07]
Universalität	B	[EI05, Wu08, Wi07, Sp07, SS07, He07, KG13, KN09]
Kompatibilität	B	[KN09, EI05, Wu08, KG13, SS07, Sp07, He07]
Selbstorganisation	B	[AGS05, Ma75, WL07b, SS11, ER07, HW06, RSW10]
Integrationsfähigkeit	B	[Du01, HW05, KS11, DEK10, EI05]
Neutralität	B	[HW05, Wi07, AG04, He07]
Dezentralisierung	B	[NK02, BK04, AGS05, Wi06]
Standardisierung	B	[NK02, AGS05, Wu08, He07]
Wissen	B	[RSW10, WT04, Gr08, NK02]
Handlungsspielraum	B	[vP66, RRH08, GW09, USK13]
Interoperabilität	B	[AGS05, SS11, RSW10]
Konvertierbarkeit	B	[KS11, Wi07, EI05]
Redundanz	B	[AGS05, RSW10, WF03]
Selbstähnlichkeit	B	[AGS05, Du01, RSW10]
Verfügbarkeit	B	[RSW10, Gr08, Wu08]

Tab. 2: Stilisierte Fakten: Indikatoren wandlungsfähiger Unternehmensarchitekturen

Besonders die deutschsprachige Literatur ist stark von den frühen Arbeiten [z.B. Re97, WS98, BWW98], die den Begriff der Wandlungsfähigkeit explizit eingeführt haben, geprägt. Die dort behandelten Indikatoren wurden von nachfolgenden, voneinander unabhängigen Forschungsprojekten als Ausgangslösung aufgegriffen, variiert und ergänzt. Dadurch erklärt sich der harte Kern der häufig gemeinsam und intensiv untersuchten Indikatoren aus der A-Kategorie (siehe Tab. 2).

4 Diskussion

Wandlungsfähigkeit über eine bestimmte Menge von Indikatoren nachzuweisen, stellt nur eine mögliche Art von Erklärungsmodellen zur Wandlungsfähigkeit dar [We15]. Tatsächlich werden mehrere Merkmale betrachtet, welche jeweils mit unterschiedlichen definitorischen Merkmalen des Betrachtungsgegenstandes korrelieren. Alternative Ansätze betrachten einzelne definitorische Indikatoren oder schlussfolgernde Indikatoren [Kr91] nach [No63]. Definitorische Indikatoren werden direkt aus der jeweiligen Definition von Wandlungsfähigkeit abgeleitet bzw. werden als konstituierende Merkmale für diesen Begriff festgesetzt. Da es jedoch unterschiedliche Definitionen gibt, besteht darüber keine Möglichkeit einen allgemein verwendbaren Indikatoren-Satz zusammenzustellen. Schlussfolgernden Indikatoren fehlt jeglicher direkte Bezug zum Begriff der Wandlungsfähigkeit. Ihr kausaler Zusammenhang ist deshalb nicht offensichtlich und nicht immer vorhanden, sodass aus ihnen kein allgemein belastbarer

Indikatoren-Satz zusammengestellt werden kann. Darüber hinaus ist die spezialisierte Betrachtung eines einzelnen korrelierenden Indikators möglich. Problematisch kann hierbei eine durch die Spezialisierung ungeeignete Übertragbarkeit auf unterschiedliche Anwendungskontexte sein. Der Einsatz einer Menge von Indikatoren kann fallspezifische Anwendungsmängel einzelner Indikatoren ausgleichen, ist unabhängig von der zugrunde liegenden Begriffsauffassung zur Wandlungsfähigkeit und besitzt trotzdem einen kausalen Bezug zur Wandlungsfähigkeit.

Der Einsatz der Methode der Stilisierten Fakten ist nur ein erster Schritt bei der Zusammenstellung eines Indikatoren-Satzes zum Nachweisen und Bewerten von wandlungsfähigen Unternehmensarchitekturen. Der Erkenntnisgewinn besteht darin, welche Indikatoren als Elemente für diese Menge überhaupt als Kandidaten in Betracht gezogen werden sollen. Es sind diejenigen Indikatoren, die durch vorangegangene Arbeiten bereits eine wissenschaftliche Belastungsprobe und Anerkennung erfahren haben. Erst in einem weiteren Schritt können die Wirkbeziehungen (und Kausalität) zwischen den ausgewählten Indikatoren, mögliche Fehl-/Doppelmessungen sowie die Aussagekraft des Indikatoren-Satzes in seiner Gesamtheit ermittelt werden. Der Einsatz stilisierter Fakten schließt nicht aus, dass bisher wenig untersuchte Indikatoren ebenfalls als universelle Indikatoren geeignet sind. Diese Analyse stellt einen nachfolgenden Optimierungsschritt dar, um den Indikatoren-Satz hinsichtlich der Aussagekraft und Messgenauigkeit zu steigern. Die Nachrückkandidaten sind als Ergebnis der Methode der Stilisierten Fakten je nach Kategorie in einer White List (auf jeden Fall untersuchen), Grey List (bei vorhandenen Kapazitäten untersuchen) oder Black List (untersuchen, wenn genug neue Fakten hinzukommen) geordnet. Ausgangspunkt muss aber die initiale Zusammenstellung der belastbarsten Einzelindikatoren sein, weil dadurch der größte Grenznutzen zu erwarten ist.

5 Zusammenfassung und Ausblick

Die Güte von Unternehmensarchitekturen ist nicht direkt messbar, sodass zwangsläufig auf Indikatoren zurückgegriffen werden muss. Einerseits ist es wichtig, dass dafür neue Indikatoren entwickelt und erprobt werden. Andererseits dürfen die allgemeinen Vorarbeiten aber nicht unberücksichtigt gelassen werden. Dafür darf nicht nur die Aussagekraft einzelner Indikatoren betrachtet sondern das Zusammenwirken von Indikatoren muss verstanden werden. Die Methode der Stilisierten Fakten kann zum einen dazu beitragen, die am stärksten belastbaren und anerkannten Indikatoren für eine weiterführende Wirkungsanalyse zusammenzustellen. Zum anderen kann sie geeignete Nachrückkandidaten anbieten, wenn die Wirkungsanalyse zu einer bestimmten Indikatorenmenge zu große inhaltliche Überschneidungen oder Fehlstellen ermittelt hat.

Literaturverzeichnis

- [AG04] Andresen, K.; Gronau, N.: Der Faktor Wandlungsfähigkeit bei der Planung neuer Fabriken - Ein Marktüberblick von Unternehmensberatungen im Bereich Fabrikplanung. *Industrie Management*, 20/04, S. 60-68, 2004.
- [AG05] Andresen, K.; Gronau, N.: Die Gestaltung von Unternehmensarchitekturen - Wieviel Flexibilität ist notwendig. *ERP Management*, 1/05, S. 30-32, 2005.
- [AGS05] Andresen, K.; Gronau, N.; Schmid, S.: Ableitung von IT-Strategien durch Bestimmung der notwendigen Wandlungsfähigkeit von Informationssystemarchitekturen. In (Ferstl, O. et al.): *Wirtschaftsinformatik 2005: eEconomy, eGovernment, eSociety*. Physica Verlag, 2005.
- [BK04] Blecker, D.; Kaluza, B.: Heterarchische Hierarchie: Ein Organisationsprinzip flexibler Produktionssysteme. TCW-Verlag, 2004.
- [BWW98] Balve, P.; Wiendahl, H.H.; Westkämper, E.: Auftragsmanagement in wandlungsfähigen Unternehmensstrukturen - Anforderungen und Ansätze. *PPS-Management* 17/98, S. 461-468, 1998.
- [DEK10] Denkena, B.; Eikötter, M.; Kröning, S.: Wandlungsfähige Produktionssysteme mit Hilfe von Prozesskettensimulation. *Agile Produktionslogistik und Transportanlagen*, ZWF - Zeitschrift für wirtschaftlichen Fabrikbetrieb, 105/10, S. 47-51, 2010.
- [Di07] Diekmann, A.: *Empirische Sozialforschung - Grundlagen, Methoden, Anwendungen*. Rowohlt, 2007.
- [Du01] Dürschmidt, S.: *Planung und Betrieb wandlungsfähiger Logistiksysteme in der variantenreichen Serienproduktion*, Herbert Utz Verlag, 2001.
- [El05] ElMaraghy, H.: Flexible and reconfigurable manufacturing systems paradigms. *International journal of flexible manufacturing systems*, 17/05, S. 261-276, 2005.
- [ER07] Eisenberg, W.; Renner, U. *Naturwissenschaftliche und technische Systeme im Fokus von Fremd- und Selbstorganisation*, Symposium 2005, Leipziger Universitätsverlag.
- [Fe06] Fettke, P.: State-of-the-art des state-of-the-art. *Wirtschaftsinformatik*, 48/06, S. 257-266, 2006.
- [Gr06] Gronau, N.: *Wandlungsfähige Informationssystemarchitekturen: Nachhaltigkeit bei organisatorischem Wandel*, Gito-Verlag, Berlin, 2006.
- [Gr08] Gronau, N.: IT-Business Alignment und Wandlungsfähigkeit von Informationssystemen, *Industrie Management*, 24/08, S. 11-13, 2008.
- [GW09] Gronau, N.; Weber, E.: *Wandlungsfähigkeit: Generische Strategien zur Handhabung von Veränderungen in der Umwelt*. Lehrstuhl für Wirtschaftsinformatik und Electronic Government, Universität Potsdam, Arbeitsbericht WI-2009-07, 2009.
- [GP12] Gunderson, L. H.; Pritchard, L.: *Resilience and the behavior of large-scale systems*, volume 60. Island Press, 2012.
- [Gu08] Gustafsson, P.; Franke, U.; Höök, D.; Johnson, P.: Quantifying IT impacts on organizational structure and business value with Extended Influence Diagrams. In:

- (Stirna, J. et al. Hrsg.) Proceedings of First IFIP WG 8.1 Working Conference: The Practice of Enterprise Modeling (PoEM2008), S. 138-152, 2008.
- [He03] Hernández, R.: Systematik der Wandlungsfähigkeit in der Fabrikplanung, VDI Verlag, 2003.
- [He07] Heger, C.: Bewertung der Wandlungsfähigkeit von Fabrikobjekten, PZH Verlag, 2007.
- [HFL09] Houy, C.; Fettke, P.; Loos, P.: Stilisierte Fakten der ereignisgesteuerten Prozesskette - Anwendung einer Methode zur Theoriebildung in der Wirtschaftsinformatik. In: (Nüttgens et al Hrsg.) EPK 2009. GI-Workshop "Geschäftsprozessmanagement mit Ereignisgesteuerten Prozessketten" (WI-EPK-09), 2009.
- [HFL11] Houy, C.; Fettke, P.; Loos, P.: Stilisierte Fakten in der gestaltungsorientierten Wirtschaftsinformatik - Allgemeine Potentiale und erste Erfahrungen. In: (Bernstein, A. et al. Hrsg.) Wirtschaftsinformatik Proceedings, 2011.
- [Ho06] Hollnagel, E.; Woods, D. D.; Leveson, N.: Resilience engineering: Concepts and precepts. Ashgate Publishing, 2006.
- [HW05] Hernandez, R.; Wiendahl, H.: Die wandlungsfähige Fabrik - Grundlagen und Planungsansätze. In (Kaluzna, B.; Blecker, T. Hrsg.) Erfolgsfaktor Flexibilität - Strategien und Konzepte für wandlungsfähige Unternehmen. E. Schmidt Verlag, 2005.
- [HW06] Hülsmann, M.; Wycisk, C.: Selbstorganisation als Ansatz zur Flexibilisierung der Kompetenzstrukturen. In: (Burmman, C. et al. Hrsg): Neue Perspektiven des strategischen Kompetenz-Managements, S 324-350. DUV Gabler 2006.
- [KG13] Krebs, M.; Gossman, D: Wandlungsfähigkeit gestalten mit System. In: (Nyhuis, P. et al. Hrsg.) Wandlungsfähige Produktion heute für morgen gestalten, PZH Verlag, 2013.
- [KN09] Klemke, T.; Nyhuis, P.: Lean changeability - evaluation and design of lean and transformable factories. World Academy of Science, Engineering and Technology, 53, S. 653-660, 2009.
- [Ko99] Koren, Y.; Heisel, U.; Jovane, F.; Moriwaki, T.; Pritschow, G.; Ulsoy, G.; Van Brussel, H.: Reconfigurable manufacturing systems. CIRP Annals-Manufacturing Technology, 48(2), S. 527-540, 1999.
- [Kr91] Kromrey, H.: Empirische Sozialforschung. Leske+Budrich, 1991.
- [KS11] Koren, Y.; Shpitalni, M.: Design of reconfigurable manufacturing systems. Journal of manufacturing systems, 29/11, S. 130-141, 2011.
- [Lo11] Loos, P.; Fettke, P.; Weißenberger, B. E.; Zelewski, S.; Heinzl, A.; Frank, U.; Iivari, J.: Welche Rolle spielen eigentlich stilisierte Fakten in der Grundlagenforschung der Wirtschaftsinformatik? Wirtschaftsinformatik, 53/11, S. 109-121, 2011.
- [Ma75] Maturana, H. R.: The organization of the living: A theory of the living organization. Int. J. Man-Machine Studies, 7/75, S. 313-332, 1975.
- [MS04] Morganwalp, J. M.; Sage, A. P.: Enterprise architecture measures of effectiveness. International Journal of Technology, Policy and Management, 4/04, S. 81-94, 2004.
- [Ne01] Neuhausen, J.: Methodik zur Gestaltung modularer Produktionssysteme für Unternehmen der Serienproduktion. Bibliothek der RWTH Aachen, 2001.

- [NK02] Nofen, D.; Klußmann, J.: Wandlungsfähigkeit durch modulare Fabrikstrukturen. *Industrie Management*, 3/02, S. 49-52, 2002.
- [No63] Nowak, S.: Correlational, definitional, and inferential indicators. *Polish Sociological Bulletin*, 2/63, S. 31-53, 1963.
- [Re97] Reinhart, G.: Innovative Prozesse und Systeme - Der Weg zu Flexibilität und Wandlungsfähigkeit. In: (Milberg, A. et al. Hrsg.) *Mit Schwung zum Aufschwung*, Verlag Moderne Industrie, S. 173-202, 1997.
- [RRH08] Ross, A.; Rhodes, D.H.; Hastings, D.: Defining Changeability: Reconciling Flexibility, Adaptability, Scalability, Modifiability, and Robustness for Maintaining System Lifecycle Value. *Systems Engineering* 3/08, S. 246-262, 2008.
- [RSW10] Röchert-Voigt, T.; Stein, M.; Weber, E.: *Wandlungsfähige Schutzstrukturen - Handlungsleitfaden*. GITO Verlag, 2010.
- [Sa10] Saat, J.: *Planung der Unternehmensarchitektur - Vorgehen-Gestaltungsgegenstand-Alternativenbewertung*. Logos Verlag, 2010.
- [Sc01] Schwerin, J.: *Wachstumsdynamik in Transformationsökonomien: Strukturähnlichkeiten seit der industriellen Revolution und ihre Bedeutung für Theorie und Politik*. Böhlau Verlag, 2001.
- [Sp07] Spur, G.: Wandlungsfähigkeit von Produktionssystemen. *ZWF*, 3/07, S. 96-97, 2007.
- [SS07] Specht, D.; Stefanska, R.: Wandlungsfähige Fabrikstrukturen als Strategie: Zur Steigerung der Wettbewerbsfähigkeit von produzierenden Unternehmen. *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, 05/07, S. 286-290, 2007.
- [SS11] Scholz-Reiter, B.; Sowade, S.: Wandlungsfähigkeit durch selbststeuernde Produktionssysteme - ein Beispiel aus der Bekleidungsindustrie. *Industrie Management*, 3/11, S. 9-12, 2011.
- [Ti94] Tichy, G.: *Konjunktur - Stilisierte Fakten, Theorie, Prognose*. Springer Verlag, 1994.
- [USK13] Ullrich, A.; Sembritzki, U; Skupsch, K.: Störungsmanagement im saisonalen Umfeld. *Productivity Management* 05/13, S. 45-48, 2013.
- [vP66] Parreren, C.v.: *Lernprozess und Lernerfolg*. Westermann Verlag, 1966.
- [We15] Weber, E.: *Erarbeitung einer Methodik der Wandlungsfähigkeit*. GITO Verlag, 2015.
- [WF03] Wiendahl, H.; Fiebig, C.: Kooperation von Fabrik- und Technologieplanung - technologiepotentiale frühzeitig erkennen. *Werkstattstechnik online*, 93/03, S. 233-237, 2003.
- [Wi06] Wilke, M.: Wandelbare automatisierte Materialflusssysteme für dynamische Produktionsstrukturen. *Utz*, 2006.
- [Wi07] Wiendahl, H.; ElMaraghy, H.; Nyhuis, P.; Zäh, M.; Wiendahl, H.; Duffie, N.; Bricke, M.: Changeable manufacturing - classification, design and operation. *CIRP Annals-Manufacturing Technology*, 2/07, S. 783-809, 2007.
- [Wi09] Winter, R.; Krcmar, H.; Sinz, E.; Zelewski, S.; Hevner, A.: Was ist eigentlich Grundlagenforschung in der Wirtschaftsinformatik? *Wirtschaftsinformatik*, 2/09, S.

223-231, 2009.

- [WL07a] Weißenberger, B.; Löhr, B.: Planung und Unternehmenserfolg: Stylized facts aus der empirischen Controllingforschung im deutschsprachigen Raum von 1990-2007. Zeitschrift für Planung & Unternehmenssteuerung, 4/07, S. 335-363, 2007.
- [WL07b] Wockenfuß, M.; Lämmer, A.: Analyse der Wandlungsfähigkeit des ERP-Systems greenax. In: (Gronau, N. et al. Hrsg.) Wandlungsfähige ERP-Systeme: Entwicklung, Auswahl und Methoden, Gito Verlag 2007.
- [WS98] Wiendahl, H.P.; Scheffczyk, H.: Wandlungsfähige Fabrikstrukturen. wt Werkstattstechnik, 4/98, S. 171-175, 1998.
- [WT04] Warneke, G.; Thurnes, C.: Wandlungsfähig durch Kompetenzmanagement. Industrie Management 2/04, S. 9-11, 2004.
- [Wu08] Wulfsberg, J.; Redlich, T.; Lehmann, J.; Bruhns, F.: Square foot manufacturing - Ein wandlungsfähiges Produktionssystem für die Fertigung von Mikroteilen. Werkstattstechnik online, S 5-98, 2008.
- [Ze11] Zelewski, S.: Stilisierte Fakten: ein Beitrag zur gestaltungsorientierten Wirtschaftsinformatik. Wirtschaftsinformatik, 2/11, S. 113-115, 2011.

Leichtgewichtige IT-Modernisierung in der öffentlichen Verwaltung mit der Hilfe von EAM

Lars Santesson¹, Tim Piontek¹, Marvin Zielaskowski¹

Abstract: Diese Arbeit beschreibt einen Ansatz, wie mithilfe von Enterprise Architecture Management (EAM) die IT der öffentlichen Verwaltung modernisiert werden kann. Der Schwerpunkt der Überlegungen liegt darauf, wie mit wenigen Architekten möglichst viel erreicht wird. Hierzu werden einige Methoden bzw. Werkzeuge vorgestellt, die darauf abzielen, schnelle und wirklich hilfreiche Ergebnisse zu erzielen. Außerdem wird darauf eingegangen, wie Unternehmensarchitekten und Lösungsarchitekten dabei zusammenarbeiten sollten.

Keywords: Enterprise Architecture Management, Unternehmensarchitektur, IT-Modernisierung, Lean EAM

1 Die aktuelle Situation

Auf der einen Seite stehen die Herausforderungen für die öffentliche Verwaltung: Neue Anforderungen wegen geänderten oder neuen Gesetzen, wegen des demografischen Wandels und wegen aktueller politischer und gesellschaftlicher Aufgabenstellungen/Probleme (Stichwort Flüchtlingskrise). Außerdem existieren hoher Innovations- und Kostendruck durch den geforderten Bürokratieabbau, von „oben“ vorgegebene Konsolidierungsmaßnahmen (Stichwort IT-Konsolidierung des Bundes [Cio15]), eine geforderte engere Zusammenarbeit über Ländergrenzen hinweg und nicht zuletzt immer größere Anforderungen der Bürger und Unternehmen im „digitalen Zeitalter“ (Stichworte Mobile, Social Media, Open Data).

Auf der anderen Seite steht die derzeitige IT der öffentlichen Verwaltung: über viele Jahrzehnte gewachsene IT-Landschaften mit mangelnder Integrationsfähigkeit für neue Anforderungen, veraltete Technologien, deren Betrieb und Wartung kostspielig ist und von immer weniger Mitarbeitern beherrscht wird, überholte Architekturparadigmen und unzeitgemäße Vorgehensweisen bei der (Weiter-)Entwicklung der Systeme.

Viele Verwaltungen können mit einer solchen IT-Landschaft nicht auf die neuen Chancen und Möglichkeiten reagieren. Sie tendieren viel mehr dazu, in einen reinen „Feuerlöschmodus“ zu schalten, indem lediglich schnellstmöglich die dringendsten Probleme gelöst werden. Dies wiederum ist einer zukunftsfähigen Gestaltung der IT-Landschaft nicht zuträglich.

¹ Capgemini Deutschland GmbH, Potsdamer Platz 5, 10785 Deutschland,
<vorname.nachname@capgemini.com>

Für eine zukunftsfähige Modernisierung dieser bestehenden IT-Landschaften werden gut geschulte IT-Architekten benötigt, die es in der Regel jedoch nur eingeschränkt gibt. Dies ist insbesondere in kleineren und mittelgroßen Verwaltungen der Fall.

Gesucht wird also ein möglichst leichtgewichtiger Ansatz, um mit wenigen Mitarbeitern möglichst viel zu erreichen.

2 Der Modernisierungsansatz

Im Folgenden wird ein Ansatz beschrieben, der auf einer agilen und pragmatischen Top-down-Vorgehensweise beruht und folgende, in diesem Abschnitt beschriebene Methoden / Vorgehensweisen beinhaltet. Es ist mit diesem Ansatz möglich, die organisationsweite Modernisierung gezielt zu steuern. Das Ziel ist es, mit wenigen Architekten eine große Wirkung zu erreichen.

2.1 Entwicklung eines groben Zielbilds

Bevor mit der Modernisierung begonnen wird, ist es für eine Organisation wichtig zu wissen, wo die Reise hingehen soll. Dazu müssen konkrete IT-Ziele erarbeitet werden, die sich natürlich an den Geschäftszielen der Organisation ausrichten sollten. Das Erreichen dieser Ziele wird durch die Definition von IT-Prinzipien sichergestellt. Diese bilden „vorweggenommene Entscheidungen“, an die sich alle IT-Vorhaben (seien es Projekte oder Veränderungen „in der Linie“) halten müssen. Beispielsweise könnte ein Prinzip lauten, dass für unterstützende Prozesse Kaufprodukte generell der Eigenentwicklung vorgezogen und neue Anforderungen rigoros daraufhin überprüft (und gegebenenfalls auch abgeschwächt) werden, wenn der Einsatz eines Kaufproduktes möglich ist. Daraus ergeben sich eine Reihe weiterer Fragestellungen, die gelöst werden müssen. Zum Beispiel, wie die Integration der Kaufprodukte (v. a. die Enterprise Systems) gestaltet werden sollte (siehe dazu [Sul13]). Für die Auswahl der richtigen Enterprise Systems siehe [ABR11].

Es empfiehlt sich, das Zielbild der IT-Landschaft mit einer Grafik zu veranschaulichen. Bei einem unserer Kunden war dies bspw. eine Grafik mit den wesentlichen fachlichen Anwendungen und den technischen Komponenten. Die folgende anonymisierte Abbildung veranschaulicht dies:

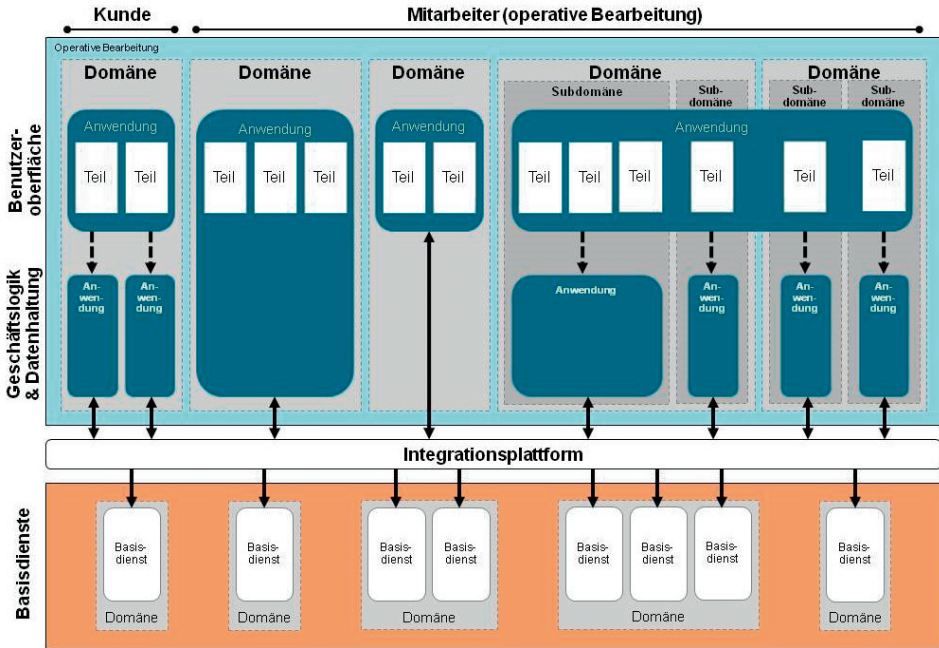


Abb. 1: Anonymisiertes Beispiel eines Zielbildes

Dieses Zielbild dient einem gemeinsamen Verständnis als Leuchtturm am Horizont. Als nächstes sollte die IT-Landschaft in Richtung dieses Leuchtturms gesteuert werden. Da sie in dieser Seefahrt-Metapher und im Hinblick auf die Ausgangslage viel mehr einen Tanker als ein Schnellboot darstellt, sind besondere Maßnahmen zu ergreifen. Auch darf unsere sehr kleine Besatzung nicht unberücksichtigt bleiben.

2.2 Gesteuerte Evolution der IT-Vorhaben in Richtung Ziel

Genauso wie auch Städte nicht abgerissen und neu aufgebaut werden, passiert das nicht für Anwendungslandschaften. Zwar werden einzelne Gebäude abgerissen und neu aufgebaut oder den aktuellen Umständen angepasst, jedoch muss dafür gesorgt werden, dass diese grundsätzlich in die Stadt passen, sprich architekturkonform sind. Eine Evolution gibt es immer. Diese Evolution wird nun aber steuerbar und an dem zuvor definierten Zielbild ausgerichtet.

Für diesen Zweck gibt es eine Vielzahl von Architekturinstrumenten, wie z. B. die Architekturprinzipien, Referenzarchitekturen, einen Sollbebauungsplan, ein Technologie-Radar und Standards für Anwendungen, Technologien und Infrastruktur. Bei der Modernisierung einer einzelnen Anwendung wird mithilfe dieser Werkzeuge sichergestellt, dass nicht nur die Nutzeranforderungen und nichtfunktionalen

Anforderungen eingeführt, sondern dass gleichzeitig ein Teil zur Erreichung des Ziels beigetragen wird. Manche IT-Vorhaben wirken dabei mehr in Richtung fachlichem Nutzen (z. B. Umsetzung neuer Anforderungen), andere mehr in Richtung IT-Nutzen (z. B. Agilität erhöhen). Ein Beispiel, wie vom Ist zum Soll transformiert werden kann, liefert [Bro14]. So oder so bewegt sich die IT-Landschaft somit Stück für Stück in Richtung des Ziels.

2.3 Heatmap zur Identifikation der kritischen Bereiche

Bei der Modernisierung der Anwendungslandschaft können nicht alle Themen mit gleicher Priorität behandelt und gleichzeitig gelöst werden, sodass man sich auf die wichtigsten Themen konzentrieren muss. Zunächst müssen diese Themen identifiziert und in einer Art Heatmap festgehalten werden. Heatmap-Themen sind Bereiche, in denen eine Modernisierung sehr dringend ist. Deren Identifikation kann nur mit den relevanten Entscheidern aus den verschiedenen Bereichen vorgenommen werden. Hierbei ist darauf zu achten, dass Themen mit Architekturelevanz identifiziert werden, also Themen, bei denen das Architekturmanagement einen Beitrag leisten kann. Bei einem unserer Kunden haben wir so die zehn wesentlichen Heatmap-Themen identifiziert und jeweils auf einer Powerpoint-Folie beschrieben. Dabei haben wir stets die gleichen Aspekte beleuchtet (Ausgangslage, Ziel, Auswirkungen wenn nicht / unvollständig gelöst, Aktueller Status, Auftrag / Zieltermin, Erforderliche Rahmenbedingungen), um eine Vergleichbarkeit zu gewährleisten und die Priorisierung der Themen zu vereinfachen. Als Unterstützung diente uns dabei ein zuvor grob erarbeitetes fachliches Domänenmodell, um die Berücksichtigung aller Bereiche zu gewährleisten. Das Dokument mit den Heatmap-Themen wird regelmäßig aktualisiert. Veränderte Anforderungen, der aktuelle Status der Themen sowie neue und ausgeschiedene Heatmap-Themen sind somit stets ersichtlich.

2.4 Keine Architekturarbeit auf Vorrat

Wie eingangs erwähnt, sollte das Zielbild zunächst nur grob entwickelt und als weit entfernter Leuchtturm verstanden werden. Das vorhandene grobe Zielbild wird nun genau an den Stellen konkretisiert, wo auch die wichtigsten Bereiche (Heatmap-Themen) identifiziert wurden. Die Architekturarbeit wird also nur dort gemacht, wo sie auch nützlich ist. Somit wird gewährleistet, dass sich die wenigen Architekten auf das Wesentliche konzentrieren und sie keine Architekturarbeit für Themen durchführen, die nicht auf der Agenda stehen. Ein Sollbebauungsplan muss bspw. zwar alle fachlichen Domänen enthalten, jedoch nur diese spezifizieren, die als nächstes bearbeitet werden. Ein anderes Beispiel ist die Erarbeitung von Referenzarchitekturen, die bei der Lösungsfindung für zukünftige Handlungsbedarfe unterstützen. Hier können benötigte/hilfreiche Referenzarchitekturen identifiziert werden (z. B. zunächst als leere Container – siehe weiter unten), die lediglich für die Heatmap-Themen näher

ausdetailliert werden. Der Unternehmensarchitekt und der Lösungsarchitekt arbeiten dabei Hand in Hand.

2.5 Der Unternehmensarchitekt steuert, der Lösungsarchitekt gestaltet

Die Aufgabe des Unternehmensarchitekten ist es, unternehmensweit zu agieren. Gerade wenn nur wenige Unternehmensarchitekten verfügbar sind, können sie sich nicht mit den Details beschäftigen, sondern sollten lediglich Akzente setzen und Hypothesen aufstellen (z. B. „In unserer stark punktuell verknüpften Anwendungslandschaft sollten wir eine zentrale Integrationsplattform einführen.“). Wenn es im Rahmen der Bearbeitung eines Heatmap-Themas nötig ist, konkretisiert der Lösungsarchitekt diese Akzente bzw. überprüft die Hypothesen. Auf gleiche Art und Weise kann z. B. die Zielarchitektur für die relevanten Bereiche konkretisiert oder die identifizierten, aber noch leeren Referenzarchitektur-Container detailliert werden.

2.6 Referenzarchitekturen für ähnliche Aufgaben und Herausforderungen

Referenzarchitekturen bieten bewährte Lösungen für wiederkehrende Problemstellungen. Sie vereinfachen somit die Lösungsfindung und wirken gleichzeitig unternehmensweit standardisierend. Idealerweise werden für alle Domänen der Anwendungslandschaft Referenzarchitekturen erarbeitet. Es ist jedoch keine 1:1 Beziehung notwendig, sondern es können auch domänenübergreifende Referenzarchitekturen existieren, wie z. B. eine „Stammdatensystem-Referenzarchitektur“, die als Grundlage in vielen Domänen gilt. Wichtig ist es, dass zunächst alle Referenzarchitekturen nur als leere Container erstellt werden. So haben wir mit den Unternehmensarchitekten einer unserer Kunden zunächst die relevanten Referenzarchitekturen identifiziert. Für alle relevanten Referenzarchitekturen haben die Unternehmensarchitekten dann zunächst erste grundlegende Prinzipien und weitere Klärungspunkte formuliert. Als nächstes sollen, entsprechend des Grundsatzes „Keine Architekturarbeit auf Vorrat“, genau jene Referenzarchitekturen gemeinsam mit den Lösungsarchitekturen erarbeitet werden, die im Rahmen der Heatmap-Themen eine Rolle spielen. Durch diese Zusammenarbeit konnten wir gewährleisten, dass die Referenzarchitekturen wirklich hilfreich sind und sie andererseits eine unternehmensweite Sichtweise einnehmen (z. B. Abgleich mit anderen unternehmensweiten Vorgaben). Denkbar ist außerdem, dass man in einer weiteren Iteration eine konkrete technische Plattform identifiziert und aufbaut, die zur Erstellung einer Lösung nach dieser Referenzarchitektur hergenommen wird. Diese stufenweise Erarbeitung der Referenzarchitekturen lässt sich entsprechend eines Reifegradmodells auffassen und veranschaulichen (siehe folgende Abbildung).

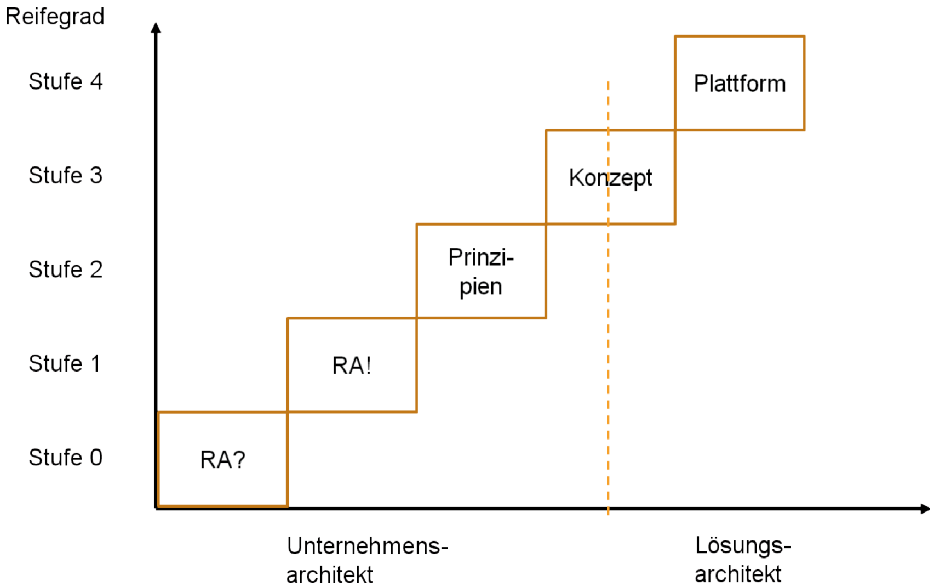


Abb. 2: Reifegradmodell für Referenzarchitekturen

Doch nicht nur Vorgaben und Hilfestellungen sind nötig, sondern deren Nutzung und Einhaltung muss ebenfalls überprüft werden (Stichwort IT-Governance).

2.7 Gewährleistung architekturkonformer IT-Vorhaben über ihren Lebenszyklus

Ein entscheidender Aspekt bei der Modernisierung in Richtung des Zielbildes ist die einheitliche Kontrolle und Bewertung von IT-Vorhaben im Hinblick auf architekturelevante Gesichtspunkte. Am Anfang eines IT-Vorhabens und bei Änderungen im Laufe der Umsetzung muss festgestellt werden, ob ein IT-Vorhaben konform zu den Architekturvorgaben ist – also einen Beitrag zur Erreichung des Ziels leistet.

Um auch dies möglichst leichtgewichtig zu halten, sollte durch die IT-Vorhaben eine verdichtete Darstellung ihrer architektonisch relevanten Aspekte erarbeitet werden. Bei einem unserer Kunden wird das z. B. durch ein standardisiertes Template-Dokument erreicht, indem eine fachliche Verortung im Domänenmodell stattfindet, die wesentlichen umzusetzenden funktionalen und nicht-funktionalen Anforderungen kurz beschrieben werden sowie standardisierte Viewpoints aufgeführt werden (für eine Definition von Viewpoint siehe [Tog11]), die die Software- und Systemarchitektur prägnant aufzeigen (z. B. Kontextdiagramme, Schnittstellendiagramme, Komponentendiagramme, usw.). Auf dieser Grundlage kann schnell identifiziert werden,

ob das IT-Vorhaben noch „on track“ ist. Diese Darstellung wird dann regelmäßig einem Organ (z. B. Architektur-Gremium) vorgestellt, welches genügend Einfluss hat, um von dem IT-Vorhaben eine veränderte Architektur einzufordern.

3 Zusammenfassung

In diesem Beitrag haben wir gezeigt, wie mit möglichst leichtgewichtigen EAM-Mitteln die IT der öffentlichen Verwaltung modernisiert werden kann. Gerade in kleinen und mittelgroßen Verwaltungen besteht oftmals ein Mangel an qualifizierten Architekten bei einer gleichzeitig oftmals komplexen und schwer zu steuernden IT-Landschaft. Wir haben gezeigt, wie auch mit wenigen Architekten, und durch die Zusammenarbeit von Unternehmens- und Lösungsarchitekten, ein großer praktischer Nutzen erzielt wird: Im ersten Schritt werden dazu IT-Ziele und IT-Architektur-Prinzipien erarbeitet, sowie ein organisationsweites, grobes Zielbild aufgestellt. Danach werden die wichtigsten Themen mit Architekturrelevanz identifiziert und priorisiert. Unternehmensarchitekten und Lösungsarchitekten detaillieren dann gemeinsam die grobe Zielarchitektur in genau den Bereichen, wo Heatmap-Themen identifiziert wurden. Als wesentliches Steuerungs- und Standardisierungswerkzeug können dabei Referenzarchitekturen hergenommen werden, wobei auch hier darauf zu achten ist, dass lediglich jene detailliert erarbeitet werden, die für die Heatmap-Themen relevant sind. Abschließend ist im Rahmen der IT-Governance die Architekturkonformität der IT-Vorhaben über ihren gesamten Lebenszyklus zu überwachen. Auch hier haben wir eine möglichst leichtgewichtige Vorgehensweise.

Literaturverzeichnis

- [Cio15] cio.bund.de (Der Bundesbeauftragte der Bundesregierung für Informationstechnik): Grobkonzept zur IT-Konsolidierung Bund, https://www.cio.bund.de/SharedDocs/Publikationen/DE/Innovative-Vorhaben/it_konsolidierung_bund_grobkonzept.pdf?__blob=publicationFile, Online-Quelle, letzter Zugriff am 07.06.2016
- [Sul13] Sultanow, E; Brockmann, C.; Pratt, R.M.E; Andresen, K.: Integrate Enterprise Systems to our Hyperconnected World: Anything, Anywhere, Anytime through architectural design, Proceedings of the 19th AMCIS, Chicago, Illinois, August 15-17, 2013.
- [ABR11] Andresen, K.; Brockmann, C.; Roztocki, N.: "Business Models for Enterprise System Providers: Towards the Solution Based Procedure", AMCIS 2011 Proceedings - All Submissions. Paper 236.
- [Bro14] Brockmann, C.: An approach to design the business model of an ERP vendor, Schriften der Forschungsvereinigung Software; 2014.
- [Tog11] The Open Group: The Open Group Architecture Framework 9.1, <http://pubs.opengroup.org/architecture/togaf9-doc/arch/>, letzter Zugriff am 08.06.2016

Design und Nutzung einer industriespezifischen Referenzarchitektur für die Telekommunikationsindustrie

Christian Czarnecki¹

Abstract: Zur Unterstützung des Transformationsbedarfs von Telekommunikationsunternehmen sind die Referenzmodelle des TM Forums in der Praxis weltweit anerkannt. Dabei findet jedoch meist eine losgelöste Nutzung für spezifische Einzelthemen statt. Daher führt dieser Artikel die bestehenden Inhalte in einer industriespezifischen, übergreifenden Referenzarchitektur zusammen. Der Fokus liegt auf den Ebenen Aufbauorganisation, Prozesse, Applikationen und Daten. Darüber hinaus werden inhaltliche Architekturdomänen zur Strukturierung angeboten. Die Referenzarchitektur ist hierarchisch aufgebaut und wird hier beispielhaft für ausgewählte, aggregierte Inhalte beschrieben. Als erste Evaluation wird die Anwendung der Referenzarchitektur in drei Praxisprojekten erläutert.

Keywords: Referenzmodellierung, Unternehmensarchitektur, Unternehmenstransformation, TM Forum, enhanced Telecom Operations Map (eTOM).

1 Einleitung

Telekommunikationsunternehmen (TK-Unternehmen) sind einem enormen Veränderungsdruck ausgesetzt [GS03, Pi06, Pi14]. Dabei sind die Fragmentierung der Wertschöpfungskette [PH11], zunehmender branchenübergreifender Wettbewerb [WZ11] sowie schnellere Entwicklungszyklen bei steigender technischer Komplexität [Br08] zu berücksichtigen. Aus wirtschaftlicher Sicht sind Kundenorientierung und Produktinnovation bei gleichzeitig sinkenden Umsätzen und steigendem Effizienzdruck zu realisieren [Pi14]. Intern führt dies bei TK-Unternehmen zu einer Vielzahl an Transformationsprojekten [CWS12].

Unternehmensarchitekturen [UM07, Ma11] und Referenzmodelle [Th06, vBr15] können zur methodischen und inhaltlichen Gestaltung dieser Transformationen genutzt werden. In der Telekommunikationsindustrie (TK-Industrie) sind die industriespezifischen Referenzmodelle des TM Forums als Standard verbreitet [Ke03, Cz13]. Das TM Forum ist ein weltweit agierendes Industriegremium, das mit über 900 Mitgliedsunternehmen eine Plattform für die gesamte Wertschöpfung der TK-Industrie bietet. Neben Referenzmodellen werden gemeinsame Forschungs- und Entwicklungsaktivitäten, Zertifizierungen, Schulungen und Konferenzen angeboten. Wie eine eigene Untersuchung von 184 Transformationsprojekten zeigt, ist eine umfangreiche Verwendung der Referenzmodelle des TM Forums gegeben [CWS12]. Dabei findet

¹ Hochschule für Telekommunikation Leipzig, Gustav-Freytag-Straße 43-45, 04277 Leipzig, czarnecki@hft-leipzig.de

jedoch überwiegend eine Nutzung einzelner Referenzmodellelemente statt, ohne dass die Abhängigkeiten und Schnittstellen aus einer übergreifenden Sicht abgebildet werden. Aus diesem Grund wird in diesem Artikel in Abschnitt 3 eine detaillierte Referenzarchitektur auf Basis der bestehenden Referenzmodelle vorgeschlagen und deren Anwendung in Abschnitt 4 erläutert. Abschnitt 2 gibt einen kurzen Überblick über den Stand der Forschung. Abschließend folgen ein Fazit und Ausblick in Abschnitt 5.

2 Stand der Forschung

Bei der industriespezifischen Entwicklung und Einführung einer Unternehmensarchitektur ist zwischen (1) der *generellen Struktur* und (2) den *industriespezifischen Inhalten* zu unterscheiden. Rahmenwerke für Unternehmensarchitekturen, sogenannte *Enterprise Architecture Frameworks* (EAF), können als Referenz genutzt werden [UM07, Ma11]. Ziel ist die fundamentale Organisation eines Unternehmens durch die Strukturierung der relevanten Komponenten und deren Abhängigkeiten [WF07]. Darüber hinaus werden Methoden zur Gestaltung, Bewertung, Einführung und kontinuierlichen Verbesserung vorgeschlagen [Ah12, Ha12]. Es existieren eine Vielzahl unterschiedlicher EAF [UM07], siehe Schekckermann [Sc04] und Matthes [Ma11] für eine Übersicht. Neben den strukturellen und methodischen Empfehlungen eines EAF sind bei der konkreten Gestaltung einer Unternehmensarchitektur inhaltliche Empfehlungen notwendig. *Referenzmodelle* liefern Lösungsvorschläge, die bei einer konkreten Anwendungsmodellierung wiederverwendet werden können [Th06, vBr15]. Dabei werden generalisierbare Empfehlungen in Bezug auf eine Problemdomäne (z. B. Industrie, Funktion) in einem Modell als Referenz bereitgestellt [FL04], siehe z. B. vom Brocke [vBr15] sowie Fettke und Loos [FL07] für eine ausführliche Diskussion.

Aus industriespezifischer Sicht stellt das TM Forum die folgenden drei Referenzmodelle als Empfehlungen für TK-Unternehmen bereit [CWS12]. (1) *Business Process Framework* (eTOM) ist eine hierarchische Strukturierung von Geschäftsprozessen von einem Ordnungsrahmen bis zu detaillierten Prozessdefinitionen [Ke03]. (2) *Application Framework* (TAM) ist eine Strukturierung von Funktionen nach logischen Applikationsgruppen. Diese werden in einem hierarchischen Funktionsbaum bereitgestellt [TM15a]. (3) *Information Framework* (SID) ist ein logisches Datenmodell, das aggregierte Datenobjekte definiert und diese in detaillierten Entity-Relationship-Modellen spezifiziert [TM15b]. Eine umfangreiche Nutzung dieser Referenzmodelle in Praxisprojekten ist dokumentiert [CWS12]. Mögliche Anwendungsszenarien sind u. a. die Entwicklung von Operations-Support-Systemen (OSS) [Mi04], das Monitoring von Next Generation Networks (NGN) [Ra11], die Gestaltung eines integrierten Netzwerkmanagements [YF07], die Effizienzsteigerung der Bereitstellung von Telekommunikationsdiensten [Br08] und die Automatisierung von Netzwerkfunktionalitäten [St13]. Übergreifende Transformationsprojekte benötigen jedoch eine Empfehlung relevanter Inhalte aus Gesamtunternehmenssicht [CWS12]. Daher wird eine Zusammenführung der einzelnen Referenzmodelle des TM Forums in eine

Unternehmensarchitektur vorgeschlagen [Cz13]. Diese bestehenden Inhalte wurden in einer detaillierten Referenzarchitektur weiterentwickelt und werden in diesem Artikel auszugsweise dargestellt (siehe Abschnitt 3).

3 Design einer Referenzarchitektur für die TK-Industrie

Eine Referenzarchitektur für die TK-Industrie sollte zwingend die als de facto Standard anerkannten Referenzmodelle des TM Forums berücksichtigen (siehe Abschnitt 2). Als grobe Struktur für die Referenzarchitektur werden die Ebenen (1) Strategie, (2) Aufbauorganisation, (3) Prozesse, (4) Applikationen, (5) Daten, (6) Netzinfrastruktur vorgeschlagen, die in Anlehnung an Winter und Fischer [WF07] als Gemeinsamkeit gängiger EAF anzusehen sind. Somit ist eine Nutzung der Referenzarchitektur mit unterschiedlichen bestehenden methodischen Unternehmensarchitekturansätzen möglich. Aus strategischer Sicht sind die Veränderungen der TK-Industrie zu berücksichtigen (siehe Abschnitt 1). Aus zeitlicher und finanzieller Perspektive stellt die Netzinfrastruktur einen begrenzenden Faktor dar [Mi04, Br08].

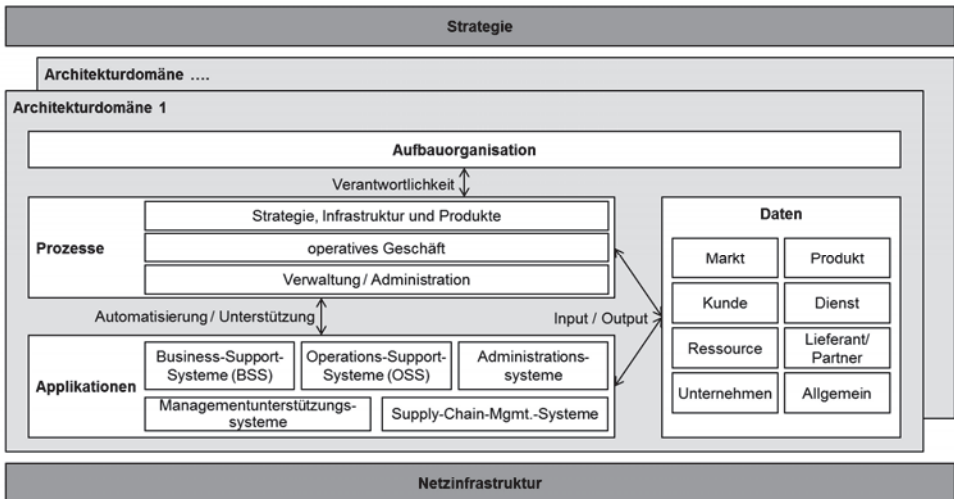


Abb. 1: Referenzarchitektur (in Anlehnung an [Cz13])

Kern der Referenzarchitektur sind Prozesse, Applikationen und Daten, die wie folgt strukturiert werden (siehe Abb. 1). Die *Prozesse* werden gemäß eTOM [Ke03] in die Schaffung der kommerziellen und technischen Voraussetzungen zum Verkauf von Telekommunikationsprodukten (*Strategie, Infrastruktur und Produkte*) sowie die Bearbeitung von Kundenanfragen und -aufträgen im Tagesgeschäft (*operatives Geschäft*) unterteilt. Diese beiden Prozessgruppen sind spezifisch für die TK-Industrie und werden durch generelle Prozesse zur *Verwaltung/Administration* ergänzt. Die

Applikationen werden in fünf Applikationsgruppen unterteilt. Als spezifisch für die TK-Industrie gilt die Unterscheidung in *Business-Support-Systeme* (BSS) und *Operations-Support-Systeme* (OSS). BSS unterstützen die kommerzielle Bearbeitung von Kunden und Produkten, während OSS die technische Umsetzung von Diensten und Ressourcen ermöglichen. Darüber hinaus werden *Managementunterstützungs-*, *Supply-Chain-Management-*, und *Administrationssysteme* unterschieden. Die *Daten* werden gemäß SID [TM15b] in die Entitätsgruppen *Markt*, *Kunde*, *Produkt*, *Dienst*, *Ressource*, *Lieferant/Partner*, *Unternehmen* und *Allgemein* strukturiert. Zur Beherrschung der technischen Komplexität bilden Produkte die kommerzielle Sicht ab, welche durch Dienste und Ressourcen auf technischer Ebene realisiert werden [Br08]. Darüber hinaus stellt die *Aufbauorganisation* einen wichtigen Bestandteil bei der organisatorischen Umsetzung der Referenzarchitektur dar. Diese ist in den Referenzmodellen des TM Forums nicht enthalten. Generell ist davon auszugehen, dass die konkrete Aufbauorganisation sich je nach Anwendungsszenario stark unterscheidet. Jedoch bietet die Referenzarchitektur Strukturierungskriterien sowie beispielhafte Detaillierungen als Referenz an.

Aufgrund der Komplexität wird als weiteres Strukturierungsmerkmal eine Unterteilung in Architekturdomänen vorgenommen und damit eine Entkopplung einzelner Architekturbereiche ermöglicht [AW08]. Dabei wird eine prozessgetriebene Aufteilung eines TK-Unternehmens in (1) Kunde-zu-Kunde-Domäne, (2) Technologie-Domäne, (3) Produkt-Domäne, (4) Kunden-Domäne und (5) Support-Domäne vorgeschlagen. Die Kunde-zu-Kunde-Domäne enthält alle Architekturelemente, die zur Bearbeitung von Geschäftsfällen notwendig sind, die direkt durch den Kunden ausgelöst werden. Dabei wird eine konsequente Ende-zu-Ende-Perspektive eingenommen, d. h. jeder Geschäftsfall beginnt und endet beim Kunden. Die Technologie-Domäne behandelt Bereitstellung, Betrieb und Wartung von Diensten und Ressourcen. Diese Entkopplung zwischen kommerzieller und technischer Sicht ist ein wichtiges Gestaltungsmerkmal, siehe Czarnecki [Cz13] und Czarnecki et al. [CWS13] für eine detaillierte Beschreibung der Architekturdomänen.

Eine Detaillierung der Referenzarchitektur erfolgt anhand der obigen Domänen. Abb. 2 fasst ausschnittsweise die Detaillierung für die Kunde-zu-Kunde-Domäne zusammen. Für die Aufbauorganisation werden *Kanal* (z. B. Call-Center, Internet), *Kundentyp* (z. B. Privatkunde, Wiederverkäufer), *Produkttyp* (z. B. Mobilfunk, Festnetz) und *geographische Struktur* (zentral und regional) als Gestaltungskriterien vorgeschlagen. Beispielhafte, detaillierte Organigramme verdeutlichen die möglichen Optionen der aufbauorganisatorischen Gestaltung anhand dieser Kriterien. Die Prozessebene wird durch sieben Referenzprozessabläufe definiert, für die entsprechende Prozessmodelle verfügbar sind [CWS13]. Auf der Applikationsebene wird die Kunde-zu-Kunde-Domäne ausschließlich durch BSS unterstützt. Dabei werden die relevanten Funktionen über das Referenzmodell TAM definiert, wobei auf der höchsten Aggregationsebene (TAM Level 0) die Funktionen *Kundenmanagement* sowie *Markt-/Vertriebsmanagement* zugeordnet sind [TM15a]. Die Datenebene wird anhand des Referenzmodells SID spezifiziert und ist in *Produkt* und *Kunde* (SID Level 0) aufgeteilt [TM15b]. Die hier beispielhaft

dargestellten Inhalte werden in der Referenzarchitektur weiter detailliert. So wird z. B. der Referenzprozessablauf *Auftrag-bis-Zahlung* durch eine Modellierung der Aktivitäten für Kundenkontaktmanagement, Auftragsabwicklung, Bereitstellung und Abrechnung detailliert, und das Datenelement *Produkt* ist gemäß SID in u. a. Produktspezifikation und Produktangebot verbunden mit detaillierten Datenmodellen unterteilt. Somit bietet die vorgeschlagene Referenzarchitektur eine Auswahl und Strukturierung der relevanten Referenzmodellelemente an und stellt die übergreifenden Zusammenhänge dar. Durch die hierarchische Strukturierung der genutzten Referenzmodelle wird eine Umsetzung auf operativer Ebene unterstützt.

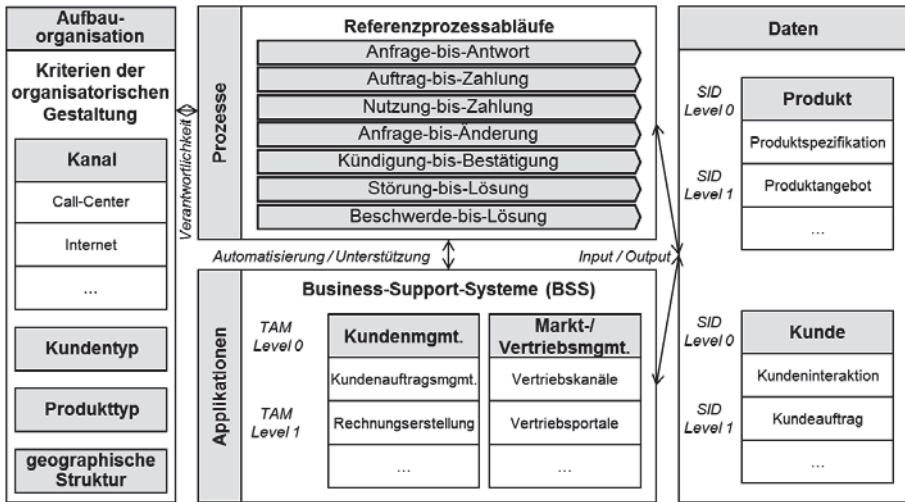


Abb. 2: Detaillierung der Referenzarchitektur für die Kunde-zu-Kunde-Domäne (Ausschnitt)

4 Nutzung der Referenzarchitektur in der TK-Industrie

Die hier dargestellte Referenzarchitektur wurde in unterschiedlichen Projekten in der TK-Industrie verwendet. Tab. 1 enthält eine Zusammenfassung von drei ausgewählten Projektbeispielen bei denen die Referenzarchitektur genutzt wurde. Gegenstand der Projektbeispiele sind unterschiedliche TK-Unternehmen aus Europa, Asien und Afrika. Generell wird in der gestaltungsorientierten Forschung die Fallstudie als gängige Evaluationsmethode vorgeschlagen [He04], wobei gerade in der Referenzmodellierung die Wiederverwendung in einem praktischen Kontext als wichtig angesehen wird [Th06]. Insofern können die hier zusammengefasst dargestellten Projektbeispiele als eine erste Bestätigung der vorgeschlagenen Referenzarchitektur verstanden werden.

Projekt	Umfang	Nutzung der Referenzarchitektur
Einführung eines OSS (Asien)	Konzeptionelles Design zur Einführung eines OSS als Teil eines Transformationsprogramms	<ul style="list-style-type: none"> - Referenzprozessabläufe und aufbauorganisatorische Zuordnung als Grundlage des Fachkonzepts - Zuordnung zu Applikationsfunktionen zur Anforderungsdefinition bei der Softwareauswahl
Einführung eines Customer-Relationship-Management-Systems (CRM-System) (Europa)	Auswahl und Umsetzung eines standardisierten CRM-Systems auf Basis von Standardsoftware	<ul style="list-style-type: none"> - Referenzprozessabläufe der Kunde-zu-Kunde Domäne - Zuordnung auf Applikationen zur Anforderungsdetailierung - Übergreifendes Datenmodell als integrierendes Element
Einführung einer Netzbetriebszentrale (Network Operations Center, NOC) (Afrika)	Entwicklung von NOC-Prozessen sowie aufbauorganisatorische und technische Einführung	<ul style="list-style-type: none"> - Referenzprozessabläufe als Grundlage für die Prozessentwicklung - Aufbauorganisatorische und technische Zuordnung auf Basis der Referenzarchitektur

Tab. 1: Projektbeispiele zur Nutzung der Referenzarchitektur

5 Fazit und Ausblick

Die Veränderungen der TK-Industrie haben zu einer Vielzahl an Transformationsprojekten bei TK-Unternehmen geführt. In der Praxis dienen dabei die Referenzmodelle des TM Forums als Empfehlung. Jedoch findet deren Nutzung häufig in losgelöster Form für einzelne Themen statt, so dass die Zusammenführung in einer Referenzarchitektur vorgeschlagen wird. Die Weiterentwicklung dieser Referenzarchitektur wird in diesem Artikel beschrieben. In der aggregierten Darstellung liegt der Schwerpunkt auf den Ebenen Aufbauorganisation, Prozesse, Applikationen und Daten. Darüber hinaus wird durch Architekturdomänen ein inhaltliches, industriespezifisches Strukturierungselement angeboten. Anhand der Kunde-zu-Kunde-Domäne wird die Detaillierung der Referenzarchitektur erläutert und die Integration der einzelnen Architekturelemente gezeigt. Aufgrund der hierarchischen Struktur der Referenzarchitektur ist eine Detaillierung bis auf operative Ebene sichergestellt. Damit kann sie als Basis für die Planung, Gestaltung und Umsetzung von übergreifenden Transformationsprojekten dienen. Die Anwendung der Referenzarchitektur wird beispielhaft und in zusammengefasster Form anhand von drei Projekten erläutert. Eine kontinuierliche Weiterentwicklung u. a. auch in Kooperation mit dem TM Forum ist geplant. Dabei können die Berücksichtigung von bestehenden Anwendungsmodellen im Sinne einer induktiven Referenzmodellierung, die Anreicherung der inhaltlich ausgerichteten Referenzarchitektur mit generellen Methoden des Enterprise Architecture Managements (EAM) sowie die kontinuierliche Anpassung an die aktuellen technischen Innovationen der TK-Industrie als Startpunkt für zukünftige Forschung dienen.

Literaturverzeichnis

- [Ah12] Ahlemann, F.: Strategic enterprise architecture management: challenges, best practices, and future developments. Springer, Berlin ; New York, 2012.
- [AW08] Aier, S., Winter, R.: Virtuelle Entkopplung von fachlichen und IT-Strukturen für das IT/Business Alignment – Grundlagen, Architekturgestaltung und Umsetzung am Beispiel der Domänenbildung. *Wirtschaftsinformatik*, 51:175–191, 2008.
- [Br08] Bruce, G., et al.: Streamlining the telco production line. *Journal of Telecommunications Management* 1:15 – 32, 2008.
- [Cz13] Czarnecki, C.: Entwicklung einer referenzmodellbasierten Unternehmensarchitektur für die Telekommunikationsindustrie. Logos-Verl, Berlin, 2013.
- [CWS12] Czarnecki, C., Winkelmann, A., Spiliopoulou, M.: Transformation in Telecommunication – Analyse und Clustering von Real-life Projekten. In: Mattfeld, D. C., Robra-Bissantz, S. (eds) *Multi-Konferenz Wirtschaftsinformatik 2012*. GITO Verlag, Braunschweig, pp 985–998, 2012.
- [CWS13] Czarnecki, C., Winkelmann, A., Spiliopoulou, M.: Referenzprozessabläufe für Telekommunikationsunternehmen: Eine Erweiterung des eTOM-Modells. *Wirtschaftsinformatik*, 55:83–97, 2013.
- [FL04] Fettke, P., Loos, P.: Referenzmodellierungsforschung. *Wirtschaftsinformatik* 46:331–340, 2004.
- [FL07] Fettke, P., Loos, P.: Reference modeling for business systems analysis. Idea Group Pub, Hershey, PA, 2007.
- [GS03] Grover, V., Saeed, K.: The telecommunication industry revisited. *Communications of the ACM* 46:119–125, 2003.
- [Ha12] Hanschke, I. *Enterprise Architecture Management - einfach und effektiv ein praktischer Leitfaden für die Einführung von EAM*. Hanser, München, 2012.
- [He04] Hevner, A. R., et al.: Design Science in Information Systems Research. *MIS Quarterly* 28:75–105, 2004.
- [Ke03] Kelly, M. B.: The TeleManagement Forum’s Enhanced Telecom Operations Map (eTOM). *Journal of Network and Systems Management* 11:109–119, 2003.
- [Ma11] Matthes, D.: *Enterprise Architecture Frameworks Kompendium*. Springer, Heidelberg, 2011.
- [Mi04] Misra, K.: *OSS for Telecom Networks: An Introduction to Network Management*. Springer, London, 2004.
- [Pi06] Picot, A.: *The Future of Telecommunications Industries*. Springer-Verlag, Berlin/Heidelberg, 2006.
- [PI14] Plunkett, J. W.: *Plunkett’s telecommunications industry almanac 2015: the only comprehensive guide to the telecommunications industry*, 2014.
- [PH11] Pousttchi, K., Hufenbach, Y.: *Wertschöpfung im Mobilfunkmarkt*.

Wirtschaftsinformatik, 53:287–300, 2011.

- [Ra11] Raouyane, B., et al.: eTOM Business Processes Conception in NGN Monitoring. In: Lin, S., Huang, X. (eds) *Advanced Research on Computer Education, Simulation and Modeling*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 133–143, 2011.
- [Sc04] Schekkerman, J.: How to survive in the jungle of enterprise architecture frameworks : creating or choosing an enterprise architecture framework. Trafford, 2004.
- [St13] Stamatelatos, M., et al.: Information Model for Managing Autonomic Functions in Future Networks. In: Pesch, D., et al. (eds) *Mobile Networks and Management*. Springer International Publishing, Cham, pp 259–272, 2013.
- [Th06] Thomas, O.: Understanding the Term Reference Model in Information Systems Research: History, Literature Analysis and Explanation. In: Bussler, C., Haller, A. (eds) *Business Process Management Workshops*. Springer Berlin / Heidelberg, pp 484–496, 2006.
- [TM15a] TM Forum: Application Framework: The Digital Services Systems Landscape (GB929 Addendum D), Version 14.5.1, 2015.
- [TM15b] TM Forum: Information Framework (SID): Concepts and Principles (GB922), Version 15.0.0, 2015.
- [UM07] Urbaczewski, L., Mrdalj, S.: A Comparison of Enterprise Architecture Frameworks. *Issues in Information Systems* 7:18–23, 2007.
- [vBr15] vom Brocke, J.: *Referenzmodellierung: Gestaltung und Verteilung von Konstruktionsprozessen*. Logos, Berlin, 2015.
- [WF07] Winter, R., Fischer, R.: Essential Layers, Artifacts, and Dependencies of Enterprise Architecture. *Journal of Enterprise Architecture* 2:7–18, 2007.
- [WZ11] Wulf, J., Zarnekow, R.: Branchenübergreifender Wettbewerb in der Telekommunikation. *Wirtschaftsinformatik* 53:277–286, 2011.
- [YF07] Yari, A. R., Fesharaki, S. H. H.: A Framework for an Integrated Network Management System Based on Enhanced Telecom Operation Map (eTOM). In: Ata, S., Hong, C. S. (eds) *Managing Next Generation Networks and Services*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 587–590, 2007.

Entwicklung eines Architekturkonzepts zum flexiblen Einsatz von Analytics

Benedict Bender¹, Marcus Grum²

Abstract: Die optimale Dimensionierung von IT-Hardware stellt Entscheider aufgrund der stetigen Weiterentwicklung zunehmend vor Herausforderungen. Dies gilt im Speziellen auch für Analytics-Infrastrukturen, die zunehmend auch neue Software zur Analyse von Daten einsetzen, welche in den Ressourcenanforderungen stark variieren. Damit eine flexible und gleichzeitig effiziente Gestaltung von Analytics-Infrastrukturen erreicht werden kann, wird ein dynamisch arbeitendes Architekturkonzept vorgeschlagen, das Aufgaben auf Basis einer systemspezifischen Entscheidungsmaxime mit Hilfe einer Eskalationsmatrix verteilt und hierfür Aufgabencharakteristiken sowie verfügbare Hardwareausstattungen entsprechend ihrer Auslastung berücksichtigt.

Keywords: Analytics, Architekturkonzept, cyber-physische Systeme, Cloud, Internet of Things

1 Einleitung

Die IT-Infrastruktur von Unternehmen wird zunehmend komplexer. Dies umfasst sowohl die Hardware- als auch die Softwarekomponenten. Im Bereich der Hardware zeichnet sich dies u.a. durch eine zunehmende Anzahl vernetzter Geräte aus, welche Teil der Unternehmensinfrastruktur, und durch eine verstärkte Heterogenität der Geräte [XHL14]. Im Softwarebereich der IT-Infrastruktur werden für die Bewerksstellung der Aufgaben für die jeweiligen Aufgabenbereiche in den Unternehmen zunehmend spezialisierte Systeme eingesetzt. Das Enterprise Architecture Management (EAM) zielt auf die Ausrichtung der betrieblichen IT-Infrastruktur bzgl. der Bedürfnisse der Geschäftsaktivitäten ab. Dabei werden beide vorstehende Perspektiven (Hard- und Software) betrachtet.

Die Bedeutung von Auswertungen der im Rahmen der Betriebstätigkeit anfallenden Daten gewinnt zunehmend an Bedeutung. Beispielsweise werden im Produktionsumfeld mit dem Einsatz verschiedener Sensoren Informationen über den physischen Produktionsprozess erhoben, welche für die Steuerung des Prozesses genutzt werden können [VF14]. Mit Hilfe von vielfältigen Analytics-Ansätzen können auf Basis derartiger Sensorwerte Entscheidungen über das weitere Verfahren innerhalb des Produktionsprozesses systematisch untersucht, optimale Maßnahmen abgeleitet und

¹ Universität Potsdam, Lehrstuhl für Wirtschaftsinformatik, Prozesse und Systeme, August-Bebel-Str. 89, 14482 Potsdam, bbender@lswi.de

² Universität Potsdam, Lehrstuhl für Wirtschaftsinformatik, Prozesse und Systeme, August-Bebel-Str. 89, 14482 Potsdam, mgrum@lswi.de

anschließend im physischen Produktionsprozess automatisiert umgesetzt werden [BHV14]. Beispielsweise kann beim Abfüllen von Getränken überprüft werden, ob die gewünschte Füllmenge in einzelnen Flaschen erreicht wurde, worin ein Abweichen begründet ist, welche Maßnahmen in systematischen Untersuchungen optimale Ergebnisse erzielen und zur Umsetzung dann in die Produktion geleitet werden.

Stark sinkende Kosten für Datenverarbeitung, Sensorik und Aktorik sowie verbesserte Internetanbindungen haben die Verbreitung von cyber-physischen Systemen gefördert, welche Daten dezentral über eigene Sensoren wahrnehmen, dezentral über eigene Prozessoren auswerten und Aktoren Maßnahmen umsetzen können. Die Ausprägungen von CPS können sehr unterschiedlich sein. Von einzelnen vernetzten Sensoren bis hin zur vollintegrierten Produktionsmaschine. In diesem Kontext wird häufig vom Internet der Dinge oder auch Industrie 4.0 im Produktionskontext gesprochen. Nach einer Schätzung von Cisco wird die Anzahl der internetfähigen Geräte in den kommenden Jahren deutlich steigern: Während 2010 noch 12,5 Milliarden Geräte im Einsatz waren, wird für 2020 bereits der Einsatz von 50 Milliarden Geräten erwartet [Ev11]. Dies soll den rasanten Anstieg an netzwerkfähigen Geräten sowie das Ausmaß der damit verbundenen Daten verdeutlichen [AIM10]. Die Rolle einer zeitnahen bzw. Echtzeitauswertung der Informationen nimmt hier einen größeren Stellenwert ein als zuvor, da Tages-, Wochen- oder Monatsberichte für eine operative Steuerung der Produktionsprozesse ungeeignet sind [Mi12]. Die Praktikabilität eines Data Warehouses zur zentralisierten Vorhaltung der Informationen ist somit zu hinterfragen, da die Entscheidungen zeitnah und im entsprechenden Steuerungssystem bzw. dezentral an den einzelnen Komponenten zu finden sind. Der Berechnungsaufwand sowie die Zeitkritikalität der Analyseaufgaben variiert und dementsprechend ist für jede Analyseaufgabe eine adäquate Verarbeitung entsprechend der verfügbaren Hardware sicherzustellen. Ein hier entwickelter Ansatz soll Hilfestellung leisten.

Beim Aufbau von Hardwareinfrastrukturen stellt das Finden einer angemessenen Balance von Flexibilität und Effizienz eine wesentliche Herausforderung dar. Dabei stellt die Dimensionierung der Hardwarekomponenten ein wesentliches Merkmal dar. Angesichts langer Einsatzzeiträume realisierter Hardwareinfrastrukturen und einer komplexen Abschätzung des zukünftigen Bedarfs, bringen Ineffizienzen in dem Aufbau langwährende Konsequenzen mit sich. Weiterhin ist die Wirtschaftlichkeit in Bezug auf die Dimensionierung von großer Bedeutung. Eine auf *Flexibilität* ausgerichtete Infrastruktur hält tendenziell mehr Ressourcen vor, resultiert somit aber in höheren Kosten und ist tendenziell nicht ausgelastet. Eine auf *Effizienz* ausgerichtete Infrastruktur hält weniger Ressourcen vor und ist somit günstiger und ausgelasteter, welches bei vielen gleichzeitigen Aufgaben zu einer verzögerten Verarbeitung führen kann. Die Frage der Dimensionierung stellt sich besonders bei dem Einsatz vieler dezentraler IT-Komponenten, welche im Kontext des Internets der Dinge oder der Industrie 4.0 zunehmend an Bedeutung gewinnen. Da die einzelnen dezentralen Komponenten, häufig auch als cyber-physische Systeme (CPS) bezeichnet werden und meist nur für ihre speziellen Einsatzzwecke ausgelegt sind, kann eine mögliche Überkapazität bei einer auf Flexibilität ausgelegten Infrastruktur nicht für weitere Aufgaben innerhalb der IT-

Infrastruktur genutzt werden [VF14]. Zentrale Infrastrukturen können deutlich universeller fungieren und auch dezentral anfallende Aufgaben bearbeiten, womit jedoch eine Übertragung der dafür notwendigen Daten und ggf. Anweisungen notwendig wird. Der hier entwickelte Ansatz soll sich diesen Herausforderungen stellen.

Im Kontext des EAM soll somit ein dynamisches Architekturkonzept entwickelt werden, welches den flexiblen Einsatz von Analytics unter Beachtung von Ressourcenbeschränkungen der unterschiedlichen IT-Infrastrukturkomponenten effizient ermöglicht. Hierfür wird im folgenden Abschnitt der Aufbau von Analytics-Infrastrukturen als Ansatzpunkt für das Architekturkonzept beschrieben. Im dritten Abschnitt wird das Architekturkonzept und damit verbundene Annahmen erläutert, aufbauend wird eine Entscheidungsmaxime gezeigt. Im fünften Abschnitt erfolgt im Rahmen eines Ausblicks der Verweis auf einen Evaluationsansatz für die entwickelte Lösung.

2 Analytics-Infrastrukturen

Als Ausgangspunkt für das flexible Analytics-Architekturkonzept dient im folgenden eine Analytics-Infrastruktur in Anlehnung an „TwinCAT Analytics“ [Pa16], welche in Abb. 1 dargestellt ist.

Es wird deutlich, dass Endkunden über diverse cyber-physische Systeme mit variierenden Ausstattungsmerkmalen verfügen, die mittels einer Netzwerk- bzw. Internet-typischen Kommunikation Informationen austauschen können. Analytics-Aufgaben können somit auf dem CPS selbst ausgeführt werden, oder aber auf eine lokale oder in eine öffentliche Cloud verlagert werden. Auf der jeweiligen Analytics-Ebene müssen hierfür eigene Prozessoren zur Auswertung und eigene Speicher zur Datenhaltung zur Verfügung gestellt werden. Für eine weitere Differenzierung können diese Ebenen entsprechend ihrer Fähigkeiten unterschieden werden [Ko10]. Zu Beginn wird davon ausgegangen, dass Architekturebenen Aufgaben tieferliegender Ebenen berechnen können, ohne koordinierende Funktion nicht jedoch umgekehrt. Beispielsweise kann die Local Cloud Berechnungen anstelle der CPS durchführen, diese jedoch nicht anstelle der Local Cloud, was zum einen in der stärker ausgeprägten Abstraktionsfähigkeit höherer Ebenen begründet ist – hier sind mehr Funktionen implementiert, die auf tieferen Ebenen nicht zur Verfügung stehen. Zum anderen kann dies durch den weiterreichenden Einflussbereich bzw. deren größeren Reichweite im Sinne ihres Umgebungsbewusstseins erklärt werden [Ko10]. So werden höhere Systeme über eine größere Anzahl weiterer Systeme gekoppelt und verfügen über zusätzliche Informationen, die in den von tieferen Systemen verfügbaren Daten nicht enthalten sind. Weiterhin können externe Hardwarekomponenten oder dritte Analysten über Schnittstellen in die Auswertung integriert werden.

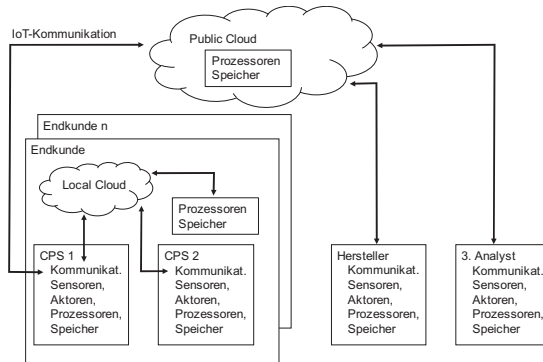


Abb. 1: Analytics-Infrastruktur im Kontext der Industrie 4.0

Die Durchführung der Aufgabenverarbeitung ist somit in Abhängigkeit von der Ausgestaltung der bei den Akteuren verbauten Hardwarekomponenten zu wählen, sodass der *Ressourcenbedarf* einer Analyticsaufgabe erfüllt wird. Ausgehend von der Annahme, dass die verfügbaren Prozessoren-, Kurz- und Langzeitspeicherelementen zu sich unterscheidenden Kostenbedingungen zum Einsatz kommen, können relevante Einheiten entsprechend ihrer *Leistungsfähigkeit* und *Verfügbarkeit* im Sinne einer Auslastung als betriebliche Ressource identifiziert werden und bedürfen einer Steuerung hinsichtlich ihrer Prozessoren, Sensoren, Aktoren und Kommunikationsmöglichkeiten. Diese entsprechen den typischen Komponenten eines CPS [Ve13]. Eine Messung auf Erfolg sollte sich somit an betrieblichen Kennzahlen orientieren, wie beispielsweise für Prozesse den Berechnungskosten je Bearbeitungszeit, für RAM der Speichernutzung je Operation, einer bereichsspezifischen Netzwerkauslastung, oder dergleichen Key Performance Indicators.

Die Verortung der Aufgabenverarbeitung ist weiter in Abhängigkeit von dem Berechnungsaufwand der jeweiligen Analyticsaufgabe zu wählen. Es ist anzunehmen, dass der Fokus von operativen CPS auf der Realisierung ihrer operativen Tätigkeit liegt, nicht jedoch auf der Auswertung von Analyticsaufgaben über den eigenen *Kontext* hinaus. So ist beispielsweise die Bestimmung des optimalen Wartungszeitpunktes einer Maschine bei der Maschine selbst sinnvoll verortet, die Bestimmung sämtlicher Wartungszeitpunkte in Anbetracht der gesamten Produktion und bestehender Abhängigkeiten jedoch auf einem besser ausgestatteten Analytics-System. Dies verdeutlicht, dass Aufgaben größerer Kontexte mit einem höheren Berechnungsaufwand und zumeist höheren Aggregationsebenen einher gehen. Ausgehend von dem Kontext, können simple Events im Rahmen eines *event-driven Analytics* vermehrt auf Ebene der CPS verortet werden, komplexere Events auf Ebenen größerer Kontexte. Im Rahmen eines *data-driven Analytics* kann davon ausgegangen werden, dass explorierende Methoden in einem engen Kontext weniger Berechnungsaufwand erfordern als Aufgaben größerer Kontexte.

Eine Verortung der Aufgabenverarbeitung hängt weiter an der *Parallelisierbarkeit* der

jeweiligen Analyticsaufgabe ab. Aufgaben, die nicht verteilt prozessiert werden können, erfordern Analyticskomponenten mit leistungsfähigeren Ausprägungen. So ist z.B. die Verarbeitung einer zeitlichen Sequenz einer Produktionsmaschine nur im Gesamten aussagekräftig, kann nicht zerteilt und verteilt berechnet werden, die Analyse unabhängiger Informationen wie zeitliche Muster verschiedener Maschinen jedoch schon.

Zusammenfassend kann eine Verortung der Analyticsaufgaben entsprechend folgender Aufgabencharakteristiken geschehen: Ressourcenbedarf, Kontext (event-driven und data-driven) und Parallelisierbarkeit. Charakteristiken einer Hardware (Leistungsfähigkeit und Verfügbarkeit) sind im Architekturkonzept ebenso genüge zu tun. Hierbei ist insbesondere zwischen statischen (z.B. Leistungsfähigkeit) und dynamischen Aspekten (z.B. Verfügbarkeit im Sinne der gegenwärtigen Auslastung) zu unterscheiden.

3 Architekturkonzept

In diesem Abschnitt erfolgt die Entwicklung des Architekturkonzepts und den damit zusammenhängenden Entscheidungshilfen. Für Entwicklung des Architekturkonzepts wird in Anlehnung an der gezeigten Analytic-Infrastruktur von einer schematischen, dreistufigen Informationsinfrastruktur ausgegangen (siehe Abb. 2), welche in Abhängigkeit von den vorhandenen Aggregationsebenen im betrieblichen Kontext beliebig erweitert werden können. Beispielsweise können private Clouds auf Produktionsbereichs-, Abteilungs- oder Unternehmensebene existieren.

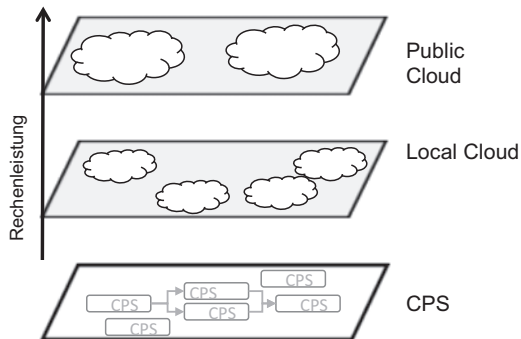


Abb. 2: Architekturkonzept mit unterschiedlichen Ebenen

Auf der untersten Ebene befinden sich alle im operativen Leistungserstellungsprozess tätigen Infrastrukturkomponenten, nicht jedoch übergeordnete Systeme, welche für mehrere Bereiche des Unternehmens tätig sind. Am Beispiel der Produktion umfasst die CPS-Ebene einzelne Maschinen, welche am Herstellungsprozess beteiligt sind. Die mittlere Ebene umfasst die im Unternehmen übergreifend Hardwarekomponenten,

welches zumeist eine zentrale Recheninfrastruktur oder sogar ein Rechenzentrum umfasst. Hier werden für das Unternehmen zentral tätige Dienste und Systeme wie z.B. ein ERP-System oder Dokumentenmanagement als lokaler Cloud-Service betrieben. Die dritte Ebene (Public Cloud) ist nicht direkt Teil der Unternehmensinfrastruktur. Diese Ebene umfasst Ressourcen ausstehender Unternehmen, welche gemietet werden können und sich durch eine hohe Skalierbarkeit auszeichnen [Gu13]. Die Rechenleistung der auf der jeweiligen Ebene zur Verfügung stehenden Komponenten steigt mit zunehmender Ebene an.

Im Kontext der Analytics stellt sich die Frage, auf welcher der Ebenen die unterschiedlichen Analyseaufgaben prozessiert werden sollten. Wie dem zweiten Abschnitt entnommen werden kann, können Analysefragestellungen und Hardware unterschiedlicher Charakterisierungen identifiziert werden. Abb. 3 stellt die Bearbeitungsdauer von Aufgaben auf den drei Ebenen des vorgestellten Architekturkonzepts in Abhängigkeit des jeweils benötigten Berechnungsaufwandes dar. In der Annahme, dass die für die Analyse erforderlichen Daten auf der untersten CPS-Ebene entstehen, ist für die Bearbeitung auf einer übergeordneten Ebene die Übertragung der hierfür erforderlichen Daten von Nöten, welche zusätzlich Zeit erfordert und in der vorliegenden Grafik als Bias interpretiert werden kann. Für die CPS-Ebene entfällt eine derartige initiale Übertragungsdauer, da eine Auswertung auf dem CPS selbst geschieht. Weiter liegt die Annahme zugrunde, dass höhere Ebenen des vorgestellten Architekturkonzepts mehr Rechenleistung der Prozessierung der jeweiligen Aufgaben zur Verfügung stellen können, eine Bearbeitung somit schneller erfolgen kann und ein Anstieg der jeweiligen Geraden flacher ausfällt. Beispielsweise ist bei der Bearbeitung einer Analyseaufgabe auf der zweiten Ebene vor Beginn der Bearbeitung die, ggf. vollständige, Übertragung der Rohdaten erforderlich, welche in Abhängigkeit von der Anbindung nennenswert Zeit in Anspruch nehmen kann. Der Anstieg mit zunehmendem Berechnungsaufwand hingegen ist deutlich geringer als auf der untersten Ebene, da leistungsfähigere Ressourcen zur Verfügung stehen. Die Bearbeitung auf der dritten Ebene (Public Cloud) setzt eine zusätzliche Übertragung voraus und ist mit einer Initialisierung verbunden, insofern die Ressourcen nicht bereits angemietet wurden und zur Verfügung stehen. Dementsprechend ergibt sich hierfür der größte Initialaufwand, dementsgegen allerdings die leistungsfähigsten Ressourcen stehen.

Die Vorteilhaftigkeit jeder Ebene ist in der Abbildung anhand der blau eingezeichneten Vorteilsbereiche ersichtlich. Die Grenzen finden sich jeweils in den Schnittpunkten der zugehörigen Funktionen, sodass leicht die zeitoptimalen Ebenen identifiziert werden können. Zur Vereinfachung ist in der Abbildung exemplarisch jeweils nur ein Vertreter je Ebene ersichtlich. Da andere Systeme gleicher Ebenen verschiedene Hardwarecharakteristiken aufweisen können und ein sich unterscheidender Kommunikationsaufwand für dessen Ansprache erforderlich ist, zeigt sich dies in Funktionen verschiedener Neigungen und Schnittpunkte mit der y-Achse. Für eine vertikale Eskalation, also der Übertragung einer Aufgabe von einem System der einen Ebene auf ein System einer nächsten Ebene, wäre dann das System der nächsten Ebene zu bevorzugen, welches für die entsprechende Aufgabe den besseren Wert aufweist.

Exemplarisch zeigt Abb. 3 die aufwärtsgerichtete, vertikale Eskalation einer Beispielaufgabe von der privaten Cloud in eine öffentliche Cloud.

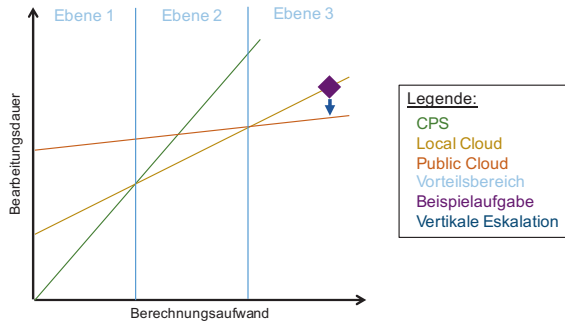


Abb. 3: Vorteilhaftigkeit identifizierter Architekturebenen

Weiterhin denkbar ist eine horizontale Eskalation, welche analog die Übertragung einer Aufgabe von einem System der einen Ebene auf ein System der gleichen Ebene bezeichnet. In der Grundversion der Aufgabendistribution (siehe nächster Abschnitt) wird diese Option zur Vereinfachung noch nicht integriert, da hierfür zusätzliche Abhängigkeiten beachtet werden müssten.

4 Entscheidungsmaxime

Angesichts verschiedener Aufgabencharakterisierungen sollen Aufgaben entsprechend ihres Bedarfs an Berechnungsaufwand und der verfügbaren Berechnungsleistungen (s. Hardwarecharakterisierung) den Architekturebenen zugeordnet werden. Als Grundlage für diese Entscheidung dient im Folgenden deren Wichtigkeit als die Bedeutung der Aufgabe für das Gesamtsystem (z.B. Betrachtung der Auswirkungen für das Gesamtsystem wenn die Aufgabe nicht ausgeführt wird), sowie die Dringlichkeit als Kehrwert der verbleibenden Zeit bis das Ergebnis der Berechnung vorliegen muss. Diese beiden Kriterien werden für die folgende Betrachtung unter dem Begriff der Priorität subsumiert. Die Priorisierung kann weiterhin anhand der für das individuelle System relevanten Dimensionen wie z.B. den Bearbeitungskosten, der erforderlichen Bearbeitungsqualität u.v.m. vorgenommen werden.

Abb. 4 zeigt vier verschiedene Strategien wie mit Aufgaben in Abhängigkeit der zuvor beschriebenen Dimensionen umgegangen werden soll, bzw. ob eine Aufgabe auf der betrachteten Architekturebene prozessiert oder auf eine höhere Architekturebene eskaliert werden soll. Da die Daten entsprechend den Annahmen auf der untersten Ebene des Architekturmodells entstehen, erfolgt die Betrachtung beginnend bei der Ebene einzelner CPS. Die dargestellte Entscheidungsmaxime kann jedoch analog für weitere Ebenen und den dort anfallenden Aufgaben angewendet werden. Der Schnittpunkt der gestrichelten Linien aus Abb. 4 kann den Funktionsschnittpunkten der jeweiligen

Ebenen zugeordnet und somit anschaulich interpretiert werden (vgl. Abb. 3).

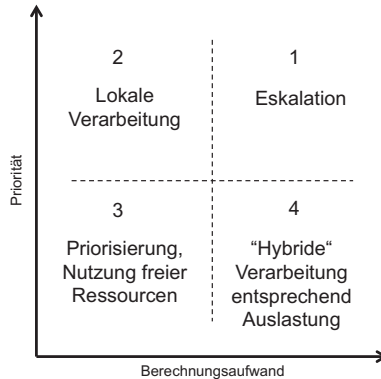


Abb. 4: Eskalationsmatrix als ebenenspezifische Entscheidungsmaxime

Ist eine Aufgabe von hoher Priorität und erfordert einen geringen Berechnungsaufwand, so sollte diese lokal verarbeitet werden, da die Ergebnisse zeitnah benötigt werden und vorhandene Ressourcen die Aufgabe bewältigen können (siehe Quadrant 2). Im Fokus stehen hier Analytics-Aufgaben, dessen Events für die Ebene adäquate Komplexität und Datenkontexte aufweisen und sich mit der Hardwareleistungsfähigkeit und -verfügbarkeit im Fit befinden.

Eine Aufgabe mit geringer Priorität und geringem Berechnungsaufwand wird dem dritten Quadranten zugeordnet und sollte entsprechend der gegenwärtigen Auslastung des CPS in Zeiten mit freien Kapazitäten bearbeitet werden. Erfolgt die Bearbeitung allerdings nicht rechtzeitig, kann diese Aufgabe zu einer Typ 2-Aufgabe werden und erfährt eine Prioritätsaufwertung. Die Hardwareleistungsfähigkeit und -verfügbarkeit spielt zunächst nur eine untergeordnete Rolle.

Eine Aufgabe mit hoher Priorität und intensivem Berechnungsaufwand (Quadrant 1), der die Kapazitäten auf der Ebene ggf. sogar überschreitet, sollte auf eine höhere Architekturebene eskaliert werden, um eine rechtzeitige Bearbeitung der umfangreichen Aufgabe sicherzustellen. Im Fokus stehen hier Analytics-Aufgaben, die für die Ebene zu komplexe Events und zu großer Datenkontexte aufweisen.

Für Aufgaben mit großer Intensität und geringer Priorität (Quadrant 4) bietet sich eine hybride Verarbeitung an. Dabei kann ein Teil der Aufgaben auf eine höhere Ebene eskaliert werden und eher weniger aufwendige Berechnungsschritte auf der Ebene selbst ausgeführt werden. Im Fokus stehen hier Analytics-Aufgaben, die für die Ebene zu komplexe Events und zu großer Datenkontexte aufweisen und zumindest teilweise relativ gut verteil- bzw. parallelisierbar sind.

Die Nutzung des vorgestellten Architekturkonzepts unter Beachtung der Ressourcen- und Zeitbedarfe der entsprechenden Aufgaben eignet sich nicht nur für Aufgaben,

welche auf Basis von Sensorik im operativen Geschäftsbetrieb (z.B. Produktion) entstehen. Als Entscheidungsmaxime ist für jeden Analytics-Teilnehmer in der beschriebenen Infrastruktur für die effiziente Auswahl der entsprechenden Verarbeitungsebene die obenstehende Eskalationsmatrix zu implementieren, sodass eine Aufgabenverteilung flexibel und zugleich dezentral realisiert werden kann.

5 Limitationen, Validierung und Ausblick

Im Zuge einer initialen Ideenformulierung findet sich im Folgenden eine Sammlung von Limitationen der gezeigten Ansätze, welche im Zuge von einer Validierung berücksichtigt und sukzessive verringert werden sollten.

Es sei angemerkt, dass für die Ermittlung der architekturebeneninherenten Vorteile initial lineare Zusammenhänge angenommen werden, welche sich in Untersuchungen bestätigen müssen. In aufbauenden Modifikationen müssen ebenso relevante Einflussfaktoren auf Steigung und Bias der jeweiligen Funktionen nähergehend untersucht werden.

Die gezeigten Ausführungen fokussieren die Bearbeitungsdauer und Berechnungsaufwand zur Verdeutlichung der Vorteile exemplarischer Architekturebenen. Andere KPIs können analog gehandhabt werden, resultieren vermutlich jedoch in variierenden Vorteilsebenen. Ein Unterschied wäre hier zu identifizieren und weiter ist ein integrierter Ansatz zu schaffen.

Vielversprechend ist weiter eine Ausgestaltung der x- und y-Achsen des Architekturkonzepts. Beispielsweise ist anzunehmen, dass die Wahl nah verorteter Systeme Auswirkungen auf den Verlauf der Ebenenfunktionen hat (z.B. durch eine bessere Anbindung und kürzere Netzwerklatenzen). Die Vermessung nah verorteter Systeme kann jedoch auf verschiedene Weisen geschehen. Beispielsweise seien hier eine physische Verortung oder eine Verortung auf Basis von Übertragungsgeschwindigkeiten aufgeführt.

Im Zuge von Validierungsläufen sind diverse Hardwarekonfigurationen und Aufgabenszenarien entsprechend des Architekturmodells zu schaffen und Aufgabenprozessierungen entsprechend der Eskalationsmatrix zu simulieren. In systematischen Simulationen können Parameter gezielt variiert werden, sodass auf die Tauglichkeit der Entscheidungsmaxime geschlossen werden kann. Anzunehmen ist, dass Aufgaben flexibel und effizient bearbeitet werden können. Gegebenenfalls können Modifikationen in der Eskalationsmatrix leicht auf eine Verbesserung geprüft werden.

Weiter sind dynamische Entwicklungen im Verlauf der Zeit zu berücksichtigen. Diese können sich von kurzfristigen Entwicklungen bspw. in Folge von aktuellen Auslastungen, bis hin zu längerfristigen Entwicklungen wie der Neugestaltung von Public-Cloud-Verträgen erstrecken. Zeitrelevante Einflussfaktoren sind somit ebenso zu

erheben sowie in den Validierungsläufen zu berücksichtigen.

Im Sinne eines Benchmarkings kann der erarbeitete Ansatz im Vergleich zu alternativen Ansätzen Störfaktoren unterzogen werden, sodass Grenzen der Ansätze identifiziert werden können und sich individuellen Stärken und Schwächen zeigen lassen.

Literaturverzeichnis

- [AIM10] Atzori, L.; Iera, A.; Morabito, G.: The Internet of Things: A survey. In *Computer Networks*, 2010, 54; S. 2787–2805.
- [BHV14] Bauernhansl, T.; Hompel, M. ten; Vogel-Heuser, B. Hrsg.: *Industrie 4.0 in Produktion, Automatisierung und Logistik. Anwendung, Technologien, Migration*. Springer Vieweg, Wiesbaden, 2014.
- [Ev11] Evans, D.: *Das Internet der Dinge. So verändert die nächste Dimension des Internet die Welt*, 2011.
- [Gu13] Gubbi, J. et al.: Internet of Things (IoT). A vision, architectural elements, and future directions. In *Future Generation Computer Systems*, 2013, 29; S. 1645–1660.
- [Ko10] Kortuem, G. et al.: Smart objects as building blocks for the Internet of things. In *IEEE Internet Computing*, 2010, 14; S. 44–51.
- [Mi12] Miorandi, D. et al.: Internet of things. Vision, applications and research challenges. In *Ad Hoc Networks*, 2012, 10; S. 1497–1516.
- [Pa16] Pascal, D.: TwinCAT beherrscht Big Data. In *PC-Control*, 2016; S. 15–18.
- [Ve13] Veigt, M. et al.: Entwicklung eines Cyber-physischen Logistiksystems. In *Industrie Management*, 2013; S. 15–18.
- [VF14] Vermesan, O.; Friess, P.: *Internet of things. From research and innovation to market deployment*. River Publishers, Aalborg, 2014.
- [XHL14] Xu, L. D.; He, W.; Li, S.: Internet of Things in Industries: A Survey. In *IEEE Transactions on Industrial Informatics*, 2014, 10; S. 2233–2243.

Lufthansa Aviation Standard: Developing an Open Group Reference Architecture for the Aviation Industry

Eldar Sultanow¹, Carsten Brockmann², Kai Schroeder³, Carsten Breithaupt⁴

Abstract: Since March 2016, architects from Lufthansa and Capgemini have been collaborating on developing an Aviation Reference Architecture (RA), designed to serve as the basis of an Open Group Standard. Central to the development and description of the RA is an Architecture Definition Document (ADD), which explains the RA in its four TOGAF-specific architecture domains, namely the Business Architecture, Application Architecture, Data Architecture and Technology Architecture. In this paper, the RA will be presented in these four aspects, and the “Revenue Management & Pricing” business function has been selected to model and describe the core elements of the RA.

Keywords: Reference Architecture, Aviation, Lufthansa

1 Problem Statement and Motivation

As a complex organization, Lufthansa constantly explores and pursues various activities to improve, standardize and make and the architecture processes and its assets more effective. As part of the IT Management activities, the Architecture Board of Lufthansa decided to invest in drawing up and developing a RA to gain insights and capitalize on the potential a RA provides. Capgemini was chosen as a partner to assist with the development and publication of the RA.

On Lufthansa’s side, the project has three business drivers: Firstly, Lufthansa intends to achieve process and time-to-market improvements by avoiding having to educate their relevant suppliers on the way it approaches, structures and delivers selected IT services internally. Secondly, architects and/or companies new to the industry can utilize the knowledge and information inherent in the RA to achieve greater productivity in shorter periods of time. Thirdly, Lufthansa can generate a structured and cascading view on an aviation industry-specific EA that transparently describes, in a modularized fashion, various capabilities, including business capabilities, components/services, applications, interoperability, and creates the opportunity to develop combined capabilities and modules in a controlled fashion.

¹ Capgemini, Bahnhofstraße 11C, 90402 Nuremberg, Germany, eldar.sultanow@capgemini.com

² Capgemini, Potsdamer Platz 5, 10785 Berlin, Germany, carsten.brockmann@capgemini.com

³ Capgemini, Luebecker Strasse 128, 22087 Hamburg, Germany, kai.schroeder@capgemini.com

⁴ Deutsche Lufthansa Aktiengesellschaft, Von-Gablenz-Straße 2-6, 50679 Köln, Germany
carsten.breithaupt@dlh.de

Using the RA, Lufthansa and Capgemini aim to enable the architectural and delivery aspects leading to improved business efficiency and effectiveness to be pursued through the use of aviation industry standards.

The content and its structure published in this RA respects the boundaries of confidentiality, thereby protecting Lufthansa's unique competitive advantage achieved through proprietary means.

2 Introduction

Reference architectures are used as a checklist for modeling, they provide a procedure for project realization, and ideally they are an international standard [Sc99, S. 3]. A RA serves as a base template for the implementation of architecture in a particular industry context or in a specific application field, and it therefore acts as a design principle, recommendation and proven reference [BD10, S. 201]. The RA defines components and their relationships to each other, as well as between associated interfaces and components. Cloutier et al. define a RA as follows [CM10]:

“Reference Architecture provides a proven template solution for an architecture for a particular domain. It also provides a common vocabulary with which to discuss implementations, often with the aim to stress commonality. A Reference Architecture often consists of a list of functions and some indication of their interfaces (or APIs) and interactions with each other and with functions located outside of the scope of the Reference Architecture.”

TOGAF, is an Enterprise Architecture Framework (EAF) and an open international standard, which was first published in 1995 and is being continuously developed and maintained by the Open Group, a consortium of more than 500 member organizations. It is freely available for internal use in order to develop an EA. TOGAF provides a framework for creating a RA, which has been used for the Lufthansa project presented here. TOGAF enables the development of a unified architecture that reflects stakeholder needs and takes account of current and future business requirements. It distinguishes between four essential architecture domains; Business, Data, Application, and Technology.

The first architecture domain, business architecture covers the entire range from descriptions of corporate strategies to business processes. The second, data architecture covers entities with their relationships that are used to implement business processes – ultimately, the organizational data model is clearly and concisely described. The third architecture domain, application architecture, includes applications, interactions and the interfaces between them as well as the entire application landscape, within which business processes run. The fourth domain, the technology architecture, covers architectural elements for setting up, running and managing the IT infrastructure – it defines the basis on which applications are procured, integrated and operated.

These four architecture domains have been considered in modeling/creating the aviation RA and the “Revenue Management & Pricing” function as the initial business domain.

Several standardization initiatives have been attempted in the past by non-profit organizations and governmental institutions. For example, the *International Air Transport Association (IATA)* has started to develop and publish the *New Distribution Capability (NDC)*, an open XML-based data transmission standard, which aims to enhance the capability of communications between airlines and travel agents [IATA14]. In a similar manner, the *International Civil Aviation Organization (ICAO)* develops infrastructures as well as binding standards for aviation, to enable the implementation by member countries. In addition, in 1999 the European Commission launched the *Single European Sky (SES)*, a European program targeting consistently higher safety standards, increased airspace capacity for air traffic and air navigation service providers that increase efficiencies for the future. This present RA project builds on and involves these standards adequately and at an appropriate level of abstraction. It takes into consideration that the aviation industry is undergoing change. For instance, new NDC and dynamic pricing approaches will have a significant impact on the aviation industry’s architectural landscape.

3 Procedure

The approach to creating the RA is divided into 4 main steps; initiation (1), consolidation and analysis of existing material (2), transformation of existing material into an architecture definition document, abbreviated here as ADD (3), and submitting the ADD to the Open Group (4).

As part of the initiation step, the domain model has been extracted and clarified from Lufthansa’s enterprise architecture, which is described by a collection of architectural artifacts bound together under the umbrella of the *Lufthansa Enterprise Architecture Framework (LEAF)*. The consolidated domain model encompasses the aviation business domains including various industry-specific subdomains.

The second step encompassed the collection, review and in-depth analysis of additional Lufthansa documents, which have been incorporated into the ADD.

The third step involves the actual transformation of all of the previously collected material into the ADD. Additional content-related discussions and exchanges between the architects developing the RA in meetings, video conferences, and other means took place. As part of these discussions and exchanges, additional information has been gathered that was relevant for editing and refining the ADD.

The last step includes the submission of the architecture definition document to the Open Group to enable the publication of a reference architecture standard for the aviation industry. Fig. 1 depicts one of the most fundamental and important preliminary results of this project, the overall domain model of aviation industry consisting of seven domains.

Highlighted here is the “Revenue Management & Pricing” business domain, which serves as the example to model the aviation reference architecture.

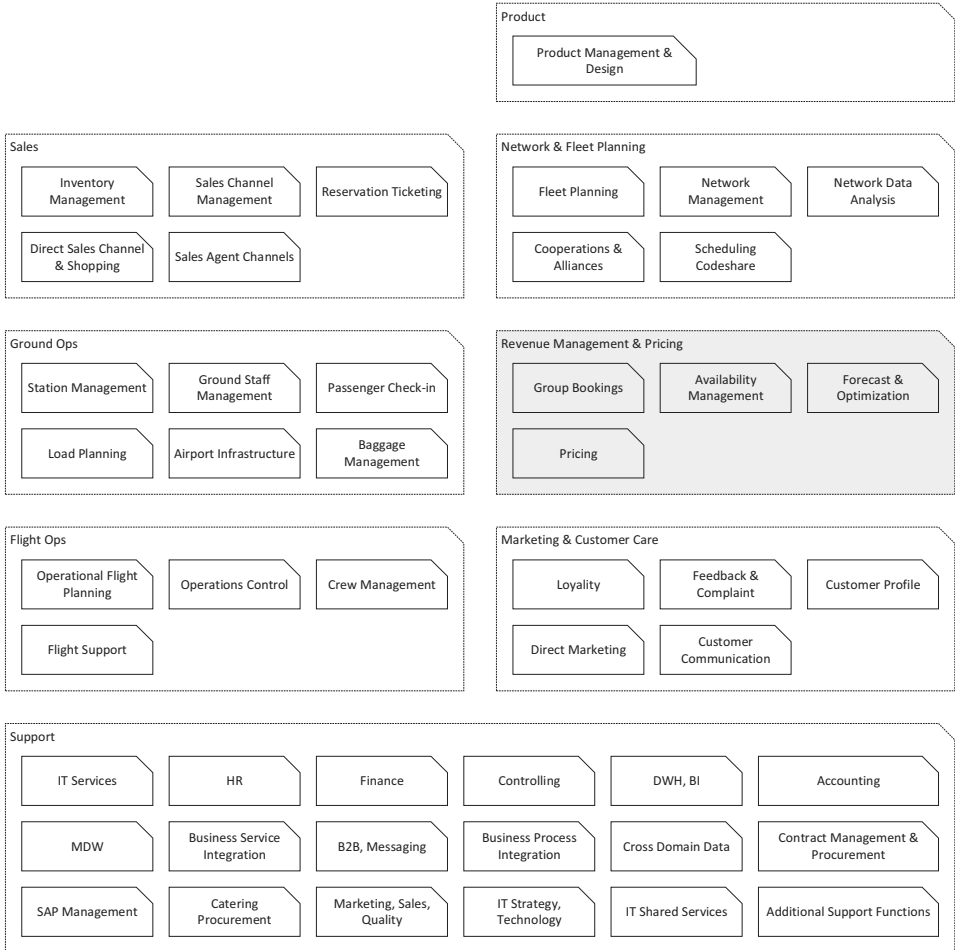


Fig. 1: Aviation Domain Model

4 Modeling the “Revenue Management & Pricing” Domain

In the following chapter, the creation of the ADD will be illustrated using the “Revenue Management & Pricing” business domain. Reference literature on airline ticket pricing and yield management in the aviation industry is provided by Boyd [Bo07], Lindenmeier [Li05] and Talluri/van Ryzin [TaRy04].

Basically, the mission behind “Revenue Management & Pricing” features the sale of the right ticket to the right passenger at the right time for the right price. Since this guiding concept has its origins in the early 70s, it is partially outdated. In fact even in today’s systems this mission translates to: “Publish the right fares beforehand and manage the sales process by opening and closing 26 booking classes corresponding to letters in the alphabet”.

However, the “Revenue Management & Pricing” business domain is undergoing significant change. These changes and the evolution of the overall domain are being progressed by most airlines in order to realize benefits by either:

- implementing an availability management solution outside an Airline Passenger Services Systems (PSS) and/or,
- moving away from the classical availability-based pricing to a more dynamic-based pricing scheme.

The models presented here focus on the enterprise architecture management views. They have been created to capture the content for each TOGAF-specific architecture domain. Hence, the aviation industry-specific model comprises the Business Architecture, Data Architecture, Application Architecture, and Technology Architecture. The domain “Revenue Management & Pricing” consists of four subdomains, “Group Bookings”, “Availability Management”, “Forecast & Optimization”, and “Pricing”.

Managing booking class availability derived from outputs of *Forecast & Optimization* (demand and no-show forecasts, bid prices), direct availability control (in case of irregularity management), decision support for revenue controlling and availability for ancillary and special services (advanced seat reservation, wheelchair, special meal, etc.) are within the scope of the *Availability Management* subdomain.

The second subdomain *Forecast & Optimization* covers forecasts of demand, bookings, cancellations, and buy-downs. In addition, no-show figures based on historical, special and expert data are also included in this domain. The domain also caters for the controlling of the forecast and optimization process, which are monitored here. The aim here is to optimize revenue (fare-mix, overbookings) to steer capacity based on average earnings, the management of availability, inventory and reservation data.

Pricing, the third subdomain, includes the provision and distribution of prices and conditions for each service and the booking class on each market. Maximizing the airline's revenue, monitoring competitor’s prices and defining the price concept and strategy also fall within this domain.

Group Bookings, the last subdomain, deals with checking the availability for groups and the determination of group pricing. This is necessitated by the fact that the process for booking of groups slightly differs from the standard booking process.

4.1 Business Architecture

The business architecture offers a management oriented perspective onto the corporate structure, as it includes a business capability view that groups capabilities by subdomains. This includes a description of business capabilities, a business process that portrays the workflow, as well as the inputs and outputs of the relevant steps.

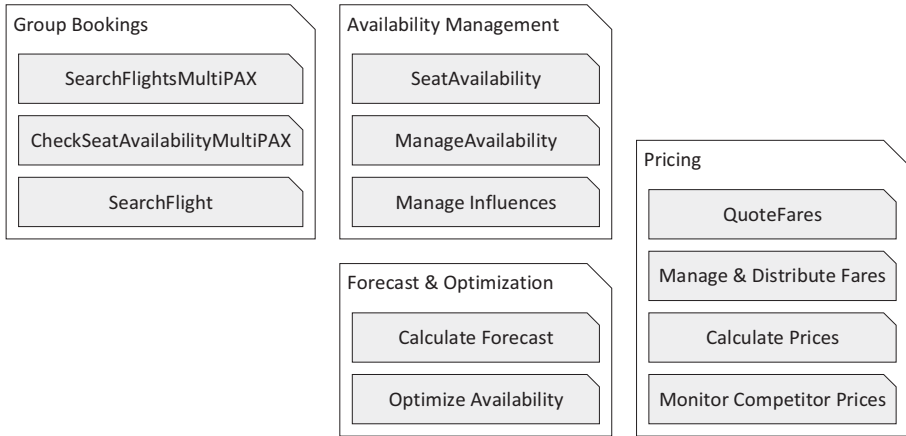


Fig. 2: Business Capability Map of the domain “Revenue Management & Pricing”

The business capabilities grouped by subdomains of the “Revenue Management & Pricing” domain are depicted in Fig. 2. The business capabilities belonging to the “Revenue Management & Pricing” domain are described in Table 1. Several of these have to be transformed into the definitions from the *NDC Implementation Guide* published by the *International Air Transport Association* [IATA14].

In general, business capabilities claim to be free of overlap, express abilities organizations require or offer and provided a logical grouping of these services. A business capability answers the question of “what” a service carries out, clearly distinguishing from the analogous question of “how” something is being carried out. All business capabilities, especially when introduced into a reference architecture, must conform to these principles.

Capability	Description
SearchFlightsMultiPAX	A flight search consist of search criteria including more than 9 passenger criteria and returns availability, fares and flight related ancillaries
CheckSeatAvailabilityMultiPAX	Check the availability of more than 9 seats for a specific flight and booking class

SearchFlight	If the user searches for flights with more than 9 passengers, the seat availability has to be checked first
SeatAvailability	A service to receive available seats by booking class
Manage Availability	(Re-)Calculates availabilities (to be presented to customers or published through the eCommerce channel) based on relevant settings and stored sold bookings
Manage Influences	Allows to adjust the influence settings, which are used as input to the calculation of availabilities and results therefore in a modified output
Calculate Forecast	Calculate the expected demand of seats based on historical bookings and current booking trend
Optimize Availability	Optimizes availabilities by utilizing forecasts and the current sold capacity to recalculate the inventory settings
FareQuote	A service to provide the fares fitting the search criteria (the available flights)
Manage & Distribute Fares	Provision and communication of the fare model including conditions to partners and fare filing
Calculate Prices	Calculate the prices to be provided for the various fares
Monitor Competitor Prices	Track and observe pricing models and current fares offered by competitors

Table 1: Description of *Revenue Management & Pricing* Business Capabilities

Business process models provide a means of developing a behavioral view of the system under investigation. The business process of the domain “Revenue Management & Pricing” is depicted in Fig. 3, which selectively illustrates the input and output of each activity.

As part of the initial step in this process, the strategy, objectives and methods are defined. During the second step, the market is analyzed and tactics are selected, followed by creating a proposal appropriately managed by the governance structures. The penultimate step is to govern capacity.

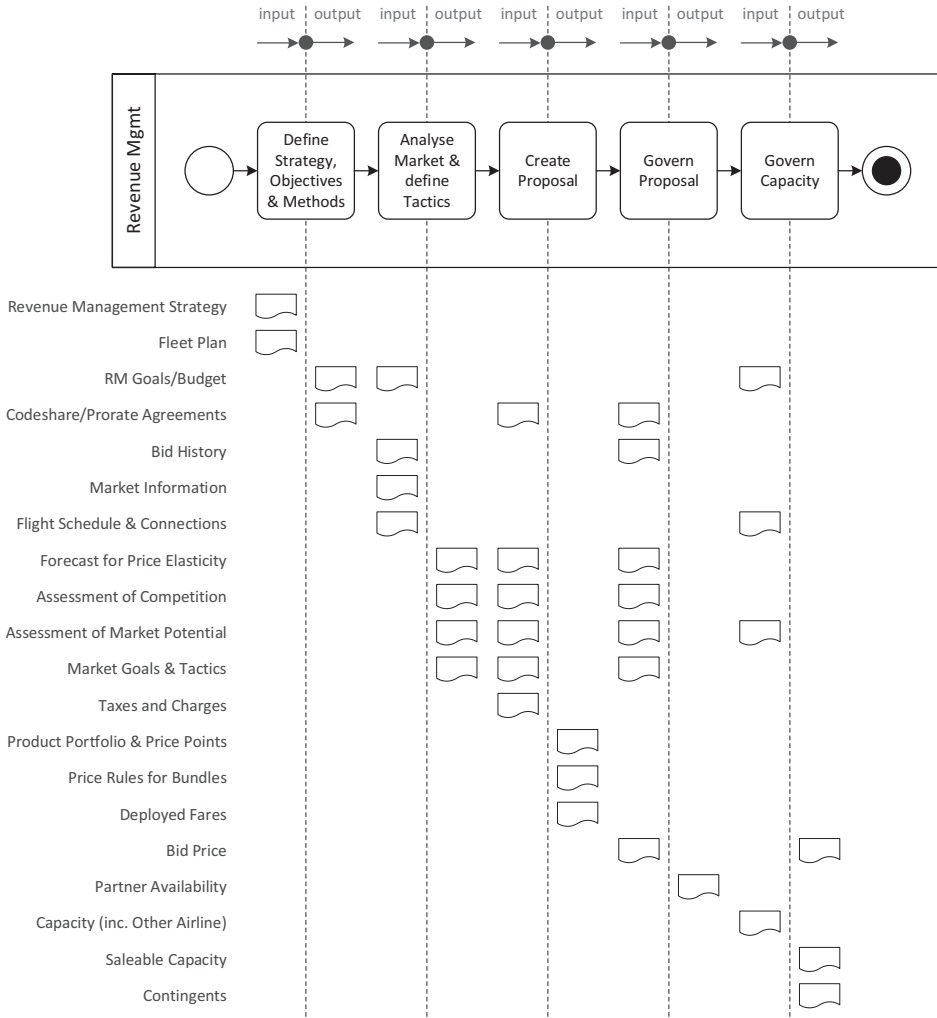


Fig. 3: Revenue Management & Pricing Business Process

4.2 Application Architecture

Fig. 4 depicts the mapping of business capabilities to applications. The main applications assigned to the domain *Group Bookings* are the *Group Offer Manager* and the *Group Sales and Reservation Tool*. A real-time-capable *Availability Calculator* represents the application, which forms core of the *Availability Management* subdomain. Similarly, a *Forecaster* is part of the *Forecast & Optimization* subdomain. The *Pricing* subdomain is covered by the *Fare Engine*, which calculates prices and distributes fares. Also part of the

Pricing subdomain is the *Market Observation Engine* that monitors published offers by competitors. Such an assignment of applications to the business capabilities within the relevant subdomain clarifies what kind of functionality a new application must cater for if it is to be considered to replace an existing application in future.

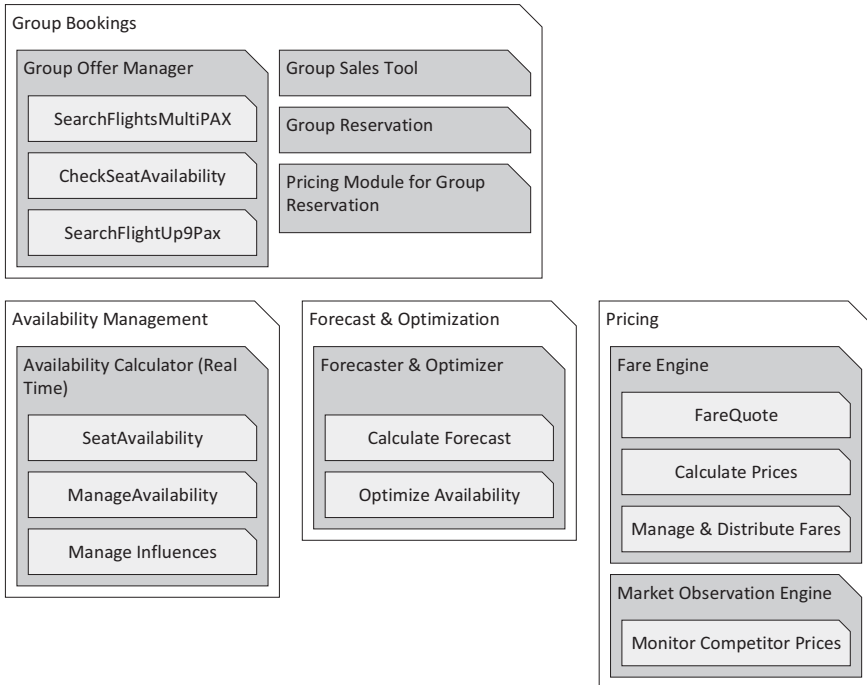


Fig. 4: Mapping Revenue Management & Pricing Capabilities to Applications

4.3 Data Architecture

The data architecture offers an insight into the organizational data model. A suitable reference point for a data model fitting organizations of the aviation industry is the *IATA Airline Industry Data Model (AIDM)*. The AIDM data model facilitates the development and maintenance of messaging standards in the aviation industry [ThHo16]. The basis is a consistent model collaboratively developed and maintained by globally distributed industry working groups that use the Sparx Enterprise Architect tool, including the “industry-agreed vocabulary, data models, and message definitions, as well as the related business process context and requirements” [ThHo16]. Oracle provides a particularly detailed data model as well, the *Oracle Airline Data Model (OADM)*. It is “a standards-based, industry-specific, prebuilt data warehouse database schema with associated analytic models and dashboard” and a key component of the Oracle Passenger Data Management Industry Solution [Orac16].

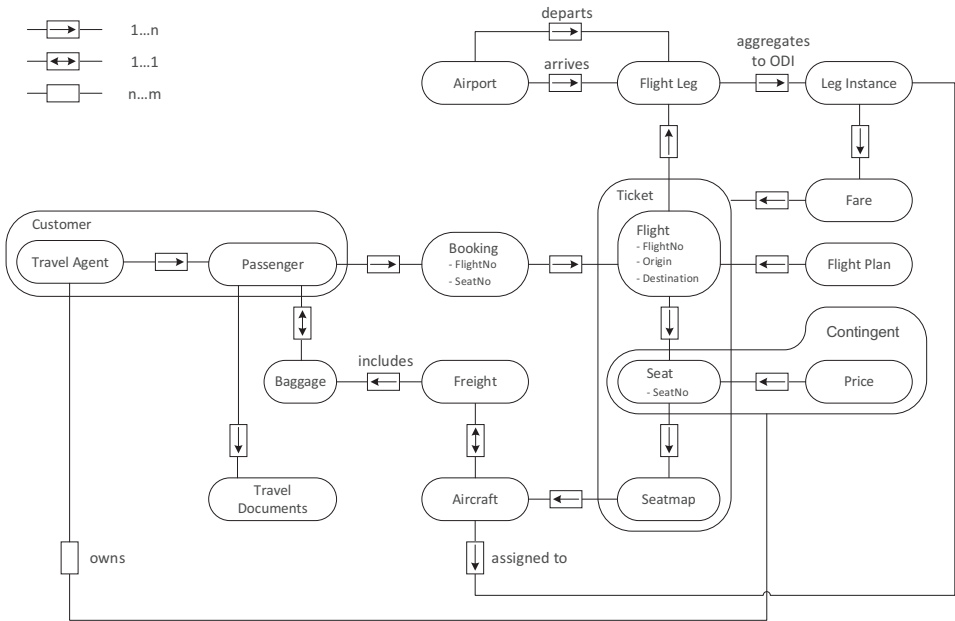


Fig. 5: Data Model including the Revenue Management & Pricing Domain

Here, a reasonably abstracted segment is displayed in Fig. 5 illustrating fundamental business entities of the Revenue Management & Pricing domain and their relationships.

4.4 Technology Architecture

The technology layer of the Revenue Management & Pricing domain is subdivided into four technological areas, *Bookings & Direct Sales* (1), *Forecast* (2), *Availability Optimization* (3), and *Availability Management* (4). The first area includes an E-Commerce Web Portal backed by a Shop System. Within the second area, namely the *Forecast* area, a data loader, integrator, store and analyzer are located, which corresponds to a typical *Business Intelligence (BI)* architecture. *Availability Optimization*, the third area, is technologically represented by a *Linear Optimization Problem Solver* such as *IBM CPLEX*. An *Availability Request Processor*, which is high throughput-oriented, forms the core part of the last area: *Availability Management*. Here, the critical aspect is the ability to process as many requests in the shortest possible timeframe. Within this technology layer, essential platforms can be found, providing various services to applications, which are located in a layer above.

5 Concluding Summary

As a summary, we conclude that a project to design an industry-specific Reference Architecture and Standard is not only beneficial for organizations of this particular industry, but a valuable activity for organizations in general, such as suppliers of services or products.

Aviation industry service providers benefit from a generic framework that serves as a guideline or template for their enterprise architecture. In addition, actors share a common vocabulary when designing solutions, information systems or interfaces between them. Companies, peripheral to the aviation industry are able to create and deliver value through using the Reference Architecture by finding deeper insights into apparent processes and industry-specific realities.

The next step of this project will be to finalize the architecture definition document and refine the model and structures with various other aviation industry organizations to complete the Open Group Standards submission process.

6 Challenges and Outlook

The content, observations and analysis of this RA have been developed with the help of a group of people under the leadership of the Lufthansa CIO Dr. Roland Schütz and the Lufthansa Corporate CIO Josef Bogdanski, mainly in the team of the Lufthansa Director for Revenue Management Dr. Jutta Rockmann.

This paper was developed under the joint leadership of Carsten Breithaupt (Lufthansa Head of Domain Architectures) and Kai Schroeder (Capgemini Germany - Head of BTech SI Architecture), the coordinator and project manager Carsten Brockmann (Capgemini), and the lead architects Eldar Sultanow (Capgemini) and Christian Vollmer (Lufthansa). A team of eleven additional architects poured their combined knowledge and wisdom into this RA, protecting commercially sensitive information, while making the content for the reader as accessible and as meaningful as possible.

References

- [BD10] Bruns, R., & Dunkel, J. (2010). *Event-Driven Architecture: Softwarearchitektur für ereignisgesteuerte Geschäftsprozesse*. Berlin, Deutschland: Springer.
- [Bo07] Boyd, E. A. (2007). *The Future of Pricing: How Airline Ticket Pricing Has Inspired a Revolution*. New York, NY: Palgrave Macmillan.
- [CM10] Cloutier, R., Muller, G., Verma, D., Nilchiani, R., Hole, E., & Bone, M. (2010): The Concept of Reference Architectures. *Systems Engineering*, 13(1), 14-27, DOI: 10.1002/sys.20129.

- [Orac16] Oracle (2016). Oracle Airline Data Model, <http://www.oracle.com/us/industries/airline-data-model-1453220.html>
- [TaRy04] Talluri, K., & van Ryzin, G. J. (2004). *The Theory and Practice of Revenue Management*. New York, NY: Springer Science+Business Media.
- [ThHo16] Thomas, M., Hofman, A. (2016). IATA Airline Industry Data Model. Dutch Sparx Enterprise Architect Event, Schiphol: The Netherlands, June 10th 2016 <http://www.eausergroup.com/component/content/article/16-next-event/128-schiphol-speakers-and-abstracts-june-2016>
- [IATA14] International Air Transport Association (2014): *Simplifying the Business: The NDC Implementation Guide*. Montreal, Canada: IATA.
- [Li07] Lindenmeier, J. (2005). *Yield-Management und Kundenzufriedenheit: Konzeptionelle Aspekte und empirische Analyse am Beispiel von Fluggesellschaften*. Wiesbaden, Germany: Deutscher Universitäts-Verlag.
- [Sc99] Schmidt, G. (1999). *Informationsmanagement: Modelle, Methoden, Techniken* (2. Auflage). Berlin, Germany: Springer.

Additional Reading

- Brockmann, C. (2014). *An approach to design the business model of an ERP vendor*. Berlin, Deutschland: Gito.
- Sultanow, E. (2015). *Real World Awareness in kollaborativen Unternehmens-prozessen*. Berlin, Deutschland: Gito.

IT-Governance und Strategisches Informations- management (ITG-SIM)

IT-Governance und Strategisches Informationsmanagement (ITG-SIM)

Daniel F. Abawi¹, Matthias Goeken², André Miede³

1 Motivation und Thema

Die verlässliche sowie effektive und effiziente Gestaltung und Steuerung der IT ist für viele Unternehmen und Organisationen sowohl aus regulatorischen Gründen als auch zur Steigerung der Wettbewerbsfähigkeit und zur Sicherung eines Beitrags der IT zum Unternehmenserfolg heute wichtiger als in der Vergangenheit. Hinzu kommt die Notwendigkeit, die IT an den aktuellen Trends der Gestaltung und Entwicklung komplexer IT-Systeme auszurichten und dabei auch neu aufkommende Technologien und Anwendungen im Sinne der Unternehmensziele zu nutzen. Verbunden mit diesen Aspekten – aber über sie hinausgehend – ist eine informationszentrische Sichtweise, die nicht die IT als Organisation oder Anwendungen und Technologien in den Mittelpunkt stellt, sondern den Vermögensgegenstand Information. Die unterschiedlichen Sichten und die mit ihnen zusammenhängenden Fragestellungen sollen Gegenstand des Workshops sein.

Die IT-Governance definiert in diesem Zusammenhang Strukturen, Prozesse und weitere Mechanismen, die darauf abzielen, Unternehmensziele und IT-Ziele abzustimmen und die konkreteren Ziele – Sicherung des Wertbeitrags sowie Planung, Steuerung und Kontrolle von IT-Compliance, IT-Risiken und IT-Sicherheit – zu realisieren. In diesem Umfeld führen auch die Planung, Steuerung und Kontrolle der Systemlandschaft und das Zusammenspiel mit der fachlichen Seite zu neuen Herausforderungen. So stellen Trends wie „Mobile“, „Cloud“, „Internet of Everything“ etc. Anforderungen sowohl in technischer Hinsicht als auch mit Blick auf Governance und Management der Unternehmens-IT. Unter dem Begriff „Information Governance“ rücken darüber hinaus Themen wie Informations- und Datenqualität und Master Data Management aber auch Information Privacy und ethische Fragen ins Blickfeld. Im Workshop sollen vor diesem Hintergrund auch grundsätzlichere Fragen beleuchtet und diskutiert werden.

Um den verschiedenen Herausforderungen ganzheitlich begegnen zu können, ist die Entwicklung zeitgemäßer Governance-Ansätze und Methoden, die Anwendung aktueller Modelle und Best-Practice-Frameworks sowie die Entwicklung neuer Ansätze erforderlich. Eine gewisse Renaissance scheinen seit einigen Jahren

¹ Hochschule der Deutschen Bundesbank, Schloss, 57627 Hachenburg, matthias.goeken@bundesbank.de

² Hochschule für Technik und Wirtschaft des Saarlandes, Fakultät für Wirtschaftswissenschaften, Waldhausweg 14, 66123 Saarbrücken, daniel.abawi@htwsaar.de

³ Hochschule für Technik und Wirtschaft des Saarlandes, Fakultät für Ingenieurwissenschaften, Goebenstraße 40, 66117 Saarbrücken, andre.miede@htwsaar.de

„Managementsysteme“ zu erfahren, wobei aktuell noch Managementsysteme für Teilaufgaben im Fokus stehen (Risikomanagementsysteme, Informationssicherheitsmanagementsysteme) und ganzheitliche Ansätze für die IT, die die verschiedenen Themen integriert betrachten, weniger im Blick von Wissenschaft und Praxis sind. Ziele, Herausforderungen und Anforderungen an Managementsysteme in der IT sowie Ansätze für ihre Entwicklung und Implementierung sind daher ein interessantes Themengebiet des Workshops.

Im Workshop werden aktuelle Forschungsarbeiten einschließlich „Work in Progress“ zum Themenkomplex IT-Governance, Information Governance sowie dem strategischen Informationsmanagement vorgestellt und diskutiert. Der Workshop gibt auch Teilnehmern aus der Praxis die Möglichkeit, von ihren Erfahrungen zu berichten.

2 Programmkomitee

Wir bedanken uns recht herzlich bei den Mitgliedern des Programmkomitees für ihre konstruktiven Beiträge zum Gelingen dieses Workshops:

- Prof. Dr. Carsten Felden, Technische Universität Bergakademie Freiberg
- Prof. Dr. Dr. h.c. Volker Herwig, Fachhochschule Erfurt
- Christoph Hochreiner, Technische Universität Wien
- Prof. Dr. Michael Klotz, Hochschule Stralsund
- Prof. Dr.-Ing. Arne Koschel, Hochschule Hannover
- Dr.-Ing. Ulrich Lampe, Technische Universität Darmstadt
- Dr. Marc Lohmann, msgGillardon AG
- Prof. Dr. Giselher Pankratz, Hochschule der Deutschen Bundesbank
- Prof. Dr. Jens Pöppelbuß, Universität Bremen
- Marco Tröbs, BearingPoint GmbH
- Prof. Dr. Thomas Widjaja, Universität Passau
- Prof. Dr. Till Winkler, Copenhagen Business School
- Prof. Dr. Alfred Zimmermann, Hochschule Reutlingen

Entwicklung eines Werkzeugs zur Standortbestimmung von IT-Abteilungen basierend auf COBIT 5

Nicola Bigler, Konrad Walser¹

Abstract: (IT-)Berater stehen immer wieder vor derselben Herausforderung, nämlich, effizient und effektiv herauszufinden, wo beim Kunden (und seiner IT) Verbesserungspotenziale vorhanden ist. Dies führte im folgenden Beitrag für den IT-Bereich zur systematischen auf dem IT-Governance-Framework COBIT beruhenden Entwicklung eines Werkzeugs und einer Methode zur Analyse von IT-Abteilungen. Ziel ist es, eine effiziente und effektive Untersuchung von IT-Abteilungen nachvollziehbar zu gestalten und die Schlussfolgerungen der Analyse (visuell) nachvollziehbar aufzeigen zu können. Zur Entwicklung wurde ein iteratives Forschungsvorgehen mit vier Durchläufen bei vier verschiedenen Firmen durchgeführt und aus wissenschaftstheoretischer Sicht wurde auf die folgenden Ansätze gesetzt: Method Engineering, Design Science sowie Action Research. Einerseits wurde eine (Beratungs-)Methode und ein Werkzeug zur Untersuchung und Analyse von IT-Abteilungen entwickelt. Letzteres lässt sich kurz als Excel-basiertes Werkzeug zur Reifegradbestimmung von COBIT-Prozesszielen bei Unternehmen charakterisieren, das über eine einfache Visualisierung u.a. auch zur Vermittlung der Resultate der IT-Analysen zuhanden der Betroffenen beiträgt.

Keywords: IT-Standortbestimmung, COBIT 5, Consulting, Method Engineering, Design Science, Werkzeug, Beratungsmethode

1 Einleitung

1.1 Problemstellung

Der vorliegende Beitrag ist klar aus der Perspektive eines IT-Beraters entwickelt und basiert auf [Bi14] (nicht veröffentlichte Master Thesis in Wirtschaftsinformatik). IT-Berater stehen typischerweise vor dem Problem, baldmöglichst wissen zu müssen, wo beim Kunden das Problem in der IT liegt, das vom IT-Berater unterstützt zu lösen ist. Gleichzeitig muss das vom Kunden geschilderte Problem methodisch transparent und systematisch verifiziert oder bestätigt werden. Verschiedene Ursachen können eine IT-Beratung initiieren. Sicherlich geht es vielfach um einen Optimierungswunsch des Kunden bezüglich IT. Grundsätzliche Gründe für IT-Beratungen können ferner in den folgenden Bereichen liegen: Neuaufbau der IT oder von Bereichen daraus, Reorganisation, Fusionen oder Unternehmenskäufe, etc. Hier geht es um die Reorganisation und im Falle einer Fusion von IT-Abteilungen um das Schaffen von Grundlagen

¹ Berner Fachhochschule, E-Government-Institut, Brückenstrasse 73, CH-3005 Bern, nicola.bigler@bigler.ch; konrad.walser@bfh.ch

zur Due Diligence sowie Positionierungen der IT in Relation zum Geschäft. Überall zählt eine konsequent transparente und effiziente sowie effektive Analyse zum Wichtigsten, um das weitere Vorgehen zu erleichtern. Zudem sind IT-Berater wie bereits erwähnt mit dem ökonomischen Problem des Ressourceneinsatzes für die Analyse im Vorfeld der beratenden Tätigkeit konfrontiert sowie dem daraus ableitbaren Einsatz im Beratungsprojekt. Dies kann je nach Größe des Beraters oder der zu untersuchenden Firma unterschiedlich ausgeprägt sein. Die Genauigkeit der Analyse impliziert danach auch die Genauigkeit des Ressourceneinsatzes für den eigentlichen Beratungsteil. Damit ist das Interesse an Werkzeugen und Methoden gegeben, welche Effizienz- und Effektivitätsgewinne in der Analysephase des Beratungszyklus ermöglichen.

1.2 Zielsetzung

Die Ziele des vorliegenden Beitrags lauten wie folgt:

- Eruiieren, wie und womit (Methode) die IT eines Unternehmens mit Hilfe von bestehenden Rahmenwerken ganzheitlich auf Verbesserungspotenzial hin analysiert und daraus effizient und effektiv Verbesserungsvorschläge abgeleitet werden können.
- Entwicklung einer entsprechenden Vorgehensmethode zur Analyse und Standortbestimmung einer IT.
- An Frameworks ausgerichtete Suche nach einem Werkzeug zur IT-Analyse.
- Anforderung an Resultat der Analyse: Problembereiche einer IT werden effektiv und effizient identifiziert.

2 Ausführliche Darstellung des Methodischen Vorgehens

Im vorliegenden Beitrag wird im Wesentlichen auf drei methodischen Grundlagen aufgesetzt: Dem Design-Science-Ansatz [HMP04], dem Method-Engineering-Ansatz [Ma95] und dem Action-Design-Research-Ansatz ([Ii07]; [BR92]; [Se11]). Aus Platzgründen wird davon ausgegangen, dass die Ansätze bekannt sind, weshalb in der Folge nur in sehr summarischer Form auf die einzelnen forschungsmethodischen Grundlagen für das Forschungsvorgehen eingegangen wird.

2.1 Design Science Ansatz nach Hevner et al. (2004)

Der Design-Science-Ansatz wurde von [HMP04] begründet.

Richtlinie nach [HMP04]	Beschreibung der Richtlinie aus Sicht des vorliegenden Beitrags
1. Design as an Artefact	Es wird ein Werkzeug und eine Vorgehensmethode entwickelt basierend auf einem Vorgehensmodell und einem Analyseinstrument basierend auf einem IT-Governance-Rahmenwerk.
2. Problem Relevance	Es wird ein reales Praxisproblem im IT-Beratungsumfeld gelöst.
3. Design Evaluation	Die Vorgehensweise folgt einer bestimmten Methode (zur Entwicklung eines Beratungsvorgehens und eines Analysewerkzeugs).
4. Research Contributions	Die Resultate münden in ein problemlösendes Artefakt. Durch die Analyse werden Forschungslücken sichtbar.
5. Research Rigour	Es wird nach anerkannten wissenschaftlichen Forschungsmethoden vorgegangen.
6. Design as a Search Process	Durch die iterative Entwicklung wird ein effizient und effektiv einsetzbares Artefakt erzeugt.
7. Communication of Research	Die Artefakte oder Ergebnisse sollen in der Praxis Anwendung finden. Es findet im Rahmen dieser Beitragspublikation eine Kommunikation der Resultate zur Diskussion unter Experten statt.

Tabelle 1: Umsetzungs-Richtlinien zur Erreichung von Rigour in problemlösungsorientierten Forschungsansätzen (nach [HMP04]) und deren Konkretisierung für das vorliegende Vorhaben.

Die Autoren positionierten problemlösungsorientierte Ansätze theoriebasierten Ansätzen gegenüber und thematisieren die Frage, wie die das Rigour-Konzept theoriebasierter Ansätze auch auf problemlösungsorientierte Ansätze übertragen werden kann. [HMP04] haben dazu sieben Richtlinien definiert, die für den vorliegenden Beitrag wie folgt konkretisiert werden können (vgl. dazu Tabelle 1 oben).

2.2 Method Engineering Process

Der Ansatz des Method Engineering Process präsentiert ein Vorgehen nach [Ma95] um eine Methode zu kreieren. Übertragen auf das vorliegende Vorhaben präsentiert die Tabelle 2 die Prozessschritte und deren Ausprägungen für den vorliegenden Beitrag ausgehend vom Method Engineering Process nach [Ma95].

#	Prozessschritte nach [Ma95]	Beschreibung
1	Document motivations	Ausgangslage und Problemstellung liefern den Input und den Scope.
2	Search for existing methods	Literatursuche nach bestehenden Frameworks und Methoden der Strukturierung.
3	Adopt existing method	Adaptieren von vorhandenen Methoden und Werkzeugen zur Durchführung einer Standortbestimmung der IT.

#	Prozessschritte nach [Ma95]	Beschreibung
4	Tailor existing method	Tailoring von vorhandenen Frameworks auf die Bedürfnisse zur Analyse der IT.
5	Develop new method	Entwicklung einer neuen Methode und eines neuen Werkzeugs zur ganzheitlichen Analyse der IT.
6	Design method application techniques	Entwerfen einer Vorgehensweise, um die identifizierten Frameworks zu konsolidieren und zur Analyse zu verwenden: Beispielsweise Entwurf eines Fragenkatalogs oder Prozessablaufs.
7	Test candidate design elements	Testen der erarbeiteten Resultate mittels Fallbeispielen oder durch Action Research.
8	Iteratively refine method design	Identifizieren von Schwachstellen und Einleiten von Verbesserungen auf Basis der Tests.

Tabelle 2: Mapping des Method Engineering Process auf die vorliegende Forschungsarbeit.

2.3 Action-Design-Research-Ansatz

Der Action-Design-Research-Ansatz nach [Ii07] und [Se11] ist zu verstehen als Ansatz eines Forschungsvorgehens im realen Praxisumfeld (in dem der Forschende selber tätig ist). Dies war im vorliegenden Beitrag deshalb der Fall, weil aus Beratersicht zu beratende und beratene Unternehmen untersucht wurden (dies betrifft den ersten der beiden Autoren dieses Beitrags). Die Schlussfolgerungen ausgehend vom angewendeten Ansatz lauten in tabellarischer Form wie folgt (Vgl. Tabelle 3).

Phase 1	Beschreibung des Problems als Basis für die weiteren Arbeiten. Dies wird durch die Problemstellung definiert.
Phase 2	Iterativer Prozess (Iterationen I bis IV) zur Entwicklung einer Lösung (Werkzeug und Methode zu dessen Einsatz). Durch eine nahe Zusammenarbeit mit der Geschäftsseite werden Lösung und Problem ständig neu evaluiert und angepasst.
Phase 3	Laufende Reflexion der geleisteten Arbeit sowie Überprüfung dazu, ob die verwendete Theorie passend ist.
Phase 4	Dokumentieren der Ergebnisse.

Tabelle 3: Phasen der Action-Research-Design-Methode.

2.4 Zirkuläres Forschungsvorgehen und entsprechende Iterationen

Die Forschung basiert auf dem zirkulären und qualitativen Ansatz der Forschung nach [Fl07]. Dies hatte vier Forschungsiterationen zur Folge. Es wurden vier KMU im Raum Bern in die Untersuchung miteinbezogen. Die Art der Iterationen im iterativen Vorgehen nach dem Method-Engineering-Ansatz und dem zirkulären Forschungsansatz lauten wie im Folgenden dargestellt.

- Iteration I basierend auf dem SAMM-Ansatz nach [Lu03] und [Lu04], der nach der ersten Iteration verworfen wurde: Die Erstdurchführung erfolgte aufgrund der generischen Methodenableitung auf einem realen Kunden-Case mit Kunde A. Aus Sicherheitsgründen, und weil es die erste Durchführung war, übernahmen die Berater Rollen aus dem betroffenen Unternehmen und spielten das Szenario im Sinne der Kundenrolle durch. Gründe für das Verwerfen des SAMM-Ansatzes: Begrifflichkeiten als Problem für Befragungen ohne Interviewer, nicht vollständige Abdeckung der IT-relevanten Aufgaben und Prozesse sowie Fachbereiche.
- Iteration II basierend auf COBIT 5 ([ISA12a]; [ISA12b]): Zweitedurchführung mit einer verbesserten Methode ausgehend von den Rückschlüssen aus Iteration I und dem adaptierten Werkzeug, basierend auf einem realen Kunden-Case mit Kunde B (Weiterbildungsinstitution, acht IT-Mitarbeiter)
- Iteration III basierend auf COBIT 5: Dritte Durchführung mit einer nochmals verbesserten Methode ausgehend von den Rückschlüssen aus Iteration II und dem adaptierten Werkzeug, basierend auf einem realen Kunden-Case mit Kunde C (Architektur- und Bauunternehmung mit vier Standorten, 60 Clients, 130 Mitarbeiter und 150 Stellenprozent IT-Mitarbeiter, Teil-IT-Outsourcing).
- Iteration IV basierend auf COBIT 5: Vierte Durchführung mit Verfeinerungen des erzielten Methodensets ausgehend von der Iteration III, basierend auf einem realen Kunden-Case mit Kunde D (ein Standort, 500 Mitarbeiter, 100 Stellenprozent IT-Mitarbeiter, kaum eigenes IT-Know-how, ca. 200 Workplaces und IT-Outsourcing).

2.5 In das Forschungsvorgehen einbezogene Rahmenwerke

Wie erwähnt resultierten zwei Lösungsansätze basierend auf Rahmenwerken:

- SAMM-Ansatz (Iteration I) zur Analyse der IT: Ansatz eines Maturitätsmodells, Soll- und Ist-Business-IT-Alignment-Status aufgrund des SAMM/Strategic Alignment Maturity Model. Dieser Ansatz musste aufgrund der oben angegebenen Gründe verworfen werden. Die Idee war, über Maturitätsmessungen zu GAP-Analysen bezüglich IT zu kommen und ausgehend davon Maßnahmen abzuleiten.
- COBIT-Ansatz (Iterationen II bis IV) zur Analyse der IT. Prozessmaturitätsmessung nach ISO/IEC 15504 – Process Maturity Model/Self Assessment Guide als Perspektive ([ISA13a], [ISA13b]). Dieser Ansatz erwies sich als brauchbar und bildete die Grundlage für das weitere Vorgehen und die entwickelten Methoden- und Werkzeug-Artefakte.

Ein weiteres Rahmenwerk, das in die Evaluation hätte miteinbezogen werden können, ist das ITIL-Rahmenwerk. Dieses wurde nicht weiter in Betracht gezogen, da es den Autoren wichtig war, zur Analyse der IT die Geschäftsperspektive einzunehmen. Dies war aus

Sicht COBIT ([ISA12a]; [ISA12b]) am ehesten gewährleistet, gerade auch wegen der Zielkaskade, über welche von Geschäfts- auf IT-Ziele rückgeschlossen werden kann. Dies ist wie erwünscht einer strukturierten Business-IT-Alignment-Methode gleichzusetzen. Zudem, dies ist ein weiterer Grund, der zur Ablehnung des ITIL-Einsatzes führte, ist das ITIL-Rahmenwerk aus Sicht des Geschäfts eher auf das Wie der IT-Implementierung ausgerichtet und weniger auf das Was. Wie erwähnt überzeugte die „hohe Flughöhe“ von COBIT 5 und dessen Ausrichtung auf das Was der IT(-Implementierung). Zur Evaluation von IT-Abteilungen existiert im COBIT-Umfeld außerdem das „Werkzeug“ des COBIT Quick Assessment Guide (vgl. [ISA13a]; [ISA13b]). Dieses Werkzeug ermöglicht, gekoppelt mit COBIT 5 ([ISA12a]; [ISA12b]), einen Überblick über die gesamte IT-Abteilung und all deren Facetten zu gewinnen. Basierend darauf ist (auf Grundlage etwa von in COBIT existierenden RACI-Charts) eine rollenbasierte Befragung möglich. Anerkanntermaßen bietet COBIT überdies Referenzen auch zu anderen Rahmenwerken, was den Vorteil hat, dass für spezifische Aufgaben auf spezifischere Rahmenwerke zurückgegriffen werden kann, sofern die Korrelation unmittelbar klar ist. Beispielsweise ist an verschiedenen Orten in COBIT 5 ein „Drill-Down auf ITIL“ möglich, um aus Sicht des Geschäfts entsprechende Wie-Fragen, z.B. bezüglich eines bestimmten Servicemanagement-Prozesses, zu beantworten. Im Weiteren existieren hierzu Mappings (Vgl. für ein Mapping der alten Versionen ITIL V3 und COBIT 4.1 [ISA07b]).

3 Resultate zu Vorgehensmethode und Werkzeug zur Analyse

3.1 Resultierende Vorgehensmethode zum Einsatz des Analysewerkzeuges auf Excel-Basis

Ausgehend von den obigen methodischen Überlegungen resultierte das folgende initiale Vorgehen für die Untersuchung einer IT-Abteilung oder eines IT-Bereichs im Hinblick auf Verbesserungspotenziale auf der Basis von COBIT 5 (Vorgehen und Werkzeug wurden im Verlauf der geschilderten Iterationen sukzessive und systematisch angepasst):

1. Auslöser (für Beratungsmandat) in IT-Abteilung oder IT-Bereich eruieren: Effizienzprobleme, mangelnde Transparenz der IT(-Tätigkeiten), etc.
2. IT-Ziele ableiten (ausgehend von der Zielkaskade von COBIT 5): Unternehmensziele führen wie erwähnt zu IT-Zielen.
3. IT-Prozesse identifizieren: IT-Ziele führen zu IT-Prozessen, welche möglicherweise verbesserungsbedürftig sind; Projekte können IT-Prozesse ergänzen.
4. Messung durchführen: Rollenbezogenen Fragebogen beantworten lassen.
5. Analyse und Ergebnisse: Problembereiche identifizieren sowie weiterführende Analysen durchführen.

Ausgehend von diesem initialen Vorgehen veränderte sich das Vorgehen und die Analysemethode basierend auf den in der zweiten Iteration eruierten Verbesserungspotenziale wie in Abbildung 1 präsentiert.

Anpassung	Beschreibung	Bezug
Einführungs-/Abschluss-workshop	Zum besseren Verständnis der Vorgehensmethodik sollte diese in einem Start-Workshop erläutert werden.	Einführung
Fragenkatalog	Am Fragenkatalog werden folgende Änderungen durchgeführt: <ul style="list-style-type: none"> • Fachbegriffe reduzieren oder erklären (im Fragetext) • Fragen aufteilen, falls mehrere Attribute abgefragt werden. Dabei kann nicht mehr berücksichtigt werden, dass nur eine Frage pro Prozessziel abgefragt wird. • Der Fragekatalog wird erweitert, damit dem Befragten der Fragenkontext klar wird. 	Doppeldeutigkeit Fachbegriffe Kontext
Auswertung	Um die Präzision der Auswertung zu erhöhen, wird diese so angepasst, dass damit nicht mehr auf IT-Prozessebene sondern auf IT-Prozessziel-Ebene gearbeitet wird. Mit dieser Maßnahme können Probleme besser eingegrenzt werden.	Struktur
IT-Ziele	In zukünftigen Zyklen wird die Möglichkeit vorgesehen, IT-Prozesse direkt auszuwählen ohne die Zielkaskade berücksichtigen zu müssen. Damit können Unternehmen nur einen bestimmten Teilbereich untersuchen lassen. Zusätzlich können so auch Analysen durchgeführt werden, wenn keine Unternehmensstrategie oder IT-Strategie vorhanden ist.	Ziele
Messung	Um die Rollenproblematik zu kompensieren, werden die vorgegebenen Rollen von COBIT zusammengefasst.	Rollen
IT-Prozesse	Es soll ein fixer Satz an Prozessen ausgewählt werden, um den Umfang einzugrenzen. Somit werden einerseits die Praxispartner entlastet, andererseits kann die Qualität der Vorgehensmethodik erhöht werden, da sich eine kleinere Anzahl von Prozessen in den Zyklen schneller weiterentwickelt. Ein weiterer Vorteil ist die gewonnene Vergleichbarkeit zwischen den Praxispartnern.	Umfang

Abbildung 1: Anpassungen nach Iteration I basierend auf COBIT ([ISA12a]; [ISA12b]).

Dies führt zum folgenden angepassten Vorgehen für die Iteration II:

1. Durchführung Start-Workshop mit Klärung Ziele, Modell und Roadmap.
2. IT-Ziele ableiten: Unternehmensziele führen zu IT-Zielen.

3. IT-Prozesse identifizieren: IT-Ziele führen zu IT-Prozessen, Projekte können IT-Prozesse ergänzen.
4. Messung durchführen: Rollenbezogene Messungen basierend auf einem Fragebogen.
5. Analyse und Ergebnisse: Problembereiche identifizieren und weiterführende Analysen durchführen; dazu Abschlussworkshop durchführen.
6. Maßnahmendefinition: Definition weiterer Schritte.

Die Iteration II führte zu den folgenden Anpassungen am Vorgehen und an der Analyse-methode.

Anpassung	Beschreibung	Bezug
Kennzahl Streuung	Die Kennzahl für die Streuung wird überprüft.	Statistische Auswertung
Grafische Auswertung	Die grafische Auswertung soll so angepasst werden, dass für jede Kennzahl die gleiche Skala verwendet werden kann.	Statistische Auswertung
Zeitmanagement	Der Zeitaufwand, insbesondere für den Abschlussworkshop wird den zukünftigen Praxispartnern neu kommuniziert.	Zeitmanagement
Naming	Für die Excel-(Pivot-)Tabelle müssen aussagekräftige Spaltenbeschriftungen verwendet werden.	Naming

Abbildung 2: Anpassungen nach Iteration II basierend auf COBIT 5 ([ISA12a]; [ISA12b]).

Änderungen erfolgten gemäß der obigen Auswertung nicht im eigentlichen ablauforientierten Vorgehen, daher bleibt dies gleich wie in Iteration II dargestellt, sondern vor allem in Hinsicht auf die statistischen Verfahren, welche im Excel-Werkzeug eingesetzt werden.

Nach der Iteration III resultierte nochmals eine Anpassung der Methode zur Evaluierung einer IT basierend auf COBIT:

1. Durchführung Einführungs-Workshop: Dies erfolgte neu mit Klärung Ziele, Erklärung Vorgehen, Roadmap-Definition und Rollendefinition.
2. IT-Ziele ableiten: Unternehmensziele führen zu IT-Zielen; dies basiert falls nötig auf der Workshop-basierten Dokumentation von IT-Zielen, beispielsweise wenn keine Unternehmensziele vorhanden sind.
3. IT-Prozesse identifizieren: IT-Ziele führen zu IT-Prozessen, Projekte können IT-Prozesse ergänzen. Zur Prozessergänzung können auch Workshops durchgeführt werden.

4. Messung durchführen: Rollenbezogene Messungen basierend auf einem Fragebogen. Für den Fragebogen sollten Antwortzeiten pro Frage von nicht länger als 30 bis 60 Sekunden anvisiert werden.
5. Analyse und Ergebnisse: Problembereiche identifizieren und weiterführende Analysen durchführen oder beschließen; dazu Abschlussworkshop durchführen, der nicht mehr als 2 Stunden dauern sollte.
6. Maßnahmendefinition: Definition weiterer Schritte (zusammen mit dem IT-Beratungskunden).

3.2 Resultierendes Excel-Analysewerkzeug basierend auf COBIT 5

Im Folgenden wird ausgehend von den drei auf COBIT basierenden Untersuchungs-Iterationen das Werkzeugset in seiner Endform skizziert. Eine erweiterte Präsentation ist aus Platzgründen nicht möglich. Das Werkzeugset kann in deutscher Sprache bei Interesse bei den Autoren eingefordert werden. Kenntnisse zur Struktur von COBIT und zu seiner Zielkaskade werden vorausgesetzt und hier nicht weiter ausgeführt.

Die Grundlage für die COBIT-basierte Analyse bildet das Process Assessment Model PAM von COBIT und darauf aufbauend der COBIT Self Assessment Guide ([ISA13a], [ISA13b]). Dieser sieht vier Stufen oder Levels von COBIT-Prozessen vor. N: Not achieved (0 bis 15% Erreichungsgrad), P: Partially achieved (15 bis 50% Erreichungsgrad). L: Largely achieved (50 bis 85% Erreichungsgrad). F: Fully achieved (85 bis 100% Erreichungsgrad). Diese vier Stufen werden auch zur Beantwortung von Fragen in einem entsprechenden Fragebogen aufgeführt. Zusätzlich wird als Antwortalternative „Keine Angabe“ angeführt.

Befragt wurden in den Unternehmen Rollen in Analogie zum Rollenmodell von COBIT 4.1 übernommen und angepasst. [ISA07a]. Das Rollenmodell von COBIT 5 ([ISA12a]; [ISA12b]) erwies sich für KMU-Befragungen als zu umfangreich und zu komplex. Es wurde deshalb als Grundlage für die Befragung verworfen. Ausgehend von den abgefragten Prozessen wurde für die Auswertung von je bis zu einem halben Dutzend Mitarbeiter eine Auswertungsmatrix – basierend auf den eingebbaren Werten wie oben dargestellt – erstellt. Im Laufe der Untersuchung wurde das Prozessrahmenwerk dahingehend reduziert, dass nur ein bestimmtes für KMU geeignetes Prozess-Set aus COBIT 5 untersucht wurde, um die Arbeit im Rahmen der gegebenen Ressourcen zu halten. Die Auflistung des dadurch resultierenden Prozesssets erfolgt weiter unten.

Personen	P1	P2	Mittelwert
Frage-ID			
APO08.2.1.3	2	2	2
APO08.4.1.3	3	4	3.5
Mittelwert	2.5	3	2.75

Prozessziel APO08 pro Person
Gesamtprozessziel APO08 Bewertung

Abbildung 3: Berechnungsbeispiel.

Der rote Bereich enthält die Fragenbewertung pro Person. Die Formel dazu lautet wie folgt.

$$\text{Fragebewertung} = \frac{\sum(\text{Antworten Person } x)}{\text{Anz. Antworten}}$$

Der grüne Bereich enthält die Prozesszielbewertung aller Personen. Die Formel dazu lautet wie folgt.

$$\text{Prozessziel pro Person} = \frac{\sum(\text{Antworten Prozessziel } x)}{\text{Anz. Antworten}}$$

Der blaue Bereich enthält die Gesamtprozesszielbewertung und berechnet sich aus dem Mittelwert der Prozesszielbewertung. Die Formel dazu lautet wie folgt.

$$\text{Gesamtprozesszielbewertung} = \frac{\sum(\text{Prozesszielbewertung})}{\text{Anz. Prozesszielbewertungen}}$$

Für die Prozessbewertung wird der Mittelwert aller Gesamtprozessziele herangezogen. Diese Formel lautet wie folgt.

$$\text{Mittelwert Gesamtprozessziele} = \frac{\sum(\text{Gesamtprozessziele})}{\text{Anz. Prozessziele}}$$

Mit der Excel-Auswertung entsteht nach dieser Vorgehensweise eine Übersicht über alle Prozesse, inkl. deren einzelner Prozessziele, wie dies in der folgenden Abbildung für die Prozesse APO01 und APO08 ersichtlich gemacht wird.

		Mittelwert von	Mittelwert von	Mittelwert von
3	Prozess	Ergebnisse P1	Ergebnisse P2	Konsolidiert
4	⊖ APO01	1.43	1.90	1.82
5	1	1.43	1.89	1.81
6	2		2.00	2.00
7	⊖ APO08	2.44	3.20	2.61
8	1	2.00	2.00	2.00
9	2	2.60	4.00	2.80
10	3	2.50	3.00	2.75

Abbildung 4: Beispiel Excel-Auswertung als Pivot-Tabelle.

Bis anhin mussten IT-Ziele aus Unternehmenszielen abgeleitet werden. Im Weiteren wurde die Excel-Methodik so erweitert, dass auch direkt IT-Prozesse ausgewählt werden konnten, ohne dass Unternehmensziele konkretisiert wurden. Ebenso wurde im Weiteren auf einem festen und reduzierten Set an COBIT-Prozessen aufgesetzt, die typischerweise für KMU von Relevanz sind. Dies umfasste die Prozesse EDM04 – Sicherstellen Ressourcenoptimierung, APO07 – Managen des Personals, APO08 – Managen von Beziehungen, APO13 – Managen der Sicherheit, BAI02 – Managen der Definition von Anforderungen, BI08 – Managen von Wissen, DSS02 – Managen von Serviceanfragen und Störungen, DSS05 – Managen von Sicherheitsservices sowie MEA03 – Überwachen, Evaluieren und Beurteilen der Compliance mit externen Anforderungen.

Grenzwert	Beschreibung
Gesamtbewertung IT-Prozess < 3 und Bewertung IT-Prozessziel < 3	Diese Grenzwerte leiten sich aus dem COBIT self assessment guide ab ([ISA13b], S. 11 f; [ISA13a]). Damit ein IT-Prozess den Fähigkeitsgrad Level 1 erreicht, muss er seinen Zweck zu 50%-100% erreichen. Somit werden IT-Prozesse und IT-Prozessziele unter der Bewertung 3 betrachtet.
Streuung > = 2	Die Streuung ist willkürlich gewählt. Eine Streuung von mehr als zwei Bewertungspunkten entspricht jedoch mehr als 50% auf der möglichen Skala. Diese Abweichung kann zur Identifikation von möglichen Verbesserungen führen.
Minimum <= 1	Die Bewertung mit einer 1 entspricht dem Erfüllungsgrad von 0% - 15%. Damit wird das Prozessziel aus Sicht einer befragten Person nur sehr schlecht unterstützt. Möglicherweise kommt dieser Wert durch Missverständnisse zustande, die es zu klären gilt.

Tabelle 4: Grenzwerte zur Analyse Iteration II.

Die Methode zur Eingrenzung der Prozesse kann aus Platzgründen nicht weiter erläutert werden. Ausgehend vom Action-Design-Research-Ansatz kann davon ausgegangen werden, dass sie schwerpunktmäßig auf Erfahrungen der beteiligten Personen (beraterseitiges Umfeld des einen der beiden Autoren) beruhte. Ausgehend von den Ergebnissen der Auswertungen der Pivot-Tabellen werden zur besseren Veranschaulichung und Visualisierung Powerpoint-Visualisierungen zur Verfügung gestellt, welche in den oben thematisierten Workshops im Vorgehen diskutiert werden können und woraus

dann Maßnahmen oder Beratungsleistungen zuhanden der IT-Berater abgeleitet werden können. In der Abbildung 5 werden einerseits die grafische Veranschaulichung der Ziel-Mittelwerte aus der Iteration III sowie andererseits tiefste und höchste Bewertungen und Kommentare dargestellt. Diese grafischen Veranschaulichungen stellen eine zentrale Grundlage für Workshop-orientierte Verbesserungsvorschläge und damit die Generierung möglicher Beratungsleistungen dar.

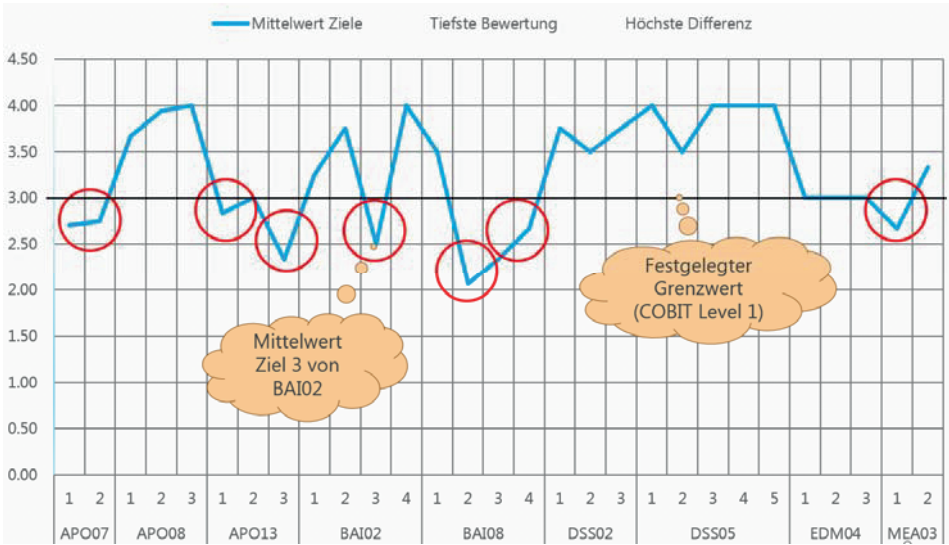


Abbildung 5: Auswertung Beispiel Mittelwert aus Iteration II der Untersuchung.

Aus der Iteration II resultierte auch die Ergänzung des Analyse-Excels um eine Kennzahl ‚Streuung der Ergebnisse‘, um auch diesen Aspekt in die Analyse und die Synthese in den Workshops übernehmen zu können. Die Breite der Auswertungen und der Excel-Files nahm damit um einige Spalten zu. Die Spaltenbezeichnungen des Excels zur Analyse lauten wie folgt: Prozess – COBIT-IT-Prozess und dazugehörige Prozessziele (als eigentliche Messgrößen); P1 – P4 – Mittelwerte der befragten Personen (hier Person 1 bis Person 4); Durchschnitt – Durchschnitt zum IT-Prozess und dessen Prozessziele; Minimum – Kleiner Wert/Angabe zu einer Frage zum IT-Prozess und dessen Prozesszielen; Spannweite – Größte Differenz zwischen kleinem und größtem Wert einer Frage zu einem IT-Prozess(ziel); Standardabweichung – größte Standardabweichung der Angaben zu einem IT-Prozessziel; Variationskoeffizient – größter Variationskoeffizient innerhalb einer Frage zu einem IT-Prozessziel. Die Standardabweichung dient insbesondere dazu herauszufinden, ob sich verschiedene Personen innerhalb einer Zielbewertung nicht einig sind. [Qu08] verwendet den Variationskoeffizient zum Vergleich der Streuung von Merkmalen, die einen stark unterschiedlichen Mittelwert aufweisen. Die Formel dazu lautet wie unten dargestellt.

$$\text{Variationskoeffizient} = \frac{\text{Standardabweichung}}{\text{Mittelwert}}$$

Der Variationskoeffizient ist somit robuster gegenüber unterschiedlichen Mittelwerten. Zusätzlich ist mit ihm ein Vergleich der Streuung zwischen zwei verschiedenen Fragen möglich, falls dies künftig eine Anforderung sein sollte.

4 Zusammenfassung und Ausblick

Die Schlussfolgerungen ausgehend von der Forschung in drei Iterationen lauten wie folgt. Erwünscht ist ein Pragmatismus der Umsetzung. Sowohl die Vorgehensmethode wie auch das Analysewerkzeug müssen praxistauglich sein. Dies wurde mit dem vorliegenden Vorhaben erreicht. Das Vorgehen lässt sich in sechs Schritten darstellen: 1. Durchführung Einführungs-Workshop mit Klärung Ziele, Erklärung Vorgehen, Roadmap-Definition und Rollendefinition. 2. IT-Ziele ableiten: Unternehmensziele führen zu IT-Zielen; dies basiert falls nötig auf der Workshop-basierten Dokumentation von IT-Zielen, beispielsweise wenn keine Unternehmensziele vorhanden sind. 3. IT-Prozesse identifizieren: IT-Ziele führen zu IT-Prozessen, Projekte können IT-Prozesse ergänzen. Zur Prozessergänzung können auch Workshops durchgeführt werden. 4. Messung durchführen: Rollenbezogene Messungen basierend auf einem Fragebogen, für den Antwortzeiten von Fragen nicht länger als 30 bis 60 Sekunden betragen sollten. 5. Analyse und Ergebnisse: Problembereiche identifizieren und weiterführende Analysen durchführen oder beschließen; dazu Abschlussworkshop durchführen, der nicht mehr als 2 Stunden dauern sollte. 6. Maßnahmen Definition: Definition weiterer Schritte. Ferner wurde zu diesem Methodenvorschlag ein auf Excel basierendes Werkzeug entwickelt, welches die Analyse von bestimmten IT-Sachverhalten basierend auf COBIT 5 Prozessen und entsprechenden Prozesszielen gemäß einem Maturitätsmodell ermöglicht.

Das Set der zu analysierenden Prozesse wurde für KMU's reduziert, aber auch wegen der Unmöglichkeit, das Werkzeug in einem ersten Aufschlag gleich für alle Prozesse zu konkretisieren. Zudem wären KMU, welche schwerpunktmäßig für die Untersuchung eingesetzt wurden, mit dem ganzen COBIT Prozess-Satz überfordert gewesen. Das Werkzeug setzt auf dem COBIT Self Assessment Guide auf und nutzt das COBIT PAM/Process Assessment Model als Maturitätsmodell für Items und Prozesse. Das Werkzeug unterstützt die grafische Visualisierung der IT-Verbesserungspotenziale. Das Werkzeug wurde (noch) nicht für alle COBIT Prozesse umgesetzt, ist jedoch dafür vorgesehen. Weitere Schritte der grafischen Umsetzung sowie deren Automatisierung sind denkbar, ebenso wie die weitergehende Tiefenschärfung des Analysewerkzeugs durch weitere Parameter. Letzteres ist indes nur mit Verlusten in der Effizienz und ev. der Effektivität des Methodeneinsatzes möglich. Die Methode wurde in der Praxis getestet und es trifft ausgehend davon die Aussage zu, dass die Methode effiziente und effektive Resultate für den Problemlösungskontext erbringt. Damit wurden für die vorliegende Untersuchung alle erforderlichen Schritte des oben thematisierten Design-Science-Vorgehens erfüllt (vgl. Kapitel 2).

Literaturverzeichnis

- [BR92] Babüroglu, O. N., Ravn, I., Normative Action Research. *Organization Studies* (1992) 19/1, S. 19-34.
- [Bi14] Bigler, N. (2014): Entwicklung eines Werkzeugs zur Standortbestimmung von IT-Abteilungen basierend auf COBIT 5, Master Thesis, Berner Fachhochschule, Bern.
- [Fl07] Flick, U. (2009): *An Introduction to Qualitative Research*. London: Sage Publications Ltd.
- [HMP04] Hevner, A. R.; March, S. T.; Park, J.: Design Science in Information Systems Research. In: *MIS Quarterly* 28 (2004) 1/ March, S. 75-105.
- [ISA07a] ISACA (2007a): COBIT 4.1. ISACA, Rolling Meadows.
- [ISA07b] ISACA (2007b): Mapping of ITIL V3 with COBIT 4.1. ISACA, Rolling Meadows.
- [ISA12a] ISACA (2012a): COBIT 5 – A Business Framework for the Governance and Management of Enterprise IT, ISACA, Rolling Meadows.
- [ISA12b] ISACA (2012b): COBIT 5 – Enabling Processes. ISACA, Rolling Meadows.
- [ISA13a] ISACA (2013a): COBIT Process Assessment Model (PAM): Using COBIT 5, ISACA, Rolling Meadows.
- [ISA13b] ISACA (2013b): Self-assessment Guide: Using COBIT 5, ISACA, Rolling Meadows.
- [Ii07] Iivari, J. A.: Paradigmatic Analysis of Information Systems As a Design Science. In: *Scandinavian Journal of Information Systems* (2007), S. 39-64.
- [Lu03] Luftman, J.: Assessing IT/Business Alignment. In: *Information Systems Management* (2003), S. 9-15.
- [Lu04] Luftman, J. (2004): Assessing Business-IT Alignment Maturity. In: Grembergen, W. V. (Hrsg.): *Strategies for Information Technology Governance*, Hershey : Idea Group Publishing.
- [Ma95] Mayer, R. J. et al. (1995): *Information Integration for Concurrent Engineering (IICE) Compendium of Methods Report*. Ohio. Human Resource Directorate Logistics Research Division.
- [Qu08] Quatember, A. (2008): *Statistik ohne Angst vor Formeln – Band 2*. München.
- [Se11] Sein, M. K.; Henfridsson, O.; Puro, S.; Rossi, M.; Lindgren, R., Action Design Research. In: *MIS Quarterly* 35 (2011) März, S. 37-56.

Anwendung von Referenzmodellen der IT-Governance: Konzeptionelle Grundlage und Anwendungsszenarien

Matthias Goeken¹, Stefanie Looso²

Abstract: Zwar liegen mit Referenzmodellen und Standards Hilfestellungen für die Umsetzung einer IT-Governance in Unternehmen vor. Allerdings bleibt vielfach unklar, wie diese herangezogen werden und als Vorlage dienen können, um bspw. ein unternehmensspezifisches IT-Governancemodell zu entwickeln. In zwei Forschungsprojekten wurde untersucht, wie die Anwendung von Referenzmodellen der IT-Governance mittels ihrer semiformalen Abbildung unterstützt werden kann. Das dort entwickelte Artefakt ist Gegenstand dieses Beitrags und wird dem DSRM-Ansatz folgend [Pe07] anhand von drei Anwendungsszenarien demonstriert. Zuvor werden kurz Motivation und Zieldefinition sowie die Vorgehensweise bei der Abbildung von semiformalen Modellinhalten in einem semantischen Netz beschrieben.

Keywords: IT-Governance, Referenzmodelle, Best Practice, Anwendungsszenarien, COBIT, ITIL

1 Einleitung

Ein Bedarf nach methodischer Unterstützung beim Management und bei der Governance der IT ist in der Praxis erkennbar, zum Beispiel bei Planung, Steuerung und Kontrolle des Technologieeinsatzes, wenn es darum geht, einen Wertbeitrag aus dem IT-Einsatz zu realisieren [Ba15], Alignment zwischen Business und IT herzustellen [CH07], IT-Risiken und die Sicherheit der IT zu managen oder Compliance der IT und durch IT zu gewährleisten [TV07, DG15]. Zur Bereitstellung dieser Unterstützung haben sich „eine Reihe von Referenzmodellen etabliert, die einen gewissen normativen Charakter aufweisen und zum Teil durch ihre praktische Anwendung empirische Bestätigung erfahren haben.“ [Kr15, 601]. Hiermit angesprochen sind unter anderem Modelle wie COBIT, ITIL oder ISO 38500, die Best-Practice oder Standard sind und unter anderem dazu dienen, unternehmensspezifische Modelle zu entwickeln, Audits, Benchmarks und Assessments durchzuführen oder Reifegrade von IT-Abteilungen festzustellen [JG11].

Allerdings erfolgt die Entwicklung unternehmensspezifischer Modelle in der Regel eher unsystematisch und das, was man als *Anwendung eines Referenzmodells* bezeichnen kann, hat in der Praxis sehr unterschiedliche Ausprägungen: So fällt darunter zum Beispiel ein sogenanntes „Cherry Picking“, was die eher beliebig erscheinende selektive Verwendung von geeigneten Komponenten im Unternehmen bezeichnet. Andere Unternehmen hingegen erstellen zunächst ein eigenes Soll-Modell unter Zuhilfenahme eines Referenzmodells oder Standards und verwenden dieses später etwa als Vorlage

¹ Hochschule der Deutschen Bundesbank, Matthias.Goeken@bundesbank.de

² Frankfurt School of Finance and Management, Management Department, S.Looso@fs.de

oder auch zum Self Assessment. In diesen Fällen ist zu berücksichtigen, welche Teile eines Modells in das unternehmensspezifische Modell eingehen und inwieweit die Erstellung systematisch auf Basis des Standards erfolgt [Lo11].

Der vorliegende Beitrag beschäftigt sich mit eben dieser systematischen Anwendung von Best-Practice-Modellen. Es wird der Frage nachgegangen, wie eine systematische Anwendung von Best-Practice-Modellen aussehen kann und wie sie zu unterstützen ist. Hierfür wird das Artefakt TosI vorgestellt, welches unter anderem ein Ergebnis aus mehreren Forschungsprojekten (SemGoRiCo und TosI³) ist, in denen Referenzmodelle der IT-Governance in „k-infinity“, einer Graphdatenbanken, abgelegt wurden. Im Rahmen dieser Projekte wurden mit Konsortialpartnern mehrere konkrete Anwendungsszenarien erarbeitet und implementiert.

Bereits an anderer Stelle konnten die Verfasser verwandte Fragen und Vorarbeiten diskutieren [LGF16], u. a. wie eine semiformalen Abbildung/Wissensrepräsentation die Modellanwendung methodisch unterstützen kann. Konkret verfolgtes Gestaltungsziel war dabei die unternehmensspezifische Anpassung und Erweiterung von Referenzmodellen und Best Practices sowie ihre Repräsentation in einem Tool. In diesem Beitrag stehen nun Anwendungsszenarien im Vordergrund, die die dort präsentierten Grundlagenaspekte vertiefen und veranschaulichen. Er legt damit den Schwerpunkt auf die Demonstration der entwickelten Lösung (des IT-Artefakts) und dient weniger dazu, Motivation, Anforderungen und Entwurf darzustellen, weshalb diese hier nur kurz betrachtet werden (Abschnitt 2). Abschnitt 3, der Hauptteil dieses Beitrags, demonstriert die Anwendung des Artefakts, indem drei unterschiedliche Szenarien dargestellt werden. Abschnitt 4 fasst die Erkenntnisse kurz zusammen und gibt einen Ausblick.

Die Szenarien fokussieren Einsatzgebiete von Best Practices im Rahmen der IT-Governance und wurden im Verlauf der beiden Forschungsprojekte erarbeitet. So thematisiert die erste Anwendung den Einsatz von COBIT-Metriken als Erweiterung zu bereits vorhandenen Metriken des Unternehmenscontrollings. Im zweiten Anwendungsszenario werden Strukturen einer IT-Governance-Organisation (aufbauorganisatorische Aspekte) abgebildet. Hierbei waren in Interviews und mehreren Workshops Verantwortlichkeiten, Entscheidungsrechte und Rechenschaftspflichten erhoben und mithilfe des Modells sowie des Tools analysiert worden. Im dritten Anwendungsszenario werden schließlich prozessorientierte Aspekte der IT-Governance betrachtet. Hierbei wird ein IT-Prozess unter Nutzung des Artefakts abgebildet. Unter Verwendung von BI-Werkzeugen werden darüber hinaus relevante Prozesskennzahlen visualisiert und zur Planung und Steuerung des IT-Prozesses verwendet. Insofern adressieren die Szenarien drei relevante Gestaltungsaspekte der IT-Governance.

³ Beide Projekte, SemGoRiCo (Semantische Governance, Risk und Compliance, Projektnr: 160/08-22) und TosI (Toolgestützte Einführung und semantische Integration von Referenzmodellen, Prozessen und Systemlandschaften zur Unterstützung der IT-Governance, 333/12-23), wurden durch die Hessen Agentur gefördert.

2 Grundlagen, Motivation, Anforderungen und Entwurf

2.1 Governanceverständnis und grundlegende Anforderungen

Eine mittlerweile recht weit verbreitete Auffassung, was unter IT-Governance zu verstehen ist, findet sich bei [DG15]: „Definition and implementation of processes, structures, and relational mechanism that enable both business and IT people to execute their responsibilities in support of business/IT alignment and the creation of value from IT-enabled business investments.“ Weiter definieren sie: “Enterprise Governance of IT can be deployed using a mixture of various structures, processes, and relational mechanisms.” [DG15, 8 und 11].

Die wesentlichen Gestaltungsaspekte einer IT-Governance sind demnach *Organisationsstrukturen* (Zuordnung von Verantwortlichkeiten, Entscheidungsrechten und Rechenschaftspflichten und dadurch die Bildung von Stellen, Gremien etc.), *Prozesse* (verstanden als IT-Prozesse, die die Abfolge und den Ablauf von Aufgaben beschreiben) [siehe auch JG11] sowie *Beziehungsmechanismen* (Maßnahmen, die die Partizipation und Kollaboration der verschiedenen Anspruchsgruppen sicherstellen sollen). Neben diesen wurden in den Forschungsprojekten *Kennzahlen* als ein weiterer relevanter Gestaltungsaspekt identifiziert. So enthalten einige Modelle der IT-Governance u. a. Metriken, um vorhandene auf das Controlling ausgerichtete Kennzahlensysteme in Richtung IT-Governance zu erweitern [IS12]. Dies ermöglicht bspw. ein quantitatives Reporting über die Leistungserstellung in der IT oder deren Reifegrad sowie die Erfolgsauswertung von konkreten Maßnahmen oder deren Planung.

Aus den genannten vier Gestaltungsaspekten resultieren bereits Anforderungen für die im Rahmen der Projekte erstellende Lösung. So sollen der gewählte Lösungsansatz, der Entwurf und die Toolimplementierung sowohl die strukturellen und die prozessualen Aspekte einer IT-Governance adressieren als auch Kennzahlen enthalten. Die Beziehungsmechanismen werden als personenbezogene und „weiche“ Aspekte nicht tiefergehend thematisiert, da sich bereits früh zeigte, dass ein modellierungsorientierter Ansatz bei der Entwicklung einer Lösung die anderen drei Gestaltungsaspekte adressieren kann, sich aber weiche Faktoren mittels semantischer Modelle nicht sinnvoll abbilden lassen.

Die in Abschnitt 3 präsentierten Anwendungsszenarien adressieren die drei genannten Gestaltungsaspekte einer IT-Governance und veranschaulichen ihre toolgestützte Abbildung in einem Governancemodell.

2.2 IT-Governance-Referenzmodelle und Anforderungen

Referenzmodelle und Best Practices der IT-Governance geben Hilfestellungen sowie inhaltliche und methodische Unterstützung bei der Definition von Zielen und Metriken sowie der Organisation zur Planung, Steuerung und Kontrolle der Unternehmens-IT vor

dem Hintergrund ihrer wichtigsten Ziele. Sie dienen unter anderem dazu, die Definition und Implementierung von Governance- und Managementprozessen an der Schnittstelle zwischen Geschäft und IT zu verbessern [JG11].

Die bereits oben genannten Referenzmodelle wie COBIT und ITIL sind in wesentlichen Teilen natürlichsprachliche, informale, höchstens semiformale Beschreibungen und stellen einen „one best way“ dar und sehen keine methodisch fundierte Hilfestellungen für die unternehmensspezifische Anpassung und Anwendung vor [Lo11].

Hieraus resultierte die Anforderung, verschiedene Mechanismen der Anpassung und Anwendung der Modelle konzeptionell klar zu trennen und eindeutig zu beschreiben und dadurch die Mechanismen systematisch zu unterstützen. Weiterhin ergibt sich daraus, dass viele Referenzmodelle der IT-Governance im Wesentlichen natürlichsprachlich und damit informal formuliert sind, die Anforderung, sie semiformal zu beschreiben bzw. abzubilden, um Mechanismen der Referenzmodellierungsforschung auf sie anwenden zu können [FL04; Lo11; LGT16].

2.3 Semiformale Beschreibung der Referenzmodelle der IT-Governance

In den Forschungsprojekten wurde mit semantischen Netzen eine Wissensrepräsentation bzw. Technologie gewählt, die es erlaubt, Modellinhalte semiformal zu beschreiben bzw. abzubilden [Re10]. Diese hat vor dem Hintergrund der Anforderungen zwei entscheidende Vorteile: Zum einen lässt sich mit semantischen Netzen gut abbilden, dass keiner der genannten drei Gestaltungsaspekte einer IT-Governance dominiert. Zum zweiten nehmen semantische Netze – anders als zum Beispiel relationale Datenbanken – keine strikte Trennung zwischen Typ- und Instanzebene bzw. Metamodellobjekten und Modellobjekten vor, was sich für die in den Projekten verfolgten Ziele als vorteilhaft erwiesen hat, da dies dem intuitiven Umgang mit relevanten Modellbegriffen als Repräsentationen von Ausschnitten der Wirklichkeit eher entspricht [LGF16].

Bei der Abbildung der Inhalte der Referenzmodelle in semantischen Netzen war es möglich, sich an den Vorgehensmodellen des Ontology Engineering zu orientieren [zum Beispiel NM01]. Bei der Entwicklung eines semantischen Netzes ist – wie auch bei der Entwicklung einer Ontologie – der erste Schritt die Ableitung einer (schwachen) Taxonomie aus Metamodellobjekten, die hier aus den vorliegenden IT-Governance-Modellen abstrahiert werden. Im Falle des als Beispiel dienenden Netzes in Abb. 1 sind dies die mit einem dunklen „C“ versehenen Konzepte (z. B. IT-Ziel, IT-Prozess, Management-(Mmt-)Praktik). Diese stehen in einer „hat_oberbegriff“ Beziehung. Neben dieser unspezifischen Über-/Unterordnungs-Beziehung werden domänenspezifische Beziehungen abgebildet. Sie beschreiben hier zum Beispiel, dass eine Rolle eine Verantwortlichkeit bezüglich einer Aktivität hat („verantwortet“) oder dass eine Metrik die Zielerreichung „misst“.⁴ In einem dritten Schritt werden dann den so gebildeten

⁴ Um die Abbildung nicht zu überfrachten wurden nur einige domänenspezifische Assoziationen beschriftet.

Strukturen aus Metamodellobjekten (Modell-)Instanzen zugeordnet („I“).

Dies erfolgt, indem für einen konkreten Prozess (z. B. für den COBIT-Prozess „Sicherstellung der Lieferung von Wertbeiträgen“) das Konzept Rolle mit der konkreten Rolle CIO instanziiert wird und konkrete Aktivitäten beschrieben werden. Diese Instanzen sind Modellobjekte aus den IT-Governance-Modellen. In der Regel verfügt ein Konzept über mehrere Instanzen und einer Instanz können mehrere Instanzen eines anderen Meta-modellobjekts zugeordnet sein. So hat das dargestellte IT-Ziel „Bereitstellung von IT-Services ...“ mehrere zugeordnete IT-Prozesse, von denen zwei dargestellt sind.

Aus Abb. 1 wird auch deutlich, dass in dem semantischen Netz Strukturen, Prozesse und Metriken enthalten sind: Es ist in der Lage, sowohl die Verantwortlichkeiten als auch die Metriken und Aktivitäten (als grundlegende Bestandteile von Prozessen) darzustellen, ohne dass eine Perspektive vernachlässigt werden müsste.

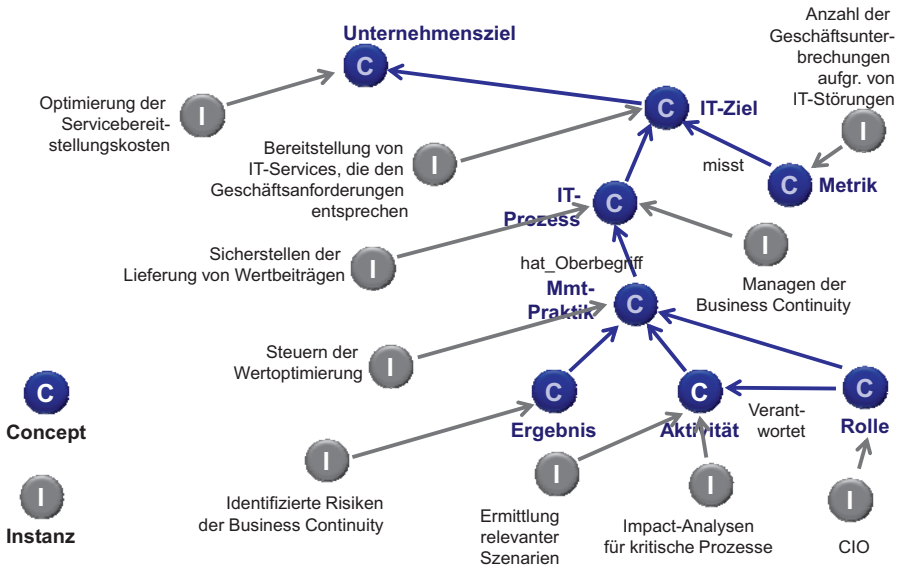


Abb. 1: Repräsentation eines Referenzmodellausschnitts als semantisches Netz (zwei COBIT-Prozesse: *Managen der Business Continuity* und *Sicherstellen der Lieferung von Wertbeiträgen*)

Das resultierende semantische Netz bildet ein Best-Practice-Modell zunächst im Original ab, also ohne unternehmensspezifische Änderungen. Dieses ist dann Ausgangspunkt für eine *Anpassung* und *Anwendung* auf Unternehmensebene und wird der sog. Referenzebene zugeordnet. Die an das Unternehmen angepassten unternehmensspezifischen Modelle werden hingegen der Unternehmensebene zugeordnet.

Für die Erstellung eines Modells der Unternehmensebene wird zunächst eine exakte Kopie des semantischen Netzes angefertigt. Der hierfür im Tool vorgesehene Mechanis-

mus wird als das Klonen des Netzes bezeichnet. Der Klon kann vom Unternehmen angepasst werden, d. h. es werden Konzepte oder Instanzen spezifiziert, verändert, gelöscht oder auch dupliziert. Dies soll möglich sein, ohne dabei die Beziehung zum semantischen Netz des Referenzmodells zu verlieren, sodass Änderungen zurückverfolgbar bleiben (ähnlich der backwards-traceability im Anforderungsmanagement). Abb. 2 zeigt einen Modellausschnitt dargestellt im Tool k-infinity.

Im nächsten Schritt, der *Anwendung* des Modells, erfolgt eine weitere Instanzierung durch die Zuordnung von Objekten aus der Organisation des Unternehmens (so wird bspw. der konkreten Rolle CIO eine konkrete Person zugeordnet oder es werden Prozesse konkreten Organisationseinheiten zugeordnet, mit dem Ergebnis, dass es eine Prozessinstanz „*Manager der Business Continuity*“ für jede relevante Niederlassung oder Betriebsstätte gibt).

Liegt nach der Durchführung der Schritte Anpassung und Anwendung ein unternehmensspezifischer und instanzierter Klon des semantischen Netzes vor, kann in weiteren Schritten die Nutzung des Modells für konkrete Anwendungsszenarien erfolgen. Mögliche Szenarien sind die Entwicklung eines Management-Dashboards als ein auf Referenzmodellen aufsetzendes IT-Reporting (3.1), die Repräsentation des strukturellen Aspekts der IT-Governance im Sinne von aufbauorganisatorischen Mechanismen (3.2) sowie die Auswahl und unternehmensspezifische Modellierung relevanter IT-Prozesse (3.3).

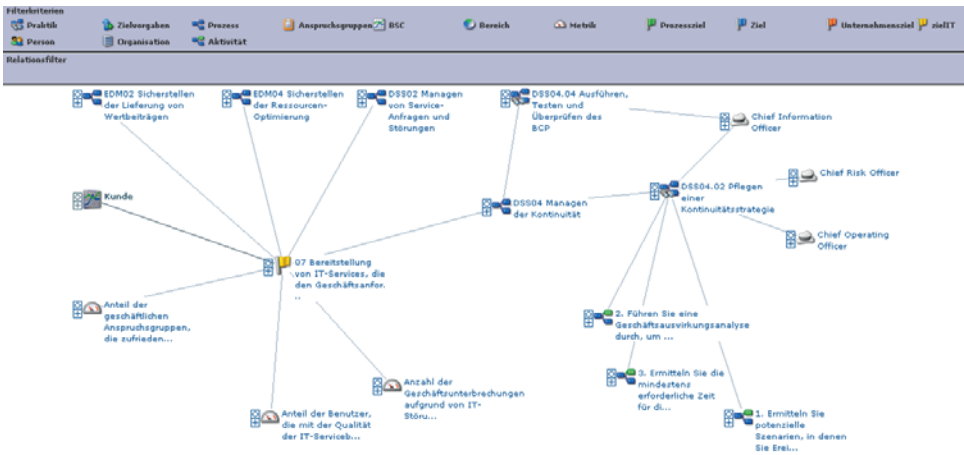


Abb. 2: Screenshot des Tools - Repräsentation des relevanten Modellausschnitts

Das in den Projekten entwickelte IT-Artefakt bekam im Projektverlauf des zweiten Forschungsprojekts den Namen TosI (siehe oben Fußnote 3). Im folgenden dritten Abschnitt werden nun verschiedene Anwendungsszenarien für TosI beschrieben.

3 Demonstration des IT-Artefakts: Drei Anwendungsszenarien

3.1 Entwicklung eines Governance-Dashboards

Im ersten Anwendungsszenario soll beispielhaft gezeigt werden, wie die in TosI abgebildeten COBIT-Metriken einzelnen Unternehmensbereichen zugeordnet werden und dann mittels eines Dashboards visualisiert werden. Als Datengrundlage hierfür wird ein Data-Warehouse eingesetzt, das die Datenstruktur des TosI-Modells multidimensional aufbereitet. Zur Auswertung der Daten und zur Gestaltung des Dashboards wurde die Daten-Analyse-Software Qlikview verwendet. Das IT-Governance-Dashboards dient so dazu, den Status von Prozessen visualisieren, was direkt zur Operationalisierung von Governanceziele beiträgt.

Über die modellierten Zusammenhänge im zugrundeliegenden semantischen Netz werden Verantwortlichkeiten für die Prozesse identifiziert und Prozesse mit aktuellen Metriken aus dem Unternehmen verknüpft. Der Entwurf soll die Anforderungen an ein modernes Führungsinformationssystem adressieren, d. h. die Integration soll eine Übersicht über den Status der IT-Governance-Prozesse bereitzustellen, dem Anwender helfen, durch komplexe Informationen zu navigieren z. B. mithilfe von „Drill-down“-Operationen Kennzahlen zu Prozessen detaillierter zu betrachten, etwa durch das schrittweises „Hereinzoomen“ auf genauere Detailtiefe von Dimensionen bei der Analyse von Daten (Jahr-Monat-Tag, Staat-Region-Stadt-Betriebsstätte). Zudem soll die Bedienung direkt und einfach sein, indem der Anwender sich selbständig und/oder geführt die für ihn relevanten Prozesse anzeigen lassen kann (Ad-hoc-Reporting; „Self-Service BI“).

Die Gestaltung des Dashboards zielt darauf, die Transparenz für die Prozessbeteiligten zu erhöhen und die Steuerung und Überwachung relevanter Prozesse zu ermöglichen. Das semantische Modell enthält bereits die Verantwortlichkeitsbeziehungen für Rollen und wird bei der Anpassung mit konkreten Personen „instanziiert“. Metriken werden in COBIT zwar nicht direkt bestimmten Rollen zugeordnet, jedoch wird der Kreis der potentiell Verantwortlichen durch das semantische Netz eingeschränkt.

Grundlage sind Prozessmetriken als ein Typ quantifizierbarer Daten, die COBIT vorschlägt. Für diese Metriken müssen Ziele und/oder Schwellwerte definiert werden, um sie evaluierbar zu machen, bevor sie schließlich visuell dargestellt werden können. Die COBIT-Metriken sind zweidimensional, d. h. sie zeigen den Zustand eines Sachverhaltes zu einem bestimmten Zeitpunkt an und die Schwellwerte geben an, wie gut der gemessene Wert ist.

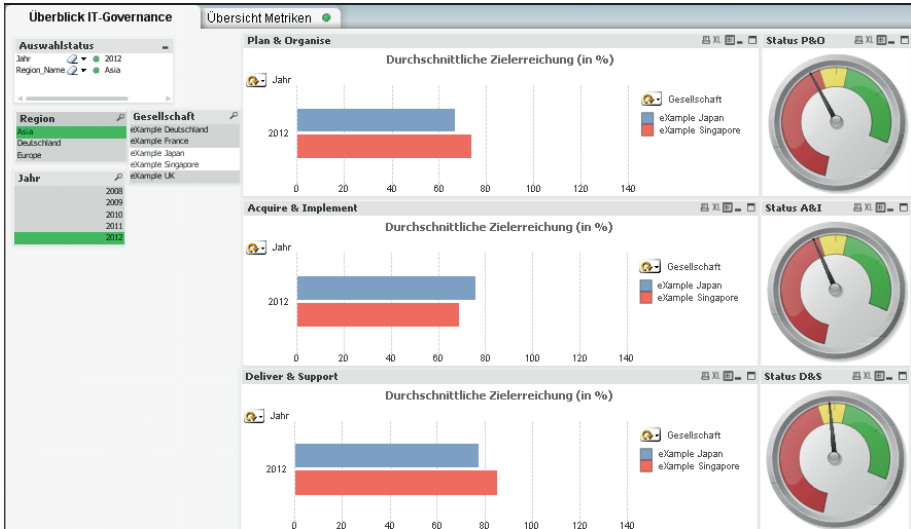


Abb. 3: Dashboard mit Tachometer-Anzeige

Für Daten dieser Klasse bieten sich klassische Geschäftsgrafiken wie Balkendiagramme an. Häufig wird in Reportinganwendungen die Tachometermetapher verwendet, die sich auch im vorliegenden Anwendungsszenario sehr gut eignet, den Status der einzelnen Prozesse zu visualisieren: Anzeige der Tachonadel bei Erfüllung im grünen, bei Erreichung eines Schwellwertes im gelben und bei Nichterfüllung einer Metrik im roten Bereich, wie Abb. 3 zeigt.

Für die Anzeige der Historie einer bestimmten Metrik kann ein Punktediagramm verwendet werden, für metrikübergreifende Statusanzeigen (z.B. Status eines gesamten Prozesses) ist ein Diagramm, wie etwa ein Kiviati-Diagramm, sinnvoll, das anzeigt, zu wie viel Prozent die durch Metriken repräsentierten Ziele eines Prozesses oder, höher aggregiert, einer Domäne, erreicht sind (vgl. Abb. 4). Es dient hier der visuellen Darstellung von Werten mehrerer, gleichwertiger Kategorien und erlaubt die gleichzeitige Darstellung mehrerer Landesgesellschaften und deren Zielerreichung mittels aus COBIT entnommener Kennzahlen.

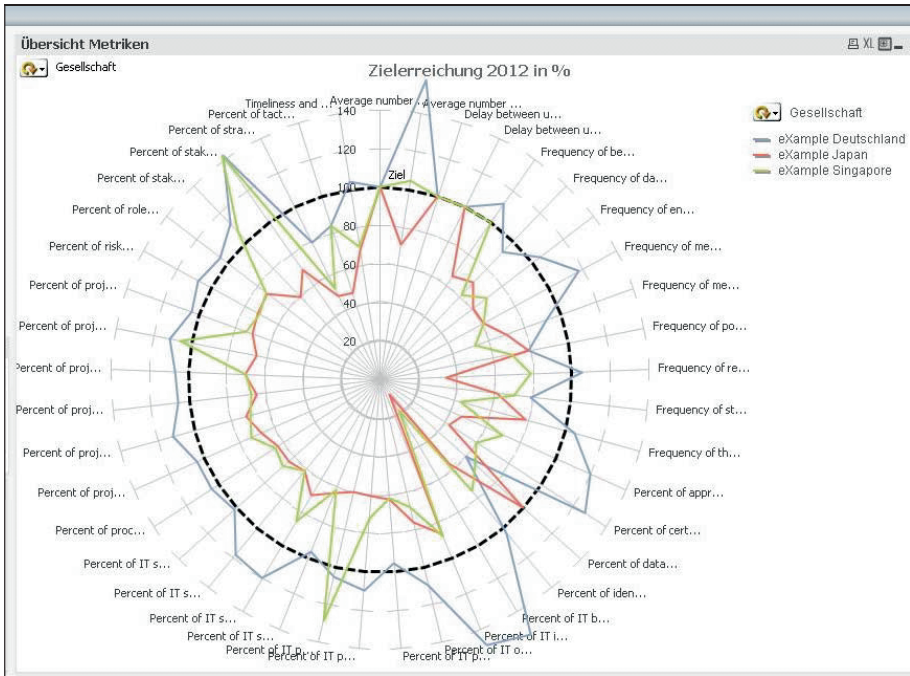


Abb. 4: Kiviati-Diagramm

3.2 Abbildung struktureller Governance-Aspekte

In einem weiteren Anwendungsszenario wurden Strukturen einer IT-Governance-Organisation (aufbauorganisatorische Aspekte) abgebildet. Hierbei wurden mittels Interviews und mehrerer Workshops Verantwortlichkeiten, Entscheidungsrechten und Rechenschaftspflichten bei einem größeren Consulting-Unternehmen erhoben und diese mit dem COBIT-Rollenmodell verglichen und ersteres hierdurch analysiert (der Unternehmensname musste in Abb. 5 verdeckt werden). Die so erstellten Ist- und Soll-Analysen wurden als Grundlage einer Modellbereichsauswahl und -variation herangezogen und schließlich implementiert. Das Werkzeug TosI konnte darüber hinaus nutzbringend sowohl zur Vorbereitung der Interviews und Workshops als auch zur Analyse und Bewertung des Ist-Zustands genutzt werden. Im Rahmen dieses Anwendungsszenarios wurde auch das semantische Netz entsprechend der erhobenen Kunden-Ist- und -Soll-Situation modelliert, sodass beispielsweise im Bereich der Zuständigkeiten (RACI-Rollen in COBIT) mehr Information im Zusammenhang direkt abrufbar ist.

Abb. 5 zeigt die Rollen und deren Beschreibung sowie, welche konkrete Person welche Rolle innehat (hier sind nur Namenskürzel sichtbar), welche Personen keine Rollen haben und eine Tabelle, in der die Rollen gemäß der Unternehmenssituation und der

Referenzmodell-Vorgaben gegenübergestellt werden. Ziel und Zweck davon war, den Abdeckungsgrad zu erheben, einen Vergleich Modell/Ist vorzunehmen und hieraus Anregungen für die Identifikation von Schwachstellen zu erhalten, die dazu dienen, ein Sollmodell zu definieren.

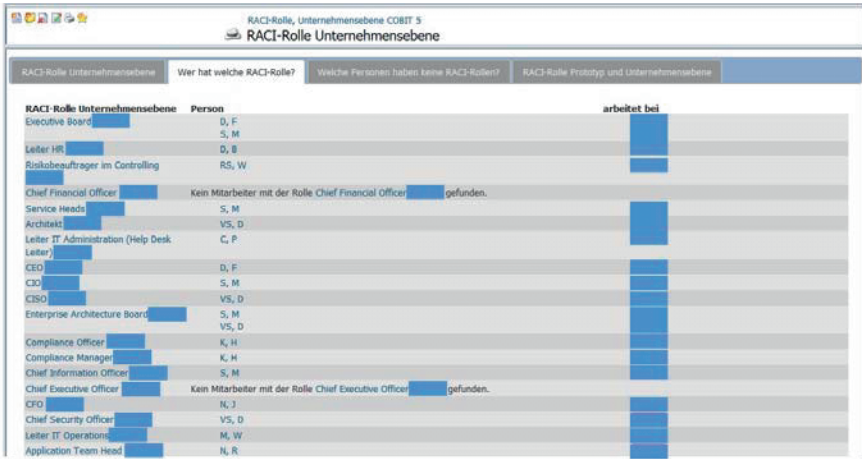


Abb. 5: Benutzeroberfläche, Sicht auf Zuständigkeiten (Unternehmensebene)

Die folgende Abb. 6 zeigt für eine ausgewählte Rolle, den „Grad der Verantwortlichkeit“ (RACI, d. h. Responsible, Accountable, Consulted, Informed Responsible gemäß [IS13]) für die Praktiken, die der Rolle und damit der konkreten Person gemäß dem unternehmensspezifischen Modell zugeordnet sind (letztere wird in dem Screenshot rechts angezeigt).

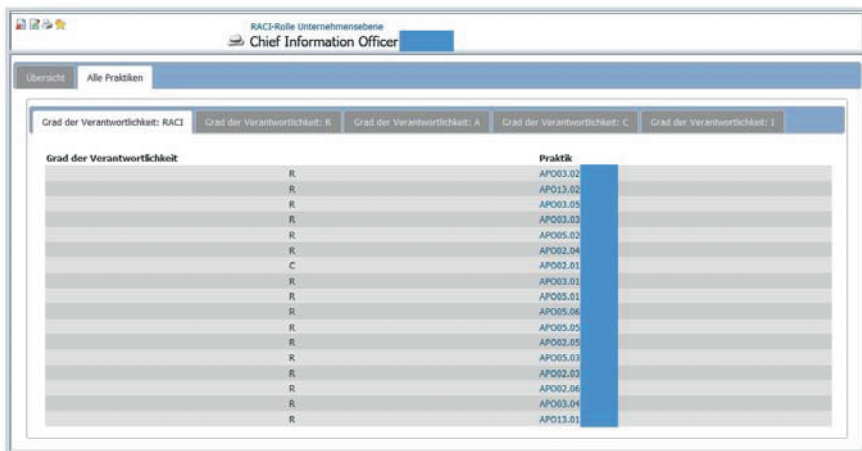


Abb. 6: Benutzeroberfläche, Sicht auf die RACI-Rolle „Chief Information Officer“ der Unternehmensberatung (Unternehmensebene)

3.3 Prozessorientierte Sicht der IT-Governance im Incidentmanagement

Im dritten Anwendungsszenario wird ein IT-Prozess unter Nutzung des Werkzeugs Tosl abgebildet. Unter Verwendung von BI-Werkzeugen werden darüber hinaus relevante Prozesskennzahlen visualisiert und zur Planung und Steuerung des Prozesses verwendet.

Fachliche Grundlage war ein Prozess des Incidentmanagements (Störungsmanagement) eines großen IT-Dienstleisters, der kaum dokumentiert war und bei dem die Prozesssteuerung und das Reporting mittels eines Tabellenkalkulationsprogramm erfolgte. Die Reports in Excel boten aufgrund ihrer schieren Größe und mangelnden Flexibilität für die fachlichen Aufgabenstellungen nur eine geringe Unterstützung. Es sollten an dieser Stelle einschlägige COBIT-Prozesse zur Ableitung eines Sollprozesses dienen und zugehörige Metriken auf ihre Eignung zur Prozesssteuerung hin untersucht werden.

Als relevanter, diese Aufgabenstellung unterstützender COBIT-Prozess wurde *DSS02: Managen von Service-Anfragen und Störungen* identifiziert, welcher der Domäne *Bereitstellen, Betreiben und Unterstützen* zugeordnet ist. Modellkomponenten von diesem sowie mit ihm in Beziehung stehende Prozesse werden in Abb. 7 dargestellt.

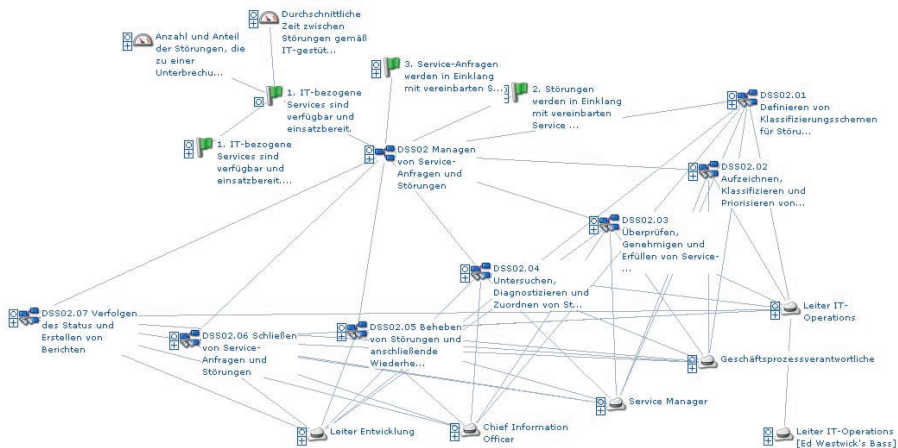


Abb. 7: DSS02 mit Zielen und Metriken sowie Prozesspraktiken und Rollen

Der nächste Schritt war das Anlegen des Unternehmensmodells. Hierzu wurde die Organisationsstruktur untersucht, um den Prozesspraktiken unternehmensspezifische Rollen bzw. konkreten Personen zuordnen zu können (dies ist im obigen Screenshot, um die Anonymität des Unternehmens zu wahren, nur mit einer Instanz der Rolle „Leiter IT-Operations“ angedeutet). Ebenfalls untersucht wurde die regionale Struktur der Serviceerbringung und es wurde eine Klassifikation der unterschiedlichen Services erstellt. Diese Daten fanden Eingang in die multidimensionalen Datenstrukturen, die die Grundlage für ein Data-Warehouse waren, das ein überarbeitetes Reporting mit Ad-hoc-

Analysen für den Prozessverantwortlichen ermöglicht. Abb. 8 zeigt eine Auswertung des so erstellten Reporting-Werkzeugs.

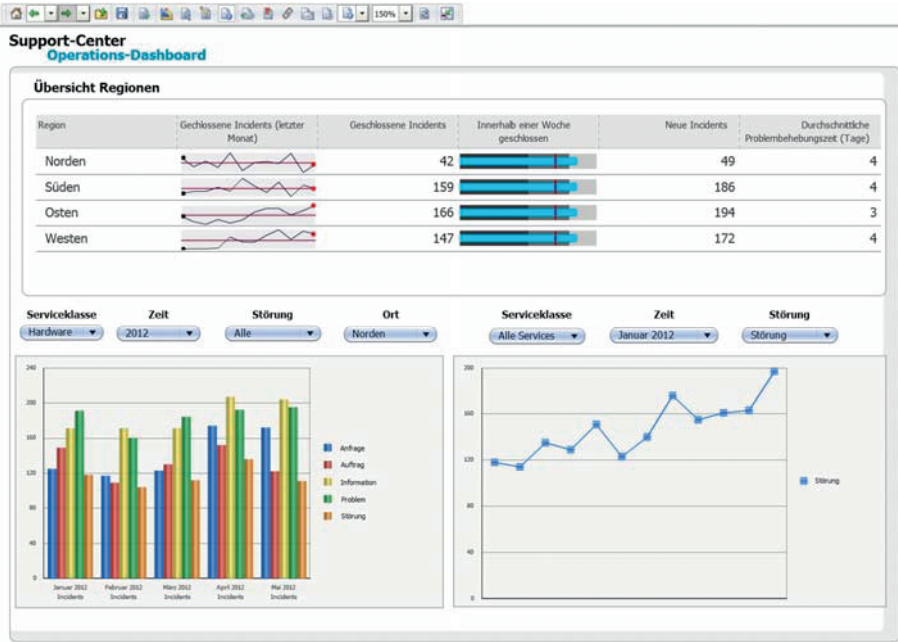


Abb. 8: Operations-Dashboard für das Incidentmanagement bzw. -reporting

4 Zusammenfassung und Ausblick

Gegenstand dieses Beitrags war ein Werkzeug, das in zwei Forschungsprojekten entwickelt wurde. Zunächst wurden grundlegende Anforderungen und grundsätzliche Ziele dargestellt. Hierbei wurde deutlich, dass in beiden Forschungsprojekten ein breiter Ansatz zur Unterstützung der IT-Governance verfolgt wurde, bei dem alle Prozesse sowie Strukturen und das Reporting über Metriken Gegenstand waren. Nach Kenntnis der Verfasser haben kommerzielle Werkzeuge einen im Vergleich engeren Fokus, bspw. auf das Management von Projektportfolien oder zur Steuerung einzelner IT-Prozesse. Mit semantischen Netzen/Ontologien wurde darüber hinaus eine für die Toolentwicklung in diesem Bereich innovative Technologie und Form der Repräsentation gewählt.

Das entwickelte Artefakt lag im Fokus dieses Beitrags. Es wurde anhand von drei Anwendungsszenarien vorgestellt und demonstriert. Gemäß der Design Science Research Methodology (DSRM) von Peffers et al. [Pe07] schließt die Demonstration eines (innovativen) IT-Artefakts an die vorangegangenen Phasen Problemdefinition/Motivation, Zieldefinition und Design an. Der Demonstration wiederum folgt die Evaluation

des IT-Artefakts. In dieser Arbeit stand die Demonstration im Mittelpunkt. Es ging darum, Einsatzmöglichkeiten und die Anwendbarkeit aufzuzeigen und das in den Forschungsprojekten TosI und SemGoRiCo entwickelte Werkzeug darzustellen. Es wurden auch keine systematischen "Lessons Learned" diskutiert. In nachfolgenden Forschungsarbeiten müssen daher Nutzen und Nützlichkeit der vorgeschlagenen Lösung vertieft diskutiert und evaluiert werden. Insbesondere sollte kritisch betrachtet werden, ob mit der systematischen Anwendung der Referenzmodelle tatsächlich ein Nutzen verbunden ist. Möglicherweise mag man es als eine Selbstverständlichkeit ansehen, dass eine systematische Herangehensweise mehr Erfolg verspricht als eine unsystematische. Für den hier gegebenen Fall der Referenzmodelle müsste dies jedoch noch belegt werden, zumal eine systematische Anwendung wahrscheinlich mit einem deutlich höheren Aufwand verbunden sein dürfte.

Literaturverzeichnis

- [Ba15] Bartsch, S.: Ein Referenzmodell zum Wertbeitrag der IT. Wiesbaden 2015.
- [CR07] Chan, Yolande E.; Reich, Blaize Horner: IT alignment: an annotated bibliography. *JIT* 22 (2007) 4, 316-396
- [DG15] De Haes, S.; van Grembergen, W.: *Enterprise Governance of Information Technology*. Springer, Heidelberg 2015.
- [FL04] Fettke, P.; Loos, P.: Referenzmodellierungsforschung. In: *Wirtschaftsinformatik*, Jg. 46 (2004), H. 5, 331–340.
- [IS12] ISACA: *COBIT 5: A Business Framework for the Governance and Management of Enterprise IT*, 2012, online verfügbar unter: www.isaca.org
- [JG11] Johannsen, W.; Goeken, M.: *Referenzmodelle der IT-Governance*. dPunkt.verlag, Heidelberg 2011.
- [Kr15] Krcmar, H. (2015): *Informationsmanagement*. Springer, Heidelberg 2015.
- [LGF16] Looso, S.; Goeken, M.; Felden, C.: Modellierung und Nutzung von Referenzmodellen der IT-Governance. Vorgehen, Fallstudie und Toolunterstützung, In: Nissen, V. et al. (Hrsg.): *Multikonferenz Wirtschaftsinformatik MKWI 2016*, Ilmenau, Germany, 2016.
- [Lo11] Looso, S.: *Best-Practice-Referenzmodelle der IT-Governance. Struktur, Anwendung und Methoden*. Frankfurt School 2011, <http://www.frankfurt-school.de/dms/dissertations/Looso0/Looso.pdf>
- [NM01] Noy, N. F.; McGuinness, D.L.: *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, 2001.
- [Pe07] Peffers, K.T.T.; Rothenberger, M.A.; Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. In: *Journal of Management Information Systems*, Jg. 24 (2007-2008), H. 3, 45–77.

- [Re10] Reichenberger, K. (2010): *Kompendium semantische Netze: Konzepte, Technologie, Modellierung*. Springer, Heidelberg 2010.
- [TK03] Tallon, P.P.; Kraemer, K.L. Investigating the relationship between strategic alignment and IT business value: the discovery of a paradox, In (Shin, H.): *Creating business value with information technology challenges and solutions*, Idea Group, Hershey: 2003, 1-22.
- [TV07] Tuttle, B.; Vandervelde, S.D.: An empirical examination of COBIT as an internal control framework for information technology. In: *International Journal of Accounting Information Systems*, Jg. 8 (2007), 240–263.

Elektronische Langzeitspeicherung als SOA-Dienst – Kernelement eines vertrauenswürdigen Informationsmanagements

Tomasz Kusber¹, Steffen Schwalm²

Abstract: Es besteht eine hohe Notwendigkeit, nicht nur in der öffentlichen Verwaltung, sondern auch in Unternehmen, Geschäftsprozesse zu digitalisieren und für die elektronischen Dokumente und Daten (Unterlagen) auch in ferner Zukunft die Lesbarkeit, Verfügbarkeit sowie die Integrität, Authentizität und Verkehrsfähigkeit gewährleisten zu müssen. Diese Anforderungen bestehen aufgrund geltender regulatorischer Vorgaben einschließlich der Verpflichtung zum Nachweis gegenüber Prüfbehörde, Gerichte, Dritten – bei gleichzeitigen Aufbewahrungsfristen zwischen 2 und 110 Jahren oder dauernd, die teilweise auch erst nach Jahrzehnten beginnen. Mit Blick auf die sinkenden Lebenszyklen der IT bestehen insofern besondere Herausforderungen an die Erhaltung des Beweiswerts der Unterlagen sowie deren Verfügbarkeit. Ein elektronischer Langzeitspeicherdienst ermöglicht den Aufbau der notwendigen technischen Komponenten, Module und Funktionen für alle relevanten IT-Verfahren und kann als eine Grundlage für ein vertrauenswürdigen Informationsmanagement bezeichnet werden. Der Beitrag stellt anhand geltender Standards und Normen sowie der langjährigen Erfahrungen der Autoren mögliche Lösungswege sowie eine beispielhafte Architektur eines Langzeitspeicherdienstes vor.

Keywords: Elektronische Langzeitspeicherung, SOA, vertrauenswürdigen Informationsmanagement, Beweiserhaltung, Vertrauensdienste, eIDAS, Records Management, Informationserhaltung, Information Governance

1 Einleitung

Die Nutzung der Informationstechnologie für Abwicklungen von Geschäftsprozessen ist allgemein etabliert. Geschäftsrelevante Unterlagen liegen zunehmend ausschließlich elektronisch vor. Elektronische Dokumente können jedoch aus sich heraus weder wahrgenommen noch gelesen werden. Sie liefern aus sich heraus auch keine Hinweise für ihre Integrität und Authentizität sowie die Ordnungsmäßigkeit im elektronischen Rechts- und Geschäftsverkehr. Gleichzeitig bestehen jedoch umfassende Dokumentations- und Aufbewahrungspflichten, deren Dauer zwischen zwei und 110 Jahre oder dauernd umfasst, die einen langfristigen Nachweis von Authentizität, Integrität und Nachvollziehbarkeit elektronischer Unterlagen erfordern. Während dieser Fristen muss es zudem möglich sein, die Dokumente Prüfbehörden oder Gerichten

¹ BearingPoint GmbH, Team Secure Information Management, Kurfürstendamm 207-208, 10719, Berlin, tomasz.kusber@bearingpoint.com

² BearingPoint GmbH, Team Secure Information Management, Kurfürstendamm 207-208, 10719, Berlin, steffen.schwalm@bearingpoint.com

vorzulegen und anhand der Daten die genannten Nachweise zu führen (Verkehrsfähigkeit). Diese Anforderungen gelten unabhängig vom konkret eingesetzten IT-Verfahren. Insofern kann eine elektronische Langzeitspeicherung als Querschnittsaufgabe bezeichnet werden, deren Umsetzung in einem verfahrensübergreifenden IT-Dienst im Sinne einer SOA-Architektur, die Nutzung von Synergieeffekten ermöglicht und Doppelaufwände vermeidet. Ein solcher Dienst muss entsprechend der geltenden Anforderungen die Erhaltung der Daten selbst, deren Lesbarkeit/Interpretierbarkeit (Informationserhaltung) sowie des Beweiswerts der Daten (Beweiswerterhaltung) ermöglichen [Ko13], [Ro07].

Der vorliegende Beitrag stellt, auf Basis der langjährigen Praxiserfahrungen der Autoren, eine Beispielarchitektur für einen elektronischen Langzeitspeicher (LZSP) als verfahrensübergreifenden IT-Dienst sowie dessen fachliche Integration in eine SOA-Architektur vor und ist folgendermaßen gegliedert: Abschnitt 2 erläutert die grundsätzlichen Anforderungen an die Aufbewahrung elektronischer Unterlagen. Der Abschnitt 3 beschreibt die wesentlichen Funktionen und beispielhafte Architektur eines Langzeitspeicherdienstes, Abschnitt 4 fasst die Rolle eines Langzeitspeicherdienstes in einem vertrauenswürdigen Informationsmanagement zusammen, ergänzt um einen Ausblick auf zukünftige Entwicklungen.

2 Grundsätzliche Anforderungen an die Aufbewahrung elektronischer Unterlagen

2.1 Grundsatz

Wie in der Einleitung dargelegt, muss entsprechend geltenden regulatorischen Vorgaben, bis zum Ablauf der geltenden Aufbewahrungsfristen der Nachweis von Authentizität, Integrität, Verkehrsfähigkeit und Nachvollziehbarkeit elektronischer Unterlagen gegenüber Gerichten, Prüfbehörden, Dritten jederzeit verlustfrei möglich sein. Besonders in hochregulierten Branchen wie Luft- und Raumfahrt, Gesundheitswesen, LifeScience/Pharma, Forschung, Transportation oder Energiewirtschaft bedeutet dies eine umfassende Herausforderung, insbesondere angesichts Aufbewahrungsfristen von 30 Jahren oder mehr, wobei die Fristen häufig erst nach Jahrzehnten beginnen, z.B. wenn das Produkt außer Produktion geht oder das Bauteil nicht mehr genutzt wird. [Ko13].

Eine wesentliche Grundlage für eine ordnungsgemäße elektronische Langzeitspeicherung bildet ein sachgerechtes Records Management innerhalb eines vertrauenswürdigen Informationsmanagement einer Behörde oder eines Unternehmens. Praktisch gewährleistet ein sachgerechtes Records Management insbesondere

- Sicherstellung der Verfügbarkeit, Sicherheit, Compliance geschäftsrelevanter Unterlagen solange diese in der Organisation benötigt werden
- Definition der notwendigen Rollen, Verantwortlichkeiten und Regularien (Policies)
- Nachvollziehbarkeit der Geschäftsprozesse auf Basis organisatorischer und technischer Maßnahmen
- Daten- und Beweiserhaltung mindestens bis zum Ablauf der geltenden Aufbewahrungsfristen

und erzeugt so Unterlagen, die als eindeutiger Nachweis von Geschäftsprozessen und geschäftlichen Entscheidungen gegenüber Dritten (Gerichte, Prüfbehörden etc.) dienen können [Wi15], [To07], [To10], [ISO-30301]. Es sichert so die Vertrauenswürdigkeit der Unterlagen und damit des Informationsmanagements [ISO-30301], [Lu12], [Ko14].

Die Nutzung kryptographischer Mittel, wie fortgeschrittene oder qualifizierte elektronischer Signaturen und qualifizierte Zeitstempel sowie künftig elektronischer Siegel, ermöglicht nach geltendem Recht die Erhaltung des für die Nachweisführung notwendigen Beweiswerts³, ohne die Verkehrsfähigkeit einzuschränken (siehe [F06], [Ro07], [BMWi07], [eIDAS]). Diese kryptographischen Mittel werden direkt am Dokument/Daten angebracht, so dass er Beweiswert, ebenso wie das Dateiformat oder die Metadaten zur inhaltlichen Beschreibung eines Dokuments, eine inhärente Eigenschaft der jeweiligen elektronischen Unterlagen bildet. Dementsprechend müssen Maßnahmen zur beweissicheren Langzeitspeicherung auch direkt an den elektronischen Unterlagen ansetzen.

Die Beweiserhaltung erfolgt durch eine Nachsignatur bzw. Neuverhashung (vgl. u.a. § 17 SigV, [eIDAS]), also der Anbringung einer neuen qualifizierten elektronischen Signatur oder eines qualifizierten Zeitstempels, sowie ggf. dem Neuverhaschen der eigentlichen Daten, sobald die Sicherheitseignung der Signatur-/Hashalgorithmen nicht mehr gegeben ist. Durch die Verwendung von Merkle-Hashbäumen gemäß [RFC 4998] bzw. [RFC 6283] kann eine wirtschaftliche Nachsignatur einer Vielzahl von Daten gewährleistet werden. Die Nachsignatur muss dabei jeweils alle vorhergehenden Signaturen und Zeitstempel einschließen. Die nachstehende Grafik zeigt eine solche Hashbaum:

³ Der Beweiswert ist terminologisch mit dem juristischen Begriff der „Beweiskraft“ gleichzusetzen.

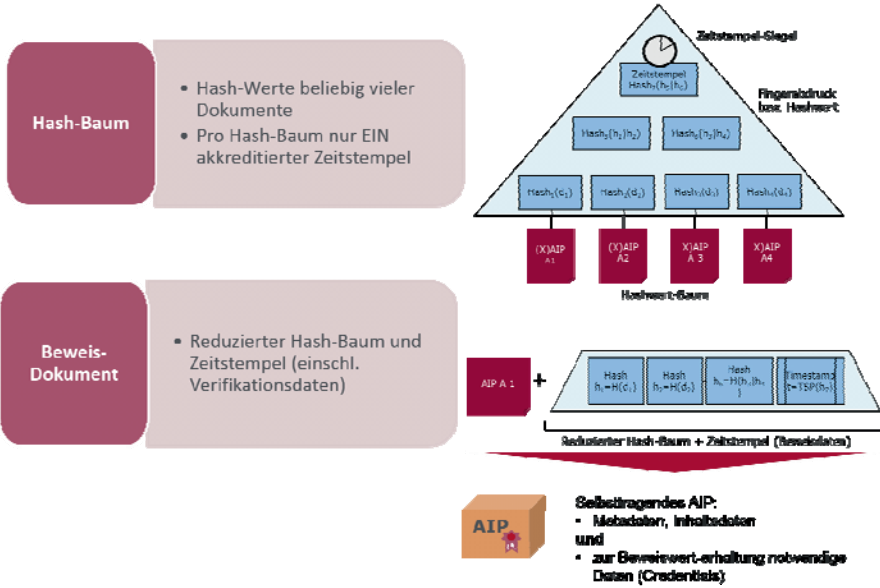


Abbildung 1: Hashbaum und Evidence Record

Daneben gilt es zum einen den Entscheidungsprozess, also die Nachvollziehbarkeit, der Unterlagen nachzuweisen⁴, zum anderen, um den Nachweis zu führen, technische Daten (Gewährleistung der langfristigen Verfügbarkeit) der aufzubewahrenden Unterlagen zu erheben und langzeitzuspeichern [Lu12]. Dies bedingt schlussendlich die Langzeitspeicherung der Unterlagen in Form sog. selbsttragender Archivpakete im Sinne geltender Standards und Normen (vgl. Kap. 2.2) auch mit Blick auf die neue [eIDAS], die in EU und EFTA einheitliche, verbindliche Maßgaben für qualifizierte elektronische Signaturen, Siegel und Zeitstempel definiert und ebenso die Erhaltung von deren Vertrauenswürdigkeit und damit die Beweiserhaltung fordert [eIDAS], [Ku,Vo,Do,Sc16]. Verbindliche ETSI-/CEN-Normen bestimmen zudem den technischen Rahmen in der Umsetzung der [eIDAS] z.B. durch einheitliche Formate. Aufgrund der künftigen Zulässigkeit von Server-/Fern- und mobilen Signaturen ist zudem eine weitere Verbreitung signierter Dokumente, wie dies außerhalb Deutschlands z.B. in Österreich oder Großbritannien bereits der Fall ist, absehbar. Gleichzeitig gewährleistet eine beweissichere elektronische Langzeitspeicherung eine langfristige Erhaltung der aufzubewahrenden, geschäftsrelevanten Unterlagen und somit deren Verfügbarkeit (siehe [Ko13], [ISO-14721], To07). Als Grundlage dienen anerkannte nationale wie internationale Standards (vgl. Kap. 2.2).

Darüber hinaus stellt sich die Aufgabe der Erhaltung der aufzubewahrenden Unterlagen unter Wahrung geltender Aufbewahrungsvorgaben wie z.B. maschinelle Auswertbarkeit,

⁴ z.B. bedingt durch das Prinzip der Aktenmäßigkeit gem. Art. 20 Abs. 3 Grundgesetz in öffentlichen Stellen.

originäre Darstellbarkeit der Daten. Es gilt also Dateiformate, Strukturen etc. teilweise jahrzehntelang so zu erhalten, dass die originäre Darstellung der Daten gewährleistet wird. Dabei ist der Grundsatz zu beachten, dass nicht das IT-Verfahren, sondern die Information den Aufbewahrungspflichten unterliegt. Insofern müssen auch Maßnahmen zur Informationserhaltung an den Daten selbst ansetzen [Borghoff03], [Ro07], [Giaretta11].

2.2 Relevante Standards und Normen

Die nachstehende Grafik zeigt die wesentliche Standards und Normen zur beweisicherten Langzeitspeicherung im Überblick:

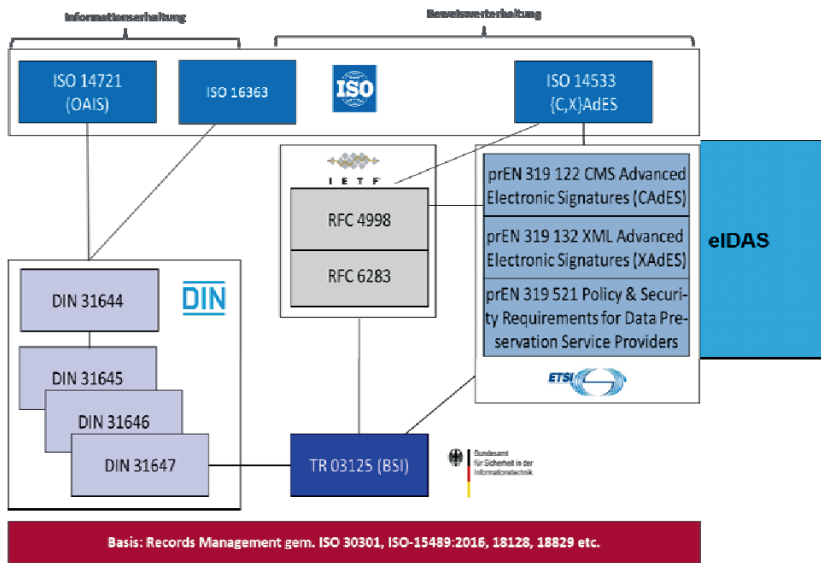


Abbildung 2: Standards und Normen zur Langzeitspeicherung

Wesentliche Basis bildet ein ordnungsgemäßes Records Management, welches anhand klarer Richtlinien, Verantwortlichkeiten und Prozesse die Identifikation und strukturierte wie anforderungsgerechte Ablage geschäftsrelevanter Unterlagen gewährleistet (vgl. [ISO30301], [ISO15489], [ISO18128], [ISO18829]). Darüber hinaus das OAIS-Modell [ISO14721] und die [ISO16363], [DIN31644], [DIN31645] die Prozesse und Informationspakete zur Informationserhaltung innerhalb eines vertrauenswürdigen digitalen Langzeitarchivs (dLZA) während die [DIN31647] die notwendigen Funktionen und Informationen zur Beweiswerterhaltung in einem OAIS-konformen dLZA beschreibt und damit die Verbindung zu den technischen Normen zur Beweiswerterhaltung bildet (vgl. [ISO14533], [EN319122], [EN319132], [EN319142], [EN319162], [RFC4998], [RFC6283]). Die [TR03125] wiederum beschreibt eine

mögliche Referenzarchitektur eines Systems zur Beweiswerterhaltung elektronischer Unterlagen mit daraus abgeleiteten Anforderungen. Sie integriert die Anforderungen aus ETSI⁵ und IETF ([RFC 4998/6293]). Im praktischen Fall des betrachteten Unternehmens war diese Referenzarchitektur ein integraler Bestandteil Langzeitspeichers – als eigenständiges Modul zur Beweiswerterhaltung gem. OAIS [Ko16]. Sowohl bei [TR-03125] als auch [ISO16363] und [DIN31644] bestehen anerkannte Zertifizierungsverfahren, um die Standardkonformität konkreter Marktlösungen zu prüfen und nachzuweisen [Ko14], [DIN31644]. Daneben floss die [TR03125] in das E-Government-Gesetz des Bundes [EGovG] ein und gilt für öffentliche Stellen als sog. Stand der Technik zur Langzeitspeicherung.

3 Aufbau einer Beispielarchitektur für einen elektronischen Langzeitspeicherdienst

3.1 Notwendige Funktionscluster gem. OAIS

Wie in Kap. 2 dargestellt benötigt ein elektronischer LZSP sowohl Funktionen zur Informations- als auch zur Beweiswerterhaltung. Ziel sollte es dabei sein, so viele wie möglich vorhandene Dienste innerhalb der IT-Infrastruktur nachzunutzen. Erfahrungsgemäß sind zahlreiche Funktionen nicht langzeitspeicherspezifisch. So wird eine Signaturerzeugung einerseits im Geschäftsprozess, andererseits zur Beweiswerterhaltung (Nachsignatur) benötigt. Gleiches gilt für Konvertierungsfunktionen z.B. von Word nach PDF/A, was beim elektronischen Versand sowie zur Informationserhaltung erforderlich ist. Um eine effiziente Umsetzung zu ermöglichen und Synergien zu nutzen, bietet sich zur Clusterung der einzelnen Funktionen erfahrungsgemäß die Orientierung an den Prozessen des OAIS-Modells sowie der [TR03125] an. Diese stellen faktisch bereits konkrete Funktionsgruppen dar.

Ingest

Das erste Funktionscluster wären also die Funktionen zur Übernahme elektronischer Unterlagen in einen elektronischen LZSP bilden. Aus Gründen der IT-Sicherheit, so insbesondere dem Schutz der im LZSP abgelegten Unterlagen kann es sinnvoll sein, die Übernahmefunktionen in Funktionen zur Anbindung von Geschäftsanwendungen bzw. sendenden Systemen im sog. Pre-Ingest und die eigentlichen Kernfunktionen zur Datenübernahme im Kern-Ingest zu trennen.

⁵ Empfehlung für Verwendung von {C,X,P}AdES als Signaturformat zu verwenden.

Um den LZSP als Dienst nutzen zu können ist es erforderlich, verschiedene Standardschnittstellen (z.B. Webservice-Stack Spezifikationen entspr. OASIS und W3C sowie ArchiveLink bei SAP-Anbindung) gekapselt in einem sog. Konnektormodul im Pre-Ingest nach außen anzubieten. Hierüber werden in der Folge die Geschäftsanwendungen an den LZSP angebunden. Die Ablage von Daten kann dabei sowohl bei Eingang eines Dokuments in der Geschäftsanwendung z.B. E-Mails, Scan etc. als auch nach Abschluss eines Geschäftsvorfalles erfolgen. Das Konnektormodul prüft auf Basis einer sicheren Authentisierung die Zugriffsberechtigung der Geschäftsanwendung und leitet den eigentlich Request an die Kern-Ingest-Komponenten weiter. Grundsätzlich ist der Befehlsvorrat der Konnektoren zwecks standardisierter und damit effizienter Anbindung von Geschäftsanwendungen die notwendigen Kernprozesse zu beschränken. Diese sind i.d.R.: [TR03125]

- Übernahme von Daten und Erzeugung Archivinformationspakete (AIP)
- Erweiterung bereits im LZSP aufbewahrter Archivinformationspakete (z.B. durch eine weitere Mail, weiteren Dokumentenstand)
- Abruf von Datenpaketen (DIP) und einzelner Datenelemente aus Archivinformationspaketen
- Abruf technischer Beweisdaten aus AIP
- Löschen von AIP.

Mit diesem Vorgehen lässt sich faktisch jedes aktuelle IT-Verfahren an einen elektronischen LZSP anbinden. Dies wurde auch bereits für SAP realisiert [Ko15]. Für die Aufnahme von Daten aus IT-Verfahren, die nicht über die notwendigen Schnittstellen verfügt resp. der Aufwand zu deren Anpassung zu hoch ist oder die aufzubewahrenden Daten in einem Filesystem ohne Geschäftsanwendung vorliegen (z.B. Windows-Fileshare) kann eine einfache Uploadfunktion umgesetzt werden.

Die Kern-Ingestkomponente basiert verfahrensbezogen auf sog. Policies, also Regeln, wie, in welcher Form, in welchem Format, Struktur etc. die Unterlagen der verschiedenen IT-Verfahren in den elektronischen LZSP zu übernehmen sind. Diese technischen Regeln (z.B. umgesetzt als Workflows) sind anhand der regulatorischen und sonstige Anforderungen an die Aufbewahrung der Unterlagen, durch die fachlich verantwortlichen Organisationseinheiten, gemeinsam mit der IT zu definieren. Anhand dieser Policies werden die vom Pre-Ingest übergebenen SIP⁶ geprüft und nur bei erfolgreicher Validierung weiterverarbeitet. Die Prüfung erfordert die Nutzung von Formatvalidierern sowohl für Metadaten (i.d.R. XML), Inhaltsdaten als auch vorhandener Signaturen und Zeitstempel. Ganz im Sinne einer diensteorientierten Architektur kann das dem LZSP immanente Krypto-Modul zur Signaturprüfung separat

⁶ SIP = Submission Information Package (vgl. [ISO14721])

angesteuert werden, ohne alle übrigen Funktionen der Middleware zu nutzen oder selbige dabei zu kompromittieren [TR03125].

Neben der Form des SIP definieren die o.g. Policies Form und Inhalt der Archivinformationspakete (AIP), die aus den SIP erzeugt werden. Hierzu gehören Module mit Funktionen zur Informationserhaltung wie z.B.:

- Konvertierung von Inhaltsdaten und Validierung der Konvertierung
- Erhebung technischer Metadaten zur Beschreibung, Validierung inhaltlicher und technischer Metadaten und Volltextindizierung
- Mapping von IDs (z.B. bei SAP siehe [Ko15])
- Erzeugung und Validierung von AIP-Container

Beweiswerterhaltung

Im Ergebnis der Erzeugung eines AIP entsprechend den Vorgaben zur Informationserhaltung in der Kern-Ingestkomponente wird dieses an die Komponente zur Beweiswerterhaltung übergeben. Hierbei empfiehlt sich, eine Anlehnung an die Referenzarchitektur der [TR-03125] und deren standardisierte Komponenten, um eine wirtschaftliche Umsetzung zu ermöglichen.

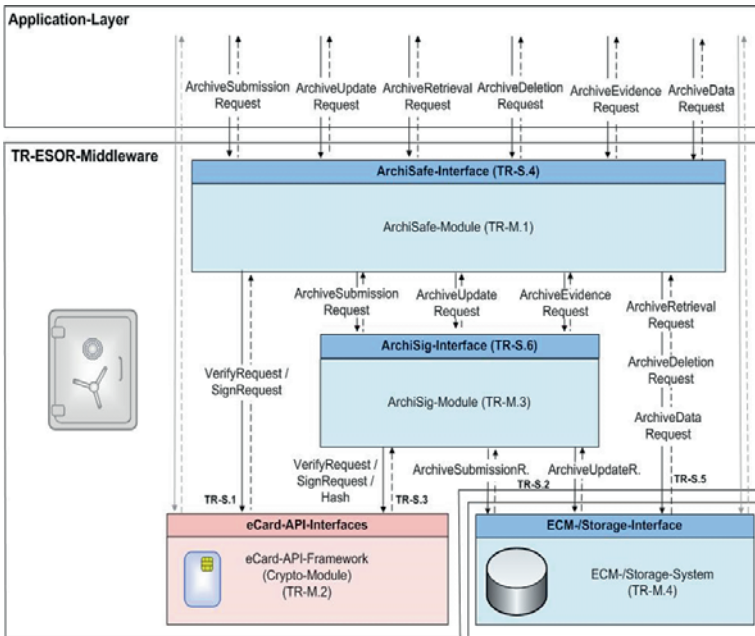


Abbildung 3: Referenzarchitektur BSI TR-03125 v1.2

Da die Middleware die Kernfunktionen zur Beweiswerterhaltung umfasst und die unmittelbar Schnittstelle zum eigentlichen Storage bildet, sind beide Cluster besonders abzusichern, z.B. durch Betrieb in einer eigenen Sicherheitszone. Um bei großen Datenmengen⁷ das Performanceproblem sowohl bei XML-Datenpaketen aufgrund der Base64-Codierung (binäre Daten) als auch ZIP- oder PDF/A-3-Containern zu vermeiden, wäre folgender Weg eine Lösungsoption. Aus der Kern-Ingestkomponente können die Inhaltsdaten in einem sicheren Zwischenspeicher, der Teil des Archivspeichers ist, abgelegt und das Link hierauf sowie ein Hashwert der Inhaltsdaten als eindeutige Repräsentation im XAIP-basierten AIP abgelegt werden. Das AIP selbst wird nun einschl. Metadaten, Credentials (beweisrelevante Daten), dem Hashwert der Inhaltsdaten und dem Link als XAIP an die Middleware übergeben. Diese führt die ersten Funktionen zur Beweiswerterhaltung durch also z.B. Signaturprüfung, Einholung der Zertifikats- und Sperrinformationen. In das Hashen der für die Beweiswerterhaltung relevanten Teile des AIP werden durch Zugriff über einen gesicherten Kanal (z.B. VPN) die Inhaltsdaten mit einbezogen, jedoch nicht ohne zuvor deren Hashwert zu prüfen, um sicherzustellen, dass die Daten im Zwischenspeicher nicht verändert wurden. Nach Aufbau des Hashbaums kann dieser sofort reduziert und damit die Evidence Records erzeugt und diese in den nichtgehashten Teilen des AIP abgelegt werden [TR03125-F]. Aufgrund des sicheren Links auf die eigentlichen Inhaltsdaten könnte, nach Abschluss der Funktionen zur Beweiswerterhaltung, vor der eigentlichen Speicher und im Archivspeicher die Unterlagen selbst, ebenso im AIP abgelegt werden. Damit wird ein vollständig selbsttragendes AIP erzeugt, welches grundsätzlich mit marktüblicher Standardsoftware lesbar und prüfbar ist. Die eindeutige AOID je AIP wird im ArchiSig-Modul erzeugt, auf Basis der [TR03125] und damit in non-proprietärer Form. Diese wird sowohl an das Cluster Datenmanagement als auch die Geschäftsanwendung (soweit vorhanden) für einen späteren Zugriff auf die AIP gegeben.

Datenmanagement (Repository)

Beim Datenmanagement handelt es sich üblicherweise um ein Standardarchivsystem, welches Funktionen zur Datenverwaltung, Zugriffskontrolle und Prozessverbindung zwischen der Kern-Ingestkomponente, der Beweiswerterhaltung sowie dem Access organisiert. Daneben realisiert es i.d.R. in der praktischen Umsetzung bzw. Produktausprägung die Funktionen zur Recherche, Workflow Zugriff, Bestandserhaltung und Systemadministration. Im Repository werden zudem, neben den Volltextindizes die inhaltlichen Metadaten der Unterlagen sowie deren eindeutige AOID für einen späteren Zugriff z.B. für Daten aus Fileshares verwaltet. Alle Zugriffe auf das Cluster Beweiswerterhaltung und Archivspeicher nach Ablage der Daten sollten aus Gründen der Komplexitätsreduktion über das Cluster Datenmanagement erfolgen.

Archivspeicher

Die AIP werden anschließend im Archivspeicher abgelegt und die erzeugte eindeutige AOID der sendenden Geschäftsanwendung sowie dem Repository des LZSP übergeben.

⁷ z.B. Geodaten

Im Sinne einer serviceorientierten Architektur bietet sich die Nutzung von Standardspeicher auch für die Langzeitspeicherung auf Basis eines Sicherheitskonzepts nach ISO 27k an, um ein flexibles Daten- und Speichermanagement zu ermöglichen und den Betrieb einer teuren, separaten Infrastruktur, wie dies z.B. für WORM notwendig wäre, zu vermeiden. Die Nutzung von WORM ist aus Sicht der Langzeitspeicherung wenig empfehlenswert, da aufgrund der hochproprietären Speicherung der Daten im WORM, eine immense Abhängigkeit vom jeweiligen Hersteller besteht, was bei teilweise jahrzehntelangen Aufbewahrungsfristen ein unabsehbares Risiko für die Verfügbarkeit der Daten impliziert. Hinzu kommt, die absehbare Datenexplosion bei der Erweiterung von AIP oder der Informations- und Beweiserhaltung. [Spi11], [Ko15].

Access (Recherche und Zugriff)

Der Access-Cluster fasst die Module und Funktionen zur Recherche, Zugriff und Darstellung der Daten (sofern dies nicht vollumfänglich durch die Geschäftsanwendung erfolgt) zusammen. In der Praxis handelt es sich meist um Teile des für das Repository eingesetzten Archivsystems. Das Rechtekonzept des Langzeitspeichers beruht auf systembasierten Benutzerrechten. Im Klartext heißt das, dass jedes zugreifende IT-Verfahren nur als Systemnutzer behandelt wird, d.h. es wird die Systemberechtigung geprüft, nicht die Berechtigung des einzelnen personellen Nutzers in der Geschäftsanwendung.

Preservation Planning und Systemadministration

Das Cluster Preservation Planning umfasst i.d.R. Funktionen zur Überwachung der technischen Entwicklung von Formaten, Datenstrukturen, Softwarefunktionen im Sinne der Informations- und Beweiserhaltung, so z.B. die Kategorisierung und Pflege von Spezifikationen, das Risikomanagement, die Bewertung eingesetzter Softwaretools oder die Überwachung von Algorithmen. Hinzu kommt die Planung und Steuerung konkreter Maßnahmen wie z.B. Migration von Formaten und Strukturen, die Nachsignatur oder Hasherneuerung.

Das Cluster Systemadministration hingegen umfasst die Fach- und technische Administration des Langzeitspeichers. Beide Cluster werden in der Praxis häufig durch Funktionen des Archivsystems (Repository) abgebildet.

3.2 Mögliche Gesamtarchitektur eines IT-Dienstes zur elektronischen Langzeitspeicherung

Die in Kap. 3.1 beschriebenen Cluster, Module und Funktionen lassen sich zu folgender Gesamtarchitektur eines Langzeitspeicherdienstes zusammenfassen. Es handelt sich um eine Überblicksdarstellung in der Struktur des OAIS-Modells und enthält aus Gründen der Übersichtlichkeit nicht alle Module im Detail.

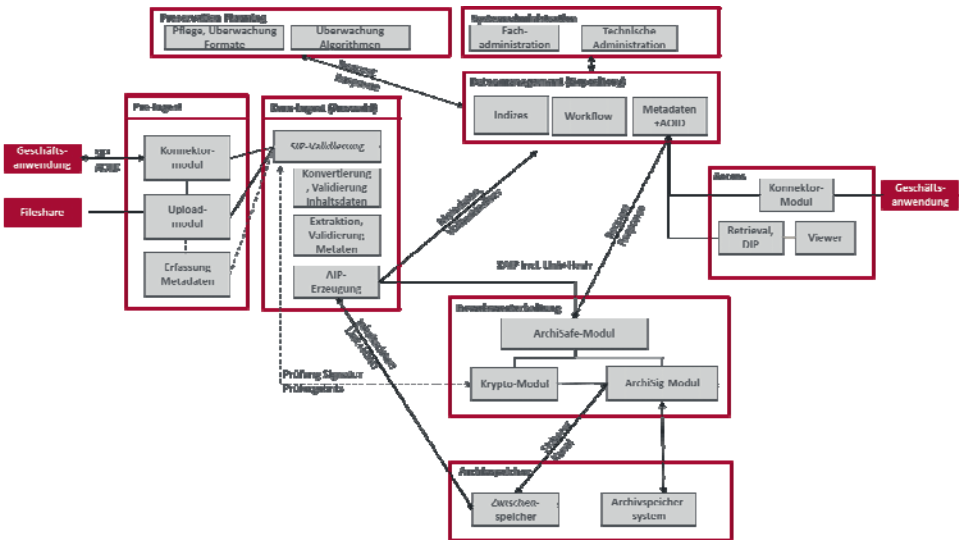


Abbildung 4: Überblick Gesamtarchitektur

Diese Architektur ist ein Beispiel, auf Basis der Projekterfahrungen der Autoren in Behörden, IT-Dienstleistern oder Unternehmen umgesetzt wurde.

4 Rolle eines elektronischen Langzeitspeicherdienstes in einem vertrauenswürdigen Informationsmanagement – Zusammenfassung und Ausblick

Vertrauenswürdigkeit eines Systems oder Unterlagen bedeutet, dass sie das sind, was sie zu sein vorgeben [DIN31644] und sind elementare Grundlage, um bestehenden Dokumentations- und Nachweispflichten erfolgreich nachzukommen [DIN31644 Kommentar]. Ein ordnungsgemäßes Records Management schafft hierfür die fachliche wie technische Basis, eine elektronische Langzeitspeicherung sichert die Authentizität, Integrität, Verkehrsfähigkeit und Nachvollziehbarkeit geschäftsrelevanter Unterlagen und damit deren Vertrauenswürdigkeit bis zum Ablauf auch jahrzehntelanger Aufbewahrungsfristen [ISO-30301], [ISO-15489], [To07], [Ro07]. Das Informationsmanagement einer Behörde oder eines Unternehmens muss sich an der Vertrauenswürdigkeit der Unterlagen messen lassen. Nur wenn diese authentisch und integer sind, können sie als valides Fundament für die Geschäftsprozesse und Entscheidungen dienen. Ein elektronischer Langzeitspeicherdienst ermöglicht den Aufbau der notwendigen technischen Komponenten, Module und Funktionen für alle

relevanten IT-Verfahren. Langzeitspeicherung ist eine Querschnittsaufgabe, die zudem nicht erst bei Abschluss eines Geschäftsvorfalles, sondern wie z.B. bei signierten Dokumenten, Mails oder gescannten Daten bei Beginn des Geschäftsprozesses (sog. frühe beweisichere Speicherung) stattfindet. Der LZSP kann als IT-Dienst insofern für den gesamten Prozess und Lebenszyklus geschäftsrelevanter Unterlagen als vertrauenswürdiger Datenraum dienen – abrufbar für alle angebotenen Geschäftsanwendungen. So kann darüber hinaus z.B. durch die Nutzung der Module des Langzeitspeichers der Eingang von Dokumenten langfristig eindeutig nachgewiesen werden (Eingangszeitstempel via Krypto-Modul und frühe Aufbewahrung), wie dies z.B. bei Versicherungen oder in Behörden realisiert wird. Die Aufbewahrung der Unterlagen in selbsttragenden AIP gewährleistet dabei die Unabhängigkeit von einer bestimmten technischen Plattform und damit Investitionssicherheit (z.B. leichter umsetzbare Migration). Die nachstehende Grafik verdeutlicht dies.

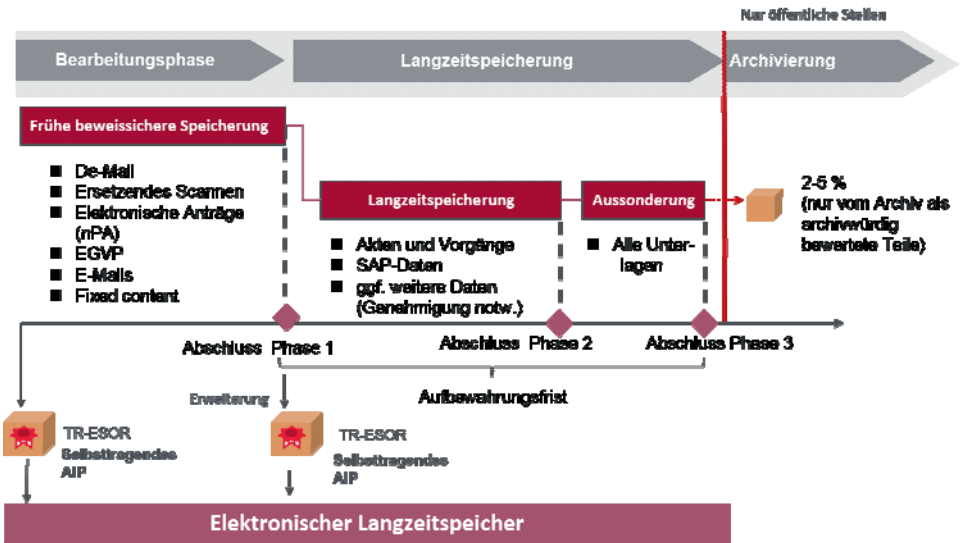


Abbildung 5: Anwendungsbereich des Langzeitspeichers

Im Hinblick auf das absehbar höhere Aufkommen signierter Dokumente im Kontext der neuen [eIDAS]-Verordnung sowie der steigenden Anforderungen an die Sicherheit geschäftsrelevanter Dokumente so z.B. in kritischen Infrastrukturen oder aufgrund der EU-Datenschutzgrundverordnung (vgl. [EU-DSG]) oder dem IT-Sicherheitsgesetz [ITSG] etc. gewinnt ein elektronischer Langzeitspeicherdienst als sicherer Datenraum eines vertrauenswürdigen Informationsmanagement eine erfolgskritische Bedeutung.

Literaturverzeichnis

- [BMWi07] Bundesministerium für Wirtschaft und Technologie (Hrsg.): Handlungsleitfaden zur Aufbewahrung elektronischer und elektronisch signierter Dokumente, Berlin 2007.
- [DIN31644] DIN 31644:2012 Information und Dokumentation — Kriterien für vertrauenswürdige digitale Langzeitarchive, 2012.
- [DIN31645] DIN 31645:2011 Information und Dokumentation – Leitfaden zur Informationsübernahme in digitale Langzeitarchive. 2011
- [DIN31646] DIN 31646:2013 Information und Dokumentation – Anforderungen an die langfristige Handhabung persistenter Identifikatoren (Persistent Identifier). 2013
- [DIN31647] DIN 31647:2015 Beweiswerterhalt kryptografisch signierter Dokumente, 2015.
- [EGOVG] Gesetz zur Förderung der elektronischen Verwaltung (E-Government-Gesetz - E-GovG) vom 25.07.2013
- [EIDAS] VERORDNUNG (EU) Nr. 910/2014 DES EUROPÄISCHEN PARLAMENTS UND DES RATES über elektronische Identifizierung und Vertrauensdienste für elektronische Transaktionen im Binnenmarkt und zur Aufhebung der Richtlinie 1999/93/EG“ vom 23.07.2014
- [EU-DSGV] VERORDNUNG (EU) 2016/679 DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung)
- [EN319122] ETSI EN 319 122 – {1,2}, Electronic Signatures and Infrastructures (ESI); CADES digital signatures, ETSI Draft (V1.1.0 (2016-02))
- [EN319132] ETSI EN 319 132 – {1,2}, Electronic Signatures and Infrastructures (ESI); XAdES digital signatures, ETSI Draft (V1.1.0 (2016-02))
- [EN319142] ETSI EN 319 142 – {1,2}, Electronic Signatures and Infrastructures (ESI); PAdES digital Signatures, ETSI Draft (V1.1.0 (2016-02))
- [EN319162] ETSI EN 319 162 – {1,2}, Electronic Signatures and Infrastructures (ESI); Associated Signature Containers (ASiC), ETSI Draft (V1.1.0 (2016-02))
- [Fisc06] S. Fischer-Dieskau: Das elektronisch signierte Dokument als Mittel zur Beweissicherung, Baden-Baden, 2006.
- [ISO13527] ISO 13527:2010, Space data and information transfer systems -- XML formatted data unit (XFDU) structure and construction rules, 2010
- [ISO-14721] ISO 14721:2012, Space data and information transfer systems — Open archival information system — Reference model, 2nd Edition, 2012
- [ISO14533] ISO 14533: Processes, data elements and documents in commerce, industry and administration – Long-term signature profiles. 2014

- [ISO16363] ISO 16363:2012. Space data and information transfer systems - Audit and certification of trustworthy digital repositories. 2012
- [ISO-21320] ISO/IEC 21320-1:2015. Information technology -- Document Container File -- Part 1 : Core
- [ISO30301] ISO 30301:2011, Information and documentation - Management systems for records - Requirements. 2011
- [ITSG] Gesetz zur Erhöhung der Sicherheit informationstechnischer Systeme (IT-Sicherheitsgesetz) vom 17. Juli 2015.
- [Ko13] U. Korte, S. Schwalm, D. Hühnlein: Vertrauenswürdige und beweiswerterhaltende Langzeitspeicherung auf Basis von DIN 31647 und BSI TR-03125, Informatik 2013, GI-LNI, P220, ISBN 978-3-88579-614-5, S. 550-566, 2013
- [Ko14] U. Korte, S. Schwalm, D. Hühnlein: Standards und Lösungen zur langfristigen Beweiswerterhaltung. DACH-Security 2014, S. 46-58. Frechen 2014
- [Ko15] U. Korte, S. Schwalm, D. Hühnlein, T. Kusber: Ersetzendes Scannen und Beweiswerterhaltung mit SAP. DACH-Security 2015. S. 72-85. Frechen 2015
- [Ko16] U. Korte, S. Schwalm, D. Hühnlein, T. Kusber: Beweiswerterhaltung im Kontext eIDAS - eine Case Study. DACH-Security 2016, Frechen 2016
- [Lu12] Schriftgutverwaltung nach DIN ISO 15489-1. Ein Leitfaden zur qualitätssicheren Aktenführung, Hrsg. von Alexandra Lutz c/o Arbeitskreis Schriftgutverwaltung im DIN NABD 15. Berlin 2012
- [RFC3161] C. Adams, P. Cain, D. Pinkas, R. Zuccherato: Internet X.509 Public Key Infrastructure – Time-Stamp Protocol (TSP), IETF RFC 3161, <http://www.ietf.org/rfc/rfc3161.txt>, 2001.
- [RFC4998] T. Gondrom, R. Brandner, U. Pordesch: Evidence Record Syntax (ERS), IETF RFC 4998, <http://www.ietf.org/rfc/rfc4998.txt>, August 2007.
- [RFC6283] A. J. Blazic, S. Saljic, T. Gondrom: Extensible Markup Language Evidence Record Syntax (XMLERS), IETF RFC 6283, <http://www.ietf.org/rfc/rfc6283.txt>, Juli 2011
- [Ro07] A. Rossnagel: Langfristige Aufbewahrung elektronischer Dokumente, Anforderungen und Trends, Baden-Baden, 2007
- [Spi11] Stephan Spitz et.al.: Kryptographie und IT-Sicherheit. Grundlagen und Anwendungen. Wiesbaden 2011
- [To07] Peter M. Toebak: Records Management. Ein Handbuch. Baden 2007
- [To10] Peter M. Toebak: Records Management. Gestaltung und Umsetzung. Baden 2010
- [TR-03125] BSI: Beweiswerterhaltung kryptographisch signierter Dokumente (TR-ESOR), TR 03125, V1.2., 2015.
- [TR-03125-F] BSI: Anlage F zu TR-03125, Formate und Protokolle, TR-03125, V1.2., 2015.
- [Wi15] Bruno Wildhaber et.al.: Leitfaden Information Governance. Zürich 2015

Eye-Tracking zur Untersuchung von Vertrauenssignalen auf Webseiten von Cloud Computing-Anbietern

Michael Adelmeyer¹, Jan Heinrich Beinke¹, Marc Walterbusch¹, Ricardo Ramos Gameiro², Peter König^{2,3} und Frank Teuteberg¹

Abstract: Durch die dynamische Bereitstellung von Ressourcen bietet Cloud Computing Unternehmen die Möglichkeit Effizienz- und Wettbewerbsvorteile zu realisieren. Aufgrund der mit der Technologie einhergehenden Delegation der Kontrolle über eigene Daten und Services haben viele Unternehmen jedoch Vorbehalte, insbesondere in Hinblick auf Datensicherheits- und Datenschutzaspekte. Durch Vertrauenssignale, wie Zertifikate, Kundenbewertungen, Referenzkunden und vertrauensrelevante Informationen wie zum Beispiel zur Ausfallsicherheit, können Cloud-Anbieter der Informationsasymmetrie entgegenwirken und Vertrauen erzeugen. Um die Wahrnehmung und den Effekt der Platzierung solcher Vertrauenssignale auf Anbieter-Webseiten zu untersuchen, wurde im vorliegenden Beitrag eine Eye-Tracking Studie durchgeführt. Die Ergebnisse der Untersuchung deuten auf eine Bestätigung der positiven Wahrnehmung sowie den positiven Auswirkungen der Vertrauenssignale hin. Hierbei werden sicherheitsbezogene Signale als positiver vertrauensbeeinflussend wahrgenommen als soziale.

Keywords: Cloud Computing, Vertrauen, Eye-Tracking, IT-Sicherheit, Vertrauenssignale

1 Einleitung und Motivation

Das Vertrauensverhältnis zwischen Anbieter und Kunde ist ein entscheidender Faktor bei der Auswahl von Cloud Computing-Dienstleistungen [WT12]. Insbesondere bei der Auslagerung von kritischen bzw. sensiblen Daten und dem damit einhergehenden Kontrollverlust über die Speicherung und Verarbeitung der Daten in der Cloud, ist die Entwicklung und Aufrechterhaltung von Vertrauen zwischen Dienstleister und Kunden essentiell. Aufgrund der im Cloud Computing-Markt herrschenden Informationsasymmetrie suchen Kunden daher nach Qualitätsindikatoren der Anbieter [BH11]. Vertrauenssignale wie Zertifikate können dabei helfen, diese Informationsasymmetrie durch „Signaling“ relevanter Informationen zu überbrücken [WT12]. Da die Informationsgewinnung bzw. Kontaktaufnahme zwischen Kunde und Cloud Service Provider (CSP) häufig über die Webseite des Anbieters stattfindet, spielt die Gestaltung dieser und die Platzierung von Vertrauenssignalen bei der initialen Vertrauensbildung eine besondere Rolle [KH02, WSW04]. Die Einbettung von vertrauensbeeinflussenden Faktoren (VBF)

¹ Universität Osnabrück, Fachgebiet Unternehmensrechnung und Wirtschaftsinformatik, Katharinenstr. 1, 49076 Osnabrück, {michael.adelmeyer; jan.beinke; marc.walterbusch; frank.teuteberg}@uni-osnabrueck.de

² Universität Osnabrück, Institut für Kognitionswissenschaft, Albrechtstraße 28, 49076 Osnabrück, {pkoenig; rramosga}@uni-osnabrueck.de

³ Institut für Neurophysiologie und Pathophysiologie, Universitätsklinikum Hamburg-Eppendorf, Hamburg

auf Webseiten von CSP ist daher bereits gängige Praxis [Ök14]. Einzelne Qualitäts- und Vertrauenssignale wie Zertifikate [LSS13], Service Level Agreements [Ad16] oder Kundenbewertungen [BTH12] wurden bereits hinsichtlich ihrer Auswirkungen auf das Kundenvertrauen untersucht. Dennoch fehlt in diesem Kontext ein Verständnis über die Wirkung im Zusammenspiel von VBF auf Webseiten von CSP. Ziel der Studie ist es daher, die Wahrnehmung von VBF auf Webseiten von CSP zu untersuchen, um so Empfehlungen für die Auswahl und Platzierung sowie Implikationen für die Identifizierung und Bewertung von VBF auf Webseiten zu geben. Dem Beitrag liegt daher folgende Forschungsfrage zugrunde:

Wie stark werden vertrauensbeeinflussende Faktoren auf Webseiten von Cloud Service Providern wahrgenommen und als solche empfunden?

Die Forschungsfrage wird unter Verwendung eines fiktiven CSP untersucht. Dabei werden die Probanden mit der Webseite des fiktiven CSP konfrontiert und die Wahrnehmung der für CSP spezifischen VBF per Eye-Tracking sowie einer anschließenden Umfrage erhoben. Die Ergebnisse zeigen, dass insbesondere sicherheitsrelevante im Vergleich zu sozialen VBF stärker und als vertrauensvoller wahrgenommen werden. In Abschnitt zwei des Beitrags werden zunächst die Grundlagen vom Vertrauen sowie zu vertrauensbeeinflussenden Faktoren dargestellt, welche im Verlauf des Beitrags untersucht werden. Abschnitt drei befasst sich mit der Forschungsmethodik der Eye-Tracking Studie; in Abschnitt vier werden die Ergebnisse analysiert. Abschnitt fünf fasst die wesentlichen Aspekte der Studie zusammen und zeigt Limitationen des Ansatzes auf.

2 Theoretischer Hintergrund

2.1 Vertrauen im Cloud Computing

Trotz einer Vielzahl von Forschungsarbeiten im Bereich des Vertrauens existiert keine einheitliche Definition des Begriffs [WT12]. Im Rahmen der Ausarbeitung wird daher die weit verbreitete Definition von Mayer et al. (1995) als Ausgangspunkt verwendet, die Vertrauen als Verhältnis zwischen Treugeber und Treuhänder definieren, bei dem der Treugeber den Aktionen des Treuhänders unterliegt und somit bewusst durch diese verletzbar ist. Dies geschieht in der Erwartung, dass der Treuhänder eine dem Treugeber wichtige bestimmte Aktion ausführt, ohne dass dieser ihn hinreichend überwachen oder kontrollieren kann [MDS95]. Aufgrund der Kontrolldelegation vom Kunden zum Anbieter im Cloud Computing und der damit einhergehenden Informationsasymmetrie ist das Vertrauensverhältnis von besonderer Bedeutung, insbesondere in Public Clouds [WT12]. Die höheren verbundenen Risiken und die gesteigerte Schwierigkeit für den Nutzer diese einzuschätzen, verstärken zudem die Notwendigkeit von Vertrauen im Cloud Computing im Vergleich zum klassischen E-Commerce [OM13]. Intensiviert wird dieser Aspekt sofern Unternehmen sensible Daten oder ganze Prozesse in die Cloud auslagern, womit sich diese in eine Abhängigkeit von den Handlungen der CSP begeben [WMT13]. Das Vertrauen des Endkunden in den CSP ist folglich ein essentieller Faktor bei der Auswahl

und Nutzung von Cloud-Services [Wa14]. Die Faktoren, die das Vertrauensverhältnis beeinflussen, sind dabei vielfältig und zudem von rechtlichen und technologischen Fragestellungen umgeben [WMT13]. Daher suchen Kunden nach geeigneten vertrauensschaffenden Signalen, die Rückschlüsse auf die Qualität des Anbieters ermöglichen, um somit die asymmetrische Informationsverteilung zu reduzieren und das Anbieterrisiko im Cloud Computing zu überbrücken [WT12]. Im Cloud Computing ist es daher neben technologischen und rechtlichen Aspekten nötig, dass CSP Verantwortung signalisieren. Soziale Mechanismen sind insbesondere dann gefragt, wenn technische Sicherungsmechanismen an ihre Grenzen stoßen [Moe11]. Im Fokus existierender Untersuchungen stehen zumeist die Rolle der Webseite und ihre Auswirkung auf das Vertrauen im Kontext des E-Commerce [KH02, WSW04]. Untersuchungen zur Auswirkung der Gestaltung von Webseiten sind ebenfalls primär im E-Commerce-Kontext einzuordnen [KBH11]. Jedoch existieren hierzu auch Studien im Bereich des Cloud Computings [Ök14], insbesondere zur Wirkung einzelner VBF [Ad16, LSS13, Wa14]. Trotz der besonderen Rolle des Vertrauens im Cloud Computing [OM13] fehlt eine gemeinsame Betrachtung der Wirkung von VBF auf diesem Gebiet. In dieser Studie wird der Fokus dabei auf organisationales Vertrauen, konkret auf das Vertrauensverhältnis zwischen Endkunden und Public Cloud-Anbietern, gelegt.

2.2 Wahrgenommene Sicherheit durch Zertifikate und Ausfallsicherheit

Sicherheitsaspekte und die damit einhergehenden Risiken spielen eine zentrale Rolle bei der Vertrauensbildung und der Auswahl von CSP [WT12]. Wahrgenommene IT-Sicherheitsrisiken werden in diesem Kontext von Ackermann et al. (2012) als die Wahrnehmung von Risiken durch Entscheidungsträger, die die Sicherheit und den Schutz der IT eines Unternehmens im Rahmen einer Auslagerung zu Cloud Computing-Diensten betreffen, definiert [Ac12]. Die wahrgenommene IT-Sicherheit lässt sich dabei in verschiedene Dimensionen unterteilen, nach Ackermann et al. (2012) bspw. weiterführend in Vertraulichkeit, Integrität, Verfügbarkeit, Performanz, Zurechenbarkeit und Wartbarkeit. Die Dimensionierung wird jedoch in der Literatur nicht trennscharf vorgenommen. So wird bspw. die Verfügbarkeit (im Folgenden: Ausfallsicherheit) sowohl als Dimension innerhalb der IT-Sicherheit angesehen [Ac12], als auch als eigenständiger vertrauensbildender Faktor neben der IT-Sicherheit [LSS13, RWZ11]. Die aufgeführte Literatur betont jedoch die Wichtigkeit der Ausfallsicherheit in diesem Kontext. Da andere Dimensionen, wie bspw. die Integrität oder Vertraulichkeit, nur bedingt messbar sind, müssen Kunden alternative VBF zu ihrer Bewertung heranziehen. Zertifikate unabhängiger Institutionen können dazu beitragen, das Vertrauen in die Sicherheit des CSP zu erhöhen und die im Markt herrschende Informationsasymmetrie zu reduzieren, da diese eine Beurteilung der IT-Sicherheitsmaßnahmen eines Anbieters ermöglichen [SS13]. Die direkte Darstellung von Cloud-spezifischen VBF wie Zertifikaten und der Ausfallsicherheit auf der Webseite eines CSP kann somit zu einer höheren wahrgenommenen Sicherheit und somit zu einem erhöhten Vertrauen beim Endkunden führen. Zudem ist dies bereits fester Bestandteil auf den Webseiten von CSP [Ök14], weshalb diese VBF

exemplarisch als direkt analysierbare Faktoren der wahrgenommenen Sicherheit adaptiert wurden. Im Kontext der Studie bezieht sich wahrgenommene Sicherheit folglich darauf, wie stark diese VBF vom Nutzer (als vertrauensvoll) wahrgenommen werden.

2.3 Soziale Präsenz durch Referenzkunden und visuelle Kundenbewertungen

Soziale Präsenz bezeichnet die Wahrnehmung von menschlicher Wärme [SWC76] und wird als Vorstufe von Vertrauen angesehen [Wa14]. Steinbrück et al. (2002) stellen in einer Studie fest, dass die Einbettung von sozialen Signalen („social cues“) wie z. B. Bilder oder Text in den Online-Auftritt eines Anbieters eine effektive Strategie beschreibt, um das Vertrauen des Endkunden in den Anbieter zu erhöhen. Portraits (bspw. Fotos von Unternehmensvertretern) helfen bei der Generierung von sozialer Präsenz, indem sie den unpersönlichen Prozess des E-Commerce in eine vertraute Situation transferieren. Der Endkunde entwickelt eine quasi-soziale Beziehung zu den dargestellten Personen, in einer ansonsten immateriellen bzw. virtuellen Umgebung [St02]. Dennoch kann sich die übermäßige Einbettung von sozialen Signalen konträr auf die Vertrauenswürdigkeit von Anbietern auswirken, da hierdurch die Fähigkeit der Kunden, gute von schlechten Anbietern objektiv zu unterscheiden, verringert wird [RSM03]. Aufgrund der asymmetrischen Informationsverteilung ist der Kunde abhängig von Informationen, die ihm direkt vom Anbieter oder indirekt bspw. durch Bewertungen von Dritten zur Verfügung gestellt werden [WT12]. Das Vertrauen in Kundenbewertungen, welches als interpersonelles Vertrauen und als soziale Komponente anzusehen ist, kann sich dabei positiv auf das Vertrauen in Online-Anbieter auswirken [BTH12]. Als weiterer VBF im Kontext der sozialen Präsenz können Referenzkunden angesehen werden [WMT13, WT12]. Referenzkunden ermöglichen es, Schlüsse über die Marktakzeptanz eines Anbieters zu ziehen. Zudem profitieren CSP vom Bekanntheitsgrad der aufgeführten Referenzkunden [WMT13]. Im Rahmen der Studie bezeichnet soziale Präsenz den Einfluss von Kundenbewertungen (inkl. Fotos) und Referenzkunden auf die Vertrauenswürdigkeit des CSP.

3 Eye-Tracking Studie

3.1 Forschungsmethodik

Die Methode des Eye-Trackings ist ein Verfahren zur Ermittlung der Blickrichtung von Personen. Hierdurch lässt sich bspw. auswerten, wie lange und in welcher Reihenfolge bestimmte Medieninhalte von Rezipienten betrachtet werden. Per Eye-Tracking können die Phasen des relativen Stillstandes der Augen bei der Betrachtung (sog. Fixationen) eines Stimulus erfasst werden [Bl13]. Augenbewegungen werden durch mehrere Faktoren wie bspw. die Auffälligkeit des Reizes (sog. Salienz), räumliche Begebenheiten oder die Aufgabe beeinflusst [Ko10]. Ziel der durchgeführten Eye-Tracking Studie sowie der Umfrage ist es Anhaltspunkte dafür zu bekommen, wie stark VBF (z. B. Zertifikate) auf

Webseiten von CSP fixiert und damit wahrgenommen werden und wie stark diese das Vertrauen und die Einstellung der Probanden zum CSP beeinflussen. Die Studie wurde als Single-Group Post-Test-only Design konzipiert und unter kontrollierten Laborbedingungen durchgeführt. Hierbei gilt es zu beachten, dass diese Versuchsanordnung primär als beschreibend verwendet werden kann. Aufgrund der Limitation durch eine fehlende Kontrollgruppe oder eines Pre-Tests ist eine Validität von kausalen Schlüssen nur bedingt gegeben [Ja15]. Da das Ziel des Eye-Trackings jedoch in der Beschreibung der Wahrnehmung von VBF durch die Probanden liegt, kann diese Versuchsanordnung Anwendung finden. Die Herstellung möglichst natürlicher Situationen ist Voraussetzung, um valide Daten durch die Erfassung von Blickrichtungen zu generieren [Bl13]. Die Probanden wurden daher über eine Vignette (eine Beschreibung einer aus einzelnen Bausteinen bestehenden hypothetischen Situation, innerhalb derer eine Entscheidung zu treffen ist [Wa14]), in eine realitätsnahe Entscheidungssituation versetzt. Die Probanden wurden damit konfrontiert personenbezogene und sensible Daten in der Cloud zu speichern. Anschließend wurde den Teilnehmern ein Stimulus in Form von drei statischen Webseiten-Screenshots („Trials“) des fiktiven CSP „Core Cloud“ (Trial eins: Zertifikate und Ausfallsicherheit, Trial zwei: Kundenbewertungen, Trial drei: Referenzkunden) präsentiert und die Fixationen der Augen mit Hilfe eines Eye-Trackers aufgezeichnet. Abschließend wurden den Teilnehmern im Rahmen des Post-Tests Fragen gestellt, um eine Verknüpfung zwischen den objektiven Metriken (hier: Fixationen) und den subjektiven Einstellungen der Probanden herstellen zu können. Um die Konstrukte „wahrgenommene Sicherheit“ und „soziale Präsenz“ und die korrespondierenden VBF zu messen, wurden geeignete Items aus der Literatur herangezogen und anhand einer fünf-Punkte-Likert-Skala von „stimme voll zu“ bis „stimme überhaupt nicht zu“ erhoben. Die gesammelten Eye-Tracking-Daten wurden aggregiert und in Form von Heatmaps grafisch aufbereitet. Dies ermöglicht eine qualitativ-deskriptive Auswertung und somit eine Aussage darüber, welche Merkmale verstärkt wahrgenommen werden [Bl13].

3.2 Operationeller Aufbau & Implementierung

Zu Beginn des Eye Trackings wurde den Probanden der standardisierte Ablauf der Studie erklärt, um mögliche Irritationen bei der Durchführung zu vermeiden. Des Weiteren erhielten diese eine Einführung in das Thema Cloud Computing, um gleiches Vorwissen bei allen Teilnehmern zu gewährleisten. Im Rahmen der technischen Umsetzung erwies sich eine Betrachtungszeit von 45 Sekunden pro Screenshot als angemessen. Eine zu lange Einführungs- und Durchführungszeit der Studie beeinflusst die Konzentration der Teilnehmer und damit auch die Ergebnisse der Studie negativ. Bei der Erstellung der Screenshots wurde darauf geachtet, dass die einzelnen Bilder nicht überladen sind, um mögliche Reizüberflutungen zu verhindern. Des Weiteren wurden die einzelnen VBF optisch getrennt dargestellt und der Fokus auf die relevanten Inhalte gelegt. Eine deutliche visuelle Trennung von VBF ist bei der Auswertung von Vorteil, da den jeweiligen VBF konkrete auszuwertende Bereiche zugeordnet werden. Aufbau und Inhalt der Webseite wurden anhand verschiedener realer Anbieter abgeleitet. Um den Aufbau der Web-

seite sowie die Auswahl und Platzierung der VBF des fiktiven CSP visuell sowie inhaltlich möglichst realistisch zu gestalten, wurde zudem Literatur herangezogen, die die Nutzung von VBF auf Webseiten thematisiert. Der Fokus lag auf den eingangs identifizierten Merkmalen Referenzkunden und Kundenbewertungen (inkl. Fotos) sowie Zertifikaten und Ausfallsicherheit. Für die Analyse der Webseiten realer CSP, auf deren Basis die Webseite des fiktiven CSP gestaltet wurde, wurde der „Magic Quadrant für Cloud Infrastructure as a Service“ von Gartner aus dem Jahr 2013 herangezogen und die Webseiten der dort aufgeführten Anbieter analysiert [Ök14] (s. Anhang A). Bezüglich der Zertifikate ergab die Analyse, dass diese bei der Mehrheit der Anbieter vorhanden waren, in der Regel sogar mehrfach. Als Zertifikate wurden EuroCloud als Cloud-spezifisches Zertifikat, ISO 27001 als allgemein anerkannter Standard zum Informationssicherheitsmanagement sowie aufgrund der hohen Bekanntheit der Zertifizierungsinstitution der TÜV ausgewählt. Hinsichtlich der Ausfallsicherheit garantiert ein Großteil der untersuchten Anbieter eine Verfügbarkeit von 99,9% oder mehr, sodass diese Zahl für die Studie adaptiert wurde. Des Weiteren gab die Mehrheit der Anbieter Referenzkunden und Kundenbewertungen an. Bei der Darstellung von Referenzkunden wurden die von Amazon Web Services angegebenen Partner übernommen und um weitere bekannte Unternehmen ergänzt. Bei der allgemeinen Gestaltung der Webseite des fiktiven Cloud Service Providers Core Cloud GmbH wurden die „Anforderungen an die Gestaltung von Webseiten“ von Meidl (2013) herangezogen [Me13]. Zudem wurde die Webseite mit neutralen Inhalten ergänzt. Auf dieser Basis wurde eine Referenz-Webseite für die in der Eye-Tracking Studie verwendeten statischen Screenshots erstellt.



Abb. 1: Webseite Trial drei Referenzkunden

Abb. 1 zeigt beispielhaft die statische Webseite des ersten Trials zu den Referenzkunden, die den Probanden während des Eye-Trackings vorgelegt wurde. Der relevante Bereich zu den vertrauensbeeinflussenden Faktoren (hier: Darstellung von Logos der Referenzkunden; zur Anschaulichkeit hier mit roter Umrandung hervorgehoben) ist als sogenannter Area of Interests vom übrigen neutralen Bereich optisch eigenständig. Der neutrale

Bereich der Webseite besteht aus für CSP-Webseiten typischen Inhalten, die sich an im Rahmen der Webseiten-Analyse identifizierten Inhalten orientieren. Obwohl diese Inhalte möglichst neutral gestaltet wurden, ist eine Beeinflussung der Probanden durch diese nicht gänzlich ausgeschlossen. Analog dazu wurden die Webseiten-Screenshots zu Trial eins (Zertifikate und Ausfallsicherheit) und Trial zwei (Kundenbewertungen) gestaltet.

3.3 Durchführung und Validierung

Insgesamt nahmen an der Studie 20 Personen teil. Davon waren zwölf männlich und acht weiblich. Das Durchschnittsalter der Versuchspersonen betrug 23,15 Jahre. Alle Teilnehmer waren Studierende an der Universität Osnabrück. Zur Ermittlung der Blickrichtung der Probanden wurde ein kopfmontierter Eye-Tracker⁴ verwendet. Eine Kalibrierung des Geräts wurde vor jedem Durchgang durchgeführt, um die technische Gültigkeit der Ergebnisse sicherzustellen [B113]. Die Validierung der Datensätze als auch die spätere Auswertung fanden unter Anwendung der Software MATLAB statt. Der üblichen Praxis folgend wurden dabei zu kurze Fixationen (< 50 ms), Fixationen, deren Dauer mehr als zwei Standardabweichungen über dem Durchschnitt lagen, Fixationen außerhalb des Bildschirms sowie die jeweils erste Fixation jeder Versuchsperson bei jedem Screenshot herausgefiltert [Ka13]. Die Datensätze des Post-Tests wurden hinsichtlich der von den Probanden benötigten Zeit sowie bestimmter Antwortmuster validiert. Insgesamt konnten alle Datensätze als vollständig und gültig bewertet werden.

4 Analyse der Ergebnisse

Bei der Auswertung und Interpretation der Eye-Tracking Daten gibt es verschiedene Möglichkeiten [B113]. Insbesondere bei der Verknüpfung der Eye-Tracking Daten mit den Ergebnissen der Umfrage entstehen aufgrund der Versuchsanordnung Interpretationsspielräume. Die hier gelieferten Ergebnisse sollen daher primär die Wahrnehmung von VBF durch die Probanden beschreiben und Erklärungsmöglichkeiten bzw. -ansätze aufzeigen. Bei der Auswertung der Screenshots wurden zunächst verschiedene Areas of Interests analog zu den VBF definiert, die bereits während der Implementierung optisch abgegrenzt wurden. Anschließend wurden alle Fixationen, die sich in diesem Bereich befanden, summiert. Als Auswertungsmethode für die Betrachtung der Webseiten wurden Heatmaps verwendet [Du07]. Heatmaps als Ausprägung sogenannter „Attention-Maps“ visualisieren die Betrachtungsintensität unterschiedlicher Bereiche eines Stimulus [B113]. Bei ihrer Verwendung gilt es jedoch zu beachten, dass Schlüsse bzw. Empfehlungen rein auf Basis solcher Auswertungen nur bedingte Validität haben und diese sich daher primär für die Beschreibung der Aufmerksamkeitsverteilung von Probanden eignen [Bo09]. Nach der Validierung der durch das Eye-Tracking generierten Rohdaten

⁴ EyeLink II von SR Research Ltd.; Samplerate: 500Hz; Genauigkeit: < 0.5°, Distanz: 80cm

wurde ein Auswertungsskript zur Erstellung einer Dichtefunktion der Fixationen der Probanden erstellt. Diese wurde anschließend über den jeweiligen Screenshot (Trial) gelegt, um die aggregierten Daten grafisch darzustellen [SM07].



Abb. 2: Trial 1, Zeit 45 Sekunden, alle Versuchspersonen

Abb. 2 umfasst die Datensätze aller Versuchspersonen über die gesamte Dauer des ersten Trials. Es lässt sich erkennen, dass die Probanden die neutralen Elemente sowie die VBF (Zertifikate, vertrauensrelevante Informationen zur Ausfallsicherheit) der Webseite intensiv wahrnehmen. Die Analyse des intuitiven Verhaltens der Versuchspersonen anhand der Darstellung der ersten fünf Sekunden über alle Datensätze für Trial eins führt zu ähnlichen Ergebnissen (Anhang. B). Sowohl neutrale Elemente (Logo, allgemeine Informationen zum CSP) als auch die VBF werden zu Beginn intensiv betrachtet.

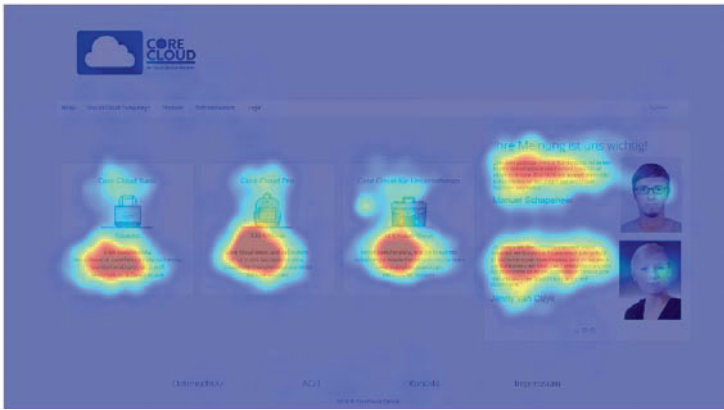


Abb. 3: Trial 2, Zeit 45 Sekunden, alle Versuchspersonen

Analog dazu lassen sich die Heatmaps zu Trial zwei und drei interpretieren. Es ist zu erkennen, dass der Text der Kundenbewertungen bzw. der neutralen Elemente bei der

Betrachtung im Vordergrund stehen (vgl. Abb. 3). Dies ist unter anderem mit dem höheren Workload beim Erfassen von Text zu erklären, dessen Rezeption entsprechend mehr Fixationen als bspw. bei Bildern benötigt [B113]. Bei der Analyse der ersten fünf Sekunden wird deutlich, dass insbesondere neutrale Bilder und die dargestellten Fotos und Bewertungen der Kunden zu Beginn im Fokus der Betrachtung lagen (Anhang C). Bei der Auswertung von Trial drei wird die Auswirkung der Verringerung des Betrachtungszeitraumes deutlich. Über den gesamten Auswertungszeitraum wirkt es, als würden die Referenzkunden kaum wahrgenommen (Anhang D). Wird der Betrachtungszeitraum auf fünf Sekunden reduziert (Abb. 4), lässt sich erkennen, dass die Referenzkunden initial intensiv fixiert werden. Erst dann wird damit begonnen den Text zu lesen, was ebenfalls auf den höheren Workload zur Erfassung des Textes zurückzuführen ist.



Abb. 4: Trial 3, Zeit 5 Sekunden, alle Versuchspersonen

Da Blickaufzeichnungen bzw. Heatmaps hinsichtlich zugrunde liegender Motive bzw. subjektiver Bewertungen wenig aussagekräftig sind, sollte ein Mehrmethoden-Design wie bspw. in Form einer Befragung (hier: Post-Test durch Umfrage) herangezogen werden [B113, Bo09]. Der Mehrwert einer kombinierten Auswertung gegenüber einer reinen Umfrage liegt in der Erhebung objektiver Metriken, welche der subjektiven Einschätzungen der Probanden aus der Umfrage gegenübergestellt wird. Werden bspw. VBF als subjektiv vertrauensbildend empfunden, so empfiehlt sich hierdurch nicht automatisch eine vorrangige Platzierung auf Webseiten. So kann bspw. der Workload zur Erfassung eines VBF in Relation zu anderen, in der Umfrage subjektiv weniger stark gewichteten Faktoren, zu hoch sein. Aufgrund der Konzeption der Versuchsanordnung als Single-Group Post-Test-only Design können auf Basis der Daten- bzw. Methodentriangulation lediglich Indikatoren und nur beschränkt Kausalitäten für bestimmte Verhaltensweisen und Einstellungen der Probanden geschlossen werden. Aufgrund der geringen Teilnehmerzahl von 20 Probanden wurden primär deskriptivstatistische Auswertungsmethoden bei der Analyse der Umfrage herangezogen (s. Tab. 1).

Konstrukt	VBF	Mittelwert (arithmetisch)	Anzahl Probanden	Standardab- weichung (σ)
Wahrgenommene Sicherheit	<i>Zertifikate</i>	3,950	20	1,202
Wahrgenommene Sicherheit	<i>Ausfallsicherheit</i>	3,725	20	0,866
Soziale Präsenz	<i>Referenzkunden</i>	3,325	20	1,228
Soziale Präsenz	<i>Kundenbewertungen</i>	2,825	20	0,936

Tab. 1: Auswertung der Konstrukte und Items der Umfrage

Die Ergebnisse der Wahrnehmung der VBF aus der Umfrage in Kombination mit dem Eye-Tracking lässt sich in folgendem Ranking darstellen (s. Tab. 2, s. Anhang E).

Konstrukt	VBF	Summe Fixationen	Mittelwert Umfrage
Soziale Präsenz	<i>Kundenbewertungen</i>	1247	2,825
Wahrgenommene Sicherheit	<i>Zertifikate</i>	896	3,950
Wahrgenommene Sicherheit	<i>Ausfallsicherheit</i>	583	3,725
Soziale Präsenz	<i>Referenzkunden</i>	424	3,325

Tab. 2: Ranking zur Wahrnehmung VBF aus Eye-Tracking und Umfrage

Die Gegenüberstellung der Mittelwerte der Einschätzung der VBF aus der Umfrage mit den Summen der absoluten Fixationen in den entsprechenden Areas of Interests aus dem Eye-Tracking ergibt verschiedene Interpretationsmöglichkeiten. Kundenbewertungen und Zertifikate wurden durch die Versuchspersonen am häufigsten betrachtet. Die Kundenbewertungen wurden innerhalb der Umfragen jedoch als weniger relevant bewertet. Dies macht die Defizite bei der Auswertung der absoluten Fixationsanzahl deutlich. Aufgrund des unterschiedlichen Workloads zur Erfassung der VBF weichen subjektive Bewertung und Fixationsanzahl zu einem gewissen Grad voneinander ab. Für die Gestaltung von Webseiten von CSP lässt sich folgern, dass Kundenbewertungen subjektiv als weniger vertrauensbildend empfunden werden und zudem aufgrund des Textes einen hohen Workload bei der Rezeption erfordern. Daher sollten diese von CSP eher nachrangig auf Webseiten platziert werden. Zertifikate sowie die Ausfallsicherheit werden in der Umfrage von den Probanden als ähnlich positiv vertrauensbeeinflussend wahrgenommen und relativ häufig fixiert. Diese sollten von CSP daher entsprechend vorrangig (gut wahrnehmbar) auf Webseiten platziert werden. In diesem Kontext lassen sich besonders die Zertifikate herausstellen, da diese als positiv eingeschätzt wurden und einen relativ geringen Workload zur Erfassung benötigen und dennoch relativ häufig fixiert wurden. Referenzkunden werden als moderat vertrauensbeeinflussend wahrgenommen und relativ wenig fixiert, daher sollten diese ebenfalls nachrangig platziert werden.

5 Fazit und Diskussion

5.1 Zusammenfassung und Diskussion

Die Ergebnisse der Studie weisen darauf hin, dass bestimmte VBF wie Zertifikate auf Webseiten von CSP im Vergleich von den Probanden als positiver vertrauensbeeinflus-

send und verstärkter wahrgenommen werden. Hierbei ist zu beachten, dass die zur Auswertung des Eye-Trackings verwendeten Heatmaps auf Basis der Fixationsanzahlen Spielraum für subjektive Interpretationen zulassen [B113]. Weitere auswertbare objektive Metriken neben der Fixationsanzahl sind bspw. Fixationsdauer oder Anzahl der Fixationswechsel [Du07]. Dennoch lässt sich in Kombination der Ergebnisse des Eye-Trackings mit der Auswertung der Umfrage zusammenfassend festhalten, dass VBF auf Webseiten von CSP eine zentrale Rolle bei der Auswahl von CSP spielen. Zudem können auf dieser Basis Implikationen zur Platzierung von VBF auf Webseiten von CSP abgeleitet werden. Sicherheitsrelevante VBF (Zertifikate und Ausfallsicherheit) des Providers wurden von den Probanden als positiver vertrauensbeeinflussend eingeschätzt als soziale (Kundenbewertungen und Referenzkunden), weshalb diese Elemente bei der Gestaltung von Webseiten von CSP vorrangig platziert werden sollten. Als weitere Implikation lässt sich ableiten, dass CSP bei der Gestaltung von Webseiten zur Vertrauensbildung nicht nur subjektive Größen, sondern vielmehr auch objektive Metriken berücksichtigen sollten. Hierbei sind insbesondere VBF mit einem relativ geringen Workload bei der Erfassung und relativ hoher positiver vertrauensbeeinflussender Wirkung von Interesse. Zudem sollte die Gestaltung von Webseiten den Nutzer nicht überfordern und VBF in Relation zu neutralem Inhalt angemessen positioniert werden.

Der Aufwand für Eye-Tracking-Studien ist vergleichsweise hoch, insbesondere durch die fortlaufende Kalibrierung (Driftkorrektur) der Geräte sowie durch den nötigen Übungsaufwand. Zudem beeinflussen Störfaktoren wie bspw. Make-up an den Augen oder Anspannung der Probanden die Messgenauigkeit. Bei der Auswertung der Eye-Tracking Ergebnisse erwies sich die angesprochene Interpretationsoffenheit der Ergebnisse als problematisch. [B113, Du07]. Die erhobenen Messgrößen können sowohl auf Stimuluseigenschaften (Komplexität) als auch auf Rezipienteneigenschaften (Interesse) hindeuten, weshalb bspw. die Auswertung der Fixationsanzahlen nur bedingt Aussagen über die Wirkung auf Rezipienten zulässt. Zudem sind die Berücksichtigung von Kontextmerkmalen als auch der subjektiven Wahrnehmung notwendig, um auf die inhaltliche Bedeutung von Messgrößen schließen zu können [B113]. Vielmehr sollten im Sinne einer Daten- bzw. Methoden-Triangulation verschiedene Analysemöglichkeiten bzw. Messgrößen der Blickbeobachtung in Betracht gezogen werden. Die dem Eye-Tracking angeschlossene Umfrage ermöglicht es, Teilen dieser Problematik beizukommen, dennoch ist eine Ausweitung auf zusätzliche Methoden bzw. Daten durch Triangulation denkbar, z.B. durch neurowissenschaftliche Methoden [RJ12].

5.2 Limitationen und Ausblick

Der Aufbau der Studie als Single-Group Post-Test-only Design birgt Gefahren für die Validität der Ergebnisse, weshalb sowohl die Ergebnisse als auch die Auswertung des Eye-Trackings und der Umfrage primär deskriptiven Charakter haben [Ja15]. Eine Versuchsanordnung analog eines Experimentaldesigns erhöht die interne als auch externe Validität der Ergebnisse [Ja15]. Auf Basis dieser Studie kann anhand unterschiedlicher

Stimuli in verschiedenen Gruppen die Wirkung der VBF auf das Vertrauen und die Nutzungsabsicht ermittelt werden. Als weitere Limitationen des Eye-Trackings lassen sich die Aspekte Teilnehmerzahl und die ungewohnte Versuchsumgebung nennen. Zudem konnten aufgrund der soft- und hardwarebedingten Einschränkungen lediglich statische Bilder untersucht werden. Ferner werden zur Untersuchung solch statischer Elemente Remote Eye-Tracking Systeme als geeigneter angesehen [B113]. Bei der Wahrnehmung der Webseite beeinflussen zudem neutrale Webseiteninhalte die Stärke der Wahrnehmung der VBF, weshalb Einflüsse durch die Gestaltung der Webseite auf die Eye-Tracking Ergebnisse unterstellt werden müssen. Hinzu kommt, dass der Fokus in diesem Rahmen auf zuvor definierten VBF lag. Durch die Analyse realer Webseiten von CSP kann dieser eingeschränkten Betrachtung entgegengewirkt werden. Die in der Literatur durchaus kontrovers diskutierte Verwendung von Studenten als Zielgruppe für Studien stellt eine weitere Limitation dar [Co12]. Da Studenten jedoch als Early Adopter einer innovativen Technologie wie Cloud Computing sowie als erfahrene Internetnutzer anzusehen sind, ist die Verwendung von Studenten als Probanden im Kontext dieser Studie angemessen [GPF01]. Aufgrund der Gestaltung der Untersuchung und den teilweise Cloud-spezifischen Faktoren ist eine Übertragbarkeit der Ergebnisse, bspw. auf andere Bereiche des E-Commerce, zudem nur bedingt gegeben.

Basierend auf den Ergebnissen der Studie ergeben sich Ansätze für zukünftige Forschung. Die beleuchtete Kundenseite sollte um Aspekte der Anbieterseite erweitert werden. In diesem Rahmen kann untersucht werden, warum Anbieter bestimmte VBF auf Webseiten platzieren, so haben kleinere oder mittlere CSP bspw. oft nicht das nötige Budget, um sich entsprechend zertifizieren zu lassen [SS13]. Offen ist, warum einzelne VBF stärker wahrgenommen werden und wie sich der tatsächliche Einfluss dieser auf das Vertrauen und die Nutzungsabsicht darstellt. Weiter sind die in dieser Studie betrachteten Konstrukte und VBF exemplarisch und können in weiteren Untersuchungen ergänzt werden. Diese können anschließend bspw. in Strukturgleichungsmodelle zum Schätzen und Testen korrelativer Zusammenhänge überführt werden.

Danksagung

Die Autoren danken den Teilnehmern der Studie, den Projektmitgliedern, explizit Christian Ahlers, Daniel Bender, Maximilian Garske und Julian Lang für ihre Mitarbeit sowie den Gutachtern für ihr substanzielles Feedback.

Literaturverzeichnis

- [Ac12] Ackermann, T. et al.: Perceived IT security risks of cloud computing: conceptualization and scale development. In: Proceedings of the International Conference on Information Systems (ICIS), 2012.
- [Ad16] Adelmeyer, M. et al.: Does the Augmentation of Service Level Agreements affect User

- Decisions in Cloud Adoption Scenarios? - An experimental Approach. Erscheint in: Proceedings of the European Conference on Information Systems (ECIS), 2016.
- [BH11] Benlian, A.; Hess, T.: The Signaling Role of IT Features in Influencing Trust and Participation in Online Communities. *International Journal of Electronic Commerce* 15/4, S. 7–56, 2011.
- [BI13] Blake, C.: *Eye-Tracking: Grundlagen und Anwendungsfelder*. Handbuch standardisierte Erhebungsverfahren in der Kommunikationswissenschaft, S. 367–387, 2013.
- [Bo09] Bojko, A. A.: Informative or Misleading? Heatmaps Deconstructed. In: Proceedings of the Human-Computer Interaction International Conference (HCII), 2009.
- [BTH12] Benlian, A.; Titah, R.; Hess, T.: Differential effects of provider recommendations and consumer reviews in e-commerce transactions: an experimental study. In: *Journal of Management Information Systems* 29/1, S. 237–272, 2012.
- [Co12] Compeau, D. et al.: Research Commentary: Generalizability of Information Systems Research Using Student Subjects - A Reflection on our Practices and Recommendations for Future Research. In: *Information Systems Research* 23/4, S. 1093–1109, 2012.
- [Du07] Duchowski, A.: *Eye tracking methodology: Theory and practice*. Springer Verlag London, 2007.
- [GPF01] Gallagher, K.; Parsons, J.; Foster, K.: A Tale of Two Studies: Replicating “Advertising Effectiveness and Content Evaluation in Print and on the Web”. In: *Journal of Advertising Research* 41/4, S. 71–81, 2001.
- [Ja15] Jackson, S.: *Research methods and statistics: A critical thinking approach*. Cengage Learning Boston, 2015.
- [Ka13] Kaspar, K. et al.: Emotions’ Impact on Viewing Behavior under Natural Conditions. In: *PLoS ONE* 8/1, 2013.
- [KBH11] Karimov, F. P., Brengman, M., Hove, L. van: The effect of website design dimensions on initial trust: a synthesis of the empirical literature. In: *Journal of Electronic Commerce Research* 12/4, S. 272–301, 2011.
- [KH02] Koufaris, M.; Hampton-Sosa, W.: *Customer trust online: examining the role of the experience with the Web-site*. Department of Statistics and Computer Information Systems Working Paper, Baruch College, New York, 2002.
- [Ko10] Kollmorgen, S. et al.: Influence of low-level stimulus features, task dependent factors, and spatial biases on overt visual attention. In: *PLoS Comput Biol* 6/5, 2010.
- [LSS13] Lansing, J.; Schneider, S.; Sunyaev, A.: Cloud Service Certifications: Measuring Consumers’ Preferences for Assurances. In: Proceedings of the European Conference on Information Systems (ECIS), 2013.
- [MDS95] Mayer, R. C.; Davis, J. H.; Schoorman, F. D.: An Integrative Model of Organizational Trust. In: *Academy of Management Review*, 20/3, S. 709–734, 1995.
- [Me13] Meidl, O.: *Global Website: Webdesign im internationalen Umfeld*. Springer

- Fachmedien Wiesbaden, 2013.
- [Moe11] Möllering, G.: Vernebeltes Vertrauen? Cloud Computing aus Sicht der Vertrauensforschung. In: *Trust in IT*. Springer Berlin Heidelberg, 39–47, 2011.
- [Ök14] Öksüz, A. et al.: Designing Trust in Websites-an Evaluation of Leading Infrastructure as a Service Providers' Websites. In: *Proceedings of the Pacific Asia Conference on Information Systems (PACIS)*, 2014.
- [OM13] Ouedraogo, M., Mouratidis, H.: Selecting a Cloud Service Provider in the age of cybercrime. In: *Computers & Security* 38, S. 3–13, 2013.
- [RJ12] Riedl, R.; Javor, A.: The biology of trust: Integrating evidence from genetics, endocrinology, and functional brain imaging. In: *Journal of Neuroscience, Psychology, and Economics* 5/2, S. 63, 2012.
- [RSM03] Riegelsberger, J.; Sasse, M.; McCarthy, J.: Shiny happy people building trust?: photos on e-commerce websites and consumer trust. In: *CHI2003*, S. 121–128, 2003.
- [RWZ11] Repschläger, J.; Wind, S.; Zarnekow, R.: Klassifikationsrahmen für die Anbietersauswahl in der Cloud. In: *Lecture Notes in Informatics*, 2011.
- [SM07] Špakov, O.; Miniotas, D.: Visualization of Eye Gaze Data using Heat Maps. In: *Electronics and Electrical Engineering* 2/74, S. 55–58, 2007.
- [SS13] Sunyaev, A.; Schneider, S.: Cloud Services Certification. In: *Communications of the ACM* 56/2, S. 33–36, 2013.
- [St02] Steinbrück, U. et al.: A Picture Says More Than A Thousand Words - Photographs As Trust Builders in E-Commerce Websites. In: *CHI2002*, S. 748–749, 2002.
- [SWC76] Short, J.; Williams, E.; Christie, B.: *The social psychology of telecommunications*. Wiley London, 1976.
- [Wa14] Walter, N. et al.: “May I help You?” Increasing Trust in Cloud Computing Providers through Social Presence and the Reduction of Information Overload. In: *Proceedings of the International Conference on Information Systems (ICIS)*, 2014.
- [WMT13] Walterbusch, M.; Martens, B.; Teuteberg, F.: Exploring Trust In Cloud Computing : A Multi- Method Approach. In: *Proceedings of the European Conference on Information Systems (ECIS)*, 2013.
- [WSW04] Wakefield, R. L.; Stocks, M. H.; Wilder, W. M.: The role of web site characteristics in initial trust formation. In: *Journal of Computer Information Systems* 45/1, S. 94–103, 2004.
- [WT12] Walterbusch, M.; Teuteberg, F.: Vertrauen im Cloud Computing. In: *HMD - Praxis der Wirtschaftsinformatik* 49/6, S. 50–59, 2012.

Anhang

Die Anhänge zum Beitrag können unter folgender URL als PDF abgerufen werden:

<http://bit.ly/Informatik2016>

Hochschule 2026

Hochschule 2026

Ulrike Lucke¹, Markus von der Heyde², Wolfgang Meyer³

Die Digitalisierung stellt moderne Bildungseinrichtungen in ihrer Entwicklung beständig vor technische, organisatorische und inhaltliche Herausforderungen: leistungsfähige und allgegenwärtige IT-Dienste, Mikro- und Makromobilität von Studierenden, Mitarbeitern und Gästen, dynamische Allokation von personalisierten Inhalten, flexible Gestaltung und Organisation von kooperativen und individuellen Lern- und Arbeitsprozessen – und das bei meist engen Ressourcen und regulativen Rahmenbedingungen. Der Workshop thematisiert diese Herausforderungen und stellt sie zur Diskussion.

Aus den eingereichten Beiträgen wurden durch ein Peer-Review sieben Beiträge zur Publikation im Tagungsband ausgewählt. Das Spektrum umfasst auch in diesem Jahr konkrete IT-Dienste und strategische Managementansätze, sowohl hochschulintern als auch im Umfeld der Hochschulen. Die Präsentationen im Workshop werden ergänzt durch einen ersten Einblick in die Ergebnisse der diesjährigen CIO-Studie.

Wir danken den GI-Fachgruppen Appliances, Personal and Smart Devices (APS) und E-Learning (ELE) für die Unterstützung des Workshops. Für die sorgfältige und engagierte Organisation des Workshops danken wir ganz herzlich Raphael Zender und Francis Zinke. Zudem gebührt den Mitgliedern des Programmkomitees für die Begutachtung der eingereichten Beiträge unser Dank:

Arndt Bode (LRZ München)
Stefan Fischer (Uni Lübeck)
Christoph Igel (DFKI)
May-Britt Kallenrode (Uni Osnabrück)
Frank Klapper (Uni Bielefeld)
Bernd Krämer (FU Hagen)
Gudrun Oevel (Uni Paderborn)
Christa Radloff (Uni Rostock)
Ulrik Schroeder (RWTH Aachen)
Ramin Yahyapour (GWDG)

Christian Erfurth (FH Jena)
Marc Göcks (Multimedia Kontor Hamburg)
Odej Kao (TU Berlin)
Udo Keschull (Uni Frankfurt)
Michael Koch (UniBW München)
Sören Lorenz (Uni Kiel)
Hans Pongratz (TU München)
Peter Schirnbacher (HU Berlin)
Bernd E. Wolfinger (Uni Hamburg)
Markus Zahn (Uni Augsburg)

¹ Universität Potsdam, Institut für Informatik, A.-Bebel-Str. 89, 14482 Potsdam, ulrike.lucke@uni-potsdam.de

² vdH-IT, Rosenweg 7, 99425 Weimar, info@vdh-it.de

³ TU Wien, Zentraler Informatikdienst, Wiedner Hauptstraße 8-10 / 020, 1040 Wien, meyer@zid.tuwien.ac.at

Eine wirtschaftsinformatische Standortbestimmung von Forschungsinformationssystemen

Sebastian Herwig¹, Stefan Schlattmann¹

Abstract: Forschungsinformationssysteme erfahren vor dem Hintergrund der stetig zunehmenden Berichts- und Auskunftspflichten von Hochschulen im Bereich der Forschung wachsende Relevanz. Die in diesem Zusammenhang verstandenen Systemlösungen unterscheiden sich jedoch etwa hinsichtlich der Ausgestaltungsform und Zweckorientierung teils deutlich. Dies ist u.a. auf ein fehlendes allgemeingültiges definitorisches Begriffsverständnis zurückzuführen. Unter Berücksichtigung des in der Wirtschaftsinformatik etablierten Begriffsverständnisses eines Informationssystems werden Wesensmerkmale von Forschungsinformationssystemen herausgearbeitet, um auf diesem Weg zur Schaffung eines einheitlichen Begriffsverständnisses beizutragen.

Keywords: Forschungsinformationssysteme, CRIS, FIS, Hochschul-IT, Forschungsberichterstattung, Forschungsdokumentation, Forschungsportal, Forschungsdatenbank

1 Einleitung

An welchen Themen wird gegenwärtig an der Forschungseinrichtung geforscht? Welche Forschungsprojekte werden von den einzelnen Disziplinen durchgeführt und von wem werden diese gefördert? Welche internationalen Forschungsoperationen unterhält die eigene Einrichtung mit Einrichtungen aus bestimmten Ländern? Für welche Forschungsaktivitäten wurden Angehörige der eigenen Einrichtung mit welchen Preisen ausgezeichnet? Welche Aktivitäten werden zur Förderung des wissenschaftlichen Nachwuchses unternommen? Aufgrund stetig zunehmender Berichts- und Auskunftspflichten im Bereich der Forschung sehen sich Hochschulen und Forschungseinrichtungen in zunehmendem und vor allem wiederkehrendem Maße mit derartigen und verschiedenen weiteren Fragestellungen konfrontiert. Um hierauf mit vertretbarem Aufwand fundierte Antworten und Auskünfte geben zu können, bedarf es eines aktuellen und detaillierten Selbstbildes hinsichtlich der eigenen Forschungsaktivitäten und -ergebnisse.

Vor diesem Hintergrund rücken Fragen eines integrierten Informationsmanagements zu Forschungsaktivitäten sowie der Verfügbarkeit geeigneter Werkzeugunterstützung für die Forschungsberichterstattung zunehmenden in den Mittelpunkt der Betrachtung von Hochschulen und Forschungseinrichtungen. Sowohl national als auch international werden daher in zunehmender Zahl an Hochschulen und Forschungseinrichtungen Forschungsinformationssysteme betrieben, eingeführt oder zumindest geplant. [RDM16], [St14] Die hierbei verstandenen Systemlösungen unterscheiden sich u. a. hinsichtlich der

¹ Westfälische Wilhelms-Universität Münster, Abteilung 6.4 – Forschungsinformationen und Forschungsberichterstattung, Schlossplatz 2 48149 Münster, [sebastian.herwig | stefan.schlattmann]@uni-muenster.de

Ausgestaltungsform und Zweckorientierung teils deutlich. Die zu diesem Thema bestehende Literatur beschränkt sich jedoch lediglich auf eine exemplarische Begriffsbestimmung und lässt ein allgemeingültiges definitorisches Begriffsverständnis sowie eine Systematisierung wesentlicher Gestaltungsdimensionen von Forschungsinformationssystemen (FIS) vermissen.

Die Zielsetzung dieses Beitrags besteht daher darin, eine Systematik grundlegender Wesensmerkmale von FIS herauszuarbeiten und auf diesem Weg die Ausprägung eines einheitlichen Begriffsverständnisses anzustoßen und zu befördern. Als theoretischer Bezugsrahmen wird hierzu das in der Wirtschaftsinformatik allgemein etablierte Begriffsverständnis von Informationssystemen herangezogen, um bereits in der Literatur bestehende Begriffsverständnisse und Wesensmerkmale einzuordnen und hinsichtlich der begrifflichen Abdeckung zu untersuchen. Die literaturgestützte Ableitung von Charakteristika wird durch praxisbezogene Erfahrungen angereichert.

2 Begriffliche Einordnung

2.1 Informationssysteme aus Sicht der Wirtschaftsinformatik

Ausgehend vom Begriffsverständnis nach [Te99] werden Informationssysteme als sozio-technische Systeme verstanden, die durch die drei Wesensmerkmale Mensch, Aufgabe und Technik geprägt werden. Ausgerichtet auf die Bewältigung betriebswirtschaftlicher Aufgabenstellungen umfasst ein Informationssystem die hierzu notwendigen technischen wie organisatorischen Mittel und Strukturen. Die notwendigen informationstechnischen Mittel – auch als Anwendungssystem bezeichnet – umfassen hierbei neben der Hardwaretechnik und Basissoftware vor allem anwendungsbezogene Softwaresysteme. Als weiterer Bestandteil umfasst das soziale System sowohl die für die Aufgabenabwicklung verantwortlichen Menschen als Aufgabenträger als auch das hierzu notwendigen organisatorische Umfeld in Form von aufbau- und ablauforganisatorischen Strukturen.

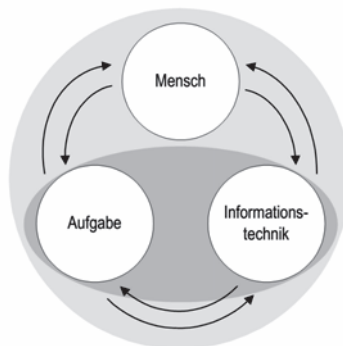


Abb. 1: Komponenten eines Informationssystems (entnommen aus [Te99])

Diese Grundbestandteile beeinflussen unter Berücksichtigung wechselseitiger Interdependenzen gleichermaßen die Gestalt eines Informationssystems (siehe Abb. 1). So lassen sich von der Aufgabenstellung Anforderungen an die erforderlichen Leistungen der Informationstechnik ableiten. Gleichzeitig wird diese maßgeblich von den involvierten Menschen sowie dem zugrundeliegenden organisatorischem Umfeld beeinflusst. Die Informationstechnik wirkt insofern auf die Gestalt eines Informationssystems ein, als dass sich Potentiale und Restriktionen für Umsetzung der Aufgabenerfüllung ergeben. Infolgedessen erfordert die Betrachtung jeglicher aufgabenspezifischer Ausprägungen von Informationssystemen eine explizite und wechselseitige Berücksichtigung sowohl der technischen als auch sozialen Systemkomponenten.

2.2 Bestehendes Begriffsverständnis von Forschungsinformationssystemen

In der gegenwärtigen Literatur werden Forschungsinformationssysteme (FIS) vielmehr exemplarisch beschrieben, ein klar abgegrenztes definitorisches Begriffsverständnis ist hingegen nicht zu finden. Hierbei werden gemeinhin unter einem Forschungsinformationssystem eine spezialisierte Datenbank oder ein Informationssystem verstanden, mit dessen Hilfe Informationen zu Forschungsaktivitäten und -ergebnissen erhoben, verwaltet und bereitgestellt werden. [Af15], [JA06], [Ho06] Die hierbei betrachteten Informationen stellen Metadaten über Forschungsaktivitäten wie bspw. Titel und Laufzeit von Forschungsprojekten oder Titel, Autoren, Veröffentlichungsjahr von Publikationen sowie vergleichbare Angaben zu Preisverleihungen, Promotionen, Habilitationen, Patenten, Forschungsinfrastrukturen sowie zu Personen und Organisationseinheiten, die diese Aktivitäten ausführen, dar und werden als Forschungsinformationen bezeichnet. Als grundlegendes Anliegen von Forschungsinformationssystemen wird die Harmonisierung und Zusammenführung von bereits bestehenden dezentralen Beständen von Forschungsinformationen herausgestellt, um diese für unterschiedliche Zwecke zugänglich und (nach)nutzbar zu machen. [Af15], [HLTB12] Ausgehend von der Vielzahl intendierter Verwendungszwecke werden im Wesentlichen drei Typen von Forschungsinformationssystemen unterschieden (vgl. auch im Folgenden [Af15]):

- **Dokumentations- und Nachweissysteme** verfolgen zumeist den Zweck der Dokumentation von durchgeführten Forschungsaktivitäten bzw. hervorgegangenen Ergebnissen. Hierbei umfassen diese Systeme überwiegend Metainformationen und bleiben auf einzelne ausgewählte Arten von Forschungsaktivitäten beschränkt. Als Beispiele hierfür lassen sich Hochschulbibliographien sowie Projektdatenbanken bzw. -portale nennen. Eine Verknüpfung zwischen den einzelnen Nachweissystemen und damit von den unterschiedlichen Forschungsaktivitäten zu einer personen-, organisations- oder themenbezogenen Gesamtschau auf die Forschungslandschaft wird zumeist nicht verfolgt.
- **Forschungsprofildienste** beschreiben Systeme oder Plattformen, die im Wesentlichen dazu dienen, Forschungsaktivitäten und -ergebnisse einzelner Personen, Einrichtungen, Communities oder Themengebiete im Sinne von Forschungsprofilen zusammenzufassen und so zu kommunizieren und zu verbreiten. Beispielhaft hier-

für lassen sich Forschungsportale oder Wissenschaftliche Netzwerkplattformen wie ResearchGate oder Academia.edu bzw. als eigenständige Softwarelösung VIVO nennen, die vorrangig der Kommunikation von Forschungsaktivitäten gegenüber der breiten Öffentlichkeit dienen. Zur Datengewinnung greifen diese Profildienste überwiegend auf öffentlich verfügbare Informationsbestände zurück, führen diese teils automatisiert personen-, einrichtungs- oder themenbezogenen zusammen und stellen diese wiederum der Öffentlichkeit bereit. Diese Art von Systemen folgt damit der Open Data-Philosophie (siehe hierzu bspw. [He12]).

- **Integrierte Forschungsinformationssysteme** haben die Erhebung, Verwaltung sowie Bereitstellung und Nutzung von Forschungsinformationen entlang administrativer Prozesse zwischen Wissenschaft und Verwaltung im Fokus. Hierbei folgen diese Systeme einem integrativen Gedanken, indem diese auf unterschiedliche Arten von Verwaltungsinformationen wie bspw. finanzielle Informationen zu Drittmittelprojekten oder organisatorische Informationen über das Personal und dessen Bezug zur Organisationsstruktur zurückgreifen, diese mit weiteren Informationen zu Forschungsaktivitäten verknüpfen bzw. anreichern und so in einen inhaltlich-forschungsbezogenen Kontext setzen. Um dies zu ermöglichen, fassen integrierte FIS ein breites Spektrum von Forschungsinformationen in einer einheitlichen Strukturierungsweise in einem oder mehreren Anwendungssystemen zusammen und erlauben es unterschiedliche Forschungsaktivitäten miteinander in Beziehung zu setzen. Integrierte FIS dienen folglich nicht nur einem singulären Zweck, sondern sind unter dem Leitmotto „einmal erfassen, mehrfach nutzen“ auf die Abwicklung vielfältiger Anwendungsfälle ausgerichtet.

3 **Forschungsinformationssysteme als Informationssysteme**

Eine Schärfung des Begriffsverständnisses von Forschungsinformationssystemen erfordert eine möglichst umfassende Betrachtung der strukturgebenden Komponenten und Wesensmerkmale. Ausgehend vom grundlegenden Begriffsverständnis eines Forschungsinformationssystems bezeichnet dieses ein aufgabenspezifisches Informationssystem, welches der Bewältigung der Aufgaben des Forschungsmanagements [LEJ12], insbesondere der Forschungsberichterstattung, dient und zu diesem Zweck die notwendigen informationstechnischen Mittel und organisatorischen Strukturen bereitstellt. Ein FIS bezieht sich folglich nicht ausschließlich auf ein einzelnes abgeschlossenes Softwaresystem. Vielmehr beschreibt es ein Anwendungssystem oder eine Anwendungssystemlandschaft sowie die organisatorische Prozesse, Regeln und Aufgabenträger, die es zur Bewältigung der forschungsmanagementbezogenen Aufgabenstellungen bedarf.

Ausgehend von diesem Verständnis lassen sich grundlegende Wesensmerkmale und Gestaltungsdimensionen von Forschungsinformationssystemen sowie deren Wechselwirkungen auf Basis des allgemeinen Informationssystemverständnisses und der damit einhergehenden Bestandteile (vgl. hierzu nochmals Abschnitt 2.1) ableiten. Zur Systematisierung der Dimensionen und deren Ausprägungen werden diese mit Hilfe eines

morphologischen Kastens zusammengefasst (vgl. Abb. 3). Hochschulen und Forschungseinrichtungen, die sich mit dem Aufbau bzw. der Einführung von Forschungsinformationen beschäftigen, kann dieser als Leitfaden zur Identifikation einer geeigneten Ausgestaltungsform eines FIS dienen. Weiterhin lassen sich auf dieser Basis unterschiedliche Systemklassen von Forschungsinformationssystemen charakterisieren und voneinander abgrenzen.

3.1 Mensch & Organisation

Bezugsrahmen und Einrichtungsbezug

Forschungsinformationssysteme bilden die Forschungsaktivitäten einer definierten Betrachtungseinheit ab. Der *Bezugsrahmen*, der in einem FIS behandelten Informationen und Prozesse, kann einerseits einrichtungsbezogen² (bspw. eine Universität oder außeruniversitäre Forschungseinrichtungen) und andererseits themen- oder disziplinenbezogen (bspw. Agrarforschung oder Philologie) abgesteckt werden. Zur weiteren Abgrenzung der Betrachtungseinheit gilt es zwischen solchen Systemen zu unterscheiden, die sich auf eine einzelne Einrichtung oder eine Gruppe von Einrichtungen beziehen (*Einrichtungsbezug*). Bei Betrachtung unterschiedlicher Gestaltungsformen von FIS wird deutlich, dass die beiden Dimension Einrichtungsbezug und Bezugsrahmen orthogonal zu einander zu sehen sind und sich in Abhängigkeit der jeweiligen Ausprägungsformen unterschiedliche Gestaltungsanforderungen an ein entsprechendes FIS ergeben.

- *Institutionelle Forschungsinformationssysteme für eine Einzeleinrichtung*: Erfassung des gesamten Forschungsspektrums einer Forschungseinrichtung. [HLTB12], [Af15]
- *Institutionelle Forschungsinformationssysteme über eine Gruppe von Einrichtungen*: zentralisierte Erfassung des Forschungsspektrums bspw. einer Region oder eines Forschungsförderers. [JO12]
- *Themenbezogene FIS über eine Gruppe von Einrichtungen*: themen- bzw. disziplinspezifische Abbildung von Forschungsaktivitäten einer Forschungs-Community oder eines fachbezogenen Forschungsförderers. [BKLS10]

Akteure

Ausgehend vom definierten Aufgabenspektrum eines FIS ergeben sich unterschiedliche *Akteure* als Aufgabenträger bzw. Adressaten. Im Fall eines *institutionellen FIS*, das etwa einer Einzeleinrichtung dient, kann als ein Nutzerkreis die Leitungsebene, welche die Universitätsleitung, aber auch die Fachbereichsleitungen und weitere Gremien der Selbstverwaltung umfasst, ausgemacht werden. Die Unterstützung von Prozessen des Forschungsmanagements sowie die damit verbundene Erfassung und Nutzung von Forschungsinformationen betrifft vor allem den Forschungsservice, die Bibliothek, die

² Eine Übersicht bestehender institutioneller Forschungsinformationssysteme bietet euroCRIS unter: <http://www.eurocris.org/activities/dris>

Transferstellen, das International Office sowie die Bereiche des akademischen Controlling, der Nachwuchsförderung sowie der Presse- und Öffentlichkeitsarbeit. Nicht zuletzt ist den Wissenschaftlern als Informationsträger über die unternommenen Forschungsaktivitäten eine wesentliche Rolle beizumessen – sowohl durch einen Beitrag zur Datenerfassung, als auch bei der aktiven Verwendung von Forschungsinformationen. Das legitime Interesse an öffentlich finanzierter Forschung von Seiten der Fördereinrichtungen, den Medien sowie der Wirtschaft und Praxis kann sinnvollerweise ebenfalls auf Basis der Informationen eines FIS bedient werden. [HLTB12]

Plattformen, die im überwiegenden Maße die Dokumentation und Verbreitung von Forschungsinformationen im Fokus haben, adressieren einen abweichenden Kreis von Akteuren. Eine Unterscheidung zwischen internen und externen Adressaten ist in diesem Fall nur schwer möglich bzw. nicht nötig. Hierbei rückt der Kreis der Wissenschaftler als Akteure in den Mittelpunkt, die sowohl als interne Nutzer eigene Forschungsprofile pflegen, als auch als externe Adressaten bspw. nach Kooperationspartnern oder bestehenden Veröffentlichungen recherchieren. Weitere in dem Sinne externe Adressaten sind vor allem die allgemeine Öffentlichkeit, Mittelgeber, sowie staatliche und privatwirtschaftliche Einrichtungen. [BKLS10]

3.2 Aufgabe

Anwendungsbereich

Ausgehend von der grundlegenden Aufgabenstellung eines FIS – die Unterstützung des Forschungsmanagements – lassen sich aus den fachlichen Bereichen des Forschungsmanagements im Wesentlichen drei grundlegende *Anwendungsbereiche* ableiten. [LEJ12], [Af15] Im Bereich der Forschungsberichterstattung liefern FIS in Gestalt von Berichts- und Auskunftssysteme Informationen zu den unternommenen Forschungsaktivitäten und erzielten Ergebnissen, um sowohl gesetzlich verankerten Berichtspflichten nachzukommen als auch Informationsbedarfe unterschiedlicher Akteure zu befriedigen. Im Hinblick auf die Sichtbarkeit, Kommunikation und den Transfer von Forschungsergebnissen können FIS ebenfalls Informationen für Zwecke der Außendarstellung bspw. den Aufbau eines Forschungsportals bereitstellen. Weiterhin ist es das Anliegen von FIS, administrative Prozesse zwischen Wissenschaft und Verwaltung zu unterstützen bzw. abzuwickeln, um auf diese Weise eine prozessbegleitende Erfassung der für die Forschungsberichterstattung und Außendarstellung benötigten Informationen zu ermöglichen.

Außendarstellung

Um die Sichtbarkeit und Kommunikation des Forschungsprofils sowie der Forschungsaktivitäten zu befördern, stellen einrichtungsweite, aber auch community-bezogene Forschungsportale ein geeignetes Werkzeug dar. Ausgestaltet als webbasierte Wissensplattformen bieten diese einen zentralen Anlaufpunkt u. a. für die Öffentlichkeit zur Recherche über Forschungsthemen, -aktivitäten und -ergebnisse. [BKLS10] Forschungsportale können hierbei in Form von einfachen Nachweissystemen umgesetzt werden, die einen spezifischen Fokus bspw. auf geförderte Projekte oder Publikationen aufweisen und die

benötigten Daten explizit für diesen Zweck erheben oder aus bestehenden Verwaltungssystemen übernehmen. Ferner können Forschungsportale auch als eine Komponente eines integrierten Forschungsinformationssystems oder mit Hilfe von Forschungsprofildiensten umgesetzt werden. Derartige Portale greifen auf den im FIS hinterlegten Teil des Datenbestands zurück, der u. a. für die Kommunikation gegenüber der Öffentlichkeit aufbereitet und freigegeben ist.

Ferner können die im FIS enthaltenen Daten zur dynamischen Einbindung auf dezentralen Webseiten der Fachbereiche, Institute und Arbeitsgruppen bereitgestellt werden. Über diesen Weg können bspw. Profile einer Person samt Projekt- oder Publikationslisten auf Basis der Daten im FIS erzeugt und in Webseiten eingebunden werden. Hiermit wird eine Mehrfacherhebung von Daten vermieden und damit der Pflegeaufwand seitens der Wissenschaftler reduziert.

Forschungsberichterstattung

Das Berichts- und Analysespektrum im Umfeld von Forschungsinformationen erstreckt sich von standardisierten Berichten über die Möglichkeit zur Erstellung von flexiblen Berichten bis hin zur Anwendung von tiefergehenden Analysemethoden aus dem Umfeld des Business Intelligence. Sowohl bei [St14] als auch [RDM16] stellen die Studienergebnisse die Forschungsberichterstattung als zentrale Funktionalität eines FIS heraus. Im Umfeld von Wissenschaftseinrichtungen bestehen vielfältige Berichterfordernisse in Form von fest verankerten Berichtspflichten als auch ad-hoc aufkommenden Anfragen gestellt durch Gremien, Ministerien oder auch die interne Leitungsebene. Als zusätzliche Unterstützung im Rahmen der strategischen Planung, bspw. zur Identifikation von interdisziplinären Schwerpunktbereichen, bieten weiterführende Datenanalysen wie Netzwerkanalyse- oder Text Mining-Ansätze umfangreiche Potentiale. Durch beliebige Variation im Aggregationsniveau können beispielsweise Co-Autoren- oder auch Kollaborationsnetzwerke extrahiert und unter Nutzung von geeigneten Vernetzungsmaßen oder auch Community-Detection-Algorithmen Beziehungsgeflechte innerhalb der Einrichtung aufgedeckt werden. [RKL12], [Sc16]

Prozessintegration und Prozessunterstützung

Ein Forschungsinformationssystem kann einen wesentlichen Beitrag zur Professionalisierung des Forschungsmanagement leisten, in dem es eine integrierte Plattform sowohl zur Abwicklung forschungsbezogener Verwaltungsprozesse als auch zur Erhebung, Bereitstellung und Nutzung von Informationen entlang dieser Prozesse bietet. Nicht zuletzt zur Vermeidung von Mehrfacherfassungen und damit zumeist einhergehender Dateninkonsistenzen gilt es das FIS in geeigneter Weise in die bestehende Anwendungslandschaft zu integrieren, in welcher mitunter einzelne Aspekte des Forschungsmanagements bereits durch andere Anwendungssysteme unterstützt werden. Die Prozessintegration bezieht sich an dieser Stelle auf die Art der Einbettung des FIS in die fachlichen Prozesse des Forschungsmanagement. Erfolgt *keine Einbindung in Prozesse*, werden die Forschungsinformationen redundant zur Verwendung in den Verwaltungsprozessen erfasst. Auch vorhandene Informationen im FIS werden nicht zur Aufgabenerledigung herangezogen. Wird ein FIS rein *organisatorisch in Prozessabläufe eingebunden*, wer-

den außerhalb von Anwendungssystemen bereits existierende organisatorische Abläufe für die Erfassung von Daten im bzw. die Bereitstellung von Informationen aus dem FIS genutzt. Schließlich ist unter *technischer Integration in Prozesse* die Abbildung von administrativen Prozessen im FIS bzw. die technische Kopplung mit anderen Anwendungssystemen zur Prozessabwicklung zu verstehen.

Der Abbildung bzw. *Unterstützung von Prozessen* in einem FIS sollte jedoch stets eine Eignungsüberprüfung vorangehen, bei welcher etwa Kriterien wie der Formalisierungs- bzw. Strukturierungsgrad oder die Fallzahl der Ausführung der Prozesse herangezogen werden. [MH12] Konkret bezogen auf die Anwendungsbereiche von FIS erscheinen auch insbesondere die Prozesse als geeignet, bei denen relevante Forschungsinformationen anfallen oder die zur Prozessabwicklung im FIS hinterlegte Informationen benötigen. Hierzu zählen bspw. das Stellen einer Drittmittelanzeige, der Antrag auf und die Berichterstattung zu einem Forschungsfreiemester, der Antrag auf Erstattung von Publikationskosten bei Open Access-Publikationen sowie die Beantragung von interner wie externer Forschungsförderung.

Art der Datenerfassung

Der Aspekt der Datenintegration nimmt eine zentrale Rolle im Rahmen der Datenerfassung ein. Dem integrativen Ansatz folgend gilt es die erforderlichen Informationen über geeignete Schnittstellen zu internen Anwendungssystemen, Importmöglichkeiten aus öffentlichen Informationsdiensten und eigene Datenerfassungsprozesse zu beschaffen. Daraus lassen sich grundsätzlich drei Möglichkeiten der Datenerfassung ableiten:

- manuelle Datenerfassung
- automatisierte Datenübernahme aus bestehenden Systemen
- nutzergesteuerte Datenübernahme aus bestehenden Systemen

Mit *manueller Datenerfassung* ist die Eingabe durch einen Benutzer gemeint, bei welcher die entsprechenden Informationen unabhängig von einer etwaigen vorherigen Erfassung in einem anderen IT-System in einem FIS erfasst werden. Die Erfassung der eigenen Publikationen durch den Wissenschaftler stellt dabei einen häufig auftretenden Anwendungsfall sowohl in institutionellen als auch in community-bezogenen FIS dar. [St14], [BKLS10]

Für einen *automatisierten Datenimport aus bestehenden Systemen* kann eine Anbindung von internen sowie externen Anwendungssystemen realisiert werden. Anwendungssysteme, die zur Erfassung von Forschungsinformationen idealerweise einbezogen werden können sind das Campusmanagement-System sowie das Identitätsmanagement-System als interne Systeme sowie öffentliche Publikations- und Projektdatenbanken. [Af15] Insbesondere für die Übernahme von öffentlich zugänglichen Forschungsinformationen kann dies neben einer Datenbankschnittstelle auch durch die Nutzung von Semantic Web Technologien realisiert werden, beispielsweise durch die Semantic-Web-Ontologie VIVO. [RKL12]

Schließlich kann ein Datenimport aus einem bestehenden System auch derart ausgestal-

tet sein, dass der jeweilige Benutzer die zu übernehmenden Informationen explizit sowie anlassbezogen auszuwählen hat. Ein sinnvoller Anwendungsfall besteht etwa darin, Wissenschaftlern den Import ihrer Publikationen oder sonstiger öffentlich verfügbarer Forschungsleistungen aus Publikations- und Projektdatenbanken zu ermöglichen.

Organisation und Turnus der Datenpflege

Neben der Art der Datenerfassung spielt die *organisatorische Zuständigkeit* eine ebenso gewichtige Rolle. So können die Daten etwa *zentral* mittels einer automatisierten Datenübernahme oder durch manuelle Eingabe von der administrierenden Verwaltungsabteilung erfasst werden. Ebenso kann die Datenerfassung *dezentral* bei den Wissenschaftlern angesiedelt werden. Weiterhin ist ebenfalls eine Kombination aus zentraler und dezentraler Erfassung und Pflege von Daten denkbar. Beispielsweise sind Informationen zu Publikationen zumeist nur bei den Wissenschaftlern bekannt, die diese dezentral erfassen. Zur Qualitätssicherung werden diese Angaben in einem weiteren Schritt durch eine zentrale Instanz wie bspw. der Bibliothek validiert. Analog dazu ist ebenso eine zentrale Datenerfassung einer Verwaltungseinrichtung oder der Bibliothek durch eine nachgelagerte Einbindung der Wissenschaftler zur Anreicherung um zentral nicht bekannte Informationen denkbar.

Die qualitative Studie von [St14] identifiziert im Hinblick auf die Zuständigkeit der Datenerfassung die Wissenschaftler, die Verwaltungseinrichtungen sowie die Bibliothek als primäre Akteure. Bezogen auf aktuell betriebene Forschungsinformationssysteme hat die Datenerfassung durch den Wissenschaftler die stärkste Verbreitung gefunden, gefolgt von den Verwaltungsabteilungen und schließlich der Bibliothek. Die etwaige Akteurübergreifende Datenerfassung fand in der Studienkonzeption jedoch keine Berücksichtigung. Letztlich wird die Art wie auch der *Turnus der Datenpflege* im Wesentlichen durch den formalen und konsequenzbehafteten Charakter des Anwendungsbereichs bestimmt, für den die erhobenen Daten bereitgestellt werden sollen. Wird ein FIS als Auskunftssystem für stetige Anfragen genutzt, ist eine kontinuierliche Aktualisierung und Pflege der Daten notwendig, wohingegen die Nutzung für zeitpunktbezogene Berichte lediglich eine mit entsprechendem Vorlauf anlassbezogene Organisation der Datenpflege notwendig ist.

Inhaltsdimensionen

Das Spektrum der zu erfassenden Forschungsinformationen ist in erheblichem Maße getrieben durch die Erfüllung von Berichtspflichten. Daher erscheint es für die Identifikation wesentlichen Inhaltsdimensionen eines FIS zielführend, unter anderem die durch der externen Berichtserfordernisse Informationsbedarfe in den Fokus zu nehmen. Als Grundlage eignet sich diesbezüglich insbesondere die Spezifikation des Kerndatensatzes Forschung, mit welcher eine für das deutsche Wissenschaftsumfeld systemweite, standardisierte Datenbasis angestrebt wird. [WR13], [WR16] Mit dem Common European Research Information Format (CERIF) besteht zudem ein Referenzdatenmodell für Forschungsinformationen, welches von euroCRIS initiiert wurde und stetig weiterentwickelt wird. Einblicke in das Spektrum der erfassten Forschungsinformationen im praktischen Umfeld verschaffen die Beiträge von [RDM16] und [St14]. Um Überschneidungs-

freiheit herzustellen, wurde – sofern hinreichend eindeutig – eine Zuordnung der Definitionen aus dem Kerndatensatz Forschung sowie gemäß CERIF zu den Ergebnisausprägungen der Studien vollzogen. Als Grundlage dient dabei der im CERIF definierte konzeptionelle Rahmen, dem eine Untergliederung in Core-Entities und Result-Entities zu Grunde liegt. [Jö12] Dieser wird ergänzt um Dimensionen aus dem Kerndatensatz Forschung sowie den qualitativen Studien (vgl. Abb. 2).

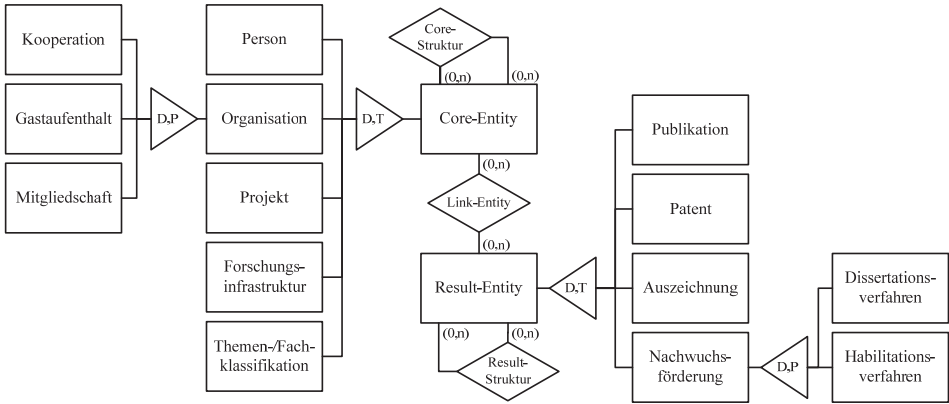


Abb. 2: Konzeptionelle Inhaltsstruktur eines Forschungsinformationssystems

3.3 Informationstechnik

Art der Datenhaltung

Bezogen auf die Datenintegration bestehen in der Literatur mit der *Integration durch Verbinden* und der *Integration durch Vereinigen* zwei Ansätze, welche die gemeinsame Datennutzung verschiedener Anwendungssysteme unter Berücksichtigung von Datenqualitätsanforderungen wie etwa dem Vorliegen in widerspruchsfreier Form zum Ziel haben. [Be01] Integration durch Verbinden meint dabei eine verteilte Datenhaltung, bei welcher der Datenaustausch zwischen den Anwendungen über Schnittstellen ermöglicht wird. Konsequenz umgesetzt entsteht ein bidirektionaler Datenaustausch, bei welchem sich die Daten verhalten, als würden sie sich in konsistenter Form in einem Datenmodell befinden. Unter Integration durch Vereinigen wird in der Literatur in überwiegender Form die gemeinsame Nutzung einer Datenbasis durch verschiedene Anwendungen verstanden. [Ju05] Ausgehend von der Literatur ergeben sich somit grundsätzlich die Datenhaltungsformen der verteilten Datenhaltung sowie der zentralen Datenhaltung.

Im Bereich von Universitäten sowie anderen Forschungseinrichtungen, die mit einem FIS eine Vielzahl von Verwendungszwecken verfolgen, werden in diesem Zusammenhang integrierte FIS als Systemansatz der Wahl propagiert. Als kennzeichnende Eigenschaft wird diesen Systemen die integrierte Speicherung jeglicher relevanter Entitäten in einem Datenmodell zugeschrieben. [AF15] Die daraus entstehende Datenbasis folgt im Kontext des FIS einer Integration durch Vereinigung. Bezogen auf die gesamte Anwen-

dungslandschaft repräsentiert durch angrenzende Systeme, wie dem Campusmanagement-, Identitätsmanagementsystem oder auch einer Hochschulbibliographie ist diese Form der Ausgestaltung hingegen einer Integration durch Verbinden (verteilte Datenhaltung) zuzuordnen. Jedoch bestehen dabei meist lediglich unidirektionale Schnittstellen zum FIS hin, weshalb diese Ausgestaltungsform an dieser Stelle als *isolierte Datenhaltung* bezeichnet wird. Aus der konsequenten Umsetzung eben dieser bidirektionalen Schnittstellen zwischen den Anwendungssystemen ergibt sich eine weitere Ausgestaltungsform der Datenhaltung, bei welcher eine separate Datenbasis für die Anwendungszwecke eines FIS nicht erforderlich ist, da die in verschiedenen Systemen bereits vorhandenen Forschungsinformationen je nach Anwendungszweck über bidirektionale Schnittstellen ausgetauscht und zusammengeführt werden können. Als *dynamische Verlinkung* betitelt, stellt das FIS lediglich die integrative Anwendungsschicht dar, welche als Oberfläche zur Aufgabenerledigung dient. Dies spiegelt somit eine stringente Verfolgung des Ansatzes Integration durch Verbinden dar. Im Gegensatz dazu sieht eine stringent umgesetzte Datenvereinigung vor, *eine Datenbasis über die Systemgrenzen* des FIS für mehrere administrative Anwendungssysteme zu schaffen. Eine derartige Ausgestaltung stellt eine Redundanz- sowie Widerspruchsfreiheit sicher. Dem entgegen steht jedoch die erhebliche Komplexität in der Umsetzung. So müssen innerhalb eines Datenmodells sämtliche Anforderungen der zu integrierenden Anwendungssysteme berücksichtigt werden.

Systemausgestaltung

Ausgehend vom dem zugrunde gelegten Verständnis eines Forschungsinformationssystem als Informationssystem und damit als sozio-technisches System bezieht sich demnach der Kontext eines FIS nicht notwendigerweise, wie oft propagiert, ausschließlich auf ein abgeschlossenes Softwaresystem. Vielmehr ergeben sich vielfältige Architekturkonzepte, etwa bestehend aus einem oder auch mehreren Anwendungssystemen, welche die forschungsbezogenen Prozesse abbilden bzw. unterstützen. Ein *monolithischer* Ansatz sieht vor, sämtliche unterstützte Funktionalitäten in einem Anwendungssystem abzubilden. In diesem Fall werden jegliche benötigten Informationen und Funktionalitäten eines FIS in einem geschlossenen Anwendungssystem hinterlegt bzw. umgesetzt. Dementgegen ist ebenso eine *modulare* Systemausgestaltung denkbar, bei welcher bestehende Anwendungssysteme, wie ein Campusmanagement-System, ein Finanzsystem oder eine Hochschulbibliographie als einzelne Komponenten eines FIS aufgefasst und durch einer Integrationsplattform miteinander verbunden werden. Im Gegensatz zum monolithischen Ansatz agiert die Integrationsplattform vielmehr als Präsentationschicht der Funktionalitäten und Daten bereits bestehender Anwendungssysteme. Darüber ist es ebenfalls denkbar, dass die Integrationsplattform neben der reinen Einbindung bestehender Anwendungen auch separate Funktionalitäten bspw. zur Verknüpfung der Daten aus den einzelnen Anwendungen bereitstellt.

Dimension		Ausprägungen										
Bezugsrahmen	einrichtungsbefugten	disziplinen- / themenbezogen										
Einrichtungsbezug	Einzeleinrichtung	Gruppe von Einrichtungen										
Akteure/Adressaten	WissenschaftlerInnen	Leitungsebene	Gremien	Akademisches Controlling/ Strategische Planung	Bibliothek	Forschungs- service	Transfer- stellen					
	International Office	Nachwuchs- förderung	Presse- & Öffentlichkeitsarbeit	Forschungsförderer	Allgemeine Öffentlichkeit	Wirtschaft & Praxis	Medien					
Inhaltsdimensionen	Personen	Organisationen	Publikationen	Projekte	Preise & Auszeichnungen	Patente	Forschungsinfrastruktur					
	Promotionsverfahren	Habilitationsverfahren	Kooperationen	Mitgliedschaften	Gastaufenthalte	Themen-/ Fachklassifikationen						
Anwendungsbereich	Außendarstellung	Forschungsberichterstattung										
Außendarstellung	zentrales Forschungsportal	dynamische Einbindung von Daten in Webseiten										
Forschungs- berichterstattung	standardisierte Berichte	ad-hoc Berichte und Auskünfte										
Prozessintegration	keine Einbindung in Prozesse	organisatorische Einbindung in Prozesse										
Prozessunterstützung	Drittmittelanzeige	Beantragung Forschungsfreimester	Erfindungs- meldung	Antragstellung für (interne) Forschungsförderung	Nominierung für Preise	Erstattung von Open Access-Gebühren						
	manuelle Datenerfassung	automatisierte Datenübernahme aus bestehenden Systemen										
Organisation der Datenpflege	zentrale Erfassung & Pflege	dezentrale Erfassung & Pflege	dezentrale Erfassung mit zentraler Qualitätssicherung									
	Turnus der Datenpflege	kontinuierlich										
Art der Datenhaltung	isolierte Datenhaltung	zentrale Datenhaltung										
Systemausgestaltung	monolithische Anwendung	modulare Anwendungslandschaft										
		dynamische Verlinkung										

Abb. 3: Gestaltungsdimensionen von Forschungsinformationssystemen

4 Zusammenfassung und Ausblick

In dem vorliegenden Beitrag wurden auf Basis bestehender Literatur aus dem Umfeld von Forschungsinformationssystemen sowie eigenen Erfahrungen aus dem Betrieb eines FIS Wesensmerkmale eben dieser Systeme abgeleitet. Insbesondere der zugrundeliegende Bezugsrahmen eines Informationssystems befördert die Betrachtung eines FIS in seinem Gesamtkontext. Der resultierende morphologische Kasten dient dabei als Systematisierung und erlaubt gleichzeitig die Ableitung bestehender Ausgestaltungsformen als auch die Entwicklung weiterführender Ansätze und Lösungen. Aufbauend auf der hier erfassten Systematisierung kann zum einen die Ableitung von Systemklassen weiter voran gebracht werden. Eine geeignete Methode ist dieser Hinsicht eine gezielte qualitative Studie über bisher betriebene Ausgestaltungsformen in Universitäten, außeruniversitären Wissenschaftseinrichtungen sowie Forschungs-Communities. Neben der weiteren Differenzierung können die hier erworbenen Erkenntnisse genutzt werden, um in der weiteren Auseinandersetzung mit der Thematik zu zentralen Begrifflichkeiten zu kommen, die schlussendlich in einer Definition von Forschungsinformationssystemen münden.

Förderung

Dieser Beitrag wurde durch die Förderung des Projekts „Landesinitiative CRIS.NRW – Unterstützungsstrukturen für die Implementierung des Kerndatensatz Forschung und von Forschungsinformationssystemlösungen an nordrhein-westfälischen Hochschulen als Beitrag zur Harmonisierung der Forschungsberichterstattung in Nordrhein-Westfalen“ durch das Ministerium für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen unter dem Aktenzeichen 212-7.01.01-(126331) ermöglicht.

Literaturverzeichnis

- [Af15] DINI AG FIS (2015): Forschungsinformationssysteme in Hochschulen und Forschungseinrichtungen. Positionspapier der Arbeitsgruppe Forschungsinformationssysteme der Deutschen Initiative für Netzwerkinformationen e.V, 2015.
- [Be01] Becker, J.: Integrationsorientierte Wirtschaftsinformatik – Auf dem Weg zu einer Theorie der Integration. Anlässlich des 60. Geburtstags von Prof Dr. Dr. h.c. August-Wilhelm Scheer, 2001.
- [BKLS10] Becker, J., Knackstedt, R., Lis, L., Stein, A.: Towards a Maturity Model for Research Portals. *ECIS 2010 Proceedings*, Paper 38, 2010
- [He12] Herb, U. (Hrsg.): Open Initiatives: Offenheit in der digitalen Welt und Wissenschaft, Universitätsverlag des Saarlandes, Saarbrücken, 2012.
- [HKM08] Hornbostel, S., Klingsporn, B., von Ins, M. (2008): Messung von Forschungsleistungen – eine Vermessenheit? In: Alexander von Humboldt-Stiftung (Hrsg.): Publikationsverhalten in unterschiedlichen Disziplinen – Beiträge zur Beurteilung von Forschungsleistung. Diskussionspapier Nr. 12 der Alexander von Humboldt-Stiftung, Bonn, 2008.
- [HLTB12] Herwig, S., Lis L., Tusch, Ch., Becker, J.: Einführung eines Forschungsinformationssystems – Anforderungen, Konzept und Umsetzung am Beispiel der Westfälischen Wilhelms-Universität

- Münster. In: Schwartze, S Tröger, B., Vogl, R. (Hrsg.): Fortschritte des integrierten Informationsmanagements an Hochschulen, 2012.
- [Ho06] Hornbostel, S.: From CRIS to CRIS: Integration and Interoperability, International Conference of Current Research Information Systems, Bergen, 2006
- [JA06] Jeffery, K.G., Asserson, A.: CRIS: Central Relating Information System, Proceedings of 8th International Conference of Current Research Information Systems, Bergen, 2006
- [JO12] Johansson, A., Ottosson, M., A.: A national Current Research Information System for Sweden, Proceedings of 11th International Conference of Current Research Information Systems, Prague, 2012
- [Jö12] Jörg, B. et al.: CERIF 1.3 Full Data Model (FDM): Introduction and Specification. Technical Report. 2012 http://www.eurocris.org/Uploads/Web%20pages/CERIF1.3/Specifications/CERIF1.3_FDM.pdf (abgerufen am 24.06.2016)
- [Ju05] Reinhard, J.: Architekturen zur Datenintegration – Gestaltungsempfehlungen auf der Basis fachkonzeptueller Anforderungen, DUV. 2005
- [LEJ12] Locker-Grütjen, O., Ehmann, B., Jongmanns, G.: Definition für ein optimales Forschungsmanagement, In: Wissenschaftsmanagement 3/2012, S.34-38. 2012
- [MH12] Zur Mühlen, M., Hansmann, H.: Workflowmanagement. In: Becker, J., Kugeler, M., Rosemann, M.(Hrsg.): Prozessmanagement – Ein Leitfaden zur prozessorientierten Organisationsgestaltung, 7. Auflage, Springer-Verlag. 2012
- [RDM16] Ribeiro, L., de Castro, P., Mennielli, M.: EUNIS – euroCRIS joint survey on CRIS and IR, Final Report, 2016
- [RKL12] Corson-Rikert, J., Krafft, D. B., Lowe, B.J.: VIVO: A Semantic Network of Researchers and Research Information as Linked Open Data. In: Bittner, S., Hornbostel, S., Scholze, F. (Hrsg.): Forschungsinformation in Deutschland: Anforderungen, Stand und Nutzen existierender Forschungsinformationssysteme. 2012
- [Sc16] Schlattmann, S.: Capturing the collaboration intensity of research institutions using social network analysis, Proceedings of 13th International Conference of Current Research Information Systems, St. Andrews. 2016
- [St14] Sticht, K.: Einsatz von Forschungsinformationssystemen an Universitäten und Hochschulen mit Promotionsrecht, HU Berlin 2014
- [Te99] Teubner, R.A.: Organisations- und Informationssystemgestaltung, Gabler Edition Wissenschaft, 1999.
- [WR13] Wissenschaftsrat : Empfehlungen zu einem Kerndatensatz Forschung (Drs. 2855-13), 2013
- [WR16] Wissenschaftsrat : Empfehlungen zur Spezifikation des Kerndatensatz Forschung (Drs. 5066-16), 2016

Erfolgsfaktoren für die Einführung integrierter Campus-Management-Systeme – eine vergleichende Literaturanalyse mit praxisbezogener Evaluation

Gunnar Auth¹, Steffen Künstler²

Abstract: Die Einführung eines integrierten Campus-Management-Systems (CMS) zur Unterstützung der Geschäftsprozesse entlang des studentischen Lebenszyklus stellt Hochschulen noch immer vor große Herausforderungen. Auch wenn sich mittlerweile das Verständnis für die erfolgskritische Rolle einer projektintegrierten Organisationsentwicklung sowie die Gefahren einer zu starken IT-Fokussierung bis hinauf in die Hochschulleitungen verbreitet hat, werden von CMS-Einführungsprojekten die zu Beginn definierten Ziele insbesondere in Hinblick auf Termine und Kosten häufig nicht erreicht oder es kommt sogar zum Projektabbruch. In dieser Arbeit werden spezifische Faktoren identifiziert, die einen Beitrag dazu leisten können, den Zielerreichungsgrad von CMS-Einführungsprojekten zu erhöhen und somit den Projekterfolg zu verbessern. Dazu wird von allgemeinen Erfolgsfaktoren für IT-Projekte sowie für die Einführung der nahe verwandten ERP-Systeme ausgegangen.

Keywords: Campus-Management-System, Erfolgsfaktor, Enterprise Resource Planning, Einführungsprojekt, IT-Projekt, Organisationsentwicklung, Hochschulmanagement.

1 Einleitung

Bei Campus-Management-Systemen (CMS) handelt es sich um sektorspezifische Anwendungssysteme, mit denen Hochschulen computergestützt die Durchführung ihrer Geschäftsprozesse im Bereich Studium und Lehre planen, steuern und überwachen. Obwohl es bis heute auch Hochschulen gibt, die ihr individuelles CMS selbst entwickeln,³ überwiegt die Zahl der Hochschulen, die ein oder mehrere kommerzielle Produkt(e) von Software-Herstellern wie bspw. HIS, Datenlotsen oder SAP einsetzen [Ra09], [EY12]. Diese Produkte werden für einen Massenmarkt und mit dem Ziel eines möglichst großen Marktanteils entwickelt und können daher als Standardsoftware betrachtet werden (vgl. [Bu14]).⁴

¹ Hochschule für Telekommunikation Leipzig, Department Wirtschaftsinformatik, Gustav-Freytag-Str. 43-45, 04277 Leipzig, gunnar.auth@hft-leipzig.de

² Ebda., steffen.kuenstler@hft-leipzig.de

³ Bspw. die Technische Universität Graz, die das von ihr zunächst für den eigenen Bedarf entwickelte CMS CAMPUSonline seit geraumer Zeit auch an andere Hochschulen lizenziert. Dieses weist dadurch zunehmend Merkmale von Standardsoftware auf. Siehe: <https://campusonline.tugraz.at>, 10.03.2016.

⁴ Eine andere Auffassung hierzu vertreten [Sp+15], die aufgrund der für jede Hochschule notwendigen, umfangreichen Anpassungen während der Einführung CMS-Produkte (noch) nicht als Standardsoftware ansehen.

Einführungsprojekte für CMS stellen angesichts ihrer Vielschichtigkeit Hochschulen und Universitäten auch heute noch vor große Herausforderungen, da zeitgleich organisatorische, strukturelle und technische Ineffizienzen eliminiert werden sollen [SKB10]. Nachteilig wirkt sich dabei in der Regel ein Mangel an Erfahrung der Hochschulen mit der Durchführung von IT-Projekten dieser Größe und Komplexität aus [Sp15+]. Obwohl sich das Verständnis für die Notwendigkeit einer professionellen Projektorganisation, die erfolgskritische Rolle einer projektintegrierten Organisationsentwicklung sowie die Gefahren einer zu starken IT-Fokussierung bis hinauf in die Hochschulleitungen verbreitet hat, werden von CMS-Einführungsprojekten die zu Beginn definierten Ziele insbesondere in Hinblick auf Termine und Kosten häufig nicht erreicht (bspw. [Me13], [Sp+15]) oder es kommt sogar zum Projektabbruch. So hat etwa die Universität Bielefeld Anfang November 2015 ein im Oktober 2013 begonnenes CMS-Einführungsprojekt abgebrochen und anschließend mit einem anderen Software-Anbieter neu gestartet [UB15].

Eine wichtige Aufgabe bei der Einführung von Standardsoftware ist die Anpassung der im Produkt implementierten Standardfunktionalität an die spezifischen Anforderungen der Anwenderorganisation, das sog. Customizing [Nü14]. Bei der CMS-Einführung ist dies eine besonders anspruchsvolle Aufgabe, da es aus Effizienz- und Kostengründen aber u. a. auch im Interesse einer wohlgeordneten Studienorganisation erforderlich ist, dass unterschiedlichste Fachbereiche ihre jeweiligen Anforderungen soweit aneinander annähern, dass sie von einem integrierten Anwendungssystem erfüllt werden können (vgl. [Au15]).

Im Bereich der Wirtschaft gibt es bei der unternehmensweiten Einführung von Standardsoftware ähnliche Problemstellungen. Am ehesten vergleichbar ist dabei die Einführung sog. ERP-Systeme (Enterprise Resource Planning-Systeme), mit denen Unternehmen ihre Ressourcen wie Material, Personal und Kapazitäten zur Durchführung der Geschäftsprozesse verwalten [Gr14a], [Au15].

Im vorliegenden Beitrag werden spezifische Faktoren identifiziert und beschrieben, die einen Beitrag dazu leisten können, den Zielerreichungsgrad von CMS-Einführungsprojekten zu erhöhen und somit den Projekterfolg zu verbessern. Dazu wird nach Beschreibung von Forschungsgegenstand und –methodik in Abschnitt 2 das komplexe Phänomen Erfolg von CMS-Einführungsprojekten in Abschnitt 3 anhand spezifischer Kriterien operationalisiert. Ausgangspunkt der weiteren Überlegungen sind allgemeine Erfolgsfaktoren für IT-Projekte sowie für die Einführung der nahe verwandten ERP-Systeme. Mittels einer vergleichenden Literaturanalyse werden in Abschnitt 4 zunächst Erfolgsfaktoren für ERP- und CMS-Einführung aus der jeweiligen Literatur extrahiert. Anschließend werden die theoretisch ermittelten Erfolgsfaktoren im Rahmen von Interviews durch Experten aus der CMS-Praxis kritisch reflektiert. In Abschnitt 6 werden die Ergebnisse der vorangegangenen Schritte zusammengeführt und anhand eines Vorgehensmodells für die CMS-Einführung strukturiert. Der Beitrag endet mit einem Fazit und Ausblick.

2 Forschungsgegenstand und -methodik

Forschungsgegenstand dieser Arbeit sind Erfolgsfaktoren für die Einführung von CMS. Unter Erfolgsfaktoren versteht man in der Betriebswirtschaftslehre nach [Fr93] Kausal-faktoren, die über Erfolg oder Misserfolg betrieblicher Aktivitäten entscheiden. Neben Einflussgrößen, die von der Umwelt bestimmt werden, sind vor allem sog. kritische Erfolgsfaktoren von Interesse, die vom Management eines Unternehmens prinzipiell beeinflusst werden können. Die Motivation der Erfolgsfaktorforschung beruht auf der Annahme, dass trotz der Multidimensionalität und Multikausalität von Erfolg einige wenige zentrale Faktoren für dessen Eintritt verantwortlich sind [SKW05]. Ihren Nutzen für die Praxis entfalten die Ergebnisse der Erfolgsfaktorforschung bei Entscheidungen über die Durchführung erfolgsversprechender Maßnahmen, wofür sie eine bessere Entscheidungsgrundlage bieten [SKW05].

Mittlerweile gibt es ein breites Spektrum an Untersuchungsgegenständen innerhalb der Erfolgsfaktorforschung. Für den vorliegenden Beitrag sind besonders die Arbeiten zum Erfolg von Projekten bzw. Projektmanagement relevant (bspw. [MB96], [GPM15]). Mertens hat über einen Zeitraum von ca. acht Jahren Probleme in IT-Projekten der öffentlichen Verwaltung untersucht (von ihm selbst als „Misserfolgsvorschung“ bezeichnet [Me12a]) und dabei auch ein Beispiel aus dem Hochschulumfeld untersucht, nämlich die bis heute nicht abgeschlossene Einführung der zentralen Internet-Plattform hochschulstart.de für die zentrale Studienplatzvergabe [Me12b].

Zudem beschäftigen sich die Wirtschaftsinformatik und ihre englischsprachige Schwesterdisziplin schon seit längerem mit Erfolgsfaktoren für Informationssysteme. Hier bilden wiederum die ERP-Systeme einen Schwerpunkt, zu dem auch Untersuchungen zählen, die speziell die Erfolgsfaktoren für die Einführung von ERP-Systemen betrachten (bspw. [ST13]).

Dem vorliegenden Beitrag liegen folgende Forschungsfragen zugrunde:

- 1) Existieren allgemeingültige Erfolgsfaktoren für die Einführung von Campus-Management-Systemen?
- 2) Lassen sich Erfolgsfaktoren aus dem Bereich der Einführung von ERP-Systemen auf die Einführung von Campus-Management-Systemen übertragen?
- 3) Was konkret sind die signifikanten Erfolgsfaktoren für die Einführung von Campus-Management-Systemen?

Zur Beantwortung dieser Fragen wurde zunächst eine qualitative Literaturanalyse durchgeführt, bei der Erfolgsfaktoren sowohl für ERP- als auch für CMS-Einführungsprojekte extrahiert und verglichen wurden. Zur Evaluation der Analyseergebnisse wurde anschließend eine qualitative Interviewstudie durchgeführt. Als Erhebungsansatz für die Interviewstudie wurde das Leitfaden-Interview gewählt. Mit der Interviewstudie wird vorrangig das Ziel verfolgt, die individuelle und subjektive Bewertung der Teilnehmer zu explorieren und darüber auf die Validität und Vollständigkeit der ermittelten Erfolgsfaktoren zu

schließen. Mit der Verwendung eines Leitfadens wurde eine möglichst gute Vergleichbarkeit zwischen den einzelnen Interviews angestrebt. Als Interviewpartner wurden Projektverantwortliche an Hochschulen ausgewählt, die ein Einführungsprojekt kürzlich durchgeführt haben bzw. sich derzeit in der Umsetzung befinden. Ebenfalls wurden externe Berater mit Spezialisierung auf die CMS-Einführung bzw. IT-Projekten im Hochschulwesen einbezogen.

3 Operationalisierung des Erfolgs von CMS-Einführungsprojekten

Um Erfolgsfaktoren bestimmen zu können, ist es zunächst erforderlich zu definieren, was unter Erfolg verstanden wird und mit welchen einheitlichen Kriterien Erfolg rückblickend gemessen bzw. überprüft werden soll. In Bezug auf Projekte stellt sich Erfolg als die möglichst vollständige Erreichung der zu Beginn definierten Projektziele dar. Dabei lässt sich zwischen Sach- und Formalzielen unterscheiden. Die Sachziele beziehen sich auf Zweck und Inhalt des Projekts, im vorliegenden Fall die Einführung eines integrierten CMS. Mit der Zielerreichung verbunden ist die Realisierung von Nutzeffekten für den Projektauftraggeber [CD02] bspw. eine möglichst effiziente Studienorganisation oder eine gute Studierbarkeit eines stark verflochtenen Modulangebots. Zusätzlich zu den Sachzielen verfolgen Projekte gemäß ihrer Definition auch immer Formalziele, die sich auf die Güte der Sachzielerreichung beziehen. Dabei handelt es sich um Ziele in Bezug auf Zeit bzw. Dauer, Kosten und Qualität der Projektdurchführung sowie des Ergebnisses [De+05].

Gewünschtes Ergebnis eines CMS-Einführungsprojektes ist die hochschulweite Nutzung eines integrierten Anwendungssystems, welches die zu Beginn definierten Anforderungen erfüllt. Damit lässt sich wiederum an die Erfolgskriterien für Informationssysteme anknüpfen, die bspw. von DeLone und McLean ([DM92], [Pe+08]) untersucht wurden. Die Erfüllung der formal definierten Anforderungen ist letztlich noch kein Garant für die subjektive Zufriedenheit des Auftraggebers, der Nutzer sowie der übrigen Stakeholder. Für diese ist primär der wahrgenommene Nutzen des Projektergebnisses ausschlaggebend [CD02]. Dieser kann sich durchaus vom zu Beginn des Projekts erwarteten Nutzen unterscheiden, bspw. wenn im Rahmen der Projektdurchführung Lösungen entstehen, die den zuvor erwarteten Nutzen übertreffen. Andererseits können sich im Projektverlauf Grenzen zeigen, die zuvor nicht erkennbar oder nicht akzeptiert waren. Stellt sich während eines Projekts eine bestimmte Verbesserungsidee trotz großer Anstrengungen als nicht realisierbar heraus, kann dies dennoch zur Akzeptanz und damit gewissermaßen auch Zufriedenheit mit einer als nicht optimal wahrgenommenen Situation führen.

Während bisher Projekterfolg aus einer Außensicht (Auftraggeber, Nutzer, Stakeholder) beleuchtet wurde, gehört zu einer ganzheitlichen Betrachtung auch die Innensicht, d. h. die Sicht der Projektmitarbeiter und –beteiligten (bspw. Key user). Aus Innensicht lässt sich der Erfolg eines Projekts zusätzlich zu den bisher genannten Kriterien nach der Qualität der Projektarbeit bemessen, die wiederum von der Planungsqualität, hinreichender

Ressourcenausstattung, Eignung und Anpassbarkeit des Standard-CMS sowie der Güte der Zusammenarbeit mit dem Anbieter beeinflusst wird.

Der Erfolg eines CMS-Einführungsprojektes bemisst sich demnach an der Erreichung der individuellen Sach- und Formalziele des Projekts, der Nutzung des eingeführten Systems und der damit verbundenen Nutzerzufriedenheit sowie der Zufriedenheit der Projektauftraggeber, Projektbeteiligten und anderer Stakeholder.

4 Vergleichende Literaturanalyse zu Erfolgsfaktoren von ERP- und CMS-Einführungsprojekten

4.1 Vorgehen

Die Literaturanalyse wurde in drei Schritten durchgeführt. Als erstes wurde nach relevanter Literatur zu ERP-Erfolgsfaktoren gesucht. Dazu wurden die Internet-Suchmaschinen Google Scholar und Microsoft Academic sowie das soziale Netzwerk ResearchGate genutzt. Gesucht wurde mit den Suchwörtern „ERP“, „Einführung“, „Erfolgsfaktor“ sowie den englischen Übersetzungen „Implementation“ und „Success factor“. Aus den Ergebnissen wurden relevante Beiträge von Hand ausgewählt und über deren Literaturverzeichnisse eine einstufige Rückwärtssuche durchgeführt. Schließlich wurde für ausgewählte Artikel mit Hilfe der Suchmaschinen noch eine Vorwärtssuche durchgeführt.

Im nächsten Schritt wurde genauso mit der Literatur zu CMS-Erfolgsfaktoren verfahren. Als Suchwörter wurden hierbei verwendet „Campus Management“ sowie ebenfalls „Einführung“ und „Erfolgsfaktor“. Auf die Suche nach englischsprachigen Beiträgen wurde verzichtet, da die Bezeichnung Campus-Management-System international weniger verbreitet ist. Stattdessen wird bspw. von „Student Information System“ gesprochen (bspw. [GP03]), wobei es dazu eine große Spannbreite an konkreten Systemausprägungen gibt. Auch verzichtet wurde auf die Suche mit älteren Begriffen wie „Hochschulinformationssystem“.

Schließlich wurde im dritten Schritt die jeweils extrahierten Erfolgsfaktoren verglichen und auf Gemeinsamkeiten und Unterschiede hin untersucht.

4.2 Erste Ergebnisse

In Anbetracht der deutlich früheren Entstehung und des deutlich größeren Verbreitungsgrads sowie Einsatzbereichs von ERP-Systemen konnten evident mehr Arbeiten mit Bezug zu Erfolgsfaktoren in diesem Bereich ausgemacht werden. Die meisten der untersuchten Publikationen ermittelten Erfolgsfaktoren auf Basis von Interviews und Fallstudien. Neben diesen überwiegend empirisch ausgerichteten Arbeiten wurden in jüngerer Zeit mehrere Literaturstudien zu ERP-Erfolgsfaktoren publiziert ([ST13], [ZF14], [No15]).

Erste wissenschaftliche Ausarbeitungen zur Ermittlung von Erfolgsfaktoren mit Fokus auf CMS finden sich verstärkt ab Mitte der 2000er Jahre. Ab diesem Zeitpunkt verbreitet sich der Begriff CMS im deutschsprachigen Hochschulraum und mehrere Produkte sind auf dem Markt erhältlich. Auffällig ist die Häufung von Arbeiten zur Internationalen Tagung Wirtschaftsinformatik im Jahr 2009 in Wien, die sich durch einen eigenen Track „Integrierte Campus-Managementsysteme“ erklärt. Als Forschungsmethode diente den jeweiligen Autoren vorrangig das qualitative Interview sowie die Fallstudie, häufig basierend auf Erfahrungen mit CMS-Einführungen, an denen die Autoren selbst beteiligt waren ([KG07], [St+07], [Br+09], [De+09], [FH09], [Ja+09], [OT11], [SAA12], [Au14], [Au15], [Sp+15] et al.).

Erwartungsgemäß fand sich ein großer Anteil der ERP-Erfolgsfaktoren auch in Arbeiten zu CMS-Erfolgsfaktoren wieder. Da ERP-Erfolgsfaktorforschung aber bereits deutlich länger betrieben wird, liegen in diesem Bereich auch differenziertere, theoretisch fundiertere, besser empirisch abgesicherte und umfassendere Ergebnisse vor. Diese lassen sich nutzen, um die bisherigen Erkenntnisse über CMS-Erfolgsfaktoren einerseits kritisch zu reflektieren und andererseits zu ergänzen. Allerdings sind die Untersuchungen von CMS-Erfolgsfaktoren – ebenfalls erwartungsgemäß – deutlich spezifischer in ihrem Hochschulbezug, d. h. in der Berücksichtigung der Besonderheiten des Organisationstypus Hochschule.

Bevor die Ergebnisse der Zusammenführung von ERP- und CMS-Erfolgsfaktoren in Abschnitt 6 im Detail dargestellt werden, wird nachfolgend eine interviewbasierte Evaluation der Ergebnisse der Literaturanalyse beschrieben, die weitere Erkenntnisse brachte.

5 Evaluation mittels Experteninterviews

Zur Überprüfung der mittels Literaturanalyse extrahierten Erfolgsfaktoren wurden diese im Rahmen einer Interviewstudie durch CMS-Experten aus der Hochschulpraxis kritisch reflektiert. Um die Vergleichbarkeit der Interviews untereinander zu erhöhen, wurden die Interviews entsprechend der Phasen eines Vorgehensmodells (vgl. Abschnitt 6) mittels Leitfaden halbstrukturiert durchgeführt. Die Auswahl der Interviewteilnehmer konzentrierte sich aus praktischen Gründen auf die Ballungsräume Berlin, Dresden sowie Leipzig. Für die Auswahl der Interviewteilnehmer waren weiterhin ausschlaggebend: a) der Bezug zu CMS, b) die Rolle innerhalb eines CMS-Einführungsprojekts und c) der Status der CMS-Einführung an der Hochschule des Interviewteilnehmers (abgeschlossen oder kurz vor Abschluss). Insgesamt wurden zur Teilnahme am Interview zehn Personen kontaktiert. Zur Durchführung des Interviews kam es schließlich mit vier Personen (siehe Tab. 1), ein bereits vereinbartes fünftes Interview kam kurzfristig nicht zustande. Die durchgeführte Studie kann angesichts des geringen Umfangs keinen Anspruch auf Allgemeingültigkeit erheben, was allerdings auch nicht bezweckt war. Sie dient vielmehr einer ersten Evaluation der Erfolgsfaktoren sowie dem Testen und Weiterentwickeln des Interview-Leitfadens als Vorbereitung einer umfassenderen empirischen Studie.

Kürzel	Bezug CMS	Rolle	Status
(I1)	Unabhängiger Berater	Berater	Mehrere Einführungen abgeschlossen
(I2)	Unabhängiger Berater	Berater	Mehrere Einführungen abgeschlossen
(I3)	Hochschulvertreter (FH)	Projekt-koordination	Einführung abgeschlossen
(I4)	Hochschulvertreter (Uni)	Projektleitung	Einführung fast abgeschlossen

Tab. 1: Charakterisierung der Interviewpartner

Im Rahmen der Interviewstudie konnte zunächst das zugrunde gelegte Vorgehensmodell zur Einführung von CMS weitestgehend bestätigt werden. Dabei erfolgte die Bestätigung sowohl direkt als auch indirekt durch die Interviewpartner. Direkt, weil vereinzelt von den Befragten konkrete positive Rückmeldung zum Modell geäußert wurde. Häufigster Kritikpunkt war, dass nicht ersichtlich werde, dass sich die einzelnen Phasen auch überlagern können. Indirekt wurde das Modell durch die Beschreibungen der Teilnehmer zum individuellen Projektverlauf bestätigt. Vor allem der geschilderte Projektverlauf bei (I3) glich sehr stark dem entwickelten Vorgehensmodell und dies, obwohl man keinen konkreten Projektmanagementansatz bei der Umsetzung berücksichtigte.

Bei der Auswertung fiel auf, dass Erfolgsfaktoren überwiegend in der ersten Projektphase Planung ausgemacht werden konnten. Hier gab es die größte Übereinstimmung zwischen den einzelnen Interviews. In den späteren Phasen divergierten die Aussagen zwischen den einzelnen Interviews viel stärker, weshalb hier nur noch vereinzelt eine Zusammenführung möglich war. Als Grund dafür, dass vor allem in der Phase Planung die Aussagen der einzelnen Teilnehmer nahezu deckungsgleich waren, ist womöglich die identische Ausgangslage auszumachen. Je größer der Projektfortschritt, desto individueller gestaltete sich der Projektverlauf. Angesichts der unterschiedlichen Projektverläufe ab der Phase Planung entwickelten sich im weiteren Projektverlauf ebenso unterschiedliche Schwerpunkte bei den erfolgsrelevanten Faktoren.

Für alle Interviewpartner von zentraler Bedeutung war die geschlossene Unterstützung durch die Hochschulleitung. Auch wenn der ursprüngliche Projektanstoß meist durch Einzelpersonen in die Hochschule eingebracht wurde, bestand zum Zeitpunkt der Bewilligung des Einführungsprojekts eine einheitlich geschlossene Position innerhalb der gesamten Hochschulleitung. Zuvor war im Vorfeld der Projektbewilligung durchgängig sehr viel Gremienarbeit und Stakeholdermanagement notwendig.

Ebenfalls ein Merkmal, das in jedem Interview festgestellt werden konnte, ist die Positionierung des CMS-Einführungsprojekts als Reorganisationsprojekt. Es zeichnete sich bereits in der Literaturanalyse ab und wurde von den Interviewteilnehmern bestätigt. In jedem Fall sah man in der Einführung eines CMS ein Reorganisationsprojekt mit Anpassung der Aufbau- sowie Ablauforganisation, die durch ein Softwaresystem lediglich unterstützt

wird. Das Softwaresystem ist zwar bei diesen Projekten keineswegs zu vernachlässigen, wurde aber als weniger kritisch im Vergleich zur Organisationsentwicklung bewertet. Dies signalisiert durchaus ein gewisses Vertrauen gegenüber den jeweils ausgewählten Standardsystemen.

Bei der Wahl der Einführungsstrategie hat sich gezeigt, dass ausschließlich eine gestufte Einführung umgesetzt wurde. Dies ist angesichts der Komplexität und des Umfangs nach einhelliger Meinung gar nicht anders realisierbar. Ein Interviewpartner thematisierte zudem den permanenten Wechsel zwischen organisatorischer Anpassung und technischer Weiterentwicklung des bereits im Betrieb befindlichen CMS an seiner Hochschule.

Eine beobachtete Entwicklung bei der CMS-Einführung ist die zunehmende Adaption allgemein anerkannter Standards, Frameworks oder Good Practices aus dem allgemeinen IT-Management. In immer mehr Ausschreibungen findet sich explizit die Anforderung zum Einsatz bestimmter Standards bspw. für Projektmanagement. Nach Aussagen (I1) und (I2) enthalten aktuelle Ausschreibungen verstärkt Verweise auf bestimmte Projektmanagementmethoden, darunter vermehrt PRINCE2. Damit verbunden ist der Aufbau einer professionellen Projektorganisation. Um vor allem den zeitlichen Ablauf derartiger Projekte steuern zu können, ist die Bedeutung einer selbstständigen und kompetenten Projektorganisation für den Gesamterfolg von allen Befragten einheitlich bestätigt worden. Die an Hochschulen hochgeschätzte Konsens- und Gremienkultur steht dem jedoch nicht selten im Wege, bspw. wenn schnelle Entscheidungen erforderlich sind. Ein gangbarer Lösungsweg ist die projektbezogene Übertragung von Verantwortung an kleine entscheidungsfähige Gremien. Diese Projektgremien müssen sich aus Vertretern des Projekts, des Anbieters sowie der Fachbereiche bzw. Fakultäten zusammensetzen. Eine konkrete Realisierung hierfür lieferte (I4). Ansätze zeigte ebenfalls (I3), wobei hier der Kanzler der Hochschule in kritischen Projektsituationen eigenverantwortlich entschied bzw. mit persönlichen Gesprächen im kleinen Kreis Lösungen herbeiführte. Ein so energisches Durchgreifen forderte auch (I2) vom eingesetzten Projektleiter. Oftmals vermisse er aber diese Führungsqualität beim Projektleiter.

6 Zusammenführung der Ergebnisse

Eine wichtige Erkenntnis aus Literaturanalyse und Interviewstudie ist die Feststellung, dass bei differenzierter Betrachtung Erfolgsfaktoren von ERP-Einführungsprojekten einen zeitlichen Bezug zu den einzelnen Projektphasen aufweisen [SN01]. Dabei kann diese Betrachtung auch auf die der Einführung nachfolgende Phasen und somit zu einer lebenszyklusorientierten Betrachtung erweitert werden [HZ11], [ST13]. Für die Zusammenführung der extrahierten ERP- und CMS-Erfolgsfaktoren wurde dieser Zusammenhang ebenfalls angenommen und dazu ein einfaches Phasenmodell auf der Basis von [ZBP05], [BB09], [SKB10] und [HN12] entworfen. Die Phasen einer CMS-Einführung lassen sich demnach wie in Abb. 1 dargestellt abgrenzen:

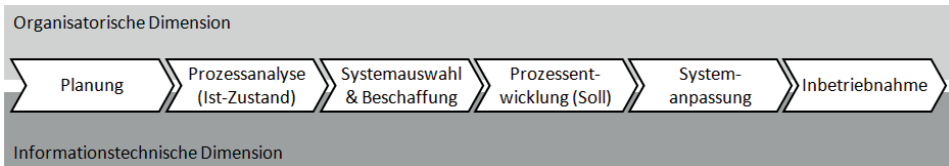


Abb. 1: Phasenmodell für die CMS-Einführung

In diese Phasen wurden die Erfolgsfaktoren für ERP- und CMS-Einführungsprojekte eingeordnet und durch Ergebnisse der Interviewstudie ergänzt. Tab. 2 zeigt alle 39 ermittelten Erfolgsfaktoren entlang der Phasen eines CMS-Einführungsprojektes und gibt dazu die jeweiligen Quellen an. Aussagen in den ERP-Quellen sind dabei teilweise sinngemäß übertragen worden. Z. B. weisen ERP-Quellen auf die Berücksichtigung von Besonderheiten der Organisationskultur hin, was als Beleg für CMS-Faktor 12) *Identifizierung hochschulspezifischer Besonderheiten* gewertet wurde. Für die Interviews werden die in Tab. 3 eingeführten Kürzel verwendet. Der Übersichtlichkeit halber werden die Interviews nur dort als Beleg aufgeführt, wo noch keine oder nur einmalige Nennung in der CMS-Literatur vorlag. Die in der letzten Spalte angegebene Klassifizierung wird aus Darstellungsgründen im Anschluss an Tab. 4 erläutert.

Die Verortung innerhalb der Phasen erfolgte jeweils dort, wo die Wirkung des Erfolgsfaktors beginnt und dieser damit für Entscheidungen bzgl. des Projekts relevant wird. Dies bedeutet jedoch nicht zwangsläufig, dass die Faktorwirkung mit Abschluss der Phase ebenfalls endet, sondern es gibt auch Faktoren, die phasenübergreifend u. U. bis zum Projektende wirksam sind wie bspw. Faktor 2) *Anhaltende Unterstützung durch die Hochschulleitung*.

Einf.-Phase	Erfolgsfaktor	Quellen CMS / Interviews	Quellen ERP	Klasse
Planung	1) Strategische Entscheidung der Hochschulleitung zur Einführung und Projektinitiierung	[Bi13], (I1), (I2), (I3), (I4)	[SL01], [SN01], [WT04], [ND06], [ST13]	2
	2) Anhaltende Unterstützung durch Hochschulleitung	[Bi13], (I2), (I3)	[SL01], [SN01], [WT04], [ND06]	2
	3) Definition eines Zielsystems auf Basis strategischer Positionierung	[St+07], [BB09], [KG07], [KI09]	[ND06], [HN12], [ST13]	1
	4) Kommunikation mit Beteiligten und Betroffenen	[Au14], [De+09], [FH09], [Ja+09], [KG07], [OT11], [SAA12]	[SL01], [SN01], [WT04], [ND06]	3
	5) Etablierung von Projektorganisation und -management	[BB09], [De+09], [KG07], [KI09], [OT11], [Bi13], [Au14]	[SN01], [MG03], [WT04], [ST13]	3
	6) Bereitstellung ausreichender Projektressourcen	[Sp+15]	[SL01], [ST13]	2
	7) Identifikation und Einbindung der Stakeholder	[De+09], [Ja+09], [KG07], [SAA12], [Gr+12], [Au14]	[SL01]	3
	8) Steuerung und Begleitung der Veränderungsprozesse	[St+07], [Br09], [De+09], [Ja+09], [KG07], [Bi13]	[SN01], [MG03], [WT04], [ND06], [KR10], [ST13]	1
	9) Einbindung der Akteure (Nutzer, Stakeholder, Projektteam)	[St+07], [De+09], [Ja+09], [KG07], [Au14]	[WT04], [ST13], [ZF14], [No15]	1
	10) Vorüberlegungen zu Einführungsstrategie und Pilotkandidaten	[Br+09], [FH09], [Ja+09], [Sp+15]	[MG03], [HN12], [ST13]	1
	11) Analyse des Mengengerüsts bzgl. Studiengängen, Ordnungen, Dokumente	[FH09], (I1)	[HN12]	1
	12) Identifizierung hochschulspezifischer Besonderheiten	[FH09], [OT11], [Gr+12], [SAA12], [Au15]	[ST13]	3
Prozessanalyse	13) Bestandsaufnahme der Ist-Prozesse	[St+07], [De+09], [FH09], [Gr+12], [OT11], [Bi13], [SA13], [Au15]	[Gr10], [HN12], [Gr14b]	3
	14) SWOT-Analyse als Basis von Optimierungen	[OT11], [SAA12], [Au15]	[HN12]	3
	15) Festlegung der Prozessmethodik	[SAA12], [SA13]	[HN12]	3
Systemauswahl & Beschaffung	16) Management funktionaler, qualitativer sowie anbieterbezogener Anforderungen	[BB09], [De+09], [Gr+12], [Au15]	[SN01], [BÇ05], [Gr10], [HN12], [Gr14b], [KR10], [ST13], [HZ11]	2
	17) Sondierung des Marktangebots	[BB09], [Au15], (I4)	[SN01], [Gr10], [HN12], [Gr14b], [KR10], [ST13]	2

Einf.-Phase	Erfolgsfaktor	Quellen CMS / Interviews	Quellen ERP	Klasse
	18) Anbieterkompetenz und -erfahrung	[BB09]	[ST13]	1
	19) Erfahrungsaustausch mit vergleichbaren Hochschulen	[Au15], (I4)	[Gr14b]	1
	20) Festlegung von Ausschreibungsformalitäten und -umfang bzw. Aufteilung in Lose	(I1), (I3)		4
	21) Akzeptanzsicherung durch Einbindung der Nutzer/Stakeholder	[BB09], [De+09], [Au14]	[Gr10], [Gr14b], [No15]	1
Prozessentwicklung	22) Entwurf der Soll-Prozesse nach Auswahlentscheidung	[Sp+15], (I2)	[HN12], [Gr14b]	2
	23) Prozessstandardisierung auf Basis von Good Practices	[FH09], [SAA12], [SA13], [Bi13]	[SL01]	3
	24) Prozessverbesserungen durch Systemfunktionalität	[SAA12], [OT11], [Au15], [K115]	[SN01]	3
Systemanpassung	25) Anpassungen der ausgelieferten Standardprozesse oder Angleichung der Ist-Prozesse	[St+07], [FH09], [OT11], [SAA12], [SA13], [K115]	[SL01], [SN01]	3
	26) Anpassung der Benutzeroberfläche und Schnittstellen	[BB09], (I4)	[BÇ05], [HN12]	2
	27) Sicherstellung von Ergonomie und Barrierefreiheit	[BB09]		3
	28) Schutz von sensiblen und personenbezogenen Daten	[BB09], [FH09]	[ST13]	3
	29) Datenmigration aus Altsystemen	[Gr+12], [SKB10], [Au15], [Sp+15]	[KR10], [ST13]	3
	30) Sicherstellung von Datenintegrität/-konsistenz und rechtssicherer Datenhaltung	[BB09], [Gr+12]	[ST13]	3
	31) Testplanung und -durchführung	(I4)	[HN12], [ST13]	2
	32) Fehlerbehebung und Systemoptimierung	(I3), (I4)		4
	33) Systemabnahme	(I3)		4
	Inbetriebnahme	34) Anbindung an Drittsysteme	[BB09], (I4)	[BÇ05], [KR10], [HN12], [ST13]
35) Abschaltung der Altsysteme			[HZ11]	2
36) Einweisung in neue Aufgaben/Rollen		[FH09], [K109]	[ST13], [ZF14]	1
37) Zielgerichtete Prozess- und Systemschulungen		[BB09], [FH09], [K109]	[HN12], [ST13]	1
38) Absicherung des (initial hohen) Supportbedarfs		[BB09], [Ja+09], [K109]	[KR10]	3
39) Fortlaufende Systemoptimierung im Betrieb		[St+07], (I3)	[KR10], [HN12], [HZ11], [ZF14], [ST13]	2

Tab. 5: Erfolgsfaktoren für die CMS-Einführung

Im Zuge der vergleichenden Analyse (ohne Interviews) wurden die CMS-Erfolgsfaktoren wie folgt klassifiziert:

- 1) Ausgewogene Nennung sowohl in ERP- als auch CMS-Literatur (bspw. Nr. 8);
- 2) Einseitige/überwiegende Nennung in ERP-Literatur (Nr. 34);
- 3) Einseitige/überwiegende Nennung in CMS-Literatur (bspw. Nr. 12).

Hinzu kommen als

- 4) Ergänzungen aus den Interviews wie bspw. Nr. 20.

Die Häufigkeitsverteilung dieser Klassen stellt sich wie folgt dar:

EF-Klasse	Anzahl	Anteil ⁵
1	10	26%
2	11	28%
3	15	38%
4	3	8%
Σ	39	100%

Tab. 6: Häufigkeitsverteilung

Diese Verteilung kann folgendermaßen interpretiert werden: Die geringe Zahl von Ergänzungen durch die Experteninterviews (Klasse 4) deutet auf eine weitgehende Vollständigkeit hin. Die CMS-Faktoren der Klasse 1 werden ähnlich häufig auch bei ERP-Projekten beschrieben und können damit als bestätigt gelten. Klasse 3 beinhaltet CMS-spezifische Faktoren, die dadurch nicht oder weniger relevant in ERP-Projekten sind. Am Interessantesten sind die Faktoren der Klasse 2, die in ERP-Projekten relevant sind, aber in CMS-Projekte bisher keine oder wenig Beachtung finden. Hier könnten Ansatzpunkte zur Vermeidung oder Lösung von heute noch in CMS-Projekten auftretenden Problemen zu finden sein.

7 Fazit und Ausblick

Für die Einführung von CMS konnten aus der einschlägigen Literatur und ergänzenden Interviews insgesamt 39 Erfolgsfaktoren identifiziert werden, deren praktische Relevanz mittels Experteninterviews validiert wurde. Dabei ließ sich ein Zusammenhang zum zeitlichen Fortschritt von Einführungsprojekten feststellen, über den die Zuordnung zu den Phasen eines Vorgehensmodells erfolgte. Zudem konnte gezeigt werden, dass sich Erfolgsfaktoren aus dem Bereich der Einführung von ERP-Systemen auf die Einführung von CMS übertragen lassen.

In der Praxis können die vorliegenden Ergebnisse als Checkliste genutzt werden, um bei CMS-Einführungsprojekten entlang der Phasen des Vorgehensmodells die Beachtung der

⁵ Gerundet.

jeweils angegebenen Faktoren abzusichern. Aus der Klassifizierung lassen sich weitere Schlüsse ziehen. An der Gesamtzahl der ermittelten Erfolgsfaktoren haben diejenigen, die bereits aus der Einführung von ERP-Systemen bekannt sind (Klasse 1 und 2) einen Anteil von rund 54 %. Daraus kann gefolgert werden, dass es dem Projekterfolg zuträglich sein könnte, wenn im Projektteam für die CMS-Einführungen Personen mitarbeiten, die über Kompetenzen und Erfahrungen mit ERP-Projekten verfügen. In diesem Fall sollte besonderes Augenmerk auf die verbleibenden Faktoren (Klasse 3 und 4, ca. 46 %) gerichtet werden, die als CMS-spezifisch klassifiziert wurden.

Bei einer Betrachtung aus Sicht des Projektmanagements fällt zudem auf, dass eine beträchtliche Anzahl der Erfolgsfaktoren sich auf Zusammenhänge beziehen, die entweder allgemein für Projekte oder speziell für Software-Einführungsprojekte längst als bedeutsam bekannt und in entsprechenden Methoden bzw. Empfehlungswerken (bspw. PRINCE2 oder PMBOK) fest verankert sind. Dies betrifft z. B. Punkte wie 7) Einbindung von Stakeholdern, 8) Steuerung und Begleitung der Veränderungsprozesse oder 38) Absicherung des (initial hohen) Supportbedarfs. Dies kann einerseits als Bestätigung für den bereits in der Einleitung angeführten Mangel an Erfahrung von Hochschulen mit der Durchführung komplexer IT-Projekte gelten, bietet andererseits aber auch Ansatzpunkte für die Überprüfung und den Aufbau spezifischer Kompetenzen vor Start eines CMS-Projekts.

Durch die geringe Größe der Stichprobe sowie die Beschränkungen bei der Auswahl der Interviewpartner ist die praxisbezogene Evaluierung im Rahmen dieser Arbeit nur begrenzt aussagekräftig. Die bisherigen Ergebnisse bieten aber die Grundlage für eine umfassendere Datenerhebung von CMS-Einführungsprojekten, um die entscheidenden Faktoren für eine erfolgreiche CMS-Einführung noch besser zu verstehen, bspw. durch Untersuchung der Signifikanz des jeweiligen Einflusses.

Literaturverzeichnis

- [Au14] Auth, G.: Zur Rolle des Stakeholder-Managements in IT -Projekten an Hochschulen - Erfahrungen aus der Einführung eines integrierten Campus-Management-Systems. In (Engstler, M. et al., Hrsg.): Tagungsband Projektmanagement und Vorgehensmodelle 2014, LNI P-236, Köllen, Bonn, 2014, S. 83-97.
- [Au15] Auth, G.: Prozessorientierte Anforderungsanalyse für die Einführung integrierter Campus-Management-Systeme. In (Abmann, Uwe et al., Hrsg.): Proceedings of Software Engineering & Management 2015. LNI P-239, Köllen, Bonn 2015, S. 446-461.
- [BB09] Bick, M.; Börgmann, K.: Referenzmodell zur Evaluation von Informationssystemen für ein integriertes Campus-Management. In: Hochschulmanagement, Nr. 4/2009, S. 108-112.
- [BÇ05] Baki, B.; Çakar, K.: Determining the ERP package selecting criteria. Business Process Management Journal, 11 (1) 2005, S. 75-86.

- [Bi13] Bick, M.: Zwischen Fachwissen und strategischer Entscheidung: Was muss die Hochschulleitung über IT wissen? CIO/IT-Governance-Modelle in deutschen Hochschulen. In (Stratmann, F., Hrsg.): IT und Organisation in Hochschulen. HIS-Selbstverlag, Hannover 2013, S. 1-21.
- [Br+09] Brune, H.; Jablonski, M.; Möhle, V.; Spitta, T.; Teßmer, M: Ein Campus-Management-System als evolutionäre Entwicklung. In [HKF09], S. 483-492.
- [Bu14] Buxmann, P.: Standardsoftware, Markt für. In (Kurbel, Karl et al., Hrsg.): Enzyklopädie der Wirtschaftsinformatik, 8. Aufl. Oldenbourg, München 2014. <http://www.enzyklopaedie-der-wirtschaftsinformatik.de>, 10.03.2016.
- [BSS98] Bancroft, N. H.; Seip, H.; Sprengel, A.: Implementing SAP R/3. 2. Aufl. Manning Publications, Greenwich, CT, USA 1998.
- [CD02] Cooke-Davies, T.: The “real” success factors on projects. International Journal of Project Management 20, 2002, S. 185-190.
- [De+05] DeLone, W.; Espinosa, J. A.; Lee, G.; Carmel, E.: Bridging Global Boundaries for IS Project Success. Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS), Maui, Hawaii, USA 2005, S. 1-10.
- [De+09] Degenhardt, L.; Gilch, H.; Stender, B.; Wannemacher, K.: Campus-Management-Systeme erfolgreich einführen. In [HKF09], S. 463-472.
- [DM92] DeLone, W. H.; McLean, E. R.: Information Systems Success: The Quest for the Dependent Variable. Information Systems Research, 3 (1) 1992, S. 60-95.
- [EY12] Ernst & Young (Hrsg.): Campus-Management zwischen Hochschulautonomie und Bologna-Reform. Studienergebnisse, 2012. [http://www.ey.com/Publication/vwLUAssets/Campus-Management_zwischen_Hochschulautonomie_und_BolognaReform_2012/\\$FILE/ErnsErnstY_Campus-Management-Studie.pdf](http://www.ey.com/Publication/vwLUAssets/Campus-Management_zwischen_Hochschulautonomie_und_BolognaReform_2012/$FILE/ErnsErnstY_Campus-Management-Studie.pdf), 08.03.2016.
- [FH09] Fischer, H.; Hartau, C.: STiNE an der Universität Hamburg – Zur Einführung eines integrierten Campus-Management-Systems. In [HKF09], S. 533-542.
- [Fr93] Fritz, W.: Die empirische Erfolgsfaktorenforschung und ihr Beitrag zum Marketing: Eine Bestandsaufnahme. Berichte des Instituts für Wirtschaftswissenschaften der TU Braunschweig, Nr. 93/12, Braunschweig, 1993.
- [GH13] Gregor, S., Hevner, A.: Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly, 37 (2) – Appendices/June 2013, S. A1-A6.
- [GP03] Gemell, M.; Pagano, R.: A Post-Implementation Evaluation of a Student Information System in the UK Higher Education Sector. Electronic Journal of Information Systems Evaluation, 6 (2), 2003, S. 95-10.
- [GPM15] GPM Deutsche Gesellschaft für Projektmanagement e. V. (Hrsg.): Ergänzung und Veränderung von Erfolgsfaktoren im Projektmanagement bei zunehmender Internationalisierung. O. O., o. J. [2015]. http://www.gpm-ipma.de/fileadmin/user_upload/Know-How/studien/Studie_Erfolgsfaktoren.pdf, 10.03.2016.
- [Gr10] Gronau, N.: ERP-Auswahl mittels ROI-Analyse – Risikoreduzierung und Nutzensteigerung. ERP Management, Nr. 3/2010, S. 17-20.

- [Gr14a] Gronau, N.: Enterprise Resource Planning-System. In (Kurbel, Karl et al., Hrsg.): Enzyklopädie der Wirtschaftsinformatik, 8. Aufl. Oldenbourg, München 2014. <http://www.enzyklopaedie-der-wirtschaftsinformatik.de>, 10.03.2016.
- [Gr14b] Gronau, N.: Die Zeit ist reif: Ein neues ERP-Auswahlverfahren. ERP Management, Nr. 10/2014, S. 45-47.
- [Gr+12] Grechenig, T.; Spitta, T.; Suppersberger, M.; Kleinert, W.; Steininger, R.; Kier, C.; Pöll, M.: Entwicklung und Betrieb eines Campus-Management-Systems – Aspekte zur Nachhaltigkeit am Beispiel TISS. In (Brand-Pook, H. et al., Hrsg.): Nachhaltiges Software Management. LNI P-209, Bonn, 2012, S. 135-152.
- [HN12] Hansmann, H.; Neumann, S.: Prozessorientierte Einführung von ERP-Systemen. In (Becker, J. et al., Hrsg.): Prozessmanagement. Springer, Berlin, Heidelberg, 2012, S. 329-366.
- [HKF09] Hansen, H.R.; Karagiannis, D.; Fill, H-G., Hrsg.: Business Services – Konzepte, Technologien, Anwendungen (Bd. 2). 9. Int. Tagung Wirtschaftsinformatik, Febr. 2009, Österr. Computer-Ges., Wien, 2009.
- [HZ11] Haddara, M.; Zach, O.: ERP Systems in SMEs: A Literature Review. 44th Annual Hawaii International Conference on System Sciences (HICSS), Maui, Hawaii, USA 2011, S. 1-10.
- [Ja+09] Janneck, M.; Adelberger, C.; Fiammingo, S.; Luka, R.: Von Eisbergen und Supertankern: Topologie eines Campus-Management-Einführungsprozesses. In [HKF09], S. 453-462.
- [KG07] Kuper, S.; Göcks, M.: Campus-Management-Systeme einführen: Faktoren erfolgreichen Projekt- und Change Managements. Wissenschaftsmanagement – Zeitschrift für Innovation, 13 (3), 2007, S. 40-46.
- [K109] Klug, H.: Erfolgsfaktoren bei der Umstellung von Informationssystemen an Hochschulen. In [HKF09], S. 473-482.
- [K115] Kleinschmidt, A.: Strategisches zu Campusmanagementsoftware. In (Cunningham, D. et al., Hrsg.): Tagungsband der GI-Jahrestagung Informatik 2015, Cottbus 2015, LNI P-246, Bonn, 2015, S. 701-712.
- [KR10] Kompalka, K.; Riha, I.: Erfolgsfaktoren und Fallstricke globaler ERP-Harmonisierungsprojekte. ERP Management, Nr. 3/2010, S. 41-44.
- [MB96] Muuns, A. K.; Bjeirmi, B. F.: The role of project management in achieving project success. Int. Journal of Project Management, 14 (2) 1996, S. 81-87.
- [MG03] Mandal, P.; Gunasekaran, A.: Issues in implementing ERP: A case study. European Journal of Operational Research, 146 (2) 2003, S. 274-283.
- [Me12a] Mertens, P.: Schwierigkeiten bei IT-Großprojekten der Öffentlichen Verwaltung. Arbeitspapier Nr. 2/2012 des Lehrstuhls Wirtschaftsinformatik I der Universität Erlangen-Nürnberg, 4. Aufl., Nürnberg 2012.
- [Me12b] Mertens, P.: Schwierigkeiten mit IT-Projekten der Öffentlichen Verwaltung – Neuere Entwicklungen. Informatik Spektrum, 35 (6), S. 433-446, 2012.

- [Me13] Meier, P.: Informatikchaos an der Uni Bern. Berner Zeitung online, 12.11.2013. <http://www.bernerzeitung.ch/region/kanton-bern/Informatikchaos--an-der-Uni-Bern/story/26148533>, 08.03.2016.
- [ND06] Nah, F. F.-H.; Delgado, S.: Critical success factors for enterprise resource planning implementation and upgrade. *Journal of Computer Information Systems*, 46 (5) 2006, S. 99-113.
- [No15] Norton, A.: Enterprise resource planning II – A review of critical success factors. *International Journal of Computer Science and Information Security*, 13 (11), November 2015, S. 5-14.
- [Nü14] Nüttgens, M.: Einführung von Anwendungssystemen. In (Kurbel, Karl et al., Hrsg.): *Enzyklopädie der Wirtschaftsinformatik*, 8. Aufl. Oldenbourg, München 2014. <http://www.enzyklopaedie-der-wirtschaftsinformatik.de>, 10.03.2016.
- [OT11] Oevel, G., Toschläger, M.: Einführung eines prozessorientierten Campusmanagement an der Universität Paderborn – ein Erfahrungsbericht. In (Degkwitz, A., Klapper, F., Hrsg.): *Prozessorientierte Hochschule*. Bock & Herchen, Bad Honnef 2011, S. 128-145.
- [Pe+08] Petter, S.; DeLone, W.; McLean, E.: Measuring information systems success: models, dimensions, measures, and interrelationships. *European Journal of Information Systems*, 17 (3) 2008, S. 236-263.
- [Ra09] Radenbach, W.: Integriertes Campus Management durch Verknüpfung spezialisierter Standardsoftware. In [HKF09], S. 503-512.
- [SAA12] Schreiter, J.; Alt, R.; Auth, G.: Business Engineering bei der Einführung von Campus-Management-Systemen – Herausforderungen und Potenziale. In (Goltz, U. et al., Hrsg.): *Tagungsband der GI-Jahrestagung Informatik 2012*, Braunschweig 2012, LNI P-208, Bonn, 2012, S. 642-656.
- [SA13] Schreiter, J., Alt, R.: Modellierungswerkzeuge zur Abbildung der Standardisierung und Individualisierung bei Hochschulprozessen. In (Horbach, M., Hrsg.): *Tagungsband der GI-Jahrestagung Informatik 2013*, Koblenz 2013. LNI P-220, Bonn, 2013, S. 320-333.
- [SKB10] Sprenger, J.; Klages, M.; Breitner, M. H.: Wirtschaftlichkeitsanalyse für die Auswahl, die Migration und den Betrieb eines Campus-Management-Systems. *Wirtschaftsinformatik*, 52 (4) 2010, S. 211-224.
- [SKW05] Schmalen, C.; Kunert, M.; Weindlmaier, H.: Erfolgsfaktorenforschung: Theoretische Grundlagen, methodische Vorgehensweise und Anwendungserfahrungen in Projekten für die Ernährungsindustrie. In (Bahrs, E. et al., Hrsg.): *Beiträge der 45. Jahrestagung der Gesellschaft für Wirtschafts- und Sozialwissenschaften des Landbaues (GEWISOLA)*, Landwirtschaftsverlag, Münster-Hiltrup, 2005, S. 351-362.
- [SL01] Skok, W.; Legge, M.: Evaluating Enterprise Resource Planning (ERP) Systems Using an Interpretive Approach. *Proceedings of the 2001 ACM SIGCPR Conference on Computer Personnel Research*. New York 2001, S. 189-197.
- [SN01] Somers, T. M.; Nelson, K.: The Impact of Critical Success Factors across the Stages of

- Enterprise Resource Planning Implementations. 34th Annual Hawaii International Conference on System Sciences (HICSS), Maui, Hawaii, USA 2001, S. 1-10.
- [Sp+15] Spitta, T.; Carolla, M.; Brune, H.; Grechenig, T.; Strobl, S.; vom Brocke, J.: Campus-Management Systeme als Administrative Systeme. Basiswissen und Fallbeispiele zur Gestaltung und Einführung. Reihe essentials, Springer Vieweg, Wiesbaden 2015.
- [St+07] Stender, B.; Jablonski, M.; Brune, H.; Möhle, V.: Campus Management von der Hochschule aus gedacht – Werkstattbericht aus der Universität Bielefeld. Wissenschaftsmanagement, 6 Nov./Dez. 2007, S. 19-26.
- [ST13] Shaul, L.; Tauber, D.: Critical Success Factors in Enterprise Resource Planning Systems: Review of the Last Decade. ACM Computing Surveys, 45 (4) August 2013, S. 1-39.
- [UB15] Universität Bielefeld (Hrsg.): Projekt Campusmanagement: Zusammenarbeit mit CAS beendet. Bielefeld 2015.
http://ekvv.uni-bielefeld.de/blog/uniintern/entry/ende_von_bis_3_0, 11.03.2016.
- [WT04] Wong, B.; Tein, D.: Critical Success Factors for ERP Projects. Journal of the Australian Institute of Project Management, 24 (1) 2004, S. 28-31.
- [ZF14] Zouine, A.; Fenies, P.: The Critical Success Factors Of The ERP System Project: A Meta-Analysis Methodology. The Journal of Applied Business Research, 30 (5) September/October 2014, S. 1407-1448.

Neue Wege des Kundensupports im IT-ServiceDesk des IT Centers der RWTH Aachen University

Ingo Hengstebeck¹, Sarah Grzemski²

Abstract: Durch schnelle Innovationsprozesse in der IT steigt die Mobilität der Anwender von IT-Dienstleistungen rapide an. Smartphones und Tablets sind zu unseren täglichen Begleitern geworden. Daraus resultiert ein Wandel der Kommunikationsgewohnheiten. Die Erwartungshaltung, innerhalb kürzester Zeit auf Anfragen eine Antwort zu bekommen, ist sehr hoch. Dabei steht die Kommunikation per E-Mail und Telefon in Konkurrenz zu Chats und Informationsportalen. Ebenso erfahren Blogs mit entsprechenden Communities eine zunehmende Beliebtheit. Dieser Trend zur Nutzung moderner Kommunikationsmittel wird auch durch die jährliche Nutzerzufriedenheitsbefragung des IT-ServiceDesk bestätigt. Auf Grund dieser Ergebnisse und der oben genannten Entwicklungen, beschäftigte sich das IT-ServiceDesk im Jahr 2015 intensiv mit Überlegungen sowie der konzeptionellen Einführung zukunftsorientierter Kommunikationskanäle. Ziel ist es, neben den etablierten Kommunikationskanälen Schritt-für-Schritt moderne Kommunikationsformen anzubieten. Dieser Artikel beschreibt die Vorgehensweise des IT-ServiceDesk, um langfristig eine entsprechende Lösung zu etablieren.

Keywords: Service Desk, Chat Support, Multi-Channel-Kommunikation, Omni-Channel-Kommunikation, Anwender-Support, Ticketportal

1 Einleitung

„Die Bedeutung der Informationstechnik (IT) im Kontext von Hochschulen ist kaum zu überschätzen. Sie durchdringt nahezu alle Bereiche von Forschung und Lehre.“ [In16]

Vor diesem Hintergrund ist das IT Center der RWTH Aachen University als eine zentrale Einrichtung der Hochschule zu verstehen, welche die Einrichtungen der RWTH, Mitarbeitende, Studierende sowie externe Partner mit insgesamt ca. 34 IT-Basis- und weiteren Individualdienstleistungen versorgt. Als der führende Anbieter von IT-Diensten ist das IT Center dafür verantwortlich, Standards in der IT-Versorgung bereitzustellen, die an einer Hochschule verfügbar sein müssen [In16].

Im Jahr 2008 entschied sich das IT Center, IT Service Management nach der IT Infrastructure Library (ITIL) einzuführen, um besser auf die Wünsche und Bedürfnisse der Kunden einzugehen. Bei ITIL handelt es sich um einen de facto Standard, welcher anhand von Good-Practice IT Service Management (ITSM) abbildet und organisatorisch

¹ IT Center/RWTH Aachen University, IT-ServiceDesk, Seffenter Weg 23, 52074 Aachen, hengstebeck@itc.rwth-aachen.de

² IT Center/RWTH Aachen University, IT-ServiceDesk, Seffenter Weg 23, 52074 Aachen, grzemski@itc.rwth-aachen.de

verankert [BHG11]. Durch die Ausrichtung nach ITSM werden folgenden Ziele verfolgt:

- Das IT Center der RWTH Aachen soll als verlässlicher Partner für alle Nutzergruppen wahrgenommen werden
- Das IT Center stellt sich als hochschulübergreifender Partner auf
- Das IT Center stärkt seine Kundenorientierung
- Die am IT Center stattfindenden Prozesse werden optimiert und es findet eine Erhöhung der Transparenz und Verbindlichkeit gegenüber den Nutzenden statt [BHG11]

Im Zuge dessen wurde 2010 das IT-ServiceDesk als zentraler Ansprechpartner etabliert. Das IT-ServiceDesk ist der erste Anlaufpunkt für alle Fragen rund um die IT-Dienste des IT Centers und ist u.a. verantwortlich für die Beantwortung definierter Anfragen sowie für die Koordination der Kommunikation zwischen den Anwendern und den Fachabteilungen bei komplexeren Fragestellungen [BHG11]. Hierdurch wird dem Wunsch der Anwender nach qualitativ hochwertigen Diensten und Support entsprochen. Um dies zu garantieren, unterliegen die Prozesse im IT-ServiceDesk einer ständigen Qualitätssicherung. Bereits im Jahr 2014 begann das IT-ServiceDesk mit der Einführung eines Qualitätsmanagementsystems nach DIN ISO 9001, welches im Mai 2016 entsprechend zertifiziert wurde. Das mit der Etablierung des IT-ServiceDesk konzipierte Reporting unterstützt das Qualitätsmanagement und die Qualitätssicherung. Defizite werden erkannt und es wird entsprechend gegengesteuert wie z.B. durch Schulungen, Ticketbesprechungen in Teammeetings und Hospitationen in den Fachabteilungen. Gleichzeitig dient das Reporting dazu, immer wiederkehrende Ereignisse, wie den Semesterstart besser zu bewerten und mit Hilfe einer Personaleinsatzplanung Engpässe zu vermeiden. Nur durch das Zusammenspiel klar definierter Prozesse kann eine hohe Serviceorientierung gewährleistet werden [Tm15].

Die Vielfalt und Komplexität der Dienstleistungen und die Heterogenität der Nutzenden stellen eine große Herausforderung für das IT-ServiceDesk dar. Die Mitarbeitenden des IT-ServiceDesk benötigen einerseits ein breites Wissen über die angebotenen Dienstleistungen und andererseits eine sehr hohe kommunikative Kompetenz, um einen professionellen und kundenorientierten Support anzubieten. Um besser auf Kundenbedürfnisse zu reagieren und ein besseres Verständnis für die Anforderungen der Kunden zu erhalten, führt das IT Center seit dem Jahr 2011 jährlich eine Nutzerzufriedenheitsumfrage durch. Die Ergebnisse der im Jahr 2013 durchgeführten Umfrage gaben den Anstoß, die Möglichkeiten moderner Kommunikationskanäle zu evaluieren. Aus den Ergebnissen der Umfrage wurde deutlich, dass sich die Nutzenden auch die Möglichkeit wünschen, neben den etablierten Kanälen (E-Mail, Telefon, persönlich) Kommunikationskanäle wie z.B. SocialMedia zu nutzen. Diese Ergebnisse wurden durch die Umfragen in den Folgejahren unterstützt. In der Umfrage im Jahr 2015 wurde explizit der Wunsch nach einem Chat Support geäußert.

2 Neue Kommunikationskanäle im IT-ServiceDesk

Auf Grund des Selbstverständnisses des IT Centers als moderner IT-Dienstleister und der Forderung der Anwender nach neuen Kommunikationskanälen setzte sich das IT-ServiceDesk intensiv mit der Etablierung anderer, modernerer Kanäle auseinander und diskutierte die Vor- und Nachteile der möglichen zusätzlichen Kommunikationskanäle wie z.B. die Eröffnung von Tickets über eine Webschnittstelle, Facebook, Blog oder Chat. Im Zusammenhang mit der Einführung eines modernen Kommunikationsansatzes stehen die beiden Begriffe Multi-Channel und Omni-Channel-Kommunikation [MP16]. Der Begriff Multi-Channel beschreibt die parallele Bereitstellung von Informationen oder Support über verschiedene Kanäle, wobei diese autark arbeiten und somit technisch voneinander getrennt sind. Bei der Omni-Channel-Kommunikation können die Kommunikationskanäle, wie bei der Multi-Channel-Kommunikations, getrennt genutzt werden, jedoch ist eine kanalübergreifende Nutzung kennzeichnend (siehe auch Abb. 1). Ein Wechsel des Kommunikationskanals ist jederzeit möglich. Diese Möglichkeit bietet den supportleistenden Einrichtungen, nach Absprache mit dem Anwender, gezielt auf den Kanal zu wechseln, der die angeforderte Hilfestellung am besten unterstützt. Dadurch wird das Supporterlebnis positiv beeinflusst und das Problem zeitnah gelöst. Ein wichtiger Unterschied besteht auch im Management aller relevanten Daten in einem zentralen System. Dadurch ist es möglich, die Kundenhistorie leicht nachzuvollziehen und es müssen keine aufwändigen technischen Lösungen entwickelt werden, um den Datenbestand in verschiedenen Systemen konsistent zu halten [MP16].

Maximum an Ingeration mit allgegenwärtiger mobile Commerce Option (Social - Local - Mobile)		NO-LINE
Zentralisierte Produkt- und Kundendatenbank, Touchpoints und (Kauf-)historien sind voll erfasst & stets aktuell. Ubiquitäre Zugriffe auf Gesamtinventar und -informationen	OMNI	Verschiedene, seperierte Absatzkanäle; ein Knotenpunkt.
Channel-Hopping (teilweise verbundene IT)	CROSS	Anwender bekommt kanalübergreifendes Kauf- bzw. Serviceerlebnis.
Mindestens zwei Kanäle (auf jeden Fall mit Online)	MULTI	Technische Systeme laufen separat. Transaktionen werden kanalorientiert abgewickelt.

Abb. 1: Multi-, Cross-, Omni-Channel-Kommunikation [MP16]

Um einen Überblick über bereits etablierte Lösungen anderer Hochschulen zu bekommen, wurden die Supportkanäle der führenden technischen Hochschulen in Europa (TU9 und IDEA League) analysiert. Zur Analyse wurden die Webseiten der

entsprechenden Hochschulen genutzt. Der Fokus lag auf der Bereitstellung von Nutzerunterstützung durch einen Chat. Bis auf die Delft University of Technology bot keine der untersuchten Hochschulen eine Unterstützung per Chat oder Blog mit angeschlossener Community an. Um einen Überblick an bereits etablierten Softwarelösungen bezüglich des Chat Supports im Speziellen und Multi-Channel und Omni-Channel zu erhalten, wurden im Jahr 2015 und 2016 verschiedene Fachkongresse wie z.B. Call Center World und ServiceDesk Forum besucht. Hier wurden in intensiven Beratungsgesprächen mit den verschiedenen Anbietern Lösungsansätze diskutiert. Dabei kristallisierte sich die Integration einer Multi-Channel-Lösung oder Omni-Channel-Lösung in das im IT Center verwendete Tickettool als besondere Fragestellung heraus. Die auf dem Markt befindlichen Lösungen beinhalten größtenteils bereits Ticketinglösungen, wobei die Integration in bestehende Lösungen nur mit viel Anpassungsaufwand möglich ist. Durch die Gespräche wurde die Entscheidung unterstützt, auf eine Multi-Channel-Lösung zu setzen.

Des Weiteren würde der direkte Einsatz einer Omni-Channel-Lösung große Veränderungen in der gesamten Supportinfrastruktur und einen hohen finanziellen Aufwand bedeuten. Ziel ist die Verfolgung einer Strategie der kleinen Schritte, um Erfahrungen mit modernen Kommunikationskanälen zu sammeln.

Bedingt durch die Entscheidung müssen die im Support tätigen Mitarbeitenden gleichzeitig mehrere Anwendungen nutzen, um Kundenanfragen zu bearbeiten. Zwischen den einzelnen Systemen besteht keine Verbindung, um eine zusammenhängende Historie der Anfragen eines einzelnen Nutzers über verschiedene Kanäle bereitzustellen oder alle Anfragen über eine Anwendung zu bearbeiten. Hier gilt es, langfristig eine entsprechende Software einzuführen oder eine komplette Integration in das bestehende Tickettool zu implementieren. Auch wenn noch keine allumfassende Lösung existiert, sammeln die Mitarbeitenden des IT-ServiceDesk schon erste Erfahrungen mit den neuen Kommunikationsmitteln.

Neben der o.g. Analyse und Marktevaluation wurden arbeitsorganisatorische Maßnahmen, wie z.B. eine angepasste Personaleinsatzplanung diskutiert, da durch die Etablierung weitere Kommunikationskanäle keine zusätzliche Arbeitsbelastung der Mitarbeitenden generiert werden darf. Diskussionen, wie man durch das Angebot neuer Kommunikationskanäle einen Mehrwert sowohl für die Anwender, wie auch für das IT Center schaffen kann, rückten im weiteren Verlauf in den Fokus der IT-ServiceDesk Leitung. Als Mehrwert für die Anwender wurden Funktionalitäten wie z.B. die Möglichkeit des Passwortresets und die einfache Übermittlung von Bildern identifiziert, aber auch die synchrone Bearbeitung von schriftlichen Anfragen. Auf Seite des IT Centers ergibt sich der Mehrwert, dass das Selbstbild des IT Centers als moderner IT-Provider verstärkt wird. Ebenso wird die Kundenorientierung des IT Centers deutlich hervorgehoben und die Geschwindigkeit der Bearbeitung von Anfragen erhöht.

Abschließend entschied sich das IT-ServiceDesk, die folgenden zusätzlichen Kommunikationskanäle einzuführen:

- Webschnittstelle zum Tickettool (Ticket-Portal)
- Facebook
- Blog
- Chat

Die oben aufgeführten zusätzlichen Kommunikationskanäle bieten im Gegensatz zu den bereits etablierten Kanälen den Vorteil, dass der Zugriff durch starke Verbreitung von Smartphones auf diese von überall möglich ist und gerade von Studierenden und jüngeren Mitarbeitenden bevorzugt werden. Neben der ständigen Zugriffsmöglichkeit, ist auch noch zu erwähnen, dass die Beantwortung von Fragen schneller stattfindet. In einer Community erhalten die Anwender von einem anderen Anwender eine Antwort und es muss nicht auf einen Supportmitarbeitenden gewartet werden oder auf die Antwort auf eine E-Mail. Ein Chat hat den Vorteil, dass bereits angezeigt wird, ob genug freie Mitarbeitende zur Beantwortung zur Verfügung stehen oder wie lange die Wartezeit ist. All diese Faktoren generieren einen hohen Mehrwert der neuen Kanäle im Vergleich zu den bereits etablierten.

2.1 Ticketportal und Nutzung von Facebook

Um möglichst zeitnah erste Erfahrungen mit der Einführung und Nutzung neuer Kommunikationskanäle zu sammeln, wurde beschlossen, Kanäle zu wählen, die sich ohne große technische und organisatorische Maßnahmen einführen ließen. Dazu zählen das Ticketportal und die Kommunikation via Facebook.

Im März 2015 startete das Ticketportal offiziell. Es ermöglicht den Anwendern, Anfragen online zu stellen, den Status zu überprüfen und gegebenenfalls zu ergänzen. Gerade bei Personen, die viele Anfragen an das IT Center stellen, erfreut sich das Portal großer Beliebtheit, da schnell ein Gesamtüberblick über den Status der Anfragen gegeben wird. Als weiteren Benefit für die Nutzenden ist hervorzuheben, dass diese bei der Erstellung einer Anfrage über das Ticketportal durch kontext-sensitive Formulare geleitet werden und anhand der Auswahl einen Link zu der entsprechenden Dokumentation angezeigt bekommen. Ferner werden vorab alle notwendigen Angaben zur Bearbeitung der Anfrage abgefragt. Die Abfrage der relevanten Information beschleunigt die Bearbeitung durch das IT-ServiceDesk. Um die Nutzung des Ticketportals weiter zu erhöhen sind für das Jahr 2016 entsprechende Marketingmaßnahmen, wie Flyer, Blogbeitrag usw. geplant. Die Anzahl der eröffneten Tickets über das Ticketportal ist in Abbildung 2 dargestellt. Im Verhältnis zu der Gesamtanzahl der Tickets über einen der anderen Eingangskanäle sind diese noch verhältnismäßig gering.

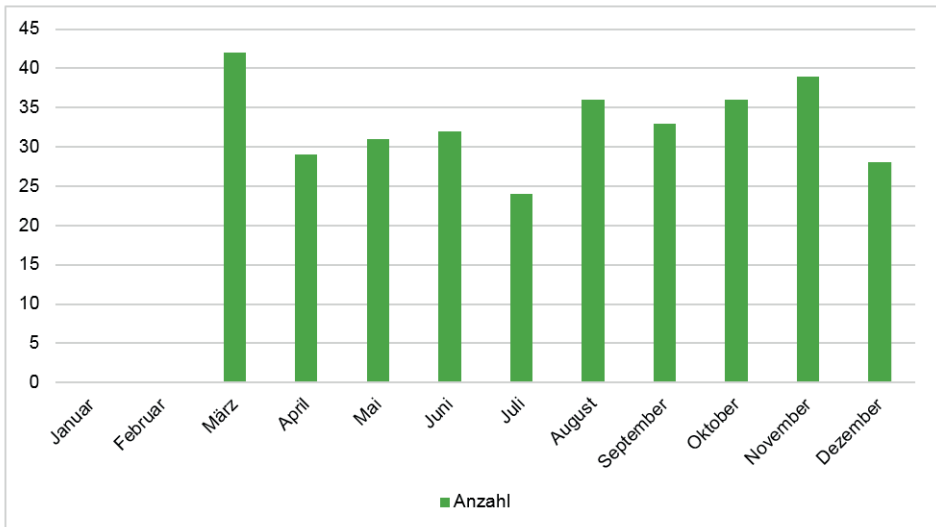


Abb. 2: Anzahl der eröffneten Tickets über das Ticketportal 2015

Der Einsatz von Facebook zur Kommunikation mit den Anwendern, im Speziellen mit Studierenden, erfolgt grundsätzlich in enger Abstimmung mit der Pressestelle der RWTH Aachen University. Das IT Center betreibt keine eigene Facebookseite, sondern nutzt den zentralen Auftritt der RWTH Aachen University. In ausführlichen Gesprächen mit den Verantwortlichen der Pressestelle wurde erläutert, worauf bei Facebookposts zu achten ist. Zielgruppen, die über Facebook angesprochen werden sollen, müssen definiert werden genauso wie die Zeit in welcher eine Rückmeldung auf mögliche Rückfragen zu Posts erfolgen muss. Die Frage, wie ein Facebookpost gestaltet sein muss, damit eine möglichst große Reichweite erzielt wird, bedurfte einiger „Pilotposts“. Wichtig ist, dass durch den Post ein Mehrwert für die Anwender geschaffen und Emotionen erzeugt werden. Im Bereich der IT ist dies eine Gratwanderung. Beispielsweise hat ein Post, der auf eine Umfrage hinweist, eine sehr viel geringere Reichweite als ein Post der darauf hinweist, dass ein großer Hardwarehersteller eine Rückrufaktion gestartet hat und im IT-ServiceDesk die entsprechenden Komponenten ausgetauscht werden können. Dies wird deutlich durch die Statistiken der beiden entsprechenden Facebookposts (vgl. Abb. 3 und 4). Facebook als Kanal zu den Anwendern wird durch das IT-ServiceDesk hauptsächlich genutzt, um auf neue Dienste oder Änderungen an bestehenden Diensten hinzuweisen.

Beitragsdetails

RWTH Aachen University
Gestern um 08:53 · Bearbeitet ·

Eure Meinung ist gefragt
Das IT Center arbeitet an der Verbesserung seines Supports und hat unter www.itc.rwth-aachen.de/umfrage bis 7.11. eine Befragung zu eurer Zufriedenheit mit dem Support des IT Centers online.
Den Fragebogen auszufüllen dauert nur 3 Minuten, die Ergebnisse könnt ihr ab 21.11. auf der Homepage des IT Centers einsehen.

Umfrage

EXZELLENT
 hervorragend
 sehr gut
 gut
 durchschnittlich
 unterirdisch

3.558 erreichte Personen **Beitrag bewerten**

Gefällt mir · Kommentieren · Teilen · **4**

3.558 erreichte Personen

Gemeindete Daten können bis zu 30 Minuten verspätet sein. X

4 „Gefällt mir“-Angaben, Kommentare und geteilte Inhalte

4 „Gefällt mir“-Angaben	4 Zu einem Beitrag	0 Zu geteilten Inhalten
0 Kommentare	0 Zu einem Beitrag	0 Zu geteilten Inhalten
0 Geteilte Inhalte	0 Zu einem Beitrag	0 Zu geteilten Inhalten

405 Klicks auf Beiträge

218 Fotoaufhufe	11 Klicks auf Links	176 Andere Klicks
-----------------	---------------------	-------------------

NEGATIVES FEEDBACK

- 0 Beitrag verbergen Alle Beiträge verbergen
- 0 Als Spam melden Seite gefällt mir nicht mehr

Abb. 3: Facebookpost Umfrage

Beitragsdetails

RWTH Aachen University
Gepostet von Anna Altelix (1) · 7. März um 14:09 ·

Apple-Rückruf: Das IT-ServiceDesk des IT Centers unterstützt beim Netzteilstecker-Tausch
Der Apple-Konzern ruft freiwillig Netzteilstecker zurück, die für den Gebrauch in Kontinentaleuropa, Australien, Neuseeland, Südkorea, Argentinien und Brasilien entwickelt wurden. In äußerst seltenen Fällen können Netzteilstecker mit zwei Stiften, die zwischen 2003 und 2015 mit Mac- und bestimmten iOS-Geräten sowie als Teil des Apple Reise-Adapter-Kit unter anderem in Mitteleuropa ausgel... Mehr anzeigen

40.584 erreichte Personen

141 Reaktionen, Kommentare und geteilte Inhalte

87 Gefällt mir	76 Zu einem Beitrag	11 Zu geteilten Inhalten
1 Haha	0 Zu einem Beitrag	1 Zu geteilten Inhalten
29 Kommentare	19 Zum Beitrag	10 Zu geteilten Inhalten
24 Geteilte Inhalte	23 Zum Beitrag	1 Zu geteilten Inhalten

6184 Klicks auf Beiträge

586 Fotoaufhufe	1046 Klicks auf Links	4552 Andere Klicks
-----------------	-----------------------	--------------------

NEGATIVES FEEDBACK

- 10 Beitrag verbergen Alle Beiträge verbergen
- 0 Als Spam melden Seite gefällt mir nicht mehr

76 **Beitrag bewerten**

14 Kommentare 23 geteilte Inhalte

Gefällt mir · Kommentieren · Teilen

Abb. 4: Facebookpost Rückrufaktion

2.2 Der IT Center Blog

Der IT Center Blog wurde 2015 als moderne Informations- und Kommunikationsplattform eingeführt, welche wöchentlich zielgruppenspezifische Beiträge bietet. Zu den Zielgruppen gehören Studierende, Mitarbeitende und Administrierende der RWTH Aachen University. Ziel ist es, einen Großteil der RWTH Aachen-Angehörigen zu erreichen. Die Unterteilung in die einzelnen Personengruppen ist notwendig, um alle Gruppen gleichermaßen anzusprechen und keine einseitige Ausrichtung der Beiträge zu fördern.

Zur Steuerung des Blogs existiert ein definierter Prozess, der die Stränge „Erstellung eines Beitrages“ sowie „zeitliche Koordination“ beinhaltet.

Ein Blog-Beitrag durchläuft beginnend bei einem Themenvorschlag bis hin zur Veröffentlichung eines Beitrages unterschiedliche Stadien und beteiligte Personengruppen, die der nachfolgenden Übersicht (vgl. Abb.5) zu entnehmen sind.

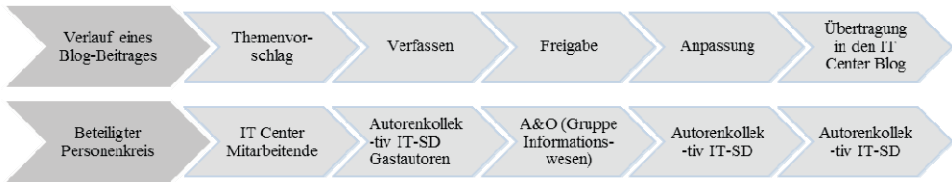


Abb. 5: Status eines Blog-Beitrages mit dem jeweiligen, beteiligten Personenkreis

Seit dem Start des Blogs werden wöchentlich Informationen für die Anwender der IT-Dienste bereitgestellt. Zu den Blog-Inhalten gehören wiederkehrende Beiträge wie z.B. zum Semesterstart (Willkommensgruß für Studierende), wochenaktuelle Meldungen (Informationen zu neuen Diensten, z.B. sciebo) und Ankündigungen von IT Center Veranstaltungen (Parallel Programming in Computational Engineering, Infotag). Des Weiteren werden Allgemeine Informationen mit Hintergrundinformationen zu den IT-Diensten (Software Shop/Download vom ftp-Server) und fachspezifische Informationen z.B. bei speziellen Updates (Microsoft Windows 10) veröffentlicht. Zeitweise werden zudem identitätsstärkende Meldungen (HiWine/HiWi am IT-ServiceDesk, Team des IT-ServiceDesk) bereitgestellt. Am 1. April 2016 wurde die Kommentarfunktion des Blogs freigeschaltet. Auch hier galt es, im Vorhinein die Antwortzeiten auf Kommentare zu definieren. Auf Kommentare erfolgt eine Antwort durch das IT-ServiceDesk innerhalb von 24 Stunden an Werktagen. Dieser Zeitraum wurde so gewählt, da auch auf Anfragen per E-Mail eine erste „menschliche“ Antwort nach spätestens 24 Stunden erfolgt. Durch die Freigabe der Kommentarfunktion wird das Ziel verfolgt, Nutzende an moderierten Diskussionen zu beteiligen und so eine aktive Community zu etablieren, in welcher sich die Anwender gegenseitig bei Fragen und Problemen unterstützen. Um den Communitygedanken an die Anwender zu kommunizieren, wird der Blog aktiv durch das IT-ServiceDesk vorgestellt. Hierzu werden Einführungsveranstaltungen für

Studierende und Austauschgespräche mit den Administratoren der verschiedenen Institute genutzt.

Die Erfahrungen aus Networking-Gruppen von Studentenwohnheimen und Institutsadministratoren haben gezeigt, dass unter Anwendern ein hohes Maß an Bereitschaft besteht, andere Anwender zu unterstützen.

2.3 Chat Support über die RWTHApp und Web Chat

Der Chat Support des IT-ServiceDesk ist über verschiedene Wege zu erreichen. Hervorzuheben ist die Möglichkeit der Nutzung über die stark verbreitete RWTHApp. Die RWTHApp ist verfügbar für iPhone, Android und Windows Phone. Für Studierende, Mitarbeitende und Besucher der RWTH Aachen bietet die RWTHApp speziell zugeschnittene Funktionen, um den universitären Alltag zu erleichtern. Dazu zählen eine Übersicht der anstehenden Termine, eine Personensuche oder der Zugriff auf das E-Learning-Portal L2P.

Die Entwicklung der RWTHApp wurde durch die Studierenden der Hochschule angestoßen. In einem Wettbewerb konnten die Studierenden ihre Ideen für zukünftige Funktionalitäten einbringen. Nachdem der Wettbewerb abgeschlossen war, wurden Grundfunktionalitäten implementiert. Alle weiteren Funktionalitäten wie z.B. der Mensaspiseplan wurden nach und nach in die RWTHApp integriert. Durch diese agile Vorgehensweise konnte die App sehr zeitnah und anwenderorientiert bereitgestellt werden. Es finden immer noch regelmäßige Austauschgespräche mit den Vertretern der Studierenden statt, um neue Funktionalitäten und deren Umsetzungsmöglichkeiten zu prüfen. Beispielsweise wird eine Lernraumsuche, die alle Lernräume der RWTH Aachen inklusive einer Anzeige der aktuellen Auslastung umfasst, diskutiert. Im Rahmen der Austauschgespräche wurde ebenfalls das Audience-Response-System mit dem Namen „Direktfeedback“, welches die Grundlage des Chat Supports ist, entwickelt. Das „Direktfeedback“ ermöglicht es Studierenden, in großen Veranstaltungen über die RWTHApp Fragen an den Dozierenden zu stellen. Dieser erhält die Fragen auf seinem Notebook und kann während der Veranstaltung darauf eingehen:

„In large-scale lectures with more than 1.000 Students, possibly in different lecture halls, interaction and communication between the students and the teachers is almost impossible. Using RWTHApp we present a way to exchange messages between students and teachers to re-enable communication during lectures but also during other events.“
[Po15]

Diese Funktionalität konnte so angepasst werden, dass die Anforderungen des Chat Supports aus Sicht der Anwender erfüllt werden.

Auch wurde der Chat Support in das Dokumentationsportal des IT Centers (<https://doc.itc.rwth-aachen.de>) integriert. Die Möglichkeit, den Chat Support über zwei verschiedene Systeme in Anspruch zu nehmen, wurde gewählt, um einerseits

Anwendern, die gerade die Webseiten des Dokumentationsportals besuchen, bei Fragestellungen zu den Inhalten zu unterstützen und andererseits, um den Anwendern auch unterwegs über die RWTHApp Hilfestellungen zu bieten. Die Mobilität der Studierenden ist ein Faktor, der in hochqualifizierten Supportprozessen Berücksichtigung finden muss.

Durch die Integration in die RWTHApp ist es auch möglich, dass Anwender den Chat Support authentifiziert nutzen können. Durch diese Authentifizierung erhält der Mitarbeitende im IT-ServiceDesk direkt nützliche Informationen über den Anfragenden und kann so gezielter helfen.

Eine besondere Herausforderung war die Entwicklung eines Backends für die Mitarbeitenden des IT-ServiceDesk, um Chats zu bearbeiten. Dieses muss einerseits sehr benutzerfreundlich gestaltet sein und andererseits gewährleisten, dass die Chats nach den bereits definierten Reportingregeln ausgewertet werden können. D.h. es muss sichergestellt sein, dass im Nachhinein ausgewertet werden kann,

- ob der Nutzende ein Studierender oder Mitarbeitender war,
- auf welches Thema sich die Frage bezog,
- ob die Frage direkt durch das IT-ServiceDesk gelöst wurde,
- und wie lange der Chat gedauert hat.

Um dies zu gewährleisten, wurde eine „lose Kopplung“ mit dem existierenden Tickettool geschaffen; d.h. nach Abschluss eines Chats wird automatisiert ein Ticket erstellt, welches alle für das Reporting notwendigen Daten enthält.

Zusätzlich wurde beachtet, dass Mitarbeitende nur eine vorher definierte Anzahl von Chats bearbeiten können. Bei einer maximalen Auslastung der Mitarbeitenden werden keine neuen Chats angenommen und die Anwender erhalten eine entsprechende Meldung. Auch wurden Sprachregelungen für den Chat Support aufgestellt und Antwortzeiten definiert. Um die Mitarbeitenden bei der Einhaltung der Sprachregelungen zu unterstützen, wird eine Software bereitgestellt, in welcher vordefinierte Textbausteine inhaltlich zentral gepflegt werden. Diese Software wird nicht nur bei der Bearbeitung von Chats genutzt, sondern auch bei der Beantwortung von Tickets. Die Nutzung der Textbausteine ist nicht verpflichtend, unterstützt die Mitarbeitenden aber dabei, inhaltlich korrekte Antworten zu verfassen.

Der „Go Live“ des Chat Supports war im November 2015. Es wurden noch keine Werbemaßnahmen eingeleitet, da erst einmal Erfahrungen mit den neuen Werkzeugen und der anderen Supportsprache gesammelt werden sollen. Von November 2015 bis 21. Juni 2016 wurde der Chat Support 223-mal genutzt. Abbildung 6 stellt die Anzahl der monatlichen Chats dar. Zur Erhöhung des Bekanntheitsgrades sind entsprechende Marketingmaßnahmen wie z.B. Facebookposts, Flyer und Meldungen auf den Webseiten des IT Centers für das Jahr 2016 geplant.

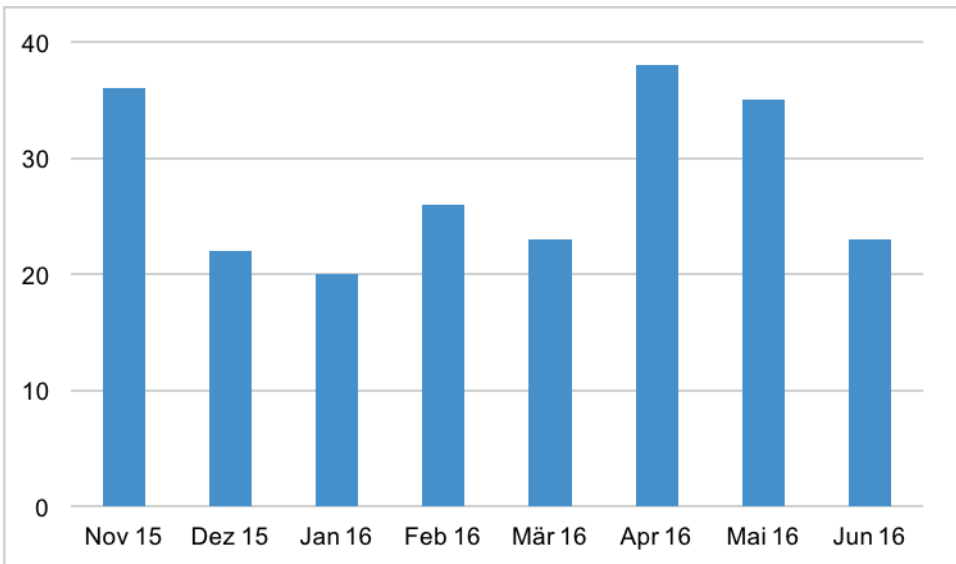


Abb. 6: Anzahl Chats pro Monat

3 Zusammenfassung und Ausblick

Für die Nutzenden von Dienstleistungen ist es heute selbstverständlich, dass sie sich bei Fragen und Problemen über eine Vielzahl verschiedener Kommunikationskanäle wie z.B. Chat, Portale, Apps und soziale Netzwerke an den Dienstleister wenden können. Diese Kommunikationskanäle haben sich bei den „big playern“ z.B. in der Telekommunikationsbranche bereits etabliert. Um das Selbstverständnis als moderner und kundenorientierter IT-Dienstleister zu festigen, verfolgt das IT Center eine offene Kommunikationsstrategie, die neue Kommunikationsmittel unterstützt und bereit ist, weitere Kommunikationskanäle einzubinden. Bei der Einführung neuer Kommunikationskanäle muss beachtet werden, dass diese auf die Zielgruppen zugeschnitten werden und berücksichtigt wird, ob der Kanal wirklich benötigt bzw. von den Zielgruppen akzeptiert wird und ob technische Supportmöglichkeiten integriert werden können.

Nicht zu vernachlässigen ist ebenfalls, dass die technische Integration der Kommunikationskanäle so erfolgt, dass notwendige Kennzahlen in bestehende Reportingsysteme einfließen und ausgewertet werden. Die aus dem Reporting gewonnenen Daten werden benötigt, um einerseits die in 2.3 aufgeführten Kennzahlen auszuwerten, und andererseits, um Änderungen im Kommunikationsverhalten der Anwender zu erkennen und entsprechende Maßnahmen zu ergreifen, so dass die am häufigsten genutzten Kanäle bestmöglich bedient werden.

Die oben beschriebenen neuen Angebote zur Kommunikation sind für das IT Center wahrscheinlich erst der Anfang. Aus Sicht der Anwender laufen zwar alle Kommunikationskanäle zentral im IT-ServiceDesk zusammen und werden dort bearbeitet. Um jedoch die Supportqualität weiter auszubauen, ist es notwendig, die bisher angestrebte Multi-Channel-Lösung in eine Omni-Channel-Lösung zu überführen. Nur so kann eine größtmögliche Kundenzufriedenheit erreicht werden. Die Transformation zu einer Omni-Channel-Lösung ist außerdem erforderlich, um die Mitarbeitenden im IT-ServiceDesk zu entlasten. Omni-Channel-Systeme sind durch den steigenden Grad der Digitalisierung in der Lage, Kundenanfragen intelligent zu steuern und so eine signifikante Verbesserung des Supports zu ermöglichen. Mit solch einer Lösung wird es möglich, weitere qualitätsfördernde Prozesse einzuführen wie z.B. skill-based Support.

Letztendlich muss es das Ziel sein, die verschiedenen Kommunikationskanäle zu synchronisieren und eine benutzerfreundliche Handhabung sowohl für Anwender wie auch für Mitarbeitende im Support bereitzustellen. Dabei muss die Benutzbarkeit im Vordergrund stehen. Nur Systeme, die einfach und verständlich zu bedienen sind, bieten eine hohe Servicequalität und verhindern die Erhöhung der Arbeitsbelastung der Mitarbeitenden durch die neuen Kommunikationskanäle.

Die strategische Vision des IT-ServiceDesk besteht darin, langfristig von einem Multi-Channel-Service-Anbieter zu einem Omni-Channel-Service-Anbieter an der Hochschule zu werden.

Literaturverzeichnis

- [Ar15] Artelt, D.: Multi-, Omni und die „richtig wichtigen“ Channel. TeleTalk 1/15, S. 27, 2015.
- [BHG11] Bischof, C.; Hengstebeck, I.; Grzemski, S.: Einführung eines Service Desk am Rechen- und Kommunikationszentrum der RWTH Aachen University. Ein Praxisbericht In (DINI, Hrsg.): Prozessorientierte Hochschule. Allgemeine Aspekte und Praxisbeispiele. DINI, Göttingen, S. 181-198, 2011.
- [GF15] Grutzeck, M.; Fajga, K.: Synchrone Kommunikationskanäle auf dem Vormarsch. TeleTalk 2/15, S. 36-37, 2015.
- [In16] Informationsverarbeitung am Hochschulen - Organisation, Dienste und Systeme, Stellungnahme der Kommission für IT-Infrastruktur für 2016-2010. Deutsche Forschungsgemeinschaft, 2016.
- [MP16] CrossScience Wissenschaftsblog für Marketing, <http://www.cross-science.de/2014/07/definition-von-multichannel-cross.html>, 10.03.2016
- [Po15] Politze, M. et al.: Facilitating Teacher-Student Communication and Interaction in Large-Scale Lectures With Smartphones and RWTHApp. In: (IATED, Hrsg.): Edulearn 15, Barcelona 2015. IATED Academy, Valencia, S. 4820-4828, 2015.

- [Tm15] Thissen, M.: Ein serviceorientiertes Zusammenspiel ist die Basis. In: Bergmann, P., (Hrsg.): Perspektivenwechsel im IT Service Management, itSMS, Frankfurt am Main, S. 202-209, 2015.

Innovationsmanagement an Hochschulen: Maßnahmen zur Unterstützung der Digitalisierung von Studium und Lehre

Jannica Budde¹, Gudrun Oevel²

Abstract: Die Digitalisierung stellt Hochschulen vor organisatorische, inhaltliche und technische Herausforderungen, insbesondere wenn Hochschulen sich auch als Ort verstehen, an dem die aktive Mitgestaltung der Entwicklung digitaler Medien durch Lehrende und Studierende bewusst gefördert wird. Ein solcher Ansatz benötigt ein explizites Innovationsmanagement als Teil der partizipativen Digitalisierungsstrategie in dessen Rahmen Maßnahmen zur Unterstützung von Innovatoren und zur Sicherstellung von nachhaltiger Verankerung notwendig sind. Die an der Universität Paderborn entwickelten Maßnahmen und ihre theoretische Rahmung werden in diesem Beitrag vorgestellt und diskutiert.

Keywords: Digitalisierung, Innovationsmanagement, Nachhaltigkeit, Transfer, Unterstützungsangebote, Lehre

1 Einleitung

Die Digitalisierung stellt Hochschulen vor organisatorische, inhaltliche und technische Herausforderungen. Insbesondere in den Kerngebieten Studium und Lehre zeigt sich dies deutlich. Der Prozess der Digitalisierung umfasst hier eine Reihe verschiedener Aspekte, wie den Ausbau der Netzinfrastruktur, die Ausstattung von Seminarräumen und die zentrale und flächendeckende Etablierung von Lern- und Campusmanagementsystemen. Diese Maßnahmen sind von zentraler Stelle, gewissermaßen ‚von oben‘, anzustoßen. Eine solche Infrastruktur muss, ebenfalls zentralisiert, auf hohem Niveau gehalten, permanent weiterentwickelt und an den Bedürfnissen von Lehrenden und Studierenden ausgerichtet werden.

Zugleich verstehen sich viele Hochschulen auch als ein Ort, an dem die Entwicklung digitaler Medien ‚von unten‘ aktiv mitgestaltet wird und werden soll. Die meisten solcher Digitalisierungsprojekte, die zum Beispiel technische Werkzeuge, neuartige Angebote für Studierende oder spezifische Blended-Learning-Konzepte (Stichwort E-Learning) entwickeln, entstehen in der Regel aus dem didaktischen Interesse einzelner

¹ Universität Paderborn, Zentrum für Informations- und Medientechnologien, Warburger Str. 100, 33098 Paderborn, jannica.budde@uni-paderborn.de

² Universität Paderborn, Zentrum für Informations- und Medientechnologien, Warburger Str. 100, 33098 Paderborn, gudrun.oevel@uni-paderborn.de

Dozenten und Dozentinnen heraus. Die Entwicklung dieser Konzepte ist demnach hoch intrinsisch motiviert und verläuft selten in einem abgestimmten Prozess zwischen didaktischen, rechtlichen und technischen Anforderungen. In vielen Projekten spielen zudem Fragen der Nachhaltigkeit und des Transfers bei Projektstart eine untergeordnete Rolle. Diese Fragen gewinnen in der Regel erst am Ende eines Projektes an Bedeutung. Insbesondere der Wunsch nach einer zentralen Verankerung konfrontiert Hochschulen, genauer die IKT-Einrichtung einer Hochschule, mit einem erheblichen Arbeitsaufwand, die Verstetigungsprozesse hemmen, wenn nicht gar aufhalten können. Konkrete Probleme ergeben sich bspw. bei der Verlagerung von einer lokalen in eine zentrale Infrastruktur, bei der Verwendung nicht-kompatibler Plattformen/Technik und Programmiersprachen, aber auch besonders in Hinsicht auf Datenschutzaspekte und fehlende technische Dokumentation. Welche Maßnahmen müssen Hochschulen demnach treffen, um solche Übergangs- und Transferphasen möglichst reibungslos zu gestalten und gleichzeitig Innovationen proaktiv zu fördern?

An der Universität Paderborn wurden im Projekt InnoVersity,³ gefördert durch den Stifterverband und die Heinz-Nixdorf-Stiftung, diese Fragestellungen systematisch untersucht und entlang der Theorien von Innovations- und Changemanagement Fördermaßnahmen auf unterschiedlichen Ebenen und für unterschiedliche Zielgruppen entwickelt. Methodisch soll dabei ein Konzept entwickelt werden, das frühzeitig die innovativen Projekte identifiziert und fördert, die sich nachhaltig in die universitätsweite Digitalisierungsstrategie einbetten lassen. Dazu wurde zunächst ein Reifegradmodell für Projekte konzipiert, aus dem sich unterschiedliche Förderungsmaßnahmen ableiten lassen. Das Reifegradmodell sowie die Maßnahmen werden mit ihrer theoretischen Einbettung in diesem Beitrag vorgestellt und diskutiert. Sie lassen sich als konkrete Elemente der Umsetzung einer partizipativen Digitalisierungsstrategie verstehen.

2 Innovationsmanagement an Hochschulen

2.1 Innovationsprozesse an Hochschulen aus Sicht des Innovationsmanagements

Innovationsmanagement beschreibt die systematische Planung, Steuerung und Kontrolle von Innovationen und ist in der Regel gesteuerter Teil der Strategie von Unternehmen und Organisationen. Hochschulen mit ihrer spezifischen Organisationsstruktur wurden bisher wenig in diesem Kontext betrachtet. Dabei lässt sich konstatieren, dass die spezifische Doppelstruktur von Top-Down- und Bottom-Up-Prozessen an Hochschulen einerseits die Entwicklung von Innovationen fördert, sich aber andererseits kontraproduktiv auf die Implementierung innerhalb der gesamten Hochschule auswirkt. Das Maß der Verbreitung bzw. Diffusion von Innovationen ist ‚von unten‘ an das professionelle Interesse einzelner Akteure sowie an die Zusammenarbeit verschiedener Akteure geknüpft [Sc99]. ‚Top-Down‘-Entwicklungen, etwa die hochschulweite

³ Mehr Informationen zum Projekt InnoVersity unter www.uni-paderborn.de/innoversity.

Einführung von einheitlichen Lernmanagementsystemen, produzieren dagegen eher Widerstände und Ablehnung. Schädler identifiziert bereits Ende der 1990er Jahre diese Problemstellung, die auch im Jahre 2016 weiterhin ihre Gültigkeit besitzt: Innovative Lehrkonzepte bzw. Tools für die Lehre werden in einem individuellen Rahmen von Lehrstühlen, Arbeitsgruppen und einzelnen Akteuren entwickelt, jedoch findet die Verbreitung an der eigenen wie an anderen Hochschulen und externen Institutionen (beispielsweise innerhalb der Fachcommunity) kaum statt. Innovationsmanagement an Hochschulen erfordert daher die Steuerung von Diffusionsprozessen ausgehend von der individuellen auf die zentrale Hochschulebene [Sc99].

Im Bereich Innovationsmanagement ergeben sich dadurch eine Reihe von Handlungsfeldern: innovative Ideen und Vorhaben erheben, dokumentieren, bewerten und gegebenenfalls begleiten, unterstützen, Innovationsträger mit anderen vernetzen und gegebenenfalls einen Weg für eine Überführung in den breiteren Einsatz und die Verankerung erarbeiten. Bei der Umsetzung dieses Ansatzes stellt sich zunächst die Frage nach Kriterien zur Bewertung von Projekten und der Identifizierung von Innovationen mit Nachhaltigkeitsoptionen. Da Innovationsmanagement immer auch Aspekte des Changemanagements beinhaltet, werden dieser konzeptionelle Ansatz und seine Anwendbarkeit im nächsten Kapitel untersucht.

2.2 Innovationsmanagement und Changeprozesse

Ein Übergang von einem lokalen innovativen Projekt in eine zentrale Infrastruktur macht einen Changeprozess notwendig. Changemanagement bezieht sich, im Gegensatz zu Projektmanagement, auf einen allgemeinen, gesamtulturellen Änderungsprozess in einer Organisation. Die Integration digitaler Medien ist dabei aber ein Aushandlungsprozess, da ohne zentrale Infrastruktur- und Personalentscheidungen „Innovationen in diesem Feld nur in lokal begrenzten einzelnen Subsystemen stattfinden und [] daher nur eine geringe Bedeutung für die gesamte Hochschule“ haben können [Br15]. Neben der Bereitstellung von Infrastruktur und Beratungsangeboten fordern Bremer et al. daher die Etablierung von Strukturen, die den Bottom-Up wie Top-Down-Strukturen einer Universität gerecht werden. Eine hochschulweite Integration von E-Learning bzw. die Digitalisierung der Hochschule insgesamt kann nur mit einer breiten Einbindung von Akteuren in einem strukturierten Changeprozess gelingen. Nach Lewin [Le58] erfolgt ein solcher Changeprozess im Sinne eines Phasenmodells, das sich folgendermaßen darstellt:

Zustand I => Unfreeze => Move => Refreeze => Zustand II.

Hierbei betrifft die Phase *Unfreeze* den Impuls zur Veränderung und die Formulierung einer klaren Strategie für die Konzeptionierung konkreter Maßnahmen, die Phase *Move* ist die eigentliche Veränderungsphase und die Phase *Refreeze* steht für die nachhaltige Verankerung bzw. Etablierung in einer Organisation. An diesem Punkt ist aus einem Projekt eine Daueraufgabe geworden. Als ein spezifisches Problem dieser Phase nennen Bremer et al. bspw. die Finanzierung [Br15], so dass Daueraufgaben häufig durch

Nachfolgeprojekte realisiert werden (müssen).

Das Phasenmodell nach Lewin wird von uns als Grundlage genutzt, um konkrete Maßnahmen für den Übergang von einem Status zum nächsten zu bestimmen. Für die Identifikation erfolgreicher Förderungsmaßnahmen müssen zudem Erfolgsfaktoren berücksichtigt werden. Kotter (1995) benennt folgende Erfolgsfaktoren für die jeweiligen Phasen [Ko95]:

- Unfreeze: Dringlichkeit, Führungskoalition, Vision und Strategie
- Move: Kommunikation, Befähigung, schnelle Erfolge
- Refreeze: Konsolidierung, Verankerung

Changemanagement in diesem Sinn verweist auf den prozessualen Charakter von Projekten und zeigt mit der Verstetigung ein deutliches Ziel auf. Die Konzeption von Maßnahmen im Rahmen eines Innovationsmanagements sollte daher nach unserer Ansicht insbesondere unter Berücksichtigung der Phasen des Changemanagements geschehen. Dies ist insbesondere für solche Fälle notwendig, wo Projekte in eine zentrale Infrastruktur eingebettet werden können bzw. sollen. Hier sollte die Übergangphase für alle beteiligten Akteure erleichtert werden, indem den Projektgruppen frühzeitig spezifische Rahmenbedingungen auferlegt und diese dafür sensibilisiert werden, Verbreitungs- und Nachhaltigkeitsstrategien als eine Art Roadmap möglichst früh in ihre Projektplanung miteinzubeziehen. Entlang dieser Roadmap sind von Seiten der Hochschule entsprechende Maßnahmen und Fortbildungsangebote zu gestalten. Diese Maßnahmen wurden am Beispiel der Universität Paderborn mit Hilfe eines Statusmodells konzeptioniert und werden im Folgenden erläutert.

3 Entwicklung eines Statusmodells

Ein Raster nach Reifegraden ist für die Entwicklung eines hochschulweiten Innovationsmanagements wünschenswert, „um den Einsatz von E-Learning-Angeboten in der Hochschullehre beurteilen und vergleichen zu können“ [Ni09] sowie um Fördermaßnahmen abzustimmen. Projekte können so effizienter bei der Weiterentwicklung unterstützt werden. Nikolopoulos betont die Bedeutung eines Reifegradmodells auch für den Aspekt der Nachhaltigkeit [Ni09].

Zur Entwicklung eines Reifegradmodelles wurden zunächst die Digitalisierungsprojekte an der Universität Paderborn systematisch erhoben und zunächst der Aktivitätsstatus (Projekt am Anfang, laufendes Projekt, Projekt etabliert, Projekt beendet) eines jeden identifizierten Projektes ermittelt (siehe Abbildung 1). Bei der Analyse des jeweiligen Status eines Digitalisierungsprojektes wurde sehr schnell deutlich, dass die Einordnung bezüglich des Aktivitätsstatus mit Finanzierungs- und Strukturaspekten sowie der Reichweite eines Angebotes korreliert. Laufende und etablierte Angebote bzw. Projekte hatten bereits mehrere Projekt- und Förderungsstufen durchlaufen, hatten in der Regel

eine gute öffentliche Rezeption und wurden deutlich weiterentwickelt. Anlaufende Projekte bzw. Angebote wurden dagegen zunächst in der Regel über hochschulinterne Fördermittel (z. B. Qualitätsverbesserungsmittel (QVM), Forschungsreserven, Preisgelder) finanziert. Fördermaßnahmen, Reichweite und Entwicklungsstufen wurden so in den lokalen Kriterienkatalog für die verschiedenen Projektphasen mit aufgenommen.

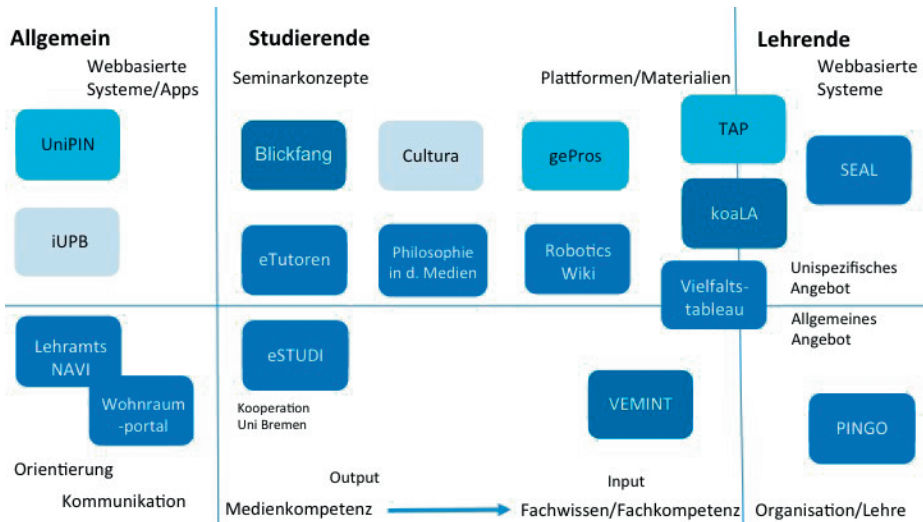


Abb. 1: Reifegradanalyse von ausgewählten Digitalisierungsprojekten

Alexander Nikolopoulos schlägt darüber hinaus für ein Raster von Reifegraden eine Anpassung des Capability Maturity Model Integration (CMMI) vor, das ursprünglich aus dem Bereich der Softwareentwicklung stammt [Ni09]. Das CMMI kennt fünf verschiedene *maturity levels* und sechs *capability levels* [Ch03]. Das CMMI lenkt den Blick einerseits auf eine oben noch nicht integrierte Vor- bzw. Anfangsphase sowie auf eine genauere Differenzierung der Phase „laufendes Projekt“ über die Grade ‚managed‘, ‚defined‘ und ‚quantitatively managed‘, die eine steigende Institutionalisierung von Prozessen abbilden. Der Status von etablierten Projekten lässt sich schließlich dem Grad ‚optimizing‘ zuordnen. Der Fokus liegt hier vor allem auf der kontinuierlichen Weiterentwicklung und Verbesserung institutionalisierter Prozesse.

Auf Basis dieses Ansatzes wurde das lokale Statusmodell ergänzt und insgesamt folgendes Reifegradmodell gebildet:

Initial: Das Projekt befindet sich in der Startphase. Ein Projektplan wurde erstellt und Projektziele wurden formuliert. Eine Finanzierung wurde beantragt bzw. wird

gesucht. Erste technische Ansätze werden entwickelt. In diesem Status werden über die Projekte in der Regel keine Informationen veröffentlicht, so dass dieser auch nicht über die Katalogisierung erschlossen werden konnte.

Anfang: Erste Projektphase, Pilot- bzw. Evaluierungsphase. Es wurde mit der eigentlichen Projektphase (Entwicklung und Evaluation) begonnen. Erste Versionen von Produkten sind verfügbar. Die strukturelle Einordnung ist innerhalb eines Lehrstuhls bzw. einer Arbeitsgruppe zu finden. Die personelle Verbindung (Leitung, Mitarbeitende) ist äußerst stark. Die Projektfinanzierung läuft über den Rahmen einer Anschubfinanzierung über interne Mittel bzw. über Drittmittel. Das Projekt bzw. Produkt ist noch wenig strukturell in die Universität eingebunden, hat wenige Entwicklungsstufen durchlaufen und hat nur eine geringe Reichweite.

Laufend: Nach der ersten Projektphase. Ein Konzept (z. B. Seminarform etc.) wird wiederholt in der Praxis erfolgreich eingesetzt. Ähnliche Projekte können mit Erfolg durchgeführt werden und es gibt einen kontinuierlichen Verbesserungsprozess, auch entlang hochschulweiter Standards. Die strukturelle Einbindung (Leitung, Mitarbeitende) ist weiterhin vorrangig innerhalb eines Lehrstuhls bzw. einer Arbeitsgruppe zu finden, die personelle Verbindung ist demnach weiterhin stark. Die Projektfinanzierung ist über die Anschubfinanzierung hinaus und das Projekt wird z. B. über (eine) weitere Entwicklungsstufe(n) mit externen Geldern und Preisen finanziert. Trotz der weiterhin engen Einbindung ist die Reichweite i. d. R. groß, d. h. es wird sowohl nach intern wie extern wahrgenommen (z. B. Öffnung der Zielgruppe innerhalb und außerhalb der Hochschule, Wahrnehmung durch Medien, in der Fachcommunity etc.).

Etabliert: Strukturelle Einbindung. Ein Konzept bzw. ein Produkt wird außerhalb eines Projektbetriebes angeboten, es gibt jedoch einen kontinuierlichen Verbesserungsprozess. Die personelle Verbindung ist schwach. Etablierte Angebote sind entweder in hochschulexterne Strukturen oder innerhalb der Universität in zentralen Einrichtungen eingebunden. Externe Einbindung umfasst z. B. die Einbindung in eine Fachcommunity über eine externe Gesellschaft bzw. einen Verein oder eine Ausgründung in ein Unternehmen. Interne Zentralisierung meint vor allem die zentrale Übernahme durch die Verwaltung, IKT-Einrichtungen, Medienzentren oder andere zentrale Einrichtungen. Die Reichweite ist entsprechend sehr hoch, sei es nach außen oder innen. In diesem Status kann von einer nachhaltigen Verbreitung gesprochen werden. Hier kann nicht mehr von einer eigentlichen Projektphase gesprochen werden. Die Lehr-/Lerninnovation ist in einem Dauerbetrieb verankert.

Die Phasen werden folgendermaßen farblich referenziert (siehe Abbildung 2):



Abb. 2: Reifegradmodell

Anhand dieses Reifegradmodells werden erste Maßnahmen im Rahmen eines Innovationsmanagements an Hochschulen im Bereich Digitalisierung von Studium und Lehre deutlich. So spielen Fragen von Finanzierung und personaler Einbindung eine wichtige Rolle in den unterschiedlichen Phasen. Zur Weiterentwicklung von Projekten von einem Status in den nächsten sind dies Fragestellungen, bei denen Projektgruppen unterstützt werden müssen, beispielsweise durch Informationsveranstaltungen und Materialien zu verschiedenen Fördermöglichkeiten. Ebenso erscheint es notwendig, Maßnahmen bezüglich der Umsetzung von Transfer- und Verbreitungsstrategien zu entwickeln, die helfen, personalgebundene Strukturen aufzulösen, soweit dies möglich ist. Möglichkeiten wären hier beispielsweise die Bündelung in zentralen Einrichtungen einer Hochschule (Rechen- und Medienzentrum, hochschuldidaktische Zentren etc.) oder die Externalisierung über Ausgründung in ein Unternehmen, Vereinsgründung, Einbindung in die Fachcommunity etc. Für Hochschulen ist gerade der erste Fall interessant. Auch für IKT-Einrichtungen ist gerade dieser Fall von großer Bedeutung, da diese i. d. R. Anlaufstellen für Implementierung in die Gesamtinfrastruktur der Hochschule auf technischer Ebene sind und entsprechend reagieren müssen. Studentische Projekte wurden zusätzlich im Statusmodell markiert, da diese einen Sonderfall darstellen. Hierzu wurden, wie in Kapitel 4 dargestellt, spezielle Maßnahmen konzipiert.

Mit Blick auf die Theorie des Changemanagements wird deutlich, dass auch einzelne Projekte, die die Lehre an bestimmten Lehrstühlen oder Instituten verbessern sollen, einem lokalen Changeprozess entsprechen. Die Phase des Unfreeze steht vor und zu Beginn des Status ‚Initial‘. Hier entsteht der Impuls zur Veränderung. Eine Vision entsteht und eine Strategie mit Projektzielen wird ausgearbeitet. Dies mündet in die Formulierung eines Projektantrages für verschiedene Finanzierungsmöglichkeiten (QVM, Preisgelder, Forschungsreserven, Drittmittel etc.) und eines Projektplanes. Der Status ‚Initial‘ (I) führt demnach in die Changephase ‚Move‘ über, in der auch die Status ‚Anfang‘ (A) und ‚Laufender Betrieb‘ (B) liegen. In diesen beiden Phasen werden Lehr-/Lerninnovationen in der Praxis umgesetzt, evaluiert und kontinuierlich verbessert. Erste Erfolge und Öffentlichkeit entstehen. Nach Bremer et al. beginnt die eigentliche Phase

des Unfreezes nach dieser Phase und führt zur Etablierung einer Innovation. Hier hat eine Innovation den Status ‚Etabliert‘ (E) erreicht, der eigentliche Projektbetrieb ist eingestellt. Aus Sicht des Changemanagements für die Gesamtorganisation Hochschule spielt sich die Refreeze-Phase als Zentralisierung einer Lehrinnovation ab. Diese Phase kann jedoch nicht innerhalb der Projektgruppe, sondern nur auf der Organisationsebene erreicht werden. (Vgl. Abbildung 3); es muss also ein Transfer der Verantwortungsebenen erfolgen.

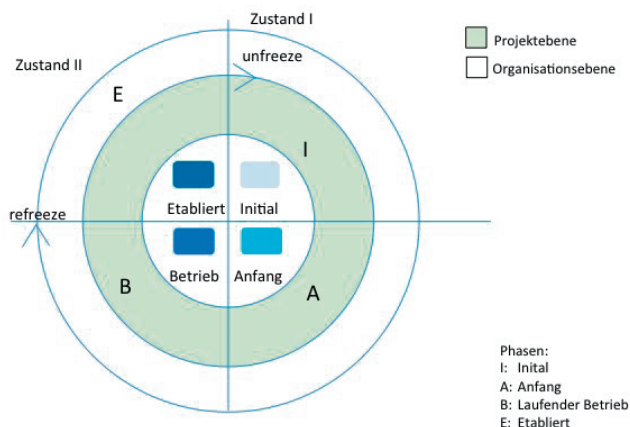


Abb. 3: Zusammenspiel von Projekt- und Organisationsebene beim Changeprozess

Der Übergang zwischen den Phasen ‚Laufender Betrieb‘ und ‚Etabliert‘ eröffnet daher eine Transferphase, die sich als hochschulweiter Changeprozess gestaltet. In dieser Phase wird der Kreis der Akteure durch Mitarbeiter und Mitarbeiterinnen zentraler IKT-Einrichtungen, aber auch beispielsweise der Hochschulverwaltung und verschiedener Stabsstellen, ergänzt. Hierbei entsteht eine heterogene Gruppe mit unterschiedlichen Vorstellungen und Interessen. In diesem Moment können unterschiedliche Interessen in Bezug auf rechtliche und technische Richtlinien, beispielsweise bezüglich des Datenschutzes, problematisch werden. Als Beispiel aus den untersuchten Projekten sei die Verwendung des Datenverkehrsanalysetools Google Analytics genannt, das für die Evaluation von Onlinetools eingesetzt wird, jedoch für den hochschulweiten Einsatz datenschutzrechtlich ein Problem darstellt.

Anhand der Untersuchung von Projektverläufen lassen sich die folgenden wichtigen Querschnittsthemen im Sinne einer erfolgreichen Transformation identifizieren:

- Hochschuldidaktische Fragestellungen und Evaluation
- Finanzierung
- Organisatorische Unterstützung
- Technik und Programmierung

- Datenschutz
- Recht und Lizenzen
- Weiterbildungsmöglichkeiten und Transfer
- Vernetzung

Zu fragen ist nun, wann welche Themen für Projektgruppen oder Akteure relevant sind und dementsprechend sinnvollerweise in Projektverläufe eingebunden werden sollten. Abbildung 4 nähert sich der ersten Frage. Diese Zuordnung gilt es anhand der Maßnahmen genauer zu evaluieren.

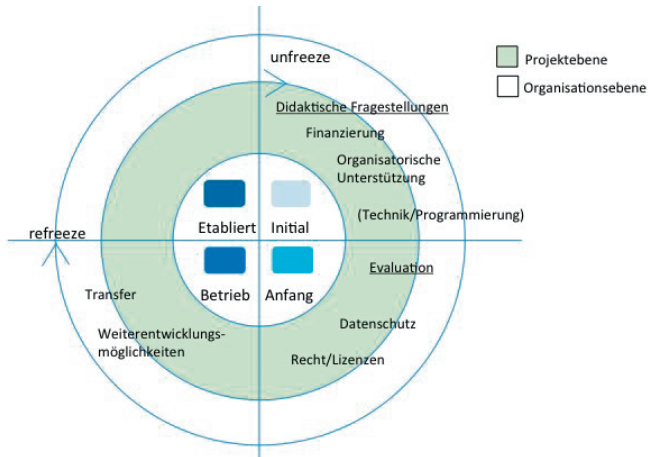


Abb. 4: Zuordnung von Projektphasen und Themen

Neben didaktischen Fragestellungen spielen für die Phase ‚Initial‘ Themen zur konkreten Umsetzung eine große Rolle, insbesondere jene, die sich auf die Finanzierung, Technik und organisatorische Unterstützung beziehen. Gesicherte finanzielle, organisatorische sowie technische Rahmenbedingungen ermöglichen die Umsetzung des in dieser Phase entworfenen Konzepts.

In der Phase ‚Anfang‘ sind nach Kotter schnelle Erfolge ein bedeutender Faktor für ein Projekt. Projektgruppen entwickeln, testen und evaluieren in dieser Phase ihr Konzept bzw. Produkt beispielsweise in Rahmen von einzelnen Seminaren. Erfahrungsgemäß werden in den meisten Fällen erst im Zuge des Übergangs in einen laufenden Betrieb, etwa die breite Umsetzung in einem Lehrstuhl oder Fachbereich, rechtliche Fragen, bspw. bezüglich des Datenschutzes und von Lizenzen, relevant. Diese Aspekte können sich in einer Verstetigungsdiskussion als kritische Faktoren herausstellen und sollten daher frühzeitig in Projekte einbezogen werden.

Die rechtlichen Themenbereiche verweisen damit auch auf den Schritt des Transfers und der Verbreitung des eigenen Konzepts. Dies wird für die Phase ‚Laufender Betrieb‘ sichtbar, wenn sich das Konzept bzw. Tool erfolgreich im eigenen Lehrstuhl oder im Fachbereich durchgesetzt hat. In dieser Phase lässt sich das Potential für eine Verankerung in zentrale Einrichtung bzw. den Transfer ermitteln. Ebenso stellt die Phase ‚Laufender Betrieb‘ die Endphase des Projektbetriebs dar, so dass überlegt werden muss, wie es weitergehen soll. Für eine zentrale Verankerung einer Innovation ist es nun u. a. wichtig, (hochschulweite) Standards zu berücksichtigen und das Projekt umfangreich für potentielle Anwenderinnen und Anwender sowie für den technischen Betrieb zu dokumentieren.

Wie deutlich geworden ist, hängen die Themenbereiche technische Umsetzung, rechtliche und didaktische Umsetzung sowie Transfermöglichkeiten stark von der Vision eines Angebotes bzw. Konzeptes ab und sollten daher in Form einer Roadmap bereits zu Anfang einer Projektarbeit thematisiert werden. Demnach muss für diese Themen frühzeitig ein Bewusstsein für solche Fragestellungen auch bei nicht-technischen Innovatoren zu geschafft werden.

4 Entwickelte Maßnahmen

Zielgruppe von Maßnahmen in einem hochschulweiten Innovationsmanagement sind nicht nur Lehrende, sondern alle Angehörige der Hochschule. Insbesondere Studierende sollen an der Universität Paderborn in die Umsetzung der partizipativen Digitalisierungsstrategie stärker miteinbezogen werden. Die bisher implementierten Maßnahmen sollen im Folgenden in Hinblick auf ihre Wirkungsebenen vorgestellt werden:

4.1 Vernetzung von Akteuren

4.1.1 Strategiekreis E-Learning

Für die Umsetzung einer Digitalisierungsstrategie sind, so zeigt auch die Analyse von Bremer et al., der Aufbau personeller Strukturen und die Einbindung möglichst vieler Akteure notwendig. Mit dem Strategiekreis E-Learning wurde ein Gremium quer zur Liniorganisation geschaffen, das sich aus verschiedenen Akteuren zusammensetzt. Neben innovativ tätigen Lehrenden sind dies u. a. ein Mitglied des Präsidiums, die E-Learning-Beauftragte der Hochschuldidaktik, die Leitung des Medien- und Rechenzentrums, die Leitung der Universitätsbibliothek, der Datenschutzbeauftragte sowie Mitarbeitende im Projekt InnoVersity. Über den Strategiekreis wurden auf der strategischen Ebene Bottom-Up- mit Top-Down-Strukturen miteinander verknüpft. Seine Aufgabe ist die Abstimmung von Maßnahmen zur Umsetzung der Digitalisierungsstrategie in Lehre und Studium.

4.1.2 Netzwerktreffen

Eine wichtige Aufgabe von Innovationsmanagement ist der Austausch von innovativ tätigen Akteuren. Dies wurde auch in vielen Interviews als Wunsch an ein hochschulweites Innovationsmanagement geäußert. Daher wird regelmäßig eine Netzwerkveranstaltung angeboten. Die Veranstaltung möchte zu einem informellen Austausch zum Thema E-Learning und Lehre anregen. Im Rahmen lockerer Gespräche können Experten, Akteure und Interessierte sich über ihre Erfahrungen in den Bereichen E-Learning, Wissensmanagement und Didaktik austauschen. Weitere informelle Angebote wie ein E-Learning-Stammtisch sind geplant. Dieses Angebot richtet sich an Projekte in allen Phasen.

4.2 Studierende

4.2.1 Projektseminar

Für das Sommersemester 2016 wurde ein Projektseminar entwickelt. Dieses Seminar begleitet die Studierenden durch den Anfang der Projektentwicklung. Lehr-/Lernziel ist die Entwicklung bzw. Konzeptionierung eines eigenen Projekts im Bereich Digitalisierung von Studium und Lehre. Der Fokus liegt daher auf der Unterstützung der Initial- und ggf. Anfangsphase von Projektideen der Studierenden.

4.2.2 Innovationspreis für Studierende

Zur Motivation wie auch zur Würdigung des studentischen Engagements in der Weiterentwicklung der Digitalisierung von Studium und Lehre wird im Sommersemester 2016 erstmalig ein Innovationspreis für Studierende ausgeschrieben. Seit 2011 lobt die Universität Paderborn bereits einen Förderpreis für Innovation und Qualität in der Lehre für Lehrende aus. Diese Auszeichnung macht das Engagement der Lehrenden für die Hochschullehre sichtbar. Über die mit dem Preis verbundenen Fördergelder konnten bereits zahlreiche innovative digitale Projekte unterstützt werden. An diesen Erfolg will der Innovationspreis für Studierende anknüpfen.

4.3 Modulare Angebote für Innovatoren zur Projektunterstützung

Kernstück des Maßnahmenkatalogs ist ein modulares Beratungs- und Informationsangebot mit unterschiedlichen, frei wählbaren Bausteinen, das sich primär an Lehrende richtet, die digitale Lehrkonzepte im Bereich Lehre entwickeln (wollen). Als Format wird pro Themengebiet ein zweistündiger Workshop abgehalten, folgende Themen sind in der Umsetzung:

- Workshop Datenschutz: Was muss man bei der Erstellung digitaler Angebote beachten?

- Workshop Programmierung: Was muss man bei der technischen Umsetzung beachten? Welche Plattformen (Sprachen) werden an der Universität unterstützt, welche Schnittstellen und vorhandenen Werkzeuge können nachgenutzt werden? Welche Richtlinien hat das Rechenzentrum?
- Workshop Finanzierung und Unterstützung: Welche Unterstützungsmöglichkeiten gibt es an der Universität für was?
- Workshop Rechtsfragen und Lizenzformen: Urheberrecht und Verwertungsrecht, CC-Lizenzen
- Workshop Transfer: Übertragbarkeit, Nachhaltigkeit, Ausgründung etc.

Daneben sollen Informationen zu diesen Themen im Web bereitgestellt und Ansprechpartner benannt werden.

Für eine optimale Abstimmung der Maßnahmen auf die Bedürfnisse der Projektakteure sollen diese Angebote im Laufe des Jahres evaluiert werden. Insbesondere die Ausrichtung an dem Status- und Phasenmodell steht dabei im Vordergrund. Wichtig ist es, bereits zu einem möglichst frühen Zeitpunkt erfolgreiche Wege und Möglichkeiten in den Status ‚Etabliert‘ aufzuzeigen und für eine langfristige Planung mit dem Gedanken der Nachhaltigkeit zu sensibilisieren.

5 Fazit

Die Digitalisierung der Hochschulen stellt eine große organisatorische Herausforderung dar. Dieser Herausforderung begegnet die Universität Paderborn mit der Etablierung eines Innovationsmanagements als Teil ihrer partizipativen Digitalisierungsstrategie. Der Ansatz eröffnet die Chance, innovativ tätige Akteure an Hochschulen besser zu unterstützen und Innovationen so nachhaltiger in die Infrastruktur einzubinden. Die Nachhaltigkeit der Digitalisierungsprojekte, ihre Verbreitung und Verstetigung, wie auch die Steigerung der Innovationstätigkeit von Hochschulangehörigen im Bereich Digitalisierung stehen bei den hier vorgestellten Maßnahmen im Vordergrund.

In diesem Beitrag wurden daher Kriterien vorgestellt, um Innovationen in der Lehre bzgl. ihres Reifegrades zu identifizieren. Anhand der Klassifikation konnten erste Maßnahmen zur Unterstützung der nachhaltigen Verankerung von Innovationen entwickelt werden. Diese Maßnahmen wurden diskutiert. Sie befinden sich derzeit an der Universität Paderborn in der Umsetzung und Evaluation und lassen sich beliebig auf andere Hochschulen übertragen. Dabei stehen in den hier vorgestellten Maßnahmen zur Umsetzung einer partizipativen Digitalisierungsstrategie die Bedürfnisse der Innovatoren, also der Lehrenden und der Studierenden, im Vordergrund.

Literaturverzeichnis

- [Br15] Bremer, C.; Ebert-Steinhübel, A.; Schlass, B.: Die Verankerung von digitalen Bildungsformaten in deutschen Hochschulen – Ein Großprojekt wie jedes andere? Edition Stifterverband, Essen, 2015.
- [Ch03] Chrissis, M.; B. Konrad, M.; Shrum, S.: CMMI. Guidelines for Process Integration and Product Improvement. Addison-Wesley, Boston, 2003.
- [Ko95] Kotter, J. P.: Leading Change: Why Transformation Efforts Fail. Harvard Business Review OnPoint Ausgabe March-April, S. 1-10, 1995.
- [Le85] Lewin, K.: Group Decision and Social Change. In (Maccoby, E.; Newcomb, T. M.; Hartley, E. L.): Readings in social psychology. Holt, Rinehart and Winston, New York, S. 197-211, 1958.
- [Ni09] Nikolopoulos, A. S.: Die Sicherung der Nachhaltigkeit von E-Learning-Angeboten in Hochschulen, Frankfurt am Main, 2009.
- [Sc99] Schädler, U.: Das Innovationspotential der Hochschulen. Chancen und Risiken der Umsetzung von Innovationen in der Lehre an deutschen Universitäten. Lang, Frankfurt am Main, 1999.

Intelligente Bildungs- und Hochschulnetze in Deutschland Wo stehen wir in 2016? Was ist zu tun bis 2020?

Christoph Igel¹, Ansgar Baums², Marc Göcks³, Martin Haag⁴, Bernd Krämer⁵, Stephan Pfisterer⁶, Siegfried Stiehl⁷, Guido Wirtz⁸, Volker Zimmermann⁹ & Jens Mühlner¹⁰

Abstract: Ausgehend von der Betrachtung des Einflusses der zunehmenden Digitalisierung auf zentrale Infrastrukturektoren wird am Beispiel von Hochschulen in Deutschland die Situation von Studium, Lehre und Weiterbildung hinsichtlich intelligenter Vernetzung dargestellt. Empfehlungen für eine von Bund und Ländern gemeinsam zu verantwortende und zeitnah zu realisierende Agenda zum Auf- und Ausbau intelligenter Bildungs- und Hochschulnetze in Deutschland werden formuliert und mit konkreten Handlungserfordernissen untersetzt.

Keywords: Nationaler IT-Gipfel, Intelligente Vernetzung, Bildungsnetze, Hochschulnetze, Status- und Fortschrittsbericht, Kooperationsverbot, Kompetenzzentren, Datenschutz, Stakeholder Peer Review, Kapazitätsverordnung, Lehrverpflichtungsverordnung, Verrechnungsmodelle

1 Einleitung

Die umfassende Digitalisierung und intelligente Vernetzung der zentralen Infrastrukturektoren Energie, Gesundheit, Verkehr, Bildung und Verwaltung ist für die Zukunft und Leistungsfähigkeit Deutschlands von elementarer Bedeutung. Nach Angaben des Bundesministeriums für Wirtschaft und Energie tragen beide Faktoren – Digitalisierung und intelligente Vernetzung - wesentlich dazu bei, zukünftig bedeutende wirtschaftspolitische und gesellschaftliche Herausforderungen wie die Energiewende,

¹ Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Educational Technology Lab, Alt-Moabit 91 C, 10559 Berlin, christoph.igel@dfki.de

² Hewlett-Packard Inc., Government Relations EMEA, Lützowplatz 15, 10115 Berlin, ansgar.baums@hp.com

³ Multimedia Kontor Hamburg GmbH, Saarlandstraße 30, 22303 Hamburg, m.goecks@mmkh.de

⁴ Hochschule Heilbronn, Fakultät für Informatik, Max-Planck-Straße 39, 74081 Heilbronn, martin.haag@hs-heilbronn.de

⁵ FernUniversität in Hagen, Fakultät für Mathematik und Informatik, Universitätsstraße 27, 58097 Hagen, bernd.kraemer@fernuni-hagen.de

⁶ Bundesministerium für Bildung und Forschung, Digitales Lernen und Medienbildung, Kapelle-Ufer 1, 10117 Berlin, stephan.pfisterer@bmbf.bund.de

⁷ Universität Hamburg, Fachbereich Informatik, Vogt-Kölln-Straße 30, 22527 Hamburg, stiehl@informatik.uni-hamburg.de

⁸ Universität Bamberg, Fakultät Wirtschaftsinformatik und Angewandte Informatik, An der Weberei 5, 96047 Bamberg, guido.wirtz@uni-bamberg.de

⁹ Neocosmo GmbH, Geschäftsführung, Starterzentrum, 66123 Saarbrücken, volker@neocosmo.de

¹⁰ T-Systems International GmbH, Innovation & Technology Management, Utbremer Straße 90, 28217 Bremen, jens.muehlner@t-systems.com

den demographischen Wandel, das lebenslange Lernen, die steigenden Verkehrsmengen, die Urbanisierung oder auch den Bürokratieabbau zu meistern [BM15].

Der Begriff „Intelligente Vernetzung“ steht dabei für eine Weiterentwicklung und optimierte Nutzung von Informations- und Kommunikationstechnologien (IKT) in eben diesen zentralen Infrastruktursektoren, die Kernbereiche von Gesellschaft und Volkswirtschaft Deutschlands – heute und in Zukunft – sind. Durch eine nahtlose, sichere, zuverlässige und schnelle Vernetzung von Geräten, Daten, Strukturen, Prozessen und Diensten unterschiedlicher Komplexität und Güte werden die dort eingesetzten IKT-Infrastrukturen zu „Intelligenten Netzen“ mit gesteigertem Mehrwert. Dieser Effekt kann durch eine sektorübergreifende Vernetzung verstärkt werden [BMW15]. Studien zufolge schaffen Intelligente Netze in Deutschland bis zum Jahr 2022 neue Wachstumsimpulse und Effizienzgewinne von bis zu 350 Mrd. Euro [B112].

Anlässlich des Nationalen IT-Gipfels der Bundesregierung wurden in den vergangenen Jahren von Expertengruppen aus Wirtschaft, Wissenschaft und Politik erstmals Zielbilder für eine „Intelligente Vernetzung in Deutschland im Jahr 2020“ für die Infrastruktursektoren entwickelt und Empfehlungen zu deren Erreichung erarbeitet. Die Zielbilder beschreiben den für 2020 angestrebten Zustand. Die Gliederung in fünf strategische Ebenen (gesellschaftlich, ökonomisch, rechtlich-regulatorisch, prozessual, technisch) erlaubt eine differenzierte Betrachtung.

Im Rahmen eines Stakeholder Peer Reviews wurden durch die Expertengruppen gemeinsam Bewertungen zum Status und Fortschritt auf dem Weg zur Erreichung dieser Zielbilder vorgelegt. Kernfragen wurden im Rahmen von Konsultationsgesprächen vertiefend kritisch beleuchtet. Zielbilder, Status, Fortschritt und Erreichungsgrad werden jährlich fortgeschrieben und dienen der Identifikation von Hemmnissen und Barrieren, der Initiierung von Verordnungen sowie Gesetzesinitiativen und der Schaffung von Awareness sowie der Förderung des politischen Dialogs auf und zwischen allen Ebenen des föderalen Systems in Deutschland.

2 Intelligente Bildungsnetze: Sachstand 2016

Metaanalysen zufolge besteht ein wachsendes Interesse an E-Learning-Anwendungen in Hochschulen. In der Schul- und Studienverwaltung ist der Einsatz von IuK vielerorts gelebte Praxis. Zudem ist alleine wegen Quantitätseffekten (vgl. Studierendenquote) und Diversitätseffekten (vgl. Inhomogenität von Lernenden) ein hoher Handlungsdruck zu konstatieren. Experten schätzen, dass etwa durch eine bundesweite E-Learning-Plattform und eine einheitliche IuK-unterstützte Verwaltungsmodernisierung in Bildungsinstitutionen jährlich rund 3 Mrd. Euro eingespart werden könnten und - bei einem 10%igen E-Learning-Anteil in Bildung und Weiterbildung p.a. - ein jährlich volkswirtschaftlicher Wachstumsbetrag von rund 2 Mrd. Euro möglich ist [B112].

Vor diesem Hintergrund haben die Mitglieder der Projektgruppe „Intelligente Bildungsnetze“ des Nationalen IT-Gipfels der Bundesregierung die Situation von Studium, Lehre und Weiterbildung an Hochschulen in Deutschland hinsichtlich Digitalisierung und intelligenter Vernetzung analysiert. Mit Blick auf Bestand, Einsatz, Nutzung und zukünftige Entwicklung erweist sich die Digitalisierung und IuK-gestützte Vernetzung von Studium, Lehre und Weiterbildung als sehr viel weniger fortgeschritten, als es technologisch möglich und gesellschaftlich wünschenswert wäre.

Die Konsequenz ist ernüchternd: Im internationalen Wettbewerb um die besten Lösungen für intelligente Bildungsnetze spielt Deutschland heute nur eine Nebenrolle und läuft in naher Zukunft Gefahr, ausschließlich Kunde international agierender Anbieter von Bildungstechnologien zu werden, nicht aber ein den internationalen - oder auch nur den europäischen - Hochschulmarkt prägender oder gar leitender Anbieter. Konkret bedeutet diese Gefahr, dass international agierende Anbieter mit ihren Plattform-, Content- und Service-Lösungen (vgl. Udacity, edX u.a.) auch den lokalen Markt in einer Weise dominieren könnten, als dass nachhaltige Abhängigkeiten entstehen.

Intelligente Bildungsnetze bieten Hochschulen vielfältige Möglichkeiten, sich den Anforderungen der vernetzten, sich rasch verändernden Welt zu stellen: Sie können gemeinsam neue digitale Studiengänge anbieten, für die jede einzelne Hochschule nicht genügend Kapazitäten oder nicht alle notwendigen Kompetenzen vorweisen kann. Schon frühe Erfahrungen der Open University UK zeigen den hohen Aufwand bei partieller ITK-Unterstützung von Studienmodellen [La99]. Durch Austausch, gemeinsame Nutzung, Pflege und Erneuerung digitaler Lehrinhalte und die Lehre unterstützende Applikationen können Kosten und Personalaufwände erheblich gesenkt und die Qualität der Lehre zugleich erhöht werden. Durch Kooperation können Hochschulen ihr Profil schärfen und ihre Stärken ausbauen, um wettbewerbsfähig zu bleiben.

Als aktuelle Treiber für intelligente Bildungsnetze werden arbeitsplatzintegrierte, multimodale Qualifizierungsszenarien ebenso identifiziert wie methodische Entwicklungen rund um Educational Data Mining und Learning Analytics, aber auch nutzerzentrierte, mitunter kleinteilige Bildungsangebote mit starken informellen Wissensanteilen und neuartigen Zertifizierungsmodellen mit Bedeutung für das berufs begleitende sowie lebenslange Lernen („Nanodegrees“).

Kurzfristige Erfordernisse werden in der bedarfsgerechten Konkretisierung von Datenschutzregeln sowie in deren Vereinheitlichung gesehen, insbesondere mit Blick auf internationale Datenschutzabkommen (auch vor dem Hintergrund von Erfordernissen der Datensouveränität). Weiterhin bedürfen Open Educational Resources sowie Open Access in Deutschland konkreter(er) Vorgaben hinsichtlich Governance und Förderbedingungen (etwa hinsichtlich des Urheberrechts mit dem Ziel der Handlungssicherheit für Lehrende). Initiativen wie das Hochschulforum Digitalisierung des Stifterverbandes für die Deutsche Wissenschaft, CHE u.a. oder auch der Antrag der Bundestagsfraktionen von CDU/CSU und SPD zur Digitalen Bildung sind zu

unterstützen und weisen in die richtige Richtung. Dies gilt auch für die Gründung zahlreicher Start-Ups zum Thema „Digitalisierung der Bildung in Deutschland“ in den letzten Jahren.

3 Intelligente Bildungsnetze: Perspektive 2020

Wenn intelligente Bildungsnetze jedoch nachhaltig, systematisch und strukturiert auf- und ausgebaut werden sollen, bedarf es nach Meinung der Mitglieder der Projektgruppe „Intelligente Bildungsnetze“ des Nationalen IT-Gipfels der Bundesregierung zusätzlicher, gemeinsamer Kraftanstrengungen von Bund und Ländern für die in den europäischen Wissenschafts- und Wirtschaftsraum eingebetteten Hochschulen in Deutschland. Dabei muss eine Agenda „Intelligente Bildungsnetze“ im Fokus stehen: zeitnah zu realisierende, konzertierte Aktionen aller Akteure mit definierter Roadmap zu nachstehenden, auf politischer Ebene dringlich anzugehenden Handlungserfordernissen:

- Gemeinsame Finanzierung intelligenter Hochschul- und Bildungsnetze durch Bund und Länder.
- Änderung bzw. wirksame Umsetzung des Artikels 91b GG (Kooperationsverbot) zu Lehre, Studium und Weiterbildung.
- Aufbau von pilotierenden Hochschulnetzwerken in nationalen und grenzüberschreitenden Regionen durch wettbewerbliche Verfahren.
- Sicherung des erfolgreichen Betriebs von Bildungsnetzen vor Ort an den Hochschulen durch entsprechende finanzielle und personelle Ressourcen.
- Etablierung eines Netzwerkes verteilter Kompetenzzentren in Deutschland (mit länderübergreifender Wirkung durch Ausroll-Strategie).
- Incentives für Treiber und Multiplikatoren von Konzeptionen, Lösungen und Services für Intelligente Bildungsnetze an Hochschulen.
- Stärkung und Steigerung einschlägiger Kompetenzstrukturen in den Bildungseinrichtungen als Voraussetzung und Bedingung für die Entfaltung von Nutz- und Mehrwerte von IuK-Infra- und Dienstestrukturen.
- Anpassung von Kapazitäts-/Lehrverpflichtungsverordnungen zur angemessenen Berücksichtigung der Aufwände von multimedialer und kooperativer Lehre.
- Konkretisierung von Datenschutzregeln auch unter Beachtung europäischer Regulierungen.
- Verrechnungsmodelle für Content- und Service-Nutzung über Landesgrenzen hinweg.

- Unterstützung technologischer Standards für Austausch und Vernetzung digitaler Bildungsinhalte.
- Förderung und Stärkung der Zusammenarbeit von Wirtschaft und Wissenschaft.
- Alle vorab genannten Aspekte sind nur durch eine deutlich verbesserte Zusammenarbeit der einzelnen Bundesländer untereinander und mit dem Bund realistisch erreichbar.

Literaturverzeichnis

- [BM15] Bundesministerium für Wirtschaft und Energie: Strategie Intelligenter Vernetzung. Berlin, 2015.
- [BI12] Bitkom & Fraunhofer ISI: Gesamtwirtschaftliche Potenziale intelligenter Netze in Deutschland. Berlin, 2012.
- [DI15a] „Deutschland Intelligent Vernetzen“ Report. www.div-report.de. 23.03.2015.
- [DI15b] „Deutschland Intelligent Vernetzen“ Konferenz. www.div-konferenz.de. 23.03.2015.
- [La99] Diana Laurillard: Educational Imperatives of the Information Age. Invited presentation, EADTU Conference. Milton Keynes. November 1999.

Gelebte Kooperationsformen zwischen Hochschulen – Möglichkeiten der Ausgestaltung kritischer Erfolgsfaktoren anhand von Praxisbeispielen

Kai Dornseiff¹, Dr. Yvonne Groening², Dr. Stephan Kassanke³

Abstract: Ausgehend von der Motivation für verstärkt auftretende Kooperationen zwischen Hochschulen stellt der vorliegende Beitrag basierend auf dem Stand der aktuellen Literatur zunächst Kooperationsformen und deren klassifizierende Merkmale vor. Am Beispiel von Kooperationen zwischen Hochschulen werden auf diese spezifische Konstellation adaptiert konkrete Erfolgsfaktoren für Kooperationen erläutert. Möglichkeiten zur Ausgestaltung dieser Faktoren werden nachfolgend anhand praktischer Erfahrungen der Autoren belegt. Die beiden Erfahrungsberichte über Kooperationen zeigen auf, wie gezielte Maßnahmen zur Ausgestaltung und Stärkung der Faktoren auch in heterogen geprägten Umfeldern dahingehend gestaltet werden können, dass ein Kooperationserfolg erzielt werden kann.

Keywords: Kooperationen, Kooperationsformen, Kooperationsprojekte, Hochschulen, IT-gestützte Verfahren, Erfolgsfaktoren, Zielidentität, Homogenität, Vertrauen, Kommunikation, Projektmanagement, Nachhaltigkeit

1 Gründe für Kooperationen zwischen Hochschulen

Universitäten und Hochschulen in Deutschland werden sich heute und voraussichtlich auch in Zukunft zwei grundlegenden Herausforderungen stellen müssen: (1) Zum einen steigt der Wettbewerbs- und Leistungsdruck unter den Hochschulen. (2) Zum anderen wird diese Entwicklung zudem durch leistungsbezogene Förderinitiativen (Exzellenzinitiativen) sowie nationale und internationale Rankings befeuert. Vor diesem Hintergrund forcieren Hochschulen ihre Profilbildung in Forschung und Lehre, um so die besten Wissenschaftler und Studierenden an sich zu binden sowie um langfristig dem „Matthäus-Effekt“⁴ folgend, das beste Forschungs- und Bildungssystem zu erreichen. Überdies muss die Finanzierung universitärer Aktivitäten unter den Auswirkungen eines steigenden Budgetdrucks der öffentlichen Hand, z. B. durch die Schuldenbremse 2020 bewältigt werden. Die in der Folge bestehende Forderung der Politik nach mehr Kooperation unter den Hochschulen zur Vermeidung redundanter Strukturen und

¹ myconsult GmbH, Geschäftsführer, Kilianstraße 20, 33098 Paderborn, dornseiff@myconsult.de

² myconsult GmbH, Geschäftsführerin, Kilianstraße 20, 33098 Paderborn, groening@myconsult.de

³ myconsult GmbH, Geschäftsführer, Kilianstraße 20, 33098 Paderborn, kassanke@myconsult.de

⁴ Der „Matthäus Effekt“ geht auf einen Bibelvers (Matthäus 25,29) zurück, welcher auf das hier skizzierte Szenario übertragen bedeutet, dass erfolgreiche Forschungsinstitutionen erfolgreiche Forscher ausbilden oder gezielt anwerben und diese wiederum erfolgreichen Nachwuchs anziehen.

Ressourcen sowie einer Steigerung der Effizienz trifft die Hochschulen bereits heute und wird auch in Zukunft präsent bleiben.

Folglich existieren vermehrt Überlegungen und Empfehlungen (vgl. [DFG10], [HRK13]), die Ressourcen von Hochschulen durch verstärkte Zusammenarbeit oder sogar Fusionen zu bündeln. So mahnte der Landesrechnungshof NRW bereits in seinem Jahresbericht 2013 eine verstärkte Zusammenarbeit der Hochschulen bei der Einführung des kaufmännischen Rechnungswesen sowie dafür benötigter Software an⁵ (vgl. [LrNW13]). Im Bereich der Lehre rät der Wissenschaftsrat in seinen Empfehlungen zur Weiterentwicklung der MINT-Bereiche an den Hochschulen des Landes Hamburg u. a. dazu, eine bessere Vernetzung der Hochschulen herbeizuführen, um eine weitere Optimierung der MINT-Bereiche zu erreichen (vgl. [WR16]).

Kooperationen im Hochschulverbund besitzen das Potenzial durch eine intelligente und nachhaltige Zusammenarbeit Arbeitslasten und notwendige Investitionen auf mehrere Hochschulen zu verteilen. Ferner kann es gelingen, gemeinsam mehr zu erreichen als eine einzelne Hochschule alleine imstande wäre. Zudem können bestehende Infrastrukturen besser ausgelastet und damit Skaleneffekte erzielt werden. Seien es die hier skizzierten Treiber rund um Kosten- und Effizienzdruck oder die aktuelle Diskussion rund um „Hochschule 4.0“ – kooperative Formen der Zusammenarbeit erscheinen als ein derzeit absolut probates Mittel. Es lässt sich prognostizieren, dass hochschulübergreifende Kooperationen auch in Zukunft ein wichtiges Fundament eines modernen Bildungssystems sein werden (vgl. [HR16]).

2 Kooperationsformen zwischen Hochschulen

Im Folgenden wird der Begriff „Kooperation“ terminologisch eingeordnet und der behandelte Betrachtungsgegenstand inhaltlich abgegrenzt. Unter einer kooperativen Zusammenarbeit oder kurz „Kooperation“ wird in diesem Beitrag eine freiwillige, oft vertraglich geregelte, zweckgerichtete Zusammenarbeit zweier oder mehrerer Hochschulen verstanden, die ein temporäres System bilden, das durch Intensität und Dauer der Zusammenarbeit sowie durch eine gemeinsame Zielrichtung geprägt ist. Übergreifendes Ziel ist eine wirtschaftliche Effizienz, eine bessere Leistungsfähigkeit und eine höhere Zielerfüllung⁶ (Vgl. [NDH97], [Sc79], [KL03]).

Während die „gemeinsame Zielrichtung“ eine inhaltliche Dimension der Kooperation beschreibt, prägen Intensität und Dauer die Gestalt der Zusammenarbeit. Diese beiden Dimensionen erlauben es, zwischen mehreren Stufen einer kooperativen Zusammenarbeit zu differenzieren. Das so definierte Feld möglicher Kooperationsformen reicht von punktueller Kooperation (geringe Bindung und Dauer), über konstanten Leistungstransfer

⁵ Auch der Landesrechnungshof Mecklenburg-Vorpommern kommt in seinem Sonderbericht über die Prüfung der Hochschulfinanzierung zu dem Zwischenfazit, dass Potenziale, welche in der übergreifenden Zusammenarbeit der Hochschulen gesehen werden, nicht genutzt werden (vgl. LrMV14).

⁶ Vgl. hierzu z.B. die Definition von Nieschlag/Dichtl/Hörsching 1997.

(mittlere Bindung und Dauer) bis hin zu einer Fusion (dauerhafte Bindung und langfristige Dauer). Abbildung 1 skizziert das Zusammenspiel der Merkmale „Dauer der Zusammenarbeit“ und „Intensität der Bindung der Beteiligten“ nach Hechler/Pasternack (vgl. [HePa13]).

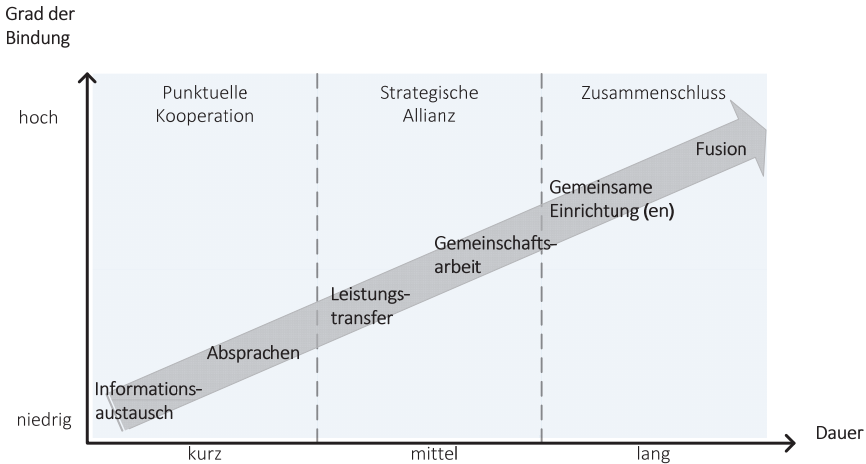


Abbildung 1: Stufen der Zusammenarbeit von Hochschulen⁷

Ein Arbeitspapier des Centrums für Hochschulentwicklung (CHE) trifft eine dreiteilige Differenzierung von Kooperationsformen in Hochschulen (vgl. [HEB07a]):

- **Kooperationen in Netzwerken** dienen als lose Verbünde eher zum Austausch von Informationen oder einem gemeinsamen Außenauftritt und beinhalten seltener die Übertragung von Kompetenzen.
- **Projektbezogene Kooperationen** beschränken sich gemäß des Projektbegriffs auf ein einmaliges Vorhaben mit klaren Zielvorgaben, definiertem Start- und Endzeitpunkt, begrenzten Ressourcen und projektspezifischer Organisation.
- **Institutionalisierte Kooperationen** definieren die Zusammenarbeit zwischen zwei oder mehr Hochschulen über festgelegte Strukturen, Abläufe und Berichte, welche häufig in einem Rahmenvertrag festgelegt sind und bei Bedarf über Projektverträge ergänzt werden.

Die Studie [HEB07b] konstatiert: Je umfangreicher und länger eine Zusammenarbeit angelegt ist, desto eher benötigt sie einen festen Rahmen sowie eine Bündelung der Ressourcen, und desto geringer fällt die Eigenständigkeit aus.

⁷ Quelle: In Anlehnung an vgl. [HePa13].

Beide zitierten Ansätze beschreiben und klassifizieren die Art der Zusammenarbeit zwischen Hochschulen und deren mögliche Differenzierungen. Unbeantwortet bleibt hingegen die Frage, aufgrund welcher Merkmale eines Vorhabens welche spezifische Form der Kooperation gewählt werden sollte, um diese erfolgreich zu gestalten.

Dieser Beitrag fokussiert auf horizontale Kooperationen⁸, die das gemeinsame Ziel der Etablierung eines (oftmals IT-gestützten) Verfahrens anstreben. Das Besondere an horizontalen Kooperationen zwischen Hochschulen ist, dass sie häufig in projektorientierter Zusammenarbeit beginnen und später in langfristig angelegten gemeinsamen Strukturen münden, welche die Ergebnisse der Aufbauarbeit des Projekts übernehmen und die Kooperation langfristig im Sinne aller Partner fortführen. Mittels der Ausbildung gemeinsamer und nachhaltig verankerter Strukturen können so dauerhaft Effizienzgewinne durch das Ausschöpfen von Skaleneffekten realisiert werden.

Eine wesentliche Grundvoraussetzung für eine langfristig angelegte partnerschaftliche Zusammenarbeit ist Vertrauen zwischen den Partnern. Doch über das Vertrauen hinaus lassen sich weitere Erfolgsfaktoren identifizieren, welche eine erfolgreiche Kooperation auszeichnen. Diese werden im folgenden Abschnitt erläutert, bevor anhand von Referenzen der praktische Umgang mit diesen Erfolgsfaktoren in erlebten Projekten wiedergegeben wird.

3 Erfolgsfaktoren von Kooperationen

Notwendige Voraussetzung für das Gelingen einer Kooperation ist die grundlegende Bereitschaft, aller Kooperationspartner zusammen arbeiten zu wollen (vgl. [CHE07b]). Diese Grundvoraussetzung ist die Basis für eine erfolgreiche, zielorientierte Zusammenarbeit. Fehlt diese basale Bereitschaft, sind alle weiteren Erfolgsfaktoren wirkungslos.

Die Bereitschaft allein führt allerdings nicht zu einem funktionierenden und später erfolgreichen Kooperationsmodell. Vielmehr müssen unterschiedliche (Erfolgs-) Faktoren vorhanden sein oder aufgebaut werden, um ein wirkungsvolles Miteinander entstehen zu lassen und auf Dauer zu erhalten. In der Literatur wird eine Vielzahl an Erfolgsfaktoren behandelt; der Fokus in diesem Beitrag liegt auf den Erfolgsfaktoren Zielidentität, strukturelle und kulturelle Ähnlichkeit, Vertrauen und wechselseitige Nachsicht, zielgerichtete Kommunikation und institutionelle Verankerung der Kooperation. Diese repräsentieren die in der Literatur häufig ermittelten Erfolgsfaktoren für Kooperationen (vgl. [Ro00]) und stellen nach Einschätzung der Autoren Schlüsselfaktoren für erfolgreiche Kooperationen zwischen Hochschulen dar.

⁸ Im Falle horizontaler Kooperationen erfolgt eine Zusammenarbeit zwischen Organisationen auf gleicher Wertschöpfungsstufe. Vertikal kooperierende Organisationen gehören unterschiedlichen Wertschöpfungsstufen an; sie agieren in einer klassischen Zulieferer-Abnehmer-Beziehung (vgl. [BDHL11]).

3.1 Zielidentität

Bereits vor der Bildung einer Kooperation sollten zwischen den potenziellen Partnern die Ziele als Erfolgsmaßstab möglichst klar und präzise formuliert werden („SMARTe Ziele“)⁹. Da sich der Kooperationserfolg nicht nur an quantitativ-ökonomischen Ergebnissen misst, sollten neben den messbaren Ergebniszielen auch qualitative Vorgehensziele definiert werden, welche sich eher an sozialen Kriterien wie zum Beispiel Zufriedenheit (Zusammenhalt, Fluktuation, Fehlzeiten) messen lassen. Denn der Erfolg wird vor allem durch die individuelle Wahrnehmung der Beteiligten bestimmt. Besonders bei beteiligungsstarken Kooperationen sind Zielgrößen über zu berücksichtigende Stakeholder erfolgskritisch.

Im Aufbau und nachfolgenden Betrieb einer Kooperation sind die Ziele regelmäßig an den aktuellen Entwicklungen des Umfelds und den Bedürfnissen der Kooperation zu spiegeln. Damit die Zielvorgaben ihren rahmengebenden Charakter nicht verlieren, ist es aus Sicht der Autoren notwendig, die Ziele für alle Partner zu jeder Zeit zugänglich zu machen und für eine hohe Transparenz und Sichtbarkeit zu sorgen. Häufig verlieren Ziele im Laufe eines Kooperationsvorhabens ihre Bedeutung und in der konkreten Ausgestaltung an Kontur; sie werden selten ernsthaft als Kontrollfunktion genutzt.

3.2 Strukturelle und kulturelle Ähnlichkeit

Hochschulen sind aufgrund ihres Grundauftrags in Forschung und Lehre per se inhärent sowohl strukturell als auch kulturell ähnlich, was sich z.B. in gemeinsamer Terminologie, Grundverständnis und Zielen äußert. Trotz ihrer grundlegenden „Nähe“ existieren Unterschiede von Aufbau- und Ablauforganisation, über diversen Hochschultypen und verschiedene Größen bis hin zu verschiedenen Führungs- und Meinungsbildungskulturen.

Eine von vornherein gegebene strukturelle und kulturelle Ähnlichkeit hilft, den Bildungsprozess einer Kooperation hin zu einem arbeitsfähigen Organisationsgebilde stark zu beschleunigen. Zudem wirkt die kulturelle Ähnlichkeit als natürlicher „Kitt“ zwischen den Kooperationspartnern. Je nach Lebensdauer und Ziel einer Kooperation kann dieses Erfolgsmerkmal aber auch kritisch bewertet werden. Beispielsweise kann die strukturelle und kooperative Ähnlichkeit zu einem Interessenkonflikt führen, wenn die Kooperation als Gemeinschaft eigentlich viele Kooperationspartner gewinnen möchte, sich in ihrem Charakter aber gegenüber Ausstehenden als „exklusiver Club“ präsentiert.

3.3 Vertrauen und wechselseitige Nachsicht

Auch wenn eine Kooperation über ein gut ausgestaltetes gemeinsames Zielgerüst verfügt, so werden im Verlauf der Zusammenarbeit immer wieder Situationen entstehen, in

⁹ Ziele sollten smart formuliert werden. SMART steht dabei für S – spezifisch; M – messbar; A – akzeptiert; R – realistisch, T – terminiert.

welchen Partikularinteressen Einzelner potenziell dominieren, bzw. die Vorherrschaft erreichen möchten. Im Sinne einer Kooperation ist es zielführend, dass ein Kompromiss geschlossen werden kann, der keine signifikante Benachteiligung Einzelner zu Gunsten von Partikularinteressen bedeutet. In solchen Situationen hilft es, wenn sich die Kooperationspartner gegenseitig ein großes Vertrauen entgegenbringen und das Zusammenwirken eine wechselseitige Nachsicht kennt.

Der Vertrauensbildungsprozess sollte bestenfalls bereits vor der Bildung der Kooperation und dem Arbeiten am gemeinsamen Ziel begonnen werden. In der gelebten Praxis sind bereits einige Hochschulen landesweit oder bundesweit über Interessenverbände oder Koordinierungsstellen mehr oder weniger intensiv vernetzt. Diese manchmal bereits seit Jahren existenten Austauschplattformen sind eine erfolgversprechende Ausgangsbasis für Kooperationsvorhaben.

3.4 Zielgerichtete Kommunikation

Der kontinuierliche, zielgerichtete und nutzenstiftende Austausch der Kooperationspartner untereinander ist im Sinne einer gewinnbringenden Zusammenarbeit elementar. Der Bedarf an Kommunikation (Art, Umfang, Häufigkeit) ist zwischen den einzelnen Partnern und ausgerichtet an der Art der Zusammenarbeit sowie an dem Ziel der Kooperation auszugestalten und formal festzuhalten. Über die „interne“ Kommunikation einer Kooperation hinaus, ist auch die Information und Kommunikation an interessierte Gruppen zu regeln. Auch die Information von und die Kommunikation mit diesen Interessengruppen entscheidet über den Erfolg einer Kooperation.

Im Zusammenhang mit den bereits beschriebenen Erfolgsfaktoren strukturelle/kulturelle Ähnlichkeit sowie Vertrauen und gegenseitige Nachsicht ist auch die zielgerichtete und vertrauensvolle Kommunikation ein wichtiger Erfolgsfaktor. Wer sich kennt und wem man zunächst vertraut, mit dem tauscht man sich auch informell tiefer und schneller aus. Eine kontinuierliche Kommunikation stellt sicher, dass Transparenz über den aktuellen Stand, (adaptierte) Ziele sowie Informationen zur Entscheidungsfindung für alle Beteiligten zugänglich sind.

3.5 Institutionelle Verankerung der Kooperationsführung

Gegenseitiges Vertrauen, wechselseitige Nachsicht und gemeinsame Ziele bilden ein starkes Fundament für eine erfolgversprechende Kooperation. Eine ausgeglichene Entscheidungs- und Eskalationskompetenz bei allen Kooperationspartnern wird als selbstverständlich vorausgesetzt. Überdies erfordert eine Kooperation ein organisationales Grundgerüst und einen Sprecher/Leiter, welcher von allen Kooperationsteilnehmern akzeptiert und in seiner Rolle als neutral und daher institutionell nicht „eingefärbt“ wahrgenommen wird.

Eine gemeinsame Wahl des Sprechers mit gleich verteilter Stimmzahl erscheint somit obligatorisch. Und doch muss häufig festgestellt werden, dass die Akzeptanz eines Sprechers gerade bei konfliktbehafteten Entscheidungen leidet. Die umfassende Neutralität wird angezweifelt und schlimmstenfalls leidet die vertrauensvolle Zusammenarbeit. Daher empfiehlt es sich, die Kooperation übergreifend durch eine neutrale Instanz leiten und steuern zu lassen. Diese sollte von allen Kooperationspartnern akzeptiert werden. Dabei kann die Rolle sowohl von Leitern der Koordinierungsstellen, Personen im Netzwerk verbundener Hochschulen oder anderen externen Kräften wahrgenommen werden.

Im Folgenden werden zwei Kooperationsprojekte, die durch die myconsult GmbH in den letzten Jahren begleitet wurden bzw. derzeit noch begleitet werden hinsichtlich der benannten Erfolgsfaktoren betrachtet. Hierbei handelt es sich erstens um das Projekt zur Einführung eines „Landesweiten, logistischen Onlineverfahrens zur Vergabe von Praktikumsplätzen im Praxissemester“ mit der Maßgabe der Institutionalisierung eines nachhaltigen Kooperationsverbunds. Der zweite Erfahrungsbericht fokussiert das Projekt „bwCMS AP10“; ein Vorhaben zur Abstimmung universitären Aktivitäten bei der Erneuerung bestehender Campusmanagementsysteme an den Universitäten des Landes Baden-Württemberg unter Rückgriff auf den Innovations- und Qualitätssicherungsfond Baden-Württemberg.

4 Institutionalisierte Kooperation am Beispiel „Praxissemester NRW“ – erster Erfahrungsbericht

Der erste Erfahrungsbericht beschreibt das Projekt des Aufbaus und der Institutionalisierung eines Kooperationsverbunds aller lehrausbildenden Hochschulen in NRW, welches im Zeitraum von Januar 2013 bis März 2015 durchgeführt wurde.

4.1 Ausgangslage der Kooperationsaktivitäten

Das Lehrerausbildungsgesetz (LABG 2009) für Nordrhein-Westfalen legt für alle Lehramtsstudierenden an nordrhein-westfälischen Universitäten mit dem verpflichtenden Praxissemester ein neues Praxiselement in der universitären Lehrerausbildung fest.

Im Zuge der Einführung des Praxissemesters im Masterstudiengang der Lehrerausbildung haben sich die elf lehrausbildenden Universitäten in Nordrhein-Westfalen und das Ministerium für Schule und Weiterbildung zunächst gemeinsam darauf verständigt ein Onlineverfahren zu entwickeln, das landesweit die Verteilung der Studierenden an die Schulen der Ausbildungsregionen bewältigt.¹⁰ Die Verantwortung des Verfahrens liegt

¹⁰ Vgl. Rahmenkonzeption zur strukturellen und inhaltlichen Ausgestaltung des Praxissemesters im lehramtsbezogenen Masterstudiengang NRW vom 14. April 2010.

dabei bei den Universitäten in Zusammenarbeit mit ca. 6.000 Schulen, den ca. 35 Zentren für schulpraktische Lehrerausbildung sowie den fünf Bezirksregierungen in NRW.¹¹

Um von Synergieeffekten profitieren und die Expertise aller beteiligten Hochschulen und Einrichtungen nutzen zu können, wurde ein Projekt zur Einführung des Praxissemesters und damit einhergehend zum Aufbau und zur Verstetigung eines Kooperationsverbunds durch den Arbeitskreis DV-Infrastruktur (DV-ISA) des Landes NRW initiiert.

4.2 Ziele der Praxissemester-Kooperation

Das initiale Kooperationsprojekt und die das Projekt überdauernde Kooperation verfolgen die Zielsetzung einer Verteilung der pro Jahr landesweit ca. 7.500 Lehramtsstudierenden im Masterstudium auf die verfügbaren Praktikumsplätze an den Schulen im Land NRW. Die übergeordneten Ziele der Entwicklung dieses Verfahrens sind daher: (1) eine von allen beteiligten Hochschulen gemeinsam getragene Entwicklung und zukünftig auch Weiterentwicklung des angestrebten Online-Verfahrens, (2) die landesweite Einführung, Institutionalisierung und bei Bedarf Anpassung einheitlicher IT-gestützter Prozesse sowie (3) die Realisierung eines gemeinsamen Betriebs der Onlineplattform.

Hierzu bedarf es landesweit abgestimmter Prozesse für die Abwicklung des Praxissemesters, eines Konsenses über die spätere regelhafte Durchführung des Verfahrens sowie langfristig stabiler kooperativer Strukturen, die die langfristige Durchführung des Praxissemesters ermöglichen und stützen.

4.3 Ausgestaltung des Kooperationsprojekts

Neben den bereits skizzierten Herausforderungen eines neuartigen Bildungsbausteins, einer Vielzahl an heterogenen Projektbeteiligten und einer prozessualen sowie informationstechnischen Dimension des zu erarbeitenden Verfahrens, war insbesondere der absolut kritische Zeitplan eine nicht zu unterschätzende Rahmenbedingung, die eine zügige und effektive Zusammenarbeit in der Kooperation notwendig machte. Hierfür ergriff die von der myconsult GmbH gestellte Projektleitung gezielt Maßnahmen, welche auf die in Abschnitt 3 identifizierten Erfolgsfaktoren abstellten. Im Folgenden werden beispielhaft Maßnahmen beschrieben, welche die Erfolgsfaktoren (1) strukturelle und kulturelle Ähnlichkeit, (2) Vertrauen und (3) neue Organisationsform fokussieren.

Aufgrund der unterschiedlichen organisatorischen Hintergründe (Universitäten in Hochschulfreiheit einerseits und Bezirksregierungen in Direktive des Schulministeriums andererseits) sowie damit einhergehenden kulturellen Unterschieden der Kooperationsbeteiligten war der Erfolgsfaktor „Ähnlichkeit“ schwach ausgeprägt und auch per se im Vorfeld nicht veränderbar. Im Fokus der Aktivitäten stand daher über das Schaffen einer

¹¹ Vgl. hierzu die Schulstatistik des Ministeriums für Schule und Weiterbildung des Landes Nordrhein-Westfalen.

uneingeschränkter Transparenz von Prozessen, Akteuren und Rahmenbedingungen ein Verständnis für die Situation „des Anderen“ aufzubauen. Über moderierte, intensive inhaltliche Diskussionen in Form von Workshops, Gesprächsrunden und fallweiser Mediation konnte eine Sensibilität für die „Andersartigkeit“, d.h. für Zwänge und Restriktionen der einzelnen Kooperationspartner erwirkt werden.

Ein weiterer benannter Faktor für das erfolgreiche Etablieren von Kooperationen ist ein hohes Maß an gegenseitigem Vertrauen; dieses in kürzester Zeit zu erzielen war eine zentrale Aufgabe im Kooperationsprojekt. Insbesondere das Aufbrechen von historisch begründeten, verhärteten Fronten zwischen hochschulseitigen Akteuren und Beteiligten der Bezirksregierung war eine große Herausforderung. Das Schaffen von **gegenseitigem Vertrauen und wechselseitiger Nachsicht** wurde initiiert durch einen gemeinsamen dreitägigen Workshop, der neben fachlichem Austausch Elemente der Teambildung enthielt. Hierdurch wurde zunächst ein kleines Kooperationssteam mit gleichem Spirit und Engagement geformt, welches den Nukleus der Kooperation bildete. Durch eine sukzessive Hinzunahme weiterer Mitglieder erlangte die Kooperation Zusammenhalt und Stabilität.

Ein Ziel des Projekts war die nachhaltige Etablierung eines Kooperationsverbunds zum gemeinsamen Betrieb des entwickelten Onlineverfahrens. Der identifizierte Erfolgsfaktor einer neuen **Organisationsform** war inhärenter Bestandteil des Auftrags und führte von Beginn an dazu, dass die Kooperationspartner ihre Aktivitäten stets an dem Aspekt der Nachhaltigkeit ausgerichtet haben und das gemeinsame Commitment zur Kooperation PVP extrem hoch ausgeprägt war.

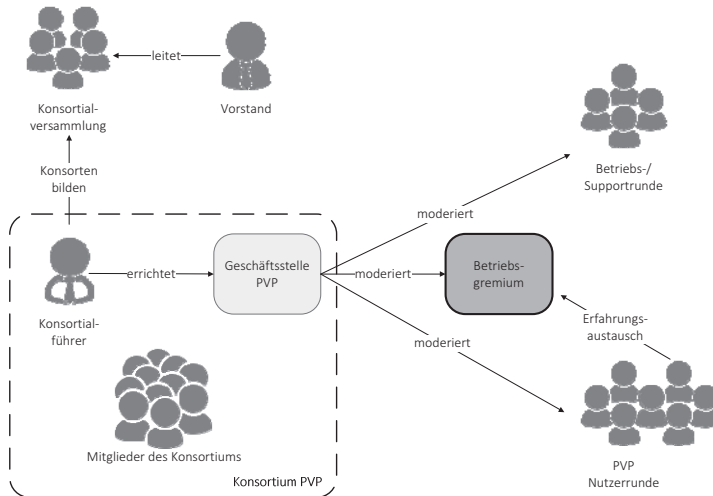


Abbildung 2: Gremienstruktur der Kooperation PVP

4.4 Resümee

Zusammenfassend lässt sich konstatieren, dass zum Start der Kooperation PVP wenige der benannten Erfolgsfaktoren stark ausgeprägt vorlagen. Dennoch hat die durch Lehrer-ausbildungsgesetz erwirkte Zielidentität sowie der enorme Zeit- und Erfolgsdruck dazu beigetragen, dass durch systematische und nachhaltige Maßnahmen der Team- und Kooperationsentwicklung ein erfolgsbringendes Fundament von Vertrauen und Nachsicht, von Homogenität und von zielführender Kommunikation geschaffen wurde, welches in einer neuen Organisationsform mündete.

5 Projektbezogene Kooperation am Beispiel „bwCMS“ – zweiter Erfahrungsbericht

5.1 Ausgangslage der Kooperationsaktivitäten

Das Land Baden-Württemberg unterstützt die Landeskooperation bwCMS, in der die neun staatlichen Universitäten gemeinsam Fragestellungen zu (1) der Abstimmung bei der Einführung eines Campus Management Systems (CMS) – trotz unterschiedlicher Softwarelieferanten) – (2) der Definition von gemeinsamen Themenschwerpunkten mit Erarbeitung von „Best Practices“ sowie (3) der Umsetzung in Entwicklungspartnerschaften erarbeiten.

5.2 Ziele der Kooperation

Die Landeskooperation bwCMS bearbeitet inzwischen unter anderem Arbeitspakete zu den Themen Dialogorientiertes Serviceverfahren (DoSV), Bewerbung und Zulassung, Geschäftsprozessmodellierung, Chipkarten- und Berechtigungsmanagement sowie elektronische Studierendenakte. Dazu werden innerhalb der Landeskooperation bwCMS einzelne Arbeitspakete (AP) gebildet und bearbeitet; diese AP bilden den inhaltlichen Rahmen, in welchem konzentriert die jeweilige Themen-/Problemstellung bearbeitet wird.

Ein Beispiel für ein solches Arbeitspaket ist das bwCMS AP10, das an dieser Stelle in den Fokus genommen wird. Das AP10 hat das Ziel der Weiterentwicklung des Einsatzes der CMS-Lösungen an den Universitäten des Landes Baden-Württemberg, des Wissens- und Erfahrungsaustauschs in Bezug auf das CMS sowie die gemeinsame Erarbeitung von Fachkonzepten und deren Umsetzung. Im Spezifischen wird hier die Weiterentwicklung des CMS Systems HISinOne der HIS eG betrachtet.

5.3 Ausgestaltung der Kooperation

Im AP10 sind die sechs HISinOne einsetzenden Universitäten des Landes Baden-Württemberg formiert. Diese weisen heterogene Fortschritte bei der Einführung von HISinOne an der jeweiligen Hochschule auf – in Teilen sind die CMS-Module bereits umfassend eingeführt, in Teilen hat die Einführung des CMS gerade erst begonnen.

Das Arbeitspaket AP10 startete im März 2015 mit folgendem Vorgehen: Der extern bestellte Projektleiter der myconsult GmbH bringt den Vorschlag ein, die Entwicklung der Anforderungen agil in inkrementellen Iterationen zu gestalten, d. h. schnell in monatlichen Abständen die Anforderungen für einen Zeitraum von vier Wochen zu konzipieren (Sprint Planning) sowie die im vorhergehenden Sprint entwickelten Funktionalitäten zu testen. Inhaltlich wurden dabei Anforderungen aus den Bereichen Bewerbung und Zulassung, Studierendenmanagement, Veranstaltungs- und Prüfungsmanagement sowie systemweite Komponenten als Anforderung in Form von Fachdesigns definiert und umgesetzt.

Bezug nehmend auf die o.a. Erfolgsfaktoren kann bereits für die Ausgangslage konstatiert werden, dass einige dieser Erfolgsfaktoren bereits stark ausgeprägt waren. So arbeiten die einzelnen Landesuniversitäten und insbesondere deren personelle Vertreter des AP10 seit langem und in verschiedenen Kontexten zusammen. Dadurch lag ein basales **Vertrauen** der Universitätsvertreter untereinander bereits vor, das einen konstruktiven Umgang im Projekt ermöglichte. Diese Zusammenarbeit ist aber nicht als „geschlossener Club“ zu verstehen, ganz im Gegenteil herrscht eine offene, begrüßende Attitüde, die es neuen Mitgliedern einfach macht, sich zu integrieren. In Kombination mit gegenseitiger **Nachsicht** und Rücksichtnahme entstand erst gar nicht die Situation, dass Partikularinteressen dominierten, vielmehr war ein **gemeinsam getragenes Ziel**, das hinsichtlich der benannten Erfolgsfaktoren betrachtet bestehende CMS HISinOne weiter zu entwickeln und dessen Funktionalität auszubauen. Eine Ausgangslage bzgl. der bestehenden Gebiete für Weiterentwicklung war im Konsortium bereits zusammen getragen worden.

Die Formulierung von Fachdesigns als Grundlage für die Entwicklung stellte sich als kommunikationsintensives Unterfangen dar. Der eigentlichen Abstimmung zwischen Kunde und Softwarelieferant vorgelagert, musste in einem ersten Schritt ein Konsens über die inhaltliche Ausgestaltung der jeweiligen Anforderung entschieden werden. Das agile Verfahren stellte dafür einen geeigneten Ansatz dar; Schwerpunktthemen konnten iterativ abgestimmt und weiter entwickelt werden. Diese Abstimmungsrunden innerhalb der Kooperation, die sich auch bei weiteren Arbeitsschritten (Test, Abnahme etc.) wiederholen, müssen koordiniert werden. Periodische Treffen des gesamten AP10 vis-a-vis haben sich innerhalb der Projektlaufzeit als elementarer Erfolgsfaktor für das Projekt erwiesen. Diese Präsenztreffen bedeuten auf der einen Seite natürlich Aufwand für Anreise etc., diesem Aufwand steht allerdings ein Forum als Gegenwert gegenüber, das insbesondere den direkten Austausch der Beteiligten ermöglicht. Der Erfolgsfaktor der (direkten) **Kommunikation** bzw. eines geeigneten Mix aus IT-gestützter Kommunikation und Präsenztreffen ist elementar im Projekt, um (1) sicherzustellen, dass eine

kontinuierliche Information aller Beteiligten erfolgt und (2) insbesondere die Präsenztreffen nicht als zusätzliche Belastung empfunden werden.

5.4 Resümee

Das AP10 wird Ende Juni 2016 abgeschlossen werden. Im Rahmen des Projekts wurden substanzielle und komplexe Anforderungen fachlich spezifiziert und durch den Softwarelieferanten in HISinOne implementiert. Diese werden als Teil des jeweiligen Releases zwei Mal pro Jahr allen Nutzern des CMS HISinOne zur Verfügung gestellt. In diesem Sinne kann das Projekt bereits als Erfolg gelten.

Die Kooperation AP10 wies bereits einige der grundlegenden o. a. Erfolgsfaktoren auf (gegenseitiges Vertrauen, Rücksichtnahme, strukturelle und kulturelle Ähnlichkeiten etc.). Durch die Schaffung passender Rahmenbedingungen, die das Eintreten weiterer Erfolgsfaktoren ermöglichten und begünstigten, gelang es, die Zusammenarbeit innerhalb der Kooperation der sechs Hochschulen untereinander als auch mit dem Softwarelieferanten gemeinsam so zu strukturieren, dass zielorientiert in kurzer Zeit Fachdesigns zu den fachlichen Anforderungen entstanden, diese vom Softwarelieferanten umgesetzt und von den Hochschulen geprüft und abgenommen wurden. Dabei kamen „neue“ bzw. in dem Verhältnis Kunde/Softwarelieferant bislang unübliche agile Herangehensweisen zum Einsatz, die sich auch in diesem Kontext als belastungsfähig und tragend erwiesen haben.

6 Fazit

Auch wenn Kooperationen zwischen Hochschulen in unterschiedlicher Ausprägung seit langem bestehen, so ist in letzter Zeit ein verstärkter Trend hin zur Bildung von Kooperationen zu erkennen. Effizienzbestrebungen, Komplexitätssteigerungen aber auch steigender Wettbewerbs- und Leistungsdruck werden in Zukunft dazu führen, dass das Eingehen von Hochschulkooperationen für eine bessere Leistungsfähigkeit weiter zunimmt.

Der Beitrag listet in der Literatur genannte Faktoren, welche auch vor dem praktischen Erfahrungshintergrund der Autoren für eine erfolgreiche Bildung von Kooperationen und eine wirkungsvolle kooperative Zusammenarbeit stehen. Dabei lässt sich konstatieren, dass Bereitschaft zur Zusammenarbeit, Vertrauen und gemeinsame Ziele wesentliche Größen für eine erfolgreiche Kooperation sind. Diese sollten bereits in der Ausgestaltung der Planung einer Kooperation berücksichtigt werden. Neben einer im Hochschulkontext i.d.R. gegebenen strukturellen und kulturellen Ähnlichkeit, kann der Erfolgsfaktor Zielidentität durch explizites Formulieren, Diskussion und gemeinsame Verabschiedung der Kooperationsziele gestärkt werden. Offene und zielgerichtete Kommunikation sowie eine institutionelle Verankerung erzeugen einen Rahmen, in dem die gemeinsame Arbeit zur Erreichung der Ziele Vertrauen und wechselseitige Nachsicht gedeihen lässt.

Dabei sind die genannten Erfolgsfaktoren nicht im Sinne eines finalen, zu absolvierenden „Katalogs“ zu verstehen – sie stellen notwendige, aber nicht hinreichende Faktoren dar. Letztendlich wird der Erfolg einer Kooperation durch das Engagement und die Beteiligung der involvierten Personen entschieden. Individuelle auf die jeweilige Kooperation abgestimmte Maßnahmen zur Realisierung der Erfolgsfaktoren können dafür ein geeignetes Umfeld und Rahmenbedingungen schaffen.

Literaturverzeichnis

- [BDHL11] Becker, T.; Dammer, I.; Howaldt, J.; Loose, A.: Netzwerkmanagement mit Kooperation zum Unternehmenserfolg; 2011.
- [HEB07a] Hener, Y.; Eckardt, P.; Brandenburg, U.: Kooperationen zwischen deutschen Hochschulen - Arbeitspapier Nr. 85; S. 24 f., 2007.
- [HEB07b] Yorck, H.; Eckardt, P.; Brandenburg, U.: Kooperationen zwischen deutschen Hochschulen - Arbeitspapier Nr. 85; S. 43 ff., 2007.
- [HRK13] Hochschulrektorenkonferenz: Entschließung des 124. Senats der Hochschulrektorenkonferenz, Perspektiven des Wissenschaftssystems; S. 5f, 2013.
- [DFG10] Deutsche Forschungsgemeinschaft: Informationsverarbeitung an Hochschulen – Organisation, Dienste und Systeme Empfehlungen der Kommission für IT-Infrastruktur für 2011–2015, S.3, 2010.
- [HePa13] Hechler, D.; Pasternack, P.: Zwischen Kooperieren und Fusionieren; Wissenschaftsmagazin 4/2013, S.14-17, 2013.
- [HR16] Horizon Report: 2016 Higher Education Edition (Hochschulausgabe); The New Media Consortium; 2016.
- [KL03] Killich, S.; Luczak, H.: Unternehmenskooperation für kleine und mittelständische Unternehmen. Lösungen für die Praxis. S.8, 2003.
- [LrNW13] Landesrechnungshof Nordrhein-Westfalen, Jahresbericht 2013, S. 99f., 2013.
- [LrMV14] Landesrechnungshof Mecklenburg-Vorpommern, Sonderbericht über die Prüfung der Hochschulfinanzierung; 2014.
- [NDH97] Nieschlag, R.; Dichtl, E.; Hörschgen, H.: Marketing; S. 1054, 1997.
- [Ro00] Royer, S.: Strategische Erfolgsfaktoren horizontaler kooperativer Wettbewerbsbeziehungen : eine auf Fallstudien basierende erfolgsorientierte Analyse am Beispiel der Automobilindustrie; S. 18ff, 2000.
- [Sc79] Schwarz, P.: Morphologie von Kooperationen und Verbänden; S. 65, S. 83, 1979.
- [WR16] Wissenschaftsrat: Empfehlungen zur Weiterentwicklung der MINT-Bereiche an den Hochschulen des Landes Hamburg; Drs. 5085-16; 2016.

Integration von Lösungen für das Bewerbungsmanagement mit Job-Portalen und beruflichen Netzwerken

Fritz Rose¹, Dirk Bernds², Michael Gassert³, Folke-Gert Stümpel⁴, Ulrike Lucke¹

Abstract: Das Personal-Management stellt ein für das Funktionieren von Hochschulen sehr wichtiges, aber im Zuge der Digitalisierungsbemühungen oft vernachlässigtes Feld dar. Die Gewinnung und Einstellung neuer Mitarbeiter ist ein aufwändiges und oft noch papierbasiertes Verfahren. Online-Bewerbungs-Lösungen können die Datenübertragung und -verarbeitung vereinfachen, jedoch bleibt der Ausgangspunkt das Ansprechen interessierter Bewerber. Die hierfür bestehenden Web-Plattformen sind bislang nicht mit der Software zur Personalverwaltung an Hochschulen gekoppelt. Der Beitrag stellt eine Lösung zur flexiblen Integration dieser bislang losgelösten Systeme vor.

Keywords: Personalverwaltung, Human Resource Management, Integration

1 Einleitung

Die fortschreitende Digitalisierung stellt aktuell eine der größten Herausforderungen für Hochschulen dar [Me12]. Während im Bereich Studium und Lehre seit etwa 15 Jahren mit umfangreichen Fördermitteln des Bundes unter den Stichworten „Multimedia“, „Neue Medien“ oder „E-Learning“ sowie derzeit v.a. mit dem Ziel der Verstetigung und Qualitätssicherung Forschungs- und Entwicklungsarbeiten zur Digitalisierung der Hochschullehre voran getrieben wurden, ist in den Bereichen Forschung und Verwaltung kaum eine systematische Weiterentwicklung zu verzeichnen [HFD15]. Zwar sind unter dem Begriff „E-Science“ einige Forschungsprojekte und Pilotlösungen entstanden, doch gerade in der Verwaltung beschränken sich die Fortschritte auf lokale Aktivitäten einzelner Hochschulen. Dabei ist angesichts steigender Studierenden- und Drittmittelzahlen gerade in der Hochschulverwaltung – und hier insbesondere im Kontakt mit „Kunden“, d.h. vorwiegend mit Studierenden und Mitarbeitern – die erforderliche Effektivität und Effizienz der Verfahren ohne moderne Online-Lösungen kaum erreichbar. Zugleich ähneln sich die Verwaltungsverfahren trotz aller lokalen und/oder regionalen Spezifika weit genug, um übergreifende Weiterentwicklungen sinnvoll erscheinen zu lassen. Die dank moderner Ansätze der Systemarchitektur verfügbaren Mechanismen zur individuellen Konfiguration und flexiblen Integration von digitalen Lösungen tragen

¹ Universität Potsdam, Institut für Informatik & Computational Science, August-Bebel-Str. 89, 14482 Potsdam

² Universität Potsdam, Zentr. Einr. f. Informationsverarb. und Komm., Am Neuen Palais 10, 14469 Potsdam

³ Universität Potsdam, Dezernat für Personal- & Rechtsangelegenheiten, Am Neuen Palais 10, 14469 Potsdam, vorname.nachname@uni-potsdam.de

⁴ HIS Hochschul-Informations-System eG, Gosseriede 9, 30159 Hannover, stuempel@his.de

ebenfalls dazu bei.

Im Bereich Campus-Management sind – vermutlich aufgrund der großen Fallzahlen sowie regulativer und organisatorischer Rahmenbedingungen – bereits zahlreiche Online-Self-Service-Lösungen sowohl für Studierende (v.a. Bewerbung, Immatrikulation, Kursbelegung, Prüfungsanmeldung, Noteneinsicht) als auch für Mitarbeiter (v.a. Kursanmeldung und Noteneintragung) verfügbar. Es gibt eine Handvoll etablierter Software-Lösungen, die derartige Verfahren mehr oder weniger umfassend, mehr oder weniger individuell, mehr oder weniger integriert unterstützen [Wi14]. Auch im Bereich des Ressourcen-Managements finden sich verwandte Produkte, die sich aber vorwiegend an Mitarbeiter in der Verwaltung richten und kaum Self-Service-Funktionen bieten. Nimmt man die in den vergangenen Jahren auf dem Workshop „Hochschule 2025“ bzw. seinen Vorgängern publizierten Ansätze als Basis, scheinen sich die Einsatzgebiete von Online-Verwaltungslösungen auf die Bewirtschaftung von Finanzmitteln zu fokussieren, insbesondere auf Drittmittelprojekte, Beschaffungen und Dienstreisen [ZKB12][Ka⁺13]. Die Gründe dafür mögen sowohl in den unmittelbaren Bedarfen der dies treibenden Wissenschaftler als auch in den überschaubaren Prozessen und gesetzlichen Vorgaben liegen.

Dieser Beitrag widmet sich dem bislang noch wenig betrachteten Gebiet der Personalverwaltung [Gr05]. Aufgrund der Komplexität der Prozesse, die v.a. aus der intensiven Verarbeitung personenbezogener Daten und den dafür etablierten Schutzmechanismen resultieren mögen, scheint dies zunächst ein schwieriges Feld zu sein. Jedoch stehen auch die Hochschulen zunehmend unter dem Druck, im Wettbewerb um die besten Köpfe bei der Ausschreibung von Stellen gegenüber der Konkurrenz (nicht nur aus der Akademia, sondern auch aus der Wirtschaft) bestehen zu können. Dies wird verschärft durch die oft unattraktive Entlohnung und befristete Einstellung, sowohl beim akademischen Personal als auch bei Mitarbeitern in Technik und Verwaltung. Nicht selten bleiben Stellen trotz mehrfacher Ausschreibung unbesetzt – ein gravierendes Problem sowohl für „Tabellenführer“ als auch im „Abstiegskampf“. Hinzu kommt, dass die etablierte Software in diesem Bereich im Wesentlichen proprietär und monolithisch ist, d.h. kaum Schnittstellen aufweist oder gar Standards unterstützt. Zwar gibt es mit Human Resource Open Standards⁵ eine (industrietriebebene) Spezifikation, die Datenstrukturen und Austauschmechanismen für die Personalverwaltung definiert, aber diese wird häufig mit individuellen Abwandlungen einzelner Anbieter genutzt. Im Ergebnis ist die Interoperabilität bestehender Lösungen sehr begrenzt. Auf der anderen Seite existieren zahlreiche Online-Portale für die Veröffentlichung von bzw. Suche nach Stellenausschreibungen (privat wie öffentlich finanziert), bis hin zu komplexen beruflichen Netzwerken (i.d.R. privat finanziert). Allerdings ist hier bislang keine Durchlässigkeit zu Verwaltungssoftware an Hochschulen zu erkennen; Daten werden über Web-Formulare manuell eingetragen, sodass angesichts der hohen Personalfuktuation bislang kein effizienter Datenaustausch möglich scheint.

⁵ <http://www.hropenstandards.org/>

Die Etablierung von Schnittstellen zwischen der Personalverwaltungs-Software an Hochschulen einerseits (insbesondere deren Online-Lösungen für das Bewerbungs-Management) und den existierenden Job-Portalen und beruflichen Netzen andererseits ist daher das Ziel der hier vorgestellten Arbeiten. In Abschnitt 2 werden zunächst die Anforderungen der Personalverwaltung und die vorhandenen Software-Lösungen analysiert. Daraus wird in Abschnitt 3 eine Integrationslösung abgeleitet, deren Architektur und Schnittstellen vorgestellt werden. Abschnitt 4 fasst das bislang Erreichte sowie weitere nötige Arbeiten zusammen.

2 Software im Personalwesen

2.1 Der Prozess der Personalgewinnung

Eine zielgerechte Personalgewinnung an Hochschulen erfordert die vernetzte Zusammenarbeit verschiedener interner Akteure wie zentraler Universitätsverwaltung, Fakultäten und Zentralen Einrichtungen, Personalvertretungen, Gleichstellungsbeauftragte sowie Schwerbehindertenvertretung. Für die hochschulspezifischen Personalkategorien der Professoren, wissenschaftlichen und künstlerischen Mitarbeiter sowie den Mitarbeitern aus Technik und Verwaltung existieren zahlreiche Verfahrensvariationen im Personalgewinnungsprozess, einerseits bedingt durch unterschiedliche gesetzliche oder tarifliche Festlegungen oder zusätzliche besondere landesverwaltungsspezifische Regelungen. Der Prozess der Personalgewinnung basiert auf klar formulierten Aufgabenverantwortungen der internen Akteure, spezifischen Anforderungskriterien in den unterschiedlichen Personalkategorien und damit jeweils verbundenen besonderen Auswahlinstrumenten. Das gemeinsame Zusammenwirken von zentralen und dezentralen Akteuren erhöht einerseits die Prozesskomplexität und führt andererseits zu zahlreichen Rückkopplungen. Abb. 6 im Anhang dieses Beitrags zeigt beispielhaft den für die Universität Potsdam erhobenen Prozess.

Zusätzlich zu den internen Akteuren bestehen mehrere Schnittstellen zu externen Akteuren. Das sind einerseits Schnittstellen zu Print- und Onlinemedien zur Veröffentlichung von Stellenausschreibungen (nachfolgend aufgegriffen als Use Case 1), gegebenenfalls unter vermittelnder Beteiligung einer Medienagentur, und andererseits als wohl wichtigste Schnittstelle diejenige zwischen Bewerber und Hochschule (Use Case 2). Über diese Schnittstelle werden weit über 95 % aller Bewerbungen in Form von E-Mails mit Anhängen in unterschiedlichen Formaten übermittelt. Aufseiten der Hochschule erfolgen Sichtung und Sortierung der per E-Mail eingegangenen Daten und Dateien notwendigerweise noch von Hand, ebenso wie deren Aufbereitung, Kommentierung und Erfassung der Kontaktdaten. Das auf diese Weise erstellte Gesamtpaket an Bewerberdaten bildet die Grundlage für die weitergehende Arbeit der Personalauswahlkommission. Spätestens hier stellt sich die Frage, ob mit elektronischen Dokumenten oder ausgedruckten und vervielfältigten Dokumenten weitergearbeitet werden muss. Am Ende des Personalauswahlprozesses gilt es, zunächst die nicht erfolgreichen Bewerber über die

Auswahlentscheidung zu informieren und die vorliegenden, nicht mehr benötigten Bewerberdaten zu vernichten bzw. zu löschen. Die prozessdokumentierenden Daten, die personenbezogenen Daten des erfolgreichen Bewerbers nebst der Protokollierung der Auswahlentscheidung bilden den Abschluss des Personalgewinnungsprozesses. Mit der Unterzeichnung des Arbeitsvertrages endet der organisationsinterne Prozess und wird mit der Anmeldung des neuen Mitarbeiters für die Vergütungsabrechnung bei einem weiteren externen Akteur zum endgültigen Abschluss gebracht.

2.2 Standardsoftware im Personalwesen

Knapp die Hälfte der deutschen Hochschulen setzt das Personalmanagementsystem HIS-PM der Hochschul-Informationssystem (HIS) eG ein ein. Dieses wurde und wird von der HIS eG gemeinsam mit den Hochschulen abgestimmt und weiterentwickelt. Ziel ist es, möglichst viele Belange des Hochschul-Personalmanagements abzubilden. Folgende Bereiche werden zurzeit vom HIS-PM abgedeckt: Stellenbewirtschaftung, Personalentwicklung, Abwicklung von Personalvorgängen, Zeiterfassung, Dienstreise-, Abwesenheits-, Bewerbungs-, Personalkosten- und Workflowmanagement sowie die Erstellung von Auswertungen und Statistiken. Es werden jeweils Oberflächen für die Personalverwaltung und – falls zweckmäßig – Web-Dialoge für die Bewerber bzw. die Beschäftigten der Hochschule angeboten.

Wichtige Kennzeichen des HIS-PM (und auch Unterscheidungsmerkmale zu anderen Software-Lösungen) sind, neben den zahlreichen Konfigurationsmöglichkeiten, die integrierte Personenverwaltung mit der Campus-Management-Software der HIS sowie die Abbildung vieler Hochschulspezifika. Hier sind zu nennen die spezifischen Personengruppen wie Professoren, Hilfskräfte, Lehrbeauftragte usw. sowie die Unterstützung von spezifischen Funktionen wie Kapazitätsberechnung und Hochschulstatistik. Nicht zuletzt bietet das HIS-PM die Möglichkeit, Personen mit mehreren (teilweise parallelen) Beschäftigungsverhältnissen und Mittelquellen zu verwalten. Das HIS-PM enthält Schnittstellen zum HIS-Finanzsystem und weiteren Haushaltssystemen, Bezügesystemen, Office-Software, Zeiterfassungsterminals sowie Verzeichnisdiensten.

Der Client-Server-Teil des HIS-PM basiert auf Microsoft Visual C++. Es ist im Produktportfolio von Microsoft fest verankert und zukunftssicher. Die Architektur der Ressourcensysteme als Microsoft-basierte Client-Server-Systeme wird entsprechend dieser Festlegungen unverändert weitergeführt. Die Java-basierten Teile des PMs werden unter Verwendung der QIS-Technologie (Qualitätssteigerung im Internet durch Selbstbedienung) entwickelt. Zurzeit wird das HIS-PM mit HISinOne integriert. Hier stehen die Personen im Fokus. Zukünftig können die Personendaten des Campus-Managements integriert mit denen des HIS-PMs verwaltet werden. Dadurch wird das redundante Pflegen von Personendaten (zum Beispiel Hilfskräften) vermieden und die Entwicklung von Schnittstellen und übergreifenden Auswertungen, die Personendaten im Fokus haben, vereinfacht. Sowohl die Java-basierten Produkte als auch die Client-Server-

Software der HIS eG folgen einem Model-View-Controller-Architekturmuster, wie Abb. 1 zeigt. Auf GX-Seite beschreibt das Interface-Objektschema (IOS) wie Objekte kommuniziert werden sollen, sowohl Menschen als auch anderen Computern gegenüber. Das Vorgangsobjektschema (VOS) definiert auf welche Art die Objekte in den Geschäftsprozessen miteinander agieren, während das Konzeptuelle Objektschema (KOS) die Beziehung der Objekte zueinander darstellt. QIS lässt den Nutzer per Browser auf eine von der View-Erzeugung generierten Oberfläche mit der Geschäftslogik interagieren. Die Operationen beider Systeme werden anschließend in einer Datenbank persistiert. HIS-PM unterstützt sowohl INFORMIX als auch PostgreSQL zur Datenhaltung.

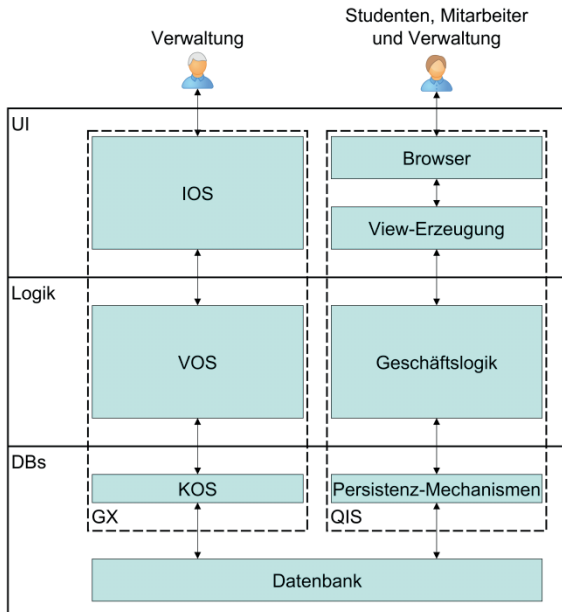


Abb. 1: Grundstruktur der HIS-Software für das Personal-Management

Die oben erwähnten Schnittstellen basieren auf unterschiedlichen Ansätzen wie dem Lesen oder Schreiben von Text- bzw. XML-Dateien, der Verwendung von Webservices oder dem Datenaustausch unter Verwendung von gemeinsamen Datenbank-Tabellen. Die Entscheidung für eine Schnittstellentechnologie ist in der Regel abhängig von den Anforderungen der durch die Schnittstelle integrierten Software.

2.3 Job-Portale und berufliche Netzwerke

Die derzeit relevanten Stellenportale lassen sich auf verschiedene Arten kategorisieren. Zum einen gibt es öffentlich finanzierte Portale, wie z.B. in Deutschland Bund.de⁶ und die

⁶ <http://www.bund.de/>

Bundesagentur für Arbeit⁷ sowie europaweit Euraxess⁸. Zum anderen gibt es die kommerziellen Angebote; am bekanntesten sind Monster⁹ und StepStone¹⁰. Zudem existieren einige berufliche Netzwerke wie Xing¹¹, LinkedIn¹² und ResearchGate¹³. Die kommerziellen Angebote sind i.d.R. weltweit tätig.

Die Portale lassen sich auch nach ihren Zielgruppen charakterisieren. Während Euraxess, academics.de¹⁴, ResearchGate und Absolventa¹⁵ sich speziell an Akademiker richten, zielen Plattformen wie Xing, LinkedIn und Jobbörse¹⁶ auf weniger spezifische Zielgruppen ab. Eine Besonderheit bei academics.de ist zudem die Kopplung an die Print- und Online-Ausgabe des Magazins DIE ZEIT.

Eine weitere Unterteilungsmöglichkeit besteht in der Unterteilung nach den für die Kunden eröffneten Möglichkeiten, mit Nutzern zu interagieren. So lassen einige Portale nur Stellenanzeigen zu, während bei anderen auch direkt bereits registrierte Bewerber gesucht werden können oder Firmen an potenziell interessierte Bewerber gezielt E-Mails schreiben können, ohne dass die Firma Informationen über die einzelnen Bewerber erhält.

Tab. 1 stellt einen Ausschnitt der für diese Arbeit betrachteten Stellenbörsen dar. Bei der Auswahl der potentiellen Partner für eine Zusammenarbeit wurde darauf geachtet, dass möglichst verschiedene Portale (öffentliche und private Stellenbörsen und berufliche Netzwerke) angebunden werden, um weitgehend sicherzustellen, dass die Architektur alle relevanten Fälle abdeckt. Weiterhin wurde beachtet, inwieweit bisher gute Erfahrungen mit den entsprechenden Portalen vorliegen.

Die in Tab. 1 dargestellten Angaben über die Anzahl der registrierten Suchenden und Stellenanzeigen sind jeweils von den Seitenbetreibern selbst angegeben. Ausnahmen bilden hierbei die ungerundeten Angaben, welche die Ergebnisse der Suche nach Stellenausschreibungen in dem jeweiligen Portal ohne spezielle Kriterien angeben. Die IVW-Visits¹⁷ (d.h. tatsächliche, zusammenhängende Nutzungsvorgänge) sind leider nur von wenigen Portalen verfügbar.

Um eine möglichst große Wirkung zu erzielen, sollte zuerst mit großen Portalen zusammengearbeitet werden, also solchen, bei denen bereits viele Nutzer registriert sind und viele Stellenanzeigen verfügbar sind.

⁷ <https://jobboerse.arbeitsagentur.de/>

⁸ <http://ec.europa.eu/euraxess/>

⁹ <http://www.monster.de/>

¹⁰ <https://www.stepstone.de/>

¹¹ <https://www.xing.com/>

¹² <https://www.linkedin.com/>

¹³ <https://www.researchgate.net/>

¹⁴ <http://academics.de/>

¹⁵ <http://www.absolventa.de/>

¹⁶ <https://www.jobboerse.de/>

¹⁷ <http://www.ivw.eu/digital/digital>

Name	# registrierte Suchende	# Stellenausschreibungen	Use Case 1: annonsieren	Use Case 2: bewerben	IVW-Visits
Bund.de		1.226	+	-	8.100.683
Bundesagentur für Arbeit	2.900.000	1.060.000	+	-	
Euraxess	37.601	4.366	+	-	
academics.de			+	-	
LinkedIn	300.000.000		+	+	
ResearchGate	8.000.000			-	
StepStone	13.600.000	500.000	+		
Monster	435.000		+	+	
Xing	15.000.000	1.000.000	+		2.734.148
Jobbörse		1.519.553			
Yourfirm	600.000				
Absolventa	80.000	7.805	+		
indeed		108.000	-	-	

Tab. 1: Übersicht relevanter Stellenbörsen und beruflicher Netzwerke

3 Integration der bestehenden Lösungen

Die Herausforderung bei der Integration der vorgestellten Lösungen besteht darin, nicht nur auf Seiten der Stellenbörsen bzw. beruflichen Netzwerke verschiedene Schnittstellen vorzufinden, sondern auch auf Seiten der Hochschule grundsätzlich offen für verschiedene Personalverwaltungs-Lösungen zu bleiben, auch wenn die vorgestellte Lösung auf Basis

des HIS-PM entwickelt wurde. Daher wurde eine Middleware entworfen, die typische Datenstrukturen und Funktionalitäten beider Akteursgruppen miteinander verbindet. Nachfolgend werden zunächst die grundlegende Architektur und danach die einzelnen Schnittstellen zu Drittanbietern vorgestellt.

3.1 Gesamtarchitektur

Die Integrationslösung hat als Hauptakteure die Zuständigen für die Personalverwaltung an der Hochschule. Diese hoheitliche Aufgabe sollte im Normalfall nicht ausgelagert werden, sodass auch die Middleware zur Anbindung von Drittlösungen in der Hochschule selbst (und zwar in unmittelbarer Anbindung an die Verwaltungs-Software) betrieben werden sollte.

Die in Abb. 2 dargestellte Programmbibliothek Job Posting Helper (JPH) beinhaltet die benötigte Funktionalität, um verschiedene Import- und Export-Plug-ins anzusprechen. Hierdurch werden sowohl mehrere parallel konfigurierte Import- als auch Export-Möglichkeiten berücksichtigt. Der Export über eine Ausgabe-Lib ermöglicht es, die Logik für den Export von Stellenausschreibungen (Use Case 1) individuell an die einzelnen Portale anzupassen, wodurch für jedes Portal eine eigene Export-Möglichkeit gegeben ist. Der Import von Bewerberdaten (Use Case 2) ist ebenfalls modular aufgebaut, damit auch andere Systeme anstelle der HIS-Datenbank für Importe genutzt werden können.

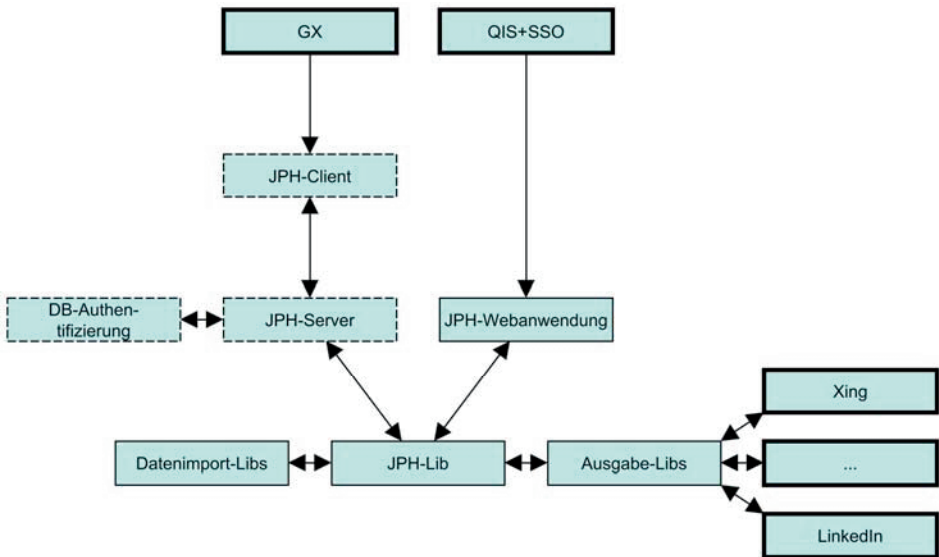


Abb. 2: Der Job PostingHelper fungiert als Middleware zwischen Personalverwaltung der Hochschulen (hier: HIS-PM) und externen Job-Portalen

Durch die Kapselung der Kernfunktionalitäten in einer Bibliothek (JPH-Lib) können auch

die Oberflächen an die Gegebenheiten einer Universität individuell angepasst werden. Prototypisch entwickelt wurde eine Anbindung als Web-Applikation unter Verwendung von Single-Sign-On (SSO) durch die Self-Service-Komponente (QIS) der HIS-Software.

Um den derzeitigen Prototypen in einer Hochschule anzubinden, welche ein anders Hochschulverwaltungssystem verwendet, muss ein neues Import-Plug-In geschrieben werden. Oberflächenseitig muss nur die Authentifizierung der Nutzer angepasst werden. Die gewählte Architektur und die genutzten Technologien zur Implementierung ermöglichen eine plattformunabhängige Nutzung und damit eine einfache Integration in andere Hochschulinfrastrukturen. Darüber hinaus erlaubt die modulare Arbeitsweise der Endpunkte zum Im- und Export eine weitaus größere Flexibilität als proprietäre Punkt-zu-Punkt-Verbindungen.

3.2 Schnittstellen zu Job-Portalen

Die Schnittstellen zu den Portalen sind sehr unterschiedlich. Bei kleinen Datenmengen wird von den Portalen empfohlen, ein Web-Formular zu nutzen um neue Stellenausschreibungen zu übermitteln. In Abb. 3 ist exemplarisch ein Ausschnitt der Weboberfläche des Portals Bund.de dargestellt. Für große Fallzahlen ist diese manuelle Vorgehensweise allerdings nicht empfehlenswert.

Für einen automatisierten Ansatz könnte man derartige Formulare (bzw. die lokalen Webseiten einzelner Hochschulen) parsen und mit den erhaltenen Daten eine eigene Bibliothek schreiben. So geschieht dies bspw. durch das Portal indeed; ein Beispiel dafür ist in Abb. 5 gezeigt. Bei einer Änderung der Darstellung entsteht jedoch ein verhältnismäßig hoher Aufwand, um den Parser anzupassen. Dadurch sind bisweilen die Angaben in derartigen Portalen unvollständig oder irreführend.

Sinnvoller sind deswegen automatisierte Anbindungen über ein echtes Application Programming Interface (API), welche von verschiedenen Portalen angeboten werden. Hierbei gibt es meist Ansätze mit XML- oder JSON-basierten Datensätzen. Diese werden je nach Portal per HTTP/FTP von einer bekannten Adresse heruntergeladen (PULL-basierte Ansätze) oder von der stellenausschreibenden Seite versandt (PUSH-basierte Ansätze). Die Unterschiede bei den Schnittstellen sind relativ gering. Unterschiede finden sich oft nur in Feinheiten. So wird beispielsweise die Angabe des relevanten Landes je nach Schnittstelle beispielsweise für Deutschland als String "National" oder als Länderkürzel "DE" angegeben.

Das Codebeispiel in Abb. 4 zeigt die Definition einer einheitlichen Datenstruktur (`LocalDataSet`) in der Middleware des Job Posting Helpers, die die nötigen Daten für die Portale bereitstellt. In dem Beispiel wird exemplarisch eine Funktion gezeigt, welche den Namen eines Exportmoduls und einer hochzuladenden Stellenanzeige entgegennimmt und die Daten der Stellenanzeige an das exportierende Modul übergibt.

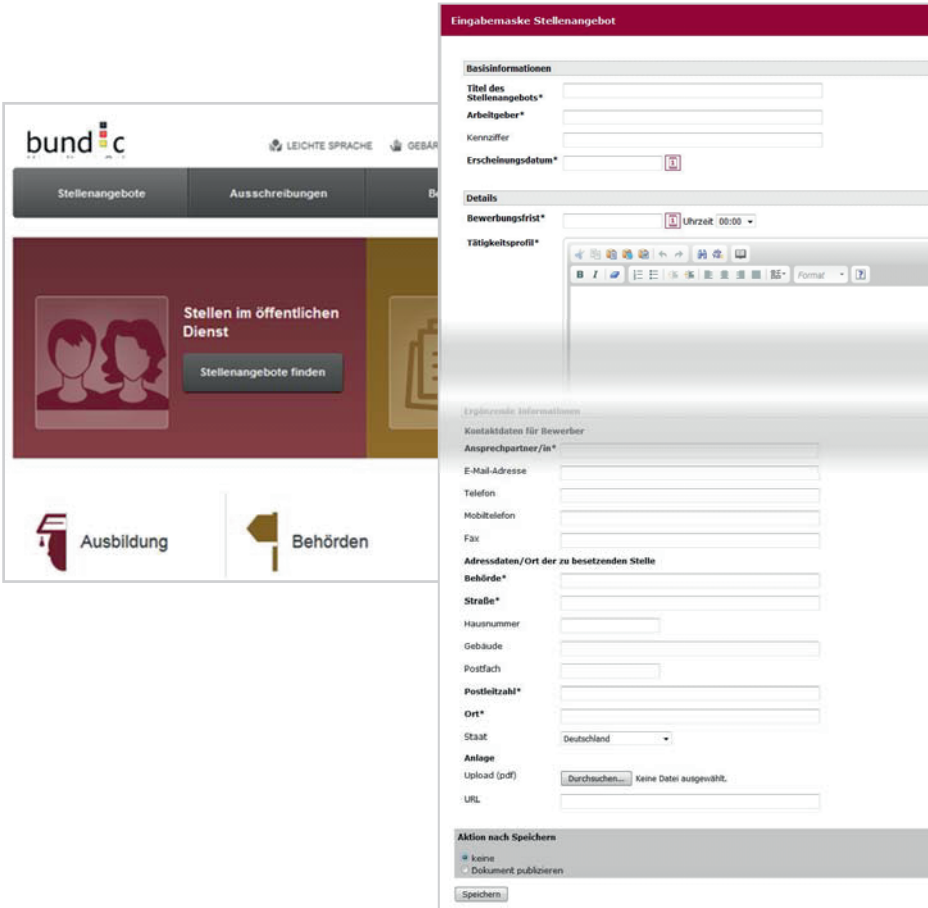


Abb. 3: Manuelle Veröffentlichung von Stellenausschreibungen im Web am Beispiel von Bund.de

```

public static boolean uploadNewJobPosting(String
moduleName, LocalDataSet dataSet) throws
ServiceNotFoundException {
    for (ExportModule module : getAllExportModules()) {
        if (module.getName().equals(moduleName)) {
            return module.uploadNewJobPosting(dataSet.getData());
        }
    }
    throw new ServiceNotFoundException();
}

```

Abb. 4: Export einer Stellenausschreibung (dataSet) über ein vorausgewähltes Modul

The screenshot shows the Indeed job portal interface. On the left is the University of Potsdam logo. The main content area displays a job listing for 'Academic Staff Member (PostDoc) at the Research Training Group Minor Cosmopolitanisms' at the University of Potsdam. The listing includes a search bar with 'Cosmopolitanisms' and 'Potsdam' entered, and a list of search filters. Below the search bar, there are several job listings, with the top one being the 'Academic Staff Member (PostDoc) at the Research Training Group Minor Cosmopolitanisms'. The job description for this position is visible, mentioning the Research Training Group Minor Cosmopolitanisms (Graduiertenkolleg) and the German Research Council (DFG). The job is for 12 positions starting on October 1st, 2016. The requisition number is 112/2016.

Abb. 5: Vergleich einer aktuellen Originalausschreibung¹⁸ und ihrer Darstellung bei indeed¹⁹

Zusätzlich gibt es einige über diesen gemeinsamen Nenner hinaus gehende Lösungen. Zum Beispiel sind bei dem Portal academics.de neben den sachlichen Angaben zu einer Stellenausschreibung auch Layout-Informationen in Form von Stylesheet-Klassen möglich, um eine Darstellung im jeweiligen Corporate Design zu erlauben. Dies setzt aber eine Kompatibilität der Stylesheet-Definitionen zwischen Job-Portal und Hochschule voraus, was im Regelfall die direkte Übernahme aus dem Web-System der Hochschule verbietet und stattdessen eine komplexe, explizite Anpassung an das Drittportal nötig macht. Diese Variante sollte daher nur bei einem Portal bzw. für Ausschreibungen von hoher strategischer Relevanz eingesetzt werden.

Eine andere Besonderheit ist der „Apply-with“-Button z.B. bei LindeIn oder Monster, der es Bewerbern ermöglicht eigene Profildaten direkt in das Bewerbungsformular eines potentiellen Arbeitgebers zu übernehmen. Diese eher seltene Unterstützung von Use Case 2 ist aufgrund des Zugriffs auf personenbezogene Daten deutlich aufwändiger umzusetzen als die oben beschriebenen Schnittstellen für Use Case 1.

Es gibt allerdings auch einige hinter diesem Entwicklungsstand zurück bleibende Möglichkeiten. So befinden sich z.B. Schnittstellen für Stellenausschreibungen derzeit noch in Entwicklung bei ResearchGate. Positiv ist hier aber anzumerken, dass dies durch ein abgestimmtes Vorgehen die tatsächliche Berücksichtigung des HR Open Standard erlauben könnte.

¹⁸ http://www.uni-potsdam.de/verwaltung/dezernat3/stellen/112_2016_RTG_12_PhD_neu.pdf

¹⁹ <http://de.indeed.com/Jobs?q=Cosmopolitanisms&l=Potsdam> (Abruf im Mai 2016)

Leider gab es auf unsere Anfragen bislang keine Rückmeldung von den jenseits der Akademia vermutlich populärsten Job Management Plattformen Xing, Monster und StepStone.

4 Fazit und Ausblick

Ziel der hier vorgestellten Arbeiten war es, die Personalbeschaffung an Hochschulen durch eine Kooperation mit Stellenbörsen und beruflichen Netzwerken zu befördern. Das erste Teilziel, die Stellenausschreibungen auf verschiedenen Stellenbörsen und in passenden beruflichen Netzwerken automatisiert und somit ohne manuellen Aufwand zu veröffentlichen, kann auf diese Weise dank der bereits zahlreich existierenden Schnittstellen der Drittanbieter erreicht werden. Zwar unterscheiden sich diese Schnittstellen geringfügig, sodass derzeit anstelle einer einheitlichen Lösung noch individuelle Plug-Ins nötig sind. Jedoch sind die Ähnlichkeiten so groß, dass der Aufwand zur Entwicklung eines neuen Plug-Ins vergleichsweise gering ist. Im Ergebnis ist eine große Entlastung in der Personalsachbearbeitung bei gleichzeitig größeren Chancen auf die Gewinnung guter Bewerber zu konstatieren.

Das zweite Teilziel, die Bewerbung eines Kandidaten durch automatisierte Übernahme seiner Profildaten aus einem beruflichen Netzwerk in die Online-Bewerbungsmaske der Hochschule zu vereinfachen, ist bislang nur bei einem Kooperationspartner möglich. Hier besteht noch deutlicher Ausbaubedarf, der aber aufgrund der Weitergabe personenbezogener Daten und der damit verbundenen Zustimmungs- bzw. Beteiligungspflichten ohnehin eine weitaus umsichtigerer Behandlung erfordert.

Die entwickelte Lösung ist derzeit auf die Personalverwaltungs-Software HIS-PM ausgerichtet. Eine Weiternutzung an anderen HIS-Hochschulen sollte deshalb problemlos möglich sein und ist sowohl mit der HIS eG als auch mit den kooperierenden Stellenbörsen und beruflichen Netzwerken ausdrücklich so konzipiert worden. Eine Übertragung auch für andere Ressourcen-Management-Systeme jenseits des spezifischen Hochschulumsfelds ist denkbar und im Architektur-Ansatz berücksichtigt, erfordert jedoch einen etwas größeren Aufwand. Durch die Kapselung als separates Modul und dessen Entwicklung in PHP erscheint die Übertragbarkeit auf andere Systeme, insbesondere wenn diese auf ähnlichen Architekturmustern oder Technologien aufbauen, jedoch als realistisch.

Wünschenswert wäre auf jeden Fall eine einheitliche Nutzung der Spezifikation im HR Open Standard sowohl durch Software zur Personalverwaltung (unabhängig vom Hochschulsektor) als auch für die Übernahme von Stellenausschreibungen durch Stellenbörsen und berufliche Netzwerke. Hier liegt derzeit noch einiges Integrationspotential unnötig brach.

Literaturverzeichnis

- [Gr 05] Grund, C.: Mitarbeiterrekrutierung über das Internet, Bonn econ discussion papers, No. 05/18, 2005.
- [HFD15] Hochschulforum Digitalisierung: 20 Thesen zur Digitalisierung der Hochschulbildung. Arbeitspapier Nr. 14, Stifterverband für die Deutsche Wissenschaft, Berlin, 2015.
- [Ka+13] Kaiser, S.; Kuhnt, E.; Lemcke, S.; Lucke, U.: Web-basierte Beschaffung. In Proc. INFORMATIK 2013, LNI P-220, Bonn : Köllen, 2013, S. 308-319.
- [Me12] Mertens, P.: Schwierigkeiten mit IT-Projekten der öffentlichen Verwaltung. Informatik-Spektrum, 35/06, S. 433-446, 2012.
- [Wi14] Wimmer, M.: Campus-Management - Anforderungen, Erfolgsfaktoren und der sichere Weg am Ziel vorbei, Campus Innovation, Hamburg, 2014
online unter <http://www.podcampus.de/nodes/pXNag>
- [ZKB12] Zakhariya, H. ; Kosch, L. ; Breitner, M. H. : Elektronische Drittmittelakte in der Hochschulverwaltung - Erkenntnisse aus Fallstudien. In Proc. 42. Jahrestagung der Gesellschaft für Informatik, LNI P-208, S. 613–626. Köllen, Bonn, 2012.

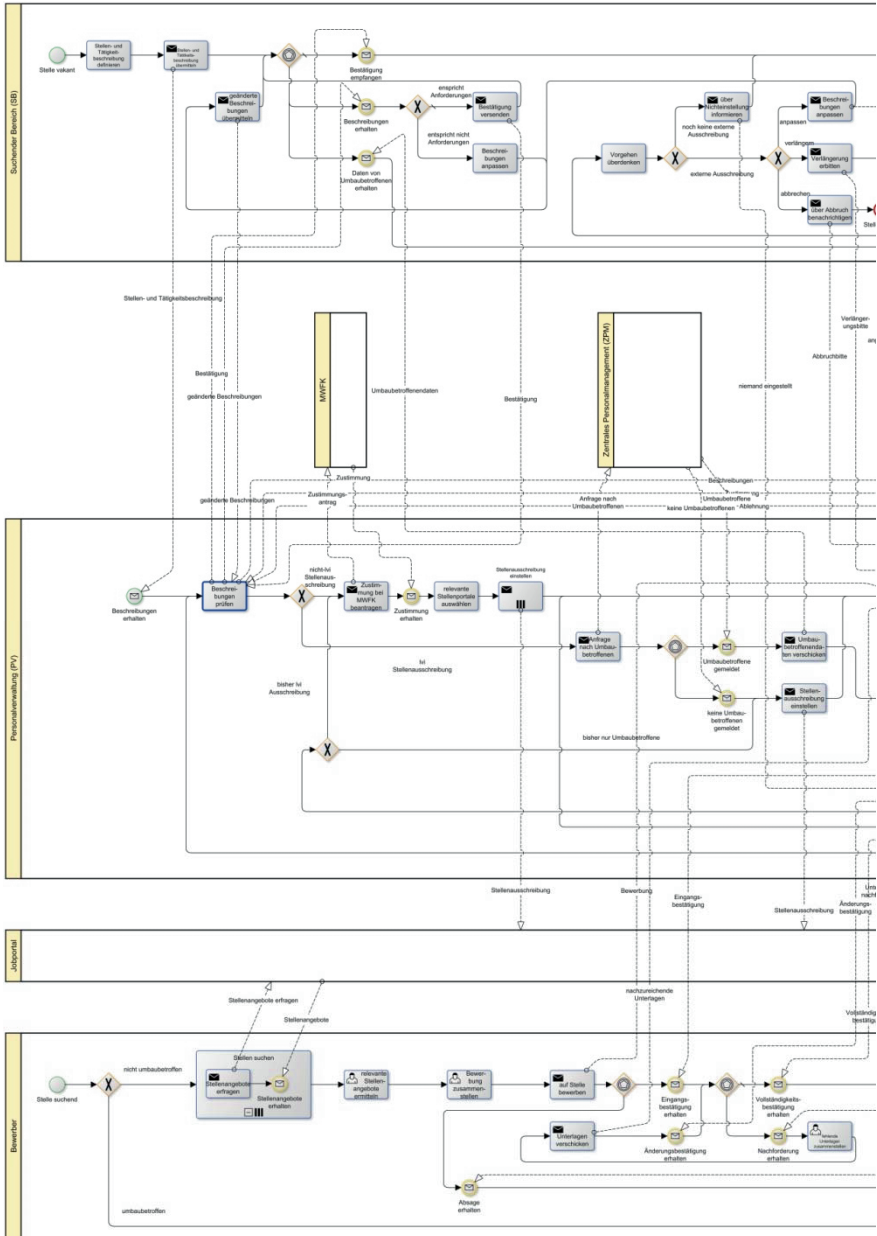
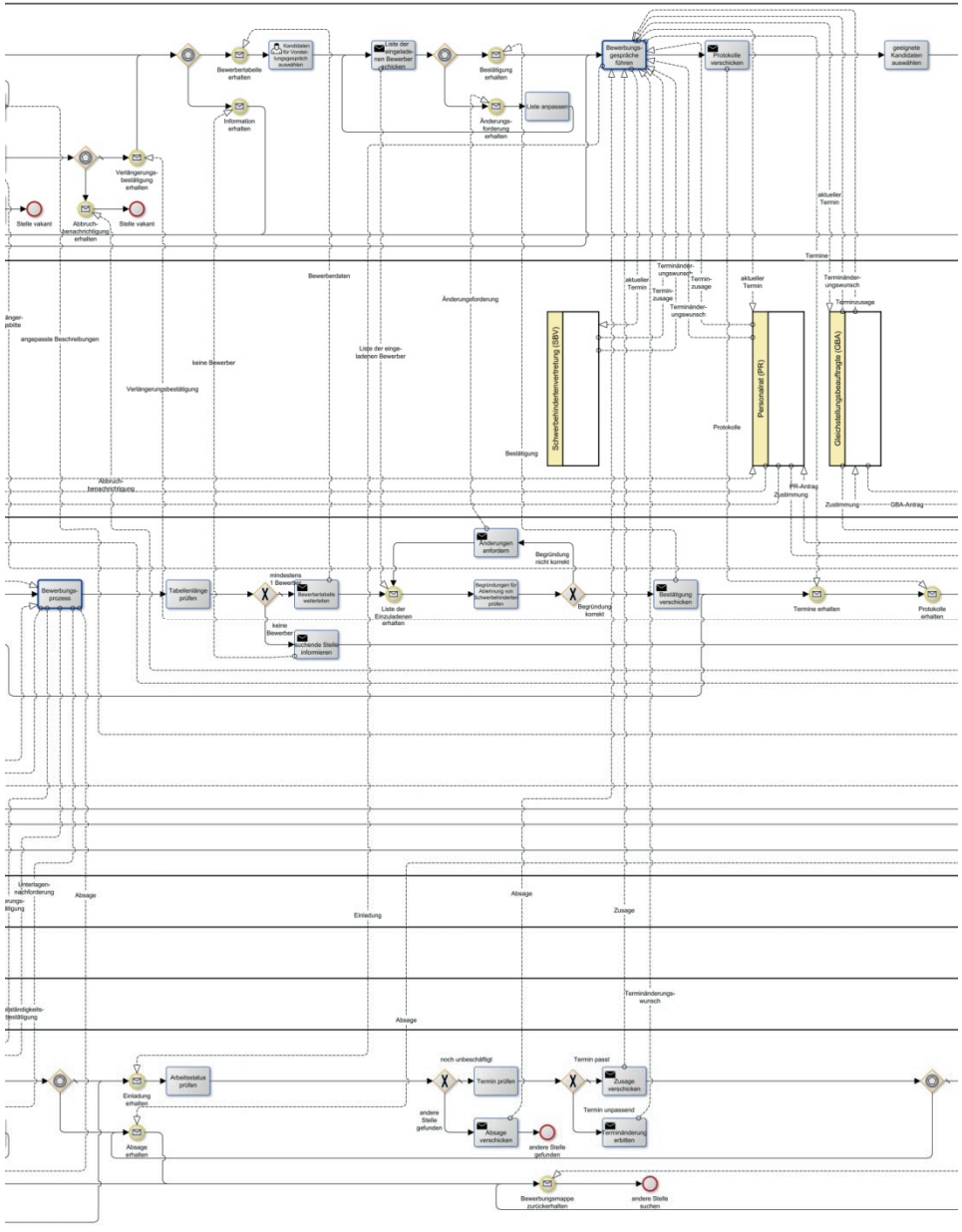
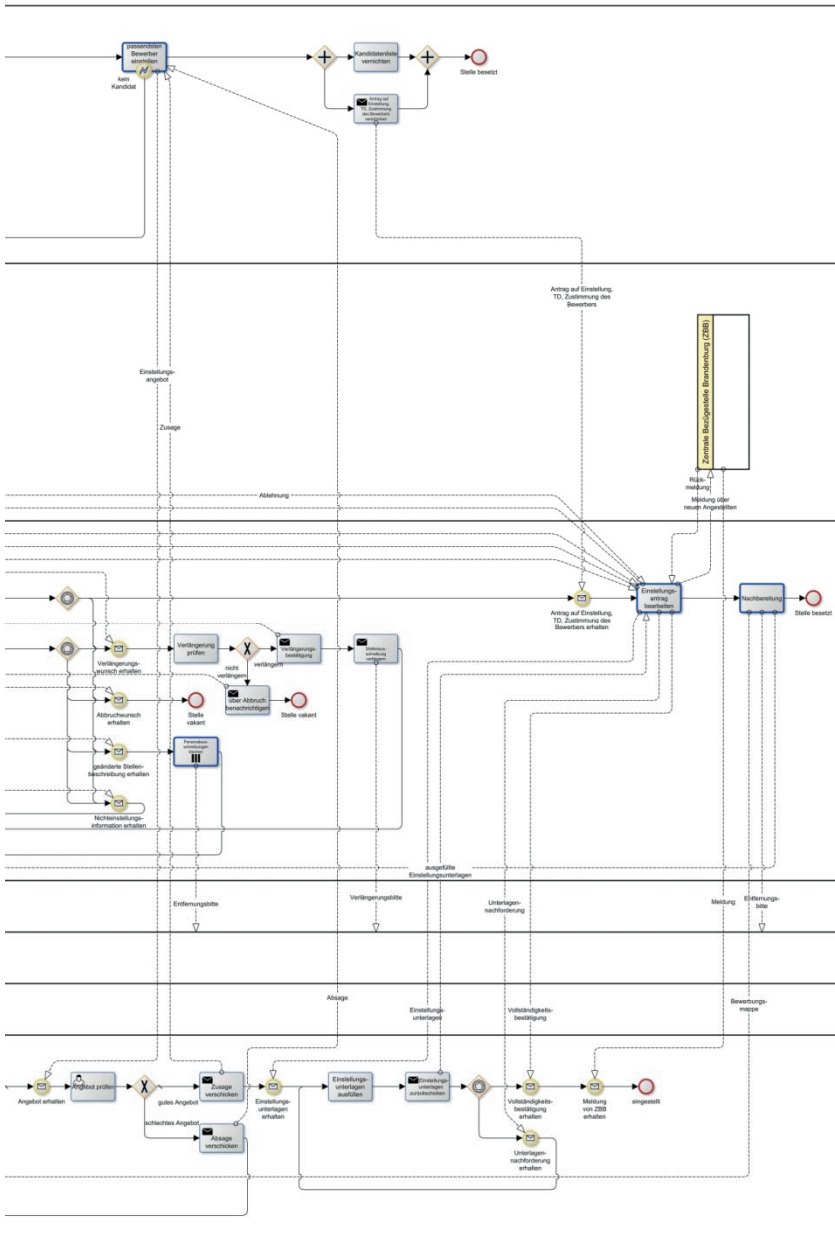


Abb. 6: Komplexität des Personalgewinnungsprozesses an Hochschulen

Das Prozessmodell ist zur besseren Lesbarkeit auf drei Seiten aufgeteilt, die nebeneinander gelegt den vollständigen Prozess ergeben.





Digitalisierung in Lehre und Lernen: des Kaisers neue Kleider?!

Digitalisierung in Lehre und Lernen: des Kaisers neue Kleider?!

Sarah Sahl¹, Alke Martens²

In den vergangenen Jahren konnten zahlreiche Initiativen und Konzepte für die Erprobung digital unterstützten Lehren und Lernens wahrgenommen werden und Begriffe wie E-Learning oder Blended Learning waren in aller Munde. Gegenwärtig redet man an vielen Punkten von Digitalisierung, weshalb man folgende Fragen stellen kann: Ist Digitalisierung das neue Kleid des E-Learnings? Ist E-Learning „tot“? Was versteht man unter Digitalisierung? Wird über die Digitalisierung in Lehre und Lernen geredet, dann können zum einen der Diskurs der technischen und didaktisch-methodischen Umsetzung digitaler Lehr-Lern-Angebote („Das neue Kleid des E-Learnings?“) und zum anderen die Digitalisierung als Bildungsgegenstand an sich identifiziert werden. Gerade in deren Verbindung ergeben sich für Forschung und Praxis verschiedene Herausforderungen, die fokussiert werden sollten: Digitale Lehr-Lern-Angebote sind nicht nur ein Thema der Informatik, sondern auch Forschungsgegenstand in anderen Fachdisziplinen, wie politische Bildung. Darüber hinaus öffnen sich die Angebote immer mehr verschiedenen Zielgruppen, unterschiedlichen Lernbiografien und Bildungswegen. Durch diese Interdisziplinarität und Öffnung der Angebote muss über eine veränderte Auseinandersetzung mit Bildungsprozessen nachgedacht werden und zugleich müssen auch bestehende Planungsprozesse (Top-Down- oder Bottom-Up-Entwicklung digitaler Lehr-Lern-Angebote) evaluiert werden.

In diesem Workshop werden verschiedene Beiträge für die Digitalisierung in Lehre und Lernen aufgegriffen und mit einem Blick auf die identifizierten Herausforderungen Grenzen und Auswirkungen diskutiert. Ein Schwerpunkt liegt hier auf der Vorbereitung und Durchführung von Prüfungen mit der Fragestellung, wie der Studierende in der individuellen Prüfungsvorbereitung durch individuelle und individualisierte E-Exams unterstützt werden kann. Ein anderer Beitrag blickt auf den gesamten Lehr-Lern-Prozess und fokussiert die Fragestellung, wie der Lernende bei der Aneignung wissenschaftlicher Konzepte unterstützt werden kann. Eine vorgestellte Herangehensweise ist das Topic Modelling mit dem Ziel, das Experten-Modell des Dozierenden und die Interaktion des Lernenden mit dem Lernstoff abzubilden. Die Entwicklung digitaler Lehr-Lern-Angebote durch Vorgehensmodelle unterstützt werden kann. Der Diskurs mit der technischen und didaktisch-methodischen Umsetzung digitaler Lehr-Lern-Angebote wird im Rahmen des Workshops mit einem Beitrag zur Gestaltung des Vorgehens bei

¹ Universität Rostock, Institut für Informatik, Lehrstuhl für Praktische Informatik, Albert-Einstein-Str.22, 18059 Rostock, sarah.sahl@uni-rostock.de

² Universität Rostock, Institut für Informatik, Lehrstuhl für Praktische Informatik, Albert-Einstein-Str.22, 18059 Rostock, alke.martens@uni-rostock.de

der Konzeption digitaler Lehr-Lern-Angebote aufgegriffen und im Rahmen analysierter Projekte diskutiert. Aus diesem Blick stellt sich die Frage nach der Unterstützung durch entsprechende IT-Infrastruktursysteme.

Vorgehensmodelle für die Konzeption digitaler Lehr-Lern-Angebote: Theoretisches Gerüst oder praktische Hilfe?

Sarah Freytag¹ Sarah Sahl² Birger Lantow³ Alke Martens⁴

Abstract: Die Digitalisierung in Studium und Lehre ist seit vielen Jahren ein für Hochschulen zu gestaltendes Thema. In einzelnen Projekten werden digitale Anteile in der Lehre etabliert, problematisch bleibt in vielen Fällen jedoch sowohl die Verstetigung der Projekte als auch die Unterstützung der Akteure bei der Konzeption und Implementierung digitaler Angebote. Mit dem vorliegenden Beitrag werden die Ergebnisse vier ausgewählter Projekte im Bereich E-Learning an der Universität Rostock auf die Prozessgestaltung bei der Konzeption der Angebote in den Blick genommen. Aufbauend auf bestehenden Vorgehensmodellen zur strukturierten Entwicklung von digitalen Lehr-Lern-Angeboten wird deren Berücksichtigung in der Praxis bewertet. Die Analyse greift dabei auf das ROME-Vorgehensmodell zurück, welches anhand der Analyse ausgewählter Projekte der Universität Rostock bewertet wird.

Keywords: Prozessmodelle, digitale Lehr-Lern-Angebote, E-Learning, Digitalisierung

1 Einleitung und Motivation

Die Einbindung neuer digitaler Lehr-Lern-Formate ist seit mehreren Jahren für Hochschulen ein Prozess, den es zu gestalten gilt. In der Vergangenheit gab es hierzu verschiedene Gestaltungsspielräume, beispielsweise durch neue, digitale Angebote, das selbstgesteuerte Lernen von Studierenden zu unterstützen oder durch digitale Angebote eine Öffnung der Hochschule auch für neue bzw. veränderte Zielgruppen zu ermöglichen. Die Möglichkeiten der Digitalisierung in Lehre und Lernen sind mit einem Blick auf die deutsche Hochschulbildung vielfältig. Auch die Universität Rostock hat in den vergangenen Jahren zahlreiche Projekte im Bereich E-Learning initiiert und umgesetzt. Mit der Digitalisierung hat sich aber auch der Blick auf Lehr-Lern-Formate verändert. Wurden in der Vergangenheit Lehr-Lern-Materialien als Produkt konzipiert und in der Lehre eingesetzt, verschieben die Möglichkeiten der Digitalisierung den Fokus vom Produkt zum Prozess: Lehr-Lern-Materialien können je nach Zielgruppe und Rahmen angepasst werden. [APW15]

¹ Universität Rostock, Institut für Informatik, Albert-Einstein-Strasse 22, 18059 Rostock, sarah.freytag@uni-rostock.de

² Universität Rostock, Institut für Informatik, Albert-Einstein-Strasse 22, 18059 Rostock, sarah.sahl@uni-rostock.de

³ Universität Rostock, Institut für Informatik, Albert-Einstein-Strasse 22, 18059 Rostock, birger.lantow@uni-rostock.de

⁴ Universität Rostock, Institut für Informatik, Albert-Einstein-Strasse 22, 18059 Rostock, alke.martens@uni-rostock.de

Diese Prozesshaftigkeit erfordert einen Blick auf die Prozesse zur Konzeption digitaler Lehr-Lern-Angebote. Mithilfe entsprechender Vorgehensmodelle kann den Akteuren eine Struktur für den Entwicklungsprozess vorgeschlagen werden, die eine Berücksichtigung aller wesentlichen Aspekte sicherstellt.

Die DIN PAS 1032-1:2004 [PA04] stellt im deutschsprachigen Raum ein solches Referenzmodell für die systematische Entwicklung von E-Learning-Angeboten dar. International bietet die ISO/IEC 19796-1:2005 eine Orientierung bei der Entwicklung von E-Learning-Angeboten. Basierend auf diesen Normen wurde das Rostocker Modell „ROME“ von Hambach et al. 2006 entwickelt [Ha08]. ROME wird als lineares Vorgehensmodell in die Phasen Analyse, Rahmenkonzept, Detailkonzept, Umsetzung, Einführung und Durchführung unterteilt, um ein digitales Lehr-Lern-Angebot zu konzipieren.⁴ In diesem Beitrag werden wir das Vorgehensmodell mit der praktischen Konzeption von digitalen Lehr-Lern-Angeboten abgleichen.

Im Rahmen des Projektes „Qualität garantieren: Professoren, Studierende und Dienstleister im Dialog für eine kompetenz- und forschungsorientierte Lehre (QualitätsDialog)“ hat die Universität Rostock 2012 im Wettbewerb „Studium Optimum“⁵ in zwei Antragsrunden universitätsweit nach Ideen zur Sicherung und Verbesserung von Studium, Lehre und Weiterbildung an der Universität Rostock gesucht. In den Jahren 2012 bis 2015 wurden 54 Projekte gefördert und begleitet, von denen 15 Projekte dem Themenbereich E-Learning/Blended Learning zugeordnet werden.⁶ Nach deren Förderende wurden zur Weiterführung der Projekte Clusterstrukturen etabliert, um die Projektergebnisse nachhaltig zu sichern und die Ergebnisse zu systematisieren. Doch auch neben der Förderung im Rahmen des Wettbewerbes „Studium Optimum“ existieren an der Universität Rostock Angebote und Konzepte zur Digitalisierung von Lehre und Lernen. Diese Projekte waren bisher nicht zentral erfasst und systematisiert worden.

Ausgehend von den bekannten Projekten wurde im ersten Schritt damit begonnen, die bisher nicht erfassten bestehenden Angebote zu identifizieren. Die Befragungen in den Fakultäten konnten uns weitere 41 Projekte und Angebote im Bereich E-Learning⁷ liefern. Um die Prozesse während der Erstellung der E-Learning-Angebote zu analysieren wurden aus den 41 Projekten vier Projekte aus verschiedenen Fakultäten ausgewählt (Elektronische Bibelkunde, Kugelwolkenmodell, MOOC, Blended Learning). Für die Analyse der ausgewählten Projekte wurden Experteninterviews durchgeführt. Mit deren Hilfe wurden unter anderem die Prozesse während der Entwicklung und Etablierung der E-Learning-Projekte, die Einbeziehung weiterer Stakeholder, die Finanzierung der Vorhaben, die genutzten Res-

⁴ Der sequentielle Ablauf dieses Modells wurde bereits bei [SM15] beschrieben.

⁵ Das Projekt QualitätsDialog wurde vom Bundesministerium für Bildung und Forschung gefördert.

⁶ Die geförderten Projekte sind im Rahmen der Cluster auf der Webseite <http://www.qualitaetsdialog.uni-rostock.de/studium-optimum/cluster-e-learning/cluster-projekte/> gelistet.

⁷ Im weiteren Verlauf wird der Begriff E-Learning als Oberbegriff für die Nutzung digitaler Medien in Lehre und Lernen genutzt und inkludiert in diesem Sinne auch Blended Learning Angebote.

sources und das didaktische Design des Angebotes erfragt. Mit dieser Datenbasis wird eine Analyse der Prozesse zur Konzeption der digitalen Lehr-Lern-Angebote vorgenommen.

2 Prozessanalysen für die Konzeption von digitalen Lehr-Lern-Angeboten

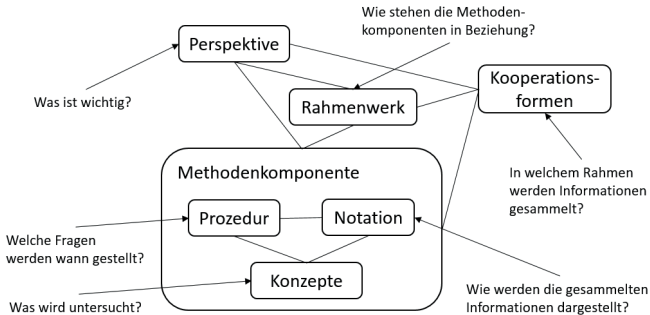


Abb. 1: Methodenmodell nach Goldkuhl et al. [GLS98]

Als Grundlage für die Analyse und den Vergleich der erhobenen Prozessmodelle mit dem ROME-Vorgehensmodell wird die Methodentheorie nach Goldkuhl et al. [GLS98] herangezogen. Die Analyse von Methoden nach Goldkuhl et al. hat sich schon an anderer Stelle bewährt (z.B. in [SK14]). Das Methodenmodell (siehe Abbildung 1) definiert die wesentlichen Aspekte zur Beschreibung einer Methode. Dabei gehen die Autoren davon aus, dass situationsabhängig unterschiedliche *Methodenkomponenten* mit jeweils eigenem Arbeitsgegenstand und Ziel zur Anwendung kommen. Wie diese *Methodenkomponenten* zur Anwendung kommen und in welchem Zusammenhang sie stehen, definiert das *Rahmenwerk* einer Methode. Diese Sichtweise unterstützt die Berücksichtigung der Interdisziplinarität und der Vielfalt von Entwicklungsprozessen für E-Learning-Angebote. Jede Methodenkomponente hat dabei jeweils eigene *Konzepte*, Abläufe (*Prozeduren*) und *Notationen*. Zusätzlich definiert eine Methode auch mögliche *Kooperationsformen* und wann sie zum Einsatz kommen. Ausgehend von einer Sicht auf ROME als Methode, lassen sich die analysierten Projekte zur Erstellung von E-Learning-Angeboten bezüglich der *Methodenkomponenten*, den *Kooperationsformen* und ihrer Kombination im *Rahmenwerk* betrachten. Die *Perspektive* ergibt sich aus der generellen Zielstellung der Einführung eines E-Learning-Angebotes. *Notationen* werden hier nicht betrachtet, da zum einen in ROME und zum anderen auch in den betrachteten Projekten keine konkreten Notationen definiert werden. *Methodenkomponenten* lassen sich anhand der jeweiligen Projektphasen (z.B. Analyse, Konzeptionierung, Umsetzung) und der dort jeweils betrachteten *Konzepte* (z.B. Lernziele, Inhalte, Lernkontrolle) identifizieren. Der Fokus liegt auf der Abfolge der Aktivitäten innerhalb der *Methodenkomponenten* (*Prozedur*) und zwischen ihnen (*Rahmenwerk*).

Ausgehend von algorithmischen Grundmustern gibt es folgende Möglichkeit der Anordnung von Aktivitäten: (1) Sequenz, (2) Parallelverarbeitung, (3) Alternative

und (4) Iteration (vgl. u.a. [LWS08]). Basierend auf diesen Grundmustern sollen im Folgenden die erhobenen Abläufe in den analysierten Projekten beschrieben werden. Eine Darstellung der Abläufe in BPMN bietet Abbildung 2.

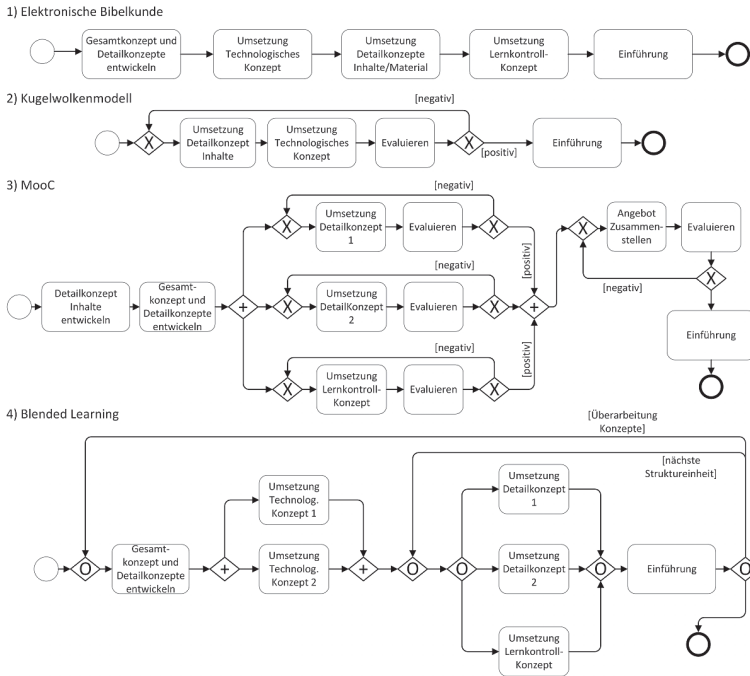


Abb. 2: Aktivitätenabfolge in den untersuchten Projekten

1) *Elektronische Bibelkunde* Das Projekt „Elektronische Bibelkunde“ zeichnet sich durch einen grundsätzlich sequentiellen Ablauf aus. Beginnend mit dem projektauslösenden Ereignis finden Workshops zur Erarbeitung des Gesamtkonzepts und der Detailkonzepte statt. Dann erfolgt die Auswahl der technischen Plattform, gefolgt von der Erarbeitung, Aufbereitung und Bereitstellung der Inhalte, der Umsetzung von Lernkontrollmechanismen und im letzten Schritt der Einführung des Angebots. Letzteres parallel über zwei Medien (CD und Lernsystem).

2) *Kugelwolkenmodell* Das Projekt zum Kugelwolkenmodell verzichtet auf einige Detailkonzepte. Es wird lediglich der darzustellende Inhalt definiert. Darauf folgt eine Iteration der Schritte Inhaltsrecherche, Aufbereitung der Inhalte und Evaluation des Ergebnisses. Nach erfolgreicher Evaluation findet die Einführung statt. Auch in diesem Projekt finden sich grundsätzlich keine parallelen Strukturen.

3) *Massive Open Online Course (MOOC)* Das betrachtete Projekt zur Erstellung eines MOOC zeichnet sich durch eine Parallelisierung der Umsetzung der entworfenen Konzepte aus. Nach der Entwicklung des Detailkonzepts für die Lehrinhalte, werden parallel das Konzept zur Lernkontrolle und die Konzepte zum Lehrmaterial (Video sowie Animation) umgesetzt. Am Ende der Umsetzung eines Kon-

zepts erfolgt eine Evaluation, die bei negativem Ausgang zu einer Iteration in der Konzeptumsetzung führt. Mit der Zusammenstellung der umgesetzten Konzepte erfolgt zunächst eine inhaltliche und dann nach der technischen Integration, eine technische Evaluation. Bei negativer inhaltlicher Evaluation wird nochmals zum Startpunkt der Konzeptumsetzung iteriert. Bei negativer technischer Evaluation zurück zur technischen Integration. Im letzten Schritt erfolgt die Einführung. Die technologische Umsetzung wird nicht in einer gesonderten Aktivität bearbeitet, da die entsprechende Plattform zur Bereitstellung der MOOCs bereits existiert.

4) *Blended Learning* Auch im untersuchten Projekt zum „Blended Learning“ zeigen sich parallel ausgeführte Aktivitäten. Nach der Konzeptentwicklung zur Lernkontrolle und zur Inhaltsdarstellung werden die technischen Voraussetzungen für die Konzepte parallel umgesetzt. Die inhaltliche Struktur wird nicht neu konzipiert, da sie durch den bereits bestehenden Kurs vorgegeben ist. Die weitere Umsetzung der Konzepte erfolgt iterativ der inhaltlichen Struktur des Kurses folgend. So werden innerhalb der Iteration Lernkontrolle und Inhaltsdarstellung für eine strukturelle Einheit (einzelne Vorlesung) umgesetzt und eingeführt. Auch hier findet sich eine parallele Umsetzung der Konzepte. Bei diesem Projekt kann ein Feedback aus der Nutzung des E-Learning-Angebots identifiziert werden. Mit dem Auftauchen von Problemen während des Kurses, erfolgt eine Iteration über die Erstellung von Konzepten und die Umsetzung der technischen Anforderungen.

Zusammenführung der Projektanalysen Insgesamt zeigt sich, dass echte Alternativen in einzelnen Prozessbeschreibungen nicht vorkamen. Die Möglichkeit der Parallelisierung von Aktivitäten wurde in einigen Projekten (3 und 4) genutzt. Wobei die Aktivität der Konzeptentwicklung, soweit sie mehrere Konzepte betrifft, auch in sich Parallelisierungspotential besitzt. Zusätzlich wurden Iterationen zur Qualitätskontrolle genutzt. Aktivitäten mussten bei einer negativen Evaluation wiederholt werden (Projekt 2, 3 und 4). Eine Evaluation nach der Einführung fand aber lediglich in Projekt 4 statt. Das ist wahrscheinlich auch dem Umstand geschuldet, dass hier mit der Erstellung des E-Learning-Angebots iterativ parallel die Nutzung erfolgte. Die anderen Projekte enden mit der Veröffentlichung. Ein rein sequentieller Verlauf war lediglich in Projekt 1 festzustellen. Es zeigt sich weiterhin, dass einzelne grundsätzlich zu entwickelnde und umzusetzende Konzepte bereits in entsprechender Reife vorlagen. In solchen Fällen konnte auf entsprechende Aktivitäten verzichtet werden. Mit Bezug auf die Methodentheorie ergeben sich daraus optionale Methodenkomponenten zur technologischen Konzeptionierung und Umsetzung sowie zur Konzeptionierung und Umsetzung der Detailkonzepte nach ROME. Projekt 2 verzichtet hier auch auf die Lernkontrolle, betrachtet aber im Vergleich zu den anderen Projekten lediglich eine inhaltliche Struktureinheit und muss in ein Gesamtkonzept eingebunden werden.

3 Vergleich der erhobenen Abfolgen mit den in ROME definierten

Im direkten Vergleich zwischen den Aktivitäten oder Funktionen, die in der Entwicklung der vier ausgewählten Projekte identifiziert wurden, und den Phasen und Phasenkomponenten des ROME-Modells gibt es einige Auffälligkeiten. Abbildung 3 stellt jeweils für die Projekte die Anzahl der den ROME-Phasen zuordenbaren Aktivitäten dar. Es zeigt sich, dass die meisten Phasenkomponenten zwar während der Entwicklung angesprochen werden, jedoch das Managen und die Evaluation der verschiedenen Phasen nur selten durchgeführt wird. Nur in zwei Projekten, dem Kugelwolkenmodell und den MOOCs, konnten solche Aktivitäten der Umsetzungsphase zugeordnet werden.

ROME-Phase		Elektr. Bibelk.	Kugelwolkenmodell	MOOC	Blended Learning
VS1 Analyse	VS1.1 Bildungsbedarf konkretisieren				
	VS1.2 Zielgruppe analysieren				
	VS1.3 Kontext analysieren				
	VS1.4 Phase managen und evaluieren				
VS2 Gesamtkonzept	VS2.1 Ziele festlegen				
	VS2.2 Didaktischen Ansatz entwickeln				
	VS2.3 Gesamtkonzept entwickeln				
	VS2.4 Phase managen und evaluieren				
VS3 Detailkonzepte	VS3.1 Lehr-/Lernziele festlegen				
	VS3.2 Detailkonzept Lehr-/Lerninhalte erstell.				
	VS3.3 Detailkonzept Lehr-/Lernformen erst.				
	VS3.4 Detailkonzept Lehr-/Lernmaterial erst.				
	VS3.5 Phase managen und evaluieren				
VS4 Umsetzung	VS4.1 Umsetzung planen				
	VS4.2 Medien- und multimediale Elemen. erst.				
	VS4.3 Lehr-/Lernmaterialien zusammenstellen				
	VS4.4 E-Learning-Angebot zusammenstellen				
	VS4.5 Phase managen und evaluieren				
VS5 Einführung	VS5.1 Einführung vorbereiten				
	VS5.2 E-Learning-Angebot technisch einführen				
	VS5.3 E-Learning-Angebot organisat. einf.				
	VS5.4 Phase managen und evaluieren				
VS6 Durchführung	VS6.1 Durchführung vorbereiten				
	VS6.2 E-Learning-Angebot durchführen				
	VS6.3 E-Learning-Angebot auswerten				
	VS6.4 Phase managen und evaluieren				
Farbcodierung	Keine Aktivität	1-2 Aktivitäten	3-4 Aktivitäten	Mehr als 4 Aktivitäten	

Abb. 3: Zuordnung der Prozessaktivitäten zu den ROME-Phasen

Der Vergleich zeigt, dass zwar die Aktivitäten eine ähnliche Reihenfolge haben wie im ROME-Modell, jedoch werden teilweise Komponenten gar nicht angesprochen oder Funktionen haben mehrere Zwecke. So wird z. B. die erste Aktivität der Elektronischen Bibelkunde genutzt, um sowohl die Analyse durchzuführen (Bildungsbedarf), das Gesamtkonzept zu entwickeln (Didaktischen Ansatz, Gesamtkonzept) und um die Detailkonzepte zu beschreiben (Lehr-/Lernziele, Detailkonzepte für

Lehr-/Lerninhalte, -formen und -material). Damit wurden drei der sechs ROME-Phasen bereits angesprochen, wobei aber einzelne Komponenten übergangen wurden. In diesem Fall kann das darauf zurückgeführt werden, dass z. B. die Zielgruppe, der Kontext und die Ziele des Projektes bereits von Beginn an feststanden. Die Evaluation und das Management von Analyse, Gesamt- und Detailkonzepten wurden dementsprechend vernachlässigt. Die darauf folgenden Funktionen in der Entwicklung der Elektronischen Bibelkunde sind fast alle der Phase der Umsetzung zuzuordnen. Dabei ist zu bemerken, dass auch mehrere Funktionen zu einer Komponente beitragen. Die letzten Funktionen betreffen die Bereitstellung des Angebotes für Studenten, wobei auch diese Phasen nicht evaluiert wurden. Dies kann der Art des Angebots geschuldet sein, da die Elektronische Bibelkunde auf der universitätsinternen Lernplattform ILIAS basiert und aus diesem Grund nicht ein weiteres mal eingeführt werden muss. An diesem Beispiel wird deutlich, dass der eigentliche Aufbau des Projektes zwar grundsätzlich das ROME-Modell widerspiegelt, jedoch gibt es keine expliziten Aktivitäten für die nach ROME definierten Phasen.

Das Kugelwolkenmodell entspricht noch weniger dem von ROME vorgeschlagenen Entwicklungsweg. Für dieses Projekt konnte nur ein minimaler Organisationsaufwand festgestellt werden, denn nur eine Person befasste sich mit der eigentlichen Entwicklung und dem Aufbau. Dies zeigt sich besonders darin, dass nur eine Aktivität in den ersten Phasen des ROME-Prinzips angesiedelt werden kann, und der Beginn der Entwicklung mit dem Festlegen der Ziele und der Entwicklung eines didaktischen Ansatzes einher ging. Für die Entwicklung von Detailkonzepten wurde keine Zeit aufgewandt, und es gab ebenfalls keine Evaluation. Die nachfolgenden Funktionen waren hauptsächlich auf die Sammlung von Informationen spezialisiert, die dann in einem Schritt in einem Programm zusammengefasst wurden. Auch bei den MOOCs, die nicht direkt über die Plattform ILIAS integriert werden, zeichnet sich ein ähnliches Bild. Die erste Funktion impliziert die Konkretisierung des Bildungsbedarfs und die Analyse der Zielgruppe, da beides durch die teilnehmende Fakultät begrenzt ist. Erst der vierte Schritt befasst sich mit den Zielen, dem didaktischen Ansatz und dem Gesamtkonzept, zusammen mit der Festlegung der Lehr- und Lernziele und den -formen. Dagegen wird die dritte Phase von ROME bereits im dritten Schritt bearbeitet, und später in weiteren Funktionen verfeinert. Die folgenden Aktivitäten sind wieder auf die Erstellung und Umsetzung des Angebotes gerichtet, wobei auch hier die Reihenfolge nicht exakt der in ROME definierten folgt. Allerdings wurde hier viel Wert auf das Managen und die Evaluation der Umsetzungsphase gelegt. Dies kann daran liegen, dass schon zuvor MOOCs entwickelt wurden und eine gewisse Professionalität bereits entstanden war. Die Aktivitäten des Blended Learning Projektes konzentrieren sich, wie die der anderen Projekte, größtenteils in der Umsetzung. Zusätzlich konnten in diesem Projekt keine Aktivitäten der Ein- oder Durchführung zugeordnet werden.

4 Zusammenfassung und Ausblick

In diesem Beitrag wurde anhand von ausgewählten Projekten aufgezeigt, wie die Konzeption digitaler Lehr-Lern-Angebote in der Praxis erfolgt. Ein besonderes Augenmerk lag bei der Analyse auf den Prozessen im konzeptionellen Vorgehen und deren Abbildung auf das Vorgehensmodell ROME. Dabei wurde deutlich, dass die Kritik an einem linearen Vorgehen bestätigt werden kann. Mit der Analyse wurde deutlich, dass einige Projekte durchaus andere Vorgehensschritte, als bei ROME beschrieben, wählen oder Iterationen bzw. Parallelisierungen im Vorgehen erkennbar werden. Weiterhin zeigte sich, dass es im Entwicklungsprozess Abweichungen von den in ROME beschriebenen Phasen gab. So wurden in einigen Projekten teilweise Phasen oder Aktivitäten ausgelassen, da sie durch Vorarbeiten oder weitere Aktivitäten schon berücksichtigt wurden. Zusätzlich wurde deutlich, dass der Schwerpunkt im Vorgehen in den Projekten auf unterschiedlichen Ebenen lag, da einige Projekte der technischen Realisierung eine größere Rolle zuwiesen. Die Analyse des Vorgehens zur Entwicklung digitaler Lehr-Lern-Angebote hebt zudem noch einmal hervor, dass der Fokus der Projekte in der Konzeption vor allem in der inhaltlichen Ebene liegt und hierbei das Vorgehen pädagogischer Planungsprozesse mitbedacht wird. Ausgehend von den Ergebnissen der Analyse bietet es sich an, eine Adaption des Vorgehensmodelles ROME weiterhin zu fokussieren und den linearen Prozess aufzubrechen und zukünftig die Struktur der Phasen zu überdenken.

Literaturverzeichnis

- [APW15] Arnold, Patricia; Prey, Gisela; Wortmann, Dennis: Digitalisierung von Hochschulbildung: E-Learning-Strategie (n) noch up to date? Zeitschrift für Hochschulentwicklung, 2015.
- [GLS98] Goldkuhl, Göran; Lind, Mikael; Seigerroth, Ulf: Method integration: the need for a learning perspective. In: Software, IEE Proceedings-. Jgg. 145. IET, S. 113–118, 1998.
- [Ha08] Hambach, Sybille: Systematische Entwicklung von E-Learning-Angeboten. Fraunhofer IRB-Verl, Stuttgart, 2008.
- [LWS08] Lehner, Franz; Wildner, Stephan; Scholz, Michael: Grundlagen der Informatik. In: Wirtschaftsinformatik, S. 119–176. Carl Hanser Verlag GmbH & Co. KG, 2008.
- [PA04] PAS 1032-1: , PAS 1032-1: Aus- und Weiterbildung unter besonderer Berücksichtigung von e-Learning - Teil 1: Referenzmodell für Qualitätsmanagement und Qualitätssicherung; Planung, Entwicklung, Durchführung und Evaluation von Bildungsprozessen und Bildungsangeboten, 2004.
- [SK14] Sandkuhl, Kurt; Koç, Hasan: Component-based method development: an experience report. In: The Practice of Enterprise Modeling, S. 164–178. Springer, 2014.
- [SM15] Sahl, Sarah; Martens, Alke: Process Models in E-Learning - Bottom-up or Top-Down? Cognition and Exploratory Learning in the Digital Age (CELDA 2015), S. 140, 2015.

Visualisierung konzeptuellen Lernens durch semantische Vernetzung sequenzieller Lehrinhalte

Robin Nicolay¹ Bastian Schwennigcke² Sarah Sahl³ Prof. Alke Martens⁴

Abstract: Akademische Ausbildung soll mehr als die reine Vermittlung von Lernstoff leisten. Das setzt voraus, dass die Lernenden theoriegeleitet arbeiten und die dabei maßgeblichen Konzepte überblicken und anwenden. Wie Lernende die Aneignung wissenschaftlicher Konzepte bewältigen und welche Herausforderungen sich ihnen dabei stellen, ist Teil der hochschuldidaktischen sowie der lerntheoretischen Forschung. Dieser Beitrag beschreibt Herausforderungen bei der Konzeptvermittlung durch Medienrezeption (Printmedien und Vorträge), sowie Aspekte des Konzepterwerbs durch Exploration und Explikation. Aus Sequenzen von Lehrstoff werden Experten-Modelle abgeleitet, die durch Themen und semantische Verbindungen repräsentiert sind. Auf dieser Basis wird ein Ansatz zur Messbarkeit der konzeptuellen Ordnung des Lehrinputs sowie der Differenzen mit den konzeptuellen Rekonstruktionsleistungen Lernender beschrieben. Hierbei kommen Verfahren der Latent-Dirichlet-Allocation und des Topic Modeling zum Einsatz.

Keywords: konzeptuelles Lernen, Konzeptrekonstruktion, Latent-Dirichlet-Allocation, Wissenstransfer, Lehrformen

1 Einleitung und Motivation

Akademische Ausbildung zielt auf den Erwerb konzeptueller Kompetenz. Bereits mit Absolvierung des ersten Studienzyklus sieht der Europäische Qualifikationsrahmen vor, dass Lernende „komplexe und unvorhersehbare Probleme [...] in einem spezialisierten Arbeits- oder Lernbereich“ bearbeiten können und dass sie dabei auf ein „kritische[s] Verständni[s] von Theorien und Grundsätzen“, die innerhalb dieses Arbeitsbereiches einschlägig sind, zurückgreifen. [PR08, Anhang II, C111/5, Niveau 6]

Der Ansatz konzeptueller Kompetenz geht über die Vermittlung und Abfrage von Lernstoffen hinaus. Es wird eine intellektuelle Entwicklung angestrebt [La12, S. 33-36], in deren Verlauf Lernende die Fähigkeit erwerben, Strategien und Lösungsansätze „durch Wahlentscheidungen“ zu motivieren, „bei denen gültiges und signifikantes Wissen eine zunehmende Rolle spielt.“ [Ku10, S. 14] Ein solches Wissen wird

¹ Universität Rostock, Institut für Informatik, Albert-Einstein-Strasse 22, 18059 Rostock, robin.nicolay@uni-rostock.de

² Universität Rostock, Institut für Informatik, Albert-Einstein-Strasse 22, 18059 Rostock, bastian.schwennigcke@uni-rostock.de

³ Universität Rostock, Institut für Informatik, Albert-Einstein-Strasse 22, 18059 Rostock, sarah.sahl@uni-rostock.de

⁴ Universität Rostock, Institut für Informatik, Albert-Einstein-Strasse 22, 18059 Rostock, alke.martens@uni-rostock.de

durch den konzeptuellen Haushalt eines Faches repräsentiert. Zu klären ist, wie die theoretischen und methodischen Konzepte eines Faches Eingang in die „Handlungs-, Operations- und Begriffsschemata“ [OR08, S. 26] findet, die Lernenden bei der Bearbeitung von Problemstellungen zur Verfügung stehen.

Dieser Vorgang ist nicht unmittelbar beobachtbar, was die Einschätzung der Wirksamkeit didaktischer Interventionen erheblich beeinträchtigt. Wichtig ist die Frage, welche Rolle Wissen bei der Auseinandersetzung mit Problemstellungen in Lernumwelten für Lernende spielt. Laurillard unterscheidet [La12]...

1. eine dualistische Wissensauffassung, die Aussagen nach dem richtig-falsch-Schema ordnet und die entsprechende Bewertung von einer Autorität bezieht,
2. eine offene Wissensauffassung, die multiple Perspektiven auf einen Gegenstand zulässt, diese aber als gleichwertig behandelt
3. und eine relativistische Wissensauffassung, die auf Positionierungen in Abgrenzung zu anderen möglichen Positionen im Umgang mit einem Gegenstand beruht.

Geprüft werden kann, ob Veränderungen im Umgang mit Wissen Auswirkungen auf die Rezeption von Wissensmedien haben. Erwartbar ist, dass ein zunehmendes Bewusstsein für Perspektivbildungen und konzeptuelle Differenzierung den sequentiellen Nachvollzug des Narrativs medialer Darstellungen in den Hintergrund treten lässt. Die Rekonstruktion semantischer, assoziativer und argumentativer Relationen gewinnt an Bedeutung. Offen ist, welche Entwicklungsschritte die Lernenden dabei machen, welche typischen Muster der Integration bzw. Ausgrenzung von Inhalten sich finden lassen, wie sich Interventionen auf Ebene der Lernumwelt auswirken und wie inhaltliche Verarbeitungsleistungen aussehen, welche mit den Intentionen der Lehrpersonen (nicht) konvergieren.

2 Technischer Ansatz: Topic Modelling

Erforscht wird, welche semantische Struktur sequenzierten Lehrinhalten wie Vorträgen und durch Lehrpläne verknüpfte Textmedien zugrunde liegt. Ziel ist eine Approximation des zugrundeliegenden Experten-Modells des Lehrenden und eine Abbildung der Interaktionen des Lernenden mit Lehrstoff auf diesem Model. Hierdurch soll die Art der studentischen Verarbeitung und Exploration von Themen und Relationen sichtbar werden. Der folgende Abschnitt beschreibt unseren Ansatz zur Inferenz eines semantischen Netzwerks aus Sequenzen von Vorlesungsmaterialien. Das semantische Netzwerk 1 versieht Sequenzen von Lehrinhalten mit einer Metastruktur über relevante Themen und deren Relationen.

Für die Inferenz einer Metastruktur auf Sequenzen von Lehrinhalten müssen sowohl relevante Themen, als auch semantische Relationen abgeleitet werden. Die folgenden drei Abschnitte beschreiben das LDA-Verfahren vorgestellt von David Blei

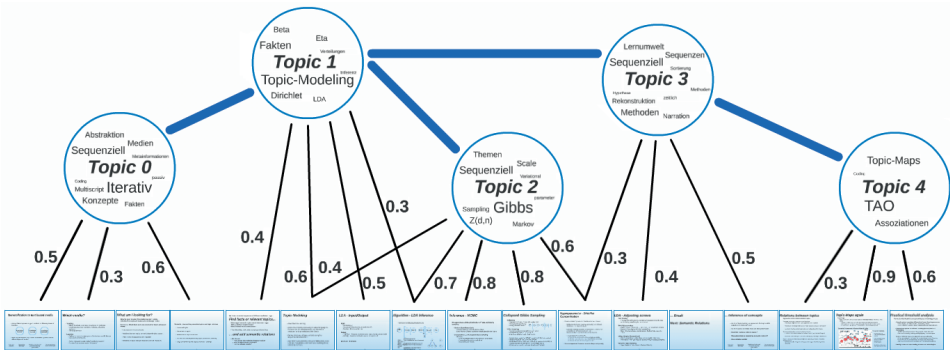


Abb. 1: Vereinfachte Darstellung eines inferierten, semantischen Netzwerks aus Themen und Themenrelationen in Form einer Topic-Map[MMM08]. Occurrences sind eine Sequenzen aus Vortragsfolien. Themen sind definiert durch "Wordclouds"relevanter Begriffe und miteinander durch Lehrverlauf und gemeinsame Occurrences verbunden.

in [BNJ03] und ein daraus entwickeltes Werkzeug als Ansatz zur Inferenz dieser Themen und Relationen.

2.1 Inferenz von Themen aus Lehrinhalten

Zur Ableitung von Themen wurden Sequenzen von Textinhalten durch Verfahren des *Topic-Modelling* analysiert. *Topic-Models* sind statistische Modelle. Sie finden unter anderem Verwendung in der Inferenz abstrakter Themen aus natürlich sprachlichen Textdokumenten. Für den Ansatz dieser Arbeit verwenden wir das Verfahren der "Latent-Dirichlet-Allocation"(kurz LDA). LDA ist eine Weiterentwicklung des "Probabilistic Latent Semantic Indexing"(kurz PLSI) beschrieben in [Ho99]. Ein Hauptmerkmal hierbei sind die hinzugefügten *Dirichlet-Priors*, deren Vorzüge in im Anschluss diskutiert werden.

Zusammengefasst inferiert LDA eine festgelegte Anzahl von Themen $T = t_1, t_2, \dots, t_k$ aus einer Menge von Dokumenten D mit Gesamtvokabular V . Hierbei liefert der Algorithmus eine diskrete Verteilung $\theta(d)$ über das Vorkommen aller Themen $t \in T$ für jedes Dokument $d \in D$. Die Summe der Wahrscheinlichkeiten des Vorkommens aller Themen t in Dokument d ergibt immer 1. Jedes Thema t selbst ist durch eine diskrete Verteilung $\beta(t)$ über die Wahrscheinlichkeit des Vorkommens jedes Wortes $w \in V$ definiert. Auch hier ist die Summe aller Wahrscheinlichkeiten des Vorkommens jedes Wortes $w \in V = 1$

Vortragsfolien arbeiten mit der Anreicherung von Inhalten und besitzen für eine Inferenz nur geringe Mengen von Text. Eine Untersuchung des *Dirichlet-Priors* im LDA-Algorithmus bietet in diesem Kontext viele Vorteile. Die Vorteile des LDA-Priors selbst beschreibt [WMM09] unter anderem, als Möglichkeit zur Anpassung der Inferenz durch Meta-Informationen und eine verbesserte Verarbeitung geringer

Textmengen. Weiterhin unterstützen Priors die Beobachtung von Themenentwicklungen über die Zeit [BL06]. Ein *Dirichlet-Prior* wird durch einen Konzentrationsvektor $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$ mit $k = |T|$ für die Verteilung von Themen auf Dokumente und $k = |V|$ für die Verteilung der Wörter auf Themen festgelegt. Für gleiche $(\alpha_1, \alpha_2, \dots, \alpha_k)$ spricht man von symmetrischen, bei ungleichen von asymmetrischen *Dirichlet-Priors*. Verschiedene Ausprägungen des Konzentrations-Parameters haben folgenden, am Beispiel der Themenverteilung auf Dokumenten $d \in D$, beschriebenen Einfluss auf die LDA-Inferenz.

1. Symmetrisch, $\alpha < 1$: Wenige Themen mit hoher Wahrscheinlichkeit
2. Symmetrisch, $\alpha > 1$: Viele Themen mit ähnlich hoher Wahrscheinlichkeit
3. Asymmetrisch, hohe $u_{x,y,z}$: Themen $t_{x,y,z}$ mit erhöhter Wahrscheinlichkeit

Während sich asymmetrische Priors zur Anpassung der Verteilung von Themen auf Dokumente eignen, reduzieren Fall

2.2 Überführung des technischen Ansatzes: Inferenz von Vortragsthemen

Für erste Versuche zur Inferenz von Themen und abgeleiteten Relationen wurden Vorlesungsfolien aus dem Studiengang Informatik an der Universität Rostock untersucht. Erste Resultate werden am praktische Beispiel der Einführungsveranstaltung zu Webtechnologien mit 49 verwendeten Folien vorgestellt. Für die Durchführung erster Tests wurde eine Python-Implementation [ld15][Da11] des Markov-Chain-Monte-Carlo-Verfahren Gibbs-Samplings[GS04] verwendet. Die Inferenz wurde für 10 Themen mit 1500 Iterationen, und mit symmetrischem Themen-Prior $\alpha = 0.1$ und symmetrischem Wort-Prior $\alpha = 0.01$ (heuristisch für Texte nach [Zh08] optimiert) durchgeführt. Die 49 Vortragsfolien hatten ein Gesamt-Vokabular von $|V| = 840$ unterschiedlichen Wörtern. Zur Schärfung des Vokabulars wurden sukzessiv folgende Filter angewendet.

1. Unifizieren von Groß- und Kleinschreibung ($|V| = 797$)
2. Entfernen von Stoppwörtern [Po15] ($|V| = 726$)
3. Entfernen grammatikalischer Verformungen [Yo15] ($|V| = 691$)
4. Entfernen von Wörtern mit einer Länge kleiner 3 und Zahlen ($|V| = 632$)
5. Entfernen regelmäßig vorkommender Wörter (Beispiel: Fußzeilen) ($|V| = 632$)
6. Entfernen von Bild-Folien ohne Text

Anschließend wurden 10 Themen Berechnet. Im Folgenden ist exemplarisch eine Auflistung einiger Themen mit ihren 5 wahrscheinlichsten Wörtern in absteigender Reihenfolge aufgelistet.

Thema 1	Thema 2	Thema 3	Thema 4	Thema 5
web	business	media	dom	user
technolog	anwend	content	css	neu
konzept	entwickl	honeycomb	java	ajax
vorles	dollar	social	xml	serv
warum	million	user	konzept	respons

Tab. 1: Top 5 Wortstämme inferierter Themen

Tabelle 1 zeigt 5 der 10 inferierten Themen. Je Thema sind die 5 häufigsten Wörter der diskreten Verteilungen über $|V|=632$ Wörter je Thema gelistet. Bereits anhand dieser Wörter sind thematische Schwerpunkte und ein thematischer Verlauf erkennbar. Die Veranstaltung begann mit der Beschreibung der Vorlesungsinhalte, Konzepte und Technologien, über eine Motivation zu späteren Berufszielen und praktischen Fragen des Alltags zu ersten fachlichen Details über Webseiten und Kommunikationstechniken.

2.3 Inferenz gemeinsamer Kontexte

In einem ersten Ansatz wurden gemeinsame Kontexte und damit Relationen zwischen Folien und Themen untersucht. Hierfür wurden Verbindungen zwischen einzelnen Vorlesungsfolien gebildet, die sich ein gemeinsames Thema teilen. Dazu wurden nur Themen in Betracht gezogen, die mit einer Wahrscheinlichkeit von über 65% (manuell festgelegt) für die jeweilige Folie erkannt wurden. Die abschließende Abbildung 2 zeigt nun Verbindungen zwischen Folien, die sich mindestens eines der 10 erkannten Themen teilen. Bereits hier ist zu erkennen, dass aufeinanderfolgende Folien durch gemeinsame Themen verknüpft sind. Teilweise werden im späteren Verlauf der Vorlesung Themen vom Anfang der Vorlesung wieder aufgegriffen. Für einige Folien wurden keine Themen mit einer ausreichend hohen Wahrscheinlichkeit erkannt. Diese besitzen keine Verbindungen.

2.4 Beobachtung studentischer Interaktionen mit Lehrinhalten

Zur Beobachtung der studentischen Verarbeitung von Lehrinhalten kommen verschiedene Methoden in Frage. In voran gegangenen Arbeiten wurde hierzu eine Möglichkeit der Beobachtungen auf Basis von "Coding" [LLG08] und der Assoziation neuer Stimuli mit bekanntem Wissen untersucht [Ni15a] [Ni15b]. Im Fokus stand hierbei die Beobachtung der studentischen Rekonstruktion von Themen und semantischen Relationen im Rahmen der Vorlesungsveranstaltung. Hierzu wurden aktuelle Vortragsfolien während der Vorlesung mit Studenten geteilt und beobachtet, ob eine Annotation durch Markieren von Stichwörtern, Verknüpfen von Folien und der Formulierung von Konflikten, Rückschlüsse auf Lernprozesse erlauben. Erste Untersuchungen zeigten, dass Studenten in Vorlesungssituationen Informationen



Abb. 2: Chorddiagram über die Relationen zwischen einzelnen Vorlesungsfolien einer Vorlesung. Die Vorlesung beginnt oben mittig und schreiten im Uhrzeigersinn voran.

nicht intuitiv verknüpfen. Aktuell werden diese Untersuchungen auf die Beobachtung von Browsingverhalten durch E-Learning Inhalte und Nachbereitung durch Studenten erweitert.

Ein Ähnlicher Ansatz zur Beobachtung der studentischen Exploration von Wissen unter Verwendung des LDA-Verfahrens wurde durch [PK13] untersucht. Hierbei wurden Webseiten entsprechend ihrer Themen durch Experten klassifiziert und mit Hilfe von LDA relevante Klassen-Begriffe extrahiert. Anschließend wurde durch die Beobachtung des Browsingverhaltens von Studenten bei der Exploration von Wissen, beim Besuch passender Webseiten ein Lernzuwachs für entsprechende Themen berechnet. Während bei [PK13] der Lernerfolg von Themen betrachtet wurde, untersuchen wir die Inferenz eines Experten-Modells aus sequenzierten Lehrmaterialien. Hierbei steht die Entwicklung von Relationen zwischen Themen im Verlauf des Lehrens, sowie die Bewegung von Studenten auf diesem in Abbildung 1 skizzierten Netzwerk während des Lernens im Vordergrund.

3 Zusammenfassung und Ausblick

Im vorliegenden Beitrag wurden die technischen Vorüberlegungen zur Anwendung des LDA als Verfahren des Topic Modelling als Möglichkeit zur Ableitung semantischer Relationen aus Lerninhalten vorgestellt und im zweiten Schritt eine Umsetzung anhand von Vorlesungsfolien für die Inferenz relevanter Begriffe umgesetzt. Dabei wurde aufgezeigt, dass das LDA-Verfahren nicht nur für eine Inferenz von relevanten Begriffen auf Webseiten genutzt werden kann, sondern auch eine Analyse der Themen von Vorlesungsfolien möglich ist. Die inferierten Themen wurden zudem in einem ersten Ansatz mittels Chorddiagramm in Relationen gebracht. Im nächsten Schritt soll nun geprüft werden, inwieweit abgeleitete Themen und Themenstrukturen mit den Intentionen von Dozierenden und Lehrveranstaltungen korrelieren. Darüber hinaus soll geprüft werden, inwieweit Vortragsfolien über Meta-Informationen verfügen, die zur Schärfung der Inferenz über Dirichlet-Priors und Textfilter genutzt werden können.

Für die weitere Arbeit ergeben sich verschiedene Fokussierungsschwerpunkte und Fragestellungen. Zum Einen gilt es, die Überlegungen in den Lehr-Lernkontext einzubetten. Mit der Betrachtung der Korrelation von Themenstruktur, Fachwissen und Intention des Dozierenden kann eine Aufmerksamkeit für die inhaltliche Strukturierung von Lehrinhalten angeregt werden. Dazu braucht es auch einen Rückbezug auf bereits vorhandene Studien mit analogen (Visualisierungs-) Methoden, die die Anreicherung konzeptuellen Wissens untersucht haben. Der Gewinn und daraus resultierende Effekte für den Lernprozess sind weitere Fragestellungen. Auf der anderen Seite ist ein empirisches Design, das für die identifizierten Fragestellungen herangezogen werden kann, zu entwickeln. Aus gegenwärtiger Sicht ist eine Interventionsstudie realisierbar, die analoge Methoden und die vorgeschlagene digitale Untersuchung konzeptuellen Wissens miteinander verbindet. Auf diese Weise sollen Differenzen zwischen dem maschinellen Abstraktionsweg und den analogen Strukturierungsleistungen Lernender und Dozierender sichtbar und validiert werden. Dies ermöglicht in weiteren Arbeiten eine Fokussierung auf die Identifizierung von Schwellen im Lernprozess, mit dem Potential die Fragestellung geeigneter Unterstützungen im Lehr-Lern-Prozess zu untersuchen.

Literaturverzeichnis

- [BL06] Blei, David M.; Lafferty, John D.: Dynamic topic models. In: Proceedings of the 23rd international conference on Machine learning. S. 113–120, 2006.
- [BNJ03] Blei, David M.; Ng, Andrew Y.; Jordan, Michael I.: Latent dirichlet allocation. the Journal of machine Learning research, 3:993–1022, 2003.
- [Da11] Darling, W. M.: A Theoretical and Practical Implementation Tutorial on Topic Modeling and Gibbs Sampling. In: In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. 2011.
- [GS04] Griffiths, T. L.; Steyvers, M.: Finding scientific topics. Proceedings of the National Academy of Sciences, 101(Supplement 1):5228–5235, 2004.

- [Ho99] Hofmann, Thomas: Probabilistic Latent Semantic Indexing. In: Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '99, ACM, New York, NY, USA, S. 50–57, 1999.
- [Ku10] Kurtz, Thomas: Der Kompetenzbegriff in der Soziologie. In (Kurtz, Thomas; Pfadenhauer, Michaela, Hrsg.): Soziologie der Kompetenz SE - 1, S. 7–25. VS Verlag für Sozialwissenschaften, Wiesbaden, 2010.
- [La12] Laurillard, Diana: Teaching as a design science: Building pedagogical patterns for learning and technology. Routledge, New York, London, 2012.
- [ld15] lda developers: LDA 1.0.3: Python Package Index: <https://perma.cc/ZJ8W-JR33>. 2015.
- [LLG08] Lee, Hyeon Woo; Lim, Kyu Yon; Grabowski, Barbara L.: Generative Learning: Principles and Implications for Making Meaning. Handbook of research on educational communications and technology, 3:111–124, 2008.
- [MMM08] Marius, Lars; Moore, Garshol; Moore, Graham: ISO 13250-2: Topic Maps — Data Model: <http://perma.cc/MPT4-MVLM>. 2008.
- [Ni15a] Nicolay, Robin; Schwennigcke, Bastian; Sahl, Sarah; Martens, Alke: Learner-Content-Interface as an Approach for self-reliant and student-centered Learning. In: Proceedings of the IADIS International Conference Cognition and Exploratory Learning in Digital Age, S. 327–331. 2015.
- [Ni15b] Nicolay, Robin; Schwennigcke, Bastian; Vetterick, Jonas; Sucharowski, Wolfgang; H. Cap, Clemens: InterLect - Lecture Content Interface. In: 7th International Conference on Computer Supported Education. S. 269–276, 2015.
- [OR08] Oelkers, Jürgen; Reusser, Kurt: Qualität entwickeln – Standards sichern – mit Differenz umgehen (Expertise im Auftrag des BMBF). Bildungsforschung, 27. BMBF, Berlin, 2008.
- [PK13] Pirolli, Peter; Kairam, Sanjay: A knowledge-tracing model of learning from a social tagging system. User Modeling and User-Adapted Interaction, 23(2-3):139–168, 2013.
- [Po15] Porter et al.: Natural Language Toolkit - NLTK 3.0 documentation: Stopwords Corpus: <https://perma.cc/F8L8-KC96>. 2015.
- [PR08] Parlament, Europäisches; Rat, Europäischer: Empfehlungen des Europäischen Parlaments und des Rates zur Einrichtung des Europäischen Qualifikationsrahmens für lebenslanges Lernen. apr 2008.
- [WMM09] Wallach, Hanna M.; Mimno, David M.; McCallum, Andrew: Rethinking LDA: Why priors matter. In: Advances in neural information processing systems. S. 1973–1981, 2009.
- [Yo15] Yoshiki Shibukawa: Snowballstemmer 1.2.1: <https://perma.cc/XPJ6-JNMF>. 2015.
- [Zh08] Zhao, Weizhong: Best Practices in Building Topic Models with LDA for Mining Regulatory Textual Documents: NCTR CTP Working Group: <https://perma.cc/8NGH-GYAL>. 2008.

Shifting the Process of Exam Preparation Towards Active Learning: A Crowdsourcing Based Approach

Agnes Koschmider¹, Detlef Buschfeld²

Abstract: Students are usually only marginally involved in the preparation of exams. They are passive peers who write an exam predefined by the lecturer. Studies in educational theory indicate, however, that active involvement instead of (passive) consumption should be adopted in learning since active learning proved to be superior to passive learning. One concept of active learning is student peer evaluation where students are mutually responsible for peer reviewing. This paper extends student peer evaluation by a Crowdsourcing based approach within the process of exam preparation. Students are equipped with more responsibilities, they design, comment and evaluate exam questions and build a considerable question pool for exams. A system tracks all interactions when typing in exam questions. Thus, individualized exams can be provided. The large set of exam questions even allows for exams that are written at an individual time. We expect that such a formative knowledge acquisition by students might improve student achievements and even reduce the number of drop-outs.

Keywords: Crowdsourcing, self-assessment, active learning, individual exams

1 Introduction

The organization of (written) exams for large lectures (with several hundred students) is an essential task at universities. Particularly, high effort must be put into the preparation and marking of these exams. Due to a reduced effort for the marking of exams, electronically-based examinations (e-exams) are becoming increasingly common. The process of e-exams is comparable to the organization of conventional exams and is depicted in Fig. 1. First, the exam questions must be prepared by the lecturer. Subsequently the questions have to be composed to an exam where quality assurance is subject to the lecturer. At a predefined time, the exams are written either on paper or at a computer-based system and the exam is supervised by the lecturer. Next, the exams must be marked. Finally, the results have to be published. The marking of e-exams is performed by a system, while paper-written exams are marked by hand. Although a significantly reduced time for marking of e-exams favors them against paper-written exams, still, e-exams have to cope with the challenge that they do not foster individual capabilities of learners.

¹ University of Cologne, Seminar of Information Systems, Pohligstrasse 1, 50969 Köln and Karlsruhe Institute of Technology, Institute AIFB, 76128 Karlsruhe, agnes.koschmider@kit.edu

² University of Cologne, Professur für Berufs- und Wirtschaftspädagogik, Herbert-Lewin-Strasse 2, 50931 Köln, detlef.buschfeld@uni-koeln.de

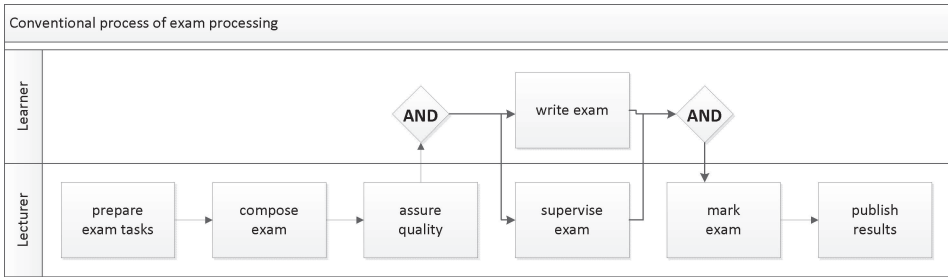


Fig. 1 Conventional process of exam processing

Looking at the process of preparation of e-exams from a role perspective, plenty a variety of activities are undertaken by the lecturer him/herself, who thus takes an active role. The students remain passive, which means that only their passive learning is demanded. Studies comparing active against passive learning, however, show that active learning outperforms passive learning [We07]. Active learning “*is a process whereby students engage in activities, such as reading, writing, discussion, or problem solving that promote analysis, synthesis, and evaluation of class content*” [UM].

To exploit the benefits from active learning also in the process of preparing exams, this paper suggests a novel procedure as modeled with BPMN in Fig. 2. Inspired by the concept of Crowdsourcing [Ho06] this paper suggests an active involvement of students into the process of exam preparation. Instead of writing exams that have been predefined by lecturers, students provide questions and solutions for exams. They indicate levels of difficulty and scores for the questions. Within the frame of exercises students mutually evaluate questions and solutions. The final quality of the exam questions and solutions is assured by the lecturer or with support of an e-exam system being equipped with corresponding review options. When an exam should be conducted, then the system composes an exam out of the students’ questions.

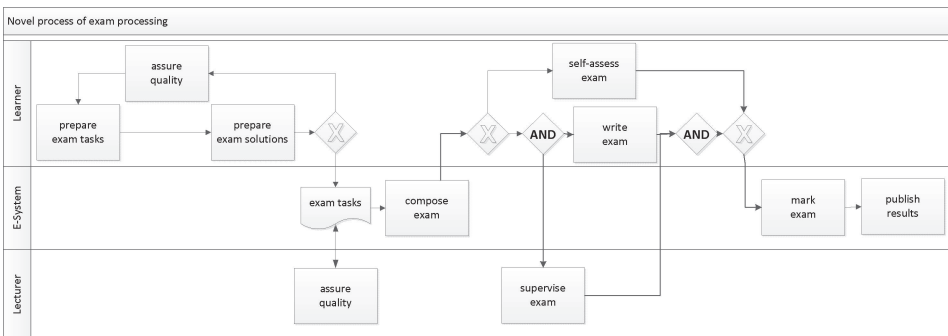


Fig. 2: The novel process of exam processing: a crowdsourcing-based approach

Such a high involvement of students into the process of exam preparation shifts the roles

of students and lecturers. The area of responsibility of lecturers decreases, while that of students increases. A Crowdsourcing based e-exam is a representative of formative learning. It fosters active learning since peers build a map of the learning material already when designing questions and commenting solutions.

Beside the benefits of active learning, such a Crowdsourcing based e-exam also supports the implementation of individual and individualized e-exams. The large set of questions is suitable to provide e-exams at individual time. When typing the questions into an e-exam system, also individual feedback can be returned to learners and lectures. Lectures can be informed about lacks of understandability with the learning material and learners can be pointed to vacant learning material. This feedback can be exploited to design individualized e-exams being capable of individual preferences with respect to the layout of questions.

The idea of Crowdsourcing based e-exams is currently implemented as a pilot study at the University of Cologne. Therefore, this paper can neither provide a proof of concept yet nor report on any significant experiences with Crowdsourcing-based individual and individualized e-exams. Instead, it discusses research challenges and reports the initial implementation of Crowdsourcing-based e-exams. Against this, the next section discusses related approaches and positions our approach while listing the differences. Research questions of Crowdsourcing based e-exams are listed in Section 3. Our pilot study is presented in Section 4. The paper ends with a conclusion in Section 5.

2 Related Work

Crowdsourcing-based individual and individualized e-exams refer to e-assessment and (student) peer evaluation.

E-assessment is a representative of formative self-assessments of online exams. Key challenges of e-assessment are the provision of a robust technical and spatial infrastructure. Usually, the question pool for e-assessment is provided by the lecturers [Ha16]. Open question pools of other universities might also be accessed in order to broaden the view of the learning material [Je16]. The key idea of a Crowdsourcing-based individual and individualized e-exam is a question pool that is not generated by lecturers but by learners. Learners also continuously improve the quality of questions.

Numerous references can be found for peer assessment [To98, Bo01, Le15]. Empirical studies show the positive effects for mutual feedback from students [Fa00]. Particularly, within student peer assessments or student peer evaluation students mutually evaluate the achievements of their peers and are involved in an active review process. Schlagwein [Sc15] recommends peer reviewing for seminar papers. The concept of Crowdsourcing-based individual and individualized e-exams goes beyond user-generated assessments or peer assessment. Within our approach additional requirements must be considered such as the preferences of learners (e.g., with respect to the layout of questions), agreement on

an appropriate level of difficulty or scales of questions.

Initial indications in favor of using Crowdsourcing in the learning process support our approach. For instance, Weld et al. [We12] recommends Crowdsourcing in combination with personalized online education in order to reach full potentials of online education. Anderson [An11] points to positive effects of Crowdsourcing for judging of answers by peers. Corneli and Mikroyannidis [Co12] found out that Crowdsourcing can offer additional richness for accreditation and assessment. To better understand how Crowdsourcing can fuel education, the taxonomy of Mitros and Kim [Mi15] should be used.

To sum up, Crowdsourcing in learning is in its infancy. However, initial empirical studies point to positive effects compared to conventional learning.

3 Research Questions

To implement a Crowdsourcing-based individual and individualized e-exam the following research questions need to be addressed:

(1) *How to prepare learners for e-exams?* To fully exploit the benefits of Crowdsourcing-based e-exams it is necessary to communicate the learning objectives to the students (for example, according to the taxonomy by Anderson / Krathwohl 2001 [An01]). Additionally, the eLearning system must be transparent to the students. Communication is required how the system tracks the tasks of students.

(2) *How to generate a sufficiently large and high-quality questions pool?* The formative involvement of students in the exam preparation requires defining appropriate complexity levels for questions (e.g., simple, advanced, complex). The levels of complexity must be checked and mutually confirmed by peers, which in turn should improve their self-assessment of the exam material. The evaluation of complexity levels might be proven within exercises.

(3) *What is a suitable incentive system?* Students will design “rich” questions only if an attractive incentive system is communicated. The incentive system should also prevent the students from false attempts. Particularly, questions should be developed, which reflect the situational application of knowledge and its justification of choice answers.

(4) *How can the effort be reduced for the validation of questions?* The effort for quality assurance of questions should be reasonable to the reduced marking effort for e-exams. This requires features that mostly automate the validation of questions.

(5) *How to guarantee the comparability of e-exams?* E-exams for the identical lecture should be comparable. For this purpose, the self-assessed complexity of questions by students should be taken into consideration as well as statistical measures, which rate the fulfillment of learning objectives per question type.

(6) *Can principles be observed from individual interactions with the e-exam system?* A wiki should be activated in the e-exam system, which allows observing individual interactions with the system. The observations should give hints how to individualize the exam.

(7) *How to integrate technically the Crowdsourcing-based e-exam into the existing technical conditions?* Finally, the individualization of e-exams should not be limited by technical conditions (e.g., data security).

4 Initial Implementation

A pilot study has been set up in order to validate the feasibility of Crowdsourcing-based individual and individualized e-exams. The opportunity of a bonus has been promised to students of the lecture “Foundations of Information Systems” (Grundlagen der Wirtschaftsinformatik) at the University of Cologne in the summer term 2016. Around 300 students attend this lecture. A bonus of at least 15 points (17% of the total score) will be granted to students if they submit to 5 out of 6 topics exam questions and solutions. The choice option 5 out of 6 should prevent that questions for only simple topics are preferred. To each question the students had to assign a complexity score (one to three stars). The learning objective of each topic was communicated to the students as well. One lecturer or teaching assistant was in charge of one topic in order to detect cheating and duplicates in the question pool. The Ilias system³ is used as IT support system.

5 Conclusion

This paper introduced a Crowdsourcing-based approach for the process of exam preparation. The Crowdsourcing-based approach shifts the responsibilities of learners and lecturers. The active involvement of students into the process of exam preparation has several advantages. First, due to a formative learning the achievements of students (with respect to marks) can be improved. Second, the upload of exam questions and results into an e-exam system allows tracking interactions of students. From these interactions individual preferences with respect to visual or graphical representation can be derived as well as individual feedback for learners and lecturers about the learning material can be provided.

After the completion of the pilot study a report will follow. We plan to investigate other stages of learning that might be suitable for Crowdsourcing-based individual and individualized approaches.

³ <http://www.ilias.de>

References

- [Lo01] Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P., Raths, J., Wittrock, M. C.: A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives, 2011.
- [An11] Anderson, M.: Crowdsourcing Higher Education: A Design Proposal for Distributed Learning, in: *Journal of Online Learning and Teaching*, Vol. 7 No. 4, pp. 576-590, 2011.
- [Fa00] Falchikov, N. and Foldfinch, J.: Student Peer Assessment in Higher Education: A Meta – Analysis Comparing Peer and Teacher Marks. *Review of educational research*, 70, pp. 272-286, 2000 b
- [Ho06] Howe, J.: The rise of crowdsourcing. *Wired Magazine*, 14, 2006
- [Mi15] Mitros, P. and Kim, J.: SPEDS. A Taxonomy for Crowdsourcing in Education, *Conference on Computer Supported Collaborative Learning*, 2015
- [To98] Topping, K.: Peer assessment between students in colleges and universities, in: *Review of Educational Research*, vol. 68/3, pp. 249-276, 1998.
- [We07] Weltman, D.: A Comparison of Traditional and Active Learning Methods: An Empirical Investigation Utilizing a Linear Mixed Model, PhD Thesis, The University of Texas at Arlington, 2007
- [We12] Weld et al.: Personalized Online Education — A Crowdsourcing Challenge, 2012 in: <http://homes.cs.washington.edu/~weld/papers/weld-hcomp12.pdf>
- [Co12] Corneli, J. and Mikroyannidis, A.: Crowdsourcing education on the Web. A rolebased analysis of online learning communities, in: Okada, A., Connolly, T., Scott, P. (Hrsg.): *Collaborative Learning 2.0: Open Educational Resources*, Hershey, PA: IGI Global, pp. 272-286, 2012.
- [UM] University of Michigan, Center for Research on Learning and Teaching, <http://www.crlt.umich.edu/tstrategies/tsal>
- [Le15] Lehmann, K., Söllner, M., Leimeister, M.: Der Wert von IT-gestütztem Peer Assessment zur Unterstützung des Lernens in einer universitären Massenlehrveranstaltung, 12. Internationale Tagung Wirtschaftsinformatik, Osnabrück, pp. 1694-1709, 2015.
- [Bo01] Boud, D.: *Introduction: Making the Move to Peer Learning*. London, Kogan Page, pp. 1-14, 2001.
- [Ha16] Hafer, J., Matthé, F.: Ein nach vorne offener Prozess. E-Assessments an Hochschulen, in: *Forschung & Lehre*, Nr. 3/16, pp.195-197, 2016.
- [Je16] Jeremias, C.: Nur am Rande betrachtet. Prüfungsrechtliche Rahmenbedingungen für elektronische Prüfungen, in: *Forschung & Lehre*, vol. 3/16, pp.202-204, 2016.
- [Sc15] Schlagwein, 2015: Students as Reviewers and Lecturers as Editors: The Peer Review with Scaffolded Assignments Model. 36th International Conference on Information Systems, AISNet, Fort Worth, 2015.

Vorlesung – Quo Vadis?
Lehre an der Schnittstelle
zwischen Mensch und Technik

Vorlesung – Quo Vadis? Lehre an der Schnittstelle zwischen Mensch und Technik

Clemens Cap¹, Robin Nicolay²

Die "Vor-Lesung" ist tot – zumindest wenn man das Wort im ganz traditionellen Wortsinn versteht: Eine Person liest aus einem Text vor, weil Bücher teuer sind und nur wenige über Bücher verfügen.

Die Weiterentwicklung der multimedialen Technologien hat dazu geführt, dass Studierenden heute hunderte Fachbücher und tausende Hochschulschriften zur Verfügung stehen. YouTube hält Lehrfilme bereit und Webinare lösen die Präsenzveranstaltung vor Ort ab. Viele Universitäten bieten MOOC (Massive Open Online Course) Veranstaltungen an, doch schon gibt es mit dem neuen Akronym SPOC (Small Private Online Course) einen Wettbewerber. Unter dem Einfluss der Technologie wandelt sich unsere Vorstellung von Lehre: Der Vorleser erhält eine neue Rolle als Mentor, Motivator und Begleiter durch ein unübersichtliches Dickicht multimedialer Informationsangebote; der reine "Stoff-Transport" wird technischen Systemen überlassen.

Hochschullehrer der Informatik finden sich oft in einer Doppelrolle wieder. Als Forscher und Technologen entwickeln sie Systeme für Kollaboration, Interaktion und multimediale Präsentation. In dieser Rolle sind sie aber häufig von den Anforderungen und Problemen der Praxis abgeschnitten. Außerdem ist mit einer – für den Anwender so wesentlichen Oberflächen- und Prozess-Optimierung – oft nur wenig wissenschaftliche Reputation zu erzielen.

Als Lehrer und Pädagogen sind sie von den Veränderungen der Vorlesung unmittelbar betroffen und sie sammeln Erfahrungen mit neuen digitalen Ansätzen. In dieser Rolle erkennen sie die zentralen Fragen für die Bewährung eines Konzepts, sind Domain-Experten für ihre Lösung und mit den dafür notwendigen Methoden, jetzt wiederum als Informatiker, technisch vertraut.

Da die rein technischen Fragen (Prozessorgeschwindigkeit, multimediale Ein-Ausgabe und Speicherung uvm.) hinreichend gelöst erscheinen, kommt gerade der Schnittstelle zwischen Maschine und Technik einerseits und dem menschlichen Lernprozess andererseits eine neue Bedeutung zu. Diese Fragen sind, wie etwa die Debatte "MOOC versus SPOC" zeigt, noch lange nicht gelöst.

Unter dem Motto "Vorlesung – Quo Vadis?" soll in diesem Workshop daher diese Schnittstelle zwischen Technik und Lernprozess aus dem Blickwinkel der Präsenzveranstaltung genauer analysiert werden.

Die drei Beiträge werfen dabei unterschiedliche Schlaglichter auf Systeme und Konzepte zu einer Weiterentwicklung des digitalen Konzepts von Vorlesung.

¹ Universität Rostock, clemens.cap@uni-rostock.de

² Universität Rostock, robin.nicolay@uni-rostock.de

Lab 360. Interaktive Lernräume zur Verbesserung der Qualität der Lehre an der Hochschule Wismar

Christopher Könitz¹, Jakob Diel², Jürgen Cleve³

Abstract: Auf der Grundlage jahrelanger Erfahrungen und Entwicklungen wurden Labore des ingenieurwissenschaftlichen Fakultät der Hochschule Wismar im zunehmenden Maße digitalisiert. Dazu werden die Labore zunächst fotografiert und in eine 360-Grad-Sphäre umgewandelt. Mittels der Content-Management-Systeme Wordpress und ILIAS können Lehrende anschließend interaktive Inhalte integrieren. Somit haben Studierende die Möglichkeit, sich einen Eindruck der Labore und der damit verbundenen Praktika zu machen. Andererseits dienen die Labore als mediendidaktische Katalysatoren. Dies wird durch die Einbindung der Lernplattform ILIAS erreicht, die das Erstellen von Lernmodulen, Selbsttests, Wikis oder Blogs unterstützt. Dadurch werden neue Lern- und Bildungspotenziale ermöglicht und die Qualität in der Lehre maßgeblich weiterentwickelt. Auf Grund der interaktiven Umsetzung und Integration in die Lehre ist zudem auch der nachhaltige Einsatz der Plattform gewährleistet.

Keywords: Interaktiv, Lernräume, 360, Panorama, Qualität, Lehre, Innovation, Nachhaltigkeit

1 Einleitung

Ein immer wieder genanntes Hauptargument für E-Learning ist das zeit- und ortsunabhängige Lernen. Jedoch gibt es an Hochschulen Lehr- und Lernszenarien, die im besonderen Maße orts- und zeitabhängig sind. An der Hochschule Wismar betrifft dies insbesondere den ingenieurwissenschaftlichen Bereich in Form der Laborpraktika. Auf der Grundlage jahrelanger Erfahrungen und Entwicklungen werden die Labore in diesem Bereich im zunehmenden Maße digitalisiert. Dazu werden die Labore zunächst fotografiert und in eine 360-Grad-Sphäre umgewandelt. Mittels der Content-Management-Systeme Wordpress und ILIAS können Lehrende anschließend interaktive Inhalte integrieren. Dabei ergeben sich für die Studierenden, Lehrenden und die Hochschule vielfältige Chancen und Potenziale:

- Studierende haben *zu jeder Zeit Zugriff* auf bisher verschlossene Räume. Damit verbunden ist auch eine erleichterte räumliche Orientierung und eine digital unterstützte Vorbereitung auf die Versuche, in der durch Informationen und Selbsttests angereicherten 360-Grad-Umgebung. Dies erleichtert insbesondere

¹ Hochschule Wismar, E-Learning-Zentrum, Philipp-Müller-Straße 14, 23966 Wismar, christopher.koenitz@hs-wismar.de

² Hochschule Wismar, E-Learning-Zentrum, Philipp-Müller-Straße 14, 23966 Wismar, jakob.diel@hs-wismar.de

³ Hochschule Wismar, E-Learning-Zentrum, Philipp-Müller-Straße 14, 23966 Wismar, juergen.cleve@hs-wismar.de

internationalen Studierenden den Zugang durch den Abbau sprachlicher Barrieren. Dieser Abbau von Barrieren kommt letztlich allen Studierenden zu Gute und unterstützt die Lernprozesse im Rahmen der Laborpraktika.

- *Lehrende entdecken* durch die Möglichkeiten der 360-Grad-Labore auch andere Tools und Darstellungsformen, wie bspw. die Selbsttests auf ILIAS. Damit transformieren die neu geschaffenen technischen Möglichkeiten auch die Form der Lehre nachhaltig und fördern die mediendidaktische Innovation und den Austausch unter den Kollegen.
- Aus der Perspektive der Hochschule wurde durch die 360-Grad-Plattform eine Möglichkeit geschaffen, um die *Qualität der Lehre nachhaltig zu verbessern* und bestehende Lernplattformen und technische Möglichkeiten⁴ aktiver zu nutzen. Außerdem nutzt die Hochschule die Plattform, um das Leben auf dem Campus Studieninteressenten näher zu bringen.

Anhand dieser cursorischen Darstellung über die Chancen und Potenziale wird deutlich, dass interaktive Lernräume viele Aspekte umfassen. Dieser Beitrag wird sich auf drei zentrale Aspekte konzentrieren. Der erste Aspekt richtet sich auf die Entwicklung und Umsetzung des technischen Frameworks. Der zweite Aspekt umfasst die Lern- und Bildungstheoretischen Möglichkeiten, die interaktive Lernräume bieten können. Schließlich richtet sich der dritte Aspekt auf den praktischen Einsatz in der Lehre.

2 Entwicklung des technischen Frameworks anhand praktischer Bedarfe

Die Entwicklung des technischen Frameworks ist eng mit der Umsetzung von Projektideen verknüpft. 2013 wurde der Grundstein für den Einsatz von 360-Gradumgebungen gelegt, da der Wunsch entstand die Jahresausstellung der Fakultät Gestaltung „DIA“⁵ medial zu dokumentieren. In enger Absprache mit den verantwortlichen Lehrenden wurde ein erstes Framework geschaffen, das als Basis die Software Pano2VR nutzte, und um JavaScript und CSS-Anpassungen erweitert wurde. Damit wurde es möglich einfache Textdateien, Bilder und Videos einzelnen Exponaten zuzuordnen. Die Panoramen wurden mittels Nodalpunktadapter und Spiegelreflexkamera in 38 Einzelbildern aufgenommen und zu einem Panorama mittels der Software Autopano Pro 3 zusammengefügt.

Mit der Ausstellung „Je t’aime MV“⁶ im Jahre 2014 wurde die Plattform überarbeitet. Statt statischen Dateien, wurde das CMS Wordpress für die Contenteinbindung integriert. Der große Vorteil dabei liegt in der einfachen Überarbeitung von Texten, die nun auch

⁴ Zu nennen seien hier beispielsweise das Learning Management System ILIAS oder das Produktionsstudio PELA (<http://www.hs-wismar.de/was/einrichtungen/service/rechenzentrum/produktionsstudio-pela/>)

⁵ Link zur Jahresausstellung der Fakultät Gestaltung DIA 13: <http://dia360.fg.hs-wismar.de/>

⁶ Link zur Ausstellung Je t’aime MV: <http://360.hs-wismar.de/jetaime-mv/>

durch die Ausstellenden vorgenommen werden konnten. 2015 wurde die Plattform um die Möglichkeit erweitert, mehrere Instanzen anzulegen. Somit konnten mehrere Rundgänge getrennt voneinander angelegt werden. In diesem Zuge wurde der Campusrundgang⁷ angelegt. Parallel begannen erste Vorgespräche mit Lehrenden an der Hochschule über die Umsetzung von Laborprototypen für die Lehre.

Für die technische Umsetzung wurde das Framework so weiterentwickelt, dass sich ILIAS-Inhalte per iframe implementieren lassen. Hierzu war zunächst eine Anpassung der Skins der Lernmodule notwendig, so dass es zwischen den Plattformen keinen großen optischen Bruch gibt. Wordpress wurde zudem mit den Plugins „Advanced Custom Fields“ und „Custom Post Type“ erweitert, um einerseits eine ID für das Panorama anzubieten und andererseits diese ID mit einem Post-Type zu verbinden. Letztlich wurde das durch Autopano und Pano2VR erzeugte Panorama um diese Inhalte angereichert. Dazu wurden die entsprechenden CSS-Dateien angepasst und die functions.php für das Darstellen der Wordpress-Inhalte erweitert.

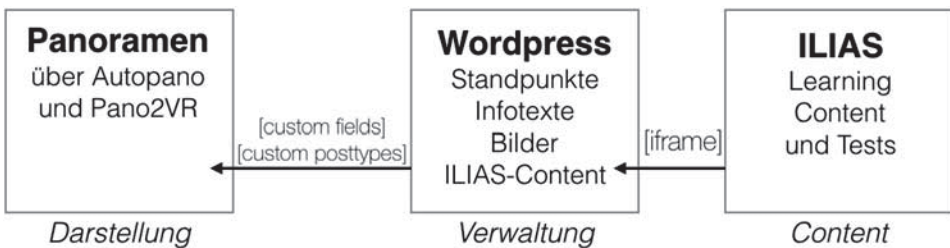


Abb. 1: Schematische Darstellung des Frameworks

3 Potenziale digital angereicherter Lernräume

Neben der technischen Umsetzung der 360-Grad-Labore stand die zentrale mediendidaktische Frage im Raum, wie man durch diese Form der Darstellung die Lernprozesse der Studierenden im Studium unterstützen kann [SLW13] [TH10]. Dabei ist zunächst festzuhalten, dass digital erweiterte Lernräume, wie die oben konturierten 360-Grad-Umgebungen, neue Lern- und Bildungspotenziale eröffnen [JM09][Un10] Dies liegt einerseits in der medialen Transformation des Raums selbst begründet. Der Raum als solcher wird ein hybrides Lernsetting und damit ein Anlass für die Lernenden, neue „Bedeutungs- und Verweisungsstrukturen“ aufzubauen [Un10]. Dazu gehören bspw. das Aneignen des Benutzerinterface oder das Navigieren über verschiedene Seiten und die damit verbundenen Plattformen. Da der digital angereicherte Raum auch immer gleichzeitig ein erlebter Raum ist [Bo63] können die digital unterstützten Lernerfahrungen

⁷ Link zum Campusrundgang: <http://360.hs-wismar.de/campus>

an die konkrete realweltliche Lernsituation (hier das Laborpraktikum) angeschlossen werden.

Außerdem eröffnen digital erweiterte Lernräume neue partizipative und kollaborative Möglichkeiten, indem sie einerseits über (Web 2.0-)Technologien und -Tools neue Formen der Zusammenarbeit und Reflexion zulassen (bspw. über Wikis oder Blogs) [IM10] [Ma14]. Informationen und Wissensbestände können somit über das Semester weiterentwickelt und reflektiert werden. Andererseits kann durch die mediale Vermittlung des Realraums auch die Partizipation am Praktikum selbst unterstützt werden. Die Studierenden können ein Wissen über den Raum und die Praktikumsaufgaben im Vorfeld durch die teilweise Überwindung der „access gap“ aufbauen. Somit können sich insbesondere ausländische Studierende⁸ intensiver auf das Praktikum und die damit verbundenen Fachtermini vorbereiten.

4 Einsatz in der Lehre

Mit der oben beschriebenen technischen Entwicklung des Frameworks sowie der mediendidaktischen Überlegungen, stellte sich letztlich die Frage, wie man diese Technologie in der Lehre einsetzen kann. Durch die Präsentation der Plattform bei offenen E-Learning-Formaten (z.B. Didaktiktage oder E-Learning-Frühstück) und persönlichen Gesprächen mit Lehrenden und Studierenden, konnten erste Szenarien und Pilotprojekte im ingenieurwissenschaftlichen Bereich entwickelt werden. Beispielhaft wollen wir zwei aktuelle Projekte kurz darstellen:

- E-Technik-Labor: Durch dieses 360-Grad-Labor wird die Vermittlung der Grundlagen der E-Technik unterstützt. Räumlich ist dieses Labor zu Fuß ca. 20 Minuten vom Hauptcampus entfernt. Insofern bot es sich durch die räumliche Trennung und den einführenden Charakter als Initialprojekt an. Derzeit sind alle versuchsrelevanten Geräte und ihre Leistungsdaten in der 360-Grad-Darstellung verfügbar (siehe Abbildung 2). Für das Wintersemester 2016/2017 ist eine engere Verknüpfung mit den Aufgaben und Versuchen via Lernmodule und Selbsttests geplant.
- Garten Malchow: ab Mai 2016 wird der Garten der Außenstelle Malchow auf der Insel Poel durch eine 360-Grad-Darstellung angereichert. Die Notwendigkeit ergibt sich durch eine große räumliche Distanz (13,4 km). Ziel ist neben der Verbesserung der Außendarstellung die Dokumentation der Entwicklung der Pflanzen mittels eingebetteter Blogs, die durch die betreuenden Studierenden gepflegt werden.

⁸ Die Zielgruppe der ausländischen Studierenden bildet mit über 890 Studierenden einen wichtigen Teil der Studierendenschaft (Insgesamt über 8500 Studierende).

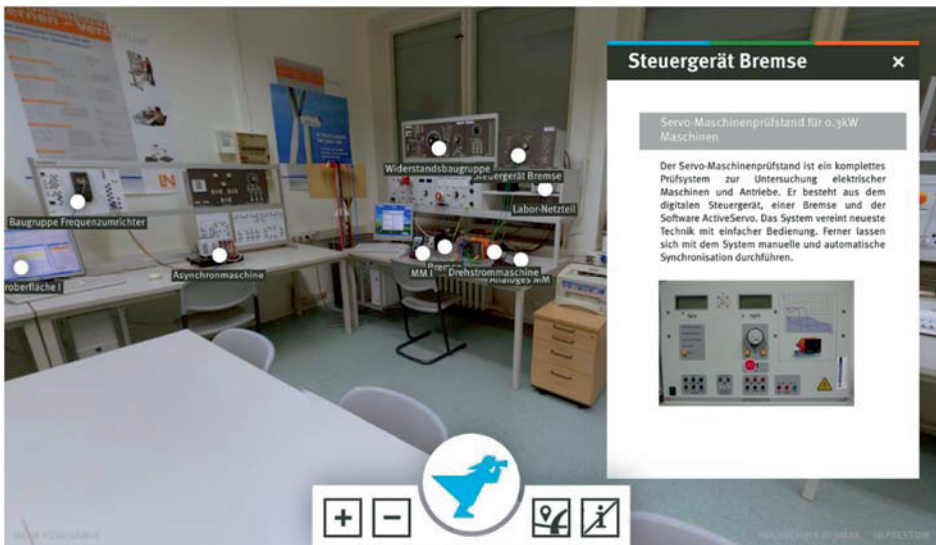


Abb. 2: Umgesetztes E-Technik-Labor

Diese beiden Beispiele zeigen insbesondere mit Blick auf die räumliche Distanz große Vorteile für die Studierenden. Doch auch campusnahe Labore werden für die Studierenden digitalisiert. Auf Grund des Initialaufwands stellt das E-Learning-Zentrum für die Einarbeitung von Inhalten eigene studentische Hilfskräfte⁹ zur Verfügung. Nach dieser Initialphase können die Inhalte durch die Lehrenden über Wordpress und die Lernplattformen leicht angepasst und erweitert werden. Mit Blick auf die oben genannten Beispiele wird dabei deutlich, dass die Entwicklung der Labore als semesterübergreifender Prozess angelegt ist. Dies liegt vor allem im hohen Initialaufwand und der knappen personellen Ressourcen begründet. Jedoch ergibt sich daraus der Vorteil, dass die 360-Grad-Umgebungen iterativ mit der Lehre verzahnt werden. Insofern erfolgt eine nachhaltige Einbindung in die Lehre bei einer gleichzeitigen mediendidaktischen Weiterentwicklung.

5 Fazit und Ausblick

Durch das langjährig eingesetzte und weiterentwickelte Framework und der entsprechenden konzeptionellen Zusammenarbeit zwischen dem E-Learning-Zentrum und den Lehrenden der Hochschule konnten erste 360-Grad-Labore entwickelt werden. Diese dienen einerseits der Zugänglichmachung von zeitlich beschränkten und örtlich entfernten Laboren der Hochschule Wismar. Somit haben Studierende die Möglichkeit, sich einen

⁹ Auf Grund der positiven Resonanz und des geschaffenen Bedarfs prüft die Fakultät Ingenieurwissenschaften, ob sie eigene Mittel für zusätzliche Hilfskräfte zur Verfügung stellen kann.

Eindruck der Labore und der damit verbundenen Praktika zu machen. Andererseits dienen die Labore als mediendidaktische Katalysatoren. Dies wird durch die Einbindung der Lernplattform ILIAS erreicht, die das Erstellen von Lernmodulen, Selbsttests, Wikis oder Blogs unterstützt. Dadurch werden neue Lern- und Bildungspotenziale ermöglicht. Auf Grund der interaktiven Umsetzung und Integration in die Lehre ist zugleich auch die Nachhaltigkeit gewährleistet. Damit verbunden ist auch die Weiterentwicklung der Möglichkeiten der Plattform. Auf Grund der steigenden Nachfrage seitens der Lehrenden werden die Produktionsprozesse in Kürze durch eine 360-Grad-Kamera erleichtert werden. Damit entfallen die mittels Nodalpunktadapter erstellten Panoramen, die viel Arbeitszeit in Anspruch nehmen. Außerdem wird gerade die Anpassung für die Mobilgeräte verbessert, so dass insbesondere Tablets und Smartphones von einer optimierten Darstellung profitieren. Mittelfristig soll die Plattform durch die Integration einer Bewegungssteuerung für Virtual-Reality-Anwendungen weiterentwickelt werden. Einerseits weil Virtual Reality durch erschwingliche Technik bereits jetzt umsetzbar ist (z.B. mittels Smartphone in einem Cardboard). Andererseits weil es bereits jetzt einen Bedarf dieser Kombination in den Bereichen Innenarchitektur und Bauingenieurwesen für die „Begehung“ von Entwürfen gibt.

Literaturverzeichnis

- [IM10] Iske, S.; Marotzki, W.: Wikis: Reflexivität, Prozessualität und Partizipation. In (Bachmair, B., Hrsg) Medienbildung in neuen Kulturräumen. VS Verlag für Sozialwissenschaften, S. 141–151, 2010.
- [Bo63] Bollnow, O. F.: Der Mensch und der Raum. In: Universitas, 18. Jg., S. 499-514, 1963.
- [JM09] Jörissen, B.; Marotzki, W.: Medienbildung - Eine Einführung: Theorie - Methoden - Analysen. UTB, Stuttgart, 2009.
- [Ma14] Mayrberger, K.: „Partizipative Mediendidaktik. Inwiefern bedarf es im Kontext einer partizipativen Medienkultur einer spezifischen Mediendidaktik?“. In (Biermann, Ralf; Fromme, Johannes; Verständig, Dan, Hrsg): Partizipative Medienkulturen. Springer Fachmedien Wiesbaden (Medienbildung und Gesellschaft), S. 261–282, 2014.
- [SLW13] Süss, D.; Lampert, C.; Wijnen, C. W.: Medienpädagogik: Ein Studienbuch zur Einführung. Springer VS, 2013.
- [TH10] Tulodziecki, G.; Herzig, B.: Mediendidaktik: Medien in Lehr- und Lernprozessen verwenden. München: kopaed. 2010.
- [Un10] Unger, A.: „Virtuelle Räume und die Hybridisierung der Alltagswelt“. In (Grell, P.; Marotzki, W.; Schelhowe, H., Hrsg): Neue digitale Kultur- und Bildungsräume. VS Verlag für Sozialwissenschaften, S. 99–117, 2010.

Multiscript – an online student-teacher collaboration platform for classroom lectures

ABM Tariqul Islam,¹ Philipp Jäcks¹ and Clemens H. Cap¹

Abstract: Online collaboration on lecture contents has gained much popularity, over the few decades, due to its potential to enhance the learning experience. We propose a novel idea of an online collaboration platform, called Multiscript (MS), for the students and the teacher, on classroom lectures. MS combines two online learning approaches into a single collaboration platform. One approach, called outside-of-class-Multiscript (OCMS), is a variant of distance online learning which is applicable for the students who want to discuss the lecture slides outside of class due to certain reasons. Another approach, called inside-of-class-Multiscript (ICMS), allows online collaboration among the students and the teacher while the lecture is going on. In OCMS, the teacher can share the slides – with audio annotations for each slides and/or a single recorded audio for the whole lecture. The students can access the slides and discuss (via text and audio chat) with their peers about the slides and annotate them, post feedback about the slide and ask questions to the teacher directly via MS platform. In ICMS, the students can create annotations for the slides and post feedback about the slides which can be read by the teacher later. MS system can be accessed by using just a web browser on any PCs, tablets, notebooks or mobile devices.

Keywords: Collaborative learning, online learning, open learning, e-learning, distance education.

1 Introduction

During the last few decades, online collaboration on classroom lecture content has been an object of interest to the researchers because of its potential to enhance the overall learning experience. While offline collaboration (face-to-face or F2F), among the teacher and the students during the class, has a certain impact on learning, quite a few researchers state that online collaboration provides an equal voice for all the students to express opinion or ask questions about a lecture [SSH06]. Hence, the students consider online collaboration as more neutral than traditional classroom discussions [St11]. Moreover, through online collaboration, the students get an opportunity to reflect on their peer's contribution (eg., comments or questions) about a lecture while they are creating or posting their own comments or questions. It helps the students to foster a certain diligence and deliberation about their peer's contribution [RS03]. Research (e.g., in [SSH06]) indicates that the way the students, being in an online collaboration, perceive the role of their colleagues, discern the social presence of them and create a certain feeling for the collaborating community, essentially affects the learning process.

¹ Department of Computer Science, University of Rostock, Albert-Einstein-Str.22, 18059 Rostock Germany.
E-mail: {tariqul.islam, philipp.jaecks, clemens.cap}@uni-rostock.de

Quite a few research (e.g., [SSH06, RS03]) state that online collaboration, among the students and the teacher, indeed plays an important role to enhance the overall learning experience by engaging the students into a comfortable, productive, equal voiced and socially well-perceived environment. Existing research works on online collaborative learning can be categorized into two main types – (i) distance online collaboration (DOC) (e.g., [BWB09]) and (ii) in-class online collaboration (ICOC) (e.g., [An06]). In case of DOC, the participants, being in geographically different locations, form small groups to collaborate on projects or home works. Whereas, ICOC mainly takes place in the class, between the students and the teachers, on lecture contents or on group works. Usually DOC or ICOC takes place via a range of devices (e.g., stationary PCs, Tablets or mobile phones). To our knowledge, the existing online collaboration platforms either support DOC or ICOC. We didn't come across a system which fully supports both types of collaboration in a single platform.

In this paper, we propose a novel online collaboration platform, called Multiscript (MS), where we combine two online learning approaches into a single collaboration platform. Our first approach, outside-of-class-Multiscript (OCMS), is a variant of distance online learning and the second approach, inside-of-class-Multiscript (ICMS), takes place during the class. OCMS is applicable when the students want to discuss the lecture slides at home because – either, after attending the lecture they still have some questions in mind to discuss with their peers, or, they were absent in the class due to medical issue or traveling purpose. Whereas, in ICMS, the collaboration among the students and the teacher takes place in the class while the lecture is going on. To our knowledge, MS is the first platform which supports both of these approaches in a single online collaboration platform. Hence, MS possesses the attributes of a *blended learning* approach [Po13] where a convergence of F2F meeting and distance learning occurs through available technology. Moreover, MS extends the F2F meeting part of *blended learning* by including the ICMS support into it.

2 Related Work

Here, we present an overview of existing works pursued on online collaborative learning. We mention in the previous section that existing works in this area can be categorized into two main types – distance online collaboration (DOC) and in-class online collaboration (ICOC). In case of DOC, quite a lot of research (e.g., [St11, SSH06, BWB09]) have been pursued which mostly investigate the effect of collaborative learning, among a group of participants, on a specific project or home work. Furthermore, there are quite a few tools available which support DOC for different purposes for different user groups.

Groupboard [Gr16] is one such DOC tool which lets participants to collaborate on a specific topic via text, audio and video chats. The collaboration in Groupboard doesn't occur directly on lecture slides (slides are not accessible), but occurs on a topic from the lecture. It also has an ICMS part with very limited functionality. Vyew [Vy16] is another DOC tool which is used mostly for collaborating on group projects or home works. This tool has the most similarity to our OCMS approach. In Vyew, teachers can upload lecture materials for real-time presentations and then students can meet online and collaborate together in real

time or separately over time. Here, users can create annotation on slides and post suggestions about a lecture content to the teacher via text chat. A couple of other notable DOC tools are Scribbler [Sc16] and Twiddla [Tw16].

Anderson et. al., in [An06], present an ICOC platform, called Classroom Presenter, in which students answer the questions provided by the teacher and submit them to the teacher via wireless connectivity. Teacher evaluates the responses, and randomly selects a few of them to annotate and display simultaneously. Here, the students and the teacher are provided with wirelessly connected stationary Tablet PCs. Hence, it lacks web based connectivity and so, it cannot reach to a wide range of users to interact with the system. It also fails to provide a reliable communication platform. Besides this, in Classroom Presenter, the student's loss of attention might occur and the specific use of a single device limits the use of technology. Caceffo and Da Rocha, in [CDR11], present another ICOC platform called Ubiquitous Classroom Response System. This approach has the most similarity to our ICMS approach. It works similar to Classroom Presenter, but uses context information to create a dynamic environment where students can use any available mobile device to answer the teacher's questions. Moreover, since it uses web based connection to support collaboration, it does not have the limitations of Classroom Presenter. Another similar approach to this system is presented by Oura et. al. in [Ou06].

3 Proposed Idea of Multiscript System

3.1 System description

We previously mention that MS brings two online collaboration approaches (outside-of-class-Multiscript (OCMS) and inside-of-class-Multiscript (ICMS)) into a single platform. Here, we describe about the functional requirements of the OCMS and ICMS respectively.

In OCMS mode, we design the MS system for the students to seamlessly collaborate, via web connection, with their peers on the lecture slides uploaded by the teacher. Here, a teacher, after delivering a lecture, can share the lecture slides by uploading it to the MS server. The students can access the lecture slides on any available PCs, tablets, notebooks or mobile devices at any time *just by using a web browser*. While sharing the lecture slides, a teacher can – attach audio annotations for each of the slides and/or attach a single audio for the whole lecture. So, the students can – listen to the audio annotations for each slide while studying the slides and/or go through all the slides while listening to the entire lecture audio in the background. Figure 1 shows an illustration of OCMS mode. In OCMS mode, the students are offered four different kinds of features, they can –

- perform text chats with their peers
- participate in audio chats with their peers
- annotate the slide for their own purpose
- post feedback about the slides and ask questions to the teacher

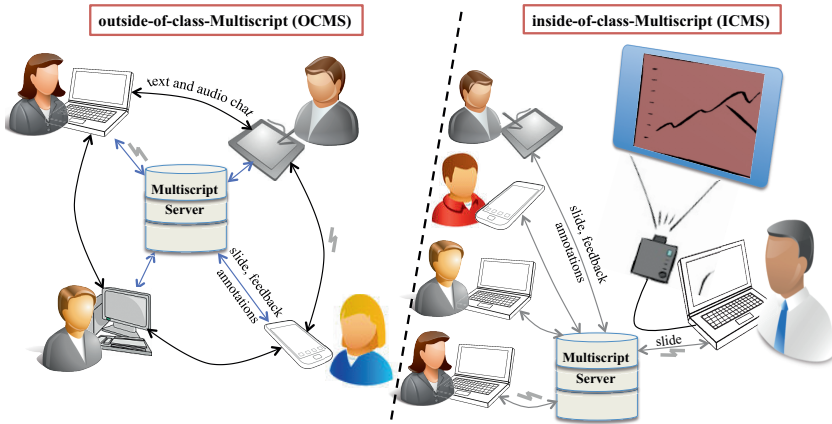


Fig. 1: Illustration of two collaboration platforms (OCMS and ICMS) of proposed Multiscript (MS) system – the left and the right part illustrate the OCMS and ICMS approaches respectively.

For the last case, the feedbacks and the questions are sent to the teacher as a 'newsletter' and the teacher, along with providing appropriate response, can also rate the feedbacks/questions. The teacher is also able to provide reason for his/her voting.

In ICMS, the students can collaborate on the lecture content while being in the classroom. Here, the students can create annotations on the slides for themselves and post feedbacks about the slides which are read by the teacher later. The teacher can response to those feedbacks and rate them similar to OCMS mode. This mode is illustrated in Figure 1. While the teacher delivers a lecture in ICMS, the slides are shown in three different modes:

- via the projector in the classroom
- on the browser of a student's device
- on his own computer's screen

In the second case, the slide displayed on a student's device is synchronized with the projector's current slide. In the third case, the teacher's screen shows a few additional features along with the slides. S/he can see the currently selected slide, a preview of the other slides and some notes for herself. The teacher can also control, from his/her machine, the number of collaborating options which will be available on the students' devices.

Apart from the functional features described above, MS system has a few other requirements which are necessary for providing the best possible user experience. The first requirement of MS is: the interaction with MS system should be fast and responsive enough to meet user's satisfaction. As a second requirement, the users consider a system as comfortable only when the user-interface is easy to use and has the minimalist design. The interface should also have the capability to fit on different screen sizes of different devices (e.g., PCs, Tablets, mobile phones). We design the MS system in such a way that it can be displayed intuitively on various screen size on a range of devices. Besides, while de-

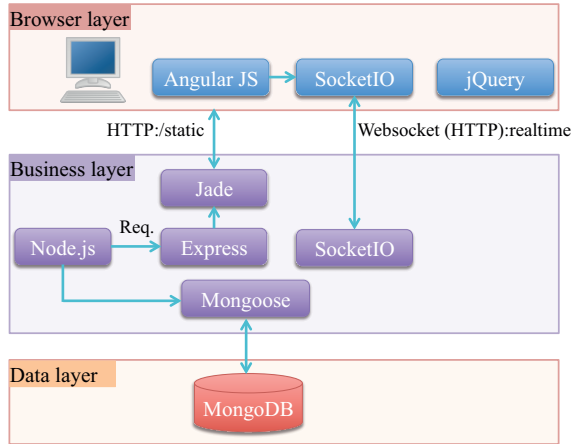


Fig. 2: Technology stack diagram of Multiscript system.

signing the MS system, we have also considered the age of the user's (specially the age of the teachers) of our system. We kept the design simple enough so that people with minimum knowledge about Internet will be able to familiarize themselves with our system with minimum effort. Moreover, MS system is designed to handle quite a large number of users at a time. In future, it might be used by a large number users, so we have considered the access peak of the system when many users start to use MS at once. The scalability of MS system allows large number of connections simultaneously. Besides this, we keep the identity of the students anonymous so that they can avoid the account creation process, and collaborate autonomously with their peers in the MS system. We have designed the MS to be modular so that, in future, new features can be integrated easily into it.

3.2 System architecture

Since we plan to support a wide range of devices (with varying screen size and operating system), rather than developing individual app for each cases, we develop a web application, which serves HTML, for the MS system. Since most of the devices (stationary or mobile) possess the capability to render HTML5, Javascript and CSS, we develop one browser-based client interface which intuitively fits on these devices. This removes the necessity to maintain different repositories for different versions (mobile phone, desktop, iOS, Android etc.) of the system. Moreover, we also consider an easy and minimal design for the user interface. Hence, to achieve a unique development platform and simplistic interface design, we choose to use established frameworks for the web-based user interface. We use Express and SocketIO framework, Node.js, Jade and AngularJS for building the web application of MS. Figure 2 outlines the technology stack of MS. In MS, we allow anonymous participation of the students so that the peers and the teacher cannot trace back to the user. To achieve this, we remove the necessity to register an individual with a university account, rather we allow everyone. Nevertheless, we define certain thresholds to prevent any misuse of the MS system with an anonymous account.

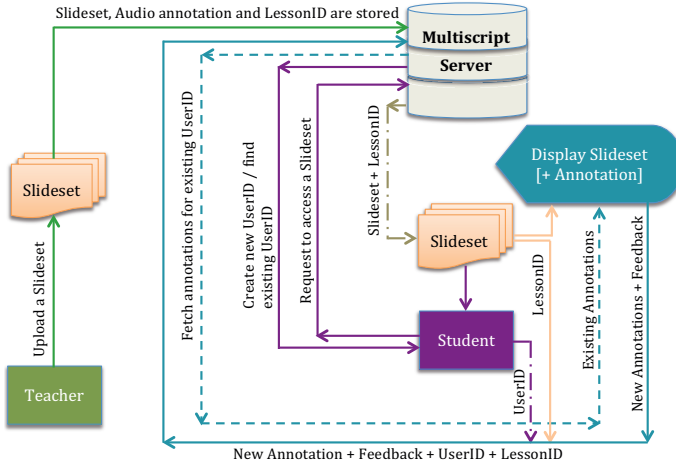


Fig. 3: Data flow within the Multiscript system.

Performance, scalability and responsiveness are the factors which matter most for a good web application, and yet, these are the requirements which are quite difficult to achieve. To develop a stable and fast web application for MS, we looked into numerous solutions for asynchronous web servers which would work with two-way communication channel in tandem. We decide to use the Node.js as our web server. It uses V8, the Google Chrome JavaScript engine. It is event-driven, stable and has the capability to perform thousands of concurrent connections. For MS, a two-way communication channel is necessary, because both, the server and the client, sends messages to each other. In MS, the client is the browser of both the students and the teacher. We use the SocketIO library as the two-way communication layer which is well implemented for Javascript and for the Node.js web server. To support audio chat in MS, we use WebRTC. We choose to use MongoDB, a non-relational database which possesses flexible scheme, to store all permanent data.

3.3 Data flow

In MS, the users are assigned hash values as UserID and a lecture content is assigned a unique LessonID. The data flow in MS are initiated – either by the *teacher* or by the *students*. Figure 3 shows the data flow diagram of MS system. When the *teacher* uploads the lecture content (called a *Slideset*) with audio annotations, an *upload* command is executed within MS. As a result, a corresponding *LessonID* is generated for the Slideset, and subsequently, the Slideset with audio annotations and LessonID are sent to the MS server to be stored. When a *student* selects a Slideset from the list of displayed Slidesets, a *request* is sent to the MS server to fetch the Slideset. Upon getting the request, the MS server fetch the Slideset and associated annotations (if already existed) for that particular student and send them to the students browser. When the students create new annotations and/or feedback for a particular slide, those are sent (along with UserID and LessonID) to the MS server to be stored.

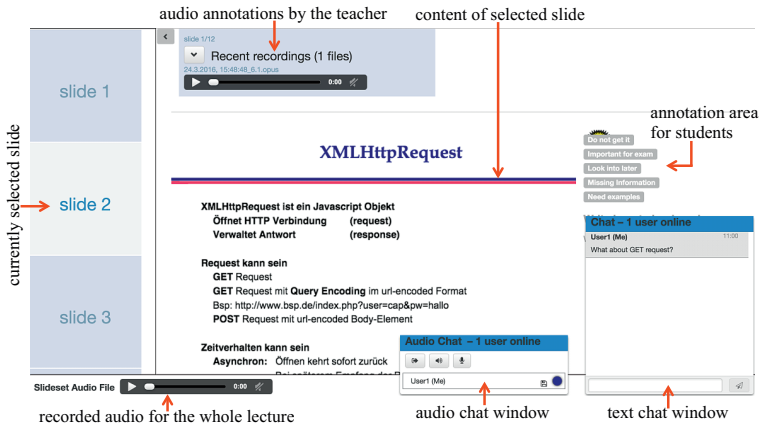


Fig. 4: OCMS collaboration window.

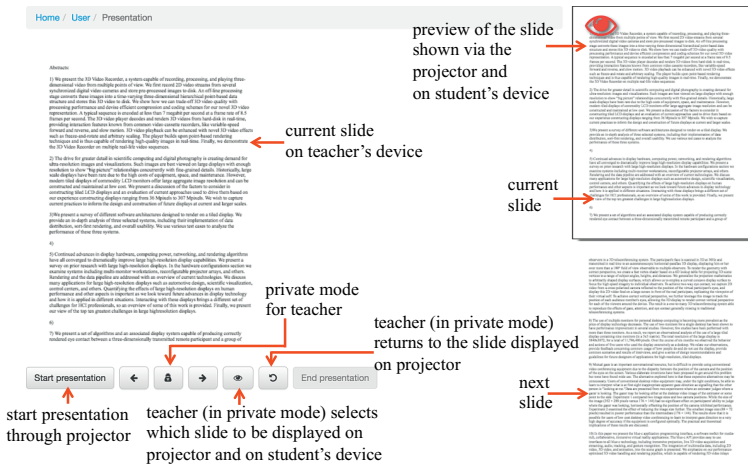


Fig. 5: ICMS collaboration window.

3.4 Development status

Currently, as of May 2016, we have completed the development of most of the features for both OCMS and ICMS mode, except the feedback feature (currently under development) for the students. An overview of the OCMS and the ICMS modes is depicted in Figure 4 and Figure 5 respectively. Figure 4 depicts that the students can look into the slide contents while listening to the audio annotations, collaborate anonymously with their peers via audio and text chat, and annotate the slides for themselves. Figure 5 shows that the teacher can start the presentation on the projector from his/her device while his/her screen shows different options (e.g., preview of current and other slides, private mode). S/he, while in private mode, can press the eye icon to display a particular slide on the projector and simultaneously, on the devices of the students. Currently, we are also developing a better

user interface by collaborating with professional interface designers which is expected to be deployed soon.

4 Conclusion

Online collaboration on lecture content has gained much popularity due to its potential in enhancing the learning process. We propose a novel idea of an online collaboration platform, called Multiscript (MS). MS offers a unique collaboration platform in which the participants can collaborate, using just a web browser, both while being inside of class and outside of class. MS is optimized to be used on different devices with varying screen size. We design the overall MS system to be fast, responsive and scalable, and keep the user interface easy and simple to meet user demand and satisfaction. We allow anonymous access for the students to persuade an autonomous participation during collaborating with peers or the teacher. We plan to test the MS system for computer science lecturers soon.

References

- [An06] Anderson, R.; O. Chung, K.M. Davis; P. Davis, C. Prince, V. Razmov; Simon, B.: Classroom Presenter – A Classroom Interaction System for Active and Collaborative Learning. Proc. Workshop Impact of Pen-Based Technology on Education (WIPTE), 2006.
- [BWB09] Brindley, Jane E.; Walti, Christine; Blaschke, Lisa M.: Creating Effective Collaborative Learning Groups in an Online Environment. The International Review of Research in Open and Distributed Learning, 10(3), 2009.
- [CDR11] Caceffo, Ricardo; Da Rocha, Heloisa Vieira: Ubiquitous Classroom Response System: An Innovative Approach to Support the Active Learning Model. Ubiquitous Learning: An International Journal, 3(1):43–55, 2011.
- [Gr16] Groupboard, <http://www.groupboard.com/products/>, Accessed: 07.05.2016.
- [Ou06] Oura, H.; Kato, H.; ; Akahori, K.: Anchored Feedback: A Classroom Feedback System Considered From An Instructor’s Viewpoint. Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, pp. 2965–2972, 2006.
- [Po13] Poon, Joanna: Blended Learning: An Institutional Approach for Enhancing Students’ Learning Experiences. MERLOT Journal of Online Learning and Teaching, 9(2):271–289, 2013.
- [RS03] Richardson, J. C.; Swan, K.: Examining social presence in online courses in relation to students’ perceived learning and satisfaction. Journal of Asynchronous Learning Networks, 7(1):68–88, 2003.
- [Sc16] Scribblar, <http://www.scribblar.com/>, Accessed: 07.05.2016.
- [SSH06] Swan, K.; Shen, J.; Hiltz, S.R.: Assessment and Collaboration in Online Learning. Asynchronous Learning Networks, 10(1):45–61, 2006.
- [St11] Strijbos, J. W.: Assessment of (Computer-Supported) Collaborative Learning. IEEE Transactions on Learning Technologies, 4(1):59–73, Jan 2011.
- [Tw16] Twiddla, <http://www.twiddla.com/>, Accessed: 07.05.2016.
- [Vy16] Vyew, <http://vyew.com/site/product>, Accessed: 07.05.2016.

SMART Lecture - die technologiegestützte Neuausrichtung der Vorlesung hin zu einer interaktiven Lernumgebung für selbstreguliertes Lernen

Iris Braun¹, Felix Kapp², Tenshi Hara¹, Frank Siegert³

Abstract: Klassische Vorlesungen an Universitäten sind weiterhin weit verbreitet, obwohl sie wegen ihres Mangels an Interaktivität kritisiert werden. Wegen der hohen Teilnehmerzahlen ist die Umsetzung von Interaktivität zwischen Dozierenden und Studierenden schwierig. Es gibt Ansätze, die dem mit technischen Hilfsmitteln entgegenwirken, beispielsweise mit klassischen Clickern, aber auch mit den Smartphones, Notebooks oder Tablets der Studierenden. Basierend auf Forschung zu selbstregulierten Lernprozessen (SRL) haben wir die existierenden Werkzeuge weiterentwickelt: anstatt nur simple Fragen zur Lehrveranstaltung bereitzustellen, haben wir das Konzept der SMART Lecture entwickelt, um diverse Interaktionsmöglichkeiten während der Lehrveranstaltung zu unterstützen. Neben Lernaufgaben unterstützt das entwickelte Tool Echtzeitrückmeldungen, metakognitive Prompts und Erhebungen in Form von Umfragen oder Fragebögen, was neue Evaluationsdurchführungsformen ermöglicht. Im vorliegenden Beitrag präsentieren wir kurz die Funktionalität und ergründen ausgehend von erfolgreichen Pilotversuchen weitere Entwicklungsmöglichkeiten der universitären Vorlesung.

Keywords: Selbstreguliertes Lernen, Audience Response Systeme, Lehrevaluation, Kognitive und metakognitive Prompts

1 Einleitung

Vorlesungen sind nach wie vor eine wichtige Lehrform. Sie haben das Ziel, Studierenden durch die strukturierte Präsentation von Expertenwissen Inhalte zu vermitteln, welche über das gängige Lehrbuchwissen hinausgehen. Seit langem wird kritisiert, dass diese Lehrform zu wenig Interaktion zwischen Lehrenden und Studierenden zulasse und Lernen als ein aktiver, konstruktiver und höchst individueller Prozess [Se03] in den Großveranstaltungen nahezu unmöglich sei. Als Konsequenz würden sich deshalb nur unzureichende Lerneffekte ergeben. Verstärkt wird das Problem durch die zunehmende Heterogenität der Studierenden. So führt bspw. an der TU Dresden die Studiengangsumstellung von Diplom auf Bachelor/Master, die zunehmende Internationalisierung und ein voranschreitendes Interesse an fachübergreifendem Wissenstransfer zwischen Fachrichtungen dazu, dass Studierende in Vorlesungen sehr unterschiedliche Ziele, Vorwissen und Herangehensweisen

¹ Professur Rechnernetze, Technische Universität Dresden (TU Dresden), {iris.braun | tenshi.hara}@tu-dresden.de

² Professur für die Psychologie des Lehrens und Lernens, TU Dresden, felix.kapp@tu-dresden.de

³ Emmy Noether-Juniorprofessur am Institut für Kern- und Teilchenphysik, TU Dresden, frank.siegert@tu-dresden.de

an die Lerninhalte haben. Das Projekt *SMART Lecture* (Self-regulated Mobile Assessment Responsive Tutoring) hat sich vorgenommen, die Veranstaltungsform Vorlesung so zu erweitern, dass Studierende beim Erreichen ihrer individuellen Lernziele besser unterstützt werden.

In Forschung und Lehre gibt es bereits verschiedene Ansätze, die Interaktivität in Vorlesungen zu erhöhen. Die Bandbreite reicht von einfachen Abstimmungssystemen (auch „Clicker“ genannt, z.B. [BSR13]) bis hin zu elaborierten Audience Response Systemen mit unterschiedlichen Frage- und Interaktionsmöglichkeiten (z.B. ARSnova [GKK13] oder SMILE [WB13]), mit welchen man z.B. die Methode der Peer Instruction [Ma97] umsetzen kann. An der Justus-Liebig-Universität Gießen wurden die Einsatzmöglichkeiten, Funktionen und Technik von 30 web- und hardware-basierten Abstimmungssystemen umfassend evaluiert und tabellarisch gegenübergestellt [Be15].

Das Konzept der SMART Lecture setzt hier an, geht aber sowohl konzeptionell als auch technisch einen Schritt weiter. Mit Hilfe des an der TU Dresden entwickelten Tools *Auditorium Mobile Classroom Service* (AMCS) wird die Vorlesung zu einer Lehrveranstaltung, welche Studierende beim Selbstregulierten Lernen unterstützt (*Self-regulated*). Dafür werden mobile Technologien wie Smartphones, Tablets oder Laptops verwendet (*Mobile*), um sowohl Studierenden eine Evaluation ihres eigenen Lernprozesses als auch Dozierenden die Evaluation der gesamten Veranstaltung zu ermöglichen (*Assessment*). Diese Daten bilden die Grundlage für die Berücksichtigung individueller Lernverläufe in der „Massenveranstaltung“ Vorlesung. Durch die Dozierenden definierte Nachrichten und Fragen werden den Studierenden in Abhängigkeit ihrer Antworten auf Umfragen und Lernaufgaben zugestellt (*Responsive*). Der Entwicklung des Tools liegt ein didaktischer Paradigmenwechsel zu Grunde: die Dozierenden werden als Gestalter von Lernumgebungen betrachtet und haben daher die Aufgabe, die Studierenden beim Erreichen ihrer individuellen Ziele zu unterstützen. Dozierenden kommt daher eher die Rolle eines Tutors als die eines klassischen „Vorlesenden“ zu (*Tutoring*).

Das zuvor erwähnte zentrale Werkzeug, nämlich AMCS, bietet zahlreiche Funktionen, die Studierende bei individuellen Lernprozessen in der Vorlesung unterstützen und gleichzeitig eine informative Evaluation der Lehrveranstaltung ermöglichen. Die Funktionalitäten des Systems wurden auf der Grundlage lernpsychologischer Forschung entwickelt (bspw. [WH98]; [Zi00]). Gemeinsam haben sie das Ziel, Studierende in Abhängigkeit individueller Bedürfnisse dabei zu unterstützen, in der Vorlesung möglichst viel zu lernen. Die einzelnen Features zielen darauf ab, den Studierenden Hilfestellungen für einen erfolgreichen Lernprozess in der Vorlesung zu geben (siehe Abb.1). Damit verändert sich aber auch die Rolle der Dozierenden: Sie sind nicht mehr nur Vortragende, sondern Konstrukteure von Lernumgebungen. Die über AMCS zur Verfügung gestellten Umfragen, Nachrichten und Lernaufgaben werden im Vorfeld durch die Dozierenden erstellt und während der Veranstaltung nach vorher genau definierten Regeln an die Endgeräte der Studierenden gesendet. Dies wird über ein die Präsentationssoftware (bspw. MS PowerPoint) erweiterndes Zusatzprogramm realisiert, welches Fragen, Lernaufgaben und Nachrichten an

die Folien koppelt, um sie dann automatisiert in der Vorlesung an die Studierenden auszuliefern. Folgend werden die Funktionalitäten erläutert.

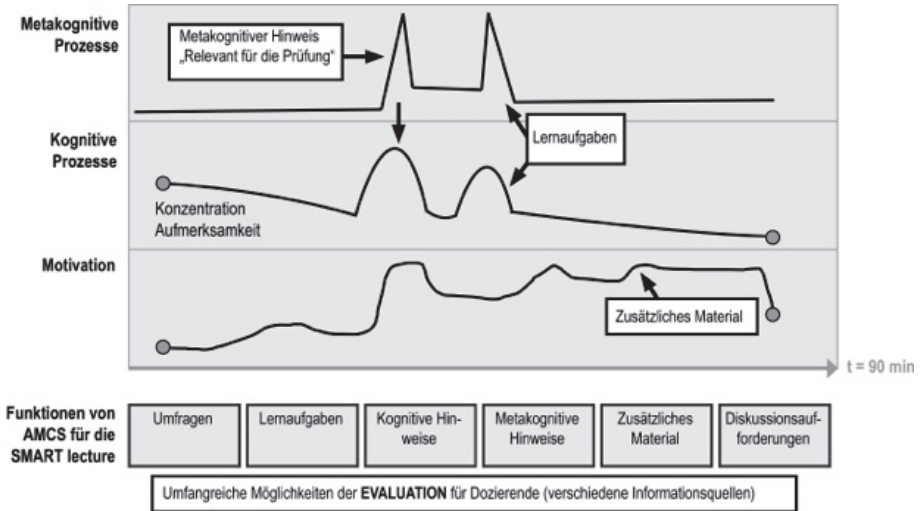


Abbildung 1: Prototypischer Lernprozess in der Vorlesung (kognitiv, metakognitiv, motivational)

2 Funktionen von AMCS

Alle Funktionalitäten werden über Smartphones bzw. andere internetfähige Geräte zur Verfügung gestellt. Sowohl die Umfragen und Nachrichten als auch die Lernaufgaben sowie die jeweiligen Ergebnisdarstellungen werden in einer webbasierten Anwendung oder in nativen Apps abgebildet. Ziel des Systems ist es, eine SMART Lecture zu ermöglichen: „Self-regulated“, „Mobile“, „Assessment“, „Responsive“ und „Tutoring“.

2.1 Abfrage von Interessen und persönlichen Zielen

Durch einige wenige Fragen werden zu Beginn der Vorlesung die persönlichen Ziele erfasst. So kann nachfolgend durchaus unterschiedlichen Zielstellungen Rechnung getragen werden. Die erhobene Information dient als Grundlage für Nachrichten und Hinweise, die zu späteren Zeitpunkten an die Studierenden geschickt werden, um ihnen bei der Regulation des eigenen Lernprozesses zu helfen. Gleichzeitig regt sie diese kurze Befragung an, sich über ihre eigenen Ziele und Interessen klar zu werden und sie ggf. zu präzisieren (bspw.: „Prüfungsvorbereitung“, „Interesse am Thema“, „Pflichtveranstaltung“).

2.2 Lernaufgaben am Anfang, in der Mitte und am Ende der Vorlesung

Interaktive Lernaufgaben können den Lernprozess sowohl bei den notwendigen kognitiven als auch metakognitiven Prozessen unterstützen ([KNK11], [MSD09]). Zeitlich am Anfang, in der Mitte und am Ende der Vorlesung angesiedelt, unterstützen sie die Studierenden bei einer aktiven Auseinandersetzung mit dem Inhalt (eine Beispielaufgabe ist in Abb. 2 als „Lernaufgabe 1“ abgebildet). Im Gegensatz zu bisherigen Clicker-Systemen, die den Dozierenden ein Meinungsbild der Studierenden liefern, erhalten die Lernenden durch das Tool zusätzlich direktes individuelles Feedback.

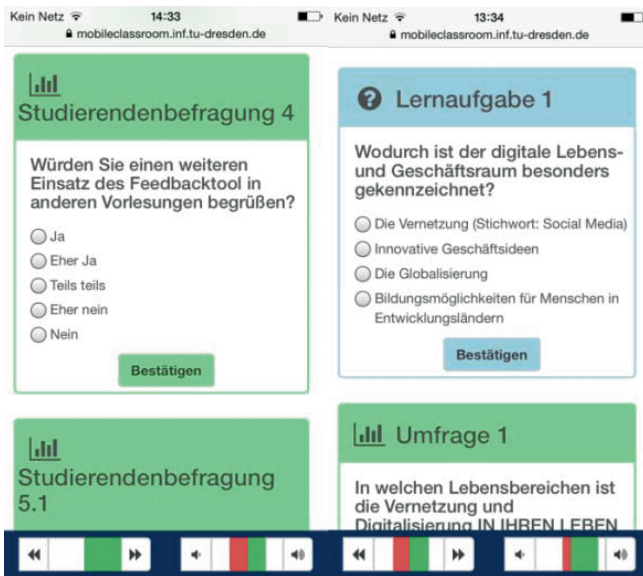


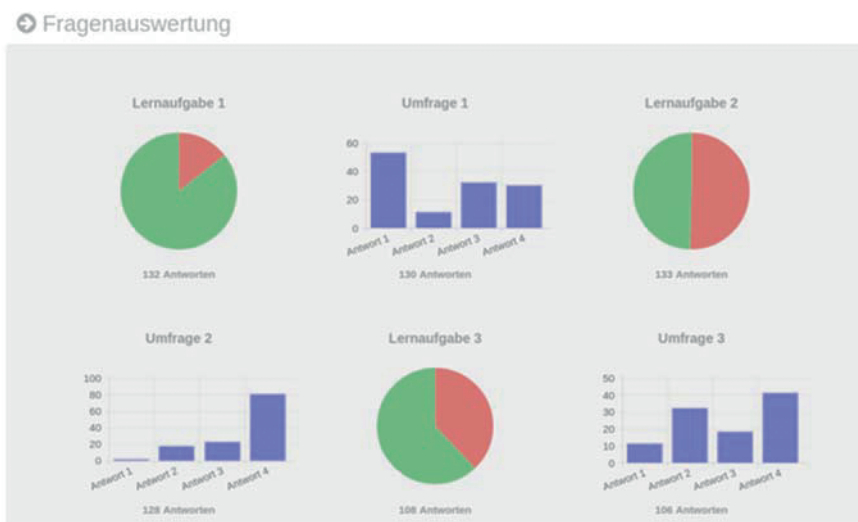
Abbildung 2: Studierendenbefragung und Lernaufgabe in AMCS

Insbesondere diese Funktionalität fördert zeitnah zur Stoffvermittlung bei den Studierenden einen aktiven, konstruktiven und höchst individuellen Wissenserwerbsprozess. Darüber hinaus dient zu Beginn der Vorlesung das Bearbeiten von Lernaufgaben dem Aktivieren von Vorwissen. Durch die Aufgaben werden die Anforderungen offengelegt und die Aufmerksamkeit auf bestimmte Inhalte gelenkt. Bspw. nach der Hälfte der Vorlesungszeit können die Studierenden anhand kurzer Lernaufgaben sowohl den thematisierten Stoff in einem ersten Durchgang üben als auch anhand des Feedbacks eine schnelle Rückmeldung über ihren Wissensstand bekommen [KBK14]. Mit einigen Lernaufgaben zum Abschluss der Lehrveranstaltung können wichtige Inhalte wiederholt sowie Rückmeldung über den Lernzuwachs gegeben werden. In Abhängigkeit ihrer Ziele können die Studierenden Konsequenzen für zukünftige Veranstaltungen hinsichtlich Aufmerksamkeits- und Motivationsregulation sowie der angewendeten Lernstrategien ziehen.

2.3 Prompts bzw. individuelle Nachrichten während der Vorlesung

Ziel der Prompts ist die Unterstützung der Studierenden bei der Aufmerksamkeitsregulation und der Zielerreichung in der Vorlesung (metakognitive Prompts). Die Prompts erfolgen in Form von Push-Nachrichten. In Abhängigkeit von Hintergrund und individuellen Zielen werden strategische Hinweise während der Vorlesung gegeben. Daten für die Auswahl der Nachrichten werden auf Grundlage eines kurzen Fragebogens zu Beginn der Vorlesung gewonnen (siehe Abschnitt 2.2). So bekommen bspw. Studierende, die sich primär für das Bestehen der Abschlussprüfung interessieren, einen Hinweis der Form „Das Thema auf der aktuellen Folie ist prüfungsrelevant. Die Professorin fragt in mündlichen Prüfungen nach der Definition ...“. Bei Interesse an einer Abschlussarbeit kann in der Veranstaltung ein Link auf Themenangebote versendet werden.

Darüber hinaus werden Nachrichten mit dem Ziel versandt, Studierende beim Wissenserwerb zu unterstützen (kognitive Prompts). So bekommen Studierende, welche bei einer Lernaufgabe zu Beginn Fehler gemacht haben, zu einem späteren Zeitpunkt bspw. folgende Nachricht: „Sie hatten bei Lernaufgabe 1 noch einen Fehler, auf der aktuellen Folie erläutert der Dozent das Konzept ... noch einmal.“ Mit Hilfe von Nachrichten können auch Diskussionen initiiert, in Ausnahmefällen sogar inszeniert werden. Es werden über Nachrichten vom System Kommentare, Argumente und/oder Rollen den verschiedenen Studierenden zugeteilt, welche sie zu bestimmten Zeitpunkten im Plenum einbringen sollen. Ziel dieses Scriptings ist es, Anteile der Vorlesungszeit optimal für Diskussionen und argumentativen Austausch zwischen den Dozierenden und den Studierenden zu nutzen, da erfahrungsgemäß aufgrund der Größe der Veranstaltung Wortmeldungen nur sehr zaghaft oder überhaupt nicht zu Stande kommen.



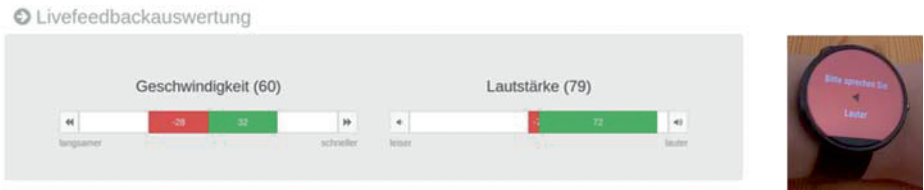


Abbildung 3: Ergebnisse von Lernaufgaben (oben) und Echtzeitrückmeldungen (unten)

2.4 Instant Feedback

Die Studierenden können den Vortragenden in Quasi-Echtzeit Rückmeldungen hinsichtlich der Redegeschwindigkeit und der Lautstärke geben. Die grundlegende Funktionalität auf Seiten der Studierenden ist im unteren Bereich von Abbildung 2 dargestellt. Die verschiedenfarbigen Balken ermöglichen die schnelle Erfassung bereits abgegebener Rückmeldungen, was einer einfachen Durchschnittsanzeige vorzuziehen ist. Das Instant Feedback erlaubt den Studierenden direktes Aufzeigen von Defiziten im Vortrag, während es Vortragenden zeitnahe Korrektur ermöglicht. Damit der Vortragende nicht ständig auf einem zweiten Gerät die Ergebnisse des Feedbacks überwachen muss, wurde AMCS mit einer Smartwatch gekoppelt. Diese vibriert und zeigt das entsprechende Feedback an, sobald ein gewisser Schwellwert überschritten wird (siehe Abb. 3: rechts unten). Der Dozent kann diesen Grenzwert selbst konfigurieren, indem er prozentuale oder absolute Werte für Rückmeldungen festlegt. Auf diese Weise wird eine starke Korrelation zwischen Auftreten, Aufzeigen und Korrektur eines Defizites erreicht.

2.5 Umfangreiche Informationen für die Evaluation

Oft werden Evaluationen durch Papierfragebögen am Ende eines Semesters realisiert. Die Fragen beziehen sich meist auf alle Veranstaltungen der Vorlesungsreihe, die Auswertung dauert lange. Daher wird in der Regel eine sinnvolle Diskussion des Feedbacks mit den Befragten unmöglich, da die Ergebnisse erst deutlich nach Ende der Veranstaltungsreihe vorliegen. Des Weiteren erschwert der zusammenfassende Charakter der Evaluationen dem Lehrenden gezielte Verbesserungen einzelner Lehrveranstaltungen. Typische Fragen der Art „Wirkte der Lehrende gut vorbereitet?“ zielen auf Charakteristiken ab, die sich teilweise deutlich von Woche zu Woche ändern können. Die geringe Wirksamkeit der Evaluationsumfragen führt im Umkehrschluss oft dazu, dass sich Studierende immer weniger an den Befragungen beteiligen, da sich der Nutzen meist erst für darauffolgende Jahrgänge einstellt. Vor diesem Hintergrund ist es das Ziel des „Assessment“, eine direkte Evaluation von Vorlesungen zu ermöglichen, um diese auch kurzfristig zu verbessern. Mit Hilfe von AMCS kann man die Vorlesung dabei zu verschiedenen **Zeitpunkten** evaluieren.

Während der Lehrveranstaltung können Lehrende auf Live Feedback zurückgreifen, um ihre Präsentationen unmittelbar zu verbessern. Da sie im Vortrag damit beschäftigt sind, Inhalte zu vermitteln und Stoff zu erklären, sollte die Bereitstellung von Evaluationsrückmeldungen auf die wirklich notwendigen Informationen reduziert werden. AMCS präsentiert daher die Ergebnisse von Echtzeitrückmeldungen (Lautstärke und Geschwindigkeit; siehe Abb. 3) und zeigt studentische Antworten auf Lernaufgaben an. Insbesondere während der Unterbrechungen, in der Studierende die Lernaufgaben beantworten (Abschnitt 2.2), können Lehrende über die Ergebnisse der Lernaufgaben prüfen, in wie weit die Lernenden das aktuelle Thema verstanden haben.

Nach einzelnen Lehrveranstaltungen sind detaillierte Analysen von Lernfragen, Umfragen und der Echtzeitrückmeldungen möglich. Diese Daten erlauben den Lehrenden eine zeitnahe Bewertung der soeben beendeten Lehrveranstaltung. Sie können bspw. das Lernverständnis der Studierenden und den Wiederholungsbedarf zu Beginn der nächsten Lehrveranstaltung abschätzen. Eine Aggregation der Echtzeitrückmeldungsdaten ermöglicht die einfache Identifikation von kritischen Momenten innerhalb der Vorlesung.

Nach dem Semester oder nach Abschluss des Kurses kann die traditionelle, normalerweise papierbasierte Evaluation digital umgesetzt werden. Befragungen mit Skalenfragen, Multiple- oder Single-Choice-Fragen sowie Freitextfragen sind mit AMCS einfach umsetzbar. Da die zugrundeliegende Datenbank die Antworten an einzelne Nutzerkonten bindet, kann beispielsweise auch der individuelle Lernfortschritt einzelner Studierender oder von Gruppen von Studierenden ermittelt werden. AMCS ermöglicht es folgende **Datenquellen** für die Evaluation zu verwenden: **Befragungen**, in welchen die Studierenden ihre Leistungen, Interessen und Motivation selbst einschätzen. **Leistungsdaten** werden über die Beantwortung der Lernaufgaben in der Vorlesung erfasst. **Echtzeitrückmeldungen und Nutzungsdaten** erlauben Rückschlüsse über die Nutzungsintensität, insbesondere wann und wie häufig Studierende Gebrauch von angebotenen Lernaufgaben und Befragungen machen. **Statistiken** über die Anzahl Studierender, die während der Veranstaltungen online und mit Lernaufgaben beschäftigt waren, erlauben interessante Rückschlüsse über die Beteiligung über den Verlauf eines Kurses hinweg.

Da für die Nutzung von AMCS lediglich die Angabe einer beliebigen Kombination aus frei wählbarem Pseudonym und Kennwort ausreicht, können die Evaluationsdaten im Vergleich zur traditionellen Evaluation als gleichwertig sicher und anonym betrachtet werden. Datenschutzbedenken bestehen lediglich auf der üblicherweise als unvermeidbar betrachteten technischen Ebene, bspw. des Logs von IP-Adresse und Zugriffszeit.

3 Pilottests

Das Konzept SMART Lecture mit AMCS als Kerntool wurde bereits in mehreren Veranstaltungen getestet (siehe [HKB15], [BKK15]). Empirische Ergebnisse aus drei universitären Vorlesungen ((n1=30 Psychologie, n2=186 Wirtschaftswissenschaften, n3=47 Phy-

sik) zeigen, dass die Studierenden AMCS sehr positiv bewerten, die Interaktivität im Vergleich zu anderen Vorlesungen als größer wahrgenommen wird und vor allem die Lernaufgaben als sehr nützlich eingeschätzt werden.

Im Wintersemester 2015 wurde AMCS über eine ganze Lehrveranstaltungsreihe hinweg eingesetzt: Studierenden der Kern- und Teilchenphysik wurden sowohl Lernaufgaben als auch Umfragen zur Verfügung gestellt. Dabei hat die Hälfte der Studierenden AMCS in mindestens 75% der Lehrveranstaltungen verwendet. Die Nutzungsmotivation seitens der Studierenden ergründete sich vornehmlich in einer Belebung der Vorlesung, der Stoffwiederholung und der individuellen Selbstkontrolle. Die Funktionalitäten während der Lehrveranstaltungen wurden sehr positiv bewertet (4,21; $n=46$; $SD=0,85$; 1: „gar nicht sinnvoll“, 5: „sehr sinnvoll“). Solche, die zwischen zwei Lehrveranstaltungen zu verwenden waren, wurden neutraler bewertet (3,31; $n=45$; $SD=0,12$). Positiv äußerten sich die Studierenden in Bezug auf den Zugewinn an Interaktivität und Rückmeldungen an den Dozenten. Das Ablenkungspotential wurde eher gering bewertet (2,21; $n=47$; $SD=0,88$). Qualitative Rückmeldungen der Studierenden zeigten transiente Kritikpunkte, wie etwa der fehlende Zugriff auf die Lernaufgaben und Rückmeldungen außerhalb von Vorlesungen, oder auch Fehlen einer $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ -Unterstützung zur Formeldarstellung.

Weiterhin wurde AMCS verwendet, um die traditionelle papierbasierte Evaluation einer Lehrveranstaltung zu ersetzen. In einem Psychologie-Seminar zu „Serious Games“ beantworteten die Studierenden am Ende der Veranstaltung eine Reihe von Evaluationsfragen, welche in Anlehnung an die Fragebögen des Zentrums für Qualitätsanalyse (ZQA) der TU Dresden zur Lehrevaluation erstellt wurden.

Die Nutzung der Smartwatch zur direkten Rückmeldung des Instant Feedback an den Lehrenden wurde im Wintersemester 2015 in der Informatik-Vorlesung „Service and Cloud Computing“ getestet. Die Dozentin berichtete im Anschluss über ihre positiven Erfahrungen und gab weitere Anregungen zur Verbesserung der Usability.

4 Ausblick

Empirische Ergebnisse aus den universitären Pilottests zeigen, dass die Studierenden AMCS sehr positiv bewerten. Auf Grundlage der ersten Erfahrungen lassen sich Möglichkeiten und Grenzen der technologiegestützten Verbesserung der universitären Lehre diskutieren. Derzeit werden weitere Pilottests sowie Laborstudien zur experimentellen Untersuchung von Aufmerksamkeitslenkung durch die Nachrichten während der Vorlesung geplant. Weiterhin soll untersucht werden, inwieweit der Einsatz von AMCS zur Erhöhung der Lernmotivation, der Verbesserung der Interaktion zwischen Dozierenden und Studierenden sowie schlussendlich zu einem Wissenszugewinn führen kann. Dabei ist bei allen Weiterentwicklungen und auch den bisher umgesetzten Funktionen wichtig, dass sich AMCS als ein zusätzliches Unterstützungsangebot versteht. Trotz des automatisierten Ablaufs einiger Prozesse vor, während und nach der Vorlesung bleiben die Beteiligten Personen - egal ob Studierende oder Lehrende - souverän. Sie können jeweils entscheiden,

ob sie das Unterstützungsangebot annehmen. So werden Leistungsdaten der Studierenden nicht personalisiert an Dritte gegeben, die Evaluationsdaten sieht der jeweilige Lehrende und kann entscheiden, ob er sie mit dem Auditorium diskutieren möchte. Technisch wird die Plattform dahingehend weiterentwickelt, dass eine Konfiguration und Auswahl der verwendeten Funktionen einfach durch den Lehrenden möglich ist. Für verschiedene Lehrformen und -situationen sollen den Dozierenden Empfehlungen für die Kombination der Features gegeben werden.

Literaturverzeichnis

- [Be15] Bernhardt, S., 2015. Vergleich unterschiedlicher Abstimmungssysteme. http://ilias.uni-giessen.de/ilias/goto.php?target=cat_25037, Stand: 23.7.2015.
- [BSR13] Brady, M., Seli, H., & Rosenthal, J., 2013. Metacognition and the influence of polling systems: How do clickers compare with low technology systems. *Educational Technology Research and Development*, 61(6), 885-902.
- [BKK15] Braun, I., Kapp, F., Körndle H., Schill, A., 2015. Onlinegestützte Audience Response Systeme: Förderung der kognitiven Aktivierung in Vorlesungen und Eröffnung neuer Evaluationsperspektiven. *Proceedings of Wissensgemeinschaften 2015 (GeNeMe 2015)*, Dresden, Germany.
- [GKK13] Gerhardt, D., Kammer, J., Knapp, D., Quibeldey-Cirkel, K., Thelen, C., & Volkmer, P. C., 2013. ARSnova: ein Audience Response System für Inverted-Classroom-Szenarien mit Unterstützung von Just-in-Time Teaching und Peer Instruction. *DeLFI 2013 – Die 11. e-Learning Fachtagung Informatik (2013)*.
- [HKB15] Hara, T., Kapp, F., Braun, I., Schill, A., 2015. Comparing Tool-supported Lecture Readings and Exercise Tutorials in classic University Setting. In *Proceedings of the 7th International Conference on Computer Supported Education (CSEDU 2015)*, Lisbon, Portugal.
- [KNK11] Kapp, F., Narciss, S., Körndle, H., Proske, A., 2011. Interaktive Lernaufgaben als Erfolgsfaktor für E-Learning. *Zeitschrift für E-Learning*, 6 (1), S.21-32.
- [KBK14] Kapp, F., Braun, I., Körndle, H., Schill, A., 2014. Metacognitive Support in University Lectures Provided via Mobile Devices. In *Proceedings of the 6th International Conference on Computer Supported Education (CSEDU 2014)*, Barcelona, Spain.
- [KNP04] Körndle, H., Narciss, S., Proske, A., 2004. Konstruktion interaktiver Lernaufgaben für die universitäre Lehre. In D. Carstensen & B. Barrios (Eds.), *Campus 2004. Kommen die digitalen Medien an den Hochschulen in die Jahre?* (pp. 57-67). Münster: Waxmann.
- [MSD09] Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., et al., 2009. Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology*, 34(1), 51-57.
- [Ma97] Mazur, E., 1997. *Peer Instruction: A User's Manual*. Upper Saddle River: Prentice Hall.
- [Se03] Seel, N.M., 2003. *Psychologie des Lernens (2nd edition)*. München: Ernst Reinhardt (UTB).

- [WB13] Weber, K., Becker, B., 2013. Formative Evaluation des mobilen Classroom-Response-Systems SMILE. In C. Bremer & D. Krömker (Hrsg.), *E-Learning zwischen Vision und Alltag*, (S. 277-289). Münster: Waxmann.
- [WH98] Winne, P. H., & Hadwin, A. F., 1998. Studying as self-regulated learning. *Metacognition in educational theory and practice*, 93, 27-30.
- [Zi00] Zimmerman, B. J., 2000. Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation*. (pp. 13-39). San Diego, CA US: Academic Press.

SAG WAS
Studentische Ausbildung und
berufliche Weiterbildung in
Allgemeinen Schlüssel-
qualifikationen

SAG WAS – Studentische Ausbildung und berufliche Weiterbildung in Allgemeinen Schlüsselqualifikationen

Karin Vosseberg¹

System- und Softwareentwicklung passiert immer weniger im dunklen Keller, im Hinterhof oder der Garage. Das Bild der Informatiker als einsame Kellerkinder ist längst passé. Komplexe Systeme entstehen in interdisziplinären Teams. Neben fachlichen Kompetenzen sind überfachliche Schlüsselqualifikationen wesentliche Erfolgsfaktoren für eine gute Produktentwicklung oder IT-Dienstleistung. Es ist ein hohes Maß an Strukturiertheit und Kommunikationsfähigkeit gefordert. Dies wird verstärkt in global agierenden Teams, wenn beispielsweise Aspekte der Interkulturalität hinzukommen. Doch welche Schlüsselqualifikationen werden benötigt und wie werden Personen darauf vorbereitet, die Schlüsselqualifikationen in der Praxis gewinnbringend für den Erfolg eines Projekts einzusetzen? Können Schlüsselqualifikationen erlernt werden? Mit Büchern? In Kursen? Learning by doing? Wie reagiert die Hochschullehre darauf? In Lehrveranstaltungen? Vertieft durch studentische Projekte? Ziel des Workshops ist, die Vermittlung von Schlüsselqualifikationen in der Aus- und Weiterbildung unter die Lupe zu nehmen. Dafür werden wir auch in dem diesjährigen SAG WAS Workshop wieder die beiden Welten – studentische Ausbildung und berufliche Weiterbildung – zusammenbringen, um einen Austausch zu ermöglichen und voneinander zu lernen.

Für den Workshop haben wir Expertinnen und Experten aus Unternehmen und Hochschulen eingeladen, die mit ihren Vorträgen erste Impulse für die Diskussion setzen. Zusätzlich werden wir im Rahmen eines *Open Space* unsere Erfahrungen mit unterschiedlichen Veranstaltungskonzepten austauschen.

Programm- und Organisationskomitee: □

Manfred Baumgärtner, Anecon Software Design und Beratung GmbH, Wien
Prof. Dr. Axel Böttcher, Hochschule München
Prof. Dr. Andreas Spillner, Hochschule Bremen
Prof. Dr. Veronika Thurner, Hochschule München
Prof. Dr. Karin Vosseberg, Hochschule Bremerhaven (Chair)

URL: www.vosseberg.net/sagwas2016/

¹ Hochschule Bremerhaven, FB 2 – SG Informatik und Wirtschaftsinformatik, An der Karlstadt 8,
27568 Bremerhaven, karin.vosseberg@hs-bremerhaven.de

Workshop-Programm

Wir starten den Workshop in der ersten Sitzung mit *SAG WAS dazu, welche allgemeinen Schlüsselqualifikationen in der Firma gefordert sind und wie diese Mitarbeiterinnen und Mitarbeiter vermittelt werden*. Erfahrungen aus Projekten zeigen immer wieder, dass nicht immer fachlich komplexe Anforderungen Projekte zum Scheitern bringen, sondern das Fehlen überfachlicher Kompetenzen das Miteinander im Projekt erschweren und den Projekterfolg verhindern. Christine Regitz von der SAP SE wird uns mitnehmen in ihren Projektalltag und uns ihre Sichtweise auf die Bedeutung von Schlüsselqualifikationen für den Projekterfolg vorstellen. Anschließend wird Renate Weichselbraun von Anecon Software Design und Beratung GmbH aus Wien in ihrem Beitrag „Den Unterschied machen die Menschen – Unternehmenswerte und Kompetenzen abseits des Fachlichen“ die Unternehmensphilosophie in ihrer Firma vorstellen und diskutieren.

In der zweiten Sitzung legen wir den Schwerpunkt auf *SAG WAS dazu, welche allgemeinen Schlüsselqualifikationen in der studentischen Ausbildung wie vermittelt werden*. □ Spätestens seit der Bologna Diskussion steht eine kompetenzorientierte Lehre im Mittelpunkt. Neben fachlichen Kompetenzen werden zunehmend auch auf überfachliche Kompetenzen gesetzt. Yvonne Sedelmaier und Dieter Landes von der Hochschule Coburg stellen in ihrem Beitrag das Zusammenspiel zwischen Erwerb von Fachwissen und beruflichen Kompetenzen in einem aktivierendes Lernsetting für die Software Engineering Ausbildung in Informatik Bachelor- und Masterstudiengängen vor. Welchen Hürden sind Studierende und Lehrende ausgesetzt? Und wie lässt sich die eigene Lehre so gestalten, dass sie die erforderlichen Fach- und Schlüsselkompetenzen auf integrierte Weise erfolgreich vermittelt und Schritt für Schritt entwickelt? Diese Fragen stellen sich Veronika Thurner, Axel Bötcher, Kathrin Schlierkamp und Daniela Zehetmeier in ihrem Beitrag „Nerd++ – Vermitteln von Schlüsselkompetenzen an angehende Softwerker/innen“.

Abschließend werden in einem *Open Space* die beiden Welten – Anforderungen aus den Firmen und die kompetenzorientierte Ausbildung an den Hochschulen – zusammengeführt. Wo liegen die Gemeinsamkeiten? Welche Anforderungen aus den Firmen können seitens der Hochschulen erfüllt oder auch nicht erfüllt werden? Wie können Firmen und Hochschulen zusammenarbeiten, um die allgemeinen Schlüsselkompetenzen von Softwareentwicklerinnen und Softwareentwicklern zu stärken und damit auch das Bild der einsamen Kellerkinder zu verändern.

Schlüsselqualifikationen in der Ausbildung des Software Engineering – ein Dilemma und ein möglicher Ausweg

Yvonne Sedelmaier¹, Dieter Landes¹

Abstract: Schlüsselqualifikationen sind ein vieldiskutiertes berufspädagogisches Konzept, da sie sowohl für den eigenen aber auch für andere Berufe und Lebensbereiche große Bedeutung haben. Der Beitrag ruft eine gängige Charakterisierung von Schlüsselqualifikationen in Erinnerung, nach der es sich dabei oft um "höhere" Fähigkeiten handelt, die Aspekte der Persönlichkeitsbildung beinhalten. Somit sind Schlüsselqualifikationen sehr komplex und in der Hochschulausbildung, neben Kompetenzen und Fachwissen, nur sehr eingeschränkt direkt und isoliert zu adressieren. Der Beitrag argumentiert, dass dies im Zusammenspiel mit beruflichen Kompetenzen und in einem aktivierenden Lernsetting erfolgen sollte. Der Beitrag skizziert einen solchen Ansatz in Form eines Software-Engineering-Projekts, in dem Master- und Bachelorstudierende der Informatik zusammenarbeiten.

Keywords: Schlüsselqualifikationen, Kompetenzen, Hochschullehre, berufliche Bildung

1 Einleitung

Seit den 1970er Jahren sind Schlüsselqualifikationen (SQ) ein vor allem in der beruflichen Bildung viel diskutierter Ansatz. Dieser Begriff bietet großen Spielraum für Definitionen und Interpretationen. Es bleibt oft unklar, ob eine Einstellung wie z.B. Teamgeist, eine Handlungsweise wie z.B. Teamarbeit oder eine Fähigkeit wie Teamfähigkeit oder ein gänzlich anderes Qualifikationselement gemeint ist.

Unter SQ werden alle Fähigkeiten zusammengefasst, die sowohl für den eigenen Beruf aber auch für andere Berufe und Lebensbereiche sehr wichtig sind. Es handelt sich dabei oft um "höhere" Fähigkeiten, die Aspekte der Persönlichkeitsbildung (Teamfähigkeit, theoretisches Denken, Selbstständigkeit, etc.) beinhalten, d.h. sie versetzen Berufstätige in die Lage, den Anforderungen der modernen Arbeitswelt zu entsprechen. „Schlüsselqualifikationen als Berufsqualifikationen sind relativ lange verwertbare funktions- und berufsübergreifende Qualifikationen zum Lösen beruflicher Probleme. Qualifikationsziel ist die berufliche Flexibilität und Mobilität.“ [Wi91, S. 56]

Mertens definiert SQ als „solche Kenntnisse, Fähigkeiten und Fertigkeiten, welche nicht unmittelbaren und begrenzten Bezug zu bestimmten, disparaten praktischen Tätigkeiten erbringen, sondern vielmehr

(a) die Eignung für eine große Zahl von Positionen und Funktionen als alternative

¹ Hochschule für Angewandte Wissenschaften Coburg, Fakultät Elektrotechnik und Informatik, Friedrich-Streib-Str. 2, 96450 Coburg, {yvonne.sedelmaier, dieter.landes}@hs-coburg.de

Optionen zum gleichen Zeitpunkt, und

(b) die Eignung für die Bewältigung einer Sequenz von (meist unvorhersehbaren) Änderungen von Anforderungen im Laufe des Lebens [sind].“ [Me74, S. 40]

Mertens unterscheidet also vier Typen von Schlüsselqualifikationen:

- Basisqualifikationen als Qualifikationen höherer Ordnung oder gemeinsame Dritte von Einzelfähigkeiten:
 - Gebiet der Denkschulung (Logik, analytisches, strukturierendes, konzeptionelles, kreatives, kooperatives Denken, etc.);
 - selbständiges Planen, Durchführen und Kontrollieren;
- Horizont- oder Horizontalqualifikationen, horizonterweiternde Qualifikationen:
 - Wissen über das Wesen von Informationen, über Gewinnung, Verstehen und Verarbeiten von Informationen;
 - Sichern eine möglichst effiziente Nutzung der Informationshorizonte der Menschheit, entweder durch Wissen von Fakten oder durch Wissen über einen raschen Zugriff auf Informationen („gewusst wo“);
- Breiterelemente sind spezielle Kenntnisse und Fertigkeiten, die über breite Felder der Tätigkeitslandschaft nachweislich als praktische Anforderung am Arbeitsplatz auftreten, also so bedeutsam zu sein scheinen, dass sie bereits zur Allgemeinbildung zählen.
- Vintagefaktoren dienen der Aufhebung der durch Innovationen entstandenen Bildungsdifferenzen zwischen den Generationen. Mertens gibt dabei der Erwachsenenbildung einen neuen Stellenwert.

2 Schlüsselqualifikationen in der Hochschulausbildung

Legt man eine dieser gängigen Definitionen von SQ zugrunde, kommt die Hochschulausbildung schnell an ihre Grenzen. Eine wesentliche Herausforderung liegt in der Komplexität der SQ und somit in der Allgemeinheit des Begriffs von SQ. Zabeck beschreibt das Dilemma mit den SQ treffend [Za89]: Werden SQ zu abstrakt formuliert, sind sie nicht umsetz- und transferierbar; werden sie zu differenziert berufsbezogen formuliert, verlieren sie den Charakter jener psychologischen Kategorien, für die sie sich ausgeben.

Um dieser Problematik zu begegnen und sowohl SQ, als auch zusätzlich Kompetenzen für Software Engineering adressieren zu können, wurde an der Hochschule Coburg ein Kompetenzprofil für Software Engineering entwickelt [SL15b, SL15a]. Dieses Kompetenzprofil beschreibt für Software Engineering spezifische überfachliche Kompetenzen, die als Kompass für die Hochschulausbildung dienen. Daraus werden Lehrziele abgeleitet und dem Primat der Didaktik [Kl64] folgend dann Methoden und

Medien in Lehrkonzepten zusammengeführt.

Um die Wirksamkeit dieser Lehrkonzepte zu evaluieren, wurde ein Bewertungsansatz entwickelt und bereits in diversen Lehrveranstaltungen des Software Engineering eingesetzt, der den Kompetenzzuwachs Studierender aufgrund einer Lehrinheit bewertet [SL14a]. Insgesamt zeigt sich, dass aktivierende und induktive Lehrkonzepte für das Training überfachlicher Kompetenzen bzw. SQ zielführend sind [Se16].

3 Beispielhafte Umsetzung in einer Lehrveranstaltung

Eine Lehrveranstaltung, in der überfachliche Kompetenzen eine entscheidende Rolle spielen, ist das Software-Engineering-Projekt [SL14b], das als Abschlussprojekt fungiert. Dabei bearbeiten Studierende ein komplettes Projekt, von der Anforderungserhebung, i.a. mit echten, fachfremden Kunden, bis zur Auslieferung. In den Teams arbeiten überwiegend Bachelorstudierende, geleitet wird jedes Team allerdings durch einen Masterstudierenden. Während die Bachelorstudierenden vorwiegend operative Aufgaben haben, kümmert sich der Masterstudierende um die Anpassung des Vorgehensmodells, sowie das Projekt- und Qualitätsmanagement.

Zum Tragen kommen im Zuge des Projekts auf Masterniveau mit „selbständigem Planen, Durchführen und Kontrollieren“ durchaus SQ im Sinne der Definition von Mertens; auf Bachelorniveau sind allerdings im Wesentlichen überfachliche Kompetenzen aus den Bereichen Angemessenheit, Zielorientierung und Problembewusstsein des Kompetenzprofils SWEBOS [SL14a, SL15a] zu nennen, die einen anderen Charakter als SQ besitzen. Die Lehrenden agieren in dieser Lehrveranstaltung im Hintergrund als Coaches, die Studierenden versuchen im Sinne des aktiven und induktiven Lernens [Pr04, PF06] Aufgaben und Problemstellungen zunächst eigenständig zu verstehen und bewältigen. Verstärkt wird der Lerneffekt zusätzlich dadurch, dass die Studierenden zum Ende des Projekts ihre Zusammenarbeit im Team in einer schriftlichen Selbstreflexion bewerten. Zudem sollen sie über weitere überfachliche kontextsensitive Kompetenzen im Software Engineering reflektieren, die im Projekt besonders relevant sind. Diese Selbstreflexion wird dann in einem gemeinsamen Abschlussgespräch des Teams, einem Post-Mortem-Review, thematisiert.

4 Fazit

Zusammenfassend ist festzuhalten, dass das Konzept der SQ einen sehr hohen Anspruch in die Hochschullehre bringt, da sich SQ nicht einfach vermitteln, sondern bestenfalls trainieren lassen. Da jedoch arbeitsplatzbezogenes Faktenwissen immer schneller veraltet, kommt in modernen Gesellschaften Bildungsinhalten mit höherem Abstraktionsniveau besondere Bedeutung zu. Folglich muss ein Studium neben Fachwissen auch allgemeine und kontext-sensitive überfachliche Kompetenzen i.V.m. SQ adressieren.

Danksagung

Das Projekt EVELIN wird vom Bundesministerium für Bildung und Forschung unter dem Förderkennzeichen 01PL12022A im Rahmen des Qualitätspakts Lehre gefördert.

Literaturverzeichnis

- [Kl64] Klafki, W.: Didaktische Analyse als Kern der Unterrichtsvorbereitung. In (Roth, H.; Blumenthal, A. Hrsg.): Didaktische Analyse. Schroedel, Hannover; S. 5–34, 1964.
- [Me74] Mertens, D.: Schlüsselqualifikationen. Thesen zur Schulung für eine moderne Gesellschaft. In Mitteilungen aus der Arbeitsmarkt- und Berufsforschung, 7; S. 36–43, 1974.
- [PF06] Prince, M. J.; Felder, R. M.: Inductive Teaching and Learning Methods: Definitions, Comparisons, and Research Bases. In Journal of Engineering Education, 95; S. 123–138, 2006.
- [Pr04] Prince, M. J.: Does Active Learning Work? A Review of the Research. In Journal of Engineering Education, 93; S. 223–231, 2004.
- [Se16] Sedelmaier, Y.: Interdisziplinäre Fachdidaktik für Software Engineering – Forschungsbasierte Entwicklung und Evaluation eines anwendungsbezogenen didaktischen Ansatzes. opus, Bamberg, 2016.
- [SL14a] Sedelmaier, Y.; Landes, D.: A Multi-Perspective Framework for Evaluating Software Engineering Education by Assessing Students' Competencies. SECAT - A Software Engineering Competency Assessment Tool. In (IEEE Hrsg.): 44th Frontiers in Education (FIE); S. 2065–2072, 2014.
- [SL14b] Sedelmaier, Y.; Landes, D.: Practicing Soft Skills in Software Engineering. In (Yu, L. Hrsg.): Overcoming Challenges in Software Engineering Education. IGI Global; S. 161–179, 2014.
- [SL15a] Sedelmaier, Y.; Landes, D.: Überfachliche Kompetenz im Software Engineering - Modellierung, Förderung und Messung in der Hochschulausbildung. In (Riegel, U. et al. Hrsg.): Kompetenzmodellierung und -messung in den Fachdidaktiken. Waxmann, Münster; S. 111–130, 2015.
- [SL15b] Sedelmaier, Y.; Landes, D.: SWEBOS - The Software Engineering Body of Skills. In International Journal of Engineering Pedagogy (iJEP), 5; S. 12–19, 2015.
- [Wi91] Wilsdorf, D.: Schlüsselqualifikationen. Die Entwicklung selbständigen Lernens und Handelns in der industriellen gewerblichen Berufsausbildung. Lexika Verlag, München, 1991.
- [Za89] Zabeck, J.: "Schlüsselqualifikationen" - Zur Kritik einer didaktischen Zielformel. In Wirtschaft und Erziehung, 41; S. 77–86, 1989.

Nerd++ – Vermitteln von Schlüsselkompetenzen an angehende Softwerker/innen

Veronika Thurner¹ Axel Böttcher¹ Kathrin Schlierkamp¹ Daniela Zehetmeier¹

Abstract: Komplexe Softwaresystem sind heute nicht mehr das Werk einsamer Keller-Nerds, sondern das Ergebnis von interdisziplinärer (und oft auch internationaler) Teamarbeit. Um für diese Arbeitsweise gerüstet zu sein benötigen angehende Software Professionals neben einer soliden fachlichen Ausbildung auch gut ausgeprägte überfachliche Schlüsselkompetenzen, wie beispielsweise Team- und Kommunikationsfähigkeit. Eine Hochschulbildung, die den Anspruch erhebt, berufsbefähigend zu sein, muss entsprechend sowohl fachliche Kompetenzen als auch überfachliche Schlüsselkompetenzen in den Studierenden entwickeln.

Keywords: Kompetenz, Kompetenzmodell, Schlüsselkompetenz, Lehre, Informatik-Studium

1 Motivation

Die Zeiten, in denen der Erfolg eines Softwareentwicklers allein am Umgang mit Code und Technik gemessen wird, sind vorbei. Selbst in den Kellern der IT-Nerds sind nun Stichworte wie Soft Skills, Teamfähigkeit und Selbstmanagement angekommen. Überfachliche Schlüsselkompetenzen erhalten daher einen immer höheren Stellenwert, wenn es um die eigene fachliche Performance geht [VSM11, Si11]. Die Orte, an denen die fachlichen mit den überfachlichen Kompetenzen verbunden, gefördert und gefordert werden, sind allerdings nach wie vor die gleichen: die Hochschulen und Universitäten.

Im Zuge des 1999 begonnenen Bologna-Reform-Prozesses stellte die Politik drei Grundforderungen an die europäischen Hochschulen:

- Förderung der Beschäftigungsfähigkeit der Studierenden,
- ein dadurch zu erzielender höherer Output an hochqualifizierten Arbeitskräften
- sowie eine neue Lehr-Lern-Kultur, die sich an den spezifischen Bedürfnissen der Studierenden orientiert.

Die Lehre an den Hochschulen soll demgemäß die Studierenden in ihr Zentrum stellen, sie aktivieren und vor allem ihre Kompetenzen fordern und gezielt fördern [Ni11].

Welche Konsequenzen hat das nun für die Lehre im Jahr 2016? Welchen Hürden sind Studierende und Lehrende ausgesetzt? Und wie lässt sich die eigene Lehre so gestalten, dass sie die erforderlichen Fach- und Schlüsselkompetenzen auf integrierte Weise erfolgreich vermittelt und Schritt für Schritt in den Studierenden entwickelt?

¹Hochschule München, Fakultät für Informatik und Mathematik, Lothstraße 64, 80335 München, <vorname>.<nachname>@hm.edu; Gefördert durch das BMBF Förderkennzeichen 01PL11025 (Projekt „Für die Zukunft gerüstet“), im Programm „Qualitätspakt Lehre“

2 Grundbegriffe

Der Begriff *Kompetenz* bedeutet seinem lateinischen Ursprung nach *zusammentreffen*. Von dieser Bedeutung ausgehend wird eine Person dann als *kompetent* bezeichnet, wenn sich in ihr ausreichende Fertigkeiten und Fähigkeiten treffen, die es ihr erlauben mit einer bestimmten Situation erfolgreich umzugehen. Kompetenz an sich meint aber noch mehr als nur den gezielten Einsatz von erlernten Fertigkeiten – ob eine Person sich wirklich kompetent verhalten kann, hängt zudem von ihrer persönlichen Bereitschaft, Fähigkeit und Zuständigkeit ab. Wollersheim spricht daher vom Konstrukt “Kompetenz”, das sowohl motivationale, kognitive als auch soziale bzw. organisationale Facetten vereint [Wo14].

Eine der bekanntesten Definitionen von Kompetenz findet sich bei Weinert: Hier wird Kompetenz ganz allgemein als verfügbare bzw. erlernbare kognitive Fertig- und Fähigkeit beschrieben, die es einem Individuum ermöglicht Probleme zu lösen: “Dabei versteht man unter Kompetenzen die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können” (siehe [We02], S. 27f). In Bezug auf den schulischen Kontext unterscheidet Weinert zudem noch zwischen fachlichen Kompetenzen, fachübergreifenden Kompetenzen sowie Handlungskompetenzen. Weinert betont, dass diese Kompetenzen nicht nur im schulischen Kontext notwendig sind, sondern auch im restlichen Lebensalltag und somit die Voraussetzungen für ein gutes und erfolgreiches Leben darstellen.

Schlüsselkompetenzen bezeichnen ferner diejenigen Kompetenzen, die die fachlichen Fähigkeiten so ergänzen, dass eine Person damit die eigenen Bedürfnisse erfüllen kann, in der Lage ist mit anderen in Gemeinschaft zu leben sowie einer nützlichen und einkommenssichernden Arbeit nachgehen kann [IdSJ09].

3 Kompetenzmodelle

Die meisten Kompetenzmodelle ordnen die einzelnen Kompetenzen in die Bereiche *Fach-, Methoden-, Sozial- und Selbstkompetenzen* ein [Pa11]. Je nach Fachdisziplin und Einsatzbereich eines Kompetenzmodells lassen sich hier zahlreiche Varianten finden.

Weicker et al. haben beispielsweise in ihrem Modell für Schlüsselkompetenzen in der Informatik neben den Methoden-, Fach- und Selbstkompetenzen auch die Kategorien *Allgemeine Informatikkompetenzen* (z. B. Formalisierungsfähigkeit) sowie *Wissenschaftskompetenzen* (z. B. wissenschaftliches Arbeiten) mit aufgenommen [Do12]. Calchera und Weber verfolgen einen anderen Ansatz: In ihrem Kompetenzmodell für die Firma AEG verzichten sie bewusst auf die Einteilung in personale, fachliche und soziale Kompetenzen. Stattdessen verwenden sie ein so genanntes Schalenmodell, dessen Kern aus affektiven Kompetenzen, die das Gefühlsleben betreffen, besteht und der von den kognitiven sowie beruflichen Kompetenzen umschlossen wird [Do12].

Das Kompetenzmodell der TU München für Lehrende im Hochschulbereich ist ein weiteres Beispiel dafür, wie unterschiedlich differenziert Kompetenzmodelle sein können. Fleischmann leitet aus der persönlichen Lehrerfahrung sowie dem Fachwissen der Psychologie und Pädagogik ein Modell ab, das lediglich die vier Bereiche “Entwickeln von Lehr- und Prüfungskonzepten, Umsetzen von Lehr- und Prüfungskonzepten, Organisieren von Lehre sowie Reflektieren und Weiterentwickeln von Lehre” enthält [FJS14].

Weitere prominente Beispiele von Kompetenzmodellen finden sich beispielsweise bei Erpenbeck, der einen Ansatz einer Kompetenzbiographie verfolgt [EH07, EvR07, Er04].

4 Problemstellung

Unsere Lehrerfahrung zeigt, dass sich viele Studierende schwer tun, die gemäß der Bloom’schen Lernzieltaxonomie [AKB01] höheren Kompetenzebenen zu erreichen, und zwar weitgehend unabhängig vom konkreten fachlichen Inhalt. Die Vermutung liegt daher nahe, dass es nicht allein fachliche Defizite sind, die ein Erreichen der gewünschten höheren Kompetenzebenen erschweren, sondern vielmehr bestimmte Schlüsselkompetenzen für diesen Entwicklungsschritt fehlen [Ze15]. Entsprechend müssen diese Schlüsselkompetenzen erst in ausreichendem Maße in den Studierenden entwickelt werden, um die Voraussetzungen dafür zu schaffen, dass sie die höheren Ebenen in ihren Fachkompetenzen überhaupt erreichen können. Ein Beispiel dafür ist die Fähigkeit des kritischen Hinterfragens, welche eine der zentralen Grundvoraussetzungen für Kompetenzen auf Ebene 5, “Evaluieren”, gemäß der Bloom’schen Taxonomie darstellt.

Schlüsselkompetenzen sind jedoch nicht nur für den erfolgreichen Studierprozess an sich, sondern insbesondere auch für das spätere Berufsleben von zentraler Bedeutung. Hierzu zählt insbesondere die Teamfähigkeit sowie ein gesundes Wechselspiel zwischen gegenpoligen Kompetenzen wie Durchsetzungskraft und Kompromissfähigkeit, oder Selbstbewusstsein und Selbstkritikalität. Grundlage für einen (lebenslangen) Lernprozess sind darüber hinaus Selbstreflexion und Kritikfähigkeit, sowohl in der Rolle des Feedback-Gebers als auch des Feedback-Nehmers. Nur die wenigsten Studierenden bringen diese Fähigkeiten beim Eintritt in die Hochschule in ausreichendem Maße mit. Da diese Fähigkeiten jedoch spätestens in der Berufstätigkeit gebraucht werden, müssen sie bis dahin entwickelt werden – also in der Zeit des Hochschulstudiums.

5 Lösungsansätze

Um gezielt die jeweiligen Entwicklungsbedarfe der Studierenden zu adressieren ist es hilfreich, zunächst ein Soll-Profil für Absolventinnen und Absolventen zu definieren, das die Schlüsselkompetenzen definiert, die von ihnen im Berufsleben erwartet werden. Neben den Erfahrungen aus der eigenen Industriezeit sind hier die Arbeitgeber aus der Industrie eine wichtige Informationsquelle, ebenso wie Profildefinitionen und Kompetenzrichtlinien einschlägiger Verbände [TBK14, Eu, Do12, Er04, IE13, IdSJ09]. Welche Fähigkeiten die

essenziell erforderlichen Schlüsselkompetenzen darstellen ist dabei abhängig vom jeweiligen Fachgebiet. Beispielsweise ist die Fähigkeit zum abstrakten Denken in der Informatik von zentraler Bedeutung, während sie in anderen Disziplinen wie z. B. sozialer Arbeit weniger wichtig ist.

In einem nächsten Schritt werden die Eingangskompetenzen der Studienanfänger/innen erfasst und den von den Arbeitgebern geäußerten Erwartungen gegenüber gestellt. Aus diesem Delta ergibt sich der Entwicklungsbedarf, der im Idealfall während des Studiums hinsichtlich der Schlüsselkompetenzen (und zusätzlich zum fachlichen Kompetenzerwerb) zu adressieren ist, um die Studierenden tatsächlich berufsbefähigend auszubilden [Ze14].

Nach diesen eher zieldefinierenden und diagnostischen Schritten folgt dann erst der eigentliche Teil der "Lehrarbeit", d. h. die identifizierten Kompetenzbedarfe in den Studierenden zu entwickeln. Traditionelle Lehransätze wie die klassische Vorlesung in Kombination mit theorielastiger Übung oder Hausaufgabe in Einzelarbeit sind dazu in der Regel nur bedingt geeignet. Entsprechend ist es bis zu einem gewissen Grad notwendig, geeignete Lehransätze zu entwickeln und umzusetzen, die Fach- und Schlüsselkompetenzen auf integrierte Weise in den Studierenden fördern.

6 Herausforderungen

Auch wenn viele Lehrende die unzureichenden Schlüsselkompetenzen ihrer Schützlinge beklagen, sehen nur wenige deren Entwicklung als Teil ihrer Lehraufgabe an. Ein gängiger Ansatz ist daher, die Förderung von Schlüsselkompetenzen auszulagern in Zusatzangeboten, die beispielsweise vom Career Center der Hochschule durchgeführt werden. Ein Problem dieser Angebote liegt jedoch darin, dass diese zusätzlich und außerhalb des eigentlichen Curriculums statt finden. Diejenigen Studierenden, die diese Angebote wirklich benötigen würden, finden daher oft überhaupt nicht zu diesen Angeboten hin, beispielsweise weil ihre Selbstreflexion nicht reicht, um überhaupt den Bedarf zu erkennen, die Eigeninitiative fehlt, um sich anzumelden und sie wegen mangelndem Zeitmanagement oder geringer Selbstdisziplin dann doch nicht hingehen.

Sinnvoll wäre also eine integrierte Entwicklung von Fach- und Schlüsselkompetenzen innerhalb fachlicher Pflichtveranstaltungen. Viele Lehrende sehen ihre Hauptaufgabe jedoch vor allem in der Vermittlung von Fachwissen und der Förderung der dazugehörigen Fachkompetenzen. Einige gehen davon aus, dass Studierende die gängigsten Sozialkompetenzen und weitere Soft Skills sowieso aus der Schule und/oder Ausbildung mitbringen. Für einige Studierende trifft das auch zu, doch die Erfahrung zeigt, dass ein großer Anteil der Studierenden hier noch Entwicklungsbedarf aufweist. Beispielsweise ist die Eigenverantwortlichkeit für das an einer Hochschule notwendige Zeit- und Selbstmanagement nicht mit dem der Schulzeit vergleichbar. Auch vielen Studierenden, die sich fachlich schwer tun, ist nicht bewusst, dass die Ursachen ihrer fachlichen Schwierigkeiten gar nicht in der Fachlichkeit, sondern vielmehr in überfachlichen Schlüsselkompetenzen begründet liegen. Entsprechend ist es notwendig, bei Lehrenden und Studierenden gleichermaßen das Bewusstsein zu schaffen für die Relevanz der erforderlichen Schlüsselkompetenzen, nicht nur für die spätere Berufstätigkeit, sondern auch für den Studierprozess an sich.

Beim Versuch, zusätzlich zu den fachlichen auch die überfachlichen Kompetenzen in der eigenen Lehre zu vermitteln stellt man schnell fest, dass die Menge der pro Zeiteinheit erlernbaren Fähigkeiten nicht beliebig nach oben erweiterbar ist. Es funktioniert also nur bedingt, auf die meist eh schon üppig bemessenen fachlichen Inhalte die überfachlichen einfach noch oben drauf zu satteln und dafür das Tempo zu erhöhen. Statt dessen ist es notwendig, sich auf Wesentliches zu konzentrieren, die Intensität des Diskurses zu vertiefen und dabei die Schlüsselkompetenzen mit zu adressieren – und dafür an anderen Stellen weniger Relevantes zu streichen oder in die Selbstlernzeit der Studierenden auszulagern.

Um die Studierenden nicht nur überwiegend kognitiv, sondern ganzheitlich zu fordern und ihre Schlüsselkompetenzen mit zu entwickeln sind neuartige Lehr-/Lernkonzepte erforderlich, die vielfältige Fähigkeiten ansprechen und Reflexionsprozesse systematisch unterstützen. Diese zu konzipieren, an die konkreten Bedarfe der jeweiligen Kohorte anzupassen und schließlich die erforderlichen Materialien vorzubereiten braucht Zeit, die während des normalen Semesterbetriebes jedoch in der Regel knapp bemessen ist. Auch wenn eine Lehrperson bereit und fähig ist, überfachliche Schlüsselkompetenzen gezielt in ihre fachliche Lehre zu integrieren, ist die Entwicklung geeigneter Formate also eine echte Herausforderung.

7 Fazit

Die integrierte Entwicklung von Fach- und Schlüsselkompetenzen in fachlichen Pflichtveranstaltungen ist möglich, erfordert aber sowohl von den Lehrenden als auch von den Studierenden einen gewissen Einsatz. Lehrende brauchen dazu insbesondere ein gewisses Maß an Selbstreflexion. Sie sind Experten ihres Fachgebietes, haben ihr Studium erfolgreich abgeschlossen und profitieren in der Regel von mehrjähriger Industrieerfahrung. Viele der auf diesem Werdegang erworbenen Schlüsselkompetenzen nehmen sie daher gar nicht mehr bewusst wahr, sondern sehen sie schlichtweg als gegeben und selbstverständlich an. Hier gilt es ein Bewusstsein für die Kompetenzen des eigenen Fachbereichs zu schaffen und zu schärfen, bevor eben jene Kompetenzen gezielt durch die Lehre bei den Studierenden gefördert werden können.

Für die Studierenden bedeutet die integrierte Entwicklung von Fach- und Schlüsselkompetenzen während des Studiums eine Intensivierung des Lernprozesses sowie eine Ausweitung desselben über die Grenzen der rein kognitiven Fähigkeiten hinaus bis hin zur eigenen Persönlichkeit. Das kann durchaus sehr fordernd sein, insbesondere wenn in der Startphase dieser Prozess noch ungewohnt und vielleicht sogar beängstigend ist. Hier ist eine vertrauensvolle Lern- und Betreuungsumgebung entscheidende Voraussetzung dafür, dass die Studierenden sich auf diesen Prozess einlassen und ihn für sich erfolgreich beschreiten können.

Literaturverzeichnis

- [AKB01] Anderson, Lorin W; Krathwohl, David R; Bloom, Benjamin Samuel: A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives. Longman, New York, 1. Auflage, 2001.

- [Do12] Doerge, Ch.: Informatische Schlüsselkompetenzen – Konzepte der Informationstechnologie im Sinne einer informatischen Allgemeinbildung (in German). Dissertation, University of Oldenburg, 2012.
- [EH07] Erpenbeck, J.; Heyse, V.: Die Kompetenzbiographie. Waxmann Verlag GmbH, 2. Auflage, 2007.
- [Er04] Erpenbeck, J.: , Kompetenzdiagnostik und Entwicklung, March 2004.
- [Eu] Europäische Kommission, Generaldirektion Bildung und Kultur: , Schlüsselkompetenzen für lebensbegleitendes Lernen – Ein europäischer Referenzrahmen (in German).
- [EvR07] Erpenbeck, J.; von Rosenstiel, L., Hrsg. Handbuch Kompetenzmessung. Schäffer-Poeschel Verlag Stuttgart, 2007.
- [FJS14] Fleischmann, A.; Jäger, C.; Strasser, A.: , Kompetenzmodell Hochschullehre, 2014.
- [IdSJ09] In der Smitten, Susanne; Jaeger, Michael, Hrsg. Studentischer Kompetenzerwerb im Kontext von Hochschulsteuerung und Profilbildung – Dokumentation zur HIS-Tagung am 03. November 2009 in Hannover (in German). Hochschul-Informations-System GmbH (HIS), Hannover, 2009.
- [IE13] IEEE: , Software Engineering Body of Knowledge, V3, 2013.
- [Ni11] Nickel, Sigrun: , Der Bologna-Prozess aus Sicht der Hochschulforschung. Arbeitspapier Nr. 148, September 2011.
- [Pa11] Paetz, N.-V.: Kompetenz in der Hochschuldidaktik. VS Verlag für Sozialwissenschaften, 2011.
- [Si11] Siwon, Peter: Die menschliche Seite des Projekterfolgs: was Softwerker über (verborgene) Denkautomatismen und -modelle in der Projektarbeit wissen müssen. dpunkt-Verlag, 2011.
- [TBK14] Thurner, Veronika; Böttcher, Axel C.H.; Kämper, Andreas: Identifying Base Competencies as Prerequisites for Software Engineering Education. In: Global Engineering Education Conference (EDUCON), 2014 IEEE. S. 1069–1076, April 2014.
- [VSM11] Vogenschow, U.; Schneider, B.; Meyrose, I.: Soft Skills für Softwareentwickler: Frage-techniken, Konfliktmanagement, Kommunikationstypen und -modelle. dpunkt-Verlag, 2011.
- [We02] Weinert, Franz E.: Leistungsmessungen in Schulen. Beltz Verlag, 2. Auflage, 2002.
- [Wo14] Wollersheim, H.-W.: Entwicklung von Schlüsselkompetenzen an Hochschulen. In: Aufbruch in die Zukunft, S. S. 448–461. Waxmann, 2014.
- [Ze14] Zehetmeier, Daniela; Kuhrmann, Marco; Böttcher, Axel; Schlierkamp, Kathrin; Thurner, Veronika: Self-Assessment of Freshmen Students’ Base Competencies. In: Global Engineering Education Conference (EDUCON), 2014 IEEE. S. 429–438, April 2014.
- [Ze15] Zehetmeier, Daniela; Böttcher, Axel; Brüggemann-Klein, Anne; Thurner, Veronika: Development of a Classification Scheme for Errors Observed in the Process of Computer Programming Education. In: International Conference on Higher Education Advances (HEAd). 2015.

Teilhabe an der allgegenwärtigen Kommunikation

Teilhabe an der allgemeinen Kommunikation

Henning Lübbecke¹

Vorwort

Zukünftig wird uns Informations- und Kommunikationstechnik überall umgeben und Arbeit, Bildung, Teilhabe, Kommunikation, Informationsgewinnung, Mobilität, Wohnen, Einkaufen etc. prägen. Wie können diese Technologien (in Zeiten des Ubiquitous Computing und des Internet der Dinge) die gleichberechtigte Teilhabe aller an der Gesellschaft sicherstellen bzw. ermöglichen?

Bei der Implementation von Webauftritten gewinnt die Berücksichtigung der Barrierefreiheitsprinzipien aufgrund zunehmender mobiler Internetnutzung eine weiterhin steigende Bedeutung. Von unterschiedlichen internationalen Organisationen, wie z. B. der Web Accessibility Initiative, wurden Barrierefreiheitsrichtlinien entwickelt, darunter die WCAG 2.0, die als Grundlage für Entwickler, aber auch für Behörden dienen und im Bereich Mobilität erweitert werden. In etlichen Ländern wurden diese Regelungen bereits vom Gesetzgeber implementiert, um Barrierefreiheit in bestimmten Bereichen (z.B. Internetauftritte von Behörden) zu gewährleisten. In diesem Zusammenhang analysieren Taisiia Makarova, Andreas Mladenow und Christine Strauss den aktuellen Forschungsstand und stellen die Ergebnisse einer empirischen Untersuchung (Korrelationsanalyse nach Spearman) vor, die einem möglichen Zusammenhang zwischen dem Barrierefreiheitsniveau und dem Suchmaschinenranking am Beispiel österreichischer Regierungsseiten nachgeht. Zur Evaluierung der Webauftritte wurden die WCAG 2.0 Erfolgsfaktoren (Konformitätsstufen AAA, AA und A) herangezogen. Der Beitrag bietet eine Grundlage für weitere Forschung auf diesem Gebiet und soll den Nutzen der Berücksichtigung von Barrierefreiheit im Internet für mehrere Nutzer- und Entwicklergruppen verdeutlichen.

Der Beitrag von Nicole Najemnik und Isabel Zorn fokussiert auf Kinder mit Behinderung im Kindergarten und zeigt auf, wie diese sowohl durch mangelnde Medienkompetenz von pädagogischen Fachkräften, als auch durch einseitiges Design von Assistenztechnologien von EntwicklerInnen behindert werden. Dafür werden Erkenntnisse aus internetsoziologischer Forschung zu (digitaler) Ungleichheit und das Konzept des Doing Disability aus den Critical Disability Studies herangezogen. Darauf aufbauend wird die Rolle (assistiver) Technologien für inklusive Medienbildung diskutiert und das Potenzial digitaler Medien für Partizipation und Teilhabe von Kindern mit Behinderung anhand aktueller Forschungsergebnisse verdeutlicht. Abschließend

¹ Arbeitskreis Inklusion in Social Media, Fachbereich Informatik und Gesellschaft,
henning.luebbecke@privatbaz.bund.de

werden Handlungsempfehlungen für die pädagogische und informationstechnische Praxis formuliert und gezeigt, wie Doing-Disability-Prozesse durch adäquates Design in der Informatik und adäquate Kompetenzvermittlung in der Pädagogik verringert werden können.

Inklusive und barrierefreie Kommunikation in mitgliedsorientierten Organisationen beinhaltet die Vorbereitung, Durchführung und Nachbereitung von Veranstaltungen, sowie die digitale Barrierefreiheit von Dokumenten und Präsentationen. Das Positionspapier von Gottfried Zimmermann, Mareen Gutscher, Jonas Häfele Lena Krächan, Damaris Rothfuß, Hina Marquart und Hermann Frank berichtet von einem gemeinsamen Projekt des Paritätischen Baden-Württemberg und der Hochschule der Medien in Stuttgart, bei dem ein Handbuch für barrierefreie Kommunikation entstehen soll. Ein besonderer Fokus wird dabei die Berücksichtigung neuerer Medien-Technologien für die Kommunikation sein.

Dieter Meiller, Karlheinz Müller, Hildegard Legat und Bernd Gerlang stellen die Rolle der Hochschule bei der Vernetzung der barrierefreien Bildung und Freizeit im regionalen Raum in ihrem Beitrag dar. Die „Offene Behindertenarbeit“ schließt sich mit örtlichen und regionalen Bildungsträgern, mit dem örtlichen Fernsehsender und der nahe gelegenen Hochschule zusammen, um grundlegende und strukturelle Verbesserungen im Bereich der Information über geeignete Kurse und Angebote voranzubringen. Die Arbeitsgruppe der Hochschule fokussiert sich dabei auf die Beratung und Entwicklung im Bereich barrierefreies Web.

Blinde greifen auf die mobile Webseite von Sozialen Netzwerken zurück, weil die reguläre Webseite aufgrund von hoher Komplexität und Inkonsistenz unzugänglich ist. Dadurch steht ihnen nicht der gleiche Informations- und Funktionsumfang zu Verfügung. Doch was genau veranlasst Blinde einen bestimmten Zugangsmodus zu wählen bzw. zu meiden? Diese Fragestellung untersucht Petra Gröber im ihrem Beitrag und beschreibt die theoretische Herleitung eines auf der IS Success-Forschung basierenden Modells zur Erfolgsmessung der Zugangsmodi von Sozialen Netzwerken (reguläre Webseite, mobile Webseite, App).

Auf der Grundlage einer Literaturstudie und von 5 Design- und Entwicklungsprojekten dreht Svetlana Matiouk die üblichen normativen Annahmen über Taubheit, die von der Sicht auf Defizite, Schwierigkeiten und Fähigkeiten geprägt sind, zur Expertise für Räumliches, Eingebettetes und Visuelles. In einer Reihe von Konzepten entstehen sogenannte „Inspiration Rooms“.

Zwei interaktive Würfel werden im Beitrag von Arne Berger, Andreas Bischof, Kevin Lefeuvre, Albrecht Kurze, Michael Storz und Sören Totzauer vorgestellt. Sie dienen als Werkzeug zum gemeinsamen Entwerfen mit Co-Designern und zum erforschen des Designspace von Smart Connected Products. Dazu wird sehr kurz Research Through Design als drittes Paradigma der HCI eingeführt, die Würfel als interaktive Werkzeuge vorgestellt und eine Case Study ihrer Verwendung in einem Co-Design Workshop mit blinden und sehbehinderten Co-Designern vorgestellt.

Die derzeit gültigen Grenzwerte zur Auslösung photosensitiver Epilepsie basieren auf Daten aus dem Jahr 1975. Die Versuche dazu wurden mit Röhrenmonitoren durchgeführt. Diese besitzen im Gegensatz zu den heute gebräuchlichen Bildschirmen eine Bildwiederholfrequenz, so dass bei der Darstellung bewegter Bilder Interferenzen zwischen den berechneten Einzelbildern und der Eigenfrequenz des Monitors entstehen. Um die tatsächlich durch das Auge wahrgenommene Frequenz zu ermitteln, wurde eine Testvorrichtung erstellt. Hierbei ergaben sich mehrere technische Probleme, deren Lösungen hier ebenfalls in Ausschnitten präsentiert werden. Mathias Haimerl stellt in seinem Beitrag die Planung der technischen Umsetzung ohne Messergebnisse oder Auswertungen vor.

Barrierefreiheit im Internet und Suchmaschinenranking – eine empirische Untersuchung

Taisiia Makarova¹, Andreas Mladenow¹, Christine Strauss¹

Abstract: Bei der Implementation von Webauftritten gewinnt die Berücksichtigung der Barrierefreiheitsprinzipien aufgrund zunehmender mobiler Internetnutzung eine weiterhin steigende Bedeutung. Von unterschiedlichen internationalen Organisationen, wie z. B. der Web Accessibility Initiative, wurden Barrierefreiheitsrichtlinien entwickelt, darunter die WCAG 2.0, die als Grundlage für Entwickler, aber auch für Behörden dienen und im Bereich Mobilität erweitert werden. In etlichen Ländern wurden diese Regelungen bereits vom Gesetzgeber implementiert, um Barrierefreiheit in bestimmten Bereichen (z.B. Internetauftritte von Behörden) zu gewährleisten. In diesem Zusammenhang analysiert der vorliegende Beitrag den aktuellen Forschungsstand und führt eine empirische Untersuchung (Korrelationsanalyse nach Spearman) durch, die einem möglichen Zusammenhang zwischen dem Barrierefreiheitsniveau und dem Suchmaschinenranking am Beispiel österreichischer Regierungsseiten nachgeht. Zur Evaluierung der Webauftritte wurden die WCAG 2.0 Erfolgsfaktoren (Konformitätsstufen AAA, AA und A) herangezogen. Für den Erfolg von Barrierefreiheit im Internet sind bereits bei der Entwicklung einer App bzw. eines Webauftritts die Richtlinien zu beachten und die Webinhalte (z.B. Typ, Größe und Komplexität) so aufzubauen, dass sie barrierefrei verwendet werden können. Der Beitrag bietet eine Grundlage für weitere Forschung auf diesem Gebiet und soll den Nutzen der Berücksichtigung von Barrierefreiheit im Internet für mehrere Nutzer- und Entwicklergruppen verdeutlichen.

Keywords: Barrierefreiheit, Web Accessibility, WCAG 2.0, WCAG 2.0 Konformitätsstufen, WCAG 2.0 Erfolgsfaktoren, WCAG 2.0 extensions, WAB-Koeffizient, Korrelationsanalyse, Barrierefreiheitsbewertung, Verbesserung der Barrierefreiheit, Suchmaschinenranking, SEO, POUR-Prinzip, Mobilität, Collaborative CSCW, Social Crowd Integration.

1 Einleitung

Mit der verstärkten Nutzung des mobilen Internets sind alle Nutzer durch kleinere Bildschirme, eine oftmals langsamere Internetverbindung und unterschiedliche operative Modalitäten mit Einschränkungen konfrontiert [Ha10]. Die Inhalte, die auf dem kleinen Bildschirm eines mobilen Gerätes gezeigt werden, müssen vorweg umgeordnet werden, damit sie lesbar sind. Dies ist jenem Prinzip ähnlich, nach dem die Bildschirmleser für blinde oder in ihrem Sehvermögen eingeschränkte Nutzer funktionieren: erst nach einer Neuordnung der Informationen werden diese vorgelesen. Obwohl festzustellen ist, dass Menschen mit Einschränkungen das Internet in den letzten Jahren intensiver nutzen können als zuvor, müssen sie dennoch zahlreiche Online-Barrieren überwinden (z.B.

¹ Universität Wien, Fakultät für Wirtschaftswissenschaften, Fachbereich eBusiness, Oskar-Morgensternplatz 1, 1090 Wien, {taisiia.makarova, andreas.mladenow, [christine.strauss](mailto:christine.strauss@univie.ac.at)}@univie.ac.at

fehlende oder mangelhafte Alternativtexte bei grafischen Inhalten, automatische Überleitungen oder Pop-Up-Fenster). Um Barrierefreiheit im Internet zu verbessern, wurden in vielen Ländern gesetzliche Rahmenbedingungen etabliert. Der Umstand, dass die Art und Weise, wie Suchmaschinen Internetseiten finden, der Internetnutzung von Menschen mit Behinderungen, die beispielsweise Bildschirmlesegeräte verwenden, stark ähnelt, bildet den Ausgangspunkt der vorliegenden Untersuchung. Grundsätzlich erschweren die zuvor genannten Mängel in der Barrierefreiheit den Vorgang der Indizierung von Internetseiten, der eine zentrale Funktion bei Suchmaschinen inne hat [Ha10]. Einige Arbeiten befassen sich zwar mit dem Zusammenhang von Barrierefreiheitsniveau und Suchmaschinenrating [MM13, EA11, Ha10, ZP03], jedoch gibt es bisher keinen empirischen Beleg.

Die vorliegende Studie analysiert einen möglichen Zusammenhang zwischen den beiden Bereichen Barrierefreiheit und Suchmaschinenranking anhand empirischer Daten und untersucht systematisch ausgewählte Internetauftritte österreichischer Behörden. Der Beitrag behandelt im nächsten Kapitel barrierefreie Webauftritte von Regierungen, Umsetzungsmaßnahmen mit Fokus auf mobile Endgeräte und die Überschneidungen WCAG 2.0 Erfolgsfaktoren und search engine optimization (SEO). Im Kernkapitel 3 wird die empirische Untersuchung zur Barrierefreiheit im Internet und dem Suchmaschinenranking durchgeführt und deren Ergebnisse beschrieben und interpretiert. Die Diskussion, ein Ausblick auf weitere mögliche Forschungsfragen (Kapitel 4) und die Zusammenfassung (Kapitel 5) bilden den Abschluss dieses Beitrags.

2 Barrierefreiheit und Suchmaschinenranking

2.1 Barrierefreie Webauftritte von Regierungsseiten

Seit 2016 müssen in Österreich nicht nur Güter und Dienstleistungen barrierefrei (ohne fremde Hilfe erreichbar) sein, sondern auch Internetauftritte von Behörden. Die gesetzlichen Grundlagen für Umsetzung und Gewährleistung der Barrierefreiheit im Internet wurden auf EU-Ebene geschaffen [Ec16]. So wurden die Kategorien der Webseiten bestimmt, darunter eben die Internetauftritte der Regierungen und sonstiger sozialer Einrichtungen und deren Dienste. Untersuchungen im Jahr 2012 haben gezeigt, dass der Anteil der barrierefreien Internetauftritte an den genannten Seiten sich auf nur knapp 30% belief [Ec16] und im Jahr 2014 noch kein untersuchter Webauftritt die WCAG 2.0 AA Norm erfüllt hat [Ce16]. In der österreichischen Gesetzgebung finden sich drei Dokumente, welche die Barrierefreiheit im Internet regulieren: das Bundes-Verfassungsgesetz, das Bundes-Behindertengleichstellungsgesetz (BGStG) und das E-Government-Gesetz (EGovG). Die Erhebung des Bundeskanzleramtes zum Thema "Barrierefreiheit im Internet" aus dem Jahr 2007 hat wesentliches Verbesserungspotenzial bei der Barrierefreiheit von behördlichen Internetauftritten in Österreich gezeigt. Verbesserungspotenzial bei den Barrierefreiheitskriterien (bewertet nach WCAG 1.0) der Priorität A wurde bei 51 % der Webseiten festgestellt. Bei den Kriterien der

Prioritätsstufen AA und AAA belief sich diese Kennzahl auf 60% bzw. 63%. In einer Studie des Kompetenznetzwerks *Informationstechnologie zur Förderung der Integration von Menschen mit Behinderungen* im Jahr 2008 wurden zehn Webseiten österreichischer Behörden (das waren im Wesentlichen die Webseiten der Bundesministerien) auf Konformität mit WCAG 2.0 detailliert untersucht. Die Ergebnisse haben gezeigt, dass viele Seiten bereits auf Prioritätsstufe A über mehrere Fehler verfügen. Zudem wurden mehrere Kritikpunkte auf den Prioritätsstufen AA und AAA festgestellt [Ce16]. In Deutschland wurden die Internetauftritte des Bundes mit dem Inkrafttreten der *Barrierefreien Informationstechnikverordnung* (BITV) bereits im Jahr 2002 zur Umsetzung der Barrierefreiheitsgrundsätze verpflichtet. Auf der Ebene der Bundesländer wurde die Verordnung teilweise übernommen und teilweise in die regionale gesetzliche Grundlage eingearbeitet [ADN05]. Wissenschaftliche Untersuchungen haben aber gezeigt, dass insbesondere in den Bundesländern auch einige Jahre nach der Einführung der BITV immer noch ein Informationsproblem bestand: Die Verwaltungen waren über die Umsetzung und Einhaltung der Barrierefreiheit nur unzureichend informiert, was in vielen Fällen zur mangelhaften Umsetzung der Barrierefreiheitsgrundsätze führte [ADN05].

Aktuelle internationale [G16, Go11] und europäische [Fu16, Pi10] Studien haben ebenfalls ergeben, dass im Wesentlichen alle Internetauftritte über eine gewisse Anzahl der Barrierefreiheitsfehler verfügen. In den USA wurde die Barrierefreiheit der Internetauftritte der Regierungen mit Section 508 im Jahr 1998 geregelt [PB10]. Der US-Justizminister ist dazu verpflichtet alle zwei Jahre dem Präsidenten einen Bericht über den aktuellen Stand der Barrierefreiheit von Internetauftritten des Bundes zu liefern. Der Nachteil an diesem Bericht besteht darin, dass die Behörden ihre eigenen Webseiten selbst bewerten müssen und nur die Ergebnisse der Einschätzungen dem Justizministerium zukommen lassen, was zu Verzerrungen führen kann [OL11]. Der letzte veröffentlichte Bericht stammt aus dem Jahr 2011 und zeigt, dass durchschnittlich 70% der Internetauftritte des Bundes im Wesentlichen barrierefrei sind und 85% der Behörden regelmäßige Barrierefreiheitsbewertungen durchführen. In der Studie von Olallere und Lazar [OL11] wurden Regierungsseiten im Jahr 2010 bewertet, wobei lediglich vier von 100 Webseiten barrierefrei waren. In Asien besteht den zuletzt durchgeführten Studien zufolge ebenfalls ein Verbesserungspotenzial im Bereich der Barrierefreiheit der Regierungsseiten. In Korea wurden auf Basis von WCAG 1.0 und Section 508 eigene Barrierefreiheitsrichtlinien festgehalten. Die Ergebnisse einer Studie aus dem Jahr 2006 haben aber gezeigt, dass lediglich eine von 39 Regierungsseiten richtlinienkonform war. Insgesamt wurden somit auf 97% der Seiten vier Hauptpunkte der Richtlinien verletzt. In China wurde eine vergleichbare Studie im Jahr 2007 durchgeführt, bei der 324 Regierungsseiten auf Konformität mit WCAG 1.0 untersucht wurden [Sh07]. Die Ergebnisse haben gezeigt, dass keine einzige sich in der Stichprobe befindliche Seite den Barrierefreiheitsgrundsätzen der Prioritätsstufe 1 entspricht. Der häufigste Fehler betraf fehlende Alternativinhalte für diverse Inhalte, die somit für Menschen mit Behinderungen nicht zugänglich sind.

2.2 Umsetzungsmaßnahmen für mobile Barrierefreiheit

Für die Umsetzung der WCAG 2.0 Richtlinien werden den vier Hauptprinzipien der Barrierefreiheit (POUR-Prinzip: (i) wahrnehmbar (perceivable), (ii) bedienbar (operable), (iii) verständlich (understandable) und (iv) robust (robust)) insgesamt zwölf Richtlinien zugeordnet (vgl. Tabelle 1). Im Rahmen von Konformitätstests wird je nach Erfüllungsgrad von Erfolgskriterien (die im Gegensatz zu den zwölf Richtlinien evaluiert werden können) eine Konformitätsstufe vergeben (AAA, AA und A, wobei AAA die höchste und A die niedrigste ist).

Bei den Techniken für die Umsetzung wird zwischen ausreichenden und empfohlenen Techniken unterschieden, wobei die Erfüllung einer ausreichenden Methode genügt, um das zugeordnete Erfolgskriterium zu erfüllen. Die Erfüllung einer empfohlenen Methode reicht dazu jedoch nicht unbedingt aus [We16, Ke13].

RL	Umsetzungsmaßnahme
1.1	Textalternativen für alle Nicht-Text-Inhalte werden so zur Verfügung gestellt, dass diese in andere vom Benutzer benötigte Formen (z. B. in Großschrift, Braille, Symbole oder einfachere Sprache) geändert werden können.
1.2	Alternativen für zeitbasierte Medien werden zur Verfügung gestellt.
1.3	Inhalte werden auf help.gv.at verschiedene Arten dargestellt, ohne dass Informationen oder Strukturen verloren gehen.
1.4	Benutzer können leichter Inhalte sehen und hören (einschließlich der Trennung zwischen Vordergrund und Hintergrund).
2.1	Funktionen sind von der Tastatur aus verfügbar sind.
2.2	Benutzern der Webseite haben ausreichend Zeit, Inhalte zu lesen und zu benutzen.
2.3	Gestaltungsarten von Inhalte, von denen bekannt ist, dass sie zu Anfällen führen, werden vermieden.
2.4	Es werden auf dem Webauftritt Möglichkeiten zur Verfügung gestellt, um Benutzer dabei zu unterstützen zu navigieren, Inhalte zu finden und zu bestimmen, wo sie sich befinden.
3.1	Text-Inhalte sind lesbar und verständlich.
3.2	Webseiten sehen vorhersehbar aus und funktionieren.
3.3	Dem User wird dabei geholfen, Fehler zu vermeiden und zu korrigieren.
4.1	Kompatibilität mit aktuellen und zukünftigen Benutzeragenten wird maximiert, einschließlich assistierender Techniken.

Tab. 1: WCAG 2.0 Richtlinien (RL) und Umsetzungsmaßnahmen nach dem POUR-Prinzip.

Durch die Verbreitung mobiler Technologien ist bei der Umsetzung von Barrierefreiheit im Internet die Nutzung mobiler Endgeräte einzubeziehen. Der Begriff „mobile Barrierefreiheit“ bezieht sich in diesem Kontext grundsätzlich auf die Barrierefreiheit von Internetseiten und Applikationen, damit diese von Menschen mit Beeinträchtigungen mithilfe mobiler Geräte genutzt werden können. Mobile Barrieren im Internet stellen allerdings sowohl für Menschen mit als auch ohne Beeinträchtigungen Hindernisse dar. Zu den häufig anzutreffenden Einschränkungen mobiler Geräte zählen [YBH11, VB11]:

- kleiner Bildschirm: Beschränkungen der Bildschirmgröße führen dazu, dass Nutzer nur Teilansichten sehen; oftmaliges Scrollen vermindert die Übersichtlichkeit.
- niedrige Eingabegeschwindigkeit: Die Eingabe mithilfe einer verkleinerten Tastatur erfolgt langsamer und steigert die Fehlerhäufigkeit.
- niedrige Bandbreite: Herunterladen von Abbildungen kann relativ viel Zeit in Anspruch nehmen oder gar nicht funktionieren.
- Farbe: Wenn das Gerät über einen Bildschirm mit beschränkter Farbwiedergabe verfügt, kann das eine Verzerrung der Webinhalte auslösen. Ähnliche Probleme können bei situationsbezogenen Einschränkungen entstehen, wie z. B. bei der Nutzung des mobilen Gerätes bei ungünstigen Lichtverhältnissen.
- fehlende Unterstützung der Markup-, Skripting- oder anderer Datenformate: Nicht nur bei den zusätzlichen Hilfsgeräten, wie z. B. bei den Bildschirmlesegeräten, sondern auch bei üblichen mobilen Geräten können bei der Wiedergabe der Markup- oder Skripting-Formate Probleme auftreten, die dem Nutzer den Zugriff auf Inhalte verwehren.

Die Folgen dieser Einschränkungen für Nutzer mobiler Geräte ohne Beeinträchtigungen und für Nutzer mit Beeinträchtigungen sind sehr ähnlich. Das Problem der niedrigen Eingabegeschwindigkeit wird z.B. auch von Nutzern mit motorischen Beeinträchtigungen erlebt, wenn sie ein alternatives Eingabegerät verwenden [YBH11]. Nutzer mit Sehbehinderungen können ebenfalls auf ähnliche Barrieren, wie ein fehlendes Zeigegerät, treffen. Im Falle fehlender Struktur bzw. Überschriften müssen die Informationen mehrmals vorgelesen werden bis die notwendigen Inhalte gefunden werden, was Informationsüberlast zur Folge haben kann. Alternative Bildbeschreibungen sind nicht nur für Nutzer mobiler Geräte mit niedriger Bandbreite vom Vorteil, sondern auch für Menschen mit Sehbehinderungen. Probleme bei der Farbwiedergabe auf mobilen Geräten ähneln jenen Barrieren, die farbenblinde Personen erleben können.

In einer empirischen Studie von Yesilada et al. [YBH11] wurde die Ähnlichkeit der Barrieren im Internet für Nutzer mobiler Geräte und Nutzer mit Behinderungen erforscht. Im Laufe der Studie wurden 61 Arten von Barrieren untersucht, wobei sich 58 % als zutreffend für beiden Nutzergruppen erwiesen haben. Die Gruppe „Nutzer mit Behinderungen“ wurde in drei Untergruppen unterteilt: blinde Nutzer, Nutzer mit Sehbehinderung und Nutzer mit motorischer Behinderung. Der Anteil der ähnlichen Barrieren lag in allen drei Untergruppen bei rund der Hälfte (50 % bzw. 55 % bzw. 42 %), die meisten Ähnlichkeiten konnten also in der Gruppe „Nutzer mit Sehbehinderungen“ festgestellt werden. Es gilt hervorzuheben, dass sich einige Barrierearten, die ursprünglich als spezifisch für Nutzer mobiler Geräte definiert wurden, auch für Menschen mit Behinderungen als relevant erwiesen haben. Es gibt aber auch Unterschiede in der Bedienbarkeitsproblematik der beiden Nutzergruppen. Physische Eingabemechanismen mit niedriger Bandbreite sind z.B. für Nutzer mit Behinderungen, jedoch nicht für Nutzer mobiler Geräte geeignet. Zusätzlich muss bei der Entwicklung einer Internetseite die Eingabegeschwindigkeit und die Fehleranzahl bei der Eingabe berücksichtigt werden. Hierbei können Unterschiede zwischen Nutzern mit und ohne Beeinträchtigungen festgestellt werden. Bei Nutzern mit Beeinträchtigungen ist die Eingabegeschwindigkeit

niedriger und die Fehleranzahl höher. Daher sind beispielsweise drei Versuche für die Passwordeingabe für Nutzer mobiler Endgeräte angemessen, für die Nutzer mit Eingabebeschränkungen jedoch nicht ausreichend. Eine zusätzliche Hürde für Nutzer mit Beeinträchtigungen kann auch die Lautstärke störender Geräusche in öffentlichen Räumen darstellen, denn diese beeinflussen die Wahrnehmung der Inhalte [Ka09].

Die Mobile Web Initiative (MWI) wurde von W3C gegründet und beschäftigt sich hauptsächlich mit der Barrierefreiheit der mobilen Versionen von Internetseiten. Die Tätigkeiten der MWI befasst sich mit der Bestimmung der Geräteinformationen, die für die Anpassung der Inhalte notwendig sind und mit der Entwicklung der bewährtesten Praktiken für das mobile Web (MWBP). Darunter fällt die Ausarbeitung der Richtlinien und der begleitenden Dokumentation zur Implementierung der Richtlinien. Die entsprechenden Prinzipien und Umsetzungsmaßnahmen sind in Tabelle 2 dargestellt [HY08].

Prinzip	Umsetzungsmaßnahme
Gestalten für das Eine Web	Inhalte bei der Nutzung unterschiedlicher Geräte thematisch kohärent gestalten; Ergreifen notwendiger Maßnahmen, um die Einschränkungen zu umgehen; Durchführung der Tests sowohl am Gerät als auch an Emulatoren
Webstandards	Berücksichtigung der formalen Standards bei der Erstellung einer Internetseite; Aufbau der logischen Struktur; Lieferung der Inhalte in Formaten, die von den Endgeräten unterstützt werden
Problemvermeidung	keine Pop-up-Fenster; keine wesentlichen Seitenänderungen ohne Bekanntgabe der Änderungen an die Nutzer; keine Tabellenlayouts, Grafiken als Abstandhalter, Frames oder Image Maps
Gerätelimits	Eigenheiten mobile Geräte in Bezug auf Funktionalitäten, die bei der Entwicklung einer Seite berücksichtigt werden müssen, beachten (Verfügbarkeit von Cookies, Objekten und Skripten, Unterstützung von Tabellen, Style-Sheets und besonderen Schriftarten sowie Farben)
Navigation	Navigation und Eingabemöglichkeiten übersichtlich und effizient gestalten
Grafik/Farben	Farbkontrast, Bilder im Hintergrund, Größe der Bilder sorgfältig einsetzen
Größe	Mobile Seite möglich klein machen, damit die Nutzer Zeit und Kosten sparen
Datentransfer	Bestimmte Funktionen von Web-Protokollen (z. B. Autorefresh, Weiterleiten und Verlinken auf externe Ressourcen) sollten eher vermieden werden
Nutzer-eingaben	Gestaltung der bedienbaren Eingabemöglichkeiten für die schnelle und fehlerfreie Dateneingabe
mobile Nutzer	Inhalte kurz und prägnant präsentieren, damit Nutzer mobiler Geräte die notwendigen Informationen innerhalb kurzer Zeit und schnell finden können

Tab. 2: Umsetzungsmaßnahmen für mobile Endgeräte nach MWBP 1.0 Bewertungskriterien

2.3 Erfolgsfaktoren von WCAG 2.0 und SEO

Um eine höhere Aufmerksamkeit bei den Usern zu erzielen, wird von Webseitenanbietern eine Verbesserung des eigenen Rankings durch SEO bezweckt. Hinweise zu möglichen Überschneidungen von WCAG 2.0 Erfolgsfaktoren von Barrierefreiheit im Internet mit

SEO bietet Tabelle 3.

Thema	Barrierefreiheit im Internet durch...	SEO durch...	EF
Keywords in ALT-Text von Bildern	lückenlose Darstellung für Nutzer mit Sehschwächen bei Verwendung von Bildschirmlesegeräten	Indexierung der Keywords	1.1.1, 1.4.5, 1.4.9
ALT-Text von Audio- und Video-Inhalten	Verstehen der Video- und Audioinhalte auch für Nutzer mit Hörbeeinträchtigungen.	Indexierung der Untertitel oder Transkriptionen, um die relevanten Dateien auszusortieren	1.1.1, 1.2.1, 1.2.2 - 1.2.8
HTML-Validierung	ordnungsgemäßen Zugriff auf Inhalte mittels Hilfsgerät; benutzerfreundlicher und zuverlässiger Zugriff auch im Falle einer niedrigen Ladegeschwindigkeit	einfacher und effizienter Zugang für die Suchmaschinen zu den Inhalten der Internetseite	4.1.1, 4.1.2
Metatags	richtige Bearbeitung der Inhalte für Bildschirmlesegeräte, Spracherkennungs- und Sprachsynthesoftware	Zuordnung relevanter Information der Suchanfrage zu einer passenden Sprache	3.1.1, 3.1.2
Keywords in den Linktexten	Erkennung des Zwecks jedes Links für die Nutzer mit Behinderungen	positiven Einfluss auf das Suchmaschinenranking bei Relevanz	2.4.4, 2.4.9
Sitemap	Übersicht aller Seiten einer Webseite ist z. B. für Nutzer mit Sehschwächen sehr hilfreich	einfacheren Zugriff der Suchmaschinen zu jeder Unterseite	2.4.5
Erste Wörter im Titel-Tag	effizientere Navigation durch die vorgegebenen keywords	Titel-Tag Verwendung für Indexierung	2.4.2
Heading-Tags <h1>...<h6>	effizienteres Navigation aufgrund klar definierter Titel	Relevanz der Inhalte durch die Struktur Webseite (h1–h6)	1.3.1, 2.4.6, 2.4.10

Tab. 3: Überschneidungen von WCAG 2.0 Erfolgsfaktoren (EF) und SEO [MM13, EA11, Ha10].

Die Details der Rankingverfahren von Suchmaschinen werden von den Anbietern nicht kommuniziert, da eine Offenlegung aktive Eingriffe und Manipulationen seitens Internetseiten-Betreibern befürchten lässt. Die Relevanz von Webauftritten wird anhand zahlreicher Rankingfaktoren bestimmt; bei Google sind das beispielweise mehr als 200 Faktoren. Zu den wichtigsten Entwicklungen der Rankingfaktoren zählen laut Searchmetrics folgende Regeln: content steht weiterhin im Fokus, relevante und holistische Inhalte werden wesentlich besser gerankt; eine höhere Anzahl an Wörtern bzw. zusätzlicher Medien, wie Bilder, wird besser bewertet, wenn dadurch die Relevanz gesteigert werden kann; eine gesunde Seiteninformationsarchitektur zählt weiterhin zu den Grundvoraussetzungen; effiziente Seitentechnik wird ebenso vorausgesetzt, darunter werden unter anderem schnelle Ladezeiten, das Vorhandensein aller gängigen Meta-Elemente sowie Überschriften verstanden.; das Vorhandensein der Keywords in der URL hat sich als eher unwichtig erwiesen; die Keywords auf der Seite spielen allerdings eine wichtige Rolle bei der Bewertung; die Verlinkungsmetriken bleiben weiterhin wichtig, der

Fokus liegt aber eher auf Qualität als auf Quantität; beim Einfluss von großen Marken resultiert die Korrelation eher aus der hohen Anzahl der Verlinkungen als aus der Marke an sich; die sogenannten Usersignale (hohe Click-through-Raten, geringe Bounce-Raten sowie längere Verweildauer auf einer Seite) spielen auch weiterhin eine wichtige Rolle.

Bei der Wahl der SEO-Techniken sollten ausschließlich White-Hat-Maßnahmen verwendet werden, wenn die Barrierefreiheit durch die Optimierung der Seiten nicht beeinträchtigt werden soll. Bei den meisten Black-Hat-Maßnahmen werden verdeckte Inhalte (wie die wiederholenden Keywords, automatische Weiterleitungen, irrelevante unsichtbare Texte etc.) eingefügt, die nur für die Suchmaschinen ersichtlich sind. Diese Inhalte stellen in der Regel keine strukturierten bzw. relevanten Informationen dar und werden von Nutzern, die z. B. mithilfe der unterstützenden Technologien auf die Seite zugreifen, als irritierend empfunden [MM13].

3 Empirische Untersuchung

3.1 Datensatz, Hypothese und Methodik

Trotz neuer Gesetzesvorgaben bestätigt die im Rahmen der empirischen Untersuchung durchgeführte Analyse, dass weiterhin Verbesserungspotenzial bei Webauftritten österreichischer Regierungsseiten hinsichtlich Barrierefreiheit im Internet besteht.

Die Stichprobe für die empirische Analyse wurde mithilfe von Daten aus dem Open Directory Project *Directory Mozilla* (DMOZ) bestimmt. Für die Zwecke der vorliegenden Untersuchung wurden diese drei Kategorien in DMOZ (World: Deutsch: Regional: Europa: Österreich: Staat; ... Bundesregierung; Landesregierungen) ausgewählt.

Den Kategorien wurden insgesamt 49 Seiten zugeordnet. Diese stellen die unbereinigte Stichprobe der vorliegenden Untersuchung dar. Diese Stichprobe musste bereinigt werden, da mitunter dieselbe Seite mehreren Kategorien zugeordnet wurde (doppelte Erfassung) oder mehrere Unterseiten einer Webseite im Webkatalog gesondert ausgewiesen wurden. Die Bereinigung der Wiederholungen führte zum Ausschluss von insgesamt sieben Seiten aus der Stichprobe. Zudem wurde in einem Fall die im Webkatalog angeführte Unterseite durch die Hauptseite ersetzt. Ferner erfolgte die Abfrage von Suchmaschinenratings für die Seiten der Stichprobe mittels "Alexa.com". Bei der Abfrage wurde festgestellt, dass für 13 Seiten wegen zu niedriger Besucherzahlen keine Ratings zur Verfügung stehen. Diese Seiten wurden aus der Stichprobe ausgeschlossen. Die bereinigte finale Stichprobe besteht aus 29 Seiten.

Die Evaluierung der Barrierefreiheit einzelner Seiten erfolgte mithilfe der Software *Total Validator*. Für die Bewertung wurde die Pro-Version mit WCAG 2.0 (nach Konformitätsstufen AAA, AA und A) als Grundlage herangezogen. Da eine stichprobenartige Überprüfung gezeigt hat, dass in der vorliegenden Untersuchung die Berücksichtigung aller Unterseiten kaum Einfluss auf die Bewertungsergebnisse hätte,

wurde eine einheitliche Obergrenze von 15 Unterseiten pro Seite festgelegt. Für jede Seite der Stichprobe wurde der sog. WAB-Koeffizient (engl. web accessibility barriers) berechnet. In der vorliegenden Arbeit wurde der WAB-Koeffizient aus den Studien von Zeng und Parmanto [ZP03] und Arrue et al. [AVA08] als Basis herangezogen und an die Rahmenbedingungen der durchzuführenden Untersuchung angepasst. Der WAB-Koeffizient wurde folgendermaßen berechnet:

$$WAB - Koeffizient = \frac{n_v}{N_v} \quad (1)$$

wobei n_v die Anzahl der Fehler und N_v die Anzahl der Warnungen (potenzieller Fehler) bezeichnet [AVA08]. Zusätzlich wurde ein bereinigter WAB-Koeffizient berechnet:

$$WAB - Koeffizient = \frac{\left(\frac{n_v}{N_v}\right)}{Np} \quad (2)$$

wobei Np die Anzahl der bewerteten Seiten bezeichnet [AVA08]. Je höher der WAB-Koeffizient, desto mehr Barrieren sind auf der Webseite vorhanden [ZP03].

Für die Analyse wurden somit folgende Variablen verwendet: Alexa Ranking; Anzahl der Fehler (gesondert für die Konformitätsstufen AAA/AA/A); Anzahl der Warnungen (gesondert für die Konformitätsstufen AAA/AA/A); WAB-Koeffizient (gesondert für die Konformitätsstufen AAA/AA/A) und bereinigter WAB-Koeffizient (gesondert für die Konformitätsstufen AAA/AA/A).

Die höchste Anzahl an Barrierefreiheitsfehler wurde auf der Webseite des Bundeslands Kärnten und die höchste Anzahl an Warnungen auf der Seite des Bundesministeriums für auswärtige Angelegenheiten festgestellt. Die wenigsten Fehler in der Konformitätsstufe AAA wies die Webseite des Bundeslands Steiermark auf und in den Konformitätsstufen AA und A die des Bundeslands Oberösterreich. Die niedrigste Anzahl an Warnungen wurde auf der Amtshelferseite (www.help.gv.at) festgestellt. Den niedrigsten bereinigten WAB-Koeffizient erreichte die Seite des Bundesministeriums für Verkehr, Innovation und Technologie. Die meisten Barrieren waren dem bereinigten WAB-Koeffizient zufolge auf der Amtshelferseite zu verzeichnen. Über das beste Alexa Ranking verfügte die Seite des Bundesministeriums für Finanzen und über das schlechteste der Österreichische Rechnungshof. Einige Studien [MM13, Ha10, ZP03] haben einen Zusammenhang zwischen dem Suchmaschinenranking und dem Barrierefreiheitsniveau einer Webseite festgestellt. Mehrere Prinzipien der Barrierefreiheit im Internet ähneln den Rankingfaktoren von Suchmaschinen. Zudem erschweren Barrieren auf einer Internetseite nicht nur Menschen mit Behinderungen, sondern auch den Suchmaschinen eine schnelle und effiziente Navigation. Es wird daher erwartet, dass zwischen dem Suchmaschinenranking (Variable *Alexa Ranking*) und dem Barrierefreiheitsniveau (Variablen *WAB-Koeffizient*, *Anzahl der Fehler und Warnungen*) der österreichischen

Regierungsseiten ein positiver (Korrelations)-Zusammenhang besteht. Die Analyse des Zusammenhanges zwischen der Barrierefreiheit und dem Suchmaschinenranking einer Webseite erfolgte mithilfe der Statistik-Software *Stata*. Zuerst wurden für alle Variablen deskriptive Zahlen errechnet. Anschließend wurden die Daten (für je zwei Variablen) mittels Streudiagramme dargestellt, um mögliche Zusammenhänge zu visualisieren. Um die Stärke der Zusammenhänge zwischen den Variablen zu quantifizieren, wurde der Rangkorrelationskoeffizient nach Spearman berechnet. Die Berechnung des Rangkorrelationskoeffizienten erfolgte für die folgenden Variablenpaare (Anzahl der Fehler – Alexa Ranking, Anzahl der Warnungen – Alexa Ranking, WAB-Koeffizient – Alexa Ranking, bereinigter WAB-Koeffizient – Alexa Ranking), und zwar gesondert für jede der drei möglichen Konformitätsstufen, nämlich AAA, AA und A. Insgesamt wurden daher 12 Berechnungen für die Variablenpaare und Konformitätsstufen angestellt. Abschließend wurde die Korrelation auf statistische Signifikanz überprüft, dabei wurde ein Signifikanzniveau von 5 % festgelegt.

3.2 Auswertung und Interpretation der Ergebnisse

Für die Erläuterung und Interpretation der Ergebnisse der Korrelationsanalyse wurden die Daten grafisch dargestellt, um einen möglichen Zusammenhang zwischen den Variablen zu erkennen. Beispielhaft sind in Abbildung 1 Zusammenhänge zwischen der Anzahl der Fehler (2a), Warnungen (2b), bereinigten WAB-Koeffizienten der Konformitätsstufen AAA (2c) bzw. AA (2d) und dem Alexa Ranking dargestellt.

Die Datenpunkte sind in allen Grafiken relativ breit verstreut, was darauf hindeutet, dass es keine bzw. nur sehr geringe Korrelation zwischen den Variablen gibt. Im Streudiagramm zum Variablenpaar "bereinigter WAB-Koeffizient" und "Alexa Ranking" lässt sich eine von links nach rechts bergab verlaufende Linie aus den Datenpunkten erkennen; dies könnte als Hinweis auf einen leicht negativen Zusammenhang gesehen werden. Eine ähnliche Verteilung der Datenpunkte ist in den Streudiagrammen zu verzeichnen, welche die Zusammenhänge für die Konformitätsstufen AA und A darstellen. Um die möglichen Zusammenhänge zu quantifizieren, wurden für jede Konformitätsstufe die Rangkorrelationskoeffizienten berechnet. Dabei wird zuerst der Zusammenhang zwischen den Variablen "Anzahl der Fehler" und "Alexa Ranking" analysiert. Basierend auf der theoretischen Analyse wurde eine positive Korrelation zwischen diesen Variablen erwartet. Der Rangkorrelationskoeffizient aller drei Konformitätsstufen befindet sich jedoch unter dem niedrigsten Richtwert 0,10, was bedeutet, dass zwischen den Variablen keine Korrelation besteht. Zudem ist der Zusammenhang in allen Fällen statistisch nicht signifikant. Somit kann die Hypothese eines positiven Zusammenhangs nicht bestätigt werden. Im nächsten Schritt wird auf den Zusammenhang zwischen den Variablen "Anzahl der Warnungen" und "Alexa Ranking" eingegangen. Es wurde ein positiver Zusammenhang erwartet. In den Konformitätsstufen AAA und A liegt der Korrelationskoeffizient unter 0,10 (statistisch nicht signifikant); das bedeutet, dass zwischen den Variablen kein Zusammenhang besteht. In der Konformitätsstufe AA beträgt der Rangkorrelationskoeffizient jedoch 0,1383, was eine

leicht positive Korrelation signalisiert. Dieser Zusammenhang ist jedoch bei einem Signifikanzniveau von 5 % nicht signifikant. Zusammenfassend kann somit festgehalten werden, dass zwischen den Variablen "Anzahl der Warnungen" und "Alexa Ranking" kein statistisch signifikanter Zusammenhang besteht. Anschließend wird auf den Zusammenhang zwischen dem WAB-Koeffizienten (bereinigt/unbereinigt) und dem Alexa Ranking eingegangen. Die Analysen haben gezeigt, dass der Rangkorrelationskoeffizient in allen drei Konformitätsstufen zwischen 0,10 und 0,30 liegt, was auf eine geringe Korrelation hindeutet. Die Richtung der Korrelation ist jedoch im Gegensatz zu den Erwartungen negativ. Beim Signifikanzniveau von 5 % ist der Zusammenhang in allen drei Fällen statistisch nicht signifikant. Somit kann die ursprüngliche Vermutung eines positiven Zusammenhangs nicht bestätigt werden.

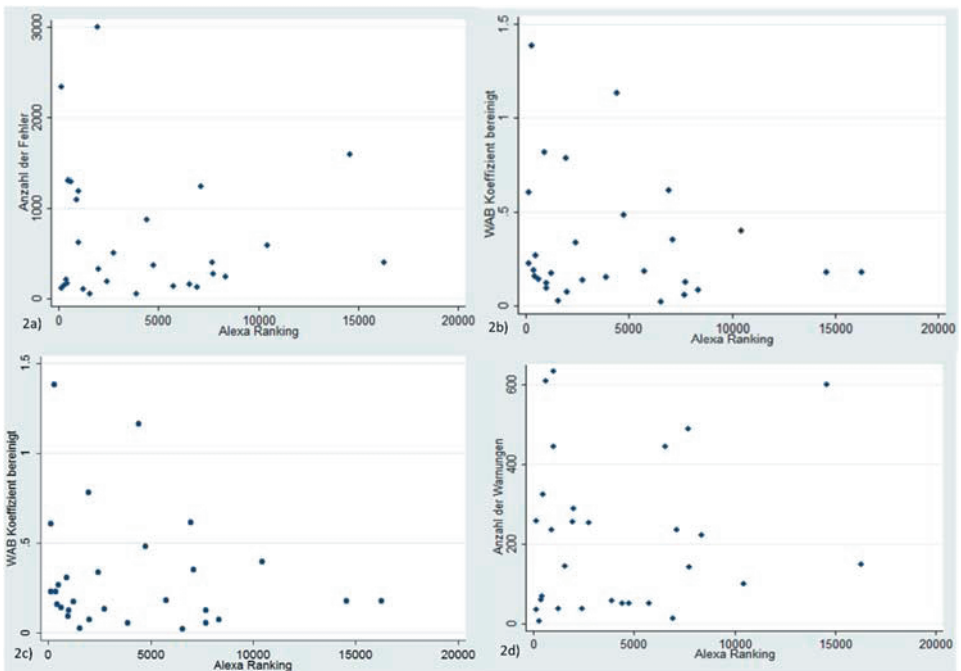


Abb. 1: Zusammenhänge der Variablen

4 Diskussion und Ausblick

Die Art und Weise, wie Suchmaschinen Internetseiten finden, ähnelt der Internetnutzung von Menschen mit Behinderungen, die beispielsweise ein Bildschirmlesegerät verwenden.

Zudem bestehen einige Überschneidungen bei den Suchmaschinenrankingfaktoren und den Barrierefreiheitsprinzipien. Im Rahmen dieser empirische Untersuchung, die Teil weiterführender Forschung ist, konnte ein positiver Zusammenhang zwischen dem Barrierefreiheitsniveau und dem Suchmaschinenranking der österreichischen Regierungsseiten basierend auf der durchgeführten Korrelationsanalyse nicht bestätigt werden, und die Korrelation hat sich zudem als statistisch nicht signifikant erwiesen. Im Zusammenhang müssen die Einschränkungen der vorliegenden Untersuchung erwähnt werden, wie z.B. die gewählten Kriterien (Alexa Ranking, WAB-Koeffizient, Regierungsseiten).

Zudem scheint die Größe der gewählten Stichprobe relevant zu sein, die im vorliegenden Fall relativ klein ist. Im Rahmen weiterer Forschungsaktivitäten könnte die Erweiterung der Stichprobe und eine Ergänzung um Internetauftritte der Regierungen in anderen Kernstaaten der Europäischen Union vorgenommen werden. Dies kann die Erreichung statistisch signifikanter Ergebnisse unterstützen. Zusätzlich wäre eine Analyse denkbar, in der Unternehmen unterschiedlicher Branchen betrachtet werden. Darüber hinaus kann die Bewertungsmethode des Barrierefreiheitsniveaus modifiziert werden; in der vorliegenden Arbeit wurde eine ausschließlich automatische Bewertung eingesetzt. Die Verwendung manueller Techniken neben den automatisierten Tools könnte hingegen ebenfalls positive Auswirkungen auf die Genauigkeit der Ergebnisse zeigen.

Der Zusammenhang zwischen der Barrierefreiheit einer Webseite und dem Suchmaschinenranking wurde bereits in einigen Studien diskutiert. So wurde anhand einer Fallstudie die Erhöhung des Besucheraufkommens nach der Einführung von Barrierefreiheit mithilfe einer empirischen Untersuchung quantifiziert [Ha10]. Mittels einer Stichprobe bestehend aus 150 Seiten unterschiedlicher Universitäten wurde in einer weiteren Studie der Zusammenhang zwischen dem Webometrics-Ranking (ein Portal, das Universitäten im Wesentlichen anhand ihrer Webpräsenz bewertet) und den Grad der Barrierefreiheit der Seite quantifiziert [EA11]. Obwohl die Bewertungsfaktoren in Webometrics nicht zur Gänze den Rankingfaktoren der Suchmaschinen entsprechen, gibt es bestimmte Überschneidungen. Darunter fallen z.B. die externen Verlinkungen, die Struktur der Seite sowie die Qualität der Inhalte.

Die vorliegende Analyse bietet einen Ausgangspunkt für zukünftige, weiterführende Forschung im Bereich der Barrierefreiheit im Internet und dem Suchmaschinenranking. Neben den webbasierten Authoring Tools, die den barrierefreien Zugang erst ermöglichen, bleibt auch die Barrierefreiheit im Internet aufgrund der auch in diesem Beitrag analysierten, wesentlichen Verbesserungspotenziale ein relevantes Thema [LSS14, Ma16]. Gerade der anhaltende Trend zur Nutzung mobiler Endgeräte [Hy15, Ml15b] und die Schwerpunktsetzung auf ortsbezogene Suche [Ml15a, Wi15] mit der Einbindung der Sozialen Crowd [MBS15, MBS16] und Collaborative Computer Supported Cooperative Work (CSCW) [No13, No14, MNS15] bieten Anknüpfungspunkte für weitere Untersuchungen in diesem Themengebiet [Ca13, MBS14]. Mobilität ist in diesem Kontext auch ein Schwerpunktthema der WCAG 2.0 extensions, die derzeit erarbeitet werden.

5 Zusammenfassung

Die herangezogenen Daten für die empirische Erhebung haben gezeigt, dass bei den ausgewählten Webauftritten österreichischer, öffentlicher Stellen weiterhin Verbesserungspotenzial besteht. Ebenso analysiert der Beitrag Überschneidungen bei den Erfolgsfaktoren von WCAG 2.0 und der SEO. Ein positiver Zusammenhang zwischen dem Barrierefreiheitsniveau von Regierungsseiten und dem Suchmaschinenranking hat sich im Rahmen dieser empirischen Untersuchung nicht bestätigt. Die weitere Schwerpunktsetzung auf ortsbezogene Suche im Internet und die Betrachtung von Webauftritten unterschiedlichen Branchen bzw. Behörden auf einer internationalen Ebene ist Gegenstand weiterer Forschungsaktivitäten.

Literaturverzeichnis

- [ADN05] Algermissen, L., Dermann, G., Niehaves, B.: Barrierefreiheit für Webseiten von Bund, Ländern und Gemeinden. *Wirtschaftsinformatik*, 47(5), S. 378-382, 2005.
- [AVA08] Arrue, M., Vigo, M., Abascal, J.: Web accessibility awareness in search engine results. *Universal Access in the Information Society*, 7(1-2), S. 103-116, 2008.
- [Ba14] Bauer, C., et al.: Fostering collaboration by location-based crowdsourcing. *LNCS 8683*, Springer, doi:10.1007/978-3-319-10831-5_13, S. 88-95, 2014.
- [Ca13] Cardonha, C., et al.: A crowdsourcing platform for the construction of accessibility maps W4A '13. *ACM*, 26, ISBN: 978-1-4503-1844-0, 2013
- [Ce16] Centre for Excellence in Universal Design.: A <http://universaldesign.ie/Technology-ICT/Measures-to-Improve-Accessibility-of-Public-Websites-in-Europe/> [Stand: 16.6.2016]
- [EA11] Elgharabawy, M. A., Ayu, M. A.: Web Content Accessibility and its Relation to Webometrics Ranking and SEO. *ICRIIS. Kuala Lumpur: IEEE*. S. 1-6, 2011.
- [Ec16] European Commission http://europa.eu/rapid/press-release_IP-16-1654_de.htm [Stand: 16.5.2016]
- [Fu16] Funka European Commission: <http://www.funka.com/en/our-assignments/position-of-trust/archive---commission-of-trust/measuring-progress-of-eaccessibility-in-europe---meac2> [Stand: 16.5.2016]
- [G16] G3ict. CRPD 2013 ICT Accessibility Progress Report. http://www.g3ict.org/resource_center/CRPD_2013_ICT_Accessibility_Progress [Stand: 16.6.2016]
- [Go11] Goodwin, M., et al.: Global Web Accessibility Analysis of National Government Portals and Ministry Web Sites. *Journal of Inf. Technology and Politics*, 8(1), S. 41-67, 2011
- [HY08] Harper, S., Yesilada, Y.: *Web Accessibility*. London: Springer, 2008.
- [Hy15] Hyben, B., et al.: Consumer acceptance on mobile shopping of textile goods in Austria:

modelling an empirical study. MoMM'15, ACM, ISBN: 978-1-4503-3493-8; S. 402-406, 2015.

- [Ha10] Hartjes, R., et al.: Veränderter Website-Traffic bei Einführung von barrierefreiem Web - eine Fallstudie. HMD - Praxis der Wirtschaftsinformatik (274), S. 85-93, 2010.
- [Ka09] Kane, S. K., et al.: Freedom to roam: a study of mobile device adoption and accessibility for people with visual and motor disabilities. SIGACCESS: ACM. S. 115-122, 2009.
- [Ke13] Kerkmann, F.: Web Accessibility. Informatik-Spektrum, 36(5), S. 455-460, 2013.
- [LSS14] Leitner, et al.: Web accessibility implementation in private sector organizations: Motivations and business impact, International Journal UAIS, S. 1-12, 2014.
- [Ma16] Makarova, T.: Web Accessibility und Search Engine Ranking. Universität Wien.
- [MBS16] Mladenow, A., Bauer, C., Strauss, C.: 'Crowd Logistics': The Contribution of Social Crowds in Logistics Activities, International Journal of Web Information Systems, 2016.
- [MBS15] Mladenow, A., et al.: Crowdsourcing in Logistics: Concepts and Applications Using the Social Crowd. DOI: <http://dx.doi.org/10.1145/2837185.2>, iiWAS '15, ACM, 2015.
- [MI15b] Mladenow, A., et al.: Online Ad-fraud in Search Engine Advertising Campaigns. Springer, LNCS 9357, doi:10.1007/978-3-319-24315-3_11, S. 109-118, 2015.
- [MI15a] Mladenow, A., et al.: Collaboration and Locality in Crowdsourcing. In Intelligent Networking and Collaborative Systems (INCOS), Int. Conf. on IEEE, S. 1-6, 2015.
- [MM13] Moreno, L., Martinez, P.: Overlapping factors in search engine optimization and web accessibility. Online Information Review, 37(4), S. 564-580, 2013
- [MNS15] Mladenow, A., Novak, N. M., Strauss, C.: Micropayments in virtuellen Welten-Prozessmodell und Nutzung bei 7-bis 12-jährigen Besuchern, LNI 246, S. 1267-1277, 2015.
- [No13] Novak, N. M., et al.: Avatar-based Innovation Processes-Are Virtual Worlds a breeding ground for Innovations?. In Proceedings of Int. Conference on Information Integration and Web-based Applications & Services ACM. doi:10.1145/2539150.2539253, S. 174, 2013.
- [No14] Novak, N. M., et al.: Virtual worlds as settings for avatar-based innovation processes, Journal of Service Science Research 6 (1), S. 71-98, 2013.
- [MM13] Moreno, L., Martinez, P.: Overlapping factors in search engine optimization and web accessibility. Online Information Review, 37(4), S. 564-580, 2013.
- [OL11] Olalere, A., Lazar, J. (2011). Accessibility of U.S. federal government home pages. Government Information Quarterly, 28(3), S. 303-309, 2011.
- [PB10] Peters, C., Bradbard, D. A.: Web accessibility: an introduction and ethical implications. Journal of Information, Communication and Ethics in Society, 8(2), S. 206-232, 2010.
- [Pi10] Piat, S. C., et al.: Visibility, accessibility and quality of Italian public health institutional websites. Italian Journal of Public Health, 7(2), S. 102-108, 2010.
- [Sh07] Shi, Y.: The accessibility of Chinese local government web sites: An exploratory study. Government Information Quarterly, 24(2), S. 377-403, 2007.

- [VB11] Vigo, M., Brajnik, G.: Automatic web accessibility metrics: Where we are and where we can go. *Interacting with Computers* (23), S. 137-155, 2011.
- [We16] Web Content Accessibility Guidelines, <https://www.w3.org/WAI/intro/wcag> [Stand: 16.5.2016]
- [Wi15] Widhalm, M, et al.: E-Appointment Plattformen zur Effizienzsteigerung und Umsatzgenerierung. *HMD Praxis der Wirtschaftsinformatik*, 52(3), S. 401-417, 2015.
- [YBH11] Yesilada, Y., Brajnik, G., Harper, S.: Barriers common to mobile and disabled web users. *Interacting with Computers* (23), S. 525-542, 2011.
- [ZP03] Zeng, X., Parmanto, B.: Evaluation of web accessibility of consumer health information websites. *AMIA Annual Symposium Proceedings*. S. 743-747, 2003.

Digitale Teilhabe statt Doing Disability: Assistive Technologien für inklusive Medienbildung im Kindergarten

Nicole Najemnik¹, Isabel Zorn²

Abstract: Der Artikel fokussiert auf Kinder mit Behinderung im Kindergarten und zeigt auf, wie diese sowohl durch mangelnde Medienkompetenz von pädagogischen Fachkräften, als auch durch einseitiges Design von Assistenztechnologien von EntwicklerInnen behindert werden. Dafür werden Erkenntnisse aus internetsoziologischer Forschung zu (digitaler) Ungleichheit und das Konzept des Doing Disability aus den Critical Disability Studies herangezogen. Darauf aufbauend wird die Rolle (assistiver) Technologien für inklusive Medienbildung diskutiert und das Potenzial digitaler Medien für Partizipation und Teilhabe von Kindern mit Behinderung anhand aktueller Forschungsergebnisse verdeutlicht. Abschließend werden Handlungsempfehlungen für die pädagogische und informationstechnische Praxis formuliert und gezeigt, wie Doing-Disability-Prozesse durch adäquates Design in der Informatik und adäquate Kompetenzvermittlung in der Pädagogik verringert werden können.

Keywords: Inklusion, Doing Disability, Kindergarten, digitale Medien, Medienkompetenz, Assistenztechnologie, digitale Teilhabe, Partizipation, inklusives Design

1 Einleitung

Wenn über die Gestaltung von und Anforderungen an Assistenztechnologien³ gesprochen wird, so sind in der Regel implizit Assistenztechnologien für ältere Menschen im Rahmen von *Ambient Assisted Living (AAL)* gemeint. Der vorliegende Artikel fokussiert auf Kinder mit Behinderung im Kindergarten und möchte aufzeigen, warum die spezifischen Bedürfnisse von *toddler netizens* [Lu99] mit Behinderung an Assistenztechnologien spezifische Berücksichtigung in Informatik und Pädagogik benötigen, um zusätzliche Barrieren zu reduzieren und Inklusion zu ermöglichen. Ausgehend vom (theoretischen) Potenzial von inklusivem Design von (Assistenz-)Technologien zielt der Artikel darauf ab, Hindernisse bei der Entwicklung und im pädagogischen Einsatz von Technologie für inklusive Bildung von Kindern mit und ohne Behinderung zu beleuchten. Mit Verweisen auf die Erkenntnisse

¹ Technische Hochschule Köln, Fakultät für Angewandte Sozialwissenschaften, Ubierring 48, 50678 Köln, nicole.najemnik@th-koeln.de

² Technische Hochschule Köln, Fakultät für Angewandte Sozialwissenschaften, Ubierring 48, 50678 Köln, isabel.zorn@th-koeln.de

³ Als Assistenztechnologien werden gemeinhin jene Technologien bezeichnet, die für Menschen mit Behinderung kompensatorische Funktionen haben und häufig von Krankenkassen als Hilfsmittel bezahlt werden. Im Kontext von spielenden Kindern halten die Autorinnen diese Definition für problematisch, da auch andere Technologien, wie beispielsweise Tablets, Spielekonsolen oder digitale Kameras, das Spiel, die Interaktion und die Kommunikation zwischen Kindern unterstützen und so inklusiv wirken können.

internetsoziologischer Auseinandersetzungen über (digitale) Ungleichheit und dem Konzept des *Doing Disability* aus den Critical Disability Studies, wird deutlich gemacht, wie Behinderung teilweise auch im Entwicklungsprozess konstruiert wird und wie Kinder mit Behinderung durch mangelnde Medienkompetenz von pädagogischen Fachkräften behindert werden. Darauf aufbauend wird die Rolle (assistiver) Technologien für inklusive Medienbildung diskutiert und die zentrale Bedeutung von medienpädagogischer Kompetenz bei pädagogischen Fachkräften für Inklusionsprozesse verdeutlicht. Anschließend wird das Potenzial digitaler Medien für Partizipation und Teilhabe von Kindern mit Behinderung anhand aktueller Forschungsergebnisse diskutiert und zukünftige Forschungsbedarfe aufgezeigt. Abschließend geben die Autorinnen Handlungsempfehlungen für die pädagogische und informationstechnische Praxis und verdeutlichen, wie *Doing-Disability*-Prozesse sowohl durch adäquates Design in der Informatik, als auch durch adäquate Kompetenzvermittlung in der Pädagogik reduziert werden können und zeigen die Notwendigkeit einer Verschränkung von Perspektiven und Methoden aus Informatik und Pädagogik im Hinblick auf Assistenztechnologie und inklusive Medienbildung im Kindergarten auf.

2 (Digitale) Ungleichheit und Behinderung

Die zunehmende Technisierung der Gesellschaft durchdringt alle Lebensbereiche und soziale, ökonomische und politische Prozesse werden vermehrt über das Internet ausgehandelt und umgesetzt [WGH12, Ba10, Th04]. Diese globalen Entwicklungen spiegeln sich in wissenschaftlichen Debatten und Analysen durch Begriffe und Definitionsversuche wie *Mediengesellschaft* [Im06], *Veralltäglichung des Internets* [Wi08] oder *Mediatisierung* [Kr07] wider. Die Allgegenwärtigkeit des Internets sowie digitaler Medien führt dazu, dass völlig neue Formen von Kommunikation und Interaktion entstehen [Wi08]. In diesem Zusammenhang lässt sich eine Verstärkung existierender gesellschaftlicher Ungleichheiten im digitalen Raum beobachten [St12]. *Behinderung* ist hierbei eine zentrale Kategorie, entlang derer ungleiche Partizipations- und Teilhabechancen reproduziert werden.

2.1 Digital Divide, Digital Inequality und Doing Disability

Digital Divide und *Digital Inequality* wurden ursprünglich auf Zugangsmöglichkeiten zum Internet(anschluss) bezogen, dem sogenannten *first-level divide*. Dann wurde ersichtlich, dass auch die Arten des Handelns im Internet (z.B. nur Spielekonsum vs. qualifizierte Informationsrecherche) Ungleichheit produzieren, was als *second-level divide* bezeichnet wird [Ha02]. Mit einer Perspektive auf die Möglichkeiten von Menschen mit Behinderung, technische Geräte für Kommunikation und Partizipation zugänglich zu haben sowie sinnvoll und versiert für Kommunikation und Bildung nutzen zu können, sehen wir, wie dadurch ebenfalls auf verschiedenen Ebenen (Zugang und Nutzungsweisen) Risiken für Ausschlüsse und Benachteiligungen bei

Partizipationsmöglichkeiten entstehen können. *Zugangsweisen* beziehen sich dann darauf, ob *geeignete* Technologien – also auch Assistenztechnologien – *bereitgestellt* werden. Hierfür sind sowohl informatische Entwicklungs- und Designprozesse relevant, um geeignete und akzeptierte Technologien zu entwickeln, wie auch pädagogische Praxen, die auch Kindern solche Geräte und damit Zugänge zu technisch vermittelter Kommunikation bereitstellen, sie im pädagogischen Alltag einsetzen und insbesondere verbalsprachlich eingeschränkte Kinder frühzeitig an diese Technologien heranführen. Bislang ist diese Aufgabe vorrangig bei LogopädInnen und Fachkräften der Unterstützten Kommunikation angesiedelt, allerdings ist es notwendig, diese auch alltäglich in pädagogisch relevanten Kontexten einzusetzen. Teilhabe fördernde Nutzungsweisen erfordern von der Informatik, Technologien passgenau für die Erfordernisse von Nutzungspraktiken und Nutzungskontexte zu entwickeln. Von der Pädagogik erfordern sie, dass auch für behinderte Kinder qualifizierte kommunikations- und inklusionsfördernde, aber insbesondere auch bildungsrelevante und alltägliche Nutzungsweisen mit entsprechenden (Assistenz-)Technologien eingeübt werden. Dies muss bereits im Kindergartenalter durch frühpädagogische Fachkräfte erfolgen, um die Entwicklung von Ungleichheiten zu vermeiden. Doing-Disability-Prozesse können somit durch adäquates Design in der Informatik wie auch durch adäquate Kompetenzvermittlung und Pädagogik reduziert werden.

2.2 Doing Disability durch pädagogische Fachkräfte im Kindergarten

Das Potenzial zur Stärkung digitaler Rechte und damit verbundener Teilhabechancen von Kindern mit Behinderung wird bislang von pädagogischen Fachkräften nur unzureichend ausgeschöpft. Gegenüber der Nutzung digitaler Medien im Kindergartenalltag lassen sich unter den pädagogischen Fachkräften große Vorbehalte und Vorurteile feststellen, weshalb eine mangelnde Bereitschaft zur Anwendung eben dieser vorherrschend ist. So wird der Medieneinsatz häufig sehr kritisch gesehen und die Fähigkeit, Medien aktiv gestalten zu können, wird nicht offensiv angestrebt. Darüber hinaus mangelt es den pädagogischen Fachkräften in Kindergärten häufig an der notwendigen medienpädagogischen Kompetenz [He16, BAB13]. Diese ist aber unabdingbar, wenn inklusive Lernszenarien umgesetzt werden sollen. Um einer weiteren Marginalisierung von Menschen mit Behinderung in der Informations- und Wissensgesellschaft entgegen zu wirken und ihnen gleiche Teilhabechancen zu ermöglichen, ist es von zentraler Bedeutung, dass sie schon früh einen selbstbestimmten Umgang mit adäquaten und relevanten digitalen Medien erlernen. Inklusive Medienbildung im Kindergarten kann hierbei einen wichtigen Beitrag leisten.

2.3 Doing Disability durch EntwicklerInnen von Assistenztechnologien

EntwicklerInnen von (Assistenz-)Technologien haben durch das von ihnen festgelegte Design Einfluss darauf, ob Menschen mit körperlichen und/oder geistigen Beeinträchtigungen im Umgang mit digitalen Medien behindert werden oder gleiche

Teilhabe- und Partizipationschancen erfahren. Häufig liegt der Fokus von EntwicklerInnen allerdings ausschließlich auf der Kompensation einer bestimmten medizinischen Beeinträchtigung, und die sozialen Inklusions- und Exklusionspotenziale von (Assistenz-)Technologien werden nicht mitgedacht. Darüber hinaus lässt sich feststellen, dass beispielsweise die Entwicklung von Assistenztechnologien für Menschen mit verbalsprachlichen Beeinträchtigungen einseitig auf diese ausgerichtet ist, und nicht beachtet wird, dass Technologien so gestaltet werden müssen, dass auch etwaige GesprächspartnerInnen sie nutzen können. Soziale Exklusion und eingeschränkte Kommunikationsmöglichkeiten werden also auch durch einseitiges Design hergestellt. Inklusives Design und ein Verständnis von universeller Bedienbarkeit, dass die Bedürfnisse *aller* NutzerInnen von Assistenztechnologien einbeziehen, kann dazu beitragen, dass Menschen mit und ohne Behinderung über gleiche Teilhabe- und Partizipationsmöglichkeiten in der Wissens- und Informationsgesellschaft verfügen [MHK10].

3 Inklusive Medienbildung und assistive Technologien

Im Zuge der Umsetzung der UN-Behindertenrechtskonvention werden in Deutschland zukünftig Kinder mit und ohne Behinderung vermehrt gemeinsam in integrativen Kindertagesstätten betreut. Somit werden deutlich mehr Kinder mit Behinderung als früher in Regeleinrichtungen durch pädagogische Fachkräfte betreut, die bislang kaum Notwendigkeit für einen Einsatz digitaler Medien im pädagogischen Alltag sahen und entsprechend über eher geringe medienpädagogische Kompetenz bezüglich des Einsatzes von digitalen Medien als Lern-, Kommunikations- und Spielmaterial in der Kindertagesstätte verfügen.

Es ist davon auszugehen, dass durch Kinder mit besonderem Förderbedarf (z.B. Lernförderung, Sprachförderung, Spielförderung, physiotherapeutische und motorische Förderung, Konzentrationsförderung) vermehrt elektronische Geräte und digitale Medien, wie beispielsweise Talker zur Unterstützten Kommunikation oder andere Assistenztechnologien, aber auch elektronische Spiele, Lerntechnologien, Hörgeräte, Stimmverstärker, elektronische Sehhilfen und viele mehr Einzug in den Kindergartenalltag halten. Große Teile des pädagogischen Fachpersonals sind kaum darauf vorbereitet, diese Geräte technisch zu bedienen, pädagogisch zu nutzen oder in inklusiven Settings zur inklusiven Gestaltung von Interaktionsszenarien zu nutzen. Gleichsam bilden diese Medien jedoch für Kinder, die von Behinderung bedroht oder behindert sind, wesentlichen Zugang zur selbstbestimmten Lebensführung, zur Partizipation an Alltag und Gesellschaft sowie zu Lernmöglichkeiten.

Dies zu ermöglichen, wird in mehreren Paragraphen der UN-Behindertenrechtskonvention gefordert (u.a. §§9, 21). Die Nutzung dieser Geräte ist insbesondere für Menschen mit Behinderung essenziell – da sie teilweise die einzige Möglichkeit für Kommunikation und Teilhabe darstellt – und muss daher frühestmöglich

erfolgen, um Entwicklungsfenster für die Entwicklung von Sehen, Hören, Kommunizieren, Interagieren sowie von Kommunikationsstrategien nicht verstreichen zu lassen. Die Kommunikation mit Kindern (und Erwachsenen) mit Behinderung verläuft bereits (aktuell und zukünftig voraussichtlich wachsend) in hohem Maße über analoge und digitale Medien (z.B. Hörhilfen, Sehhilfen, Talker, Videokommunikation, Smartphone, Fotografie), sodass auch nicht-behinderte Menschen den Umgang und die Kommunikation mit diesen Medien erlernen müssen, um inklusiv kommunizieren zu können.

Ausgehend von der Hypothese, dass digitale Medien und Technologien sowohl als Assistenztechnologien, als auch als Kompensationstechnologien, aber auch als Lern- und Kommunikationsmittel in einer mediatisierten Gesellschaft großes Potenzial für die Kompensation von Einschränkungen, wie auch für erhöhte Partizipation und Interaktion bereits von kleinen Kindern Einzug in Bildungskontexte halten und halten werden, stellt sich die Frage, wie die pädagogischen Fachkräfte in der frühkindlichen Bildung darauf vorbereitet werden müssen, diese für inklusives Lernen einsetzen zu können. Es gibt dazu bislang kaum Forschung im deutschsprachigen Raum, es fehlt sowohl an Konzepten, als auch an wissenschaftlicher Fundierung.

Um Kommunikation und Interaktion zwischen Kindern mit und ohne Behinderung zu fördern und somit zu Inklusion beizutragen, ist Medienkompetenz sowohl beim pädagogischen Fachpersonal, als auch bei den Kindern notwendig. Medienkompetenz beinhaltet nach Baacke [Ba96] neben der Medienkritikfähigkeit auch die Medienkunde, die Mediennutzung und die Mediengestaltung. Für pädagogisches Personal ist hier noch die medienpädagogische Kompetenz als bedeutsam zu ergänzen, also die Fähigkeit, den Medieneinsatz pädagogisch beurteilen und gestalten zu können. Dazu zählt auch die Fähigkeit, Medien didaktisch sinnvoll zur Gestaltung von Lernsituationen einzusetzen.

Um inklusive Spiel- und Lernszenarien gestalten zu können, ist es erforderlich, dass die dazu notwendigen Geräte (z.B. Talker) die Situation nicht behindern. Dazu zählt auch, dass sich die Konzeption der „NutzerInnen“ dieser Geräte nicht nur auf das Kind mit Behinderung beschränken darf, sondern auch die anderen KommunikationspartnerInnen als Nutzende im Blick behalten muss [Fu16]. Hier ist daher eine Anforderung an informatische Entwicklungsprozesse zu sehen, die in Abschnitt fünf ausgeführt wird.

Inklusive Medienbildung bezeichnet bislang medienpädagogische Angebote, die Menschen mit und ohne Behinderung zusammen Bildungsprojekte ermöglicht [Sc15, Bo12]. Perspektiven auf die Integration assistiver Technologien sind dabei bislang kaum ausgebildet. Hier und bei der medien- und technikbezogenen Weiterbildung der frühpädagogischen Fachkräfte ist Forschungsbedarf zu konstatieren, um den Einsatz assistiver Technologien für inklusive Bildungsprozesse zu fördern.

4 Das Potenzial digitaler Medien für Partizipation und Teilhabe von Kindern mit Behinderung im Kindergarten – Forschungsbedarf und ausgewählte Fallstudien

Die sozialwissenschaftliche und medienpädagogische Forschung beschäftigt sich intensiv mit dem Einsatz digitaler Medien im schulischen Bereich [z.B. Wa14]. Auch Fragen nach inklusiver Medienbildung in Grundschulen und weiterführenden Schulen finden zunehmend mehr Beachtung [z.B. MH16]. Eine eklatante Forschungslücke lässt sich hingegen in Bezug auf (assistive) Technologien und inklusives Lernen im Kindergarten feststellen. Es existiert kaum Forschung dazu, wie (assistive) Technologien in der frühkindlichen Erziehung so eingesetzt werden können, dass sie das gemeinsame Lernen von Kindern mit und ohne Behinderung fördern und etwaige Lernschwierigkeiten zu kompensieren vermögen. Es gibt bislang lediglich vereinzelte Fallstudien, die Kinder im Vorschulalter mit Behinderung in die Untersuchung einbeziehen und die positiven Effekte digitaler Medien zur Kompensation von Lernschwierigkeiten oder physischen Beeinträchtigungen sowie im Hinblick auf gesteigerte Partizipationsmöglichkeiten aufzeigen. Die Ergebnisse von drei Studien, die jeweils unterschiedliche Technologie genutzt haben, aber zu ähnlichen Ergebnissen gekommen sind, sollen im Folgenden skizziert werden sollen.

Eine von Zorn et al. durchgeführte Studie [Zo16] liefert auf mehreren Ebenen Ergebnisse im Hinblick auf das Potenzial von digitalen Medien für inklusives Lernen im Kindergarten. So wird anhand des Beispiels von Tablets gezeigt, dass der Einsatz digitaler Medien zu mehr Partizipation und Teilhabe von Kindern mit Behinderung führen kann. Darüber hinaus liefert die Studie Hinweise darauf, dass der geeignete Einsatz von digitalen Medien inklusives Lernen und Spielen fördert, dadurch Langeweile bei Kindern mit Beeinträchtigungen reduziert und somit pathologischem Verhalten, wie beispielsweise ständigem Augenkratzen, entgegen wirken kann. Außerdem verdeutlicht die Studie auch, dass pädagogische Fachkräfte, trotz aller positiven Effekte, starke Vorbehalte gegenüber dem Einsatz digitaler Medien im Kindergarten haben.

Ferner verdeutlichen Lin und Chang [LC14], dass digitale Medien darüber hinaus auch zu mehr physischer Aktivität von Kindern mit körperlichen Beeinträchtigungen und damit zu gesteigerten Partizipations- und Interaktionsmöglichkeiten beitragen können. In ihrer Studie zeigen sie anhand des Beispiels der MaKey-MaKey Leiterplatte auf, wie Kinder im Vorschulalter mit zerebraler Kinderlähmung, die bislang wenig für physische Aktivitäten motiviert werden konnten, durch den Einsatz von MaKey-MaKey ihre physische Aktivität kontinuierlich steigerten. Dieses Beispiel unterstreicht darüber hinaus auch, dass (Assistenz-)Technologie, die individuell an die spezifischen Bedürfnisse von Menschen mit körperlichen und/oder geistigen Beeinträchtigungen angepasst werden kann, nicht zwangsläufig hochpreisig muss.

Retalis et al. [Re14] haben gezeigt, wie der Microsoft Kinect Sensor in Verbindung mit Kinems-Lernspielen so genutzt werden kann, dass Kinder mit dem Aufmerksamkeits-Defizit-Syndrom mit Hyperaktivität (ADHS) bisher nur schwach ausgeprägte Fähigkeiten ausbauen und mit Spaß trainieren können. Die in einem Kinderkrankenhaus durchgeführte Studie berücksichtigt Kinder im Alter von vier bis acht Jahren, die im Laufe der Studie alle bestimmte Fähigkeiten verbessern konnten, wie beispielsweise ihr Reaktions- oder Konzentrationsvermögen, ihre Problemlösungskompetenzen oder ihre Koordinationsfähigkeiten. Neben den Lernerfolgen unterstreicht die Studie allerdings besonders die hohe Motivation und die Freude der StudienteilnehmerInnen, mit Kinect/Kinems zu spielen und zu lernen.

Alle drei genannten Fallstudien haben gemeinsam, dass Technologie verwendet wurde, die Kinder mit und ohne Behinderung gemeinsam nutzen können. Damit sind sie Beispiele dafür, dass der gezielte Einsatz von Technologie bereits im Kindergarten für inklusives Lernen und Spielen und damit zur Grundsteinlegung von digitaler Teilhabe und gleichen Kommunikationschancen genutzt werden kann.

5 Conclusio: Handlungsempfehlungen für Pädagogik und Informatik zur Vermeidung von Doing-Disability-Prozessen

In Anbetracht aktueller Forschungsergebnisse lassen sich zwei zentrale Schlussfolgerungen im Hinblick auf assistive Technologien und inklusive Medienbildung im Kindergarten ziehen: Erstens benötigen pädagogische Fachkräfte geeignete Ausbildungsmöglichkeiten, um Kompetenzen zum Einsatz von (Assistenz-)Technologien für inklusive Settings zu erwerben und so Doing-Disability-Prozesse gegenüber jenen Kindern, die Assistenztechnologien nutzen, zu vermeiden. Medienpädagogische Kompetenz bei frühpädagogischen Fachkräften ist erforderlich, um durch inklusiven und bildungsfördernden Medien- und Technikeinsatz Zugang und Nutzungsweisen zu fördern und somit (digitale) Ungleichheiten zu reduzieren, statt sie noch weiter zu verstärken. Zweitens werden für die Konzipierung von Hard- und Software Kenntnisse in Universal Design sowie über die Vermeidung von strukturellen Diskriminierungen durch ungeeignete Technologien benötigt, die in der Aus- und Weiterbildung von InformatikerInnen vermittelt werden können. Ein Beispiel für die Vermeidung von technikbedingten Doing-Disability-Prozessen sind Kenntnisse zu Co-Creation bzw. inklusivem und partizipativem Design.

Wenn assistive Technologien entwickelt werden, so müssen dafür die Nutzungskontexte und die jeweiligen NutzerInnen konzeptualisiert werden. Bei der Nutzung eines Talkers zur Unterstützten Kommunikation zeigt sich, dass nicht nur die nicht-sprechende Person die Nutzerin ist, sondern dass auch die mit ihr kommunizierenden Personen Nutzende sind [Fu16]. Auch sie müssen mit dem Gerät interagieren, auf Ausgaben warten, und die Maschinenstimme verstehen. Die Interaktion verläuft anders als gewöhnlich, direkter Blickkontakt wird selten aufrechterhalten, die Beteiligten fokussieren eher das Gerät.

Eine Entwicklung, die alle Beteiligten als Nutzende konzipiert, würde daher möglicherweise einen Talker mit zwei Bildschirmen (vorne und hinten) konzipieren, um insbesondere pädagogischen Fachkräften zu ermöglichen, Gerät *und* Kind im Blick behalten zu können. Um solche Entwicklungen zu ermöglichen, sind interdisziplinäre Teams in Co-Creation Prozessen erforderlich. Werden die tatsächlichen Nutzungskontexte nicht adäquat im Entwicklungsprozess berücksichtigt, so besteht die Gefahr, bereits durch das Design der Technologie Doing-Disability-Prozesse in der Nutzung der Technologie (z.B. Kommunikationshilfe) zu provozieren.

Als zwei Best-Practice-Beispiele für inklusives Design von (assistiven) Technologien sei hier auf die Fallstudien von Fletcher-Watson et al. [F116] sowie Sampath et al. [SAI13] verwiesen. Fletcher-Watson et al. zeigen anhand des Beispiels der Entwicklung einer iPad-App für Kinder im Vorschulalter mit Autismus, wie alle beteiligten NutzerInnen in den Entwicklungsprozess einbezogen werden können. So wurden nicht nur Kinder mit Autismus in Designfragen involviert, sondern darüber hinaus auch Kinder ohne Behinderung, Eltern, pädagogische Fachkräfte sowie LogopädInnen. Auch Sampath et al. haben sowohl von Autismus betroffene Kinder, als auch ihre LehrerInnen und Eltern in den Designprozess von unterstützenden Apps einbezogen, das Design immer wieder adaptiert und an die Bedürfnisse *aller* NutzerInnen angepasst.

Benton und Johnson [BJ15] liefern einen Überblick über insgesamt 39 Fallstudien aus aller Welt, die jeweils unterschiedliche Spielarten von partizipativem Design für die Entwicklung (assistiver) Technologien für Kinder mit körperlichen und/oder geistigen Beeinträchtigungen angewendet haben. Die Autorinnen unterstreichen in diesem Zusammenhang nicht nur den Gewinn von partizipativem Design für EntwicklerInnen, die so adäquate Geräte und Software herstellen können, sondern besonders auch die positiven Effekte für die Kinder, die in Designprozesse einbezogen werden. So verweisen die Fallstudien beispielsweise auf eine Stärkung des Selbstbewusstseins, gesteigerte Teamfähigkeiten, bessere Selbstwahrnehmung, Gefühle von Kompetenz und Stolz und ebenso: Spaß.

Mit der Entwicklung solcher interdisziplinären Methoden zur Erfassung der Bedarfe von Menschen mit Behinderung und der Gestaltung bildender inklusiver Interaktionen können Doing-Disability-Prozesse möglicherweise präventiv reduziert und Teilhabechancen erhöht werden – und zwar von Anfang an.

Literaturverzeichnis

- [Ba96] Baacke, D.: Medienkompetenz als Netzwerk: Reichweite und Fokussierung eines Begriffs, der Konjunktur hat. Medien praktisch 2/1996, 4-10, 1996.
- [Ba10] Baym, Nancy K.: Personal Connections in the Digital Age. Polity Press, Cambridge u.a., 2010.
- [BJ15] Benton, L.; Johnson, H.: Widening participation in technology design: A review of the

involvement of children with special education needs and disabilities. *International Journal of Child-Computer Interaction*, 3-4/2015, 23-40, 2015.

- [Bo12] Bosse, I.: *Medienbildung im Zeitalter der Inklusion*. Landesanstalt für Medien Nordrhein-Westfalen, Düsseldorf, 2012.
- [BAB13] Brüggemann, M.; Averteck, I.; Breiter, A.: *Förderung von Medienkompetenz in Bremer Kindertageseinrichtungen. Bestandsaufnahme und Befragung von Fachkräften in Bremen und Bremerhaven zur frühen Medienbildung*, Institut für Informationsmanagement Bremen GmbH, Bremen, 2013.
- [Fl16] Fletcher-Watson, S. et al.: *Designing for young children with autism spectrum disorder: A case study of an iPad app*. *International Journal of Child-Computer Interaction*, 2016.
- [Fu16] Fuzesi, P.: *Assistive Technologies and Shading the Maps of Responsibility*. Paper in a Workshop on Assistive Technologies and Transhumanism viewed from a Multidisciplinary and STS Perspective, Graz, 2016.
- [Ha02] Hargittai, E.: *Second-Level Digital Divide. Differences in People's Online Skills*. First Monday 4/2002, http://firstmonday.org/issues/issue7_4/hargittai/index.html, 2002.
- [He16] Hernwall, P.: 'We have to be professional' - Swedish preschool teachers' conceptualisation of digital media. *Nordic Journal of Digital Literacy*, 1/2016, 5-23, 2016.
- [Im06] Imhof, K. et al.: *Demokratie in der Mediengesellschaft*. VS Verlag für Sozialwissenschaften, Wiesbaden, 2006.
- [Kr07] Krotz, F.: *Mediatisierung: Fallstudien zum Wandel von Kommunikation*. VS Verlag für Sozialwissenschaften, Wiesbaden, 2007.
- [LC14] Lin, C.-Y.; Chang, Y.-M.: *Increase in physical activities in kindergarten children with cerebral palsy by employing MaKey-MaKey based task systems*. *Research in Developmental Disabilities*, 9/2014, 1963-1969, 2014.
- [Lu99] Luke, C.: *What Next? Toddler Netizens, Playstation Thumb, Techno-literacies*. *Contemporary Issues in Early Childhood* 1/1999, 95-100, 1999.
- [MH16] Maich, K.; Hall, C.: *Implementing iPads in the Inclusive Classroom Setting*. *Intervention in School and Clinic*, 3/2016, 145-150, 2016.
- [MHK10] Mankoff, J.; Hayes, G. R.; Kasnitz, D.: *Disability Studies as a Source of Critical Inquiry for the Field of Assistive Technology*. In (Barreto, A.; Hanson, V. L.): *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility*, 3-10, 2010.
- [Re14] Retalis, S. et al.: *Empowering children with ADHD Learning Disabilities with the Kinems Kinect learning games*. *Proceedings of the European Conference on Games Based Learning*, 469-477, 2014.
- [SAI13] Sampath, H.; Agarwal, R.; Indurkha, B.: *Assistive Technology for Children with Autism – Lessons for Interaction Design*. *Joint Proceedings of the 11th Asia Pacific Conference on Computer Human Interaction and the 5th Indian Conference on Human Computer Interaction*, 325-333, 2013.

- [Sc15] Schluchter, J.-R.: Medienbildung als Perspektive für Inklusion. Modelle und Reflexionen für die pädagogische Praxis. kopaed, München, 2015.
- [St12] Stegbauer, C.: Ungleichheit. Medien- und Kommunikationssoziologische Perspektiven. Springer VS, Wiesbaden, 2012.
- [Th04] Thiedeke, U.: Kosmopoliten: Einführung in eine Soziologie des Cyberspace. VS Verlag für Sozialwissenschaften, Wiesbaden, 2004.
- [Wa14] Walling, D. R.: Designing Learning for Tablet Classrooms. Innovations in Instruction. Springer, Cham u.a., 2014.
- [WGH12] Wallner, C.; Gruber, O.; Herczeg, P.: Kommunikative Partizipation als Sprecher: zum Standing unterschiedlicher Akteure in mediatisierten öffentlichen Diskursen. In (Stegbauer, C.): Ungleichheit. Medien- und Kommunikationssoziologische Perspektiven, Springer VS, Wiesbaden, 37-56, 2012.
- [Wi08] Willems, H.: Weltweite Welten. Internet-Figurationen aus wissenssoziologischer Perspektive. VS Verlag für Sozialwissenschaften, Wiesbaden, 2008.
- [Zo16] Zorn, I. et al.: Potentials of digital technology for participation of special needs children in kindergarten. In (Miesenberger, K. et al.): Computers Helping People with Special Needs Proceedings. Springer International Publishing. Peer-reviewed conference paper, Status: accepted, 2016 i.E..

Barrierefreie Kommunikation beim Paritätischen Wohlfahrtsverband

Gottfried Zimmermann¹, Mareen Gutscher², Jonas Häfele², Lena Krächan², Damaris Rothfuß², Hina Marquart³, Hermann Frank⁴

Abstract: Eine inklusive und barrierefreie Kommunikation in mitgliedsorientierten Organisationen beinhaltet die Vorbereitung, Durchführung und Nachbereitung von Veranstaltungen, sowie die digitale Barrierefreiheit von Dokumenten und Präsentationen. Dieses Positionspapier berichtet von einem gemeinsamen Projekt des Paritätischen Baden-Württemberg und der Hochschule der Medien in Stuttgart, bei dem ein Handbuch für barrierefreie Kommunikation entstehen soll. Ein besonderer Fokus wird dabei die Berücksichtigung neuerer Medien-Technologien für die Kommunikation sein.

Keywords: Digitale Barrierefreiheit, barrierefreie Präsentationen, Inklusion, barrierefreie Kommunikation, Mitgliederversammlung

1 Der Paritätische Baden-Württemberg

Der Paritätische Wohlfahrtsverband Baden-Württemberg⁵ ist einer der Spitzenverbände der Freien Wohlfahrtspflege und Dachverband von über 840 Mitgliedsorganisationen mit rund 4.000 sozialen Diensten und Einrichtungen aus allen Bereich der sozialen Arbeit und des sozialen Engagements. Er ist konfessionell und parteipolitisch ungebunden und steht für sachkundige und innovative soziale Arbeit zum Wohle des einzelnen Menschen und der Gesellschaft. Die Mitglieder sind gleichberechtigt – das symbolisiert der Begriff PARITÄT. Der Paritätische steht für Vielfalt, Offenheit und Toleranz sowie Solidarität und Partizipation.

2 Ausgangssituation

Die Arbeit beim Paritätischen ist von Kommunikation auf vielen Ebenen geprägt, die in

¹ Hochschule der Medien, Responsive Media Experience Research Group, Nobelstr. 10, 70569 Stuttgart, gzimmermann@acm.org

² Hochschule der Medien, Studiengang Computer Science & Media, Nobelstr. 10, 70569 Stuttgart, {mg159, jh176, lk105, dr056}@hdm-stuttgart.de

³ DER PARITÄTISCHE Baden-Württemberg e.V., Stabsstelle Presse- und Öffentlichkeitsarbeit, Hauptstr. 28, 70563 Stuttgart, marquart@paritaet-bw.de

⁴ DER PARITÄTISCHE Baden-Württemberg e.V., Stabsstelle Grundsatzfragen, Hauptstr. 28, 70563 Stuttgart, frank@paritaet-bw.de

⁵ <http://www.paritaet-bw.de>

der Regel auch Menschen mit Behinderungen einschließt. Kommunikation geschieht v.a. durch schriftliche Dokumente, mündlich bei Besprechungen, und bei offiziellen Versammlungen mit Beschlussfassung. Eine besondere Herausforderung liegt darin, dass es allen Teilnehmern möglich sein soll, vollständig an der Kommunikation teilzuhaben, auch wenn sie unterschiedliche Einschränkungen haben.

Zur Verdeutlichung wollen wir eine Präsentation in einer Mitgliederversammlung als Beispiel nehmen. Soll diese Präsentation barrierefrei erfolgen, so muss u.a. an Folgendes gedacht werden: Für gehörlose Menschen ist es wichtig, dass gesprochene Information in Schrift oder Gebärdensprache übersetzt wird. Für schwerhörige Teilnehmer sollte es eine Verstärkung des Tons durch assistive Hörhilfen geben. Blinde und stark sehbehinderte Teilnehmer brauchen eine Beschreibung der gezeigten Bilder und Diagramme. Kognitive eingeschränkte Menschen sollten eine Fassung in leichter Sprache und mit illustrierenden Bildern erhalten. Und mobilitätseingeschränkte Teilnehmer sollten die Möglichkeit haben, von zuhause aus der Präsentation zu folgen. Nicht alle dieser Informationen müssen notwendigerweise live bei der Präsentation ausgetauscht werden. Manche Informationen können auch schriftlich als Teil des Präsentationsdokuments festgehalten werden, wobei die Präsentation dann möglichst vorab allen Teilnehmern zukommen sollte.

Analoge Überlegungen müssen erfolgen, um allen Teilnehmern an einer Mitgliederversammlung die Wortmeldung und die Teilnahme an Abstimmungen (z.B. Wahlen) zu ermöglichen. Dabei ist außerdem an Menschen mit Körperbehinderungen zu denken, die eventuell ihren Arm nicht heben können, um sich zu Wort zu melden. Und an Menschen mit Sprechbehinderung, die nicht oder nur schlecht verständlich reden können.

3 Abgrenzung von existierenden Handreichungen

Die Wünsche und Überlegungen des Paritätischen sind nichts grundlegend Neues. Es gibt bereits eine Vielzahl an Leitfäden und Handreichungen zum Thema Barrierefreiheit. Die meisten Leitfäden wurden für bestimmte Zielgruppen entwickelt (zum Beispiel [LVGL15]), und der Fokus liegt oft auf baulichen Maßnahmen.

Der *Barriere-Checker* [PH13] vom Paritätischen Hessen betont die Notwendigkeit von baulicher Barrierefreiheit. Mit ihm hat bereits eine erste Sensibilisierung für das Thema im Verband stattgefunden. Auch eine Checkliste mit allgemeinen Hinweisen ist enthalten. Im Handbuch *Events für alle – Qualitätsstufen für barrierefreie Veranstaltungen* [FHE05] werden drei Qualitätsstufen hinsichtlich unbedingt notwendiger, notwendiger und wünschenswerter Ausstattung beschrieben und in Bezug auf unterschiedliche Veranstaltungsarten evaluiert. Daraus resultiert eine Unterteilung in verschiedene Grade an Barrierefreiheit, die durch das Ergreifen zutreffender Maßnahmen erlangt werden können. Viele Universitäten und andere öffentliche Einrichtungen haben eigene Richtlinien verfasst, um Hinweise zu geben, wie

Veranstaltungen in ihrem Bereich barrierefrei zu gestalten sind (zum Beispiel [RU16], [CU10]).

Der nun beim Paritätischen Wohlfahrtsverband entstehende Leitfaden baut auf bewährte Lösungen zur barrierefreien Kommunikation für Veranstaltungen auf. Er soll durch den gesamten Prozess und Ablauf einer Veranstaltung leiten. Andere Handbücher, unter anderem der *Leitfaden des Bundeskompetenzzentrums Barrierefreiheit* [BKB12], werden referenziert und deren Ergebnisse teilweise mit aufgegriffen. Der Leitfaden enthält jedoch neue Inhalte und Vorschläge, wenn es um den Einsatz neuer technischer Möglichkeiten und Assistenzsysteme geht. Gerade die Bereiche Teilnahme von zu Hause, Live-Streaming und Fragen aus dem Publikum werden hierbei vertieft betrachtet und neue Methoden der inklusiven Teilnahme diskutiert. Durch die Integration von Beispieldokumenten in den Leitfaden soll Veranstaltern nicht nur auf theoretischem Weg gezeigt werden, wie eine barrierefreie Kommunikation insbesondere bei Veranstaltungen zu erreichen ist, sondern auch eine praxisnahe Unterstützung gegeben werden.

4 Ziel und Vorgehensweise

Die in Kap. 2 geschilderten Überlegungen haben im Paritätischen Wohlfahrtsverband den Wunsch aufkommen lassen, ein Handbuch darüber zu verfassen, wie Kommunikation im eigenen und in den Mitgliedsverbänden am besten gestaltet werden sollte, um weitestgehend barrierefrei zu sein. Das Handbuch soll sich mit praktischen Richtlinien insbesondere an die Vereinsleitungen und das Vereinspersonal richten, die im Rahmen von Sozialarbeit mit ganz unterschiedlichen Menschen kommunizieren. Mit diesem Wunsch trat der Paritätische an die Hochschule der Medien heran und bat um ein gemeinsames Projekt, um dieses Handbuch in einem benutzerzentrierten Prozess zu entwickeln und fertigzustellen. Dieses Projekt wird zurzeit von vier Studierenden des Masterstudiengangs *Computer Science and Media* durchgeführt.⁶ Die Studierenden hatten zuvor ein Seminar zum barrierefreien Design von Informations- und Kommunikationstechnologien erfolgreich absolviert.

Ziel ist es, mit dem Leitfaden das Thema Barrierefreiheit weiter in das Blickfeld der Gesellschaft zu rücken. Durch den Einflussbereich und die Größe des Paritätischen ergibt sich eine hohe Reichweite, sodass ein Bewusstsein für die Bedarfe beim Thema Kommunikation für alle geschaffen werden kann. Denn trotz vieler Richtlinien, Gesetze und Beschlüssen ist in der Allgemeinheit vielmals noch keine ausreichende Sensibilisierung vorhanden, was barrierefreie Kommunikation konkret bedeutet und wie sie umgesetzt werden kann.

Es ist entscheidend, dass bei der Erstellung eines solchen Handbuchs alle Menschen beteiligt werden, die später davon betroffen sind. Deshalb wurden zunächst Workshops mit Teilnehmern durchgeführt, die jeweils stellvertretend für eine ganze Gruppe von

⁶ Die Studierenden sind Mitautoren dieses Beitrags

Menschen mit Behinderungen stehen. Die Workshops zur Anforderungsanalyse wurden in sechs Gruppen⁷ durchgeführt: Sehbehinderte und blinde Menschen; Menschen mit Körper- und Sprechbehinderung; Menschen mit Multiple Sklerose; gehörlose und ertaubte Menschen; schwerhörige Menschen; und Menschen mit kognitiven Einschränkungen. Für die Organisation der Workshops und die Gewinnung von Teilnehmern war es vorteilhaft, dass entsprechende Personen über die Mitgliedsorganisationen kontaktiert werden konnten. Kap. 5 gibt einen Einblick in die vorläufigen Ergebnisse der Workshops.

Ein besonderer Fokus des Projekts werden Überlegungen sein, wie moderne Technologien unterstützend für die Kommunikationsprozesse und Dokumente beim Paritätischen eingesetzt werden können. Dabei soll stets der Mensch im Mittelpunkt stehen, nach dessen Anforderungen sich die Technik richten soll, und nicht umgekehrt. Wir haben bereits festgestellt, dass bei den Betroffenen die Bereitschaft zum Einsatz neuer Technologien unterschiedlich ausgeprägt ist. Wir werden deshalb, wo wir den Einsatz neuer Technologien erwägen, diese immer nur als Alternative bzw. Ergänzung zu herkömmlichen Methoden empfehlen.

Das Ergebnis des Projekts wird ein Handbuch mit dem Titel „Richtlinien zur barrierefreien Kommunikation“ sein, das möglichst kurzgefasst sein soll. Dazu wird es einen Annex geben, der Begleitmaterial zur Entstehung des Handbuchs mit Hintergrundinformationen enthalten wird. Beide, das Handbuch und der Annex, werden – in barrierefreiem Format – der Öffentlichkeit zugänglich gemacht werden. Die offizielle Herausgabe durch den Paritätischen wird an dessen Mitgliederversammlung am 14. Oktober stattfinden. Kap. 6 beschreibt die vorgesehene Struktur des Handbuchs.

Ein Projektblog⁸ informiert in regelmäßigen Abständen über den Fortschritt des Projekts. Leser können Feedback zu Berichten und Zwischenergebnissen geben. Außerdem ist der Blog als Plattform des Austauschs gedacht, auf der Erfahrungen und Wünsche geteilt werden. Somit soll der Blog zum Erkenntnisgewinn des Projekts beitragen.

Im Workshop „Teilhabe an der allgegenwärtigen Kommunikation“ werden die Inhalte des Handbuchs und des Annexes als Entwurf vorhanden sein, so dass wir darüber diskutieren können. Von besonderem Interesse werden auch die gemachten Erfahrungen im Projekt, und der Entstehungsprozess des Handbuchs sein.

⁷ Die Aufteilung in sechs Gruppen folgte pragmatischen Gesichtspunkten, da das Projekt an den Semesterzeitplan der Hochschule gebunden ist. Man hätte manche Gruppen durchaus noch weiter aufteilen können, um noch mehr Zeit für die Erfassung der individuellen Bedürfnisse von Untergruppen zu haben. Zum Beispiel hätte man folgende Gruppen getrennt behandeln können: blinde und sehbehinderte Menschen; Menschen mit Körper- und Sprechbehinderung; Menschen mit unterschiedlichen kognitiven Einschränkungen (z.B. Aphasie vs. Lernbehinderung vs. Down-Syndrom). Man hätte auch Menschen mit bestimmten Mehrfachbehinderungen (Sehbehinderung + Schwerhörigkeit) für sich betrachten können. Eine ausführliche Betrachtung der menschlichen Fähigkeiten und möglichen Einschränkungen wird in [ISO14] Kap. 7 gegeben.

⁸ <http://gpii.eu/punktimquadrat/>

5 Vorläufige Ergebnisse der Workshops

Das Ziel der Workshops war es, die Barrieren und gleichzeitig die Bedürfnisse der verschiedenen Personengruppen in Bezug auf die Kommunikation im Rahmen von Veranstaltungen zu erfassen. Damit die Gespräche einem roten Faden folgen konnten, haben wir die Workshops anhand des Ablaufs einer jährlichen Mitgliederversammlung des Paritätischen strukturiert. Von der Einladung, über die Anmeldung, die Veranstaltung an sich mit ihren Vorträgen und Kulturbeiträgen, bis hin zu sämtlichen Informationsmaterialien, wurden alle wichtigen Bestandteile einer typischen Veranstaltung nacheinander abgearbeitet.

Ein Anliegen, das sowohl bei der einleitenden Frage nach der Erwartungshaltung, aber auch in der Abschlussrunde besonders auffiel, ist das Ideal, dass alle an einer Veranstaltung beteiligten Personen ein Bewusstsein für die einzelnen Barrieren und folglich die Bedarfe der einzelnen Teilnehmer entwickeln. Zudem wurde deutlich, dass die Teilnehmer derzeit auf sehr viele Barrieren bei Veranstaltungen stoßen, da das Bewusstsein der Veranstalter für Barrierefreiheit oft fehlt.

Aus den Workshops ging hervor, dass bei den Betroffenen die Bereitschaft zum Einsatz neuer Technologien zur Unterstützung der Kommunikation unterschiedlich ausgeprägt ist. Einzelne Personen nutzen diese bereits und sind dementsprechend aufgeschlossen. Die meisten Teilnehmer konnten sich die gewinnbringende Nutzung von modernen Technologien und Geräten wie beispielsweise einem Smartphone oder einem Tablet und den zugehörigen Anwendungen nur schlecht vorstellen. Dies liegt unter anderem daran, dass viele neue Entwicklungen wie zum Beispiel ein Gebärden-Avatar noch nicht reibungslos funktionieren und der Einsatz neuer Technologien oft nur einen Kompromiss darstellt, der nicht als signifikante Verbesserung wahrgenommen wird. Weiterhin können sich einige Personen nicht vorstellen, dass eine Technologie ihnen helfen könnte, da sie sie noch nicht erproben konnten. An dieser Stelle sollte erwähnt werden, dass die meisten unserer Workshop-Teilnehmer zu einer Generation zählen, in der die moderne Technik bis dato eine untergeordnete Rolle gespielt hat. Die jüngeren Teilnehmer unserer Workshops nutzen einige moderne Hilfsmittel und sind auch eher offen für Neues. Diese Beobachtung müsste in weiteren Workshops mit heterogenen Gruppen bestätigt werden, die jedoch im gegebenen zeitlichen Rahmen des Projekts nicht durchgeführt werden können. Aus den Workshops ging jedoch deutlich hervor, dass es nicht den *einen* richtigen Weg geben kann. Deshalb wird in dem Leitfaden der Einsatz neuer Technologien immer nur als Alternative bzw. Ergänzung zu herkömmlichen Methoden empfohlen (vgl. Kap. 4).

6 Vorgesehene Struktur des Handbuchs für barrierefreie Kommunikation

Die Struktur des Leitfadens orientiert sich an dem chronologischen Ablauf einer

Veranstaltung, woraus sich die Teile „Vor der Veranstaltung“, „Während der Veranstaltung“ und „Nach der Veranstaltung“ ergeben. In dem Teil zur *Vorbereitungsphase* geht es um die Wahl eines geeigneten Veranstaltungsortes, der eine barrierefreie Durchführung einer Veranstaltung ermöglicht. Dies beinhaltet ebenfalls die Prüfung der vorhandenen technischen Voraussetzungen und zusätzlich zu organisierende Maßnahmen. In dem Teil zur *Einladungsphase* geht es darum, wie Referenten und Teilnehmer eingeladen werden können. Dies beinhaltet die entsprechende Vorbereitung von Dokumenten, Vorträgen, Vortragsunterlagen, etc. In dieser Phase ist es äußerst wichtig, ein Bewusstsein und ein Selbstverständnis für Barrierefreiheit zu schaffen. Bei der Anmeldung ist es wichtig, die benötigten Dienste abzufragen, die in der Einladung bereits angeboten wurden, um den Bedarf für die Veranstaltung festzustellen.

Der Punkt *Veranstaltungen* behandelt die Bereiche Vorträge, Interaktion mit dem Publikum, Kommunikation während den Pausen und Abstimmungen. Hier ist das Verständnis der Inhalte für alle Teilnehmer sicherzustellen. Was muss beachtet werden, um Vortragsunterlagen und den Vortrag an sich barrierefrei zu gestalten? Für das Publikum spielt hier besonders die aktive Teilnahme eine Rolle, beispielsweise auch durch Live-Streaming von Zuhause.

In der *Nachbereitung* ist es wichtig, die entstandenen Informationen für die Rezipienten aufzubereiten und anzubieten. Dafür bietet der Leitfaden praktische Maßnahmen.

Danksagung

Wir danken den Selbstvertretern aus den folgenden Vereinen für die engagierte Unterstützung des Projekts:

- AMSEL, Aktion Multiple Sklerose Erkrankter Landesverband der DMSG in Baden-Württemberg e.V.
- Landesverband Baden-Württemberg der Lebenshilfe für Menschen mit Behinderung e.V.
- Körperbehinderten-Verein Stuttgart e.V.
- Blinden- und Sehbehindertenverband Württemberg e.V.
- Deutscher Schwerhörigenbund Landesverband Baden-Württemberg e.V.

Literaturverzeichnis

- [BKB12] Bundeskompetenzzentrum Barrierefreiheit (BKB). (2012). Handreichung und Checkliste für barrierefreie Veranstaltungen. Bundeskompetenzzentrum Barrierefreiheit (BKB). Retrieved June 20, 2016, from http://www.barrierefreiheit.de/handreichung_und_checkliste_f%C3%BCr_barrierefreie_Veranstaltungen.html

- [CU10] Cornell University. (2010). Guide to planning an accessible event. Cornell University. Retrieved June 20, 2016, from http://sds.cornell.edu/Guide_to_Planning_an_Accessible_Event.pdf
- [FHE05] Institut Verkehr und Raum des Fachbereichs Verkehrs- und Transportwesens der Fachhochschule Erfurt. (2005). Events für Alle - Qualitätsstufen für barrierefreie Veranstaltungen. FH Erfurt. Retrieved June 20, 2016, from <https://www.fh-erfurt.de/fhe>
- [ISO14] ISO/IEC. (2014, December 1). ISO/IEC Guide 71. Guide for addressing accessibility in standards. Second edition. Retrieved June 20, 2016, from <http://www.iso.org/guide71>, 2016-06-07.
- [LVGL15] Landesverband der Gehörlosen in BW. (2015, May 14). Landesverband der Gehörlosen Baden-Württemberg - Ein Leitfaden zur barrierefreien Kommunikation im kulturellen Bereich. Retrieved June 29, 2015, from <http://www.lv-gl-bw.de/aktuelle-news/items/ein-leitfaden-zur-barrierefreien-kommunikation-im-kulturellen-bereich.html>.
- [PH13] Der Paritätische Hessen. (2013, October 21). Der Barriere-Checker. Der Paritätische Hessen. Retrieved June 20, 2016, from <http://www.paritaet-hessen.org/ueberuns/aktuelles/details/show-news/der-barriere-checker.html>
- [RU16] Ryerson University. (2016, June 20). Tools & Resources - Accessibility - Ryerson University. Retrieved June 20, 2016, from <http://www.ryerson.ca/accessibility/tools/communicationsevents.html>

Barrierefreiheit im Web für Bildungseinrichtungen

Dieter Meiller¹, Karlheinz Müller², Hildegard Legat³ und Bernd Gerlang⁴

Abstract: Bisher gibt es im regionalen Raum keine Vernetzung im Bereich barrierefreie Bildung und Freizeit. Die „Offene Behindertenarbeit“ schließt sich daher mit örtlichen und regionalen Bildungsträgern, mit dem örtlichen Fernsehsender und der nahe gelegenen Hochschule zusammen, um grundlegende und strukturelle Verbesserungen voranzubringen. Die Arbeitsgruppe der Hochschule fokussiert sich dabei auf die Beratung und Entwicklung im Bereich barrierefreies Web.

Keywords: Barrierefreiheit, Bildung, Accessibility, Menschen mit Behinderungen, Inklusion



Abb. 1: Projektlogo

1 Anforderungen an ein barrierefreies Web

Eine wesentliche Voraussetzung, um Menschen mit Behinderungen die Teilnahme an Bildungsangeboten zu ermöglichen, ist es, dass diese sich über geeignete Kurse und Angebote informieren können. Dies geschieht vorwiegend über die Web-Auftritte der entsprechenden Einrichtungen. Die Webangebote müssen dabei bestimmte technische Voraussetzungen erfüllen, um den sehr unterschiedlichen Behinderungsarten gerecht zu werden [W309]. Neu entstehende Web-Auftritte sollen möglichst von Anfang an entsprechend entwickelt, bestehende Angebote überprüft und angepasst werden. Dies wird mittlerweile sogar gesetzlich vorgeschrieben [Bj11]. So setzen blinde Menschen technische Hilfen ein, um zu navigieren und um eine Webseite zu lesen. Es gibt sogenannte Braille-Zeilen, die den Text auf der Seite in Blindenschrift darstellen. Menschen können dann Buchstaben der Blindenschrift erfühlen. Personen mit einem eingeschränkten Bewegungsapparat können oft keine Maus oder Tastatur zur Navigation bedienen und sind auf andere technische Hilfsmittel angewiesen. Um die Navigation mit solchen Hilfsmitteln zu ermöglichen, muss die Technik der Webseite entsprechend

¹ OTH Amberg-Weiden, Fakultät EMI, Kaiser-Wilhelm-Ring 23, 92224 Amberg, d.meiller@oth-aw.de

² OTH Amberg-Weiden, Fakultät EMI, Kaiser-Wilhelm-Ring 23, 92224 Amberg, k.mueller@oth-aw.de

³ OBA der Lebenshilfe A.-S. e. V, Fallweg 43, 92224 Amberg, hildegard.legat@lebenshilfe-amberg.de

⁴ OTH Amberg-Weiden, Fakultät EMI, Kaiser-Wilhelm-Ring 23, 92224 Amberg, be.gerlang@oth-aw.de

beschaffen sein und bestimmte Anforderungen erfüllen. Zusätzlich zu den technischen Voraussetzungen muss eine barrierefreie Webseite besonders gestaltet werden. Beispielsweise soll der Kontrast zwischen Text und Hintergrund groß sein, auch sollen keine schnellen Animationen ablaufen. Menschen mit Leseschwierigkeiten profitieren zusätzlich von einer einfach und klar gehaltenen Sprache und von eindeutigen Symbolen. Folgende Arten von Behinderungen sollen bei der inhaltlichen und technischen Realisierung der Webangebote berücksichtigt werden [OI13].

Sehen: Schlechte Sehleistung (Visus) oder weniger als 5% (Blindheit); Rot-Grün-Schwäche

Hören: Angeborene oder erworbene Gehörlosigkeit (die Gebärdensprache hat eine andere Grammatik als die Schriftsprache)

Bewegen: Eingeschränkter Bewegungsapparat oder vollständige Bewegungsunfähigkeit

Geistige Einschränkungen: (Legastenie, Dyslexie) Verständnisschwierigkeiten bei komplexen Sachverhalten

Mehrfachbehinderungen: Meist treten die genannten Einschränkungen mehrfach auf.

2 Arbeitsschritte

Zur Umsetzung der beschriebenen Arbeiten wurde an der Hochschule eine halbe Stelle für einen Projekttechniker geschaffen, die ab Oktober 2015 besetzt wurde. Dieser soll die Arbeitspakete zusammen mit studentischen Hilfskräften und Studierenden im Rahmen von Projektarbeiten abarbeiten. Die geleisteten Arbeiten werden dokumentiert, damit eine Weiterführung der Arbeiten durch andere Personen möglich wird. Die fachlichen Koordinatoren sind zwei Professoren, jeweils zuständig für die grafischen und die technischen Arbeiten. Folgende Arbeitspakete werden ausgeführt:

2.1 Umsetzung einer barrierefreien Projekt-Webseite

Auf Basis eines Content-Management-Systems wurde eine barrierefreie Projekt-Webseite erstellt, die im weiteren Verlauf in Kooperation mit den anderen Projektpartnern angepasst und gefüllt werden kann. Diese soll zugleich als „Best Practice“-Beispiel dienen. Dazu gehört der Entwurf eines gut wahrnehmbaren Designs, die Auswahl geeigneter Farben und Schriften und die Codierung eines sogenannten Responsive-Webdesign Templates: Hierbei handelt es sich um eine Vorlage, die das Aussehen der gesamten Seiten bestimmt. Diese passt sich an verschiedene Endgeräte an, also an verschiedene Bildschirmgrößen, auch an Handys und Tablet-Computer sowie besondere Hilfsgeräte für Menschen mit Behinderungen. Der Quellcode des Templates soll als Open Source Projekt veröffentlicht und so auch anderen Bildungseinrichtungen zur Verfügung gestellt werden. Der aktuelle Stand der Forschung und Technik zum Thema Barrierefreie Webseiten sowie die rechtlichen Rahmenbedingungen wird recherchiert und dokumentiert. Die Erkenntnisse sollen später auf der Projektwebseite zugänglich gemacht werden.

2.2 Überprüfung von Webseiten auf Barrierefreiheit

Die Webseiten der Projektpartner werden auf Barrierefreiheit überprüft. Dies erfolgt mit technischen Methoden wie dem Colour Contrast Analyser (CCA), der Web Accessibility Toolbar (WAT) oder dem Accessibility Viewer (aViewer) [Pg16] sowie in Kooperation mit Testnutzern, also Menschen mit Behinderungen. Die Projektpartner sowie weitere interessierte Institutionen werden bei der Überarbeitung und Erstellung von barrierefreien Angeboten im Web unterstützt und beraten.

2.3 Entwicklung einer Symbolschrift

Als Orientierungshilfe für Menschen mit geistigen Einschränkungen soll eine Symbolschrift entworfen und umgesetzt werden, die auf Webseiten sowie auf gedruckten Medien einsetzbar ist. Vorhandene Symbolschriften sind für unser Vorhaben ungeeignet, wie von uns durchgeführte Nutzerbefragungen ergeben haben. Bliss [Bl65] beispielsweise wurde als zu abstrakt empfunden, zudem keine geeigneten Symbole für unsere Inhalte gefunden werden. Nach dem Entwurf von Symbolen soll ein digitaler Font als Reinzeichnung realisiert werden. Die fertige Symbolschrift soll dann unter einer Creative-Commons Lizenz im Internet veröffentlicht werden. Hierbei ist die rechtliche Situation genau zu überprüfen.

2.4 Durchführung von Benutzertests zusammen mit Menschen mit Behinderung

Wie erwähnt, sind für die Bedürfnisse von Menschen mit Behinderungen spezielle Ein- und Ausgabehilfen zu berücksichtigen. Dazu ist es nötig, einige Geräte für Testzwecke anzuschaffen und Tests in Kooperation mit betroffenen Menschen durchzuführen [Sm05]. Die Symbolschrift und die Internetseiten werden in Zusammenarbeit mit Menschen mit Behinderungen auf Lesbarkeit und Verständlichkeit überprüft und überarbeitet. Hierzu wird auch ein Eye-Tracking-System eingesetzt, welches es ermöglicht, den Blickverlauf der Probanden auf der Webseite zu verfolgen.

3 Bisher gesammelte Erfahrungen

Man kann keine praxistauglichen „Produkte“ von den Studierenden erwarten, allerdings von Mitarbeitern schon, deshalb: Die Einstellung von Projekt-Personal ist auch vorteilhaft für den Fortschritt von Projekten mit sozialem Engagement. Man muss trennen zwischen Prüfungsleistungen und Erwartungen des Projektpartners. Die zeitliche und räumliche Zusammenarbeit mit Menschen mit Behinderungen kann problematisch sein, allerdings ist der zwischenmenschliche Kontakt problemlos. Man sollte frühzeitig die Rechtsabteilung einbinden, einen Kooperationsvertrag ausarbeiten und dort auch verhandeln. Projekte müssen wachsen, man muss mit kleinen Schritten vorangehen und nicht zu viel erwarten. Engagement zahlt sich aus.

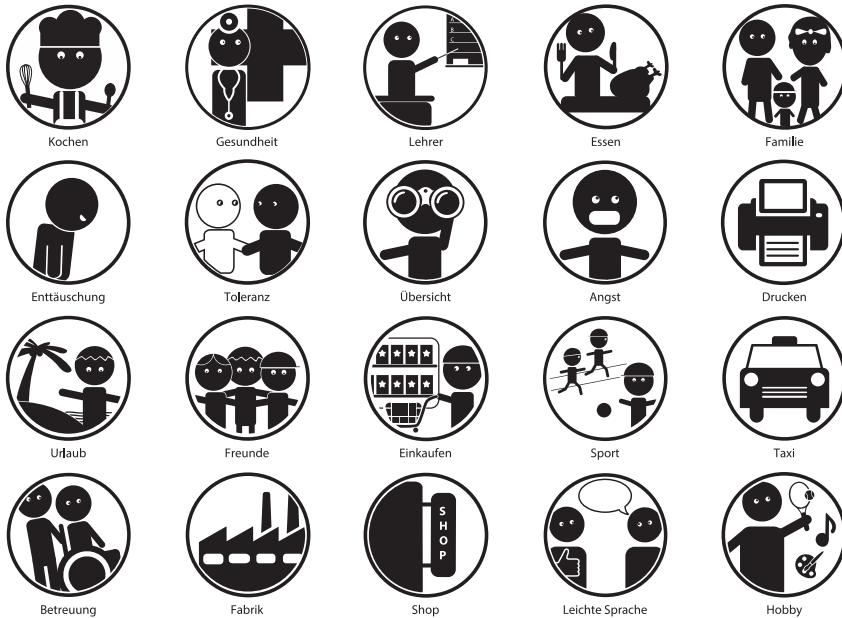


Abb. 2: Erster Entwurf einer Symbolschrift

Literaturverzeichnis

- [Bl65] Bliss, C. K.: Semantography (Blissymbolics): A Simple System of 100 Logical Pictorial Symbols, which Can be Operated and Read Like 1+ 2. Semantography (Blissymbolics) Publications, 1965.
- [Bj11] Verordnung zur Schaffung barrierefreier Informationstechnik nach dem Behindertengleichstellungsgesetz (BITV 2.0), http://www.gesetze-im-internet.de/bitv_2_0/, Stand: 05.06.2016.
- [Ol13] De Oliveira, D.: Barrierefreiheit im Internet. epubli GmbH, 2013.
- [Pg16] The Paciello Group: Resources | The Paciello Group – Your Accessibility Partner (WCAG 2.0/508 audits, VPAT, usability and accessible user experience), <https://www.paciellogroup.com/resources/>, Stand: 05.06.2016.
- [Sm05] Small, J., Schallau, P. Brown, K. & Appleyard, R.: Web accessibility for people with cognitive disabilities. In CHI '05: CHI '05 extended abstracts on Human factors in computing systems. ACM, New York, S. 1793-1796, 2005.
- [W309] W3C (2009), Web Content Accessibility Guidelines (WCAG) 2.0, <http://www.w3.org/Translations/WCAG20-de>, Stand: 05.06.2016.

Zufriedenheit blinder Nutzer mit Zugangsmethoden zu Sozialen Netzwerken: eine theoretische Modellentwicklung

Petra Gröber¹

Abstract: Blinde greifen auf die mobile Webseite von Sozialen Netzwerken zurück, weil die reguläre Webseite aufgrund von hoher Komplexität und Inkonsistenz unzugänglich ist. Dadurch steht ihnen nicht der gleiche Informations- und Funktionsumfang zu Verfügung. Doch was genau veranlasst Blinde einen bestimmten Zugangsmodus zu wählen bzw. zu meiden? Dieses Paper beschreibt die theoretische Herleitung eines auf der IS Success-Forschung basierenden Modells zur Erfolgsmessung der Zugangsmodi von Sozialen Netzwerken (reguläre Webseite, mobile Webseite, App).

Keywords: Soziale Netzwerke, blinde Nutzer, Nutzerzufriedenheit, theoretisches Modell

1 Einleitung

Soziale Netzwerke (SN) ermöglichen eine weltweite Kontaktpflege, erlauben eine Selbstdarstellung und vermitteln eine Art Gemeinschaftssinn. Im privaten Kontext sind SN, z.B. Facebook, sehr erfolgreich. Aber auch Unternehmen nutzen öffentliche SN zur Kundenwerbung, Kundenbindung, Kundenbefragung oder zur Mitarbeiterakquise und vermehrt interne SN für das Wissensmanagement innerhalb des Unternehmens [He10].

Die meisten SN haben eine reguläre Webseite und bieten zusätzlich mobile Webseiten und Apps an, um von überall auf das SN zugreifen zu können. Die komplexen regulären Webseiten sind für Blinde schwer zugänglich [Ba14], deshalb nutzen sie, auch am Desktop-Computer, die mobile Webseite. Eine mobile Webseite ist aber nur eine Ergänzung und muss den Anforderungen eines mobilen Endgerätes genügen, das bedeutet das Layout und die Interaktionen sind einfach und klar gehalten. Für die Nutzung eines Screenreaders ist die mobile Webseite daher besser geeignet, allerdings sind die Inhalte und Funktionalitäten nicht im gleichen Umfang wie auf der regulären Webseite vorhanden [Gr15]. Nutzer, die nur die mobile Webseite verwenden, sind benachteiligt, denn bestimmte Informationen und Funktionalitäten stehen ihnen nicht zur Verfügung. Um volle Zugänglichkeit zu SN zu gewährleisten, müsste entweder die reguläre Webseite umfassend zugänglich gemacht werden oder die mobilen Versionen den gleichen Informations- und Funktionsgehalt wie die reguläre Webseite anbieten.

Menschen mit Behinderungen dürfen nicht ausgeschlossen werden. Bei dem stark visuell geprägten Internet sind Blinde allerdings enormen Schwierigkeiten ausgesetzt. Laut

¹ Universität Rostock, Institut für Informatik, Albert-Einstein-Str. 22, 18059 Rostock, petra.groeber@uni-rostock.de

Babu [Ba14] ist eine erfolgreiche Nutzung von Facebook nicht möglich, so dass Blinde Webseiten meiden, wenn diese zu komplex und Funktionalitäten nicht nutzbar sind.

Doch was genau veranlasst einen Blinden einen bestimmten Zugangsmodus (reguläre Webseite, mobile Webseite, App) zu nutzen oder vielmehr nicht zu nutzen? Soziale Netzwerke sind Informationssysteme (IS), daher wurde, basierend auf der *IS Success* Forschung, ein Modell entwickelt, das den Erfolg eines Zugangsmodus misst. Angewandt auf die jeweils drei Zugangsmodi soll das Modell zeigen, warum bestimmte Zugangsmodi bei blinden Nutzern erfolgreicher sind als andere. Dieses Paper beschreibt die theoretische Herleitung sowie die einzelnen Bestandteile des Modells..

2 Theorie

Die IS Success-Forschung bringt zwei parallele Entwicklungen hervor – das *Technology Acceptance Model (TAM)* [Da86] und das *IS Success Model* [DM03]. Das TAM basiert auf einem verhaltenstheoretischen Ansatz und hat das Ziel, Nutzung von IS zu erklären und zu prognostizieren. Das IS Success Model basiert darauf, dass Zufriedenheit mit einer Aufgabe, positive Auswirkung auf dessen Ausführung hat. Während das TAM als Vorhersagemodell genutzt werden kann, aber dabei unzureichend auf Systemeigenschaften eingeht, geht das IS Success Model zwar auf Systemeigenschaften ein, kann aber die Nutzung eines Systems nicht prognostizieren.

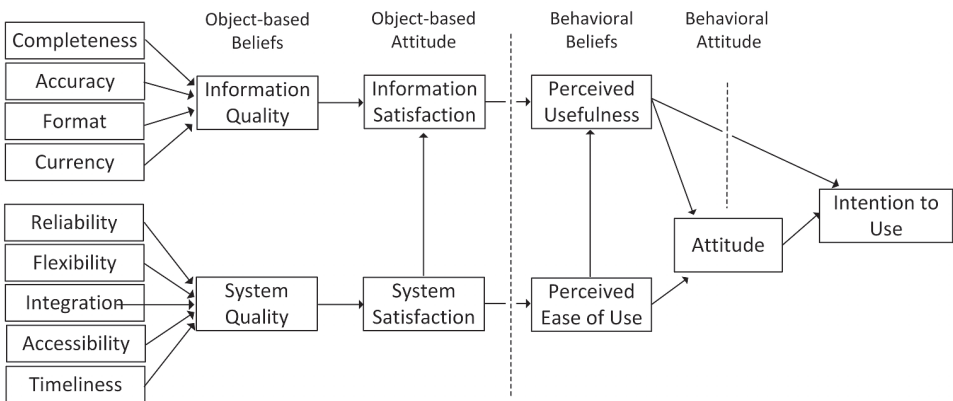


Abb.: 1: Kombiniertes Modell von Wixom & Todd.

Wixom & Todd [WT05] kombinieren die beiden Entwicklungen, um deren Vorteile zu vereinen (Abb.: 1). Rechts im Modell ist das TAM und links das IS Success Model erkennbar. Die Verknüpfung beide Modelle besteht in der Annahme, dass die *object-based attitudes* auf die *behavioral beliefs* wirken. Das bedeutet, die Zufriedenheit mit

den vom System produzierten Informationen beeinflusst die Wahrgenommene Nützlichkeit und die Zufriedenheit mit dem System selbst beeinflusst die Wahrgenommene Einfachheit. Mit anderen Worten: (1) Je zufriedener ein Nutzer mit den bereitgestellten Informationen ist, desto nützlicher wird die Anwendung wahrgenommen und (2) Je zufriedener ein Nutzer mit dem System selbst ist, desto leichter wird die Bedienung der Anwendung wahrgenommen.

Aber ist es nicht anders herum? Die nützlichen Informationen und die einfache Bedienung, führen zu Zufriedenheit: Je nützlicher die bereitgestellten Informationen und je einfacher die Bedienung des Systems, desto zufriedener ist der Nutzer.

In dem resultierenden Modell (Abb.: 2, grau hinterlegt) sind nun die *Attitudes* rechts und die *Beliefs* links dargestellt. Die *object-based Beliefs* beeinflussen die *behavioral Beliefs*, die dann auf die *object-based Attitudes* wirken. Das bedeutet, glaubt ein Nutzer die Eigenschaften eines Systems sind von hoher Qualität, beeinflusst dies die wahrgenommene Nützlichkeit und Einfachheit des Systems, und somit den Glauben ein System erfolgreich nutzen zu können. Mit diesem positiven Glauben steigt die Nutzerzufriedenheit (object-based Attitude) und somit die Intention das System zu nutzen.

Diese Wirkungskette – *externen Variablen* (siehe 3.2) – *beliefs* – *attitudes* – *Nutzung eines Systems*, findet sich auch in [DT91], die zeigen, dass die Erfolgsmessung eines IS den folgenden typischen Verlauf hat: *Causal Factors* → *Beliefs* → *Attitude* → *Use*.

Das Konstrukt *Attitude* wurde aus dem Modell entfernt, da sowohl die Wahrgenommene Nützlichkeit als auch die Wahrgenommene Einfachheit direkten Einfluss auf das Nutzerverhalten haben. Außerdem ist der Erklärungsgehalt mit oder ohne dem Konstrukt *Attitude* gleich gut.

3 Konstruktdefinitionen

Im Folgenden werden die Konstrukte des in Abb.: 2 dargestellten Modells beschrieben. Dabei werden die abhängigen und die unabhängigen Variablen unterschieden. Für die Messung der Konstrukte wird auf valide Skalen zurückgegriffen. Die entsprechenden Indikatoren sind in den Tabellen Tab. 1 und Tab. 2 zusammengefasst.

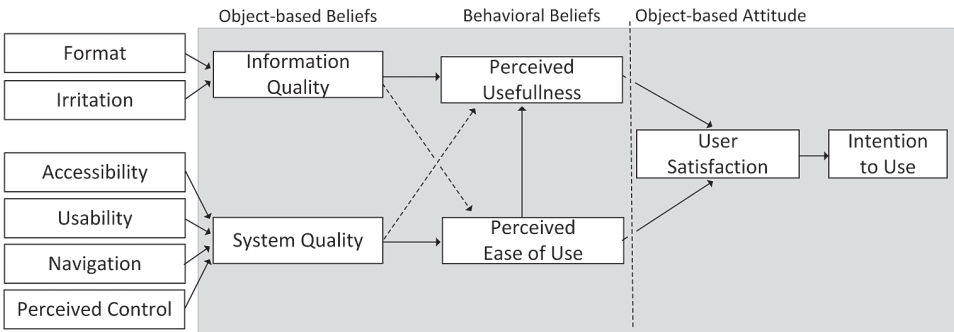


Abb.: 2: Modell zur Erfolgsmessung eines Zugangsmodus (reguläre und mobile Webseite, App).

3.1 Abhängige Variablen

Nutzungsintention Die Nutzerintention ist definiert als der Grad der Absicht den betrachteten Zugangsmodus eines SN weiter zu nutzen [WD85].

Nutzerzufriedenheit Die Nutzerzufriedenheit ist der Grad der Zufriedenheit mit den betrachteten Zugangsmodus eines SN [DM03].

Wahrgenommene Nützlichkeit Die Nutzung eines bestimmten Zugangsmodus erleichtert unter Umständen die Durchführung einer bestimmten Aufgabe, z.B. nutzen Blinde die mobile Webseite, weil diese am zugänglichsten ist, allerdings wechseln sie den Zugangsmodus um bestimmte Funktionen, z.B. Fotos hochladen, besser oder überhaupt nutzen zu können [Vo16]. Die Wahrgenommene Nützlichkeit ist der Grad, zu welchem ein Nutzer glaubt die Verwendung eines bestimmten Zugangsmodus verbessert die Durchführung einer bestimmten Aufgabe [Da86].

Wahrgenommene Einfachheit Für Blinde ist die Web-Nutzung mit hohem mentalen Aufwand verbunden. Die reguläre Facebook-Seite z.B. ist für Blinde schwierig und zeitaufwendig zu bedienen [Vo16], [Ma12]. Die Wahrgenommene Einfachheit ist der Grad, zu welchem ein Nutzer glaubt, die Verwendung eines bestimmten Zugangsmodus bedeutet am wenigsten physischen und mentalen Aufwand und Anstrengung [DBW89].

Informationsqualität Da es sich bei SN um nutzergenerierte Inhalte handelt, lässt sich die inhaltliche Qualität schwer messen und vergleichen. Zudem liegt der Fokus dieser Studie im Vergleich der verschiedenen Zugangsmodi, wo gleiche Inhalte unterschiedlich dargestellt werden. Daher ist die Informationsqualität definiert als die Qualität der Darstellung von Informationen mit einem bestimmten Zugangsmodus.

Systemqualität Bei SN handelt es sich um Webseiten bzw. Web-Applikationen für mobile Endgeräte. Für einer erfolgreiche Nutzung spielt die Interaktion mit der Webseite bzw. mit der App eine entscheidende Rolle, demnach bezieht sich die Systemqualität auf die Qualität der Interaktionsmöglichkeiten mit dem betrachteten Zugangsmodus.

<p>Nutzerintention [Bh01] Ich werde die ... nicht weiter benutzen. Ich bevorzuge die Nutzung der ... im Vergleich zu den anderen Zugangsmodi. Ich beabsichtige die ... weiter zu nutzen, anstatt diese nicht weiter zu verwenden.</p>
<p>Nutzerzufriedenheit [Bh01], [MYZ02] Unter Berücksichtigung verschiedener Aspekte bin ich mit der ... sehr zufrieden. Die ... entspricht im Großen und Ganzen meinen Erwartungen. Ich bin frustriert über die</p>
<p>Wahrgenommene Nützlichkeit [Ve00] Über die ... auf das SN zuzugreifen, ist nützlicher als die anderen Zugangsmodi zu benutzen. Im Vergleich zu den anderen Zugangsmethoden, erlaubt mir die Nutzung der ...: ... eine effizientere Nutzung. ... eine effektivere Nutzung.</p>
<p>Wahrgenommene Einfachheit [MK01] Die ... zu benutzen, erfordert einen hohen mentalen Aufwand. Die Bedienung der ... ist schwer zu erlernen. Meine Interaktion mit der ... ist klar und verständlich. Es ist unmöglich die ... ohne Hilfe zu benutzen.</p>
<p>Informationsqualität [WT05] Insgesamt bewerte ich die Darstellung von Informationen auf der ... als sehr gut. Grundsätzlich ist die Informationsdarstellung auf der ... qualitativ hochwertig. Insgesamt gebe ich der ... für die Informationsdarstellung eine sehr gute Note.</p>
<p>Systemqualität [WT05] Bezüglich der Interaktionen, bewerte ich die ... als sehr gut. Insgesamt ist die Interaktion mit der ... qualitativ hochwertig. Insgesamt gebe ich der Interaktion mit der ... eine sehr gute Bewertung.</p>

Tab. 1: Die Indikatoren zur Messung der abhängigen Variablen. In den Lücken des Textes werden jeweils die entsprechenden Zugangsmodi (reguläre Webseite, mobile Webseite, App) ergänzt

3.2 Unabhängige Variablen

Entscheidend für eine gute Darstellung von Inhalten ist das zugrunde liegende Layout einer Webseite. Die Nutzung der regulären Webseite von Facebook ist für Blinde eher frustrierend, da sie aufgrund der Komplexität und Inkonsistenz als verwirrend, unklar, unverständlich und unvorhersehbar wahrgenommen wird. Die mobile Version dagegen wird als einfach, klar, sortiert und verständlich empfunden und somit dessen Nutzung als eher Freude bringend beurteilt [RJ15]. Für ein erfolgreiches Nutzungserlebnis sind neben einer übersichtlichen Struktur aber auch eine klare und intuitive Interaktion erforderlich. Diese ist für blinde Nutzer bei Facebook mitunter sehr zeitintensiv. Nach [Vo16] sind Blinde daher frustriert und fühlen sich ausgeschlossen, wenn Inhalte nicht zugänglich sind.

Im folgenden werde die Faktoren erläutert, die die Informations- und die Systemqualität beeinflussen. Die Indikatoren zur Messung der Faktoren sind Tab 2. zusammengefasst. Bei allen Faktoren sind Unterschiede sowohl zwischen den drei Zugangsmodi als auch zwischen den beiden Nutzergruppen zu erwarten.

Format bezieht sich darauf, wie die Informationsdarstellung wahrgenommen wird. Aufgrund der unterschiedlichen mentalen Modelle blinder und sehender Nutzer, wird die Darstellung der Informationen sehr wahrscheinlich unterschiedlich wahrgenommen und bewertet.

Irritation oder auch *Web-Organisation* bezieht sich darauf, inwieweit eine Webseite als unklar, unübersichtlich und irritierend empfunden wird [Lu02]. Der Faktor gibt Aufschluss darüber, wie unterschiedlich die Struktur der verschiedenen Zugangsmodi von Blinden und Sehenden wahrgenommen wird.

Zugänglichkeit ist die Einfachheit in einem System auf Informationen zugreifen zu können [WT05] und ist von sehr hoher Bedeutung, insbesondere für blinde Nutzer, wenn sie auf komplexen Webseiten surfen wollen [Ba14].

Benutzerfreundlichkeit Neben der Zugänglichkeit ist auch die Benutzerfreundlichkeit von großer Bedeutung, denn auch wenn ein System als zugänglich gilt, ist es nicht unbedingt benutzerfreundlich. McKinney et al. [MYZ02] definieren Benutzerfreundlichkeit über die visuelle Attraktivität, die Konsistenz, Spaß und einfache Nutzung. Da die visuelle Attraktivität bei blinden Nutzern eine untergeordnete Rolle spielt, wird dieser Aspekt aufgrund der Vergleichbarkeit zwischen sehenden und blinden Nutzern ausgeklammert.

Navigation bewertet die angebotenen Links zu vorhandenen Informationen. Für ein zufriedenstellendes Nutzererlebnis ist die Navigation ein wesentlicher Bestandteil [MYZ02]. Dabei spielen die Zweckmäßigkeit der Links, ihre eindeutige Beschreibung, die leichte Auffindbarkeit, das einfache hin und her springen sowie die Erreichbarkeit gewünschter Informationen eine Rolle.

Wahrgenommene Kontrolle Wahrgenommene Kontrolle bezieht sich darauf, wie sicher und komfortabel sich ein Nutzer auf einer Webseite fühlt. Aufgrund der Schwierigkeiten, mit denen blinder Nutzer auf komplexen Webseiten konfrontiert sind, ist zu erwarten, dass sie sich im Vergleich zur mobilen Webseite auf der regulären Webseite weniger sicher und komfortabler fühlen.

4 Ausblick

Das vorgestellte Modell ist eine theoretische Vorüberlegung zur Erfolgsmessung der Zugangsmodi von SN. Basierend auf diesem Modell werden mittels einer Online-Umfrage blinde und sehende SN-Nutzer zu den verschiedenen Zugangsmodi (reguläre Webseite, mobile Webseite, App) befragt. Für die Operationalisierung des Modells wird

auf valide Indikatoren aus der Literatur zurückgegriffen (siehe Tab. 1 und Tab. 2.). Aufgrund des gewählten Messinstrumentes ist eine hohe Teilnehmerzahl notwendig, daher wird die Umfrage auf Deutsch und Englisch angeboten, um diese international zu streuen. Die beiden Nutzergruppen und die drei unterschiedlichen Zugangsmodi erlauben einen Vergleich zwischen sechs Gruppe: (1) blinde Nutzer der regulären Webseite, (2) blinde Nutzer der mobilen Webseite, (3) blinde Nutzer der App, (4) sehende Nutzer der regulären Webseite, (5) sehende Nutzer der mobilen Webseite, (6) sehende Nutzer der App. Dieser Vergleich zwischen sehenden und blinden Nutzern sowie zwischen den Zugangsmodi hilft zu verstehen, worauf es bei der Entwicklung von SN ankommt, um diese für Blinde zugänglich zu machen.

<p>Format [WT05] Die bereitgestellten Beiträge und Daten auf der ... sind: sehr gut formatiert. sehr gut gestaltet. sehr ansprechend.</p>
<p>Irritation [AHP14], [Lu02], [CW99] Die ... nervt mich. Ich finde die ... verwirrend. Die ... ist chaotisch. Insgesamt sieht die ... organisiert aus.</p>
<p>Accessibility [WT05] Mit der ... sind Informationen jederzeit abrufbar für mich. Die ... ist sehr zugänglich. Die ... ist unzugänglich.</p>
<p>Perceived Control [SZ08] Auf der ... ist es einfach dorthin zu gehen, wo ich hingehen will. Während meines Besuchs auf der ... habe ich ein hohes Maß an Kontrolle. Die ... ist überhaupt nicht handhabbar.</p>
<p>Navigation [MYZ02] Auf der ... ist es einfach, zwischen den Seiten hin und her zu springen. Die ... unterstützt es mit nur wenigen Klicks Informationen zu finden. Im Allgemeinen ist die ... einfach zu navigieren. Die ... verfügt über ausreichende Suchmöglichkeiten.</p>
<p>Usability [Br96] Ich finde die ... unnötig komplex. Es gibt zu viel Inkonsistenz auf der Ich finde die ... sehr umständlich zu bedienen. Ich fühle mich mit der Nutzung der ... sehr vertraut. Die ... besitzt ein klares Design. Die ... ist benutzerfreundlich.</p>

Tab. 2: Die Indikatoren zur Messung der externen Faktoren. In den Lücken des Textes werden jeweils die entsprechenden Zugangsmodi (reguläre Webseite, mobile Webseite, App) ergänzt.

Literaturverzeichnis

- [AHP14] Ahn, T.; Hong, M.; Pedersen, P. M.: Effects of perceived interactivity and web organization on user attitudes. *European Sport Management Quarterly*. 2/14; S. 111–128, 2014.
- [Ba14] Babu, R.: Can Blind People Use Social Media Effectively? A Qualitative Field Study of Facebook Usability. *American Journal of Information Systems*. 2/2; S. 33–41, 2014.
- [Bh01] Bhattacharjee, A.: Understanding information systems continuance: an expectation-confirmation model. *MIS Quarterly*. 3/25; S. 351–370, 2001.
- [Br96] Brooke, J.: SUS: A quick and dirty usability scale. In (Jordan, P. W. et al. Hrsg): *Usability Evaluation in Industry*. Taylor & Francis, London; S. 189–194, 1996.
- [CW99] Chen, Q.; Wells, W. D.: Attitude toward the Site. *Journal of Advertising Research*. 5/39; S. 27–37, 1999.
- [Da86] Davis, F. D.: A technology acceptance model for empirically testing new end-user information systems: Theory and results. Doktorarbeit, 1986.
- [DBW89] Davis, F. D.; Bagozzi, R. P.; Warshaw, P. R.: User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*. 8/35; S. 982–1003, 1989.
- [DM03] DeLone, W. H.; McLean, E. R.: The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems*. 4/19; S. 9–30, 2003.
- [DT91] Doll, W. J.; Torkzadeh, G.: The Measurement of End-User Computing Satisfaction: Theoretical and Methodological Issues. *MIS Quarterly*. 1/15; S. 5–10, 1991.
- [Gr15] Gröber, P.: What are the Differences? A Comparison of the Regular and Mobile Websites of Online Social Networks. In (Kouroupetroglou, G. Hrsg): *Proceedings of ICEAPVI*; S. 87–94, 2015.
- [He10] Heidemann, J.: Online Social Networks – Ein sozialer und technischer Überblick. *Informatik-Spektrum*. 3/33; S. 262–271, 2010.
- [Lu02] Luo, X.: Uses and Gratifications Theory and E-Consumer Behaviors. *Journal of Interactive Advertising*. 2/2; S. 34–41, 2002.
- [Ma12] Martiniello, N. et al.: Accessibility of Social Media for Students Who Are Blind or Have Low Vision. *Canadian Blind Monitor*. 28/32, 2012.
- [MK01] Moon, J.-W.; Kim, Y.-G.: Extending the TAM for a World-Wide-Web context. *Information & Management*. 4/38; S. 217–230, 2001.
- [MYZ02] McKinney, V. R.; Yoon, K.; Zahedi, F.: The Measurement of Web-Customer Satisfaction: An Expectation and Disconfirmation Approach. *Information Systems Research*. 3/13; S. 296–315, 2002.
- [RJ15] Ramayah, B.; Jaafar, A.: Understanding the impact of web layout and emotional changes towards navigation behaviour among visually impaired users. *Journal of Theoretical and Applied Information Technology*. 3/76; S. 289–295, 2015.

- [SZ08] Song, J. H.; Zinkhan, G. M.: Determinants of Perceived Web Site Interactivity. *Journal of Marketing*. 2/72; S. 99–113, 2008.
- [Ve00] Venkatesh, V.: Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. *Information Systems Research*. 4/11; S. 342, 2000.
- [Vo16] Voykinska, V. et al.: How Blind People Interact with Visual Content on Social Networking Services. In (Gergle, D. et al. Hrsg): *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, New York, NY, USA; S. 1584–1595, 2016.
- [WD85] Warshaw, P. R.; Davis, F. D.: Disentangling behavioral intention and behavioral expectation. *Journal of Experimental Social Psychology*. 3/21; S. 213–228, 1985.
- [WT05] Wixom, B. H.; Todd, P. A.: A Theoretical Integration of User Satisfaction and Technology Acceptance. *Information Systems Research*. 1/16; S. 85–102, 2005.

Re-imagining Deafness as a Resource for Design

Svetlana Matiouk^{1,2,3}

Abstract: This contribution sets out the necessity to re-imagine the notion of deafness in order to open up the potentials of this multifaceted cultural and linguistic phenomenon for technology design. The paper challenges common normative assumptions about deafness that are derived from some audiometric hearing “norm” and therefore bound our imagination by drawing attention to inabilities, difficulties and deficiencies. Based on a literature review of cognitive studies and long-term fieldwork, a variety of benefits arising from deafness and use of sign language are discussed. As a result of this discussion I propose considering signing deaf individuals as experts in visual, spatial and embodied matters. I argue, therefore, that this perspective is especially beneficial for designing innovative technologies and thinking out of the box. To exemplify this statement a couple of concepts for inspiration that emerged from five design and development projects are depicted. I call these concepts *inspiration rooms*.

Keywords: Deafness, Reconfiguration, Innovation, Design, Co-Design, Inspiration Rooms

1 Introduction

It is widely accepted that co-designing with people who are potential users of co-created products not only helps to gain more knowledge about the context of use, but also can foster creative processes and thinking, cf. (Muller 2002). However, what exactly do hearing designers gain when co-designing with signing deaf participants? What kind of “creative gains”⁴ may arise from this collaboration? In order to be able to properly answer these questions we firstly need to question the common notion of deafness. This paper aims at shedding light on this matter. It therefore proposes opening up deafness and the day-to-day experience with visual-gestural languages – i.e. sign languages – as a resource for design by considering deaf signers as experts for visual, spatial and embodied matters. Following this reconfiguration of deafness and as a result of five co-design projects the paper exemplarily introduces concepts – called *inspiration rooms* – that are based on the unique linguistic and cultural experience of deaf signers. These *inspiration rooms* are not only serving the argument in this paper, they moreover can be actually employed in design practice to stimulate creative thinking.

¹ RWTH Aachen, Informatik 5, Ahornstr. 55, 52056 Aachen

² Competence Centre for Sign Language and Gesture (SignGes), Theaterplatz 14, 52062 Aachen, s.matiouk@signges.rwth-aachen.de

³ Fraunhofer FIT, User-Centered Ubiquitous Computing, Schloss Birlinghoven, 53754 Sankt Augustin, svetlana.matiouk@fit.fraunhofer.de

⁴ By using this term I refer to the book “Deaf Gain” (Bauman and Murray 2014) where it has already been employed in the context of deaf studies.

2 Background

The traditional notion of deafness as a disability to perceive auditory phonetic language, cf. e.g. (Elzouki et al. 2012), shapes a perspective that draws attention to dysfunctions, deficiencies, and difficulties. This perspective and an accompanying set of practices are socially constructed and widely institutionalised. They originate from the “primordial fear of difference” (Fiedler 1996, p. 44), from the deepest anxiety to deviate from an assumed “norm” and to “lack” certain desired abilities, cf. (Rothschild 2006).

As a consequence, the prevailing social attitude generally is to make the apparently uncomfortable otherness disappear. “Using whatever scientific, technological, social, or legal measures available, the aim is to make disability just go away.” (Rothschild 2006, p. 229) One prominent example of this aspiration delivers the “treatment” of deafness by cochlear implants.

In the field of human-computer interactions (HCI), therefore, the majority of approaches and applications that address the needs of signing deaf users are focused on their inability to hear. These applications usually aim to support deaf signers w.r.t. difficulties some of them experience when reading and writing. The reason for such difficulties lies in the fact that scripture is a notation of acoustic features of spoken languages that are not directly accessible to prelingually deaf people, i.e. people who became deaf before acquiring a spoken language. The “natural” form of communication of prelingually deaf people resides instead in the visual-gestural domain, i.e. through a sign language.

It is, therefore, only logical that the majority of the scientific contributions in HCI are concerned with the accessible representation of textual content. Two main lines of research can be identified here. Firstly, there are contributions aimed at questions related to the use of pre-recorded sign language videos (Debevc, Kosec, and Holzinger 2011; Kaibel et al. 2006), or signing avatars (Cox et al. 2002; Kennaway, Glauert, and Zwitterlood 2007; Schnepf et al. 2012). Secondly, there are attempts to use text simplification (Vettori and Mich 2011). The last approach is, however, highly controversial in the deaf community. Moreover, there are case studies of various experimental applications available such as a visual display and a haptic chair for the enhancement of musical experience (Nanayakkara et al. 2009), or mobile systems for the visualisation of ambient sounds to deaf users (Matthews, Fong, and Mankoff 2005).

In addition to the scientific case studies there are accessibility guidelines available from the World Federation of the Deaf (WFD 2014) and the Web Accessibility Initiative (Caldwell et al. 2008) of the World Wide Web Consortium. Although these guidelines address important accessibility qualities for interactive products and may be helpful for the formal assessment of such products e.g. by governments, they are less suitable for being used in design practice, especially when it comes to thinking out of the box and creating innovation.

In conclusion, all these approaches to the accessibility for deaf users in HCI take the “classic” perspective on deafness as stated above. Based on a series of qualitative investigations set out in the next section as well as on the insights from the literature review on cognition in deaf signers, I argue for the need to re-imagine the notion of deafness. Such reconfiguration of deafness towards opening up its gains instead of solely highlighting difficulties is not only a necessity for being able to adequately address signing deaf experience, in fact it creates a space of opportunities and, thus, may provide the impetus for innovation.

3 Methodological Approach in this Contribution

This contribution is based on a literature review of cognitive studies and a more than three year long field research conducted in the context of five Web co-design and development projects in Germany: Vibelle (completed in 2012), Gateway (completed in 2013), DeafTrain (completed in 2015), DeafExist (in progress) and DeafMentoring (in progress). The websites resulting from the four of five projects can be accessed online via the following URLs: <http://vibelle.de/>, <http://gateway-online.de/>, <http://deaftrain.de/> and <http://deafexist.de/>. The development in the fifth project DeafMentoring is in its initial stage with the first series of explorative co-design sessions having just passed. Although the DeafMentoring website <http://deafmentoring.de/> already presents its purpose relevant information, it is not yet a result of the appropriate co-design activities. All five projects share the same general objective that is to facilitate the entry of signing deaf people into working life. During the projects, deaf and hearing co-designers closely work together on the creation of websites that support project operations and serve the overall objective.

United by the common general objective, the five projects address different topics, and each of them serves its own specific purpose. The project Vibelle was concerned with the development of a multimedia information and eLearning platform aiming at the increasing vocational qualification of deaf adolescents. An information portal for deaf and blind students was developed in the project Gateway. The DeafTrain portal was designed to facilitate the entry of young signing deaf people into working life, in particular by educating and sensitising their teachers. For this purpose blended learning seminars were developed, i.e. the seminars fuse classroom and eLearning activity. The DeafExist project was founded in support of signing deaf entrepreneurs in their business start-ups. Finally, the DeafMentoring project is aimed at supporting signing deaf teenagers in the planning of their professional development by connecting them with experienced and successful deaf signers. Thereby a peer-counselling practice and appropriate supportive tools are elaborated in the project.

During the field research a combination of qualitative approaches to data elicitation has been employed. Participant observations, group discussions and interviews have been applied in a variety of situations at work and outside. The situations at work include:

- Seven co-design workshops, cf. (Sanders and Stappers 2008), for about four hours each with the average four signing deaf and three hearing participants;
- Two design studios, cf. (Ungar and White 2008), for about one and a half hours and with three participants each (five of whom were deaf signers);
- 13 qualitative interviews with four signing deaf experts in Deaf Education, one young deaf intern, two signing deaf external volunteers, as well as three hearing experts in Deaf Education, one hearing graphic designer, and two hearing developers who were involved in the co-design projects;
- Five formative usability evaluations with gestural think aloud protocols (GTAP), cf. (Roberts and Fels 2006);
- More than ten group discussions with on average six signing deaf participants (22 different deaf signers in total).

Altogether 25 signing deaf people have participated at different stages of the study so far. Most of them, except for two external volunteers, were involved in the project work and had specific roles, e.g. conception and organisation of further training for deaf people, and production of content also in the form of sign language videos.

In order to facilitate the collective creativity between signing deaf and hearing co-designers the main language of communication was the German Sign Language (DGS⁵). However, because not all hearing team members have strong DGS skills, in almost all sessions sign language interpreters were employed. To achieve mutual visibility attendees were sitting according to a circular or semi-circular layout. As exemplified in Figure 1 when brainstorming or discussing certain design ideas in workshops, group discussions and at times in interviews, the ideas were simultaneously sketched.

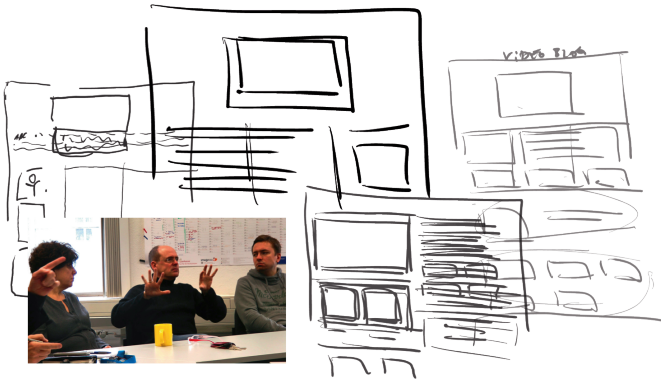


Figure 1. A snapshot from a co-design workshop and examples of sketches. The participant in the centre explains his idea. The explanations are simultaneously sketched and made visible for everyone.

⁵ DGS stands for Deutsche Gebärdensprache.

This field research has mainly been conducted within the projects DeafTrain, DeafExist and DeafMentoring. The co-design results and experience of the older projects Vibelle and Gateway have been accessed through the qualitative interviews with the members of these projects as well as group discussions about the resulting websites.

4 Re-imagining Deafness

A growing amount of scientific work on human cognition, e.g. in such fields as psycholinguistics or neuroscience, provides evidence of enhanced cognitive, psychological and social capabilities. To give a couple of examples of particular enhancements of visual cognition in deaf individuals there are studies supporting a higher level of peripheral attention (Bavelier, Dye, and Hauser 2006; Dye, Hauser, and Bavelier 2009), or of facial processing (McCullough and Emmorey 1997), or a better skill in mental transformation like mental rotation of spatial images (Emmorey, Klima, and Hickok 1998) compared to hearing people. These cognitive gains of deaf people originate mainly from three circumstances. Firstly, it is the ability of visual and auditory cortical regions to reorganise across modalities, which is called cross-modal plasticity (Finney et al. 2003). Therefore, auditory cortical regions of prelingually deaf people begin to process varieties of visual input extending the visual processing capacities already provided by the visual cortical regions, cf. (Campbell and MacSweeney 2004; Bavelier, Dye, and Hauser 2006). Secondly, the development of a visuospatial rehearsal loop in the short-term memory in deaf signers (Wilson and Emmorey 1997) leads to effects on visual perception as well (Wilson and Emmorey 2003; Dye, Hauser, and Bavelier 2009). Thirdly, it is the everyday use of sign language (Emmorey 2002).

In addition to the extensive research on visual cognition there were certain psychological and social benefits of deaf signers documented, such as the benefits related to interpersonal relationships originating from the maintaining of eye contact, or embodiment of reciprocity when together with others navigating through space, cf. (Bauman and Murray 2014).

Moreover, the everyday use of sign language contributes to the development of fine motor skills and facial expressiveness (Emmorey 2002). Other studies report an enhancement of tactile sensitivity (Levänen and Hamdorf 2001).

Taking together the scientific insights summarised above we can now state that deafness is not solely an experience of not hearing, but rather a “distinct way of being in the world, one that opens up perceptions, perspectives, and insights that are less common to the majority of hearing persons” (Bauman and Murray 2014, p. xv). Moreover, we can acknowledge certain superior abilities in signing deaf people. Therefore, I suggest considering deaf signers as experts in visual, spatial, and embodied matters. In the next section I argue that this perspective may be particularly beneficial in the area of technology design.

5 Deafness as a Resource for Design

The potential impact of considering deaf people as experts in visual, spatial and gestural matters resides at different levels. Firstly, signing deaf people themselves may benefit from an overall improvement of the user experience that is coevolving with a general growth of understanding of deaf experience by hearing technology designers. Deaf signers may furthermore benefit from being empowered to contribute and consequently being valued as experts. Secondly, hearing design practitioners may gain inspiration and insights from other ways of being. They may gain new perspectives and ways to interpret their designs and learn from the experts in visual-gestural language, e.g. to more effectively communicate with each other, as it has been shown in the case of professional communication among students in architecture by Klaudia Grote et al. (Grote et al. 2016). Moreover, co-designing with signing deaf participants may foster divergent thinking (Raïke, Pylvänen, and Rainò 2014), which is partially grounded in the use of sign language and its effects on the formation of stronger connections between syntagmatically related concepts, i.e. related through associations, cf. (Grote 2013). The insights above are additionally supported by the analysis of the qualitative data gathered within our field study. Extrapolating these insights points to the third potential long-term impact, namely on society in general. In fact, societal gains may arise for instance from the innovation boost resulting from collaboration between deaf and hearing co-designers, and as a potential consequence, a better user experience not only for deaf signers but for some non-deaf people as well.

My approach in supporting these “creative gains” is twofold. Firstly, in order to actually enable fully-fledged active collaboration between deaf and hearing co-designers such co-design approaches can be adopted as co-design workshops, design studios, and group discussions. Secondly, to support hearing designers without available resources for co-design activities, a number of “rooms” for inspiration are being developed. These *inspiration rooms* are aimed at explaining the deaf perspectives and at igniting the imagination of hearing designers. In the following, three examples of *inspiration rooms* are depicted to illustrate the idea.

Most Important in the Centre

This inspiration room is dedicated to the social and linguistic practice of signing deaf people. This practice puts emphasis on the visual field when perceiving the world, or communicating with other people, and on the role of the central perspective.

During one of the sessions a signing deaf participant mentioned: “*I’m looking in the centre. For me, there is the most important information in the centre. And this is what is viewed first. And the rim is then rather less interesting. ... There, where I hold my view, I get the important information.*”



Figure 2. The screenshot shows an interactive globe navigation in a state after multiple click actions: first on the “eLearning” item, then on “Mathematik”, followed by “Kurs I Grundrechenarten” (i.e course I - basic arithmetics). In this state the globe menu provides several affordances: going to more general categories, opening current topic described in sign language in the central globe and navigating to the two more specific topics represented by the globes “Addition und Substraktion” and “Multiplikation und Division”.

Design thinking in this direction may lead to innovative, and unconventional design solutions, such as the navigation principle developed for the Vibelle website. Figure 2 shows the so-called “globe menu” that presents a short sign language description of the contents below in the central globe. The different contents are aligned in space around the central globe.

The Semantic Dimension of Space

When using sign languages, meanings and semantic relationships are created in space. Thereby not only the locations of signs, but also the dynamics of their production play an important role. Therefore, the day-to-day use of sign language trains the spatial perception, including the visuospatial short-term memory, making deaf signers be more attentive to the visual signals produced – e.g. by interactive systems – and to their possible meanings in space and time.

To give an example, in particular for the attentiveness to the spatial component of meaning, in one of the workshops the following situation occurred. When co-experimenting with different layouts for the DeafTrain website, signing deaf participants explicitly refused to accept any layout with vertical menu alignments in order to depict more or less equally important categories of a website. Figure 3 shows the horizontal menu implemented on the website.



Figure 3. Main menu and its subpages fly-out menu are both horizontally aligned in the header of the page. The screenshot shows the state of the menu on the page “Seminare” when mouse is over this main menu item: All eight subtopics unfold horizontally under the selected menu item.

The reluctance of signing deaf participants to put equally important items, into a vertical structure can be explained from the linguistic perspective by the fact, that sign language users linguistically relate true or imagined objects or people to each other, i.e. it uses space to invoke patterns of meaning. In this regard, arranging items vertically means that items have an order, or priority. In contrast, organizing items horizontally signifies a similar level of importance. This simple co-design experiment suggests that deaf users, due to the modality of their language, experience a stronger conventionalization when appropriating spatially arranged design elements.

Futures to See and Feel

Within a heated discussion about the differences between the linguistic and cultural experience of deaf and hearing people in relation to semantic categorization during one of the workshops, one of the interpreters stopped the impassioned signing participants, who started to sign very fast, in order to interpret: *“We don’t know what is coming. We cannot hear that,”* and then added the following: *“I think ... this is incredibly important – and it is also not clear to hearing people – that sounds, and they are not solely an alarm function, are the future: it is already happening! ... If two people are arguing, and the door slams shut with a loud bang, then I ((points to herself)) know why (because I just have heard it). This is what completely evades deaf people: The future is always a mystery!”*

The deaf participants expressed their appreciation for these words with signs SUPER and APPLAUSE. This particular debate culminated in a conclusion that this situation, with respect to the use of information and communication technologies, in particular webpages for deaf people, is quite the same. Therefore, e.g. any information that is concealed in some hierarchical structures appears to be not inline with the syntagmatic, i.e. associative, thought structures shaped by sign language use.

This concept may be seen in relation to what Paul Dourish and Graham Button (Dourish and Button 1998) called “accountability” of an interface, i.e. “first, the very way in which it is organised provides to others the means to recognize it as what it is (accountable), and, second, does so within the very fact of its production (reflexivity), rather than within some wider frame of ‘social meaning’.” (Dourish and Button 1998, p. 398)

Taking together the concepts of accountability and of futures to see and feel, the characteristic of an interactive system to represent its own behavior can now be thought in terms of (visually or tactile) describing what will or may happen in the context of its use in the near future. Following this line of thought may inspire designers to create novel meaningful solutions.

6 Conclusion and Future Work

Instead of focusing on the inability to hear this paper proposed re-imagining the common notion of deafness, and by doing so facilitating “creative gains” at different levels. This reconfiguration of deafness is grounded in a variety of studies from the cognitive sciences, psychology and sociology, as well as qualitative investigations outlined in this paper. These studies provide a foundation for the recognition of the specific superior abilities in signing deaf people compared to hearing w.r.t. cognitive, social and psychological gains. Therefore, for purposes of technology design, I propose considering signing deaf individuals as experts in visual, spatial and embodied matters.

In this paper I argue that this reconfiguration of deafness may create a space of opportunities for technology design and, as a consequence, boost innovation. Moreover, in order to facilitate collaboration between hearing and deaf co-designers there is a need for a higher awareness especially among hearing participants of the linguistic and cultural perspectives related to deafness. To facilitate this understanding and stimulate imagination, I am elaborating a set of empirically grounded concepts for inspiration collected in, as I call them, *inspiration rooms*. Three of these concepts are depicted in this paper to exemplify their potential for the design thinking and creating innovation. The elaboration of *inspiration rooms* has not yet been completed, therefore, it is one of the further research activities.

A more in depth discussion on best practices of how to facilitate a fully-fledged participation of deaf signers in predominantly hearing development teams, is, however, out of the scope of this paper, but is planned as a separate contribution in the future. Another interesting direction for future work is about the limits of the societal impact of co-designing with signing deaf participants. It is for instance quite straightforward that this approach cannot promise well-accessible solutions for deaf-blind or blind users. However, it is conceivable that there are people with some other sensory disabilities who may benefit from its results.

Acknowledgements

I am very grateful to all participants who provided a boundless source of inspiration and insight: Without you this research would not be possible! I would like to cordially thank Klaudia Grote and Florian Kramer for inspiring cooperation and support. This work receives a great support from Prof. Matthias Jarke whom I especially would like to thank. I am also very grateful to Alexander Boden and Sebastian Denef for their inspiring feedback on my work, and to my colleagues at Fraunhofer FIT for fruitful collaboration. I appreciate the contribution made by Ross Redfern in proofreading my work. The projects were funded by the Federal Ministry of Labour and Social Affairs in Germany (Bundesministerium für Arbeit und Soziales). This particular contribution was funded by the Informatik 5 department at the RWTH Aachen University.

References

- Bauman, H-Dirksen L, and Joseph J Murray. 2014. *Deaf Gain: Raising the Stakes for Human Diversity*. University of Minnesota Press.
- Bavelier, Daphne, Matthew W G Dye, and Peter C Hauser. 2006. "Do Deaf Individuals See Better?" *Trends in Cognitive Sciences* 10 (11): 512–518.
- Caldwell, Ben, Michael Cooper, Loretta Guarino Reid, Gregg Vanderheiden, Wendy Chisholm, John Slatin, and Jason White. 2008. "Web Content Accessibility Guidelines (WCAG) 2.0." <http://www.w3.org/TR/WCAG20/>.
- Campbell, Ruth, and Mairéad MacSweeney. 2004. "Neuroimaging Studies of Cross-Modal Plasticity and Language Processing in Deaf People." *The Handbook of Multisensory Processes: 773–784*.
- Cox, Stephen, Michael Lincoln, Judy Tryggvason, Melanie Nakisa, Mark Wells, Marcus Tutt, and Sanja Abbott. 2002. "Tessa, a System to Aid Communication with Deaf People." In *Proceedings of the Fifth International ACM Conference on Assistive Technologies*, 205–212. Assets '02. New York, NY, USA: ACM.
- Debevc, Matjaž, Primož Kosec, and Andreas Holzinger. 2011. "Improving Multimodal Web Accessibility for Deaf People: Sign Language Interpreter Module." *Multimedia Tools Appl.* 54 (1) (August): 181–199.
- Dourish, Paul, and Graham Button. 1998. "On 'Technomethodology': Foundational Relationships between Ethnomethodology and System Design." *Human-Computer Interaction* 13 (4): 395–432.
- Dye, Matthew W G, Peter C Hauser, and Daphne Bavelier. 2009. "Is Visual Selective Attention in Deaf Individuals Enhanced or Deficient? The Case of the Useful Field of View." *PLoS One* 4 (5): e5640.
- Elzouki, Abdelaziz Y, Harb A Harfi, F B Stapleton, William Oh, H Nazer, and Richard J Whitley. 2012. *Textbook of Clinical Pediatrics*. Springer Science & Business Media.
- Emmorey, Karen. 2002. *Language, Cognition, and the Brain: Insights from Sign Language Research*. Lawrence Erlbaum Associates, Inc.
- Emmorey, Karen, Edward Klima, and Gregory Hickok. 1998. "Mental Rotation within Linguistic and Non-Linguistic Domains in Users of American Sign Language." *Cognition* 68 (3): 221–246.
- Fiedler, Leslie A. 1996. "Pity and Fear: Images of the Disabled In Literature and the Popular Arts." In *Tyranny of the Normal: Essays on Bioethics, Theology & Myth*, 33–47. Boston: David R. Godine Publisher.
- Finney, Eva M, Brett A Clementz, Gregory Hickok, and Karen R Dobkins. 2003. "Visual Stimuli Activate Auditory Cortex in Deaf Subjects: Evidence from MEG." *Neuroreport* 14 (11): 1425–1427.
- Grote, Klaudia. 2013. "'Modality Relativity': The Influence of Sign Language and Spoken

Language on Conceptual Categorization.” Hochschulbibliothek der Rheinisch-Westfälischen Technischen Hochschule Aachen.

- Grote, Klaudia, Horst Sieprath, Hannah Groninger, Irene Mittelberg, Thomas H Schmitz, and Klaus Willmes. 2016. “Space, Body & DeSign – Ein DeafGain-Projekt zur Verbesserung der visuellen Wahrnehmung und der räumlichen Beschreibungskompetenzen taubloser Menschen.” *DAS ZEICHEN* (102): 104–109.
- Kaibel, Andreas, Klaudia Grote, Kilian Knoerzer, Horst Sieprath, and Florian Kramer. 2006. “Hypertext in Sign Language.” In 9th ERCIM Workshop “User Interfaces For All”. Königswinter, Germany.
- Kennaway, J R, J R W Glauert, and I Zwitterlood. 2007. “Providing Signed Content on the Internet by Synthesized Animation.” *ACM Trans. Comput.-Hum. Interact.* 14 (3) (September).
- Levänen, Sari, and Dorothea Hamdorf. 2001. “Feeling Vibrations: Enhanced Tactile Sensitivity in Congenitally Deaf Humans.” *Neuroscience Letters* 301 (1): 75–77.
- Matthews, Tara, Janette Fong, and Jennifer Mankoff. 2005. “Visualizing Non-Speech Sounds for the Deaf.” In *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility*, 52–59. ACM Press.
- McCullough, Stephen, and Karen Emmorey. 1997. “Face Processing by Deaf ASL Signers: Evidence for Expertise in Distinguishing Local Features.” *Journal of Deaf Studies and Deaf Education*: 212–222.
- Muller, Michael J. 2002. “Participatory Design: The Third Space in HCI.” In *The Human-Computer Interaction Handbook*, edited by Julie A. Jacko and Andrew Sears, 1051–1068. L. Erlbaum Associates Inc., Hillsdale, NJ, USA.
- Nanayakkara, Suranga, Elizabeth Taylor, Lonce Wyse, and S H Ong. 2009. “An Enhanced Musical Experience for the Deaf: Design and Evaluation of a Music Display and a Haptic Chair.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 337–346.
- Raike, Antti, Suvi Pylvänen, and Päivi Rainò. 2014. “Co-Design from Divergent Thinking.” In *Deaf Gain: Raising the Stakes for Human Diversity*, edited by H-Dirksen L Bauman and Joseph J Murray, 402–420. Minneapolis; London: University of Minnesota Press.
- Roberts, Vera Louise, and Deborah I Fels. 2006. “Methods for Inclusion: Employing Think Aloud Protocols in Software Usability Studies with Individuals Who Are Deaf.” *Int. J. Hum.-Comput. Stud.* 64 (6): 489–501.
- Rothschild, Joan. 2006. “Prenatal Diagnosis and Disability: The Ultimate ‘Technological Fix’?” In , edited by Arno Bammé, Günter Getzinger, and Bernhard Wieser, 54:217–244. *Yearbook of the Institute for Advanced Studies on Science, Technology and Society*. Profil Verlag GmbH München Wien.
- Sanders, Elizabeth B-N, and Pieter Jan Stappers. 2008. “Co-Creation and the New Landscapes of Design.” *Co-Design* 4 (1): 5–18.

- Schnepf, Jerry C, Rosalee J Wolfe, John C McDonald, and Jorge A Toro. 2012. "Combining Emotion and Facial Nonmanual Signals in Synthesized American Sign Language." In Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility, 249–250.
- Ungar, Jim, and Jeff White. 2008. "Agile User Centered Design: Enter the Design Studio-a Case Study." In CHI'08 Extended Abstracts on Human Factors in Computing Systems, 2167–2178.
- Vettori, Chiara, and Ornella Mich. 2011. "Supporting Deaf Children's Reading Skills: The Many Challenges of Text Simplification." In The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility, 283–284.
- WFD. 2014. "Working Document on Adoption and Adaptation of Technologies and Accessibility."
- Wilson, Margaret, and Karen Emmorey. 1997. "A Visuospatial 'Phonological Loop' in Working Memory: Evidence from American Sign Language." *Memory & Cognition* 25 (3): 313–320.
- Wilson, Margaret, and Karen Emmorey. 2003. "The Effect of Irrelevant Visual Input on Working Memory for Sign Language." *Journal of Deaf Studies and Deaf Education* 8 (2): 97–103.

Loaded Dice: Ein Co-Design Werkzeug zum Entwerfen mit blinden und sehbehinderten Partizipanten im IoT

Arne Berger¹, Andreas Bischof¹, Kevin Lefeuvre¹, Albrecht Kurze¹, Michael Storz¹, Sören Totzauer¹

Abstract: Zwei interaktive Würfel werden vorgestellt, die als *Werkzeug zum Entwerfen* dazu dienen gemeinsam mit Co-Designern den Designspace von *Smart Connected Products* zu erforschen. Diese Würfel sind als *Forschungswerkzeuge* dazu gemacht, Co-Design Aktivitäten zu unterstützen. Dazu wird sehr kurz in Research Through Design als drittes Paradigma der HCI eingeführt, die Würfel als interaktive Werkzeuge vorgestellt und eine Case Study ihrer Verwendung in einem Co-Design Workshop mit blinden und sehbehinderten Co-Designern vorgestellt.

Keywords: Research Through Design, Experience Design, Co-Design Games, Internet of Things

1 Einleitung

Wir stellen im folgenden zwei interaktive Würfel vor, die als *Werkzeug zum Entwerfen* dazu dienen, gemeinsam mit Co-Designern den Designspace von *Smart Connected Products* zu erforschen. Diese Würfel sind als *Forschungswerkzeuge* dazu gemacht verschiedene Co-Design Aktivitäten wie Workshops oder Feldstudien zu unterstützen.

Die Notwendigkeit für solche Werkzeuge wird dabei zunächst mit Hilfe des Forschungsparadigmas *Research Through Design* begründet. Daran anschließend wird der Designspace *Smart Connected Products* kurz umrissen und die Rolle dieser Würfel als Werkzeug zum Erforschen dieses Designspace beschrieben. Abschließend werden Verwendung und Möglichkeiten dieser Werkzeuge beispielhaft an einem mit blinden und sehbehinderten Nutzern durchgeführten Co-Design Workshop beschrieben.

2 Third Wave HCI: Research Through Design als Forschungsparadigma in der Human-Computer-Interaction

“Nutzerzentrierte” Entwicklung ist Ziel und Bedingung vieler Technikentwicklungsprojekte. Die Analyse von Nutzerbedürfnissen durch Usability-Forschung (“human factors”) oder durch Participatory Design nach skandinavischen Vorbildern [Bø93] sind dafür etablierte Verfahren. Die weiterhin zunehmende Verbreitung digitaler und vor allem vernetzter Technik stellt eine Herausforderung dar, der

¹ Technische Universität Chemnitz, Fakultät Informatik, Professur Medieninformatik, 09107 Chemnitz
vorname.nachname@informatik.tu-chemnitz.de

Gestaltungsprozesse noch integrativer begegnen müssen [Ha08]. Die Komplexität technischer Lösungsangebote für soziale Problemstellungen erfordert einen dritten Schritt der Erfassung von menschlichen Aktivitäten über Ergonomie und Akzeptanz hinaus, ein phänomenologisches Verständnis von Alltagswelten und Entwicklungsprozessen, wie es von Ansätzen des so genannten Third Wave Paradigms der Mensch-Technik-Interaktionsforschung vertreten wird (ibid.). Insbesondere Research Through Design [Ga12] bildet hier ein adäquates methodologisches Dach, unter dem eine dem dritten Paradigma der HCI angemessene fallbasierte und phänomenologische Forschung möglich ist. Der Forschungsprozess orientiert sich dabei an innovativen Artefakten, die forschungspraktisch aus Verwendungskontexten erwachsen, auf neue verweisen und empirisch auf ihre Bedeutung und Angemessenheit hin untersucht werden.

Forschungspraktisch erfordert dies iterativ und reflexiv verlaufender Entwicklungsprozesse. Diese müssen dabei immer wieder mit neuen Daten aus dem Feld kontrastiert werden. Die methodischen Paradigmen Research Through Design [Ga12] und Grounded Theory [Su07] beschreiben solche Prozesse als netzwerkförmig und prozessual, was inhaltlich über entwurfspraktisch notwendige Prozesse wie Scrum genauso hinausgeht wie über ein an Konsumptionsprozesse gekoppeltes "Design Thinking".

3 Partizipation: Werkzeuge zum Entwerfen entwerfen

3.1 Smart Connected Devices

IoT Produkte wie sie unter den Schlagworten Smart Home, Quantified Self oder Ambient Assisted Living entwickelt, erforscht und vermarktet werden, sind von ihren Entwicklerinnen festgeschriebene Kombinationen von Sensoren und Aktuatoren in einem oder mehreren Regelkreisen. Durch die Messung verschiedener physikalischer Größen in der Umwelt durch Sensoren und Rückwirkungen in die Welt physikalischer Größen durch Aktuatoren ergeben sich IoT-Produkte mit innovativen Verwendungsmöglichkeiten. Es ergeben sich zudem eine Reihe an multimodalen Interaktionsmöglichkeiten. Diese können weit über das Desktop-Paradigma mit Tastatur/Maus und Displays oder kalte, uniforme Oberflächen im mobile HCI Paradigma hinausgehen. Dies bedeutet z.B. die Nutzung haptischer Elemente (Schalter, Regler) und weiterer Feedbackkanäle (z.B. Vibration) um weitere bzw. alternative (nicht-visuelle) Sinne ansprechen können.

3.2 Co-Design Support Tools

Um im Bereich von Smart Connected Products überhaupt gemeinsam mit potentiellen Verwendern entwickeln zu können, bedarf es nicht nur methodischer, sondern auch technischer Hilfsmittel. Prototyping Baukästen für Connected Things wie Blinky Blocks [Ki11], Siftables [Me07] oder littleBits [Bd09] die in den meisten Fällen explizit zusätzliches Workshopmaterial erfordern, aber selbst keine Workshopformate vorschlagen

stehen dabei nicht im Fokus unserer Arbeit. Die vorliegende Arbeit orientiert sich stattdessen an ähnlichen Ansätzen zu Co-Design Support Tools, die mit und für speziell zugeschnittene Workshopformate für kognitiv benachteiligte Anwenderinnen entwickelt wurden, wie eTextile Weaving [Gi15] und andere [ME04, Tw15]. Mit entsprechenden Co-Design Support Tools, welche die Möglichkeiten und Grenzen von IoT Technologie inkorporieren, können Verwender im Rahmen dazu passend entworfener Workshops Funktionalitäten spielerisch explorieren und kombinieren. Der Einsatz solcher Werkzeuge zum Entwerfen dient in diesem Setting nicht dem Messen von Akzeptanzwerten fertiger Produkte. Die Entwurfswerkzeuge sind dabei methodische Katalysatoren, die mögliche Funktionalitäten zukünftiger Produkte verkörpern. In den dazu passenden Workshops setzen wir diese Objekte ein, um sowohl eine kontinuierliche transdisziplinäre Anwenderintegration als auch eine übergreifende interdisziplinäre Forschungskultur zu ermöglichen [Be15].

Werkzeuge zum Entwerfen wie die in der hier beschriebenen Case Study ermöglichen Workshops in denen Verwender Funktionen, Nützlichkeiten und Nutzungszusammenhänge aus eigenen konkreten sozialen Kontexten heraus entwickeln können. Die Teilnehmer werden nicht mit fertigen Szenarien konfrontiert, sondern selbst zu Entwerfern eigener Anwendungen. Sie werden dabei nicht mit Baukästen konfrontiert deren Verwendung zunächst der Verblendung von Elektronik mit selbstgebauten Umhüllungen bedürfen.

So lassen sich eine umfassende Beteiligung der Verwender im Entwerfen und Entwickeln von Smart Connected Products im Internet of Things methodisch integriert bearbeiten. Durch die übergreifende Methodologie des Research Through Design werden diese Entwürfe und ihre Iterationen als integrierter Entwicklungsprozess verstanden. Die Erkenntnisse aus diesen Prozessen dienen schließlich als Wissensbasis, Inspiration und methodisch fundierte Basis für den Entwurf nützlicher, gewünschter, inklusiver Produkte.

4 Case Study: interaktive Würfel

Um gemeinsam mit Co-Designern mögliche Probleme und Lösungen ihrer Erlebenswirklichkeit offen zu Explorieren und gemeinsam mit ihnen sinnvolle Anwendungen für fernkommunizierende interaktive Produkte, die mit Sensoren und/oder Aktuatoren ausgestattet sind, zu entwickeln wurde ein vernetztes Würfelpaar "Loaded Dice" (Abbildung 1) entworfen. Dieses Würfelpaar besteht aus einem Sensorwürfel und einem Aktuatorwürfel mit jeweils einem Sensor bzw. Aktuator auf jeder der 6 Seiten. Bei der Verbindung von einem Sensor mit einem Aktuator werden auf eine explorative Art Anwendungsszenarien für Lösungen von alltäglichen Problemen der im Workshop involvierten Anwendern exploriert, erwogen und simuliert.

Im Sensorwürfel können dazu sechs Sensoren (Temperatur, Licht, Bewegung, Distanz, Schall, Drehregler) und im Aktuatorenwürfel sechs Aktuatoren (LED-Bargraph, Peltier-Thermofläche, Power-LED, Lautsprecher, Ventilator, Vibrationsmotor) genutzt werden

(Siehe Abbildung 2).



Abbildung 1: Interaktive Werkzeuge – Loaded Dice; links Aktuatorwürfel, rechts Sensorwürfel

Die Würfel werden verwendet, indem jeweils ein Sensor und ein Aktuator ausgewählt werden, um miteinander verbunden zu interagieren. Wählt eine Anwenderin beispielsweise den Bewegungssensor und die Peltier-Thermofläche als Aktuator, wird diese warm, wenn vor dem Bewegungssensor Bewegung registriert wird. Weil erst die Kommunikation zwischen “smarten” Objekten diese wirklich smart macht, können diese Würfel frei untereinander verbunden werden und Sensor-Aktuator-Verknüpfungen frei kombiniert und ausprobiert werden. Diese Verbindung erlaubt es, schnell und flexibel unterschiedliche Sensor-Aktuator-Kombinationen zu erproben, sowohl in Co-Design Aktivitäten von Forschern und Partizipanten, als auch in der Zusammenarbeit mehrerer Nutzer oder in der räumlichen Verteilung.

Die Besonderheiten unseres Ansatzes liegen dabei außerdem in der haptischen Semiotik. So war es ein Ziel, die Funktion der einzelnen Sensoren und Aktuatoren begreifbar zu machen (im wahrsten Sinne des Wortes) in dem jeweilige Funktionen “erfühlbar” werden und damit speziell den Anforderungen von Workshops mit blinden und sehbehinderten Co-Designerinnen gerecht zu werden. Da die Form auf die elementare Funktion hinweist und kein unmittelbar festgeschriebener Verwendungszusammenhang von einem Sensor und einem Aktuator vorgegeben wird, können technische Aspekte “vergessen” werden. Co-Designer werden so bei der freien Assoziation und einem auf technischen Möglichkeiten (anstelle von Notwendigkeiten) basierenden Denkprozess unterstützt. Durch die haptische Semiotik lassen sich zudem sprachliche Erklärungen eines Workshopleiters auf ein Minimum reduzieren.

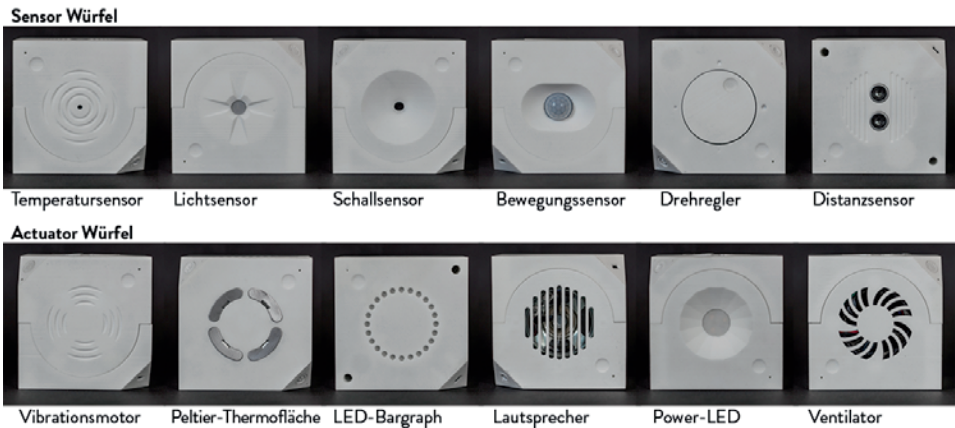


Abbildung 2: Alle Seiten des Sensorwürfels (oben) und des Aktuatorwürfels (unten)

4.1 Workshops

Eine erste explorative Erprobung des interaktiven Würfelpaars fand mit zwei Schulklassen mit insgesamt elf blinden und sehbehinderten Schülern statt. Die Workshops wurden mit Hilfe von Feldnotizen dokumentiert und qualitativ ausgewertet. Diese zwei ersten Workshops haben gezeigt, dass dieses Werkzeug als Methode zum Co-Design mit diesen Nutzergruppen sehr tauglich ist. Die Teilnehmenden waren in der Lage, die Funktionen des Würfelpaars selbstständig zu erkennen und zu explorieren. Darüber hinaus hat die methodische Kombination aus Gruppendiskussion, freier Exploration und moderierter Ideenbildung gezeigt, dass sie selbstläufige Interpretationen und Bedeutungszuschreibungen der Verwender hervorbringt. Es wurde deutlich, dass die interaktiven Entwurfswerkzeuge assoziative Explorationsmodi ermöglichen, die am sensorischen Erleben der Teilnehmer ansetzen.

Insbesondere spielerische Explorationsformen führten zu einer größeren Bandbreite sensorischer Assoziationen. So hat die Exploration mit den Loaded Dice zu einer Reihe konkreter Co-Designs geführt. Beispielhaft ist hier abhängig vom Wetter modulierter Vogelgesang als Interaktionsmuster für einen sensorbasierten Wetterbericht über die unmittelbare Wohnumgebung blinder Nutzerinnen. Dies kann nun als Basis für inklusive, vor allem aber möglichst wenig stigmatisierende Smart Connected Products dienen. Unsere Ergebnisse zeigen, dass partizipative Gestaltungsmethoden mit eingeschränkt sehenden Teilnehmern von unserem Entwurfswerkzeug enorm profitieren. Die semiotisch angereicherte und vor allem modulare Repräsentation von IoT-Funktionen hat das Erleben, Verstehen und Gestalten solcher Funktionalitäten auch im Vergleich mit anderen Tools signifikant erhöht. Damit lassen sich Nutzergruppen wie Blinde und Sehbehinderte und ihre Lebenswelten besser in partizipative Entwicklungsprozesse integrieren.

Danksagung

Wir danken allen Co-Designerinnen für die gemeinsame Arbeit, sowie allen Lehrerinnen und dem Direktor der Landesschule für Blinde und Sehbehinderte für ihre Unterstützung. Die Forschungsarbeit wird vom Bundesministerium für Bildung und Forschung (BMBF) unter dem Förderkennzeichen FKZ 16SV7116 gefördert.

Literaturverzeichnis

- [Bd09] Bdeir, A. 2009. Electronics as material: littleBits. In Proc. International Conference on Tangible and Embedded Interaction. ACM, New York, NY, USA, 397-400.
- [Be15] Berger, A., Heidt, M. & Eibl, M. 2015. Conduplicated Symmetries: Renegotiating the Material Basis of Prototype Research. Proc. ICoRD 2015. Springer. 71-78
- [Bø93] Bødker, S., Granbaek, K., & Kyng, M. 1993. Cooperative Design: Techniques and Experiences from the Scandinavian Scene. Participatory Design: Principles and Practices.
- [Ga12] Gaver, W. 2012. What should we expect from research through design?. In Proc. SIGCHI conference on human factors in computing systems (pp. 937-946). ACM.
- [Gi15] Giles, E. and van der Linden, J. 2015. Imagining Future Technologies: eTextile Weaving Workshops with Blind and Visually Impaired People. Proc. of the 2015 ACM SIGCHI Conference on Creativity and Cognition.
- [Ha08] Harrison, S., Tatar, D., & Sengers, P. (2008). The three paradigms of HCI. International Journal of Human-Computer Interaction 1-21.
- [Ki11] Kirby, B., Ashley-Rollman, M., & Copen Goldstein, S. 2011. Blinky blocks: a physical ensemble programming platform. In Proc. CHI '11 Extended Abstracts on Human Factors in Computing Systems. ACM, New York, NY, USA, 1111-1116.
- [ME04] McElligott, J. and van Leeuwen, L. 2004. Designing Sound Tools and Toys for Blind and Visually Impaired Children. Proc IDC 2004. ACM Press. 65-72.
- [Me07] Merrill, D., Kalanithi, J., Maes, P. 2007. Siftables: towards sensor network user interfaces. In Proceedings of the 1st international conference on Tangible and embedded interaction. ACM, New York, NY, USA, 75-78.
- [Su07] Suchman, L. 2007: Human-Machine Reconfigurations. Plans and Situated Actions. Cambridge University Press, New York
- [Tw15] Twyman, M., Mullenbach, J., Shultz, C., Colgate, J.E. and Piper, A.M. 2015. Designing Wearable Haptic Information Displays for People with Vision Impairments. Proc. TEI 2015. ACM Press. 341-34.

Interferenzmessung auf Bildschirmen zur Revision überholter Grenzwerte für photosensitive Epilepsie

Mathias Haimerl¹

Abstract: Die derzeit gültigen Grenzwerte zur Auslösung photosensitiver Epilepsie basieren auf Daten aus dem Jahr 1975 und Versuche dazu wurden mit Röhrenmonitoren durchgeführt. Diese besitzen im Gegensatz zu den heute gebräuchlichen Bildschirmen eine Bildwiederholfrequenz, so dass bei der Darstellung bewegter Bilder Interferenzen zwischen den berechneten Einzelbildern und der Eigenfrequenz des Monitors entstehen. Um die tatsächlich durch das Auge wahrgenommene Frequenz zu ermitteln, wurde eine Testvorrichtung erstellt. Hierbei ergaben sich mehrere technische Probleme, deren Lösungen hier ebenfalls in Ausschnitten präsentiert werden. Es wird die Planung der technischen Umsetzung ohne Messergebnisse oder Auswertungen vorgestellt.

Keywords: Epilepsie, Photosensitiv, Speichertechnik, CUDA

1 Motivation

Barrierefreiheit in der IT ist ein Thema, das sowohl gesellschaftlich als auch technologisch immer weitreichender diskutiert wird. Durch die wachsende Teilhabe von Menschen mit Behinderung werden die Ersteller technischer Bedienoberflächen vor immer größere Herausforderungen gestellt. Daher wurden bereits Richtlinien geschaffen um die Barrierefreiheit mess- und testbar zu machen (vgl. [Ve02, IS12]).

Leider sind die verwendeten Grenzwerte zum Teil nur schwer nachzuvollziehen oder wurden auf Basis heute nicht mehr gültiger Versuche erstellt. Auffällig sind besonders die Grenzwerte zur photosensitiven Epilepsie, bei denen die Web Content Accessibility Guidelines (WCAG) einen Grenzwert von 3 Hz vorschreiben, wobei jedoch der am häufigsten zitierte Frequenzbereich zur Auslösung epileptischer Anfälle zwischen 10-20 Hz liegt [Mo08]. Die WCAG stützen sich auf Grenzwerte, die spezifisch unter Zuhilfenahme von Bildschirmen ermittelt wurden um auf den Einsatzzweck angepasste Grenzwerte zu erhalten. Die verwendeten Daten stammen aus Forschungen von *Harding* und *Binnie* [Bi02], die 2002 mehrere Artikel über den Einfluss von Fernseh Bildschirmen erstellten. Die Basis der verwendeten Daten stammt teilweise aus dem Jahr 1975, was bedeutet, dass diese Daten ausschließlich Röhrenmonitore (CRT) behandeln. Auch in den Papers von 2002 ist kein Hinweis auf eine Differenzierung zwischen CRT und Flachbildschirmen (LCD) zu finden. Dies stellt jedoch ein schwerwiegendes Problem dar, da die Darstellung der Bilder technisch völlig unterschiedlich gelöst ist: Während bei LCD die einzelnen Pixel eine neue Farbe zugewiesen bekommen und damit das Bild effektiv beim Übergang eines Frames auf den nächsten nur "angepasst" wird, so wird beim CRT mit einer festen Bildwiederholfrequenz das komplette Bild mehrmals pro Sekunde neu gezeichnet. Damit erhält bereits

¹ Technische Hochschule Ingolstadt, Fakultät EI, Esplanade 10, 85049 Ingolstadt, mathias.haimerl@thi.de

ein Standbild mit einer gewissen Helligkeit eine vom Betrachter wahrgenommene Frequenz. Wird zusätzlich ein bewegtes Bild angezeigt, so muss es daher zu Interferenzen kommen und die perzipierte Frequenz weicht drastisch von der ab, die an den Monitor übermittelt wird. Um diese These zu beweisen wurde eine Testvorrichtung erstellt, die die Frequenzen der vom Bildschirm dargestellten Bilder messen kann.

2 Testaufbau

Um die dargestellte Frequenz inklusive der Eigenfrequenz des Bildschirms zu messen, muss das System kamerabasiert sein. Die Kamera muss die Bilder anschließend speichern, damit die Berechnungen für die Auswertung nicht in Echtzeit durchgeführt werden müssen. Für die Berechnungen muss schließlich eine performante Möglichkeit der Bildverarbeitung gefunden werden.

2.1 Auswahl der Kamera

Als Voraussetzung für die Wahl der Kamera galten folgende Annahmen:

- **Framerate:** Um Frequenzen im Rahmen der gebräuchlichsten Bildwiederholfräquenzen von CRT messen zu können, muss die Kamera die doppelte Framerate besitzen, damit auch Zwischenbilder² mit der resultierenden Schwarzphase aufgenommen werden können. Da CRT Frequenzen bis zu 100 Hz normal sind, muss die Kamera für diesen Testaufbau 200 fps aufnehmen können.
- **Auflösung:** Die Auflösung der Kamera darf nicht höher sein als die des Bildschirms, da ansonsten unter Umständen nicht die einzelnen Pixel des Monitors aufgenommen werden, sondern einzelne Grundfarben. Dadurch werden die Intensitätswerte verfälscht. Die Auflösung muss jedoch hoch genug sein um kleine Bereiche abzudecken. Eine vollständige Übereinstimmung der Auflösungen von Kamera und Bildschirm ist nicht notwendig, da das menschliche Auge bei normalem Abstand zum Monitor keine einzelnen Pixel unterscheidet sondern Bereiche wahrnimmt, so wie auch nicht die dargestellten Grundfarben, sondern die additive Mischung derselben perzipiert werden.
- **Portabilität:** Zur Durchführung der Testreihe muss die Kamera möglicherweise an verschiedenen Computern betrieben werden, da eine große Auswahl an Monitoren bedeutet, dass das Testsystem portabel gestaltet werden muss. Somit wäre ein standardisierter Anschluss wünschenswert.

Nach einer Vorbesprechung zur Auswahl einer geeigneten Kamera mit der Firma *PCO*, wurden die Kameras der Firma *Basler* empfohlen. Diese stellten aufgrund der zuvor Beschriebenen Anforderungen eine *aca1300-200uc*³ zur Verfügung. Diese Kamera besitzt

² Zwischen dem Bildaufbau zweier Bilder dargestelltes Bild

³ <http://www.baslerweb.com/en/products/cameras/area-scan-cameras/ace/aca1300-200uc>

eine Framerate von bis zu 203 fps bei einer Auflösung von 1280 x 1024 Pixeln. Der Anschluss für Datenübertragung und Stromversorgung basiert auf einer USB 3.0 Schnittstelle und nutzt dort den Standard *USB3 Vision*. Dieser Standard ermöglicht es, Bildinformationen über USB3 mit einer Datenmenge von bis zu 350 MiB/s zu übertragen [AI13]. Als Objektiv wurde ein Standardobjektiv einer *Canon* Spiegelreflexkamera verwendet⁴, da dieses Lichtstark ist und zum Zeitpunkt der ersten Versuche verfügbar war. Es besitzt eine Brennweite von 18-55 mm und benötigt somit einen Mindestabstand von ca. 90 cm von einem 15"-Monitor.

2.2 Aufbau der Haltevorrichtung

Die Kamera muss exakt auf die Bildfläche des Bildschirms ausgerichtet werden. Damit diese Einstellung vorgenommen werden kann und auch bei länger andauernden Versuchen gewährleistet werden kann, wurde eine 5-Punkt-Fixierung der Kamera am Monitor entwickelt. Die Basis stellt ein Standard-Kamera-Stativ dar. Auf diesem wurde ein Käfig befestigt, in den die Kamera eingeschraubt wurde. Für einen perfekten Sitz wurde dieser Käfig speziell für die Kamera mittels 3D-Drucker erstellt.

An den vier Eckpunkten des Käfigs wurden Carbonstäbe angebracht, deren Länge anpassbar ist. Diese Stäbe besitzen an ihren Vorderseiten Winkel, die sich an nahezu jedem Monitor anbringen lassen. Diese Winkel werden mit Aluminiumstäben an den Ecken des Monitors fixiert. Der Abstand zum Monitor kann mit den Carbonstäben eingestellt und fixiert werden. Da Carbon leicht und unempfindlich gegenüber Stauchung ist, wird das System durch diesen Aufbau portabel und lässt eine optimale Fixierung der Kamera zu.

2.3 Kalibrierung

Um die Einstellung von Abstand, Ausrichtung und Fokus der Kamera sowie des Objektivs zu erleichtern, wurde ein einfaches Programm erstellt, das die äußersten 5 x 5 Pixel in den vier Ecken anzeigt, so dass die Kamera darauf ausgerichtet werden kann. Zusätzlich wird in der Mitte des Bildschirms ein schwarz-weißes Testbild angezeigt, das einerseits die Fokussierung des Objektivs unterstützt und andererseits die Helligkeitskalibrierung ermöglicht. Da Monitore, abhängig von der Helligkeitseinstellung und der lokalen Lichtverhältnisse kein reines schwarz bzw. weiß anzeigen können, muss für die Durchführung der Tests der hellstmögliche Wert als "optimalweiß" bzw. der dunkelste darstellbare Wert als "optimalschwarz" ermittelt werden. Damit lassen sich die während der Versuchsdurchführung aufgenommenen Helligkeitswerte skalieren. Hierbei ist anzumerken, dass CRT-Monitore durch Absetzen der Beleuchtung einzelner Bildpunkte ein dunkleres "optimalschwarz" darstellen können, da die Beleuchtung von LCD-Monitoren dauerhaft aktiviert ist. Somit ist anzunehmen, dass die Differenz zwischen den hellsten und dunkelsten Werten bei CRT-Monitoren größer ist als bei LCD-Monitoren.

⁴ Canon EF-S 18-55mm f/3.5-5.6 III

3 Testdurchführung

3.1 Speicherung der Daten

Eine besondere Herausforderung bestand im Speichern der Daten für die spätere Verwendung. Die von der Kamera übertragenen Bilder werden als unkomprimierte Rastergrafiken gespeichert, um zu verhindern, dass zusätzliche Prozessorlast durch Konversion oder Kompression erzeugt wird. Da die Verarbeitung der Werte anschließend ebenfalls auf Pixeln basiert, müssten komprimierte Grafiken vor der Verarbeitung erneut dekomprimiert werden. Daher bot sich diese Art der unkomprimierten Speicherung an.

Durch die große Anzahl an Bildern pro Sekunde und der hohen Auflösung fällt eine Datenmenge von $24\text{Bit} \cdot 1280\text{px} \cdot 1024\text{px} \cdot 200\text{fps} \approx 752\text{MiB/s}$ an. Diese Datenmenge zu speichern stellt hohe Ansprüche an die verwendete Hardware, so dass hierfür vorab eine Evaluation möglicher Speichermedien durchgeführt werden musste. Da die anfallende Datenmenge über die Möglichkeiten des internen Write-Cache von Festplatten hinausgeht, wurde dieser für die Versuche deaktiviert. Als Testdaten wurden 10 Blöcke mit jeweils 1 GB Größe auf die Festplatte geschrieben. Außer Acht gelassen wurden hierbei die Schreibperformanzen verschiedener Dateisysteme.

- **Festplatte:** Der erste Versuch wurde mit einer konventionellen Magnetfestplatte durchgeführt. Bei den Versuchen konnte ein Durchsatz von ca. $77,5\text{MiB/s}$ gemessen werden. Damit ist diese Geschwindigkeit bei Weitem nicht ausreichend, so dass andere Speichermedien evaluiert wurden.
- **SSD:** Der nächste Versuch wurde mit einer Solid-State-Drive (SSD) Festplatte durchgeführt. Das verwendete Modell *Crucial M4 64GB* konnte mit deaktiviertem Write-Cache nur eine Schreibgeschwindigkeit von $11,8\text{MiB/s}$ erreichen. Da diese Festplatte eine der ersten Budget-SSDs war, ist es möglich, dass dieser schlechte Wert der Verarbeitung geschuldet ist. Diese Vermutung wird auch durch die bei *UserBenchmark* ausgewerteten Werte untermauert⁵, bei denen die besten und schlechtesten Werte sehr stark differieren. Zusätzlich lassen sich in vielen Foren Hinweise darauf finden, dass diese SSD nur bei der Verwendung der optimalen Blockgröße mit maximaler Geschwindigkeit arbeitet, welche beim Test nicht verwendet wurde.
- **SSD RAID:** Um die Schreibgeschwindigkeit zu erweitern wurde ein Redundant Array of Independent Disks (RAID) im Redundanz-Level 0 erstellt. Dadurch soll die Schreibgeschwindigkeit durch paralleles Schreiben auf mehrere Festplatten erhöht werden. Als Festplatten wurden vier *SanDisk SSD Plus 120GB* verwendet, mit einem nominalen Datendurchsatz von 120GB/s bei schreibendem Zugriff⁶. Als RAID-Controller diente ein *HP Smart Array P410i*. Bei diesem Versuch konnten Werte von 340MiB/s gemessen werden. Da dieser RAID-Controller mit einem optionalen Hardware Write-Cache nachgerüstet werden kann, der die Schreibgeschwindigkeit verbessern soll, wurde ein Flash-Based Write-Cache (FBWC) nachgerüstet. Dieser konnte den Datendurchsatz auf 440MiB/s erhöhen.

⁵ <http://ssd.userbenchmark.com/SpeedTest/1215/M4-CT064M4SSD2>

⁶ <https://www.sandisk.de/home/ssd/ssd-plus>

- M.2 SSD:** Da diese Geschwindigkeit noch nicht ausreichte, und die SATA III Schnittstelle ein Limit von 600 MiB/s hat, musste eine alternative Schnittstelle gefunden werden. Der Nachfolger der in Laptops verwendeten *Mini-PCIe*-Schnittstelle ist die sog. *M.2*-Schnittstelle, die auf modernen Motherboards verbaut ist. Auf dieser Schnittstelle wurde der Versuch mit einer *Samsung SM951* durchgeführt, einer SSD mit *M.2*-Schnittstelle. Dabei konnte eine Schreibgeschwindigkeit von ca. $1,4 \text{ GiB/s}$ gemessen werden, also mehr als die dreifache Geschwindigkeit des SSD-RAID-Verbunds.

Der Vergleich der Schreibgeschwindigkeiten findet sich in Abb. 1.

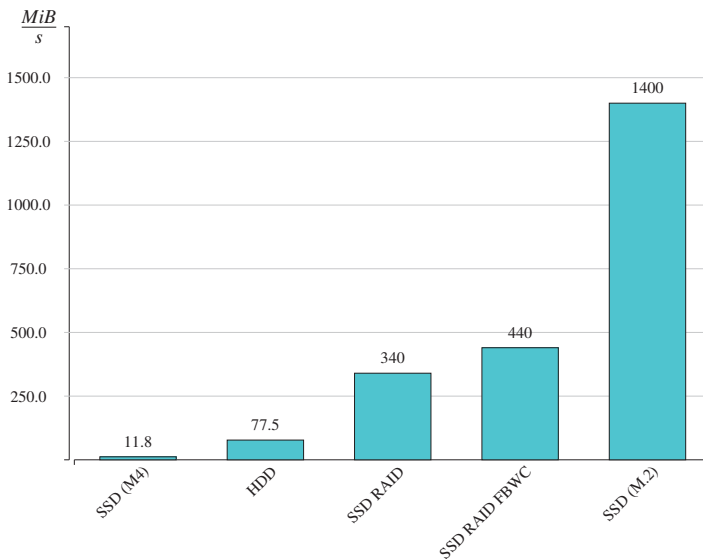


Abb. 1: Vergleich der Schreibgeschwindigkeiten verschiedener Medien

3.2 Berechnung der Intensitätsmatrizen

Eine Rastergrafik kann generell als zweidimensionale Matrix der Form $I \in \text{mat}(M, N)$ dargestellt werden, wobei jeder Bildpunkt $p \in I$ einen Wert der jeweiligen Farbtiefe annimmt. In diesem Fall nimmt ein Bildpunkt des Farbbildes einen 24-Bit Farbwert an, also $p_{m,n} = (r, g, b)$ wobei gilt $m \in [0, M[; n \in [0, N[; r, g, b \in [0, 255]$. Aus den Farbwerten lassen sich die vom Auge wahrgenommenen Intensitätswerte berechnen und auf eine Intensitätsmatrix $Y \in \text{mat}(M, N)$ abbilden, so dass wir eine Funktion $f: I \mapsto Y$ definieren können. Aufgrund der Verteilung der Zapfen im Auge werden verschiedene Farben mit einer unterschiedlichen Intensität wahrgenommen [BD80]. Die wahrgenommenen Anteile der Farben wurden von der *International Telecommunication Union, Radiocommunication Sector (ITU-R)* in den Standard *ITU-R BT-601* zur Übertragung digitaler Videosignale übernommen⁷. Daraus lässt sich die Funktion f zur Berechnung eines 8 Bit Graustufen-

⁷ <http://www-inst.eecs.berkeley.edu/cs150/Documents/ITU601.PDF>

bildes als Repräsentation der Intensitäten wie in (1) definieren:

$$f : I \mapsto Y$$

Sei $p_{m,n} = (r, g, b)$ die RGB-Kodierung des Pixels an Indexposition (m, n) der Matrix I

Wir definieren den Intensitätswert des entsprechenden Pixels der Matrix Y wie folgt:

$$f(p_{m,n}) = 0.299 \cdot r + 0.587 \cdot g + 0.114 \cdot b \tag{1}$$

3.2.1 Durchführung der Berechnungen

Durch die Anzahl der aufgenommenen Bilder pro Versuchsdurchführung, muss eine große Menge an Bildern verarbeitet werden. Mit einer derzeit üblichen 4-Kern-CPU dauern diese Berechnungen verhältnismäßig lange. In den ersten Versuchen mit sequenzieller Bearbeitung wurden für die Berechnung der Identitätsmatrizen von 3 Bildern Berechnungszeiten um die 8 Sekunden gemessen. Da die Bildpunkte unabhängig voneinander berechnet werden können, lassen sich die Berechnungen hochgradig parallelisieren. Durch eine einfache Parallelisierung auf der CPU mit C++11-Threads konnte diese Zeit auf ca. 4 Sekunden halbiert werden. Da jedoch für einen Versuchsdurchlauf mehrere tausend Bilder berechnet werden müssen, ist diese Zeit zu lange um große Versuchsreihen in akzeptabler Zeit abzuarbeiten. Als Lösung wurde die Berechnung auf der Graphics Processing Unit (GPU) einer modernen Grafikkarte favorisiert. Hierfür wurde eine *Nvidia GeForce GTX 980 Ti* ausgewählt, deren GPU 2816 arithmetisch-logische Einheiten (ALU) besitzt. Die Ansteuerung des Grafikspeichers und der GPU wurde über die proprietäre Schnittstelle *CUDA* von *Nvidia* umgesetzt. Durch die hochgradige Parallelisierung sank die Berechnungsdauer auf 0,007 Sekunden für die selben drei Testbilder.

3.3 Ermittlung der Frequenzen

Über die Intensitätsmatrizen können die Veränderungen in den Bildern im relevanten Frequenzbereich ermittelt werden. Mit der verwendeten Kamera lassen sich theoretisch Werte bis 200 Hz ermitteln. Die Versuche werden jedoch lediglich den Bereich von 1-50 Hz abdecken, da dort die Frequenzen liegen, die photosensitive Menschen stimulieren. Dafür müssen alle Bilder, die innerhalb des Ermittlungszeitraums aufgenommen wurden miteinander verglichen und die Veränderungen berechnet werden. Nach der pixelweisen Berechnung müssen die Bereiche gefunden werden, in denen die Grenzwerte der WCAG einer Intensitätsänderung von 80 % in 25 % des Bereiches von 10 % des Sichtfelds überschreiten⁸.

⁸ <https://www.w3.org/TR/UNDERSTANDING-WCAG20/seizure-does-not-violate.html>

4 Zusammenfassung und Ausblick

Durch die Berechnung der Frequenzen über Intensitätsmatritzen fehlen in den Versuchen bisher die “roten Blitze”. Die WCAG sehen vor, dass außer den Änderungen der perzierten Luminanz der Wechsel eines gesättigten Rottons (definiert als $\frac{PR}{PR+PG+PB} > 0,8$) bei mindestens einem der beiden zu vergleichenden Werte überprüft werden muss, da rote Blitze den visuellen Cortex ebenfalls in potentiell Epilepsie auslösender Art stimulieren. Dafür werden anstatt Intensitätsmatritzen Masken zur Identifikation gesättigter Rotwerte erstellt, die in die Frequenzberechnung mineinbezogen werden.

Wenn in der Untersuchung der Ergebnisse zwischen LCD und CRT ein starker Unterschied der dargestellten Frequenzen auftritt, so kann dieses Ergebnis als Beweis dienen, dass Menschen mit photosensitiver Epilepsie durch die Vermeidung von CRT das Anfallsrisiko bei der Verwendung von IT-Systemen deutlich senken können. Zusätzlich sollte ein Unterschied zwischen den effektiven Frequenzen der CRT-Bilder bei verschiedenen Bildwiederholfrequenzen messbar sein. Das bedeutet, dass es sich tatsächlich um Interferenzen des dargestellten Bildes mit der Bildschirmfrequenz handelt. Damit müsste eine Revision der bestehenden Grenzwerte angedacht werden.

Literaturverzeichnis

- [AI13] AIA: USB3 Vision Standard Specification. Bericht, Global Association for Vision Information, Ann Arbor, MI, USA, jan 2013.
- [BD80] Bowmaker, JK; Dartnall, HJk: Visual pigments of rods and cones in a human retina. The Journal of physiology, 298(1):501–511, Marz 1980.
- [Bi02] Binnie, CD; Emmett, J; Gardiner, P; Harrison, D; Wilkins, AJ: Characterizing the flashing television images that precipitate seizures. SMPTE journal, 111(6-7):323–329, 2002.
- [IS12] ISO: Information technology – W3C Web Content Accessibility Guidelines (WCAG) 2.0. ISO ISO/IEC 40500:2012, International Organization for Standardization, Geneva, Switzerland, 2012.
- [Mo08] Mothersill, Ian W.; Cenusa, Margarethe; Grunwald, Thomas; Krämer, Günter: Visuelle Reflexanfälle: Wie gefährlich sind Fernsehen sowie Video- und Computerspiele? 2008(2):28ff, Juni 2008.
- [Ve02] Verordnung zur Schaffung barrierefreier Informationstechnik nach dem Behindertengleichstellungsgesetz (BITV). Bundesministerium des Innern, Juli 2002.

Workshop
Informatik-Begeisterung
2016

Workshop: Informatik-Begeisterung 2016

Judith Michael¹ und Ingo Scholtes²

Mit den für die jeweilige Zielgruppe passenden Inhalten kann man Kindergartenkinder, Schülerinnen und Schüler unterschiedlicher Schulstufen, Studierende aber auch die Allgemeinheit für die Informatik begeistern. Neben der fachlichen Expertise ist die Fähigkeit der populärwissenschaftlichen Präsentation von komplexen Inhalten zunehmend bedeutsam.

Der Workshop „Informatik – Begeisterung 2016“ hat das Ziel wirksame Kommunikationspraktiken und spannende Informatikinhalte für unterschiedliche Altersstufen und Bildungshintergründe auszutauschen und zu diskutieren. Es wurden 15 Beiträge als wissenschaftlicher Kurzbeitrag, populärwissenschaftlicher Kurzbeitrag bzw. Zusammenfassung/Abstract eines Vortrags eingereicht, 13 Beiträge werden im Rahmen des Workshops vorgestellt und diskutiert.

Die Beiträge umfassen eine Reihe von spannenden Themen und Projekten:

- Die Bundesweiten Informatikwettbewerbe (BWINF) bieten Angebote vom Einstieg bis hin zur Förderung von Spitztalenten (Deutschland)
- Das Science Camp Informatik am Karlsruher Institut für Technologie (KIT) vermittelt Schülerinnen Spaß am Programmieren
- Initiativen wie IT Feriapraktika, Workshops für Schülerinnen und Schüler oder Weiterbildung für Lehrerinnen und Lehrer an der Fakultät für Technische Wissenschaften (TEWI) der Alpen-Adria-Universität Klagenfurt
- Der Einsatz von Unterrichtsbausteinen zur kooperativen Entwicklung und Erprobung von Informatikstunden an Grundschulen (Universität Münster und Grundschule Gronau)
- Die Wiener Zauberschule der Informatik (WIZIK) für Kinder der Primarstufe
- Anmerkungen zu einem begeisternden Informatikunterricht: nach dem Staunen Interesse wecken, das zu forschend-entdeckendem Unterricht motiviert
- Die „Informatik-Werkstatt“ an der Alpen-Adria-Universität Klagenfurt bietet Einblicke in Bereiche der Informatik wie Verschlüsselung, Modellierung, Algorithmen, Programmierung oder Netzwerke

¹ Alpen-Adria-Universität Klagenfurt, Institut für Angewandte Informatik, Universitätsstraße 65-67, 9020 Klagenfurt, Österreich, judith.michael@aau.at

² ETH Zürich, Lehrstuhl für Systemgestaltung, Weinbergstrasse 56/58, 8092 Zürich, Schweiz, ischoltes@ethz.ch

- Programmierkursen in Java für Flüchtlinge (KIT) – ein aktiver Lernansatz für heterogene Gruppen
- Die Initiative IFIT (Österreich) begeistert den Nachwuchs für IT und Technik
- Im Denkwerk-Projekt modellieren und analysieren Schülerinnen und Schüler Prozesse auf Nachhaltigkeit (Universität Hildesheim)
- Die Veranstaltungsreihe „Pi and More“ mit Vorträge und Workshops zu dem Single-Board-Computer Raspberry Pi und ähnlichen Systemen (Universität Trier)
- Der Weg von der Problemanalyse (Sketch) bis zur formalen Beschreibung eines Algorithmus (in Scratch) für Kinder (Pädagogische Hochschule Linz)
- Begabtenförderung mit Begeisterung für Robotik im Talentehaus Niederösterreich

Ein besonderer Dank geht an die Autorinnen und Autoren der eingereichten Beiträge und an die Mitglieder des Programmkomitees für ihre hervorragende Unterstützung bei der Begutachtung und Auswahl der Beiträge.

Organisation

Judith Michael
Ingo Scholtes

Alpen-Adria-Universität Klagenfurt (GI Juniorfellow)
ETH Zürich (GI Juniorfellow)

Programmkomitee

Thomas Bartoschek
Torsten Brinda

Universität Münster (GI@School)
Universität Duisburg-Essen, GI Fachbereich "Informatik und
Ausbildung, Didaktik der Informatik"

Jörg Desel
Gerald Futschek
Rainer Gemulla
Lutz Hellmig

FernUni Hagen
TU Wien, OCG
Universität Mannheim (GI-Juniorfellow)
Uni Rostock, GI Fachausschuss „Informatische Bildung in
Schulen“

Juraj Hromkovic
Agnes Koschmider
Bernhard Löwenstein
Walid Maalej
Judith Michael
Wolfgang Pohl
Sebastian Schäfer
Katharina Anna Zweig

Informationstechnologie und Ausbildung, ETH Zürich
KIT (GI-Juniorfellow)
Institut zur Förderung des IT-Nachwuchses, IFIT
Universität Hamburg (GI-Juniorfellow)
Alpen-Adria-Universität Klagenfurt (AAU)
Bundeswettbewerb Informatik
Öffentlichkeitsarbeit der Informatik Fakultät des KIT
TU Kaiserslautern (GI-Juniorfellow)

Klagenfurt und Zürich, im Juni 2016

Judith Michael und Ingo Scholtes

BWINF: Informatik mit Begeisterung entdecken

Wolfgang Pohl¹

Abstract: Die Bundesweiten Informatikwettbewerbe (BWINF) stellen vom Informatik-Einstieg bis hin zur Förderung von Spitzentalenten ein kontinuierliches Angebot an Wettbewerbsformaten für Schülerinnen und Schüler zum Thema Informatik bereit. Abgerundet wird dies durch begleitende Maßnahmen zur Weiterbildung und Förderung der Teilnehmerinnen und Teilnehmer. Durch die Wettbewerbsteilnahmen entdecken die jungen Leute das Fach Informatik und erfahren Begeisterung, indem sie ihr Potenzial, aber auch Gleichgesinnte kennen lernen.

Keywords: Informatikwettbewerbe, Schülerwettbewerbe, Talentförderung

1 Bundesweite Informatikwettbewerbe: Die Pyramide

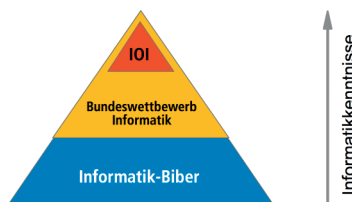


Abb. 1: Bundesweite Informatikwettbewerbe: Pyramidales Fördersystem

Das Projekt „Bundesweite Informatikwettbewerbe“ (BWINF) ist im Sinne der Vereinigung von Breiten- und Spitzenförderung pyramidal aufgebaut (vgl. Abbildung 1) und besteht aus drei Wettbewerbsstufen:

- Auf der Eingangsstufe steht der Informatik-Biber [PSH09]. Dieses Online-Format wird seit 2007 jährlich angeboten, bislang für Schülerinnen und Schüler der Sekundarstufe. Mit Aufgaben, die keine fachlichen Vorkenntnisse erfordern, und stark begrenztem Teilnahmeaufwand (für SchülerInnen, aber auch für Lehrkräfte) weckt dieses Format Interesse für Informatik und klärt über Inhalte der Informatik auf. Teilnehmende können ihr Potenzial für das Fach Informatik entdecken und werden motiviert, sich näher mit der Disziplin auseinanderzusetzen. Am Informatik-Biber nahmen im Jahr 2015 in Deutschland knapp 250.000 Kinder und Jugendliche teil.
- Die zentrale Stufe bildet der Bundeswettbewerb Informatik (BwInf) [Poh06]. 1980 gegründet, wird er seit 1984 jährlich als Aufgabenwettbewerb durchgeführt. Dieser Leistungswettbewerb steht Jugendlichen bis 21 Jahren offen, die noch keine Ausbildung abgeschlossen bzw. noch kein Studium aufgenommen haben. Ein Wettbewerbserfolg setzt erste Fachkenntnisse wie Wissen über grundlegende algorithmische

¹ BWINF (Bundesweite Informatikwettbewerbe), Reuterstr. 159, 53113 Bonn, pohl@bwinf.de

Prinzipien und erste Programmierfertigkeiten voraus. Neben der fachlichen Vertiefung trägt der BwInf dazu bei, Jugendliche mit besonderem Potenzial zu erkennen, und ermöglicht den Teilnehmenden, ihre Begabung zu erfahren und auszuloten.

- Die besten Teilnehmenden des BwInf haben schließlich die Chance, sich über ein Auswahlverfahren für das deutsche Team bei der Internationalen Informatikolympiade (IOI) zu qualifizieren. Diese dritte Wettbewerbsstufe fördert Jugendliche mit außerordentlichem Potenzial und vermittelt Fachkenntnisse auf Hochschulniveau.

2 Wirkung von Wettbewerbsteilnahme und -durchführung

Bei allen BWINF-Wettbewerbsstufen handelt es sich um Aufgabenwettbewerbe. Die Teilnehmenden bearbeiten vorgegebene Aufgaben und erhalten eine Bewertung ihrer Leistung. Je nach Bewertung werden sie mit Preisen, Auszeichnungen und – bei Wettbewerben mit mehreren Runden wie dem BwInf – mit der Berechtigung zur Teilnahme an einer weiteren Runde belohnt. So können inhaltliche und methodische Kompetenzen im Fach des Wettbewerbs überprüft, erkannt und ausgezeichnet werden. Außerdem kommen in Leistungsformaten außerfachliche Kompetenzen wie Ausdauer, Selbstdisziplin und die Fähigkeit zu organisiertem Arbeiten zum Tragen.

Darüber hinaus kann eine Wettbewerbsteilnahme bzw. die Durchführung eines Wettbewerbs viele weitere, indirekte Effekte haben. Am Beispiel der Bundesweiten Informatikwettbewerbe werden zwei wesentliche Effekte genannt.

2.1 Informatik entdecken

Im Informatik-Biber erfolgt häufig zum ersten Mal eine Begegnung mit Inhalten der Informatik. Ohne Vorkenntnisse lösen die Teilnehmenden Fragestellungen, die ihnen wie logische Puzzles oder Knobelaufgaben vorkommen, die aber immer einen klaren Informatikgehalt haben. Letzterer wird in den Aufgabenheften dokumentiert und beschrieben, die BWINF nach jeder Austragung eines Informatik-Biber publiziert². Die Vorstellungen der Kinder und Jugendlichen von den Inhalten des Fachs Informatik wird dadurch positiv beeinflusst. Außerdem liefern die Aufgaben häufig Anregungen für den Informatik-Fachunterricht. Tabelle 1 belegt die Effekte des Informatik-Biber mit den Aussagen einiger Lehrkräfte.

Im Bundeswettbewerb Informatik gibt es ähnliche Effekte. Wer Programmierfertigkeiten mit Informatikkenntnissen gleichsetzt, erfährt bei der Bearbeitung der Aufgaben, aber auch durch die Vorgaben zur Strukturierung einer Einsendung und die von BWINF veröffentlichten Beispiellösungen, dass die Lösung eines Informatik-Problems eine Modellierung und eine Diskussion der Lösungsqualität beinhaltet. Im Rahmen des IOI-Auswahlverfahrens lernen die Jugendlichen neue Fachinhalte kennen und ihr hohes Potenzial im nationalen und internationalen Vergleich mit anderen Spitztalenten einzuschätzen.

² www.informatik-biber.de/download

Außerdem finde ich die Aufgabenauswahl sehr gelungen, es macht den Schülerinnen und Schülern sehr viel Spaß. ... *Tobias K., Essen*

O-Ton einer Sechstklässlerin, die heute am Biber-Wettbewerb teilgenommen hat: „Das war voll cool, können wir das öfter machen?“ *Marcus R., Soest*

Wenn man aber genau hinschaut, erkennt man sehr wohl, wo Informatik drinsteckt. Eine Aufgabe aus dem letzten Jahr habe ich als Aufhänger für eine Informatikklausur der Jahrgangsstufe 12 verwendet. *Josef J., Illertissen*

Tab. 1: Statements von Lehrkräften zur Teilnahme am Informatik-Biber

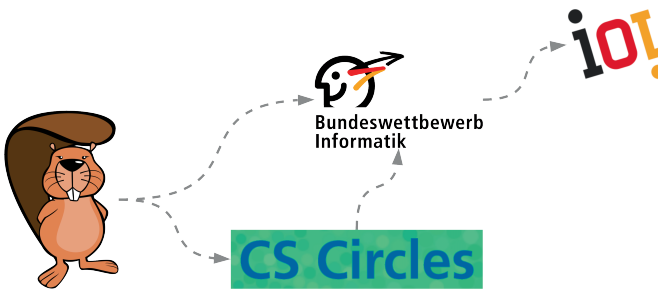


Abb. 2: Einbindung von Lernangeboten in die „BWINF-Wettbewerbskarriere“

Dabei besteht in der BWINF-Wettbewerbspyramide derzeit noch eine Lücke. Wenn eine Teilnahme am Informatik-Biber das Interesse an Informatik erfolgreich geweckt hat, müssen zur Teilnahme am Bundeswettbewerb Informatik erste Fachkenntnisse erworben werden. Da dies zur Zeit durch die Schulbildung nicht sichergestellt ist, arbeitet BWINF am Aufbau eigener Lernmodule. Die deutsche Variante des Online-Kurses „Computer Science Circles“³ ist hierbei ein erster Schritt und bietet eine leicht zugängliche Möglichkeit zum Erwerb grundlegender Programmierkenntnisse (vgl. Abbildung 2).

2.2 Begeisterung durch Austausch

In Ergänzung des Wettbewerbsangebots werden insbesondere die BwInf-Teilnehmenden weiter gefördert. Schon in den ersten BwInf-Finalrunden, die als Kolloquium mit Einzelprüfungen und Gruppenarbeit organisiert sind, wurde deutlich, dass die Teilnehmenden vom Austausch mit Gleichgesinnten besonders profitieren. Dieser Effekt sollte aber nicht nur den wenigen Finalisten zugänglich sein. BWINF hat deshalb schon früh nach Partnern gesucht, die sog. Teilnehmerworkshops ausrichten und den Jugendlichen, die Interesse und Talent durch eine erfolgreiche Teilnahme an der ersten BwInf-Runde nachgewiesen haben, sowohl eine Gelegenheit zum gegenseitigen Kennenlernen als auch die begehrte Möglichkeit liefern, weitergehende Fachkenntnisse zu erwerben. Vor über 25 Jahren wurde das erste „Jugendforum Informatik“ für BwInf-Teilnehmende aus Baden-Württemberg

³ csc.cemc.uwaterloo.edu/0_de



Abb. 3: Zwei Teilnehmerinnen am BwInf-Workshop „Girls at Google“

organisiert. Langjährige BWINF-Partner bei der Ausrichtung solcher Workshops sind u. a. das Hasso-Plattner-Institut in Potsdam, die RWTH Aachen, die TU Dortmund und das Max-Planck-Institut für Informatik sowie Google Deutschland mit einem speziellen Angebot für Teilnehmerinnen (s. Abbildung 3). Das Potenzial solcher Veranstaltungen wird am besten in den Äußerungen von Teilnehmenden erkennbar (s. Tabelle 2).

Mir wird das Camp in positiver Erinnerung bleiben. Vor allem die Teamaufgaben machten Spaß, ... *Daniel, Friedrichroda*

Im Team zu Arbeiten war für mich eine ganz neue Erfahrung. Die Aufgaben waren gut durchdacht und regten zum weiteren Nachdenken an. *Tobias, Regensburg*

Ein cooles Camp, würde ich jederzeit wieder machen. *Tobias, Lünen*

Während des BwInf-Camps 2015 konnte ich [...] den geförderten Austausch untereinander genießen. *Thilo, Heidelberg*

Tab. 2: Statements von Teilnehmenden an BwInf-Workshops

3 Ausblick

Die Bundesweiten Informatikwettbewerbe leisten durch ihre Wettbewerbe und die ergänzenden Angebote einen wesentlichen Beitrag, die Begeisterung für Informatik bei Kindern und Jugendlichen nachhaltig zu wecken. Um dies zu verstärken, wird der Informatik-Biber ab 2016 auch für die Klassenstufen 3 und 4 angeboten, mit eigenen altersgerechten Aufgaben. Außerdem werden ein „Junior-BwInf“ mit Online-Programmieraufgaben und weitere Lernangebote die Lücke zwischen Biber und BwInf schließen.

Literatur

[Poh06] Wolfgang Pohl. Wettbewerb im Silberglanz. *LOG IN*, 26(141/142):10–13, 2006.

[PSH09] Wolfgang Pohl, Kirsten Schlüter und Hans-Werner Hein. Informatik-Biber: Informatik-Einstieg und mehr. In Bernhard Koerber, Hrsg., *Zukunft braucht Herkunft: 25 Jahre INFOS – Informatik und Schule*, Seiten 38–49, Bonn, 2009. Gesellschaft für Informatik.

Das Science Camp Informatik – hier können Mädchen Informatik neu erleben

Lena Winter¹, Sebastian Schäfer² und Michael Gauß³

Abstract: Das Science Camp Informatik bietet Mädchen im Alter von 14 bis 16 Jahren die Chance Informatik einmal anders als in der Schule kennenzulernen. Im Camp kann jede Schülerin ihr eigenes Projekt umsetzen und erleben, dass Informatik mehr ist als Programmiersprachen zu beherrschen.

Keywords: Science Camp Informatik, Sommercamp für Schülerinnen

1 Einleitung

Das Science Camp Informatik ist ein einwöchiges Sommercamp für Schülerinnen zwischen 14 und 16 Jahren, das am Karlsruher Institut für Technologie (KIT) von der KIT-Fakultät für Informatik gemeinsam mit der Fachschaft für Mathematik und Informatik sowie dem Zentrum für Mediales Lernen veranstaltet wird. Das Programm besteht aus einem Workshop und Exkursionen rund um das Thema Informatik. Die Zielsetzung des Camps ist es in erster Linie, mit den gegen Informatik bestehenden Vorurteilen aufzuräumen - im Speziellen, dass Informatik "nur etwas für Jungs" sei.

Das Science Camp Informatik wird in den Sommerferien 2016 zum zweiten Mal am Karlsruher Institut für Technologie (KIT) angeboten.

2 Motivation zur Konzeption eines Informatik-Camps für Mädchen

Die Grundidee zum Science Camp Informatik ist es, ein Programm zu bieten, in dem Schülerinnen an die Informatik herangeführt werden, ohne dass sie den Konkurrenzdruck durch männliche Teilnehmer fürchten müssen.

Die Idee zu dem Projekt stammt von studentischer Seite und ist darin motiviert, dass weniger Frauen als Männer eine Laufbahn im Informatikbereich wählen. So waren laut Statistischem Bundesamt 2015 an allen Hochschulen in Deutschland nur 14,8% aller 80399 Studienanfänger im Fachgebiet Informatik weiblich.

¹ Fachschaft Informatik und Mathematik, Am Fasanengarten 5, 76131 Karlsruhe, lena.winter@fsmi.uni-karlsruhe.de

² KIT-Fakultät für Informatik, Dekanat, Am Fasanengarten 5, 76131 Karlsruhe, sebastian.schaefer@kit.edu

³ Zentrum für Mediales Lernen, Karl-Friedrich-Str. 17, 76133 Karlsruhe michael.gauss3@kit.edu

Ein Grund dafür könnte der geschlechtsspezifische Umgang mit dem Computer sein. Jungen sind eher bereit, sich Computerkenntnisse autodidaktisch anzueignen bzw. sich durch spielerische Nutzungen zu erschließen. Mädchen hingegen gehen praktischer und zielgerichteter vor und nutzen den Computer vor allem zweckbezogen zum Lösen bestimmter Aufgaben. [Ti07]

Zusätzlich ist Informatik eine Disziplin, deren Inhalte in Schulen schlecht kommuniziert werden, wodurch Informatik bei Schülerinnen und Schülern oftmals mit reinem Programmieren gleich gesetzt wird.

Die Idee des Science Camps ist es, eine sichere und entspannte Atmosphäre für die Schülerinnen zu bieten, in der sie praktische Erfahrung mit dem Computer sammeln können und dabei lernen, dass Informatik weit mehr als nur Programmieren ist.

2.1 Umsetzung

Um die Umsetzung der Maßnahmen kümmern sich die drei Partner Fachschaft Mathematik und Informatik, KIT-Fakultät für Informatik sowie das Zentrum für Mediales Lernen gemeinsam. Das Zentrum wird größtenteils im Vorfeld des Camps aktiv und koordiniert die Bewerbung an Schulen sowie die gesamte Teilnehmerabwicklung; sodass sich die Studentinnen aus der Fachschaft sowie die Mitarbeiter der Fakultät ganz auf die inhaltlichen Arbeiten konzentrieren können. Gemeinsam mit den betreuenden Studentinnen verbringen die Jugendlichen jeden Camptag am KIT, wo sie während der Workshopphasen an bereitgestellten Laptops arbeiten. Um das Programm abwechslungsreich zu gestalten, wird in der Regel an jedem Tag des Camps eine Exkursion entweder zu einer Forschungseinrichtung des KIT oder zu einer Firma in Karlsruhe angeboten.

2.2 Projektarbeit

Jede Teilnehmerin soll in die Lage versetzt werden, ihre eigene Webseite zu gestalten und mit Javascript dieser Webseite einfache Funktionalitäten zu geben.

Die Wahl eines Projektes, das in einer Woche umgesetzt werden kann, hat sich dabei als relativ schwierig erwiesen, da die möglichen Projekte folgende Anforderungen erfüllen mussten:

1. Es soll sich um ein einfaches Projekt handeln, für das möglichst wenige unterschiedliche Programme benötigt werden. So kann die Komplexität für die Jugendlichen auf einem angemessenen Maß gehalten werden.
2. Die Schülerinnen sollen auch nach Ende des Science Camps an ihrem Projekt weiterarbeiten können.
3. Die Schülerinnen sollen bei der Gestaltung größtmögliche Freiheit haben.

4. Die Ergebnisse ihres Handelns sollen für die Jugendlichen über ein visuelles Feedback sofort nachvollziehbar sein.
5. Ziel ist es, dass die Schülerinnen verstehen, wie Sie über einen Code ihre Idee gestalten und umsetzen können.

Mit einer einfachen Webseite, die auf HTML, CSS und JavaScript basiert, können die oben genannten Anforderungen erfüllt werden.

Die Projektarbeit wurde in drei Teile aufgeteilt. Am ersten Tag des Workshops stehen die Grundlagen von HTML im Fokus. Am zweiten Tag erhalten die Jugendlichen eine Einführung in CSS, am dritten Tag ist eine Einführung in JavaScript vorgesehen. Für die restlichen zwei Tage wurden Anleitungen für verschiedenen Features mit unterschiedlichen Schwierigkeitsgraden erstellt, welche sich die Schülerinnen eigenständig und modular erarbeiten.

Um für die Jugendlichen deutlich zu machen, dass es sich beim Science Camp Informatik nicht um eine weitere Schulwoche in den Sommerferien handelt, war es wichtig, während des gesamten Workshops auf eine entspannte Atmosphäre zu achten. Jede Schülerin konnte in ihrem eigenen Tempo arbeiten und kleine Pausen selbst gestalten. Um diese Binnendifferenzierung zu gewährleisten, werden die Teilnehmerinnen immer von zwei Personen aus der Fachschaft betreut. Damit kommen auf eine Betreuerin sechs Schülerinnen, dank dieses guten Betreuungsverhältnisses ist gewährleistet, alle aufkommenden Fragen und Probleme sofort und individuell behandeln zu können.

2.3 Exkursionen

Ziel der Ausflüge ist es, den Schülerinnen die Aufgabenfelder von Informatikerinnen in verschiedenen Firmen vorzustellen. Dazu treffen die Mädchen Fachkräfte aus Industrie und Forschung, die von ihrem Arbeitsalltag berichteten und ihren Werdegang vorstellen.

2015 besuchten die Teilnehmerinnen die Firmen SAP und Gameforge und lernten dort Informatikerinnen aus unterschiedlichsten Fachbereichen kennen. Daneben konnten die Jugendlichen ein Forschungsprojekt aus dem Bereich der Medizinrobotik sowohl in der Anwendung im Städtischen Klinikum Karlsruhe, als auch im entsprechenden Forschungslabor am KIT kennen lernen.

In 2016 sind Exkursionen zur Zentrale der Drogeriemarktkette dm und dem Cyberforum geplant. Das Cyberforum ermöglicht uns mehrere Startups in Karlsruhe zu besuchen. So können die Teilnehmenden Gründeratmosphäre spüren und Selbstständigkeit als weitere Zukunftsmöglichkeit für Informatikerinnen begreifen. Darüber hinaus werden wir das FZI Forschungszentrum Informatik besuchen. Dort wird neben einer Führung auch ein Robotikworkshop angeboten. In diesem Rahmen dürfen die Mädchen einfache Programmieraufgaben mit Nao-Robotern lösen.

2.4 Abschluss der Camp Woche

Den Abschluss des Programms bildet eine kurze Präsentation, in der die Teilnehmerinnen ihren Eltern sowie weiteren Interessierten ihre Projekte und die Erlebnisse des Science Camps vorstellen können. Anschließend an die Präsentation werden bei Kaffee und Kuchen noch ungezwungen die Eindrücke der Woche ausgetauscht.

Nach der Camp-Woche können die Mädchen alle ihre Projekte und Unterlagen auf USB-Sticks mitnehmen. So haben sie die Möglichkeit ihre Webprojekte zu Hause weiter zu bearbeiten.

3 Auswertung

Eine Befragung der Teilnehmerinnen und deren Eltern ergaben, dass der Pilotdurchlauf des Science Camps Informatik von den Teilnehmenden gut angenommen wurde. Gemeinsam mit dem Lehrstuhl für Lehr-Lernforschung am KIT wird für den nächsten Durchgang eine fundierte Evaluation angestrebt.

Literaturverzeichnis

[Ti07] Tigges, A.: Geschlecht und Digitale Medien, Springer VS, S.72ff., 2007.

INFORMATIK – wie geil ist denn das!

Claudia Steinberger¹

Abstract: Dieser Beitrag präsentiert einen Mix von online und offline Maßnahmen, welcher in den letzten Jahren an der Fakultät für Technische Wissenschaften (TEWI) der Alpen-Adria-Universität Klagenfurt im Bereich Studierendenmarketing umgesetzt wurden. Ziel der Maßnahmen war es, das Image der Informatik bei SchülerInnen zu verbessern, die Begeisterung bei jungen Menschen an der Informatik zu wecken und die Zahl der StudienanfängerInnen an der Technischen Fakultät zu steigern.

1 Einleitung und Überblick

Informatik ist allgegenwärtig, ihre Anwendungsmöglichkeiten sind nahezu unerschöpflich und verändern unseren Alltag. Doch die Informatik hat in der breiten Öffentlichkeit einen schlechteren Ruf, als sie verdient. Viele verbinden das Bild von InformatikerInnen immer noch mit „Nerds im stillen Kämmerlein“. Die Zahl von StudienanfängerInnen in der Informatik stagniert, die Abbrecherquote ist hoch und der Frauenanteil gering. Dabei ist die Informatik eine höchst kreative Wissenschaft mit einer Fülle verschiedener Ansätze, die das allgemeine Leben nachhaltig beeinflussen.

Die Industrie, auch in Österreich, sucht dringend InformatikerInnen. Gefragt sind logisches Denken, kreative Lösungsstrategien und Teamarbeit. Die Zahl der StudienanfängerInnen und AbsolventInnen muss wieder wachsen, besonders in einer peripheren Region, die wie Kärnten seit Jahrzehnten unter einem ‚Braindrain‘, also der Abwanderung junger, gut gebildeter Menschen leidet.

Über 5300 KärntnerInnen ziehen im Schnitt pro Jahr in ein anderes Bundesland, das Binnenwanderungsdefizit beträgt gut 1000 Personen – pro Jahr. Vor allem die Jugend verlässt Kärnten: Zwei von drei sind zwischen 15 und 34. Sie zieht es vor allem fürs Studium in die Steiermark und nach Wien. Andererseits studieren in Relation wenige Nicht-KärntnerInnen in Kärnten. Laut einer Studie des IHS [IHS14] werden Kärnten bis 2030 an die 17.000 AkademikerInnen fehlen.

Daher war es der jungen Fakultät für Technische Wissenschaften (TEWI) in den letzten Jahren ein großes Anliegen, den Bekanntheitsgrad ihres einschlägigen Studienangebots Informatik, Informationsmanagement, Informationstechnik und Technische Mathematik in der Region aber auch überregional zu steigern, das Image dieser Studien zu verbessern

¹ Alpen-Adria-Universität Klagenfurt, Angewandte Informatik, Universitätsstraße 65-67, 9020 Klagenfurt, claudia.steinberger@aau.at

und so mittelfristig eine Erhöhung der AbsolventInnenzahlen zu erreichen. Zielgruppen waren besonders SchülerInnen (Sekundarstufe 1 und 2) sowie deren Eltern und LehrerInnen. Abbildung 1 zeigt einen Überblick der Maßnahmen, die sich besonders zur Erreichung dieses Ziels geeignet zeigten. Dabei wurde mit einem Mix aus online und offline Maßnahmen gearbeitet.

Maßnahmen	Wirkung	Zielgruppe
Technik LIVE	SchülerInnen erleben spannende Informatik-Themen in Workshops in Kleingruppen. SchülerInnen lernen den Campus kennen	Sekundarstufe 1 + 2 Transitiv: Eltern
Inf2school	Workshops für Informatik LehrerInnen machen Appetit auf neue Inhalte im Unterricht. Fördert Networking Schule/Uni	Informatik-LehrerInnen
TEWI Ferialpraktika	30-50 PraktikantInnen jährlich, Kennenlernen des Campus/Studiengebiets und Weitertransport von spannenden Themen im Freundeskreis	Sekundarstufe 2
Programmierwettbewerbe	Erreicht viele SchülerInnen mit Vorkenntnissen (HTLs)	Sekundarstufe 2
TEWI Studierendenportal	Fokussierte Info für Studieninteressierte	Sekundarstufe 2
TEWI on Social Media	LIVE Ticker über Aktivitäten an TEWI FB, Twitter, TEWI-Blog, Youtube	Alle
Messen / Tag der offenen Tür	Technik LIVE Bereiche mit Exponaten zum Ausprobieren	Sekundarstufe 1+2
Pressekooperationen	Top Themen in regionalen Medien	Eltern
Talentecamp, NaWi Tage, Semi Hightech	Für besonders interessierte SchülerInnen	Sekundarstufe 1
Lange Nacht der Forschung	Herausheben von Stationen der TEWI	Alle

Abb. 1: Maßnahmen im TEWI Studierenden-Marketing

Im Folgenden gehen wir auf zwei dieser Maßnahmen genauer ein.

2 Im Fokus: Technik LIVE

Einen Schwerpunkt bei den PR-Studierendenmaßnahmen bildeten Schülerveranstaltungen der Technik-LIVE Serie, die seit 2010 einmal jährlich für die Sekundarstufe 1 und einmal jährlich für die Sekundarstufe 2 veranstaltet wurden. Unter anderem wurden auch reine Mädchenworkshops organisiert. Unter dem Motto ‚Informatik, Mathematik und Informationstechnik zum Mitmachen‘ konnten pro Veranstaltung 150 bis 250 SchülerInnen aus bis zu 20 parallelen Workshops ihre Lieblingsthemen wählen und aktiv werden (siehe <http://technik-live.aau.at>; <http://technik-live-junior.aau.at>; <http://technik-live-girls.aau.at>).

WorkshopLeiterInnen kamen aus dem Kreis der ProfessorInnen, Universitäts-AssistentInnen, ProjektmitarbeiterInnen und Masterstudierenden, die ihre Forschungsthemen schülergerecht in „90 Minuten-Bissen“ aufbereiteten (siehe z.B. Abbildung 2). Die Veranstaltungen der Technik-LIVE Reihe waren nicht nur für 'Technik-Freaks' (z.B. von Technischen Schulen) gedacht, sondern hatten auch zum Ziel, bislang technikferne SchülerInnen zu begeistern. Es sollte insbesondere diesen SchülerInnen bewusst gemacht werden, wie viel Spaß technische Wissenschaften bereiten können und wie lebensnah diese sind. Workshops und Diskussionen wurden für unterschiedliche Vorkenntnisse angeboten (Levels L1, L2), so dass für alle SchülerInnen Interessantes zur Auswahl stand. Einige Workshops wurden bewusst parallel für Mädchen und Burschen angeboten.

- [Workshop 1: Kryptographie und Steganographie \(L1\)](#)
- [Workshop 2: Kryptographie leicht gemacht \(L1\)](#)
- [Workshop 3: Fox Security \(L2\)](#)
- [Workshop 4: Lego Mindstorm - easy going \(L1\)](#)
- [Workshop 5: Lego Mindstorm - sophisticated \(L2\)](#)
- [Workshop 6: Informatik Werkstatt\(L1\)](#)
- [Workshop 7: Zaubereien im Web auf der Spur \(L2\)](#)
- [Workshop 8: Spiele programmieren mit Gamedmaker \(L1\)](#)
- [Workshop 9: Mach mal was Lustiges: Gamification \(L1\)](#)
- [Workshop 10: Game Programming mit Scratch \(L1\)](#)
- [Workshop 11: Computerspiele von innen betrachtet \(L1\)](#)
- [Workshop 12: Finde den Videoclip! \(L1\)](#)
- [Workshop 13: Media-Sync: Synchronisiertes Crowd-Fernsehen \(L1\)](#)
- [Workshop 14: Photoshopen ohne Photoshop - Tricks in der Bildbearbeitung \(L2\)](#)
- [Workshop 15: Von Springern und Tierheimen: Warum Ordnung ganz schön schwer ist \(L1\)](#)
- [Workshop 16: Mehlkäfer, das Wetter, und was hat das mit Mathematik zu tun? \(L1\)](#)
- [Workshop 17: Intelligente Wohnumgebungen via Sprache steuern \(L1\)](#)
- [Workshop 18: Rasende Geschwindigkeit = rasender Puls \(L1\)](#)
- [Workshop 19: Zuerst Schüler\(in\) und dann Student\(in\) - wie passiert Studieren? \(L1\)](#)
- [Workshop 20: Zuerst Student\(in\) und dann Forscher\(in\) - wie passiert Wissenschaft \(L1\)](#)

Abb. 2: Workshopangebot der Technik LIVE 2015

Die Veranstaltungen der Technik-LIVE Serie wurden regelmäßig über eine eigene Homepage bekannt gemacht, Live Ticker über die Social Media Kanäle (FP, Twitter, TEWI Blog) verschickt und kurze Videos der Veranstaltungen über den TEWI Youtube Kanal bereitgestellt und auch von den Homepages der teilnehmenden Schulen auf deren Homepages verlinkt. Die jährlichen Evaluierungen von StudienanfängerInnen der AAU zeigten, dass diese Technik LIVE Veranstaltungen SchülerInnen für ein Studium an der AAU motivieren konnten.

3 Im Fokus: Ferialpraktika an der TEWI

Seit einigen Jahren werden an der TEWI Ferialpraktika angeboten, die interessierten SchülerInnen ab 16 Jahren einen 4-wöchigen Einblick in aktuelle Forschungsthemen bieten und die Möglichkeit geben, zu

- prüfen, ob ein IT-Studium für sie in Frage kommt und den IT Bereich entdecken.
- erleben, wie eine Universität funktioniert und Wissen/Forschung entsteht.
- spannende und vielfältige IT-Berufe hautnah und direkt miterleben.
- herausfinden wo ihre Stärken liegen und welche verborgenen Talente in ihnen schlummern.

Schon in der Bewerbungsphase im Februar werden InteressentInnen Projektthemen angeboten, auf die sie sich mit einem Motivationsschreiben fokussiert bewerben können (siehe <http://ferialpraktikum.aau.at>). 2016 haben sich beispielsweise über 120 Personen beworben. Neben den jährlich 30-45 ausgewählten SchülerInnen werden über deren Freundeskreis noch viele weitere Jugendliche mit Informationen über Standort, Forschungsthemen und Studienmöglichkeiten versorgt. Diese ‚Informationen von Freunden‘ stellen laut Erstsemestrigenumfrage eine wesentliche Grundlage für Studienwahlentscheidungen dar.

4 Hinter den Kulissen

Neben den Fakten über die an der TEWI durchgeführten Maßnahmen stellt sich natürlich auch die Frage nach notwendigen Organisationsstrukturen, dem Personal- und Finanzaufwand und dem Wirkungsgrad dieser Maßnahmen. Darüber soll im Rahmen der Präsentation im Workshop ebenfalls berichtet und diskutiert werden.

Literaturverzeichnis

[IHS14] IHSK: Brain Drain in Kärnten-Motive, Ursachen und mögliche Handlungsoptionen, 2014.

Erste Erfahrungen beim Einsatz von Unterrichtsbausteinen zur kooperativen Entwicklung und Erprobung von Informatikstunden an Grundschulen

Alexander Best¹, Uwe Thierschmann²

Abstract: Im Rahmen des Projekts „Informatik in der Grundschule (IGS)“ am Arbeitsbereich Didaktik der Informatik der Universität Münster werden zur Förderung informatischer Bildung im Primarbereich sogenannte Unterrichtsbausteine entwickelt. Diese sind auf einen zeitlichen Umfang von zwei bis vier Unterrichtsstunden à 45 Min. ausgelegt und bieten sowohl für Schülerinnen und Schüler³ als auch Lehrkräfte die Möglichkeit eines ersten Kennenlernens mit der Informatik. Da ein wesentliches Kriterium der Bausteine eine starke Verknüpfung zwischen Planung und Durchführung ist, wird bei deren Entwicklung früh, eng und kontinuierlich mit Lehrkräften kooperiert. In diesem Beitrag werden zunächst einige Ziele und Merkmale des Projekts IGS vorgestellt. Anschließend wird der Entwicklungs- und Erprobungsprozess anhand eines exemplarischen Bausteins zur Einführung in die Programmierung mit dem Bee-Bot in einer zweiten Klasse an einer nordrhein-westfälischen Grundschule skizziert.

Keywords: Informatik in der Grundschule; Primarbereich; Bee-Bot; Unterrichtsbausteine; IGS

1 Einleitung

Ein Blick auf aktuelle Tagungs- und Workshopthemen im Bereich der Schulinformatik zeigt deutlich: Informatik im Primarbereich boomt! Dabei handelt es sich keineswegs um ein neues Forschungsfeld. So widmete etwa die LOG IN 2003 der informatischen Bildung im Primarbereich ein Themenheft [KMP03]. Auch Andreas Schwill stellte 2001 in „Ab wann kann man mit Kindern Informatik machen?“ Überlegungen zur Thematik an [Sc01]. Doch wodurch ist das steigende Interesse in den letzten Jahren zu erklären?

Insbesondere die folgenden drei Entwicklungen haben auch außerhalb der Fachgemeinschaft das Interesse an informatischer Bildung im Primarbereich geweckt:

(1) Das Projekt „Computer Science Unplugged“, (2) die Entwicklung der grafischen Programmiersprache SCRATCH sowie (3) die Einführung des Pflichtfachs Computing für SuS im Alter zwischen 5/6 und 15/16 (*key stages 1-4*) in England.

Wie alle bisherigen Projekte zur Informatik im Primarbereich ist auch das Projekt IGS

¹ Westfälische Wilhelms-Universität Münster, Didaktik der Informatik, Fliegenerstr. 21, 48149 Münster, alexander.best@uni-muenster.de

² Lindenschule Gronau, Städtische Gemeinschaftsgrundschule, Sparenbergstraße 14, 48599 Gronau, uwe.thierschmann@gmx.de

³ Im Folgenden mit SuS abgekürzt.

durch diese Entwicklungen geprägt worden. Darüber hinaus werden jedoch auch Schwerpunkte gesetzt, die es von anderen Projekten abgrenzen. Einige sollen im Folgenden kurz vorgestellt werden.

2 Einige Ziele und Merkmale des Projekts IGS

Bereits im Rahmen von „Informatik im Kontext (IniK)“ wurde angemerkt, dass die „Einstellung der Lehrenden gegenüber der Neuerung [...] für die Umsetzung entscheidend“ sei [DKW11, S. 100]. Ein Ziel des Projekts IGS ist es deshalb, die Lehrkräfte bei der Entwicklung von Unterricht früh einzubinden. Hierfür war es nötig, ein geeignetes Format für den Austausch zwischen Grund- und Hochschule zu finden.

Das in vielen Projekten eingesetzte Modulkonzept stellte sich für die Zusammenarbeit als nicht zielführend heraus. Dafür scheinen zwei Ursachen maßgeblich zu sein. Zum einen sieht das Konzept vor, dass „Unterrichtsreihen als Sequenzen zusammengestellt werden“ [Hu02, Folie 7]. Eine Unterrichtssequenz oder gar -reihe hätte jedoch den zeitlich verfügbaren Rahmen gesprengt. Auch „Verzahnungsmöglichkeiten zwischen verschiedenen Modulen“ oder „modulübergreifende Fragestellungen“ [Hu02] konnten folglich nicht berücksichtigt werden. Des Weiteren wurde das Modulkonzept mit Blick auf Informatiklehrkräfte entwickelt. Deren fachliche und fachdidaktische Qualifikation fehlt den Grundschullehrkräften jedoch. In Gesprächen mit den Lehrkräften wurde der Wunsch nach einem Format geäußert, welches sich an den schriftlichen Unterrichtsplanungen der zweiten Lehramtsausbildungsphase orientiert.

Kooperativ wurde nun eine geeignete Vorlage für die Ausarbeitung gesucht, welches den genannten Anforderungen entsprach. Hierbei stellte sich „Fit von klein auf – Unterrichtsbausteine“ der BKK Gesundheitsförderung für Kitas und Grundschulen als geeignet heraus [BKK16].

Weiterhin zeichnet sich das Projekt IGS durch den zyklischen Entwicklungs- und Erprobungsprozess der Bausteine (Design Based) sowie das selbstständige Unterrichten dieser durch die Lehrkräfte aus. In Rahmen einer wissenschaftlichen Begleitforschung werden die Vorstellungen der Grundschullehrkräfte zur Informatik bzw. Informatikunterricht (*teachers' beliefs*) sowie messbare Veränderung durch das Unterrichten der Bausteine (*conceptual changes*) empirisch qualitativ erfasst.

3 Exemplarischer Entwicklungs- und Erprobungsprozess

3.1 Entwicklung des Bausteins „Programmieren lernen mit dem Bee-Bot“

In einem ersten Treffen mit der Lehrkraft *Uwe* wurde das Kernanliegen des Bausteins festgelegt: Die Einführung in Grundkonzepte der Programmierung. Die Grundkonzepte

sollten sich auf die Teillernziele (1) das Verständnis der Programmspeicherung sowie die (2) Umsetzung und (3) Beschreibung eines Programms in einer festgelegten Notation beschränken. Der zeitliche Rahmen wurde auf zwei Stunden à 45 Min. festgesetzt. Anschließend entschieden wir uns unter Berücksichtigung der Lerngruppe, die Bee-Bots⁴ als Werkzeug einzusetzen. Für die Erarbeitungsphasen wurde die Sozialform der Partnerarbeit gewählt, während die übrigen Unterrichtsphasen im Stuhlkreis stattfinden sollten. Es folgte die Ausgestaltung des Bausteins in zwei weiteren Treffen:

Die Matte, auf der sich der Bee-Bot bewegt, sollte lediglich mit farblichen Feldern ausgestaltet sein. Eine Verknüpfung mit Rechenoperationen oder Buchstabenrätseln würde die SuS in dieser Lerngruppe überfordern und möglicherweise vom eigentlichen Kernanliegen ablenken. Ein Startfeld sowie die Ausrichtung des Bee-Bots wurden vorgegeben, um Ergebnisse besser vergleichen zu können.

Der Einstieg sollte offen in einer Probierphase stattfinden. Die SuS konnten den Bee-Bot mit seinen Funktionen so zunächst kennenlernen. Uwe beobachtete intensiv diesen Prozess, um Erkenntnisse der SuS anschließend aufgreifen zu können. Insbesondere wurde antizipiert, dass die Programmspeicherung ein neues Konstrukt für die SuS darstellen würde. Bisherige Erfahrungen der SuS mit Werkzeugen funktionierten nach dem Prinzip direkter Reaktion auf eine Aktion. Bei der Programmsteuerung müssen Reaktionen zunächst antizipiert werden, bevor eine Aktion stattfinden kann. Zur Vermeidung eines zu starken Bruchs mit ihrer diesbezüglichen Erfahrungswelt, verzichteten wir auf die Verwendung der Begriffe „Speichern“ bzw. „Löschen“ und wählten die Alltagsbegriffe „Merken“ und „Vergessen“. Um eine freundschaftliche Bindung zum Bee-Bot zu fördern und mögliche Sprachbarrieren zu umgehen, wurde er „Friedolin“ genannt.

Im Verlauf der Stunde sollte das rein spielerische Ausprobieren des Bee-Bots in die gezielte Steuerung übergehen. Hierfür wurde beschlossen, dass Programme in Form von sogenannten Merktzetteln⁵ von den SuS festgehalten bzw. von der Lehrkraft vorgegeben werden sollten.

3.2 Erprobung des Bausteins „Programmieren lernen mit dem Bee-Bot“

In der Erprobung zeigte sich, dass die intensive Kooperation zwischen Grund- und Hochschule bei der Entwicklung des Bausteins einen sehr positiven Effekt hatte. Es fanden keine gravierenden Abweichungen zur Planung statt. Das zuvor antizipierte Verhalten der SuS erwies sich als weitestgehend zutreffend. In der Nachbesprechung wurden dennoch einige Änderungen an der Planung vorgenommen:

Der zeitliche Umfang der Stunde wurde auf zwei Doppelstunden à 90 Min. erweitert. Der Grund hierfür liegt im Erreichen des Teillernziels 1 (s.o.). In der Erprobung wurde

⁴ <http://www.tts-group.co.uk/bee-bot-rechargeable-floor-robot/1001794.html>, Stand: 07.06.2016

⁵ Diese Bezeichnung kann sich evtl. ändern.

erkennt, dass es sich hierbei eigentlich um 3 Teillernziele handelt, in denen die Programme der SuS zunehmend komplexer werden. (1a) Die SuS erkennen zunächst, dass sie nach dem Ausführen einer Bewegung den „Vergessen“-Knopf drücken können, um den Programmspeicher zu leeren. (1b) Sie planen anschließend eine Route, die sie mit dem Bee-Bot in mehreren Etappen ablaufen. Der „Vergessen“-Knopf wird weiterhin eingesetzt, jedoch nicht nach jeder Bewegung. (1c) Die SuS programmieren die gesamte Route ohne Einsatz des „Vergessen“-Knopfes.

Weiterhin wurde der Begriff „Matte“ in „Spielfeld“ umbenannt. Dies gibt den SuS einen deutlicheren Funktionsbezug. Zudem haben wir die vier Kanten des Spielfelds mit Symbolen versehen. Dies wird den SuS erleichtern, sich über die Blickrichtung des Bee-Bot auszutauschen (bspw. „Friedolin guckt zur Sonne“). Diese Änderungen wurden in die Unterrichtsplanung/-materialien eingearbeitet und werden demnächst erprobt.

4 Reflexion und Ausblick

Sowohl das Format der Unterrichtsbausteine als auch der Prozess der engen Kooperation zwischen Grund- und Hochschule erwiesen sich im Entwicklungs- und Erprobungsprozess als hilfreich. Neben der Entwicklung/Erprobung von neuen Bausteinen mit neuen Kooperationspartnern soll in Zukunft an der Systematisierung des Entwicklungs-/Erprobungsprozesses gearbeitet werden. Besonders der zeitliche Aufwand muss verkürzt werden. Weiterhin soll untersucht werden, ob eine zu starke Bindung eines Bausteins an eine spezifische Lerngruppe stattfindet.

Literaturverzeichnis

- [BKK16] Fit von klein auf. BKK Gesundheitsförderung für Kitas und Grundschulen, www.fitvonkleinauf.de, Stand: 11.05.2016.
- [DKW11] Diethelm, I.; Koubek, J.; Witten, H.: IniK – Informatik im Kontext. Entwicklungen, Merkmale und Perspektiven. LOG IN 169/170, S. 97-105, 2011.
- [KMP03] Koerber, B.; Müller, J.; Peters, I. R. (Hrsg.): Informatische Bildung im Primarbereich. LOG IN 121, LOG IN Verlag, 2003.
- [Hu02] Humbert, L.: Das Modulkonzept – ein zeitgemäßer Ansatz zur informatischen Bildung für alle Schülerinnen. In (Schwill, A. Hrsg.): *informatica didactica* 5, 2002, www.informaticadidactica.de/index.php?page=Humbert2002, Stand: 11.05.2016.
- [Sc01] Schwill, A.: Ab wann kann man mit Kindern Informatik machen? Eine Studie über informatische Fähigkeiten von Kindern. In (Keil-Slawik, R; Magenheimer, J. Hrsg.): *Informatikunterricht und Medienbildung. INFOS 2001. 9. GI-Fachtagung Informatik und Schule, 17.-20. September 2001 in Paderborn*. Köllen Verlag + Druck, Bonn, S. 13-30, 2001.

Informatisches Denken in der Primarstufe: die Wiener Zauberschule der Informatik (WIZIK)

Johann Stockinger¹, Gerald Futschek²

Abstract: Die Wiener Zauberschule der Informatik (WIZIK) führt Kinder der Primarstufe an die Denkweise der Informatik heran und vermittelt ihnen erste informatische Kompetenzen. Die Kinder lernen spielerisch verschiedene Problemlösungsstrategien kennen und erhalten einen ersten Einblick in die Grundlagen logischen und prozessorientierten Denkens.

Keywords: Informatik, Informatisches Denken, Computational Thinking, Primarstufe, Computer Science Unplugged, Biber der Informatik

1 Einleitung

Für das 21. Jahrhundert werden besondere Fähigkeiten und Kompetenzen benötigen. Dazu zählen Computerkenntnisse und besonders Kompetenzen im kreativ-innovativen Bereich, gekoppelt mit Problemlösungskompetenzen, siehe z.B. auch die Schlüsselkompetenzen der Europäischen Kommission [EK16]. Das von der Wirtschaftsagentur Wien in den Jahren 2014/15 geförderte Projekt WIZIK möchte diese Kompetenzen bei SchülerInnen der Primarstufe fördern und dabei Begeisterung für die Informatik wecken.

2 Entstehungsgeschichte

Die Österreichische Computer Gesellschaft OCG hatte in den Jahren 2012/13 das von der Wirtschaftsagentur Wien geförderte Projekt „ITAKE – die Informationstechnologie des Alltags für Kinder erfahrbar machen“ [OCG13] durchgeführt. Zu den Zielen zählte die Förderung der kreativen Fähigkeiten von Kindern als Voraussetzung für eine kreativ-innovative Gesellschaft. Bei der Umsetzung wurde auf einen spielerischen Zugang zur Robotertechnologie gesetzt. Im vorschulischen Bereich wurden dabei Bee-Bots eingesetzt (tastenprogrammierbare Bienenroboter) und in der Primarstufe in erster Linie das LEGO® WeDo™ Construction Kit (Version 1.0) in Kombination mit der kindergerechten Programmiersprache Scratch. Damit konnten neben der Steuerung der roboterartigen Gebilde zusätzlich auch Animationen auf dem Bildschirm gestaltet werden und so die virtuelle Welt mit der realen Außenwelt verknüpft werden.

¹ Österreichische Computer Gesellschaft, Wollzeile 1, 1010 Wien, johann.stockinger@ocg.at

² Technische Universität Wien, Institut für Softwaretechnik und Interaktive Systeme, Karlsplatz 13, 1040 Wien, gerald.futschek@tuwien.ac.at

Die in diesem Rahmen an den Schulen durchgeführten Workshops kamen sowohl bei den LehrerInnen als auch bei SchülerInnen sehr gut an. Allerdings ist bei derartigen Projekten die Nachhaltigkeit nicht immer zufriedenstellend. Wiener Volksschulklassen sind üblicherweise nur mit zwei PCs ausgestattet. Die Anschaffung und Wartung der Roboterbausätze für die Schulen ist nicht nur mit finanziellem Aufwand verbunden, sondern auch mit einem erheblichen Wartungsaufwand. Zusätzlich müssten die LehrerInnen entsprechend eingeschult werden.

Aus diesen Erfahrungen heraus entstand die Überlegung, nach Möglichkeiten zu suchen, das sogenannte „informatische Denken“ („Computational Thinking“) mit weniger Aufwand, dabei aber umso rascher zu vermitteln. Rein technologisch gesehen sollte dafür ein sehr niederschwelliger Zugang gefunden werden.

3 Die Umsetzung

Bald war klar, dass für Kinder eine attraktive Bezeichnung notwendig ist; eine „Zauberschule“ bot sich dafür an. Eine Recherche ergab, dass die Technische Universität Aachen schon ein derartiges Konzept unter der Bezeichnung „Zauberschule Informatik“ [BLS11] umgesetzt hatte. In Absprache mit den dortigen Initiatoren nannten wir unsere Initiative „Wiener Zauberschule der Informatik (WIZIK)“. Als Zielgruppe sind die Kinder der dritten und vierten Schulstufe vorgesehen.

Bei der Auswahl der Aktivitäten legten wir Wert darauf, dass möglichst viele Aktivitäten eine Koordination und Kooperation der Kinder erfordern. Auch sollten alle Aktivitäten einen direkten Bezug zur Informatik haben. Zudem sollten sie inhaltlich als Ergänzung und Bereicherung der Grundkompetenzen „Lesen, Schreiben, Rechnen“ dienen. Für die LehrerInnen sollten sich die Informatik Aktivitäten möglichst nahtlos in ihr bestehendes Lehrangebot integrieren lassen.

3.1 Aufgaben aus Computer Science Unplugged

Unter den Aktivitäten von Computer Science Unplugged [CSU15] finden sich einige, die auch schon in der Primärstufe gut vermittelt werden können. Als Einstieg eignen sich besonders gut die Punktekarten für die Darstellung der binären Zahlen und für das binäre Zählen. Die Kinder sind mit Begeisterung dabei und lernen dabei auch einiges über die Geschichte des Zählens und ein besseres Verständnis für das in der Schule gelernte Dezimalsystem [JG16a,JG16b]. Ergänzend dazu kann ein selbstgebasteltes Flip-Flop vorgeführt werden, bei dem durch eine wiederholt herablaufende Kugel automatisch binär gezählt wird.

Sortieralgorithmen können sehr gut mit Filmdosen umgesetzt werden, die z.B. mit Sand unterschiedlich gefüllt werden. Als Hilfsmittel bekommen die SchülerInnen nur einen Bleistift und ein Lineal und müssen damit eigenständig einen Lösungsweg finden, um

die Filmdosen nach dem Gewicht zu sortieren. Wichtig ist, dass ein Lösungsweg selbständig entdeckt wird, siehe auch [FM10]. Erst danach können je nach Situation auch die Grundideen weiterer Sortieralgorithmen besprochen werden. Eine Schulklasse hatte sogar eine eigene Variante des Bubblesort-Algorithmus getanzt, Video siehe [IV15].

Mit dem Beispiel des „armen Landkartenmalers“, der mit möglichst wenigen Farben Landkarten einfärben soll, kann das Vierfarben-Problem angerissen werden. Gleichzeitig wird die zeichnerische und malerische Kreativität gefördert [CSU15].

Natürlich darf auch der „Zaubertrick“ nicht fehlen, bei dem aus einer Anzahl aufgelegter Karten Fehler entdeckt und korrigiert werden können [CSU15].

3.2 Die Welt der Geheimcodes

Die Übermittlung von geheimen Botschaften übt für Kinder eine große Faszination aus. Dabei eignen sich geschichtliche Beispiele besonders gut, angefangen von einer Skytale über die Cäsar-Verschlüsselung bis hin zum Polybios-Quadrat, dem Freimaurer- und dem Kreuzcode. Verschlüsselungsverfahren siehe z.B. [CTO16].

3.3 Aufgaben aus dem Biber der Informatik Wettbewerb

Informatisches Denken ("Computational Thinking") wird seit 10 Jahren auch im internationalen Wettbewerb "Biber der Informatik" vermittelt, siehe [DF08]. Eine spezielle Kategorie von Aufgaben ist auch für die Primarstufe geeignet. Aus den Aufgaben der Vorjahre wurden einige herausgesucht, die sich auch gut in eine gebastelte, d.h. „begreifbare“ Version umwandeln lassen. Dazu gehörte z.B. ein Frage und Antwort-Spiel zum Auffinden eines gewünschten Strandbildes zu dem vergrößerte Bilder gebastelt wurden, die die Kinder angreifen und dazu eigene Fragestellungen formulieren können.

3.4 Durchführung

Alle Materialien zu den Beispielen befinden sich in „Zauberboxen“, von denen zehn Stück hergestellt wurden, sodass sie auch von interessierten LehrerInnen ausgeborgt werden können. Die Wiener Zauberschule der Informatik besuchte damit ca. 50 Wiener Volksschulen. Die Zauberboxen werden nach und nach mit weiteren interessanten Materialien gefüllt.

Aus Kapazitätsgründen konnten an den meisten Schulen nur einmalige Halbtagskurse abgehalten werden. Die Nettounterrichtszeit betrug dabei zwei Stunden. Um ein etwas besser abgerundetes Bild der Informatik zu erhalten, wären aber mindestens zwei derartige Workshops notwendig.

Natürlich musste auch das Äußere des jeweiligen „Informatikzauberers“ entsprechend gestaltet sein. Er trat mit Hut, Umhang und Zauberstab auf und konnte so die Aufmerksamkeit der Kinder sofort auf sich lenken.

3.5 Nachhaltigkeit

Aus den Rückmeldungen der LehrerInnen konnte entnommen werden, dass sie sehr interessiert daran sind, zumindest einige Aufgaben direkt in ihre eigene Unterrichtsgestaltung einzubinden. Das Projekt wurde auch an den beiden Wiener Pädagogischen Hochschulen vorgeführt. Interessensanfragen aus anderen Bundesländer liegen vor. Ein sehr ähnliches Ziel wie die Wiener Zauberschule der Informatik verfolgt das Forschungsprojekt „Informatik – ein Kinderspiel?!“ an der Alpen-Adria-Universität Klagenfurt [SPR14].

4 Literaturverzeichnis

- [BLS11] Bergner N., Leonhardt T., Schroeder U.: Zauberschule Informatik- Einblick in die Welt der Informatik für Kinder im Grundschulalter. Informatik mit Kopf, Herz und Hand - Praxisbeiträge zur INFOS 2011. Berlin, ZfL-Verlag, 2011. S. 132-141.
- [CSU15] Computer Science Unplugged, <http://csunplugged.org/books/>, 2015.
- [CTO16] CrypTool-Online, <http://www.cryptool-online.org>, Stand: 15.5.2016
- [DF08] Dagiene, V., Futschek, G.: Bebras international contest on informatics and computer literacy: Criteria for good tasks. In: Informatics Education - Supporting Computational Thinking. Springer Berlin Heidelberg, 2008. S. 19-30.
- [Eu16] Europäische Kommission. Schlüsselkompetenzen. http://ec.europa.eu/education/policy/school/competences_de.htm, Stand 15.05.2016.
- [FM10] Futschek G., Moschitz J.: Developing algorithmic thinking by inventing and playing algorithms. In: Constructionist Approaches to Creative Learning, Thinking and Education: Lessons for the 21st Century (Constructionism 2010), 2010.
- [Ga16a] Gallenbacher J.: Was ist binär? Zahlendarstellung im Binärsystem einfach erklärt, c't 7/2016, S. 144.
- [Ga16b] Gallenbacher J.: Malnehmen leichtgemacht. Multiplikation im Binärsystem mit Kindern ausprobieren, c't 10/2016, S. 158.
- [IV15] Informatiktag: Volksschultanz, <https://www.youtube.com/watch?v=WHdHq0aoqXU>, Stand: 15.05.2016.
- [OCG13] Projekt-Homepage ITAKE. 2013. <http://www.ocg.at/de/itake>, Stand: 15.05.2016.
- [SPR14] Sabitzer B., Pasterk S., Reçi E.: Informatics – A Child’s Play?!, Informatics--A Child's Play. In: Proceedings of the 6th International Conference on Education and New Learning Technologies (EDULEARN). 2014.

Anmerkungen zu einem begeisternden Informatikunterricht

Peter Micheuz¹

Abstract: Es darf als hohe Kunst der Didaktik angesehen werden, Schülerinnen und Schüler für ein Schulfach zu begeistern. Das mag in einigen Fällen und nicht selten unbeabsichtigt gelingen, ein Kochrezept dafür gibt es nicht, auch nicht im Fach Informatik. Möglicherweise ist es pragmatischer, sich über einen Informatikunterricht und attraktive Inhalte Gedanken zu machen, der bei den Lernenden, wenn schon nicht Begeisterungstürme hervorruft, wenigstens temporäres Interesse weckt. Neben persönlichen Erfahrungen im Informatikunterricht werden zwei Beispiele zum Thema Zaubertricks kurz beschrieben. Diese haben bei den Schülerinnen und Schülern nach einer Phase des Staunens jenes Interesse geweckt, das sie zu forschend-entdeckendem Unterricht und zur eigenständigen Bearbeitung motivieren sollte.

Keywords: Informatik, Unterricht, Motivation, Interesse, Zaubertrick.

1 Einleitung

Die Frage, wie Menschen im Allgemeinen und Lernende im Besonderen für eine Sache begeistert werden können, ist eine Domäne der Motivationsforschung. Schülerinnen und Schüler für ein Fach zu begeistern, ist eine hohe Kunst und keinesfalls alltägliches Unterrichtshandwerk. Es gibt kein Patentrezept und keine Planbarkeit für begeisternden (Informatik)Unterricht. Sehr wohl aber können extrinsische Faktoren benannt werden, die mit großer Wahrscheinlichkeit bei den Lernenden keine Begeisterung auslösen. Dies ist dann der Fall, wenn wesentliche notwendige Kriterien für einen gelingenden Unterricht, über die es weitestgehend einen empirisch abgesicherten Konsens gibt, nicht erfüllt werden. Konterkariert wird diese These allerdings dadurch, dass das Interesse und im besten Fall auch Begeisterung nicht nur, aber auch für das Fach Informatik intrinsisch begründet sind. Dabei kann der Auslöser von einer vom Fach begeisterten Lehrkraft kommen, muss aber nicht.

Möglicherweise mehr als in anderen Fächern gibt es in Informatik das (intrinsische) Bestreben, etwas um seiner selbst willen zu tun, weil die Sache begeistert, Spaß macht oder herausfordert. Bei der extrinsischen Motivation steht dagegen der Wunsch im Vordergrund, bestimmte Leistungen zu erbringen, weil man sich davon Vorteile wie die Vorbereitung auf ein Studium, das Erreichen beruflicher Ziele oder gute Noten im (verpflichtenden) Informatikunterricht verspricht.

¹ Alpen-Adria-Universität Klagenfurt, Institut für Informatik Fachdidaktik, Universitätsstraße 65-67,
9020 Klagenfurt, peter.micheuz@aau.at

„Wie haben Sie es nur geschafft, meinen Sohn so für Informatik zu begeistern? Er sitzt jetzt Tag und Nacht nur mehr vor dem Computer.“ Diese - fiktive, aber möglicherweise bereits in ähnlicher Form und manchmal getätigte - Aussage an einem Elternsprechtag ist in hohem Maße interpretationsoffen und wirft einige Fragen auf: Wie alt ist der besagte Schüler? Ist Informatik für ihn ein Pflicht-/Wahl- oder Freifach? Auf wie viele Schüler in der gleichen Klasse trifft dies zu? Was ist der wahre Auslöser der Begeisterung? Ist es der Informatiklehrer, der inspirierenden Unterricht erteilt? Oder sind es die attraktiven Informatikthemen? Oder ist den Eltern nicht klar, dass ihr Kind einer Computerspielsucht anheimgefallen ist und sich gar nicht mit Informatik beschäftigt? Was heißt in diesem Zusammenhang überhaupt „sich mit Informatik beschäftigen“?

Auf jeden Fall darf das Thema „Informatikunterricht und Motivation“ nach wie vor als ein interdisziplinäres und vielschichtiges Forschungsdesiderat bezeichnet werden. Es finden sich zwar vereinzelt theoretische Abhandlungen darüber, jedoch sind dem Autor noch keine diesbezüglich empirisch abgesicherten Studien und Aktionsforschungen bekannt. Gespräche mit Kollegen und Selbsterfahrungen belegen, dass hochmotivierte und begeisterte Schüler im Informatikunterricht heute mehr als früher die Ausnahme und nicht die Regel darstellen. Die notwendigen Bedingungsfaktoren für „begeisterten“ Unterricht wie (subjektiv) interessante Aufgabenstellungen, gruppendynamische Prozesse, ein kompetitives, leistungsmotivierendes Klima in der Lerngruppe oder einfach nur intrinsische Motivation können nur situativ beurteilt und beforscht werden.

2 Beobachtungen und eigene Erfahrungen

Jede Lehrkraft, die über einen sehr langen Zeitraum Informatik (vor 1985 hieß das Fach in Österreich Elektronische Datenverarbeitung) unterrichtet hat, kann subjektiv über unterschiedliche Erfahrungen berichten und sich als Teil einer retrospektiven Aktionsforschung (des eigenen Unterrichts) sehen. In meinem speziellen Fall reicht die Pionierzeit bis in das Jahr 1981 zurück, als ich mich als begeisterter EDV-Lehrer zusammen mit 16-jährigen begeisterten Schülerinnen und Schülern an einem ganz neuen Fach versuchte. Rückblickend darf dieses Langzeitprojekt im Rahmen eines Schulversuches im Angesicht der heutigen erdrückenden Fülle an informatischen Lehr- und Lerninhalten als ein kleines Wunder gesehen werden, weil damals mit so wenig Soft- und Hardware der Spannungsbogen drei Schuljahre bis zur ersten mündlichen EDV-Matura aufrecht erhalten werden konnte. Der Pioniergeist machte es damals auch ohne Scratch, Informatik-Biber, Roboter und Drohnen möglich. Ich behaupte, dass die Attraktivität und der „Begeisterungsindex“ relativ zeitstabil sind. Die These, die neuen Möglichkeiten zur Attraktivierung des Informatikunterrichts vermögen besser zu begeistern und bei vielen Schülerinnen und Schülern nachhaltiges, extrinsisch induziertes Interesse zu wecken, muss wissenschaftlich erst geprüft werden.

Der Informatikunterricht, der durch aktuelle Angebote und Werkzeuge (blockbasierte

Sprachen, mächtige Frameworks, CS Unplugged Aktivitäten, Maker-Szene, Physical Computing inklusive Robotik) ein überbordendes Angebot an Möglichkeiten bereitstellen kann, garantiert per se noch kein anhaltendes Interesse. Dazu bedarf es beispielsweise im Falle des Physical Computing neben ausstattungsmäßigen Rahmenbedingungen (Alltagstauglichkeit der notwendigen Devices) geeigneter personeller Voraussetzungen seitens der Lehrkräfte (Ausbildung und/oder viel intrinsische Motivation) und Schülerinnen und Schüler, die nicht nur kurzfristig begeistert sind, sondern sich anhaltend und motiviert mit diversen Themen inhaltlichen auseinandersetzen. Bei all dem Überangebot an vermeintlich attraktiven Lerninhalten müssen auch entwicklungspsychologische (altersgemäße) und gruppensdynamische Überlegungen berücksichtigt werden. Unterrichtsarbeit ist auch Beziehungsarbeit, und Lernprozesse sind stark beeinflusst durch soziale Interaktionen. Schließlich kennt jede länger im Beruf stehende Lehrkraft das Diversitäts-Phänomen auf Klassen- und Schülersdrebene nur allzu gut. Neben Informatikgruppen mit günstigstenfalls mehrheitlich interessierten Schülerinnen und Schülern gibt es natürlich auch weitestgehend begeisterungsresistente Klassen.

Eine Informatikgruppe zu Beginn des 3. Jahrtausends ist mir in bester Erinnerung, da ich sie im Rahmen eines österreichischen Informatik-Bewerbes für ein Internet-Projekt begeistern und in einen Flow versetzen konnte. Dazu hat die unscharfe Aufgabenstellung, die Schülersvertretungswahl als Webservice für die Schulen österreichweit anzubieten, zusammen mit der Erwähnung (noch in den Anfängen befindlicher) webbasierter Sprachwerkzeuge ausgereicht. Alles andere ist Schulgeschichte. Dieses Projekt „Cybervoting“ wurde mit zwei Preisen ausgezeichnet und anlässlich einer IFIP-Konferenz 2002 in Dortmund auch einem internationalen Publikum präsentiert.

3 Können diese Beispiele begeistern?

3.1 Eine wundersame Schokoladeproduktion



Abb. 1: Schokoladestiegel-Vermehrung: Aus 20 werden 20 + 1 Stücke

Eine verblüffende Animation im Internet, die die Betrachter in Staunen versetzen soll, kann zum Ausgangspunkt genommen werden, den Entdeckergeist der Schülerinnen und Schüler zu wecken, und sie anschließend mit der Aufgabe zu betrauen, diese Animation mit Geldscheinen und mit einer anderen Web-Technik als mit einer animierten GIF-Grafik nachzubauen. Zum Zeitpunkt der Vorstellung dieser Aufgabe in einer Informatikgruppe (10. Jahrgangsstufe) wurde trotz intensiver Google-Suche keine

Entzauberung im Internet gefunden. Es blieb einem interessierten (oder gar begeisterten?) Team vorbehalten, dieses Phänomen zunächst zu analysieren und anschließend eine webbasierte Lösung einer Banknotenvermehrungsmaschine (bevorzugterweise mit 500 Euro Banknoten) zu realisieren. Der hoffentlich bisher begeisterte Leser dieses Beitrages ist herzlich eingeladen, dies dem Team nachzumachen.

3.2 Ein Zahlentrick mit präparierten Kärtchen

Ein weiterer verblüffender (oder gar begeisternder?) Trick wurde von einem anderen Team im Rahmen desselben Unterrichtsprojekts dankend (und begeistert?) angenommen, analysiert und entzaubert. Das „Trickopfer“ merkt sich eine Zahl zwischen 1 und 60 und sagt dem „Trickkünstler“ ob sie auf den Karten A bis F vorkommt. Die Antwort kann schnell gegeben werden, weil lediglich die Zahlen in den linken oberen Ecken zusammengezählt werden müssen.

Karte A	Karte B	Karte C
16 21 26 31 52 57	1 11 21 31 41 51	4 13 22 31 44 53
17 22 27 48 53 58	3 13 23 33 43 53	5 14 23 36 45 54
18 23 28 49 54 59	5 15 25 35 45 55	6 15 28 37 46 55
19 24 29 50 55 60	7 17 27 37 47 57	7 20 29 38 47 60
20 25 30 51 56	9 19 29 39 49 59	12 21 30 39 52

Karte D	Karte E	Karte F
8 13 26 31 44 57	2 11 22 31 42 51	32 37 42 47 52 57
9 14 27 40 45 58	3 14 23 34 43 54	33 38 43 48 53 58
10 15 28 41 46 59	6 15 26 35 46 55	34 39 44 49 54 59
11 24 29 42 47 60	7 18 27 38 47 58	35 40 45 50 55 60
12 25 30 43 56	10 19 30 39 50 59	36 41 46 51 56



Abb. 2: Vorbereitete Kärtchen und Demonstration anlässlich eines Tages der offenen Tür

Der intellektuelle Reiz dieser Aufgabe liegt weniger auf phänomenologischer Ebene, sondern in der (allenfalls automatischen) Produktion von Kärtchen in anderen Zahlenräumen, z.B. von 1 bis 32. Für diese Herausforderung ist das Verständnis von binären Zahlen unerlässlich. Dadurch ist der informatische Bildungswert dieser Aufgabe als hoch einzustufen. Im Unterschied zur „Schokoladenfabrik“ ist dieser Trick im Internet in unterschiedlichen Darstellungen präsent.

3.3 Schlussanmerkungen

Beide Tricks wurden neben anderen Stationen (Parity-Bit-Fehlerkorrektur etc.) im Rahmen eines Tages der offenen Tür 2015 staunenden Eltern und möglicherweise auch begeisterten Kindern aus den umliegenden Grundschulen vorgestellt. Tage der offenen Tür haben den Ruf, dass möglicherweise oft „Feiertagsdidaktik“ und zu viel schulalltagsferne Show betrieben wird. Mag das in anderen Fächern der Fall sein. Was die Schulinformatik anlangt, wurden - in der Theorie jedenfalls - nur alltagstaugliche Informatikthemen vorgestellt. Ob diese später im Informatikunterricht in der einen oder

anderen Form mit Begeisterung aufgenommen werden, hängt wohl (auch) von den betroffenen Akteuren ab. Von den Schülerinnen und Schülern möglicherweise mehr als von den Lehrkräften. Oder ist es doch umgekehrt?

Programming for Refugees – An Active Learning Approach for Teaching Java to Heterogeneous Groups

Niklas Kühl¹ and Jonas Lehner²

Keywords: Teaching, Programming, Refugee Crisis, Europe, Integration

1 Introduction

The refugee crisis was one of the main challenges of Europe in the last year. It forced millions of people from the middle-east to flee their homeland due to conflict, poverty or substantial threat and migrate to Europe. Germany played an important role in welcoming 475,000 refugees in 2015 [Bu16]—and faced huge political and logistical challenges associated with it. The goal was to enable a fast and successful integration of the newly arriving people, regardless of age, gender and standard of knowledge. Education is one of the key success factors of such an integration [Ro99, Ko03, SW07]. The paper at hand illustrates an initiative to allow refugees to make first steps in the fields of computer science and programming, loose potential fear of contact with IT and learn a valuable, sought-after new skill. In teaching *Java*, we address the most popular programming language [TI16] and deliver an entrance to object-oriented and modern approaches. In the resulting paper we describe two different iterations (2, 3), their challenges as well as solutions and lessons learned. Finally, we conclude and deliver an outlook for future activities (4).

2 First Iteration

Because of the momentousness of the refugee crisis in September 2015, we identified — amongst others—the three immediate challenges which refugees were facing. First, they were not able to work upon arrival since they had to wait to be processed (which could take up to six months). Second, the refugees were willing to integrate, but were excluded of most common activities due to a lack of money and language skills. Third, they could not commit to long term programs since they needed to stay flexible for administrative interactions.

In order to address those challenges in an efficient way, we tried to offer a first iteration of a programming course as fast as possible. We decided to reuse adapted material from

¹ Karlsruhe Institute of Technology (KIT), Karlsruhe Service Research Institute (KSRI), Englerstr. 11, 76131 Karlsruhe, Germany, kuehl@kit.edu

² Karlsruhe Institute of Technology (KIT), Institute of Applied Informatics and Formal Description Methods (AIFB), Kaiserstr. 89, 76133 Karlsruhe, Germany, jonas.lehner@kit.edu

the Java programming lectures for Business Engineers from the soon starting winter term, which is based on Ratz et al. [Ra14]. The material (slides and exercises) had a scientific and theoretical background for university students with a high GPA. We offered a weekly course starting in September 2015 with four hours per session and a total of four sessions. Each session was separated into lecture and exercise phases, usually alternating to teach new content first and then let participants implement it respectively. The exercises took place in the large computer labs of the university. We witnessed a fast decrease in the number of participants on the one hand (c.f. Figure 3) and had to frequently repeat basic content in the sessions 2 to 4 on the other.

3 Second Iteration

While our focus in the beginning was to come up with a working concept, we reconsidered the basic concept after conducting the course twice.

Organizational Since most refugees are regularly transferred to other camps, which impedes long-term planning for the participants, we adjusted the duration of the course to three weeks while increasing the number of sessions to two per week. We noticed a significant lack of attentiveness towards the end of each session and hence shortened each session to two hours. Furthermore we reduced the group size to about 15 people to improve the cohesion.

Content Considering the heterogeneous target group—some of the participants never used a computer before—we decided to radically cut the amount of information and focus on very basic topics. Unlike in the regular university course we have no pressure to cover a certain content which enables us to move on in a slower pace, allowing us to take care of individual skills.

Teaching Method After we did frontal teaching in the first iteration of the course we identified a lack of motivation within the participants. While this kind of teaching style works well for some topics in university, we recognized that the refugees needs were different: The separation of lecture and exercise phases was not effective. To avoid this we now use the concept of *active learning*, which Silberman describes as: “Above all, students need to ‘do it’—figure things out by themselves, come up with examples, try out skills, and do assignments that depend on the knowledge they already have or must acquire.”[Si96] To implement this teaching method we adopted successful concepts from websites like “Rails for Zombies³” and “Codecademy⁴” and followed the steps described by Hazzan et al. [HLR15]:

³ <http://railsforzombies.org/>, last received 12-05-2016

⁴ <https://www.codecademy.com/>, last received 12-05-2016

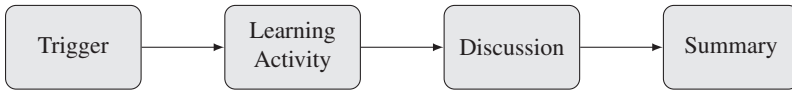


Fig. 1: Active-learning based teaching model [HLR15].

At the beginning of each session a *trigger* is presented which in our case consists of a simple working Java program that solves a certain problem. As *learning activity* the participants are confronted with the source code of the program. The teacher explains each command and encourages the learners to experiment with the code. He also provides simple tasks within the given source code. While performing the learning activity, the refugees work in small groups of 2-3. After about 80 minutes the participants present their results which leads to a *discussion* about the presented material. Each session is concluded by a summary of the covered topics given by the instructor. As Hazzan et. al propose, this summary can be designed in different ways, such as a poster, a so called cheat sheet (like in our case), a mind map, or else.



Fig. 2: Frontal teaching during first iteration (left) vs. team work (active learning) during second iteration (right).

Trough the improved concept, the role of the instructor shifted from a lecturer presenting information to a supportive and encouraging tutor who motivates the participants to explore and experiment by themselves.

4 Conclusion

Due to our feedback analysis, the course was taken as a valuable offer by the refugees. The interest of the refugees was aroused and they were thankful for the opportunity. One of the refugees was later able to successfully apply to an internship at SAP. For all of them, the idea of studying at a university was of higher interest than before and we got asked frequently on what their academic possibilities might be. As witnessed in the second iteration, the number of participants was more constant than before (c.f. Fig. 3). Besides the changing of the teaching method, there are other possible reasons for the decreasing number of participants in the first iteration. For instance, they could have moved to a different location, which made a commute too far. Unfortunately, we were not able to gain any information on why previous participants discontinued to attend. Nonetheless, our con-

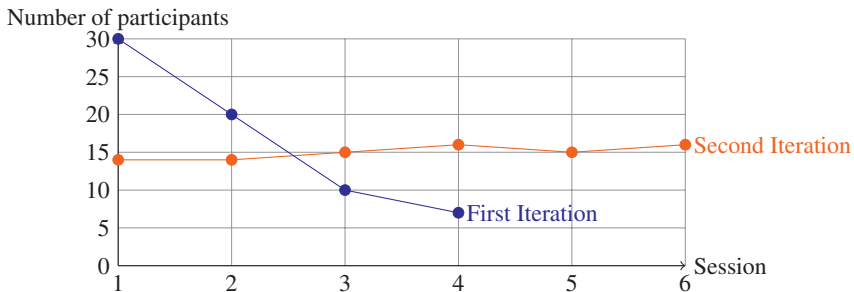


Fig. 3: Number of participants in both iterations

stant number of participants in the second iteration can lead to the conclusion that the new teaching method was more suitable and motivating.

To integrate the experiences which we gained during the first two iterations and to further promote the concept we founded a non-profit association called “EduRef – Education for Refugees e.V.” by the end of 2015. More courses will be held in 2016 covering further topics in the field of computer science and others.

References

- [Bu16] Bundesamt fuer Migration und Fluechtlinge (BAMF): , Aktuelle Zahlen zu Asyl. <https://www.bamf.de/SharedDocs/Anlagen/DE/Downloads/Infothek/Statistik/Asyl/statistik-anlage-teil-4-aktuelle-zahlen-zu-asyl.pdf>, 2016. Last received 12-05-2016.
- [HLR15] Hazzan, Orit; Lapidot, Tami; Ragonis, Noa: Guide to teaching computer science: an activity-based approach. Springer, 2015.
- [Ko03] Korac, Maja: Integration and How We Facilitate it A Comparative Study of the Settlement Experiences of Refugees in Italy and the Netherlands. *Sociology*, 37(1):51–68, 2003.
- [Ra14] Ratz, Dietmar; Scheffler, Jens; Seese, Detlef; Wiesenberger, Jan: Grundkurs Programmieren in Java. Carl Hanser Verlag GmbH Co KG, 2014.
- [Ro99] Rooth, Dan-Olof: Refugee Immigrants in Sweden - educational investments and labour market integration. Dissertation, Lund University, 1999.
- [Si96] Silberman, Mel: Active Learning: 101 Strategies to Teach Any Subject. Pearson Higher Education, Boston, 1996.
- [SW07] Stevenson, Jacqueline; Willott, John: The aspiration and access to higher education of teenage refugees in the UK. *Compare: A Journal of Comparative and International Education*, 37(5):671–687, 2007.
- [TI16] TIOBE Index for May 2016. http://www.tiobe.com/tiobe_index, 2016. Last received 12-05-2016.

Informatik-Werkstatt: Entwicklungen und Erfahrungen einer Lern- und Lehrwerkstatt für Informatik

Stefan Pasterk¹, Heike Demarle-Meusel¹, Barbara Sabitzer¹ und Andreas Bollin¹

Abstract: Neue Technologien und im Speziellen die Informatik haben in der heutigen Gesellschaft einen starken Einfluss auf das tägliche Leben. Damit wird die Informatik auch für Schulen und den Unterricht immer wichtiger. Um Kindern jeden Alters verschiedene Themen der Informatik näher zu bringen wurde im Sparkling Science Projekt „Informatik – Ein Kinderspiel!“ die „Informatik-Werkstatt“ ins Leben gerufen. Durch verschiedene mobile sowie stationäre Angebote wird es interessierten Kindern ab vier Jahren und ihren Eltern, Jugendlichen, sowie Schulklassen und ihren LehrerInnen ermöglicht, Einblicke in viele Bereiche der Informatik wie Verschlüsselung, Modellierung, Algorithmen, Programmierung oder Netzwerke zu erhalten und dafür zu begeistern. Dieser Artikel gibt einen Überblick über die Entwicklungen und Angebote der Informatik-Werkstatt, das Evaluationskonzept, und berichtet über die gesammelten Erfahrungen.

Keywords: Lernwerkstatt, Lehrwerkstatt, Informatik, Volksschule

1 Einleitung

Dadurch, dass die Informatik immer größere Auswirkungen auf das tägliche Leben in der heutigen Gesellschaft hat, wächst ihre Rolle auch in der Ausbildung. In Österreich ist Informatik jedoch nur in der 9. Schulstufe als Pflichtfach vorgesehen – oft zu spät um noch Interesse an Technik wecken zu können. Daher bieten viele Schulen Informatik schon in der Sekundarstufe I oder als Wahlfach an, meist jedoch auf die Nutzung digitaler Technologien oder Programmierung reduziert. Das Regionale Fachdidaktik-Zentrum (RFDZ) für Informatik in Kärnten versucht daher, Kinder und Jugendliche durch verschiedene Projekte schon früh für die vielfältigen Themen der Informatik zu begeistern. Dazu wurde im Verlauf des Sparkling Science Projekts „Informatik – Ein Kinderspiel?!“ [SPR14] eine „Informatik-Werkstatt“² eröffnet, in der Kinder und Jugendliche aber auch interessierte Erwachsene verschiedene Bereiche und Konzepte der Informatik entdecken und sich mit den einen oder anderen Inhalten intensiver befassen können. Von Beginn an wurde das Angebot laufend weiter entwickelt und Erfahrungsberichte aller Beteiligten eingeholt, evaluiert und in der Planung der Werkstatt berücksichtigt. Im folgenden Abschnitt wird die Entwicklung von einem mobilen Angebot über eine Sommerwerkstatt hin zu einer wöchentlichen Lernwerkstatt beschrieben. Abschnitt 3 enthält

¹ Alpen-Adria-Universität Klagenfurt, Institut für Informatikdidaktik, Universitätsstraße 65-67, 9020 Klagenfurt, {Vorname.Nachname}@aau.at.

² <http://informatikwerkstatt.aau.at>

Erfahrungsberichte und Analysen zu Beobachtungen der Informatik-Werkstatt, und Abschnitt 4 fasst die Ergebnisse und Schlussfolgerungen zusammen.

2 Organisation und Ausbau der Informatik-Werkstatt

Die Informatik-Werkstatt durchlief mehrere Phasen der Entwicklung und ist seither im Entwicklungsplan der Universität verankert. Sie begann als eine mobile Variante, wurde 2014 um die Sommerwerkstatt erweitert und findet seit Beginn des Schuljahres 2015/2016 zusätzlich einmal wöchentlich statt. Diese drei Entwicklungsstadien werden hier kurz erläutert.

Die mobile Werkstatt wurde im Jahr 2013 als Angebot für Schulen ins Leben gerufen. Für Interessierte standen verschiedene Themen wie Verschlüsselung, Binärzahlen, Netzwerke, Sortieren, und Hardware zur Wahl, deren Aufbereitung zum Teil auf anderen Projekten wie „Computer Science Unplugged“ [Be09] und „Informatik erLeben“ [MBH10] beruhte. Die Vermittlung der Inhalte wurde dabei an die Kriterien des Lehrkonzepts „COOL Informatics“ [Sa14] angepasst. Dieses Konzept basiert auf neurodidaktischen Erkenntnissen und hebt die vier Prinzipien „Entdecken (discovery)“, „Kooperation (cooperation)“, „Individualität (individuality)“ und „Aktivität (activity)“ hervor.

Die Sommerwerkstatt wurde erstmals im Juli 2014 an der Alpen-Adria-Universität Klagenfurt durchgeführt. Dabei konnten Kinder, Jugendliche, Eltern und LehrerInnen verschiedene Bereiche der Informatik an mehreren thematischen Stationen erkunden. MitarbeiterInnen des Instituts für Informatikdidaktik und sieben PraktikantInnen zwischen 16 und 19 Jahren führten die insgesamt 76 BesucherInnen durch die Stationen und arbeiteten neues Unterrichtsmaterial aus [PA16]. Im Jahr 2015 wurde die Sommerwerkstatt inhaltlich (z.B. um Roboter) erweitert. Die insgesamt 312 TeilnehmerInnen wurden von 13 PraktikantInnen im Alter von 16 bis 19 Jahren begleitet. Zusätzliche Thementage wie Girls-/Boys Days oder 3-D Druck machten das Programm noch spannender [PA16].

Die Jahreswerkstatt mit monatlichen Schwerpunkten ist seit Beginn des Schuljahres 2015/16 an einem Nachmittag pro Woche geöffnet. Die BesucherInnen werden dabei von ProjektmitarbeiterInnen sowie Informatik- und Lehramtsstudierenden begleitet.

Für 2016 sind bereits weitere Ausbaustufen geplant. So findet im Rahmen der Sommerwerkstatt nun auch ein einwöchiges „Informatik-Camp“ statt, in dem die vormittags erlernten Informatikkonzepte nachmittags in der *Kinderbetreuung* spielerisch aufgearbeitet werden. Langfristig wird die Informatik-Werkstatt in die offene Lern- und Lehrwerkstatt der *School of Education (SoE)* der Universität Klagenfurt integriert werden.

3 Evaluation

3.1 Evaluationskonzept und Forschungsinteresse

Ein Hauptanliegen der Informatik-Werkstatt ist es, das Interesse an Informatik zu steigern und informatisches Denken bewusst zu machen. Die Fragestellungen (siehe Tab. 1) konzentrieren sich auf die Bewertung der Angebote und die dadurch erzielten Outcomes.

Fragestellung	Indikatoren	Erhebungsmethode
Wie werden die Angebote bewertet?	Bewertung der Stationen und Materialien; Bewertung der Inhalte	Fragebogen; Interviews; SWOT-Analyse
Kann das Interesse an Informatik durch die Angebote gesteigert werden?	Prä-Post Vergleiche; Erreichen unentschlossener oder nicht interessierter Personengruppen; Anwendung informatischer Inhalte im Unterricht	Fragebogen; Einzelfallanalyse;
Kann informatisches Wissen aufgebaut werden?	Wissenszuwachs; Transfer des Wissens in Produkte	Selbstkonstruierte Wissensüberprüfung; Entstandene Produkte; Beobachtung
Welche Prinzipien des Lehrkonzepts sind besonders wirksam, um Informatik zu verstehen?	Anwendung Prinzipien (Discovery, Cooperation, Individuality, Activity)	Fragebogen Interviews Beobachtung

Tab. 1: Überblick über das Evaluationskonzept und die zugrundeliegenden Fragestellungen.

Die Datenerhebung für die Evaluation findet in allen drei Settings statt. Tab. 1 gibt zusätzlich einen Überblick über die Hauptfragestellungen der Evaluation. Je nach Forschungsinteresse werden in den verschiedenen Settings der Werkstatt unterschiedliche Schwerpunkte gelegt. Die Daten werden im Sinne der Triangulation aus der Perspektive der drei Zielgruppen (SchülerInnen, LehrerInnen, PraktikantInnen) erhoben. Es werden sowohl quantitative als auch qualitative Verfahren zur Datenerhebung angewendet. Als Erhebungsmethoden werden Fragebögen, Interviews und Beobachtungen eingesetzt.

3.2 Ergebnisse

Einige Ergebnisse aus der Sommerwerkstatt wurden bereits publiziert [Pa16]. An dieser Stelle gehen wir auf die Sichtweisen der drei verschiedenen Zielgruppen ein.

Von den *LehrerInnen* werden die Angebote sehr positiv bewertet und als eine Bereicherung für den eigenen Unterricht wahrgenommen. Sie berichten auch davon, dass ein Wissenszuwachs sowohl bei den SchülerInnen aber auch bei den LehrerInnen selbst beobachtet werden konnte. Die Interviews in den Partnerschulen zeigen aber auch, dass

die Bedürfnisse nach Art und Umfang der Unterstützung abhängig sind vom Schultyp bzw. dem Unterrichtsfach (Informatik vs. andere). Die mobilen Angebote werden sehr gerne angenommen, jedoch die eigenverantwortliche Einbindung informatischer Inhalte in den Unterricht funktionierte suboptimal. Daher wurde für die Partnerschulen ein Kinder-Kongress ins Leben gerufen, bei dem LehrerInnen gemeinsam mit ihren SchülerInnen Konzepte im Unterricht bearbeiten und auf dem Kongress präsentieren.

Aus *SchülerInnensicht* werden die Angebote der Informatik-Werkstatt sehr positiv bewertet. Die einzelnen Stationen sind ansprechend und die Inhalte werden verständlich vermittelt. Eigenen Angaben zufolge schätzen die TeilnehmerInnen ihr Wissen in Informatik nach dem Besuch der Informatik-Werkstatt höher ein als zuvor. Auch beim Interesse konnten Zuwächse festgestellt werden. Beim Lernen in der Sommerwerkstatt wurden insbesondere die Unterstützung der TutorInnen und das individuelle Lernen als hilfreich eingestuft.

Auch aus *Studierenden-* und *PraktikantInnensicht* wurde die Informatik-Werkstatt sehr positiv bewertet. PraktikantInnen konnten dadurch sogar für ein Informatikstudium motiviert werden. Nach Einschätzung der PraktikantInnen haben den TeilnehmerInnen beim Verständnis der informatischen Inhalte besonders die Erklärungen und die kindgerechten Materialien geholfen. Bezogen auf die vier COOL Prinzipien sehen die meisten PraktikantInnen den höchsten Lerneffekt durch die Kombination dieser Prinzipien gegeben. Die PraktikantInnen schätzten die Wirkung der Informatik-Werkstatt auf die Besucher so ein, dass deren Interesse an Informatik gesteigert werden konnte.

4 Fazit

Aufgrund steigender sozio-technischer Herausforderungen, wird es immer wichtiger Menschen für technische Fächer zu begeistern. Die in Klagenfurt implementierte Informatik-Werkstatt liefert hier nachweislich einen bedeutenden Beitrag dazu. Die begleitenden Studien zeigen aber auch, dass wir mit den Bemühungen nicht ruhen dürfen. Es bedarf weiterer Aufklärungsarbeit in Bezug auf informatische Denkweisen beim Lehrkörper in den Schulen, und noch breiteren Angeboten – Schritte die wir in Klagenfurt bereits seit 2014 setzen.

Literaturverzeichnis

- [Be09] Bell, T.; Alexander, J.; Freeman, I.; Grimley, M.: Computer Science Unplugged: School Students Doing Real Computing Without Computers. In: New Zealand Journal of Applied Computing and Information Technology 13(1), 20–29 (2009)
- [MBH10] Mittermeir, R.; Bischof, E.; Hodnigg, K.: Showing core-concepts of informatics to kids and their teachers. In: 4th International Conference on Informatics in Secondary Schools - Evolution and Perspectives (ISSEP): Teaching Fundamentals Concepts of Informatics.

pp. 143–154. Springer Berlin Heidelberg (2010)

- [Pa16] Pasterk, S.; Sabitzer, B.; Demarle-Meusel, H.; Bollin, A.: Informatics-Lab: Attracting Primary School Pupils for Computer Science. In: Proc. of LACCEI Int. Multi-Conference for Engineering, Education, and Technology, San José, Costa Rica, 2016.
- [Sa14] Sabitzer, B.: A Neurodidactical Approach to Cooperative and Cross-curricular Open Learning: COOL Informatics, Habilitation thesis, AAU Klagenfurt, 2014.
- [SPR14] Sabitzer, B.; Pasterk, S.; Reci, E.: Informatics - A Childs Play?! In: Proceedings of the International Conference on Education and New Learning Technologies, EDULEARN. StudienVerlag, pp. 1081–1090, 2014.

Initiative IFIT – Den Nachwuchs für IT und Technik begeistern!

Bernhard Löwenstein¹

Abstract: Obwohl unbestritten ist, dass die Informationstechnologie (IT) in Zukunft eine immer bedeutendere Rolle spielen wird, gibt es in Österreich kaum flächendeckende Initiativen, um mehr Kinder und Jugendliche dafür zu begeistern. Das rief 2011 eine Gruppe von Idealisten auf den Plan, um eine Organisation zu gründen, die sich des Themas annimmt: Das Institut zur Förderung des IT-Nachwuchses (IFIT) war geboren – und versucht seither seine Mission in die Realität umzusetzen: Den Nachwuchs für IT und Technik zu begeistern!

Keywords: Institut zur Förderung des IT-Nachwuchses, IFIT, MINT, Robotik, Bee-Bot, Lego Mindstorms EV3, Arduino-basierte Roboter, NAO

1 Einleitung

Es gibt in Österreich zwar verschiedenste MINT-Initiativen, den meisten dieser Projekte ist allerdings gemein, dass es sich um zeitlich oder regional beschränkte Förderprogramme handelt. Individuelle Kursangebote, die sich ganz nach den Rahmenbedingungen einer Bildungseinrichtung richten, sind da schon deutlich seltener – speziell wenn sie in ganz Österreich abrufbar sein sollen. Mit dem Institut zur Förderung des IT-Nachwuchses gibt es seit Ende 2011 eine Initiative, die diese Lücke zu füllen versucht.

2 Über IFIT

Aktuell ist das Institut zur Förderung des IT-Nachwuchses die größte aktive MINT-Förderorganisation, die altersgerecht gestaltete Technologie-Workshops für Kinder und Jugendliche in ganz Österreich organisiert und durchführt. Tabelle 1 gibt einen Überblick über das aktuelle Kursprogramm.

Kursangebot	Altersgruppe
Elektrizität & Elektronik-Workshop	ab 2. Schulstufe
Mechanik-Workshop	ab 3. Schulstufe
Programmier-Workshops	
Minecraft	ab 3. Schulstufe
Scratch	ab 3. Schulstufe
Roboter-Workshops	

¹ Institut zur Förderung des IT-Nachwuchses, Vogelsangweg 4, 3270 Scheibbs, b.loewenstein@gmx.at

Arduino-basierte Roboter	ab 6. Schulstufe
Bee-Bot	ab Kindergarten
Lego Mindstorms EV3	ab 7. Schulstufe
NAO	ab 4. Schulstufe
Roboter allgemein	ab 1. Schulstufe
Technikwissen unplugged-Workshops	ab 1. Schulstufe
Trickfilm-Workshop	ab 1. Schulstufe
Verschlüsselungs-Workshop	ab 3. Schulstufe

Tab. 1: Kursprogramm von IFIT

Seit der Gründung Ende 2011 wurden 445 Workshops mit 6.697 Teilnehmenden in Österreich und Deutschland durchgeführt (Stand: 18. Juni 2016). Anhand von Abbildung 1 lässt sich die stetige Steigerung der jährlichen Kursanzahlen deutlich nachvollziehen. Heuer wird wohl erstmals die 200er Marke überschritten werden.

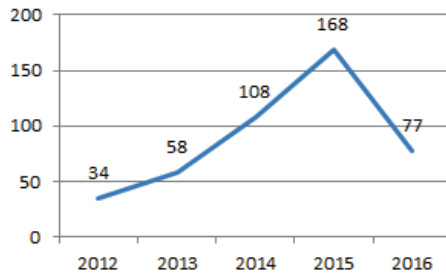


Abb. 1: Entwicklung der Workshopzahlen (Stand: 18. Juni 2016)

Die Tätigkeiten des in Scheibbs ansässigen Instituts beschränken sich mittlerweile nicht mehr nur auf die Organisation und Durchführung von Technologie-Workshops, sondern sind vielfältig, wie die folgende Auflistung zeigt:

- Weiterbildung von Lehrenden
- Durchführung der Lehrveranstaltung "Spezielle Kapitel der Schulinformatik" an der Technischen Universität Wien
- Vorträge bei Konferenzen und Publikation von Fachartikeln
- Beratung von Bildungseinrichtungen, Organisationen und Unternehmen hinsichtlich ihrer MINT-Förderstrategie
- Auftritte bei Veranstaltungen

Der rein ehrenamtlich geführte Verein finanziert sich fast ausschließlich privat. Lediglich rund 5 % des Jahresbudgets werden durch Zuwendungen der öffentlichen Hand gedeckt. Trotz intensiver Bemühungen gelang es dem Vereinsvorstand bisher nicht, bedeutende Förderbeträge zu lukrieren. Zur Anschaffung des erforderlichen

Materials investierte IFIT in den vergangenen Jahren rund 50.000 Euro.

Als Lehrende setzt das Institut auf einen Mix aus entsprechend fachlich, pädagogisch und didaktisch qualifizierten Personen. Das Spektrum reicht von Mechatronikern und Informatikern bis hin zu Kunstschaffenden und Psychologen. Genau dieser Mix ermöglicht es IFIT ganzheitliche Kurskonzepte zu erstellen.

Wesentlicher Erfolgsfaktor ist die maximale Flexibilität hinsichtlich der Kursgestaltung:

- Alter: 4 Jahre ⇔ erwachsen
- technisches Level: Kindergarten ⇔ professionelle Softwareentwicklung
- Kursdauer: 45 Minuten ⇔ mehrere Semester
- Veranstaltungsrahmen: Schulkurs ⇔ Ferienprogramm ⇔ Abendveranstaltung
- spezielle Förderprogramme: Hochbegabte ⇔ Mädchen ⇔ jugendliche Straffällige

3 Eingesetzte Robotiktechnologien

Die positive Entwicklung von IFIT hängt sicherlich mit der Schwerpunktsetzung auf die Robotik zusammen. Speziell in den ersten beiden Jahren wurden fast ausschließlich solche Workshops durchgeführt. Restliche Kursangebote kamen erst später hinzu. Heute nehmen die Roboter-Workshops immer noch einen Anteil von rund 80 % ein, weshalb die von IFIT eingesetzten Robotiktechnologien in Folge vorgestellt werden sollen.

3.1 Bee-Bot

Die Bodenroboter der britischen TTS Group zeichnen sich durch eine einfache und kindgerechte Gestaltung aus. Programmiert werden sie über die Tasten auf ihrem Rücken. Möglich sind Bewegungen nach vorne und zurück sowie Drehungen nach links und rechts. Dadurch ergeben sich einfache Ablaufmuster ohne komplizierte Winkel. Die Bienenroboter eignen sich ausgezeichnet für einen spielerischen Einstieg in die IT.

Pia Brüner: „Ich finde es immer wieder faszinierend, wie sich innerhalb weniger Stunden eine völlig neue räumliche Vorstellungskraft bei den Kindern entwickelt. Noch beeindruckender ist, wieviel Freude die Kinder am Umgang mit den Bienenrobotern haben. Am Ende höre ich oft: 'Können die Bienen morgen wieder kommen?'“

3.2 Lego Mindstorms EV3

Die Baukästen des dänischen Lego-Konzerns beinhalten neben unzähligen Lego-Technik-Bauteilen einen Mikrocontroller, an den sich Aktoren und Sensoren anschließen

lassen. Die Kinder und Jugendlichen können damit unterschiedlichste Robotermodelle konstruieren und anschließend programmieren. Für die verschiedenen Altersgruppen stehen unterschiedliche Umgebungen und Sprachen bereit.

Klaus Marterbauer: „Meiner Erfahrung nach sind so gut wie alle SchülerInnen von Lego Mindstorms EV3 begeistert – sei es wegen der Kreativität beim Bauen und Programmieren, durch das Erreichen von (realistischen) Zielen oder schlicht das Ausprobieren von etwas Neuem. Gerade beim Benützen der Software ist interessant zu beobachten, wie die Kids immer mutiger werden und die Angst vor Fehlern ablegen.“

3.3 Arduino-basierte Roboter

Bei Arduino handelt es sich um eine aktuell sehr populäre Physical Computing-Plattform, für die es unzählige Aktoren und Sensoren gibt. Sowohl die Hardware als auch die Software sind im Sinne von Open Source quelloffen. Die Programmierung erfolgt in einer C-ähnlichen Programmiersprache. Interessierte können unterschiedlichste Elektronikprojekte damit umsetzen.

Nikola Kodzic: „Es dauert zwar eine Weile, alle Sensoren vorzustellen und gemeinsam auszuprobieren, aber danach bin ich immer wieder erstaunt, auf welche kreativen Ideen die Teilnehmenden kommen.“

3.4 NAO

Beim NAO der japanischen Firma SoftBank Robotics handelt es sich um den fortschrittlichsten humanoiden Roboter für den Ausbildungsbereich. Er hat zahlreiche Motoren und Sensoren an Bord. Neben Sprach-, Gesichts- und Objekterkennung kann der NAO-Roboter auch mehrsprachig kommunizieren. Die Programmierung ist in unterschiedlichen Umgebungen und Sprachen möglich. Trotz der hohen technischen Komplexität ist die NAO-Programmierung selbst für Kinder einfach möglich.

Alexander Löwenstein: „Wo auch immer wir mit den beiden NAOs hinkommen, die Kinder und Jugendlichen sind von den niedlich wirkenden Humanoiden sofort angetan. Ich hätte niemals gedacht, dass sich die SchülerInnen durch zwei Maschinen so für das Thema Programmierung begeistern lassen.“

4 Ausblick

Obwohl die mittlerweile erreichte Kursanzahl der ehrenamtlichen Führung von IFIT einen gewaltigen Arbeitseinsatz abverlangt, ist das Institut zur Förderung des IT-Nachwuchses weiter auf Expansionskurs. So fand es mit dem Robotikspezialisten noDNA einen starken Partner, um die Kursaktivitäten zukünftig noch weiter ausbauen zu können.

Das Hildesheimer Denkwerk-Projekt – Schülerinnen und Schüler modellieren und analysieren Geschäftsprozesse

Thorsten Schoormann¹, Dennis Behrens² und Ralf Knackstedt³

Abstract: In dem von der Robert Bosch Stiftung⁴ geförderten Projekt „SchülerUni: Geschäftsprozesse nachhaltig gestalten“ werden, in Zusammenarbeit mit Schülerinnen und Schülern, Konzepte und Verfahren betrachtet, mit denen Nachhaltigkeit in Geschäftsprozessen integriert werden kann. Mit Geschäftsprozessen werden wertschöpfende Abläufe von Unternehmen fokussiert, die ökonomischen, ökologischen und sozialen Ziele gerecht werden sollen. Das Projekt adressiert zwei Ebenen. In der „Prozessmodellierung“ (I) werden Prozesse erhoben, modelliert und auf Nachhaltigkeit analysiert. In der „methodischen Weiterentwicklung“ (II) werden innovative Ansätze entwickelt, um die Modellierungstechniken selbst sinnvoll mit Nachhaltigkeitsaspekten zu ergänzen.

Keywords: Hildesheimer Denkwerk, SchuelerUni, Geschäftsprozesse, Nachhaltigkeit

1 Hildesheimer Denkwerk

Überblick. Das Hildesheimer Denkwerk-Projekt „SchülerUni – Nachhaltige Geschäftsprozesse gestalten“ (Laufzeit 08/2014 bis 07/2017) wird unter der Leitung der Abteilung Informationssysteme und Unternehmensmodellierung (Wirtschaftsinformatik) umgesetzt und durch die Abteilung Wirtschaftswissenschaft und ihre Didaktik wissenschaftlich sowie durch das Gleichstellungsbüro der Universität Hildesheim beratend begleitet. Beteiligt sind Wissenschaftler_innen, studentische Hilfskräfte, Schüler_innen und Lehrkräfte sowie regionale Praxisunternehmen.

Thema. Nachhaltigkeit ist ein weit verbreiteter Begriff, der sowohl von der Gesellschaft als auch von Unternehmen und Organisationen diskutiert wird [Bu14]. Im Rahmen des Denkwerks wird nicht die Nachhaltigkeit im Sinne einer langfristigen Nutzung adressiert, sondern die, die unter dem Begriff „Sustainability“ diskutiert wird. Dabei sollen Dimensionen für Ökonomie, Ökologie und Soziales gleichberechtigt betrachtet werden.

Zur Integration nachhaltiger Ansätze in Organisationen, können Werkzeuge einen unterstützenden Beitrag leisten, die es ermöglichen, anhand eines Ist-Zustandes, Potenziale und Maßnahmen abzuleiten. Das Modellieren von Geschäftsprozessen stellt

¹ {Universität Hildesheim, Informationssysteme und Unternehmensmodellierung, 31141 Hildesheim}
thorsten.schoormann@uni-hildesheim.de

² dennis.behrens@uni-hildesheim.de

³ ralf.knackstedt@uni-hildesheim.de

⁴ Robert Bosch Stiftung, <http://www.bosch-stiftung.de/content/language1/html/58818.asp>, Zugriff: 13.05.2016.

ein zentrales Konzept der Wirtschaftsinformatik dar und unterstützt die Betrachtung und Analyse wertschöpfender Abläufe. Damit Abläufe möglichst standardisiert und eindeutig visualisiert werden, existieren verschiedene Modellierungstechniken (z. B. EPK, BPMN).

Ein Ziel des Projektes ist es, gemeinsam mit Schüler_innen die offene Forschungsfrage zu bearbeiten, inwiefern Modellierungstechniken das Dokumentieren und das Analysieren von Prozessen hinsichtlich verschiedener Nachhaltigkeitsaspekte unterstützen.

Erlernte Methoden und Instrumente lassen sich in alltägliche Situationen übertragen, indem Schüler_innen z. B. (private) Abläufe mit den methodischen Werkzeugen kritisch reflektieren. Zudem werden weitere Forschungsmethoden (z. B. Interview und Fragebogen) angewendet, realistische Einblicke in das wissenschaftliche Arbeiten ermöglicht und soziale Kompetenzen durch das Arbeiten in schulübergreifenden Gruppen gestärkt.

2 Projektaufbau und -inhalte

Das Projekt adressiert zwei Ebenen (vgl. Tab. 1). Die *Prozessmodellierung* (Ebene I) befasst sich mit der Erhebung, Modellierung und (Nachhaltigkeits-)Analyse von Prozessen. Die *methodische Weiterentwicklung* (Ebene II) eröffnet die Möglichkeit, kreative Ansätze zur Anpassung der Modellierungstechnik selbst zu entwickeln.

	Einführung	Zieldefinition	Vorbereitung	Durchführung	Auswertung	Präsentation
Ebene I	Prozessmodellierung	Modellierungsgegenstand	Interviewleitfäden und Terminorganisation	Prozessmodellierung und -konsolidierung	Abstimmung der Modelle mit Praxispartnern	Symposium
Ebene II	Forschungsfrage	Ideen zur Weiterentwicklung	Evaluationsunterlagen	Evaluation eigener Ansätze	Statistische Aufbereitung der Resultate	Symposium

Tab. 1: Allgemeiner Projektaufbau

Ebene 1 – Prozessmodellierung. Im Folgenden werden konkrete Umsetzungen und exemplarische Resultate der Ebene zur Prozessmodellierung skizziert.

Prozessmodellierung. Zur Dokumentation der Prozesse wird die Modellierungstechnik der erweiterten Ereignisgesteuerten Prozesskette (eEPK) eingeübt, die in der Unternehmenspraxis als weit verbreitet gilt (z. B. [Kr10] [SNZ95]). Damit die erstellten Modelle analysiert werden können, erhalten die Schüler_innen einen Verbesserungskatalog der allgemeine Ansätze wie z. B. Beschleunigung, Zusammenlegung, Automatisierung und Anpassung der Reihenfolge adressiert. Mit Hilfe dieser Ansätze können bereits – eher wirtschaftliche – Maßnahmen zur

Verbesserung identifiziert und erarbeitet werden.

Nachhaltigkeit als Modellierungsgegenstand. Zur Berücksichtigung nachhaltiger Aspekte, muss zunächst der breite Begriff der Nachhaltigkeit eingeführt und abgegrenzt werden, der aktuell durch verschiedene Ansätze geprägt ist. Im Rahmen des Denkwerks stützen wir uns insbesondere auf das Drei-Säulen-Modell [PAM05] (Ökologie, Ökonomie und Soziales), den drei Nachhaltigkeitsstrategien [Hu95] (Effizienz, Konsistenz und Suffizienz) sowie dem historischen Ursprung des Begriffs im Brundtland-Bericht [Wo87].

Aufnahme, Dokumentation und Verbesserung von Prozessen. Im Anschluss an die Erhebung (z. B. mit Fragebögen und Interviews) sowie der Visualisierung von Prozessen, können die erlernten Kompetenzen praxisnah angewendet werden. Im Rahmen einer Exkursion zur Mensa der Universität Hildesheim kann überprüft werden, wie die Essenzubereitung in Großküchen nachhaltiger gestaltet werden kann. Dazu teilten sich die insgesamt 40 Schüler_innen in Teams auf, um spezifische Bereiche wie z. B. das Kassenwesen, die Menüplanung, das Hygienemanagement oder die Warenannahme, zu bearbeiten. Dazu standen entsprechende Expert_innen für Interviews zur Verfügung.

Prozessmodellierung und -konsolidierung. Anschließend müssen die gewonnen Informationen ausgewertet werden, um diese entsprechend dokumentieren zu können. Dabei entstehen vor allem Herausforderungen hinsichtlich (a) der Anwendung der eEPK, (b) der Entscheidung darüber welche Passagen und Aussagen für den Ablauf relevant sind sowie (c) die Zusammensetzung des großen Ganzen aus vielen erhobenen Details.

Nachhaltigkeitsanalyse. Auf Basis der Ist-Modelle können nun Verbesserungspotenziale hinsichtlich der Dimensionen der Nachhaltigkeit ermittelt werden. Um diese zu visualisieren, fließen die Potenziale in die Soll-Modellierung, in der z. B. Funktionen für das Spenden von nicht-verwendeter Ware (Soziales), das Einbeziehen regionaler Lieferanten (Ökologie) oder das Parallelisieren von Abläufen (Ökonomie) vorgeschlagen wurde.

Symposium. Die Ergebnisse werden regelmäßig zum Ende jedes Halbjahres sowohl den Praxispartnern als auch den anderen Schüler_innen, den Lehrer_innen und den beteiligten Wissenschaftler_innen vorgestellt und gemeinsam diskutiert.

Ebene 2 – Methodische Weiterentwicklung. Im Folgenden werden konkrete Umsetzungen und erarbeitete Beispiele der methodischen Weiterentwicklung erläutert.

Forschungsfrage definieren. In der zweiten Ebene stellen sich die Projektbeteiligten gemeinsam die Frage, inwiefern die Modellierungstechnik selbst für die Visualisierung der Nachhaltigkeitsanalyse geeignet ist.

Ideen zur Weiterentwicklung. In kleinen Teams werden Maßnahmen zur Anpassung der Modellierungstechnik, um Konzepte der nachhaltigen Entwicklung, konzipiert.

Grundsätzlich erfolgt dabei (a) die Integration neuer Symbole, (b) die Bewertung der Nachhaltigkeit durch bspw. Ampel-Metaphern sowie (c) die Integration von Checklisten, die während des Modellierens beachtet und verwendet werden sollen. Resultate adressieren z. B. Ideen zur Visualisierung eines hohen Papierverbrauchs (Ökologie), einseitiger Arbeitsbelastungen (Soziales) oder zeitintensiver Funktionen (Ökonomie).

Evaluation der Weiterentwicklung. Um zu zeigen, dass die eignen Weiterentwicklungen nützlich sind, werden z. B. Aspekte der Verständlichkeit der Prozessmodelle (Multiple-Choice-Fragebögen), des Zeitaufwandes (Dauer der Erstellung und Interpretation), der Ideenentwicklung (Untersuchung inwiefern Nachhaltigkeit adressiert wird) sowie der Modellqualität (Beurteilungen durch unabhängige Expert_innen) untersucht.

Verkaufspräsentation der Ergebnisse. Da die Teams Konzepte entwickelten, die einerseits explizit die Notation der Modellierung betreffen und andererseits Leistungen für den Konstruktionsvorgang enthalten, erfolgte die Vorstellung der Ergebnisse im Rahmen einer „Verkaufspräsentation“. In dieser werden Ergebnisse einer Jury – bestehend aus Wissenschaftler_innen und Lehrer_innen – vorgestellt, die zum „Kauf“ überzeugen soll.

3 Zwischenresultate und Reflektion

In der Ebene der *Prozessmodellierung (I)* konnten verschiedene praktisch-einsatzbare Weiterentwicklungsmaßnahmen (Soll-Modelle) durch die Schüler_innen aufgezeigt werden wie z. B. eine mobile Applikation mit Vorbestelloptionen zur besseren Ressourcenplanung, die Neu-Sortierung der Ware nach Mindesthaltbarkeit, das optionale Drucken von Kassenbelegen, die Verbesserung der Kommunikation durch IT-Systeme oder die Verwendung umweltschonender Chemikalien im Hygienemanagement.

Besonders bei der *methodischen Weiterentwicklung (II)* war erkenntlich, dass das Lösen von einzelnen Prozessen eine Herausforderung für die Schüler_innen darstellt. Das Abstrahieren von einem konkreten Prozess auf einen allgemeinen Lösungsansatz fiel schwer. Trotzdem konnten Weiterentwicklungen (z. B. neue Symbole) entwickelt werden.

Danksagung. Das Hildesheimer Denkwerk (32.5.6021.0079.0) wird ermöglicht durch die Förderung der Robert Bosch Stiftung. Wir danken dem Förderer für die Unterstützung.

Literaturverzeichnis

- [BU14] Nachhaltigkeit – Ein politisches Leitprinzip, <https://www.bundesregierung.de/Content/DE/StatischeSeiten/Breg/Nachhaltigkeit/0-Buehne/2014-01-03-ein-politisches-leitprinzip.html>, Stand: 20.06.2016.
- [Hu95] Huber, J.: Nachhaltige Entwicklung durch Suffizienz, Effizienz und Konsistenz. In (Fritz, P.; Huber, J.; Levi, H.W. Hrsg.): Nachhaltigkeit in naturwissenschaftlicher und

sozialwissenschaftlicher Perspektive, Stuttgart, Hirzel, S. 31-46, 1995.

- [Kr10] Krcmar, H.: Informationsmanagement. 5. Auflage, Berlin, 2010.
- [PAM05] Pope, J.; Annandale D.; Morrision-Saunders A.: Conceptualising sustainability assessment. In: Environmental impact assessment review 24(6), S. 595-616, 2005.
- [SNZ95] Scheer, A.W.; Nüttgens, M.; Zimmermann, V.: Rahmenkonzept für ein integriertes Geschäftsprozessmanagement. In: Wirtschaftsinformatik 5, S. 426-434, 1995.
- [Wo87] World Commission on Environment and Development. Our Common Future. Oxford, Oxford University Press, 1987.

Pi and More — Eine Veranstaltungsreihe rund um “kleine Computer”

Daniel Fett¹ und Guido Schmitz²

Abstract: *Pi and More* ist eine Veranstaltungsreihe zu dem Single-Board-Computer *Raspberry Pi* und ähnlichen Systemen. Ziel der Veranstaltungen ist es, Informatik und Technik praktisch zu vermitteln, indem ein Ort geschaffen wird, an dem sich Schüler, Lehrer, Studierende, interessierte Laien, Maker und professionelle Anwender austauschen und voneinander lernen können.

Die halbjährliche Veranstaltung hat sich mit bis zu 300 Besuchern pro Event als feste Größe etabliert und ist in ihrem Konzept einzigartig im deutschsprachigen Raum.

1 Einführung

Moderne Rechnersysteme sind sehr komplex, hochintegriert und bieten kaum Ansatzpunkte für Laien, die Funktionsweise verstehen zu können. Konzipiert mit vielen Abstraktionsebenen sehen Endanwender und Entwickler große Bereiche eines solchen Systems nur als „black box“. Dies trifft insbesondere auf die Hardware-Ebene zu, die in modernen Systemen primär mit hochkomplexen Schnittstellen konzipiert ist und somit den meisten Interessierten verschlossen bleibt.

Dies war in der Vergangenheit anders: Mit den Hobby-PCs der 1980er-Jahre, wie beispielsweise dem *Commodore 64*, etablierte sich eine große Gemeinschaft von Tüftlern und Bastlern, die für ihre Rechner mit Begeisterung selbst Hardware entwickelten. Diese konnten sie mit den damals einfach gehaltenen Schnittstellen der Rechner verbinden und mit selbstgeschriebener Software direkt ansteuern. Nachdem diese Generation von PCs auf dem Markt durch modernere Systeme abgelöst worden war, gab es kaum noch Systeme, die einen so einfachen Zugang zu Hard- und Softwareentwicklung ermöglichten.

Im Jahr 2005 erschien das Mikrocontrollerboard *Arduino*, das für interessierte Personen ohne Hintergrund in Elektronik und Programmierung eine Möglichkeit bot, interaktive Systeme zu entwickeln. Durch eine einfache Entwicklungsumgebung wurde die Schwelle zum Einstieg in die Hardwareprogrammierung stark gesenkt.

2012 brachte die Raspberry Pi Foundation, eine gemeinnützige Stiftung in England, den *Raspberry Pi* auf den Markt, der einerseits ein vollwertiger Linux-PC ist und andererseits (wie der Arduino) leichten Zugang zu hardwarenaher Programmierung mit *general purpose input/outputs (GPIOs)* sowohl in der Programmiersprache *Python* als auch mit der grafischen Programmierumgebung *Scratch* ermöglicht. Primär entwickelt um britischen Schulen eine kostengünstige Möglichkeit zu geben, einen solch universellen Rechner für den Schulunterricht einzusetzen, schloss der Raspberry Pi die Lücke zwischen Elektronik

¹ Universität Trier, Universitätsring 15, 54286 Trier, fett@uni-trier.de

² Universität Trier, Universitätsring 15, 54286 Trier, schmitzg@uni-trier.de

und Software und motiviert zunehmend mehr Menschen weltweit, sich mit Hardwareentwicklung und Programmierung zu beschäftigen.

2 Veranstaltungsreihe *Pi and More*

Initiiert durch die Raspberry Pi Foundation entstanden in England sogenannte *Raspberry Jams*, dezentrale Veranstaltungen für ein breites Publikum, bei denen sich Interessierte treffen, um sich über Hard- und Softwareentwicklung mit dem Raspberry Pi auszutauschen.

Aus der Idee heraus, eine ähnliche Veranstaltung auch nach Deutschland zu bringen, entstand die Veranstaltungsreihe *Pi and More*. Wie der Name andeutet, ist die Veranstaltung bewusst nicht auf den Raspberry Pi beschränkt, sondern bezieht vielmehr auch andere Single-Board-Computer wie den Arduino mit ein.

Im August 2012, ein halbes Jahr nach dem Verkaufsstart des Raspberry Pis, fand das erste *Pi and More* mit einem Programm aus fünf Vorträgen und Workshops an der Universität Trier statt. Seitdem wird die Veranstaltung halbjährlich fortgeführt, wobei die große Sommerausgabe immer an der Universität Trier stattfindet und eine kleinere Winterausgabe an einer anderen Hochschule in der Großregion.

Seit 2012 entwickelte sich *Pi and More* mit stetig steigenden Besucherzahlen (von anfänglich 60 auf heute 300 Besucher pro Event) zu einer festen Größe in der Community. Das Programm der einzelnen Events wuchs ebenfalls auf zuletzt über 30 Beiträge.

Die Veranstaltung steht im Wesentlichen auf drei Säulen: Vorträge, Workshops und eine Projektausstellung. Bei den *Vorträgen* werden die Themen von den Referenten meist ohne Interaktion mit dem Publikum vorgestellt. Während Vorträge in großen Hörsälen stattfinden, werden *Workshops* in kleineren Räumen mit beschränkten Teilnehmerzahlen (z.B. 20 Teilnehmer) durchgeführt. Bei Workshops werden konkrete Projekte mit den Besuchern zusammen durchgeführt – beispielsweise die Inbetriebnahme des Raspberry Pis, Programmierung von Minecraft oder der Aufbau elektronischer Schaltungen. Schließlich gibt es mit der *Projektausstellung* für die Teilnehmer die Möglichkeit, eigene Software oder Hardware im Betrieb zu demonstrieren. So finden sich hier beispielsweise Roboter, Hausautomationslösungen und elektronisch animierte Kunstobjekte.

Pi and More zielt darauf ab, ein Forum zu bieten, in dem sich Schüler, Lehrer, Studenten, interessierte Laien, Maker und professionelle Anwender austauschen können. Tatsächlich haben Besucherumfragen in der Vergangenheit bestätigt, dass diese Zielgruppen alle zu etwa gleichen Teilen vertreten sind.

Zur besseren Organisation und Finanzierung der Veranstaltung wurde der gemeinnützige Verein *CMD – Computer, Menschen, Dinge e. V.* gegründet. Dieser organisiert zusammen mit Informatik-Fachschaften und -Lehrstühlen der jeweiligen Hochschule das Event. Durch die meist kostenlose Bereitstellung der Infrastruktur durch die Hochschulen ist der Eintritt zu den Veranstaltungen kostenlos. Dies ist ein wichtiger Baustein des Konzepts, um das Angebot niedrigschwellig zu gestalten. Andere Kosten, beispielsweise für Workshopmaterialien, werden über Sponsoren gedeckt.

FROM SKETCH TO SCRATCH - schrittweise zu „computational thinking“ geführt werden

Alois Bachinger¹ und Anton J. Knierzinger²

Abstract: Grundsätzlich ist es nicht notwendig, dass sich Kinder im Kindergarten und in der Grundschule mit Informatik beschäftigen. Was wir im Workshop aber zeigen wollen ist, dass es sehr sinnvoll ist, sich mit Strukturen des allgemeinen Problemlösens und mit technischen Aspekten der kindlichen Spielwelt und des Lernalltags zu beschäftigen. Die Gründe dafür liegen nicht nur in der Omnipräsenz von digitaler Technik, der auch Kinder in diesem Alter ausgesetzt sind. Vielmehr vergrößert die Entwicklung von algorithmischen Denken die persönlichen und methodischen Kompetenzen. Nicht zuletzt aber ermöglicht die Beschäftigung mit Codieren die Umsetzung didaktischer Modelle, die kreativen, und motivierenden Unterricht ermöglichen. Der Weg von der Problemanalyse (Sketch) bis zur formalen Beschreibung eines Algorithmus (in Scratch) kann für Kinder lustig sein und ihre Kreativität anregen.

Ausgehend von der Bedeutung die Algorithmen in unserer Gesellschaft erreicht haben, soll unser 7-Stufen Modell zum Unterricht von Coding im Grundschulalter präsentiert und die Werkzeuge Faktoren, die Kinder in diesem Alter für dieses Thema begeistern können, gezeigt und diskutiert.

Keywords: Kindergarten, Grundschule, Codieren, algorithmisches Denken, Didaktik

1 Einleitung

Begeisterung für die Beschäftigung mit einem Inhalt oder einer Methode ist ein Grundelement erfolgreichen Lernens. Wir verfolgen seit 1980 die Integration von Informationstechnik in der Schule. Die Frage, warum wir in Deutschland und Österreich zu wenig Nachwuchs in technischen Berufen auf allen Ebenen haben, hängt sehr stark mit der Einstellung der Schüler zu Technik und ihrer Begeisterung dafür zusammen. Die Entwicklung von „Computational Thinking“ ist ein Weg, Schüler zu begeistern.

2 Was ist „computational thinking“?

Die Bezeichnungen für arbeiten mit Computern in der Schule hat sich im Lauf der Zeit immer wieder verändert. Ursprünglich war es Informatik, später dann Informatische Grundbildung, usw. Für unsere Arbeit passt wohl am besten: computational thinking. Diesen Begriff beschreibt Jeanette M. Wing [Wi08] als: „... it represents a universally

¹ Pädagogische Hochschule Linz, Salesianumweg 3, 4020 Linz, alois.bachinger@ph-linz.at

² Pädagogische Hochschule Linz, Figulystr. 1, 4020 Linz, anton.knierzinger@ph-linz.at

applicable attitude and skill set everyone, not just computer scientists, would be eager to learn and use“. Es geht also sowohl um Einstellungen als auch um Fertigkeiten und zwar solche, die sich an die Allgemeinheit und nicht nur an Spezialisten wenden. Computational thinking heißt viel mehr als Computer programmieren zu können. Es braucht dazu Denken auf verschiedenen Ebenen der Abstraktion. Dazu gehören:

- Konzepte zu entwickeln statt programmieren zu lernen,
- kreative Prozesse statt erlernte Fähigkeiten anwenden,
- in der Art von Menschen nicht in der von Computern zu denken,
- dabei mathematische und technische Kompetenzen zu entwickeln und
- Ideen, nicht Werkzeuge in den Vordergrund zu stellen.

3 Didaktische Begründung

Aus unseren Erfahrungen und Untersuchungen, aber auch aus Literaturhinweisen, kann abgeleitet werden, dass computational thinking sowohl die Methodenkompetenzen (Umgang mit IT und digitalen Medien, Sprachkompetenz) wie auch die persönlichen Kompetenzen (Kommunikationsfähigkeit und Problemlösekompetenz, Selbstbewusstsein und Selbsteinschätzung) erweitert.

4 Faktoren für Begeisterung

Resnicks 4P's [Re16] gibt folgende Faktoren für kreatives Lernen an:

- **projects:** Das Arbeiten soll zielorientiert, fächerübergreifend, praxisorientiert und vor allem kreativ sein.
- **peers:** Lernen floriert als soziale Aktivität, wenn Leute Ideen austauschen, an gemeinsamen Zielen arbeiten und ihre Ergebnisse zusammenlegen.
- **passion:** Wer an Projekten arbeitet, die ihm sinnvoll erscheinen, Spass machen, und herausfordernd sind, der arbeitet länger, härter und effektiver.
- **play:** Lernen soll Spass machen, experimentieren erlauben, zu seinen Grenzen führen und immer wieder probieren erlauben.

In den Aufgabenstellungen sollen sich Situationen mit spielerischem Charakter mit Bezügen zu Alltag, Technik und Phantasiewelten mischen und zu ergänzen.

5 7 Stufen im Programmierunterricht für Anfänger

Dieses Konzept stellt 7 Stufen im Unterricht von Problemlösen und Codieren dar. Von Alltagssituationen (Realität) aus wird über dingliche Darstellung und Mischformen mit steigender Abstraktion eine rein maschinelle Repräsentation (Virtualität) von Algorithmen erreicht. Die Ausprägung der einzelnen Stufen muss an Vorwissen, Alter und Lernziele angepasst werden, ist aber zumindest in den Grundelementen auch im Kindergartenalter möglich. Das Konzept stellt so einen Leitfaden für Lehrende dar. Verwendet werden vor allem Algorithmen, die Bewegung von Objekten codieren. In den Aufgabenstellungen sollen sich Situationen mit spielerischem Charakter mit Bezügen zu Alltag, Technik und Phantasiewelten mischen.

Die, in den folgenden sieben Stufen vorgeschlagene, Vorgangsweise steigert das notwendige Abstraktionsniveau stufenweise. Die Bandbreite der Einsatzszenarien reicht von 5 Jahren unter Berücksichtigung der entwicklungspsychologischen Veränderungen bis ins Erwachsenenalter. Bei jüngeren Kindern wird man sich stärker auf die ersten Stufen konzentrieren und nicht alle Stufen durchlaufen. Bei Erwachsenen, etwa in der Lehrerbildung, genügt es die ersten Stufen kurz zu streifen.

real:

Algorithmen werden im Alltag gefunden. Sie werden durch reale Dinge dargestellt (z.B. Bewegungen von Menschen, Objekten)

blended:

Algorithmen werden sowohl real wie auch softwaremäßig repräsentiert

virtuell:

Der Algorithmus, sein Ablauf und seine Wirkung laufen ausschließlich am Computer ab.

Zur didaktischen Umsetzung werden sieben Stufen vorgeschlagen:

Stufe 1: Algorithmen im Alltag.

Algorithmen sind Ketten von Anweisungen, Algorithmen entdecken, in Sprache beschreiben lernen, einzelne Schritte definieren können.

Stufe 2: Algorithmen als Anweisungen.

Lernende geben sich gegenseitig einfache Befehle zur Bewegungssteuerung, Bewegungen werden als Befehlsketten dokumentiert, Darstellung von Algorithmen als Symbole auf Würfeln und Befehlskarten.

Stufe 3: Steuerung eines tastenprogrammierbaren Roboters.

Als Roboter wird BeeBot, eine kleine, mit Tasten programmierbare „Biene“ verwendet. Die Eingabe des Algorithmus zu seiner Bewegungssteuerung erfolgt noch direkt am Gerät, zuerst step-by-step dann als Befehlsfolge. Schwierigkeitsstufen:

- 1: Freies Agieren in der Ebene ohne mathematisches Bezugssystem
- 2: Einführung eines Rasters (Koordinatensystems)

Stufe 4: Algorithmen in maschineller Darstellung.

Erste Umsetzung in maschinelle Darstellung erfolgt mit Darstellung der Bewegung des Roboters (Biene) am PC durch die Beebot-App, RunMarco-App und CargoBot-App.

Stufe 5: Algorithmen in einer einfachen Programmiersprache.

Verwendet wird vorerst Scratch junior. Als Aufgabenstellungen werden Spielsituationen und das Erzählen von Geschichten vorgeschlagen. Die Schüler werden animiert, selbst neue Problemstellungen zu erfinden und diese zu lösen.

Stufe 6: Konstruktion und Programmierung von Robotern.

Ausgehend von einem über Computersteuerung programmierbaren, fertigen Roboter (z.B. Auto), wird zu Roboterbaukästen (z.B. Cubelets) übergegangen. Einführung von Sensoren, Aktoren und Controls. Die Programmierung erfolgt über vorhanden Tools. Vorstellung von einfachen Programmstrukturen (Alternative, Iteration, Module)

Stufe 7: Programmieren in Scratch.

Scratch ist eine vollwertige Programmiersprache, die Einführung in das Programmieren in allen Altersstufen bis hin auf Universitätsniveau und auf allen Abstraktionsstufen ermöglicht. Mit ihr können auch die gebauten Roboter gesteuert werden.

6 Erfahrungsberichte

Dieses Konzept wurde bisher in zahlreichen, unterschiedlichen Settings erfolgreich durchgeführt. Erwähnt seien: KinderUNI, Talente Oberösterreich, Lange Nacht der Forschung, Studiengänge der Grundschullehrerbildung, Lehrerfortbildungsveranstaltungen für Primar- und Sekundarstufe, Pädagogische Veranstaltungen (Fachmessen Interpädagogica und Bildung online), Fachtagungen und auch in der Erwachsenenbildung (z.B. Sprachkompetenzentwicklung von Asylwerbern).

Literaturverzeichnis

- [KG15] Knierzinger, A.; Gradinarova, B.: Learning to Code, Coding to Learn - The world of algorithms in higher education. In: (Elmas, M., et.al., Eds.): ICQH Proceedings Book, University of Sakarya, icqh.net, Stand 30.12.2015

- [Re16] Resnick, M.: Give P's a Chance: Projects, Peers, Passion, Play, <http://web.media.mit.edu/~mres/papers/constructionism-2014.pdf>, Stand: 13.5.2016
- [Wi08] Wing, J., Computational Thinking and thinking about computing, <http://rsta.royalsocietypublishing.org/content/366/1881/3717>, Stand: 13.5.16

Begeisterung für Robotik – ein ganzheitlicher Ansatz der (Hoch-)Begabtenförderung mittels Blended-Learning am Beispiel des Talentehauses NÖ

Tanja Tomitsch¹ und Thomas Aschinger²

Abstract: Das Talentehaus Niederösterreich (NÖ) fördert (hoch-)begabte Jugendliche im MINT-Bereich (Mathematik, Informatik, Naturwissenschaft, Technik). Um die Jugendlichen optimal fördern zu können, wird dazu ein ganzheitlicher Ansatz verfolgt, der neben den fachlichen Inhalten auch soziales Lernen, Zeitmanagement, Projektmanagement, psychologische Begleitung usw. beinhaltet. Eine wichtige Frage hierbei ist, wie man die anfängliche Motivation der TeilnehmerInnen über den gesamten Lehrgang nutzen und fördern kann.

Keywords: Talentehaus NÖ, MINT, Robotik, STEM, Hochbegabung, Blended-Learning, E-Learning, Roboterprogrammierung, Roboterkonstruktion

1 Das Talentehaus Niederösterreich (NÖ)

Im Herbst 2013 öffnete das Talentehaus NÖ³als Projekt der NÖ Landesakademie⁴ [Nö16] mit der Förderrichtung MINT (Mathematik, Informatik, Naturwissenschaft, Technik) für (hoch-) begabte Kinder und Jugendliche seine Tore [Ta16].

1.1 Zielsetzung

Zielsetzung des Talentehauses NÖ ist es, (hoch-)begabten Kindern und Jugendlichen eine hochwertige, schulbegleitende Förderung zu bieten.

Für herausragende Leistungen auf einem Fachgebiet ist eine gezielte, ganzheitliche und stringente Förderung mit höchstem Anspruch ab dem Kindesalter erforderlich [Br08]. Durch das Talentehaus NÖ werden die Voraussetzungen geschaffen, sich im internationalen Wettbewerb zu behaupten. Die Kinder und Jugendlichen sollen im Sinne einer nachhaltigen Qualifizierung die Möglichkeit erhalten, sich mit den Besten auf Augenhöhe messen zu können [Ta16].

¹ Talentehaus NÖ, Neue Herrngasse 17A, 3100 St. Pölten, tanja.tomitsch@noe-lak.at

² Talentehaus NÖ, Neue Herrngasse 17A, 3100 St. Pölten, thomas.aschinger@noe-lak.at

³ www.talentehaus.at

⁴ www.noe-lak.at

1.2 Angebot

Kinder und Jugendliche, die am TalenteHaus NÖ gefördert werden, profitieren von einer flexiblen, ganzheitlichen und maßgeschneiderten Individualförderung in Form von Präsenzstunden und betreuten E-Learning-Angeboten. Das Lernen in der Gruppe wird neben individueller Betreuung ebenso gefördert wie fokussierte Selbst-Recherche unter Einbeziehung der aktuellsten Inhalte aus dem World Wide Web. Interdisziplinarität sowie Anwendungsorientiertheit garantieren eine Exzellenzförderung am Puls der Zeit.

Die Förderung zur Exzellenz erfolgt durch ExpertInnen unter Stärkung von Eigenverantwortlichkeit, professionellem Zeitmanagement und individuellem Lernen. Interesse, Motivation und Fleiß der Kinder und Jugendlichen sind Voraussetzung.

Für die Ausbildung im MINT-Bereich wurde das Fachgebiet der Robotik ausgewählt, da es sich um ein sehr umfassendes, technisches Querschnittsthema handelt. Derzeit laufen in diesem Bereich zwei Ausbildungsschienen:

- **Roboterprogrammierung:** Durch die Programmierung wird der Roboter zum Leben erweckt. Dieser Lehrgang ist auf die Erstellung und Anwendung von Software in der Robotik fokussiert.
- **Roboterkonstruktion:** Die Konstruktion und Elektronik bilden das Grundgerüst in der Robotik. Dieser Lehrgang setzt den Schwerpunkt auf die Konzeptionierung und Umsetzung der Hardware.

Zusätzlich zur Förderung im fachlichen Bereich ist die individuelle, persönlichkeitsbildende Weiterentwicklung der TeilnehmerInnen eine wichtige Komponente des TalenteHauses NÖ. Ergänzt werden die beiden Schwerpunkte durch eine psychologische Betreuung. Diese beinhaltet eine Testung, um eine zielgerichtete und individuell abgestimmte Förderung der einzelnen TeilnehmerInnen zu gewährleisten. Darüber hinaus wird eine kostenlose Beratung zur Besprechung der Testergebnisse sowohl für die TeilnehmerInnen als auch deren Erziehungsberechtigten angeboten.

Die Inhaltsvermittlung erfolgt beim TalenteHaus NÖ sowohl in Präsenzeinheiten als auch mittels E-Learning. Die Präsenzeinheiten finden ein Mal im Monat statt. Die E-Learning Aktivitäten erfolgen laufend zwischen den Präsenzterminen. Durch diesen Blended-Learning-Ansatz ist für die TeilnehmerInnen mit einem Arbeitsaufwand von ca. 10 Stunden pro Woche zu rechnen [Ta16].

1.3 Aufnahmeverfahren

Da das TalenteHaus NÖ nur eine begrenzte Anzahl an TeilnehmerInnen aufnehmen kann, muss vor dem Beginn der Ausbildung die Eignung individuell festgestellt werden.

Vor dem Start eines neuen Jahrgangs finden 1,5-tägige Aufnahmeworkshops an Wochenenden statt. Ein Team bestehend aus FachexpertInnen, PsychologInnen und PädagogInnen betreuen die BewerberInnen. Unterschiedlichste Aufgaben sind zu lösen und begleitend dazu werden die BewerberInnen psychologisch und pädagogisch beobachtet und beurteilt.

Das Ergebnis aus der vorliegenden Fachkompetenz im technischen Bereich in Kombination mit der sozialen und psychologischen Beurteilung ergibt ein Gesamtbild jedes und jeder einzelnen [Ta16].

2 Motivation in Blended-Learning-Umgebungen

Bezüglich des Zusammenhangs zwischen Motivation und Blended-Learning gibt es in der Literatur unterschiedliche Hypothesen und Ergebnisse [St16]. Bezüglich des Zusammenhanges dieser Variablen bei besonders begabten bzw. hochbegabten Jugendlichen existieren kaum Ergebnisse.

Eine hohe Leistungsmotivation ist eine wichtige Grundvoraussetzung um im Talentehaus NÖ aufgenommen zu werden. Diese Voraussetzung wird bereits vor Aufnahme im Zuge des Aufnahmeworkshops mittels FLM 7-13 (Fragebogen zur Leistungsmotivation) [PW07] getestet. Um den Lehrgang über die ganzen 3 Semester zu absolvieren, muss versucht werden, diese anfänglich vorhandene Leistungsmotivation für den Lehrgang nutzbar zu machen und zu verhindern, dass die teilnehmenden Jugendlichen über die Zeit ihre Motivation verlieren. Die Gefahr des Verlustes ist bei Blended-Learning-Ansätzen (besonders bei geringem Präsenzanteil) vor allem aufgrund des stark verminderten direkten Kontaktes sowie aufgrund der daraus resultierenden gefühlten leichteren Verschiebbarkeit der Aufgaben, da zum Beispiel die zeitnahe Kontrolle fehlt, erhöht [SK12].

Am Talentehaus NÖ wird laut Aussagen der Jugendlichen und deren Eltern der Motivationsverlust vor allem dadurch verhindert, dass gut bekannte und vertraute Betreuungspersonen zur Verfügung stehen, welche die TeilnehmerInnen „jederzeit“ per Mail oder Telefon erreichen können. Dabei ist es wichtig, von den Betreuungspersonen zeitnahe und qualifizierte Hilfestellungen sowie Bewertungen von abgegebenen Aufgaben zu bekommen. Diese zeitnahen Bewertungen sind entscheidend, um einen anderen Grundsatz des Talentehauses NÖ erfüllen zu können, nämlich die Möglichkeit für die Jugendlichen sich selbst stets verbessern zu können und innerhalb der vorgegebenen Deadlines die E-Learning-Aufgaben mehrmals abgeben zu können. Ein Vergleich zwischen Erstabgabe und Letztabgabe zeigt in vielen Fällen, wie die Jugendlichen innerhalb einer Aufgabe ihre Kompetenzen, Erkenntnisse und Fertigkeiten erweitern können. Als unterstützende Maßnahme werden die Jugendlichen zusätzlich zu ihrer fachlichen Ausbildung psychologisch betreut und ihnen damit unter anderem die Möglichkeit geboten sich beraten zu lassen. Des Weiteren werden die Jugendlichen in Zeit- und Projektmanagement geschult, um ihnen das Werkzeug an die Hand zu geben

sich ihre Zeit und Aufgaben selbst einteilen zu können und damit ihrer Eigenverantwortung gerecht werden zu können.

Der Beziehungsaufbau zwischen den TeilnehmerInnen untereinander und den Betreuungspersonen erfolgt am TalenteHaus NÖ hauptsächlich durch Exkursionen zu namhaften Partnern in Industrie und Wissenschaft und durch Blockwochen, in denen die Jugendlichen die Möglichkeit bekommen einige Tage gemeinsam zu verbringen und Projekte zu erarbeiten. Auch ein heuer zum ersten Mal durchgeführter Versuch zeigte sich erfolgsversprechend. So wurden aus den Lehrgängen Roboterprogrammierung und Roboterkonstruktion gemischte Teams gebildet, die gemeinsam an der ECER 2016 und an zwei damit verbundenen Wettbewerben teilnahmen. Aufbauend auf die im TalenteHaus NÖ gelernten Inhalte eigneten sich die Wettbewerbsteams in Eigenregie neue Kompetenzen mit einem hohen Maß an Begeisterung und Durchhaltevermögen an. Durch diesen Ansatz des eigenverantwortlichen Arbeitens und Lernens belegten sie im PRIA Open insgesamt den 3. Platz, der sich aus dem 3. Platz in den sogenannten Seeding Rounds und dem 2. Platz in den Double Elimination Rounds ergab [Nö16]. Aufgrund ihrer herausragenden Ausdauer und ihres Fleißes erarbeiteten sie in wenigen Tagen die Grundlagen der Unterwasserrobotik. Dadurch gelang es ihnen bei dem PRIA Underwater Wettbewerb den ersten Platz zu erreichen und mit einem Award prämiert zu werden. Zusätzlich halfen sie die Entwicklung des Underwater-Roboters voranzutreiben, wofür die Jugendlichen von der Jury mit dem Judges' choice Award ausgezeichnet wurden. Auch zum Thema wissenschaftliches Arbeiten konnten sich die Jugendlichen des TalenteHauses NÖ hervortun und wurden ausgewählt, ihr englischsprachiges Paper über das zukunftsweisende Thema „Photovoltaic plants – Energy to Infinity“ allen TeilnehmerInnen der ECER 2016 zu präsentieren [PR16].

Ob sich diese beobachteten Effekte auch messtheoretisch nachweisen lassen, gilt es in nächster Zeit systematisch zu erheben und zu berechnen. Die Erarbeitung dieser Fragestellung ist relevant, um in Zukunft Dropouts aufgrund von sinkender Motivation zu verringern.

Literaturverzeichnis

- [Br08] Beer, Rudolf: EVA nach Heinz Klippert - Praxistools zum Methoden- und Kommunikationstraining, Methodenüberblick zum handlungsorientierten Unterricht, 2008.
- [Nö16] NÖ Landesakademie, www.noe-lak.at, Stand: 16.05.2016
- [PW07] Petermann, Franz; Winkel, Sandra.: Fragebogen zur Leistungsmotivation - Harcourt Testservice, 2007.
- [PR16] Practical Robotics Institute Austria (PRIA), <https://pria.at/ecer>, Stand: 16.05.2016
- [St16] Students Motivation in a blended classroom, www.teachthought.com, 16.05.2016

- [SK12] Schober, Alexander; Keller, Lars: Impact factors for learner motivation in Blended Learning environments. International Journal of Emerging Technologies in Learning. Vol.7/2012, 2012.
- [Ta16] Talentehaus NÖ, www.talentehaus.at, Stand: 16.05.2016

Umweltinformatik zwischen
Nachhaltigkeit und Wandel
(UINW 2016)

Environmental Informatics
between Sustainability and
Change

4. Workshop: Umweltinformatik zwischen Nachhaltigkeit und Wandel (UINW 2016)

Stefan Naumann¹, Kristina Voigt²

Abstract: Die Bedeutung der Informatik in den Naturwissenschaften, den Lebenswissenschaften, den Umweltwissenschaften und für die Energiewende ist unbestritten. Umweltinformatik ist ein wichtiges Feld der angewandten Informatik. Dies wurde bereits in den achtziger Jahren erkannt und in der Gründung des Fachausschuss Umweltinformatik 1986 umgesetzt. Ziel des Workshops ist der Erfahrungsaustausch zur Analyse und Lösung von Umweltproblemen mit Hilfe aktueller Methoden der Informatik und der Kommunikationstechnologien.

Keywords: Umweltinformatik, Nachhaltigkeit, Energie, Fachausschuss Umweltinformatik

1 Eingereichte Fachbeiträge

Für den Workshop wurden insgesamt 7 Beiträge eingereicht, die nach eingehender Begutachtung sämtlich als für den Workshop geeignet bewertet wurden. Die Themen reichen von Beiträgen aus dem Bereich der Energie und der Wirtschaft über Fragen der Nachhaltigkeit von Software bis hin zur ökologischen Bewertung der Durchführung einer Informatik-Konferenz. Damit werden kontemporäre Themen der Umweltinformatik vorgestellt, die den Bezug Umwelt – Nachhaltigkeit – Mensch von verschiedenen Seiten her reflektieren. Mit dem 4. Workshop der Serie Umweltinformatik zwischen Nachhaltigkeit und Wandel wird somit die Brücke zwischen Informatik – Mensch und Umwelt geschlagen.

2 Programmkomitee

- Andreas Filler, Hochschule Trier, Umwelt-Campus Birkenfeld
- Peter Fischer-Stabel, Hochschule Trier, Umwelt-Campus Birkenfeld
- Ulrike Freitag, Condat GmbH
- Werner Geiger, Karlsruhe Institute of Technology
- Patrick Hitzelsberger, Luxembourg Institute of Science and Technology

¹ Hochschule Trier, Umwelt-Campus Birkenfeld, Institut für Softwaresysteme, Postfach 1380, D-55761 Birkenfeld, s.naumann@umwelt-campus.de

² Dr. Kristina Voigt, Helmholtz Zentrum München, Institute of Computational Biology, Ingolstädter Landstr. 1, 85764 Neuherberg, kvoigt@helmholtz-muenchen.de

- Eva Kern, Leuphana Universität Lüneburg
- Stefan Naumann, Hochschule Trier, Umwelt-Campus Birkenfeld (Chair)
- Werner Pillmann, International Society for Environmental Protection
- Wolf-Fritz Riekert, Stuttgart Media University
- Kristina Voigt, Helmholtz Zentrum München (Chair)
- Jochen Wittmann, Hochschule für Technik und Wirtschaft Berlin
- Volker Wohlgemuth, Hochschule für Technik und Wirtschaft Berlin

Auswirkungen von Energieeinsparungen auf den Gasbedarf und den Gaspreis in Deutschland bis 2040

Maik Günther¹, Michael Schimpf²

Abstract: Erdgas ist vielfältig anwendbar und kommt in der Wärmeerzeugung, der Warmwasserbereitung, der Stromerzeugung und bei Industrieanwendungen zum Einsatz. Die EU verfolgt ambitionierte Ziele zur Energieeinsparung. Dies hat auch Auswirkungen auf den zukünftigen Bedarf an Erdgas in Deutschland. 2015 wurden in Deutschland 21% des Primärenergiebedarfs mit Erdgas gedeckt. In dieser Arbeit wird der Einfluss der Energieeinsparung auf den Erdgasbedarf in Deutschland bis 2040 untersucht. Hierzu werden drei Szenarien erzeugt, für die u.a. verschiedene Renovierungsquoten im Gebäudebestand und Effizienzsteigerungen in der Industrie unterstellt werden. Anhand dieser drei Bedarfsszenarien werden zudem die Auswirkungen auf die Gasflüsse von und nach Deutschland sowie auf den Gaspreis am Handelspunkt NetConnect Germany (NCG) bis 2040 untersucht. Dabei findet das weltweite Gasmarktmodell WEGA der Stadtwerke München Anwendung.

Keywords: Gasbedarf, Energieeinsparung, Gasmarktmodell, WEGA, NCG, Gaspreis, Gasflüsse

1 Einleitung

Erdgas spielt eine wichtige Rolle in der Energieversorgung und seine Anwendungsfelder sind vielfältig: Wärmeerzeugung, Warmwasserbereitung, Stromerzeugung und Industrieanwendungen. Dem Erdgasbedarf wird das weltweit größte Wachstumspotenzial unter den fossilen Brennstoffen zugeschrieben. So rechnet BP weltweit mit einem jährlichen Wachstum von 1,8% bis 2035 [BP16]. Im Jahr 2013 hatte Erdgas in Europa einen Anteil von 23% am gesamten Primärenergiebedarf [Eu16]. Deutschland ist das europäische Land mit dem größten Erdgasbedarf. Im Jahr 2015 betrug dieser knapp 82 Mrd. m³ [AG16].

Die EU hat ambitionierte Ziele zur Energieeinsparung. So soll der Primärenergiebedarf bis 2020 um 20% [Eu11] gegenüber den ursprünglichen Projektionen sinken und bis 2030 strebt man eine Reduktion um 30% an [Eu14]. Energieeinsparungen sind beispielsweise durch veränderte Prozesse in Gewerbe und Industrie, durch eine bessere Wärmedämmung bei Wohngebäuden oder durch die Erneuerung von Heizungssystemen zu realisieren. Die langfristige Entwicklung von Bedarfstreibern wie Produktion und Energieeffizienz in der Industrie oder Renovierungsquoten im Gebäudebestand sind unsicher. Daher werden in dieser Arbeit drei Szenarien für den Erdgasbedarf in Deutschland bis 2040 erstellt. Ein Basisszenario sowie jeweils ein Szenario mit einem hierzu höheren bzw. einem

¹ Stadtwerke München GmbH, Emmy-Noether-Str. 2, 80992 München, guenther.maik@swm.de

² Stadtwerke München GmbH, Emmy-Noether-Str. 2, 80992 München, schimpf.michael@swm.de

niedrigeren Erdgasbedarf. Methodisch wird der Gasbedarf für die Erzeugung von Strom und Fernwärme (FW) mit Hilfe eines fundamentalen Strommarktmodells berechnet. Für die Sektoren Haushalte, Gewerbe und Dienstleistungen (GuD) oder Industrie werden Studien und Statistiken analysiert sowie Experteninterviews geführt. Aus den gewonnenen Erkenntnissen werden induktive Schlüsse zum zukünftigen Gasbedarf gezogen.

Ausgehend von den drei Szenarien für den Gasbedarf werden in dieser Arbeit auch deren Auswirkungen auf die Gasflüsse von und nach Deutschland bis 2040 aufgezeigt. Zudem wird ihr Einfluss auf die Gaspreise am Handelspunkt NetConnect Germany (NCG) analysiert. Hierzu wird das weltweite Gasmarktmodell WEGA³ der Stadtwerke München GmbH (SWM) eingesetzt. Der jährliche Gasbedarf Deutschlands ist dort ein Eingangsparameter und kann beliebig geändert werden. Diese Jahreswerte werden mit Hilfe von hinterlegten Profilen auf Tageswerte umgerechnet, wobei der Temperaturverlauf eines Durchschnittsjahres, ohne Kältewellen oder besonders milde Winter, zu Grunde liegt.

Der Aufbau dieser Arbeit gestaltet sich wie folgt: In Kapitel 2 werden die drei Szenarien für die Entwicklung des Gasbedarfs in Deutschland bis 2040 beschrieben. Hierbei wird auch auf die Treiber eingegangen, die den Gasbedarf in den Szenarien beeinflussen. In Kapitel 3 wird das Gasmarktmodell WEGA erläutert, mit dem anschließend in Kapitel 4 die Berechnungen zu den Gaspreisen und Gasflüssen durchgeführt werden. Die Ergebnisse werden ebenfalls in Kapitel 4 erörtert. Diese Arbeit endet mit einer Zusammenfassung.

2 Gasbedarf in Deutschland bis 2040

In diesem Abschnitt werden die drei Szenarien für die Entwicklung des Gasbedarfs in Deutschland bis 2040 vorgestellt. Hierbei handelt es sich zum einen um ein Basisszenario, welches die zentrale Sicht darstellt. Um Sensitivitäten mit WEGA zu rechnen, werden vom Basisszenario ein Hochszenario mit einem höheren sowie ein Tiefszenario mit einem niedrigeren Gasbedarf abgeleitet. Die Erstellung der drei Szenarien erfolgt dabei getrennt nach den folgenden Sektoren, die nun erläutert werden:

- Erzeugung von Strom- und FW,
- Sonstiges,
- Haushalte,
- Gewerbe und Dienstleistung (GuD),
- Industrie.

Im Sektor für die Erzeugung von Strom und FW werden analog zu Eurostat [Eu16]

³ WEGA steht für Weltweites GAsmarktmodell.

Erzeugungsanlagen von Energieversorgern sowie die Eigenerzeugung der Industrie berücksichtigt. Der Gasbedarf für Erzeugungsanlagen von Energieversorgern wird mit Hilfe des fundamentalen Strommarktmodells der SWM errechnet [SRF14] und auf die Datenbasis von Eurostat hochgerechnet [Eu16], da das fundamentale Strommarktmodell beispielsweise nicht alle KWK-Anlagen enthält. Eigenerzeugungsanlagen der Industrie sind dort ebenfalls nicht abgebildet. Daher wird ihr Gasbedarf aus der Entwicklung bei den Energieversorgern sowie bei der Industrieproduktion abgeleitet. Der so ermittelte Gasbedarf wird für das Hoch- und Tiefszenario unverändert übernommen. Zwar würde beispielsweise der Gasbedarf aus dem Sektor für Strom und FW bei einem höheren Gaspreis im Hochszenario leicht sinken. Dieser Effekt ist jedoch gering, da die Preisunterschiede zwischen dem Basisszenario und dem Hoch- sowie Tiefszenario verhältnismäßig gering ausfallen (siehe Ergebnisse in Kapitel 4). Bis 2040 erhöht sich der Bedarf aus dem Sektor für die Erzeugung von Strom- und FW um 50%.

Der Sektor Sonstiges umfasst u.a. den Transportbereich auf der Straße und zur See. Zum Transportbereich wird aber auch der Transport von Erdgas selbst gezählt. Für Deutschland ist dies nur der Transport per Pipeline, da Deutschland über kein LNG-Importterminal verfügt und der Bau eines Terminals nicht wahrscheinlich ist. Der Sektor Sonstiges umfasst auch den Eigenverbrauch bei der Förderung von Gas sowie den Gasbedarf in Raffinerien. Bei diesem Sektor wurde der Gasbedarf zunächst für die drei Szenarien separat ermittelt. Da das Gesamtniveau gering ist, fallen Unterschiede zwischen den drei Szenarien nicht ins Gewicht. Zur Vereinfachung wurde der Bedarfsverlauf aus dem Basisszenario auch für das Hoch- und Tiefszenario übernommen. Er steigt von 2016 bis 2040 um 3% an.

Im Haushaltssektor gibt es zahlreiche Bedarfstreiber für Gas, die die Raumwärme- und Heißwasserbereitung betreffen. Entwicklungen bei Kochgas können wegen ihres geringen Anteils vernachlässigt werden. Der Bedarf an Kochgas wird daher konstant fortgeführt. Ein relevanter Bedarfstreiber ist die Bevölkerungszahl. Eine plausible Entwicklung kann für diese Arbeit aus Daten des Statistischen Bundesamtes abgeleitet werden [St15]. Im Jahr 2040 werden 77,4 Mio. Einwohner in Deutschland unterstellt. Auf den Gasbedarf hat ebenfalls der Gebäudetyp einen Einfluss [Sh13]. Hierzu gehören die Typen Einfamilienhaus (EFH), Zweifamilienhaus (ZFH) und Mehrfamilienhaus/Apartmenthaus (MFH), wobei ihr Energiebedarf auch vom Alter des Gebäudes abhängt. Historische Daten zum Gebäudebestand, auf denen dann Abriss- und Renovierungsquoten aufsetzen, stammen vom Statistischen Bundesamt [St14b]. Die derzeitigen Abrissquoten je Jahr von 0,2% für EFH/ZFH und 0,3% für MFH werden bis 2040 fortgeführt [BB15]. Da ein steigender Bedarf nach neuen Wohnungen bis 2030 besteht [BB15], wurden jährliche Neubauquoten für EFH/ZFH von 0,7% bis 2030 und 0,3% ab 2030 sowie für MFH von 0,6% bis 2030 und 0,2% ab 2030 unterstellt. Bei den Renovierungsquoten wird von vollständigen Renovierungen (äquivalent) ausgegangen [KDG12], [EHB13]. Die Ergebnisse sind in Tabelle 1 aufgeführt. Während alle zuvor genannten Bedarfstreiber in den Szenarien Basis, Hoch und Tief eine identische Entwicklung haben, gibt es bei den Renovierungsquoten Unterschiede. Hiermit soll in den Szenarien beispielsweise die Wirkung von verschiedenen Maßnahmen der Politik zur Energieeinsparung nachgebildet

werden. Ebenfalls wichtig für den Gasbedarf ist die verwendete Heizungstechnologie. Gasheizungen hatten 2014 einen Anteil von etwa 48% [BD14]. Bei Neubauten waren es in 2014 53% für EFH/ZFH sowie 55% für MFH [St14a]. Zwar werden Gasheizungen durch andere Technologien wie Wärmepumpen ersetzt, doch besonders Ölheizungen werden andererseits von Gasheizungen verdrängt [Sh13]. Das jährliche Wachstum des Anteils von Gasheizungen im Gebäudebestand (renoviert wie auch unrenoviert) ist in Tabelle 1 ersichtlich.

Szenario	Parameter	Bis 2030	Ab 2030
Basis	Renovierungsquote EFH/ZFH	1,3%	2,0%
Hoch	Renovierungsquote EFH/ZFH	0,5%	1,0%
Tief	Renovierungsquote EFH/ZFH	2,0%	2,5%
Basis	Renovierungsquote MFH	1,7%	2,0%
Hoch	Renovierungsquote MFH	0,8%	1,0%
Tief	Renovierungsquote MFH	2,0%	2,5%
Basis	Anteil Gasheizungen	0,3%	0,1%
Hoch	Anteil Gasheizungen	0,4%	0,2%
Tief	Anteil Gasheizungen	0,0%	-0,2%

Tab. 1: Jährliche prozentuale Änderung der Bedarfstreiber im Haushaltssektor

Abbildung 1 zeigt die Entwicklung des Gasbedarfs im Haushaltssektor. Im Basisszenario sinkt der Gasbedarf von 2016 bis 2040 um 26%, im Hochszenario um 7% und im Tiefszenario um 41%. Der starke Rückgang des Gasbedarfs im Tiefszenario basiert alleinig auf den Annahmen, dass jährliche Renovierungsquoten von 2,0% und ab 2030 sogar 2,5% umgesetzt werden können. Zudem wird unterstellt, dass Gasheizungen bis 2030 keine Zuwächse im Gebäudebestand erzielen und danach vor allem von Wärmepumpen verdrängt werden.

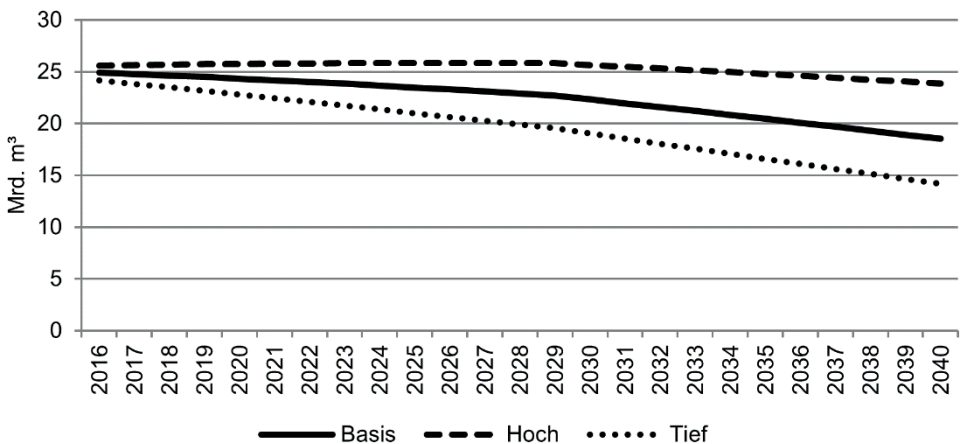


Abb. 1: Gasbedarf für Haushalte in Deutschland in den Szenarien Basis, Hoch und Tief

Die Gebäude im GuD-Sektor werden in dieser Arbeit als MFH betrachtet. Somit können die Daten für die Bedarfstreiber der MFH aus dem Haushaltssektor auf den GuD-Sektor übertragen werden. Dies gilt für die Raumwärme und die Warmwasserbereitung. Der Anteil der Prozesswärme ist bei GuD gering und wird daher konstant fortgeführt. Das Ergebnis ist in Abbildung 2 dargestellt. Der Gasbedarf bis 2040 sinkt im Basisszenario um 29%, im Hochszenario um 9% und im Tiefszenario um 44%.

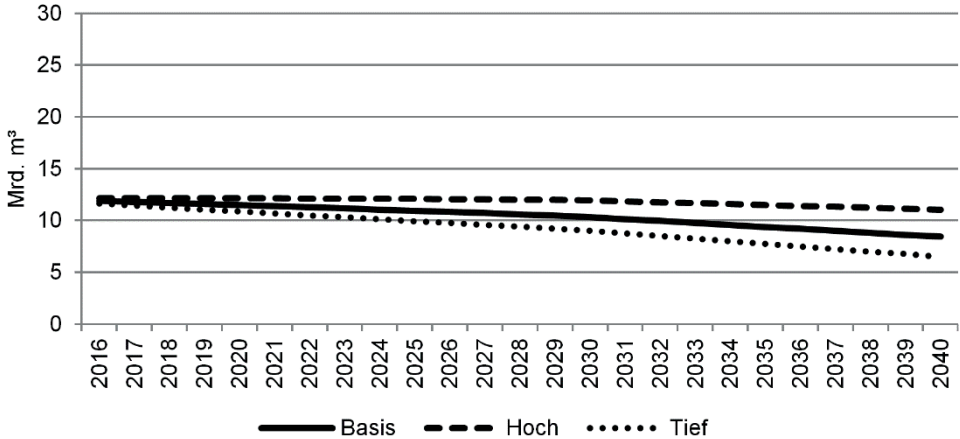


Abb. 2: Gasbedarf für GuD in Deutschland in den Szenarien Basis, Hoch und Tief

Der Industriesektor wird zur Projektion des zukünftigen Gasbedarfs in Subsektoren zerlegt, um die individuellen Besonderheiten der Subsektoren besser berücksichtigen zu können. Hierzu zählen die Subsektoren: Chemie/Petrochemie, Ernährung und Tabak, Eisen und Stahl, Nichteisenmetalle, nichtmetallische Mineralien, Maschinenbau, Papierindustrie, Fahrzeugbau sowie Sonstiges. Die Bildung der Subsektoren orientiert sich an der Branchenaufteilung des Statistischen Bundesamtes [St08] sowie des Umweltbundesamtes [Fe13]. Anhand dieser Quellen sowie aus den Veröffentlichungen von AGEB [AG13], [AG15], Eurostat [Eu16], Prognos [Pr13], Remus et al. [Re13], Schlesinger et al. [Sc14] und Wörtler et al. [Wö13] wurden individuell für jeden Subsektor die jährlichen Entwicklungsraten der Produktion und der Energieintensität mit Bezug auf den Erdgaseinsatz bis 2040 abgeleitet. So führt ein Anstieg der Produktion zwar zu einem höheren Erdgasbedarf, ein Anstieg der Energieeffizienz dämpft diesen Effekt jedoch. Das Basisszenario enthält genau diese Ableitungen. Im Hochszenario ist die Veränderung der jährlichen Energieintensität mit Bezug auf den Erdgaseinsatz je Subsektor jeweils um 0,7% nach oben korrigiert, wogegen sie im Tiefszenario jeweils um 0,7% nach unten korrigiert ist. Abbildung 3 zeigt die Entwicklung des Gasbedarfs im Industriesektor. Dieser sinkt im Basisszenario bis 2040 um 17%. Der Gasbedarf im Hochszenario verläuft fast konstant und steigt bis 2040 um 3%. Im Tiefszenario sinkt der Gasbedarf bis 2040 um 37%.

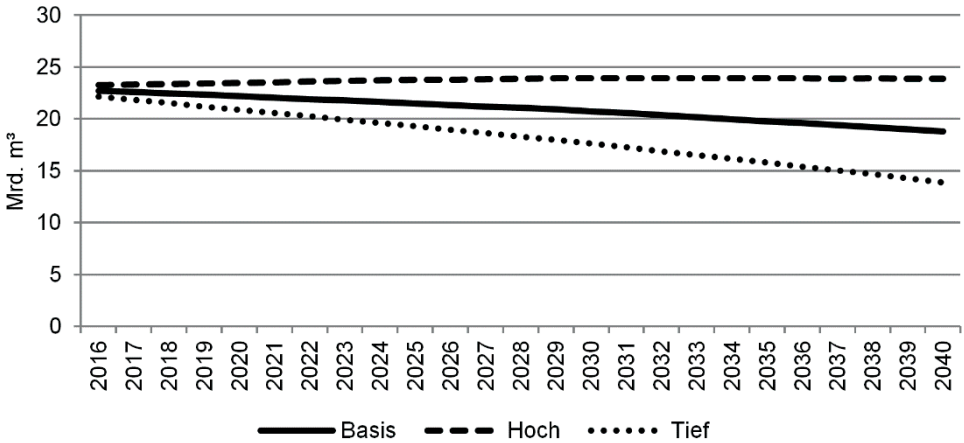


Abb. 3: Gasbedarf für Industrie in Deutschland in den Szenarien Basis, Hoch und Tief

In Abbildung 4 ist der Verlauf des jährlichen Gasbedarfs für alle drei Szenarien bis 2040 dargestellt. Man erkennt, dass sich die Gasbedarfsentwicklungen im Hoch- und Tiefszenario nur langsam vom Gasbedarfsverlauf im Basisszenario entfernen. Es braucht viele Jahre, bis Gebäude renoviert sind, ein signifikanter Anteil an Neubauten errichtet ist und sich neue Technologien durchgesetzt haben. Für das Basis- und Tiefszenario gilt, dass der Rückgang des Gasbedarfs bei Haushalten, GuD, Industrie und Sonstiges nicht vom Kraftwerks- und FW-Sektor kompensiert werden kann. Der Bedarf sinkt bis 2040 im Basisszenario um 5% und im Tiefszenario um 18%. Im Hochszenario steigt er um 8%.

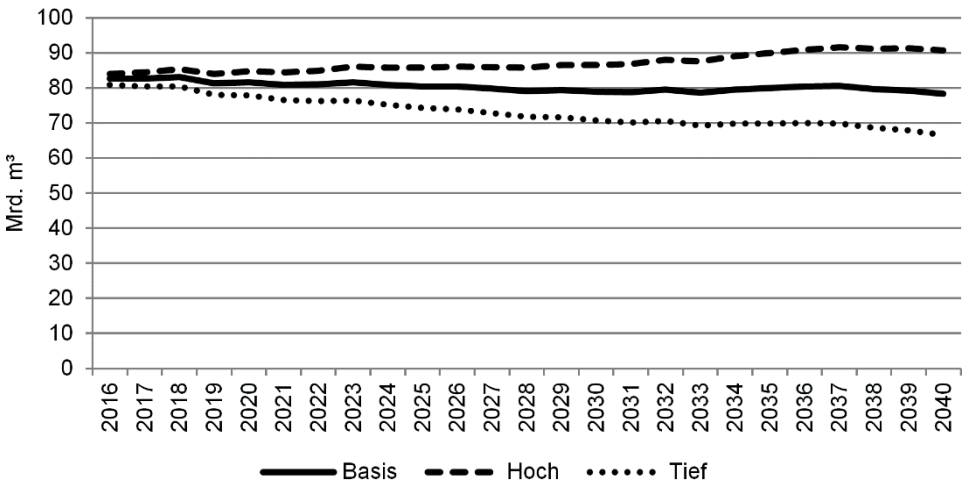


Abb. 4: Gasbedarf für Deutschland in den Szenarien Basis, Hoch und Tief

Für das Basisszenario ist der Gasbedarf bis 2040 getrennt nach Sektoren in Abbildung 5 dargestellt. Ab 2030 hat der Kraftwerks- und FW-Sektor einen größeren Anteil am

Gesamtbedarf als der Industriesektor. Die langfristig wachsende Erzeugung von Strom aus Erdgas führt zu diesem Anstieg, während alle anderen Sektoren einen rückläufigen Gasbedarf bis 2040 haben. Der Anstieg im Kraftwerks- und FW-Sektor basiert auf Annahmen zum Mindestausbaupfad für Erneuerbare, zur Stromnachfrage, zu Brennstoffpreisen, etc. Andere Annahmen bewirken einen abweichenden Verlauf. Ein sehr hoher Gaspreis würde beispielsweise zu einer geringeren Gasnachfrage im Kraftwerks- und FW-Sektor führen, da unter sonst gleichen Annahmen mehr Erneuerbare zugebaut werden.

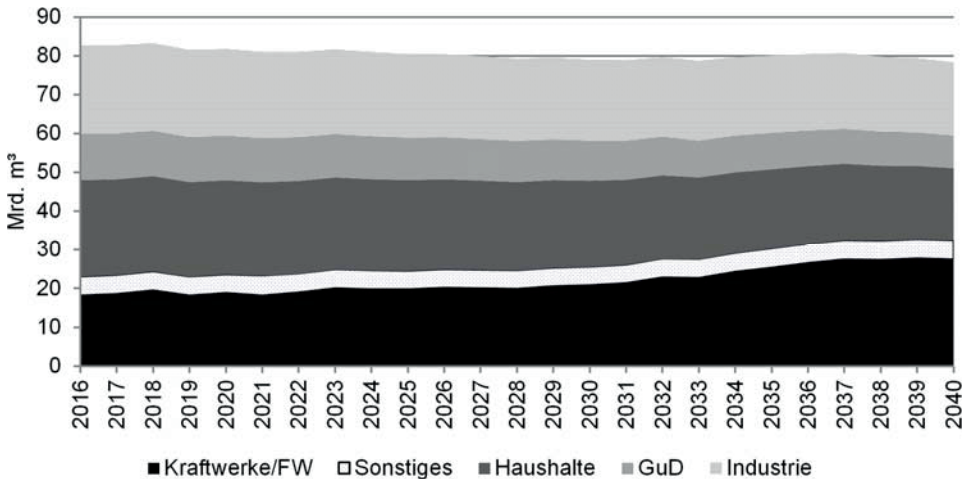


Abb. 5: Gasbedarf für Deutschland im Basisszenario nach Sektoren

3 Weltweites Gasmarktmodell

Bei WEGA ergeben sich die Gasflüsse und die daraus resultierenden Gaspreise unter perfekter Vorausschau und vollständigem Wettbewerb, wobei zahlreiche Nebenbedingungen zu beachten sind. Marktmacht von Produzenten sowie Investitionsentscheidungen sind fest vorgegeben und werden nicht vom Modell bestimmt. Zur Abbildung von Marktmacht in WEGA können jedoch Szenarien erstellt werden. Hierbei werden beispielsweise Preisvorgaben von Produzenten und ein etwaiges Zurückhalten von Mengen als Eingangsdaten im Modell hinterlegt. Mit WEGA kann ein Szenario bis 2040 in täglicher Auflösung mit einem detaillierten Pipelinenetz in etwa 25 Minuten gerechnet werden. Das Konzept von WEGA entspricht in den Grundzügen dem Gasmarktmodell TIGER des EWI [HJW14]. Für einen Überblick über Gasmarktmodelle sei auf die Arbeiten von Chyong und Hobbs [CH14], Holz et al. [HHK08] sowie auf die Arbeit des Energy Modeling Forums [EM07] verwiesen.

WEGA basiert auf Linearer Programmierung (LP) und wurde mit Hilfe der FICO Xpress Optimization Suite erstellt. Der Quellcode des Modells wurde von Pöyry Management Consulting (UK) Ltd. entwickelt und 2013 von den SWM erworben. Das kommerzielle Modell von Pöyry trägt den Namen Pegasus [DaSa04], [Pö12]. Die SWM haben jedoch auch die zugehörige Datenbank von Pöyry erworben und mit eigenen Daten verändert. Zu den SWM-eigenen Annahmen zählt z.B. der Gasbedarf aus dem fundamentalen Strommarktmodell der SWM. Zudem nutzen die SWM öffentlichen Quellen sowie Daten kommerzieller Anbieter. So haben die Unternehmen PIRA Energy Group, Wood Mackenzie, IHS und Bloomberg New Energy Finance z.T. eigene Gasmarktmodelle und stellen den SWM relevante Daten kostenpflichtig zur Verfügung. Für Plausibilitätsprüfungen werden regelmäßig die Future-Preise relevanter Märkte nachgerechnet. Zudem werden die Parameter und Ergebnisse mit Experten diskutiert.

WEGA ist zwar ein weltweites Gasmarktmodell, hat aber einen Fokus auf Europa. Daher ist Europa besonders detailliert abgebildet. Die Notwendigkeit für ein weltweites Gasmarktmodell besteht vor dem Hintergrund, dass durch den Transport von Flüssiggas (LNG) per Schiff ein weltweiter Handel von Erdgas möglich ist. Laut BP wird LNG bis 2035 einen größeren Anteil am weltweiten Gashandel haben als Pipelinegas [BP16]. Der Transport von Erdgas ist also nicht nur an Pipelines mit festen Anfangs- und Endpunkten geknüpft. Modelle, die z.B. nur Nordamerika oder Europa betrachten und LNG-Flüsse fest vorgeben, bilden die Realität daher nicht ausreichend ab.

Der weltweite Gasmarkt ist in WEGA als ein Netz aus verschiedenen Typen von Knoten und Kanten aufgebaut. Bei den Knoten unterscheidet man Bedarfszonen (Länder, Regionen oder Handlungspunkte, z.B. NCG), Gasfelder innerhalb der Bedarfszonen, Gasfelder mit Pipelineanbindung, Anlandepunkte für Pipelines (z.B. Greifswald für Nord Stream), Gasfelder mit Anbindung an LNG-Exportterminals, LNG-Importterminals, Gasspeicher sowie Flexibilitätsoptionen innerhalb einer Bedarfszone (z.B. Coal-Gas-Switch, abschaltbare Kunden, Lastabwürfe zur Vermeidung einer Unlösbarkeit des Optimierungsproblems). Für die Verbindung zwischen den verschiedenen Arten von Knoten gibt es wiederum eine Vielzahl an unterschiedlichen Kanten. Zwischen zwei Bedarfszonen gibt es beispielsweise eine Kante, die einen Interkonnektor darstellt. Sie erlaubt einen Gasfluss in eine vorgegebene Richtung. Für einen Interkonnektor, der Gasflüsse in beide Richtungen ermöglicht, werden daher zwei Kanten erzeugt.

Knoten und Kanten haben je nach Typ unterschiedliche Eigenschaften. Hierzu zählen Kapazitätsrestriktionen, Kosten, Profile für die Produktion in täglicher Auflösung, Ein- und Ausspeicherleistungen. WEGA enthält zudem eine Datenbank langfristiger Gaslieferverträge mit Take-or-Pay Klauseln, Flexibilitätsoptionen und Preisformeln. Wegen der Vielzahl an Daten ist die Business Intelligence Software Tableau in das Gasmarktmodell integriert. Mit ihr können direkt in der Modelloberfläche Eingangsdaten und Rechenergebnisse analysiert werden.

WEGA optimiert jedes Gasjahr separat unter perfekter Vorausschau. Dabei wird das Gasjahr in einzelne Tage ‚d‘ zerlegt. Die Zielfunktion ist in Formel 1 dargestellt. Hier gilt

es, die Gesamtkosten ‚KGesamt‘ eines Gasjahres zu minimieren. Diese setzen sich aus den tagesscharfen Produktionskosten ‚KProd_d‘, Transportkosten ‚KTransp_d‘, Speicherkosten ‚KSpei_d‘ sowie etwaigen Flexibilitätskosten ‚KFlex_d‘ zusammen.

$$\text{Min KGesamt} = \sum_{d=1}^D (K\text{Prod}_d + K\text{Transp}_d + K\text{Spei}_d + K\text{Flex}_d) \quad (1)$$

Eine harte Nebenbedingung ist die Deckung des Gasbedarfs jeder Bedarfszone ‚z‘ an jedem Tag ‚d‘. Nach Formel 2 wird der tägliche Gasbedarf einer Bedarfszone ‚z‘ durch die Zu- und Abflüsse der zur Bedarfszone gehörenden Pipelines, LNG-Terminals, Interkonnektoren, Speicher und Flexibilitätsoptionen gedeckt. Weiterhin gibt es Zuflüssen durch Eigenproduktion innerhalb der Bedarfszone. In Formel 2 meint ‚Ein‘ jeweils einen Zufluss zur Bedarfszone und ‚Aus‘ einen Abfluss.

$$\begin{aligned} \text{Bedarf}_{z,d} = & \text{PipelineEin}_{z,d} - \text{PipelineAus}_{z,d} + \\ & \text{LNGEin}_{z,d} - \text{LNGOut}_{z,d} + \\ & \text{InterkonnektorEin}_{z,d} - \text{InterkonnektorAus}_{z,d} + \\ & \text{SpeicherEin}_{z,d} - \text{SpeicherAus}_{z,d} + \\ & \text{FlexibilitätsoptionEin}_{z,d} + \text{EigenproduktionEin}_{z,d} \quad \forall z \in Z, d \in D \end{aligned} \quad (2)$$

Zusätzliche Nebenbedingungen betreffen die zuvor genannten zahlreichen Eigenschaften der Knoten, Kanten und Verträge. Beispielsweise darf die maximale Kapazität einer Pipeline oder eines LNG-Importterminals nicht überschritten werden und vertraglichen Vereinbarungen aus den langfristigen Lieferverträgen sind einzuhalten. Für weitere Ausführungen zu WEGA sei auf [Gül16] verwiesen.

4 Experimente und Ergebnisse

Bei den folgenden Berechnungen wird jeweils nur der Gasbedarf in Deutschland geändert. Passt man zusätzliche Parameter wie beispielsweise die Produktionskosten, den Ölpreis oder die weltweite LNG-Exportkapazität an, so ergeben sich andere NCG-Preise.

Wegen interner Restriktionen sind nachfolgend keine NCG-Preise dargestellt. Abbildung 6 zeigt jedoch die prozentualen Abweichungen der NCG-Preise für das Hoch- und Tief-szenario im Vergleich zum Basisszenario von 2020 bis 2040. Der Unterschied im Gasbedarf zwischen Tief- und Basisszenario wird bis 2040 immer größer (z.B. -14,1% in 2040), die jährlichen NCG-Preise des Tiefszenarios sinkt jedoch in keinem Jahr um mehr

als 3,1% gegenüber dem Basisszenario. Dies liegt daran, dass der nordwesteuropäische Gasmarkt im Basisszenario bereits sehr gut versorgt ist und Zugriff auf eine ausreichende Zahl relativ günstiger Quellen hat. Eine Absenkung des Gasbedarfs führt somit nur zu einer unterproportionalen Preisreduktion. Die NCG-Preise im Hochszenario entfernen sich in den 20ern weniger stark von Basisszenario als das Tiefszenario. Wiederum dämpft die sehr gute Versorgung des Gasmarktes zunächst die Preise. Ab Anfang der 30er Jahre ist der Bedarfsanstieg im Hochszenario jedoch so stark, dass zunehmend Gas aus teureren Quellen benötigt wird und die NCG-Preise im Vergleich zum Basisszenario stärker ansteigen. Grundsätzlich kann festgehalten werden, dass die NCG-Preise nicht sehr sensitiv auf Änderungen des Gasbedarfs in Deutschland reagieren, solange der Bedarf innerhalb der Grenzen des Hoch- und Tiefszenarios liegt.

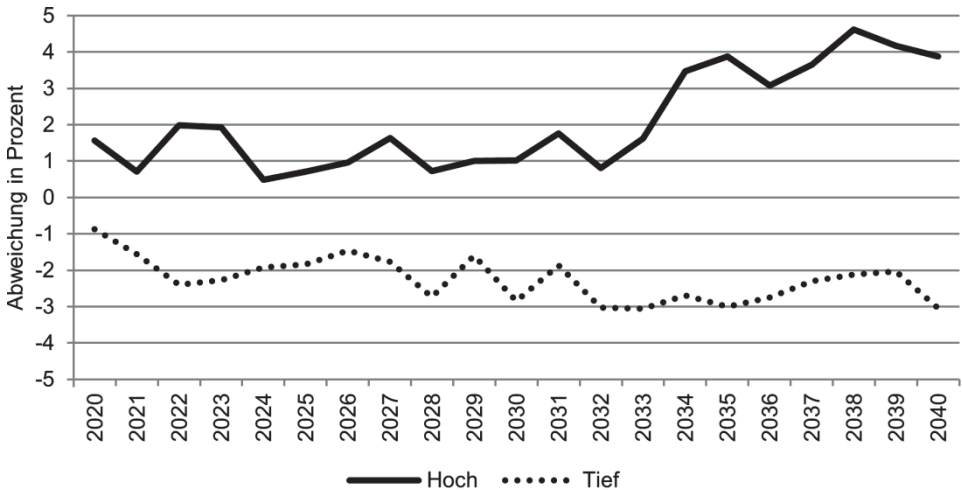


Abb. 6: Abweichungen der jährlichen NCG-Preise in Bezug auf das Basisszenario in Prozent

Abbildung 7 zeigt die Gasflüsse von und nach Deutschland in den drei Szenarien. Die einheimische Produktion sowie Gasflüsse in und aus einheimischen Speichern sind nicht dargestellt. Mit Transitgas sind vor allem Durchleitungen von Ost nach West z.B. nach Frankreich gemeint. Im Basis- und Hochszenario sind die Transitgasflüsse annähernd identisch und nehmen in Zukunft leicht ab. Lediglich im Tiefszenario steigen sie Ende der 30er Jahre durch vermehrte Exporte in die Schweiz an. In allen drei Szenarien nehmen Importe aus Russland, Polen, Tschechien und Österreich ab Mitte der 20er Jahre zu. Hierbei handelt es sich um direkte Gaslieferungen aus Russland über die Nord Stream Pipeline sowie um Gaslieferungen aus Russland und dem kaspischen Raum, welche über Polen, Tschechien und Österreich nach Deutschland gelangen. Importe aus Norwegen gehen in allen drei Szenarien ab Mitte der 20er Jahre zurück, da die Produktion in Norwegen ab diesem Zeitpunkt langsam sinkt und das Gas dann zunehmend in andere westeuropäische Staaten exportiert wird. Importe aus Westeuropa gehen im Basisszenario langsam zurück. Vor allem weil die Produktionsmengen aus dem Gasfeld Groningen in den Niederlanden sinken. Dieser Trend wird im Hochszenario etwas aufgefangen, da

durch zusätzliche LNG-Importe in die Niederlande und Belgien mehr Gas nach Deutschland exportiert werden kann. Im Tiefszenario sind Gasimporte aus Westeuropa ab 2032 nicht mehr erforderlich, um den Gasbedarf in Deutschland zu decken. In Summe zeigen sich vor allem bei den Importen aus Westeuropa sowie aus Russland, Osteuropa und Österreich größere Unterschiede zwischen den drei Szenarien. Norwegische Importe gehen in allen Szenarien fast gleichmäßig zurück und Exporte aus Deutschland steigen nur im Tiefszenario Ende der 30er Jahre stärker an als in den anderen beiden Szenarien.

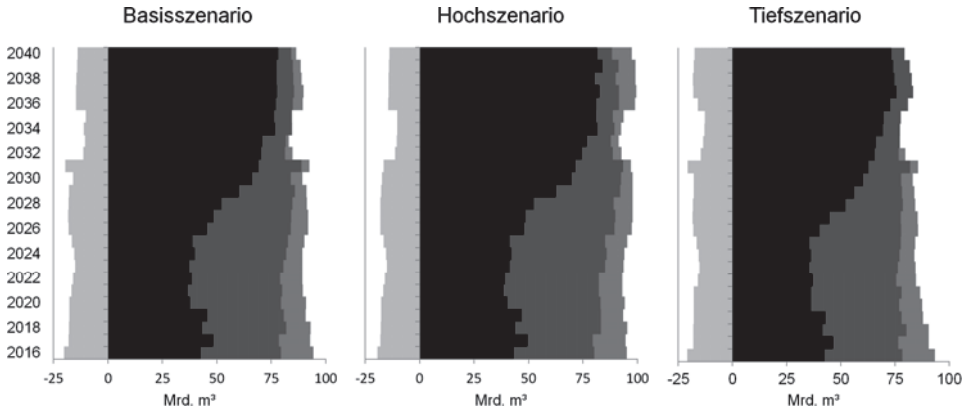


Abb. 7: Gasflüsse von und nach Deutschland für die Szenarien Basis, Hoch und Tief

5 Zusammenfassung

In dieser Arbeit wurden drei Szenarien für den Gasbedarf in Deutschland bis 2040 vorgestellt. Ausgehend von einem Basisszenario, welches die bestmögliche Sicht darstellt, wurden ein Hoch- sowie ein Tiefszenario entwickelt. Der Gasbedarf in allen drei Szenarien wurde dabei getrennt nach den Sektoren Haushalte, GuD, Industrie, Kraftwerke und FW sowie Sonstiges erstellt. Es wurde deutlich, dass sich die Gasbedarfe der drei Szenarien nur langsam voneinander entfernen. Es müssen erst Jahre vergehen, bis Maßnahmen zur Energieeinsparung in größerem Stil umgesetzt sind.

Ausgehend von den Bedarfsszenarien wurden Gaspreise und Gasflüsse mit Hilfe des Gasmarktmodells WEGA berechnet. Es hat sich gezeigt, dass die NCG-Preise der Szenarien nur langsam auseinander driften. Bei den Gasflüssen gibt es die größten Unterschiede bei den Importen von Gas aus dem russischen und kaspischen Raum sowie bei den LNG-Importen nach Westeuropa.

Aktuell finden Untersuchungen hinsichtlich der saisonalen Effekte von Einsparmaßnahmen statt. Zudem sollen zukünftig neben Deutschland weitere Länder untersucht werden.

Literaturverzeichnis

- [AG13] AGEB: Anwendungsbilanzen für die Endenergiesektoren in Deutschland in den Jahren 2011 und 2012 mit Zeitreihen von 2008 bis 2012, Nr. 23/11, 2013.
- [AG15] AGEB: Ausgewählte Effizienzindikatoren zur Energiebilanz Deutschland, 2015.
- [AG16] AGEB: Energieverbrauch in Deutschland im Jahr 2015, 2016.
- [BB15] BBSR: Wohnungsmarktprognose 2030, BBSR-Analysen 07, 2015.
- [BD14] BDEW: Wie heizt Deutschland?, 2015.
- [BP16] BP: Energy Outlook 2035, 2016.
- [CH14] Chyong, C. K.; Hobbs, B. F.: Strategic Eurasian Natural Gas Market Model for Energy Security and Policy Analysis. *Energy Economics*, 44/09, S. 198-211, 2014.
- [DaSa04] Davies, G.; Sarsfield-Hall, R.: Gas SCR – Cost Benefit Analysis for a Demand-Side Response Mechanism. A report to Ofgem, Pöyry, 2004.
- [EHB13] Enseling, A.; Hinz, E.; Born, R.: Energetische Sanierung des Gebäudebestandes privater Haushalte, 2013.
- [EM07] EMF: Prices and Trade in a Globalizing Natural Gas Market, EMF Report 23, 2007.
- [Eu11] Europäische Kommission: Energieeffizienzplan 2011, KOM(2011) 109, 2011.
- [Eu14] Europäische Kommission: Pressemitteilung, IP/14/856, 2014.
- [Eu16] Eurostat Database, <http://ec.europa.eu/eurostat/data/database>, Stand 11.04.2016.
- [Fe13] Felix, C. M.; et al.: Politikszenerarien für den Klimaschutz VI, FB 1730, 2013.
- [Gü16] Günther, M.: Practical Application of a Worldwide Gasmarket Model at Stadtwerke München. In: (Dörner, K. u.a., Hrsg.): OR Proc. (OR 2015), Springer, im Druck, 2016.
- [HHK08] Holz, F.; von Hirschhausen, C.; Kemfert, C.: A Strategic Model of European Gas Supply (GASMOD). *Energy Economics*, 30/08, S. 766-788, 2008.
- [HJW14] Hecking, H.; John, C.; Weiser, F.: An Embargo of Russian Gas and Security of Supply in Europe, EWI, 2014.
- [KDG12] Krauß, N.; Deilmann, C.; Gruhler, K.: Wo steht der deutsche Gebäudebestand energetisch? Kurzberichte aus der Bauforschung, 53/12, S.40-50, 2012.
- [Pö12] Pöyry: How will intermittency change Europe's gas markets?, 2012.
- [Pr13] Prognos: Die deutsche chemische Industrie 2030, VCI-Prognos-Studie, 2013.
- [Re13] Remus, R.; et al.: Best Available Techniques (BAT) Reference Document for Iron and Steel Production, JRC Reference Report, 2013.
- [Sc14] Schlesinger, M.; et al.: Entwicklung der Energiemärkte, Nr. 57/12, 2014.
- [Sh13] Shell: Wärmestudie im Wohnungssektor - wie heizen wir morgen?, 2013.
- [SRF14] Schaber, K.; Roth, H.; Fallahnejad, M.: Can the Gas Sector Provide the Flexibility to the

Power Sector for the Integration of Renewables? In: 13th WIW, S. 46-51, 2014.

- [St08] Statistisches Bundesamt: Gliederung der Klassifikation der Wirtschaftszweige, 2008.
- [St14a] Statistisches Bundesamt: Bautätigkeit, Fachserie 5, Reihe 1, 2014.
- [St14b] Statistisches Bundesamt: Bestand an Wohnungen, Fachserie 5, Reihe 3, 2014.
- [St15] Statistisches Bundesamt: Bevölkerung Deutschlands bis 2060, 2015.
- [Wö13] Wörtler, M.; et al.: Steel's Contribution to a Low-Carbon Europe 2050, 2013.

IT-Unterstützung von BHKW-Prüfständen in der angewandten Forschung

Dominik Schöner,¹ Richard Pump,¹ Christian Schmicke,² Jan P. Minnrich,² Henrik Rüscher,² Volker Ahlers,¹ Arne Koschel¹

Abstract: In modernen Forschungsprüfständen spielt Software bei deren Überwachung und Steuerung sowie bei der Analyse von Messdaten eine große Rolle. Das dynamische Anforderungsumfeld und die interdisziplinäre Zusammenarbeit, sowie die Integration domänenspezifischer Werkzeuge und Komponenten stellen dabei hohe Ansprüche an Flexibilität, Wartbarkeit und Nutzerfreundlichkeit von Softwarearchitektur und Systemkomponenten.

Die in diesem Beitrag vorgestellte Architektur löst diese Herausforderungen mittels eines serviceorientierten Ansatzes und einer klaren Aufteilung in drei Schichten: Einer hardwarenahen Ebene zur Anbindung der Sensoren und Aktoren des Prüfstands, dem virtuellen LabVIEW-Leitstand als Client und der diese verbindenden Serviceinfrastruktur mit Geschäftslogik und Datenhaltung.

Keywords: Elektromobilität, Mobiles Mikro-BHKW, Prüfstand, Embedded, RTOS-UH, SOA, LabVIEW, Automatisierung, KWKK

1 Einleitung

Mit verschiedenen Initiativen soll die Bundesrepublik Deutschland zum „Leitmarkt und Leitanbieter für Elektromobilität“ [RB16] werden, steht jedoch vor Diskussionen um die nicht ausreichend vorhandene Ladeinfrastruktur und die vergleichsweise geringe Reichweite von Elektrofahrzeugen. Durch die Integration von Range Extendern (RE) in reinen Elektrofahrzeugen als zusätzliches, verbrennungsmotorisches Aggregat zur Ladung der Traktionsbatterie, kann die Reichweite von Elektrofahrzeugen deutlich erhöht werden. RE stellen eine Brückentechnologie dar, deren Einsatzschwerpunkt sich von der Unterstützung des Antriebs hin zur überwiegenden Nutzung im Thermo- und Energiemanagement des E-Fahrzeugs verlagert, um deren Energieeffizienz zu erhöhen und ihren Betrieb so ressourcenschonender und CO₂-effizienter zu gestalten. Zudem kann die Nachhaltigkeit von RE und vergleichbaren Konzepten durch den Einsatz von Kraftstoffen aus nachwachsenden Rohstoffen zusätzlich erhöht werden.

Bei der Weiterentwicklung von RE steht die thermische Konditionierung (Kühlen und Heizen) sowohl des Fahrgastinnenraums als auch der Batterie im Fokus, da dies nicht nur den Komfort für den Fahrgast erhöht, sondern verbessert durch Vorkonditionierung auch die Batteriereichweite, wie in Simulationen [Mi14] nachgewiesen. Neue Konzepte

¹ Hochschule Hannover, Fakultät IV, Abteilung Informatik, Ricklinger Stadtweg 120, 30459 Hannover, Dominik.Schoener@hs-hannover.de

² Hochschule Hannover, Fakultät II, Abteilung Maschinenbau, Bismarckstraße 2, 30173 Hannover

für verbrennungsmotorische RE setzen auf die Kraft-Wärme-Kopplung (KWK) und können daher auch als Mikro-Blockheizkraftwerk (Mikro-BHKW) bezeichnet werden.

Seit 2013 wird an der Hochschule Hannover der Ansatz verfolgt, eine Nutzung des REs als mobiles mikro-BHKW (mmBHKW) zur dezentralen KWK oder Kraft-Wärme-Kälte-Kopplung (KWKK), auch im Gebäude, zu ermöglichen [Rü14]. Durch die Bereitstellung von thermischer Energie zum Heizen oder Kühlen zusätzlich zum Drehmoment lässt sich so der Gesamtwirkungsgrad der im Kraftstoff gespeicherten Energie erhöhen [Mi15] und somit die CO₂-Bilanz verbessern. Die Skalierung des BHKWs zu kompakteren, mobil im Fahrzeug integrierten Aggregaten, die auch in anderen Bereichen (z. B. im Freizeitbereich oder der Gebäudetechnik) angewendet werden können, ist Teil des an den Hochschulen Hannover und Ostfalia ansässigen Forschungsschwerpunktes „Skalierbarkeit mobiler mikro-Blockheizkraftwerke“ [Sc14].

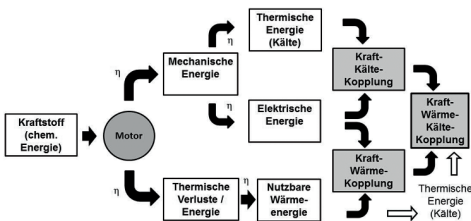


Abb. 1: Wirkungsgradketten für die Kennfelduntersuchung. Grafik: HS-Hannover

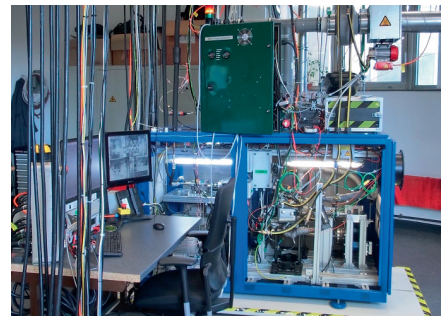


Abb. 2: 15kW Prüfstand an der Hochschule Hannover. Foto: HS-Hannover

Die an der Hochschule Hannover untersuchten mmBHKW-Konzepte grenzen sich dabei von üblichen, klassischen RE-Konzepten dadurch ab, dass ihre Hauptaufgabe in der Bereitstellung von Wärme- und Kälteströmen zur thermischen Konditionierung der Batterie und des Fahrgastraumes liegt. Um diese Technologie in Elektrofahrzeuge und Gebäude integrieren zu können, ist die Ermittlung relevanter Kennwerte (Abb. 1) notwendig, aus welchen sich Anforderungen hinsichtlich der Nutzung, Baugröße und Leistungsfähigkeit solcher Aggregate ergeben. Daraus abgeleitete Betriebsstrategien für mmBHKW-Konzepte in E-Fahrzeugen können dabei durch Steigerung der Energieeffizienz und des Komforts einen Beitrag zur Schonung von Ressourcen wie z.B. fossilen Brennstoffen leisten.

Im Rahmen des Forschungsschwerpunktes besteht hierzu eine enge Kooperation der Fachrichtungen Elektrotechnik, Maschinenbau und Informatik, wodurch es Studierenden dieser Fachrichtungen ermöglicht wird, sich in studentischen Projekten oder Abschlussarbeiten an der Bearbeitung von fachübergreifenden Aufgabenstellungen zu beteiligen.

Um die Machbarkeit dieses mmBHKW-Konzeptes zu überprüfen, wurden an der Hochschule Hannover zwei Versuchsstände im Leistungsbereich 1 – 15 kW (Abb. 2) aufgebaut. Mit Hilfe dieser können verschiedene Betriebsmodi, wie z. B. das Heizen mittels des Verbrennungsmotors bei gleichzeitigem Laden der Batterie oder das Laden der Batterie mit paralleler Kühlung des Innenraums, simuliert werden. Die Ergebnisse aus den Messun-

gen während verschiedener Versuche und Tests dienen hierbei u.a. der Einschätzung von Wirkungsgraden, Leistungen, Packaging und Gewicht [Hal16] für den Bau eines Prototyps.

In dieser Arbeit wird ein Konzept vorgestellt, mit dem die bisherige Aufzeichnung der Messwerte über einen Datenlogger und anschließender Offline-Analyse auf ein eingebettetes System umgestellt werden kann, um Messdaten im laufenden Betrieb analysieren und die Steuerung automatisieren zu können. Hierfür war es nötig, eine Systemarchitektur zu entwerfen, welche mit möglichst geringen Anpassungen dynamisch auf Veränderungen am Prüfstand reagieren kann und dennoch einen hohen Bedienkomfort bietet.

Im folgenden Abschnitt 2 werden die sich aus diesem Szenario ergebenden Anforderungen und das dafür entworfene Architekturkonzept vorgestellt. Die darauffolgenden Abschnitte gehen auf dessen Hauptkomponenten, das Mikrocontrollerboard (Abs. 3.1), die Servicearchitektur mit der Geschäftslogik (Abs. 3.2) und den Leitstand (Abs. 3.3), ein. Anschließend ordnet Abschnitt 4 diese Arbeit in den wissenschaftlichen Kontext ein, bevor Abschnitt 5 ein Fazit zu den erzielten Ergebnissen zieht und einen Ausblick auf die geplante Weiterführung und mögliche Weiterentwicklungen gibt.

2 Architekturkonzept

Der Prüfstandssoftware liegen umfangreiche funktionale Anforderungen und Rahmenbedingungen zugrunde, die einen großen Einfluss auf die Architektur haben. Insbesondere der Kontext des Forschungsprojektes beeinflusste Komponenten und Aufbau der Softwarearchitektur stark. Bei der Softwarearchitektur für den 15 kW Prüfstand steht vor allem die spätere Verwendung und Weiterentwicklung durch Mitarbeiter und Studierende aus den Fachbereichen Maschinenbau, Elektrotechnik und Informatik im Mittelpunkt.

Im Folgenden werden die Anforderungen an das System vorgestellt und mit einem Bezeichner (A1... An) versehen, mit dem sie im restlichen Dokument referenziert werden.

2.1 Anforderungen

Die Prüfstandssoftware soll die Untersuchungen am Prüfstand umfangreich unterstützen und dessen Anbindung an weiterverarbeitende Systeme vereinfachen (A1). Dazu muss die Software die Vielzahl an verschiedenen Messdaten des Prüfstandes visualisieren (A2) und die Steuerung der am Prüfstand vorhandenen Aktoren ermöglichen (A3). Durch einheitliche Schnittstellen (A4) soll die Software die Möglichkeit bieten, Steuer- und Regelprozesse mit Hilfe von weiteren Software-Komponenten zu automatisieren (A5), sowie die parallele Arbeit von mehreren Workstations erlauben (A6). Zur erneuten Auswertung muss die Software zudem die gefahrenen Versuche aufzeichnen und jederzeit zur weiteren Analyse wiedergeben können (A7), ohne dass der eigentliche Prüfstand benötigt wird. Um den sicheren Betrieb zu gewährleisten, muss die Software außerdem in der Lage sein, Über- oder Unterschreitungen von Grenzwerten zu erkennen (A8) und auf diese entsprechend, z.B. durch eine Warnmeldung an den Prüfstandsbediener, reagieren können.

Ferner muss das System auch ohne weitreichende Informatikkenntnisse leicht und verständlich bedienbar sein (A9), weshalb Werkzeuge verwendet werden sollen, die bereits in den entsprechenden Domänen genutzt werden (A10). Dabei muss es jedoch auch so flexibel und anpassbar sein, dass Änderungen an der Konfiguration des Prüfstandes, z.B. durch den Umbau von Sensoren, ohne großen Aufwand eingepflegt werden können (A11).

Da der Prüfstand im Rahmen eines bis 2018 angelegten Forschungsprojektes betrieben wird, ist außerdem zu gewährleisten, dass eine Erweiterung des Funktionsumfangs der Software ohne Beeinträchtigung der laufenden Systeme möglich ist, was eine starke Kapselung und lose Kopplung der einzelnen Komponenten voraussetzt (A12). Zusätzliche Komponenten zur Bilanzierung von Regelkreisen (A13), Aufzeichnung und Wiedergabe von Testläufen oder Automatisierung der Prüfstandssteuerung können so Schritt für Schritt eingebaut oder ausgetauscht werden.

2.2 Resultierende Architektur

Um die teils analogen, teils digitalen Sensor- und Steuersignale von der Hardwareebene in eine übersichtliche, schematische Visualisierung des Prüfstands zu überführen, bietet sich unter den gegebenen Voraussetzungen eine Aufteilung des Systems auf drei Kernkomponenten (Abb. 3) an [Du08]. Eine Oberfläche in LabVIEW bietet ein dem Anwender bekanntes Frontend (vgl. Anforderung A2), ein Application Server stellt eine Serviceinfrastruktur mit Geschäftslogik (vgl. A12) und ein Mikrocontrollerboard dient als hardwareseitiges Backend (vgl. A1). Der Application Server und das LabVIEW-Plugin entlasten zudem den Microcontroller, sodass problemlos komplexe Berechnungen zur Steuerung und Regelungen möglich sind, ohne kritische Abläufe auf dem Board zu beeinträchtigen.

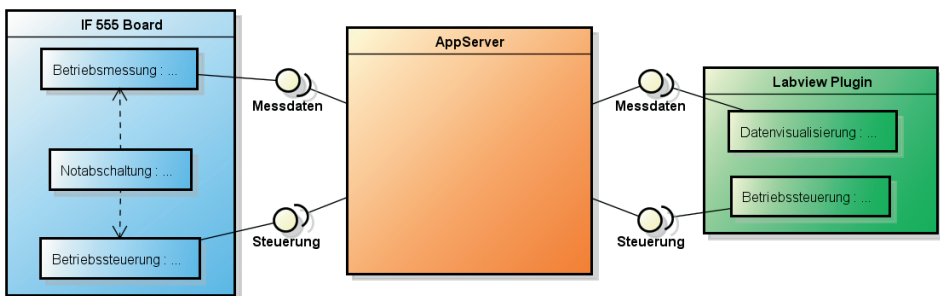


Abb. 3: Die drei Hauptkomponenten des Architekturkonzepts.

Ein IF-555-Board [Ha12] der Firma IEP aus Langenhagen, welches direkt am Prüfstand installiert wurde, bildet die Grundlage des Backends. Auf diesem befindet sich eine Motorola MPC563-CPU, auf welcher das Echtzeitbetriebssystem RTOS-UH [Ge06] (vgl. A10) läuft, wodurch Softwaremodule nebenläufig ausgeführt werden können. Dies spiegelt sich im Design und der Implementierung der Boardkomponenten (siehe Abschnitt 3.1) wieder. An das Board sind mehrere I/O-Module über CAN-Bus angeschlossen, um die Anzahl verfügbarer Sensor- und Aktoranschlüsse signifikant zu erhöhen.

Vom Board werden zwei Interfaces über TCP/IP zur Verfügung gestellt: Zum einen die Messdaten der über 100, an verschiedenen Stellen angeschlossenen, Sensoren (z.B. PT-100-Temperaturfühler und Drehzahlmesser) und zum anderen die Steuerung von mehr als 50 möglichen Aktoren (z.B. Servomotoren und Ventile) (vgl. A3). Dadurch ist eine komplette Überwachung des Prüfstandes und Automatisierung durch eine externe Regelelementkomponente möglich (vgl. A1). Die Interfaces bieten dabei wenig Abstraktion; so werden Sensordaten z.B. als Rohdaten (nicht in SI-Basiseinheiten überführt) anhand der Anschlussstelle aggregiert übertragen. Um später Sensoren und Aktoren beliebig hinzufügen oder entfernen zu können (vgl. A11), spielt es bei den Interfaces des Boards keine Rolle, ob tatsächlich ein Sensor oder Aktor angeschlossen ist. Somit bilden die Interfaces die Anschlussmöglichkeiten des IF-555-Boards [Ha12] und dessen Erweiterungen ab.

Die TCP/IP Interfaces des Boards werden durch einen Application Server genutzt, welcher die Schnittstelle des Prüfstands zu sämtlichen anderen Systemen, wie z.B. dem in LabVIEW realisierten Leitstand, darstellt. Services auf dem Application Server nehmen hierbei u.a. die Rolle eines Adapters ein, welcher zum einen eine stärker abstrahierte und standardisierte RESTful API zum Zugriff auf Sensoren und Aktoren des Prüfstands bietet (vgl. A4) und zum anderen auch den parallelen Zugriff mehrerer Clients auf diesen ermöglicht (vgl. A6), ohne dass die begrenzten Ressourcen des Boards hierdurch zusätzlich beansprucht werden. Des Weiteren bietet der Server zusätzliche Services an, welche Nutzern z.B. erlauben, die Konfiguration für Sensoren und Aktoren über ein leicht zu bedienendes Webinterface zu bearbeiten (vgl. A9) und zuvor in einer Datenbank aufgezeichnete Testläufe erneut wiederzugeben (vgl. A7). Die serviceorientierte Architektur der einzelnen, in Form von Microservices implementierten, Softwarekomponenten auf dem Application Server bietet dabei eine hohe Flexibilität (vgl. A12), sowohl was fachliche Anforderungen und die Integration mit anderen Systemen [St15] als auch die Skalierbarkeit des Systems und dessen Portierbarkeit anbelangt. Dies wird in Abschnitt 3.2 näher betrachtet.

Als virtuelle Bedienoberfläche des Prüfstands dient schließlich ein in LabVIEW (vgl. A10) realisierter Leitstand, welcher die aufbereiteten Sensordaten über die Interfaces des Application Servers bezieht und diese auch nutzt, um Steuerbefehle an die Aktoren zu übermitteln (vgl. A3). Die zentrale Ansicht des Leitstands stellt dabei die Visualisierung sämtlicher Komponenten des Prüfstands in Form eines Fließbildes, wie in Abbildung 8 zu sehen, dar (vgl. A2), welches in enger Zusammenarbeit mit der Abteilung Maschinenbau erstellt wurde. Mit der Entscheidung, für diesen Teil des Systems auf LabVIEW zurückzugreifen, wurde der Grundstein für ein nutzerfreundliches Frontend gelegt, welches sowohl von der Zielgruppe an Nutzern aus Maschinenbau und Elektrotechnik als auch den Entwicklern aus dem Bereich Informatik ohne große Einstiegshürden genutzt, gewartet und erweitert werden kann (vgl. A9). Parallel zu diesem in Abschnitt 3.3 näher beschriebenen Leitstand, wird im Rahmen einer Masterarbeit im Bereich Elektrotechnik eine ebenfalls in LabVIEW umgesetzte automatisierte Steuerung für den Prüfstand erarbeitet (vgl. A5), welche auf dieselben Schnittstellen zurückgreift.

3 Umsetzung

3.1 Hardwarenahes Backend

Das Backend der Prüfstandssoftware wurde in enger Zusammenarbeit mit Endanwendern des Systems entwickelt, wodurch der rege Austausch domänenspezifischer Informationen gefördert wurde. Dementsprechend bilden die Anforderungen an die Software des IF-555-Boards (Abb. 4) die zur Prüfstandoperation grundlegend notwendigen Operationen ab.

Neben den bereits genannten funktionalen Anforderungen existieren aufgrund der Projektstruktur weitere Qualitätsanforderungen, welche Einfluss auf das Architekturdesign und die Implementierung haben. Die komplette Boardsoftware muss, wie bereits erwähnt, einfach veränderbar und erweiterbar sein um Studierenden ohne große Einarbeitungszeit die Arbeit an der Software zu erlauben (vgl. A12). Durch das IF-555 Board besteht zudem die Rahmenbedingung der Implementierung in Ansi-C-89 unter der Verwendung von speziellen Bibliotheken. Auch die gute Unterstützung von nebenläufigen Tasks durch das Betriebssystem RTOS-UH [Ge06] (vgl. A10) soll im Architekturdesign genutzt werden, um Module – ähnlich zu Klassen in Java – semantisch orientiert zu entwerfen (vgl. A12).

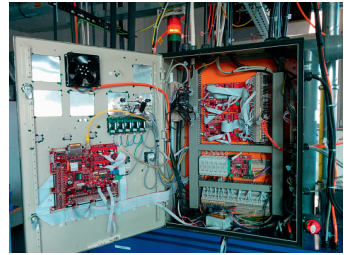


Abb. 4: Montiertes IF-555 Board mit Zusatzmodulen am Prüfstand.

Anhand der Anforderungen an die Software und die Architektur entstand ein modulares Design, in welchem die Komponenten wenig Abhängigkeiten untereinander besitzen und die Funktionalität über wenige, die anderen Komponenten verwendenden, Programme erfüllt wird. Abbildung 5 zeigt die Komponenten des Mess-Steuer-Regelsystems (MSR).

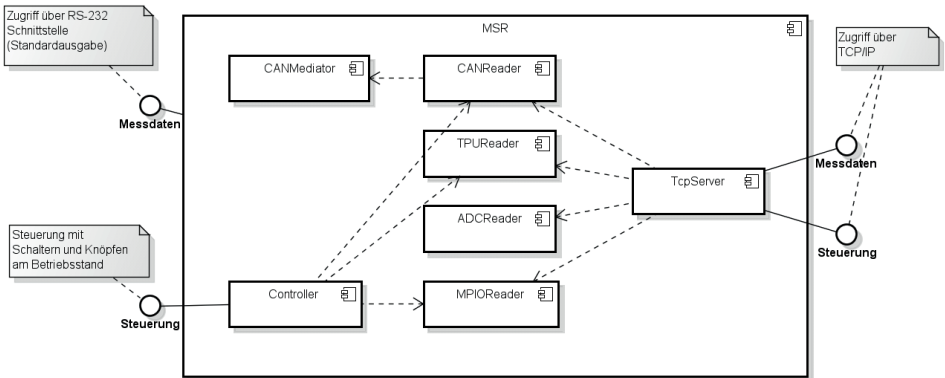


Abb. 5: Komponenten der Boardsoftware.

Die Grundlage des MSR bilden die verschiedenen *Reader*, welche anhand der Ein-/Ausgabestellen des IF-555 Boards gruppiert sind. Die Reader enthalten nebenläufige Tasks, welche die Sensordaten der abstrahierten Ein-/Ausgabestelle zur Verfügung stellen (vgl.

A2). Weiterhin bieten sie Funktionen zur Ansteuerung eventueller Ausgänge (vgl. A3). Die Reader verwenden für den nebenläufigen Betrieb die umfangreichen Mittel des RTOS-UH [Ge06], um die Stärken des Betriebssystems auszunutzen.

Der TCPServer greift auf alle Reader zu, um deren Mess- und Steuerfunktionen über eine FTP-ähnliche Schnittstelle zur Verfügung zu stellen (vgl. A1). Der Server bietet softwaretechnische Interfaces für Messdaten und Steuerung, welche vom AppServer, aber auch anderen Anwendungen genutzt werden können (vgl. A5).

Zur Steuerung des Prüfstandes über ein Kontrollpanel wird das physikalische Interface des Prüfstandes durch den Controller bereitstellt. Der Controller besteht aus einem Eventhandler, um auf Signale des Panels zu reagieren und entsprechende Steuersignale zu senden.

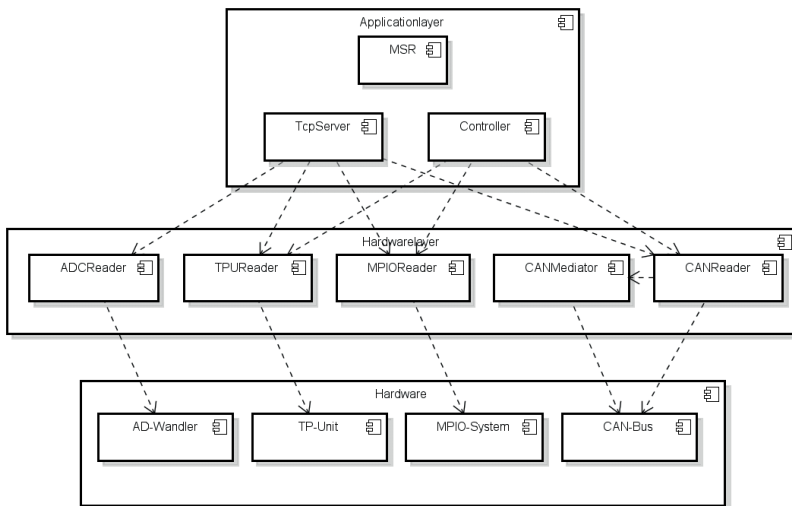


Abb. 6: Schichtarchitektur des Systems.

Durch die lose Kopplung und den schichtartigen Aufbau (Abb. 6) der Systemarchitektur sind Änderungen am System einfach durchzuführen und nachzuvollziehen (vgl. A12), wodurch eine zukünftige Weiterentwicklung durch Studierende möglich ist und das Testen der Komponenten erleichtert wird.

3.2 Servicearchitektur

Durch seine serviceorientierte Architektur bietet der Application Server (vgl. A10) einen hohen Grad an Modularität und Flexibilität, da sämtliche Services untereinander und nach außen über RESTful APIs (vgl. A4) kommunizieren (Abb. 7). Die Anbindung von Komponenten mit anderen, ggf. nicht standardisierten Protokollen kann dabei leicht in eigene Adapter gekapselt und über vereinheitlichte Schnittstellen verfügbar gemacht werden.

Der eng gefasste Aufgabenbereich und die dadurch kompakt gehaltene Codebasis der einzelnen Microservices ermöglichen es hierbei, mit geringem Aufwand auf Änderungen z.B.

am Kommunikationsprotokoll mit dem Board oder den Anforderungen an die Aufbereitung von Sensordaten zu reagieren [Ne15] (vgl. A12). Ein derart hoher Grad an Adaptivität innerhalb der Anwendungssoftware ist besonders in einem Forschungsszenario wie dem hier zugrunde liegenden Mikro-BHKW Prüfstand wichtig, um mit dem stetigen Wandel der Anforderungen, welche sich aus den verschiedenen Arbeiten am Prüfstand durch verschiedene Nutzer ergeben, Schritt zu halten.

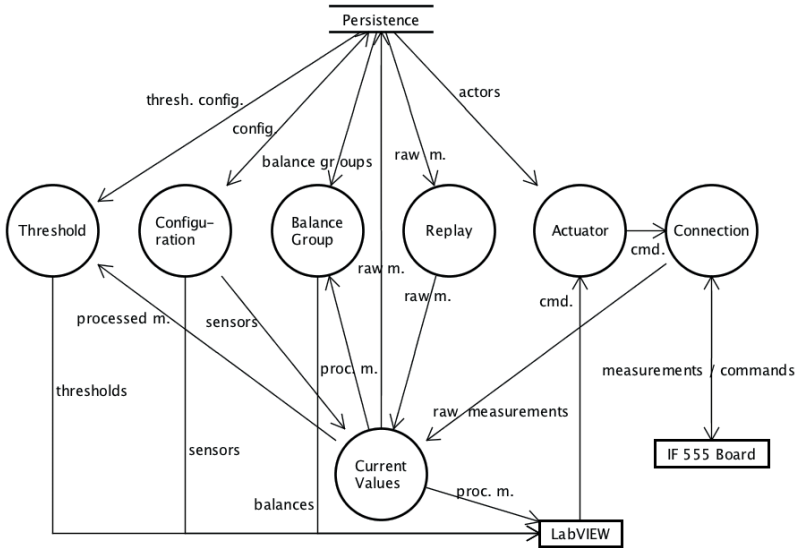


Abb. 7: Datenflüsse zwischen den einzelnen Services des Application Servers.

3.2.1 Basisdienste

Für dieses Szenario wurde ein Satz an Basisdiensten erarbeitet, der die Infrastruktur für sämtliche weiteren Dienste bildet und aus zwei zentralen Komponenten besteht:

- Der `Connection` Service stellt die Schnittstelle zwischen der Servicearchitektur des Application Servers und der Software auf dem Board dar und ist für die Verwaltung der Verbindungen zwischen beiden Systemen zuständig (vgl. A4). Er kann mehrere Verbindungen zu verschiedenen (ggf. virtuellen) Boards handhaben und sowohl deren Daten anderen Services zur Verfügung stellen als auch Steuerbefehle von diesen an die Boards weiterleiten (vgl. A1).
- Die aktuellen Rohdaten von Sensoren der jeweiligen Prüfstände werden vom Service `CurrentValues` in ggf. aufbereiteter Form anderen Services oder Clients zur Verfügung gestellt (vgl. A6).

Ergänzt werden die zentralen Services um einen generischen und agnostischen [Er14] `Persistence` Service, der u.a. zur Speicherung der Messwerte (vgl. A7) und zugehörigen Metadaten wie SI-Einheiten genutzt wird.

3.2.2 Erweiterte Funktionalitäten

Weitere Services erweitern den Funktionsumfang, wie u.a. der `BalanceGroup` Service zur Berechnung von Energie- oder Stoffbilanzen (vgl. A13) im Prüfstand. Um Störungen frühzeitig erkennen zu können, überwacht der `Threshold` Service kontinuierlich die Messwerte von `CurrentValues` auf die Über-/Unterschreitung von Messwerten (vgl. A8).

In der Datenbank aufgezeichnete Testläufe können mit Hilfe des `Replay` Service wiedergegeben werden (vgl. A7), um diese im Nachgang ausführlicher analysieren zu können. Durch die Speicherung von Rohdaten können dabei auch neue Bilanzkreise eingeführt oder bereits vorhandene Formeln korrigiert werden. Zudem kann direkt zu relevanten Zeitpunkten gesprungen werden, ohne z.B. die Aufwärmphase abwarten zu müssen.

Die über ein Webinterface änderbaren (vgl. A9) Sensorkonfigurationen werden versioniert durch den `Configuration` Service verwaltet (vgl. A11), wodurch diese auch für aufgezeichnete Testläufe verfügbar bleiben. Als Rückkanal für Steuerbefehle zur Regelung des Prüfstands mittels der Aktoren dient der `Actuator` Service (vgl. A3), welcher diese auf ihre Plausibilität prüfen kann. Hierdurch kann der Prüfstand von beliebigen Clients oder Services z.B. bei Grenzwertüberschreitungen automatisiert gesteuert werden (vgl. A5).

3.3 Leitstand

Zur Überwachung und (teilweise manuellen) Steuerung des Prüfstands wurde in enger Zusammenarbeit mit den Anwendern aus Maschinenbau und Elektrotechnik ein virtueller Leitstand in LabVIEW entwickelt, welcher die bisher genutzten Messinstrumente und Bedienelemente weitestgehend ersetzt (vgl. A2). Die Anbindung der LabVIEW-Oberfläche an den Prüfstand erfolgt dabei mittels HTTP-Aufrufen an die entsprechenden RESTful APIs der Services auf dem Application Server (vgl. A4).

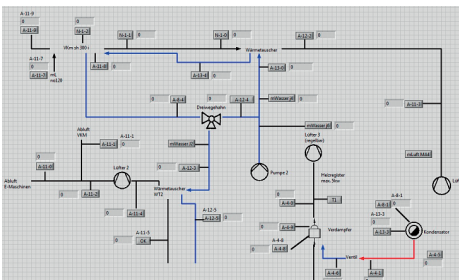


Abb. 8: Fließbild des Prüfstands in LabVIEW

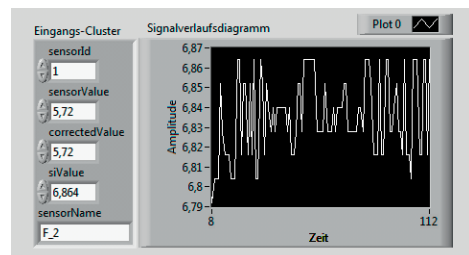


Abb. 9: Detailansicht eines Sensors

Nach dem Verbindungsaufbau werden die über `CurrentValues` ausgelesenen Messwerte sowie die Bilanzen von `BalanceGroup` (vgl. A13) in einem schematischen Fließbild des Prüfstands (Abb. 8) dargestellt (vgl. A10) und in einem konfigurierbaren Intervall aktualisiert. LabVIEW zeichnet dabei zusätzlich den Verlauf der Messwerte der einzelnen Sensoren auf, sodass dieser in der Detailansicht (Abb. 9) als Plot dargestellt werden kann. Zusätzlich werden in dieser Ansicht auch die rohen, die korrigierten und die in die ent-

sprechenden SI-Einheiten umgerechneten Sensorwerte angezeigt, um mögliche Fehler in den Korrektur- und Umrechnungsformeln erkennen zu können.

Weitere Bedienelemente im Fließbild erlauben zudem die Ansteuerung der Aktoren (vgl. A3), welche im Hintergrund über den Actuator Service läuft und so bei Bedarf die Regelung verschiedener Teilsysteme des Prüfstands von mehreren, verschiedenen Workstations aus ermöglicht (vgl. A6). Warnmeldungen zu von Threshold erkannten Grenzwertüberschreitungen werden ebenfalls in dieser Oberfläche angezeigt (vgl. A8), sodass auf diese umgehend reagiert werden kann.

Bei Änderungen (z.B. der Sensorkonfiguration) müssen diese manuell in LabVIEW nachgezogen werden (vgl. A11), was durch wiederverwendbare Bausteine mit geringem Aufwand möglich ist. Dies erlaubt die Wartung des Frontends und die Verwaltung der Sensor-, Aktor- und Bilanzkonfigurationen ohne weitreichende Programmierkenntnisse (vgl. A9).

4 Verwandte Arbeiten

In [Rü14] werden die grundlegenden Konzepte und Ideen zur Skalierbarkeit und Anwendung von mmBHKW in Fahrzeugen vorgestellt, welche die Basis der Forschungsarbeiten rund um die Prüfstände der Hochschule Hannover bilden, für welche die hier präsentierte Softwarearchitektur ausgelegt wurde. Ein Werkzeug zur Simulation des Einsatzes von mmBHKW in konkreten Szenarien hinsichtlich dessen Auswirkung auf die CO₂-Bilanz wurde vorgestellt [RSG15].

Ein ähnliches Konzept zu dem des mmBHKWs wurde in [Bo12] dargelegt, wobei hier lediglich die KWK ohne die zusätzliche Nutzung des Aggregats zur Kühlung betrachtet wird und dieses auch fest im Fahrzeug verbaut ist.

[PL14] beschreibt eine allgemeine Softwarearchitektur zur Testautomatisierung für Prüfstände, jedoch mit einem Fokus auf komplette Antriebsstränge. Die dort vorgestellte Architektur weist zwar eine in Teilen vergleichbare Struktur auf, ist in ihrem Anwendungsschwerpunkt jedoch auf Szenarien mit einem anderen Maßstab an Anforderungen ausgelegt als das hier präsentierte Konzept.

Zur Smart-Grid-Integration wurde in [Ap12] eine serviceorientierte Referenzarchitektur vorgestellt, die auch Komponenten für flexible Erzeuger, wie z.B. mmBHKW enthält. Die hier vorgestellte Prüfstandsarchitektur kann mit geringen Modifikationen in die Referenzarchitektur eingepasst werden, um ein mmBHKW direkt in ein Smart-Grid zu integrieren.

Eine cloudbasierte, serviceorientierte Architektur mit Microservices wird in [Ho16] auf Grund ihrer hohen Flexibilität und Adaptivität als Lösung für IT-Systeme mit hohen Ansprüchen an Robustheit und Stabilität gegenüber unerwarteten Ereignissen präsentiert, mit der diese in einen als „anti-fragil“ [Ho16] bezeichneten Zustand gebracht werden können. Aus diesem Grund wird dort auch deren Einsatz z.B. in Smart Grid Szenarien als besonders kritische Infrastruktur diskutiert.

Die in diesem Beitrag eingeführte Architektur ist auf die Integration der Systeme eines Prüfstands für mmBHKW ausgerichtet, bietet aber auch gute Anknüpfungspunkte für weitere Entwicklungen hinsichtlich der Anbindung eines Prototypen an bestehende und in Entwicklung befindliche Smart Grid Infrastrukturen und vergleichbare Systeme.

5 Fazit und Ausblick

In diesem Beitrag wurde ein Architekturkonzept zur softwaregestützten Überwachung und Steuerung des mmBHKW-Prüfstands der Hochschule Hannover vorgestellt. Dabei wurden die Anforderungen im Rahmen des interdisziplinären Forschungsprojekts analysiert und darauf aufbauend eine dreischichtige Softwarearchitektur aus hardwarenaher Prüfstandssteuerung, serviceorientierter Geschäftslogik und LabVIEW-Frontend entworfen.

Es wurde dargelegt, wie durch die klar definierten und abgegrenzten Zuständigkeitsbereiche der einzelnen Komponenten die Ansprüche an die Flexibilität und Anpassbarkeit des Systems an das dynamische Anforderungsumfeld eines Forschungsprüfstands erfüllt werden können. Die verwendete Servicearchitektur erlaubt zudem die Integration und Anbindung domänenspezifischer Anwendungen, wie am Beispiel der RTOS-basierten Prüfstandssteuerung sowie des virtuellen Leitstands in LabVIEW gezeigt wurde.

In unseren zukünftigen Arbeiten sollen u.a. folgende Themen betrachtet werden: Vollautomatisierung des Prüfstandes; Integration in Smart Grid Szenarien; Möglichkeiten zur Emissionsreduktion und Effizienzsteigerung; Performanz- und Skalierbarkeitsanalysen sowie der Einsatz von Complex Event Processing und Cloud Computing Technologien, ggf. basierend auf einer ereignisgetriebenen Architektur [BD10].

6 Danksagungen

Dieses Projekt wurde gefördert durch die VolkswagenStiftung und das Nds. Ministerium für Wissenschaft und Kultur (Fkz. VWZN2891). Wir bedanken uns bei den Studierenden des Bachelorprojekts und den Mitgliedern des Forschungsschwerpunkts „Skalierbarkeit mobiler Mikro-Blockheizkraftwerke“ an der Hochschule Hannover für ihre Unterstützung und die gute Zusammenarbeit.

Literaturverzeichnis

- [Ap12] Appellrath, Hans-Jürgen; Bischofs, Ludger; Beenken, Petra; Uslar, Mathias: IT-Architekturentwicklung im Smart Grid. Springer, 2012.
- [BD10] Bruns, Ralf; Dunkel, Jürgen: Event-Driven Architecture: Softwarearchitektur für ereignisgesteuerte Geschäftsprozesse. Xpert.press. Springer, Berlin [u.a.], 2010.
- [Bo12] Bouvy, Claude; Baltzer, Sidney; Ernst, Christian; Eckstein, Lutz: Range Extender als Mobile Kraft-Wärme-Kopplungseinheit. ATZ - Automobiltechnische Zeitschrift, 114(10):764–769, 2012.

- [Du08] Dunkel, Jürgen; Eberhart, Andreas; Fischer, Stefan; Kleiner, Carsten; Koschel, Arne: Systemarchitekturen für Verteilte Anwendungen: Client-Server, Multi-Tier, SOA, Event Driven Architectures, P2P, Grid, Web 2.0. Carl Hanser, München, 2008.
- [Er14] Erl, Thomas: Next generation SOA: A concise introduction to service technology & service-orientation. Prentice Hall/PearsonPTR, 2014.
- [Ge06] Gerth, Wilfried, , RTOS-UH, 2006. IEP.
- [Ha12] Hadler, Andreas und Koerth, Klaus, , Dokumentation IF555-3, 2012. IEP.
- [Ha16] Hanif, Haider Iqbal; Minnrich, Jan Peter; R. P. Schmicke, Christian; Rüscher, Henrik; Gusig, Lars-Oliver: Bauraum- und gewichtstechnische Untersuchung einer mobilen mikro-PCU zur Bereitstellung von Wärme, Kälte und elektrischen Strom im E-Fahrzeug. In: 10. Tagung – Wärmemanagement des Kraftfahrzeugs, 09. - 10.06.2016. Potsdam, 2016.
- [Ho16] Hole, Kjell Jørgen: Anti-fragile ICT Systems. Springer International Publishing, Cham, 2016.
- [Mi14] Minnrich, Jan P.; Rüscher, Henrik; Schmicke, Christian R. P.; Gusig, Lars-Oliver: Integration einer mikro-PCU im Thermomanagement eines Elektrofahrzeugs unter Berücksichtigung von Reichweite und Emissionen. In: 9. Tagung — Wärmemanagement des Kraftfahrzeugs, 05 - 06.06.2014. Potsdam, 2014.
- [Mi15] Minnrich, J. P.; Schmicke, Christian R.P.; Rüscher, Henrik; Gusig, Lars-Oliver: Multi-fuel Application in Alternative Range-Extender-Concepts. In: International Conference IDTechEx, 28. - 29.04.2015. Berlin, 2015.
- [Ne15] Newman, Sam: Building microservices: Designing fine-grained systems. O'Reilly, 2015.
- [PL14] Paulweber, Michael; Lebert, Klaus: Mess- und Prüfstandstechnik: Antriebsstrangentwicklung · Hybridisierung · Elektrifizierung. Springer Fachmedien Wiesbaden, Wiesbaden, Kapitel Softwaresicht – Prüfstand, S. 273–376, 2014.
- [RB16] Referat Öffentlichkeitsarbeit; Bundesministerium für Wirtschaft und Energie: , BMWi - Leitmarkt und Leitanbieter. <http://www.bmwi.de/DE/Themen/Industrie/Elektromobilitaet/leitmarkt-und-leitanbieter.html>, 2016. online; abgerufen 10/05/2016.
- [RSG15] Rüscher, Henrik; Schmicke, Christian R.P.; Gusig, Lars-Oliver: Applicability and scalability of mobile mCHP units in mid-size battery electric vehicles and detached houses with different energy standards. In: International Conference on Sustainability in Energy and Buildings 2015, 01. - 03.07.2015. Lisabon, Portugal, 2015.
- [Rü14] Rüscher, Henrik; Schmicke, Christian R. P.; Minnrich, Jan P.; Gusig, Lars-Oliver: Weiterentwicklung von Range Extendern zu mobilen mikro-Blockheizkraftwerken in Fahrzeugen und Gebäuden. In: Techniktagung Kraft-Wärme-Kopplungssysteme, 29. - 30.04.2014. Berlin, 2014.
- [Sc14] Schmicke, Christian R.P.; Rüscher, Henrik; Minnrich, Jan P.; Gusig, Lars-Oliver: Strategies for combined use of power conditioning units in vehicles and buildings. In: International Conference on Sustainability in Energy and Buildings 2014, 25 - 27.06.14. Cardiff, Wales, 2014.
- [St15] Starke, Gernot: Effektive Softwarearchitekturen: Ein praktischer Leitfaden. Hanser, München, 7.. Auflage, 2015.

Energieeinsparungen im Gebäudebetrieb durch visualisiertes Feedback an Nutzer: Datenerfassung und Datenvisualisierung in Nicht-Wohngebäuden

Stefan Naumann, Andrea Christian, Christoph Göttert, Klaus-Uwe Gollmer, Rainer Michels, Stefan Ruffler¹

Abstract: Öffentliche Gebäude tragen signifikant zu Energieverbräuchen in den Bereichen Strom, Wärme und Kälte und damit zu CO₂-Emissionen bei. In Rahmen eines am Umwelt-Campus Birkenfeld der Hochschule Trier durchgeführten Projekts wurden umfangreiche Messsysteme installiert, um diese Verbräuche zu erfassen und zu visualisieren. Am Beispiel eines Hörsaals wird gezeigt, wie sich solche Messungen veranschaulichen lassen, und wie gleichzeitig der Lernkomfort durch Rückmeldung des aktuellen CO₂-Gehaltes im Raum verbessert werden kann.

Keywords: Gebäudetechnik, Verbrauchsvisualisierung, Messsystem, Hochschulen

1 Motivation

Öffentliche und private Gebäude verbrauchen etwa 40% der gesamten Energie in der Europäischen Union, davon entfallen etwa 37% auf öffentliche Gebäude [WRK10]. Hochschulen sind öffentliche Gebäude und haben damit auch die Herausforderung, den Energiekonsum insbesondere in den Bereichen Strom, Wärme und Kälte zu analysieren und zu reduzieren. Hierzu ist es erforderlich, diese Verbräuche zunächst systematisch zu erheben – what you can't measure you can't manage –, und hieraus entsprechende Maßnahmen zu entwickeln. Dabei spielen vorhandene Gebäudeleittechnik, Automatisierungsmaßnahmen und auch die Aktivitäten der Nutzer des Gebäudes gleichermaßen eine wichtige Rolle. So wurde 2006 von Karatzas und Bassoukos ein System vorgestellt, das insbesondere im Schulumfeld das Energiemanagement unterstützt [KB06].

Im hier vorgestellten Projekt wurden am Umwelt-Campus Birkenfeld, einem Standort der Hochschule Trier, umfangreiche Messdaten erhoben, aus denen Einsparpotenziale abgeleitet werden konnten. In diesem Beitrag wird die genutzte Messinfrastruktur beschrieben und es werden ausgewählte Ergebnisse vorgestellt. Grundlegende Informationen zu dem korrespondierenden Projekt finden sich in [Ko13]. Einen umfassenden Überblick über Informationssysteme zum Gebäude-Energieverbrauch findet sich in [LOP08].

¹ Hochschule Trier, Umwelt-Campus Birkenfeld, Institut für Softwaresysteme, Postfach 1380, 55761 Birkenfeld, s.naumann@umwelt-campus.de

2 Darstellung des Gebäudekomplexes

Der Umwelt-Campus Birkenfeld ist ein aus einem ehemaligen amerikanischen Militärhospital hervorgegangener Gebäudekomplex. Die Gebäudestruktur beinhaltet sowohl renovierte als auch vollständig neu errichtete, modern ausgestaltete Gebäudeteile mit innovativer Anlagentechnik.

Die Heizwärmeversorgung erfolgt über ein Fernwärmenetz durch ein ca. 500 m entfernt gelegenes Biomasse-Heizkraftwerk. Im Versorgungsbereich der 2002 neu in Betrieb genommenen Gebäude wird das Fernwärmenetz durch eine 260 m² große solarthermische Anlage unterstützt. Sie speist einen Wärmespeicher, der jahreszeitabhängig die Wärme verschiedenen Verbrauchern zuführt. Im Frühjahr und Herbst kann der Wärmespeicher einen Teil der Gebäude heizen. Im Sommer nutzt eine Adsorptionskältemaschine die Wärme der solarthermischen Anlage zur Erzeugung von Kaltwasser zur Kühlung der Frischluft und zur Kühlung des Gebäudes über Rohrschlangen in Wänden und in Fußböden. Die 1996 renovierten Bestandsgebäude beziehen ihre Wärme nur aus der Fernwärmeversorgung. Zur Unterstützung der Warmwasserbereitung sind ebenfalls solarthermische Kollektoren installiert.

Zur Frischluftversorgung befindet sich in 4 m Tiefe ein Erdkollektor aus 2 Stahlbetonrohren mit 1,5 m Durchmesser. Unter Nutzung der Erdwärme bzw. -kälte wird die Außenluft zur Gebäudebelüftung über den Erdkollektor angesaugt. Im Winter ermöglicht der Erdkollektor eine Temperaturerhöhung und im Sommer eine Temperaturabsenkung. Liegt die Temperatur der Außenluft im Bereich der gewünschten Zulufttemperatur, wird der Erdkollektor nicht genutzt und kann sich regenerieren. In diesem Fall wird die Außenluft direkt über einen Betonzuluftschacht angesaugt.

Bevor die Abluft der Lüftungsanlage in die Umwelt abgegeben wird, kommt es zu einer Wärmerückgewinnung der in der Abluft enthaltenen Wärme. Ein rotierender Wärmetauscher entzieht der Abluft die Wärmeenergie und führt sie der frischen Zuluft zu. Ein Teil der noch verbleibenden Restwärme der Abluft wird über Wärmetauscher in einem Massivabsorber einer Wärmepumpe zugeführt und auf ein höheres Temperaturniveau gebracht. Diese Wärme wird im Winter zur schnelleren Regeneration des Erdreiches um den Erdkollektor genutzt.

Die Steuerung der Gesamtheizungs- und Lüftungsanlage erfolgt über ein Gebäudeleittechniksystem. Über ein zweites Messwerterfassungssystem werden weitere Sensorsignale aufgeschaltet, die nicht zwingend zur Anlagensteuerung notwendig sind, aber zusätzliche Informationen zur Ermittlung der Energieeffizienz einzelner Anlagenkomponenten bieten.

Tabelle 1: Nutzung und Topologie Hochschule Trier, Standort Umwelt-Campus Birkenfeld

Bauherr	Landesbetrieb LBB, Niederlassung Idar-Oberstein
Betreiber	Hochschule Trier, Standort Umwelt-Campus Birkenfeld, Rheinland-Pfalz
Bauphase	1996 – 2003
Ursprüngliche Gebäudenutzung	Konversion eines ehemaligen Militärhospitals (ursprüngliche Fertigstellung 1964)
Beginn Studienbetrieb	1996
Gesamtnutzungsfläche	25.678 m ²
Anzahl Räume	867
Studierende	ca. 2.700
Mitarbeiter	ca. 300
Gesamtenergiebedarfe am Beispiel des Jahres 2014	
Elektrischer Energiebedarf	1.078 MWh/a
Wärmebedarf	1.679 MWh/a
Kälte genutzt	221 MWh/a

3 Untersuchungsabgrenzung

Um den Untersuchungsrahmen eingrenzen zu können, wurden vor Projektbeginn sowohl die Nutzergruppen also auch die untersuchten Räume festgelegt. Seitens der Nutzer waren dies Dozentinnen und Dozenten, Mitarbeiterinnen und Mitarbeiter sowie Studierende. Als Raumtypen wurden Hörsäle und Seminarräume, Büroräume, ein Technikum, Flure, Besprechungs- und Sozialräume festgelegt (siehe auch Abbildung 1).



Abbildung 1. Untersuchte Raumtypen (Hörsaal, Technikum, Büroraum)

4 Messdatenerhebung

Aufbauend auf bereits verfügbaren Messeinrichtungen im Bereich Wärme und Strom wurden im Rahmen des Projektes weitere Sensoren installiert, welche eine granularere Erhebung von Messdaten ermöglichen. Insbesondere wurden weitere Kältemengenzähler für das Gebäude eingebaut, Stromzähler ergänzt, und Heizkostenverteiler (HKV) an jedem Heizkörper montiert, um raumscharf die Wärmebedarfe ermitteln zu können. Abbildung 2 zeigt die Verteilung der verschiedenen Messpunkte auf dem Campusgelände. Dabei wird die Wärmemessung noch unterteilt in Wärme Heizung sowie Wärme Lüftung sowie Wärme Lüftung.

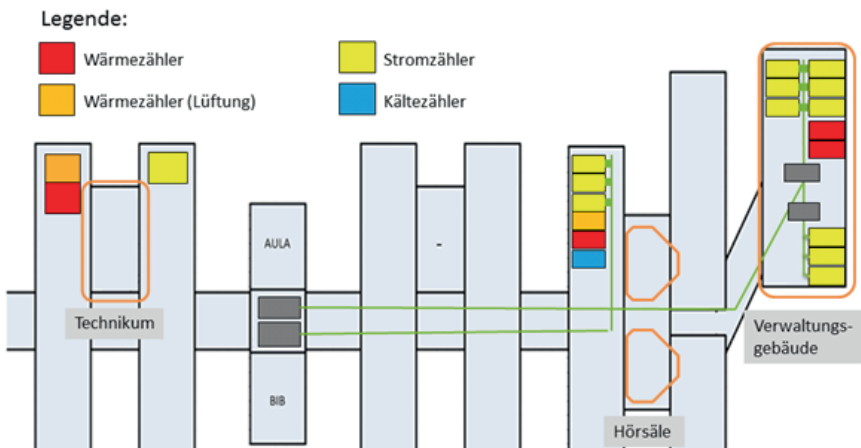


Abbildung 2. Überblick über die in den Gebäuden installierte Messtechnik



Abbildung 3. Eigenentwickelter Multisensor

Insgesamt wurden über 20 zusätzliche Sensoren im Rahmen des Projekts installiert und deren Daten zentral erfasst. Abbildung 4 zeigt die verschiedenen Datenquellen, Übertragungswege und auch Datenformate des Projekts.

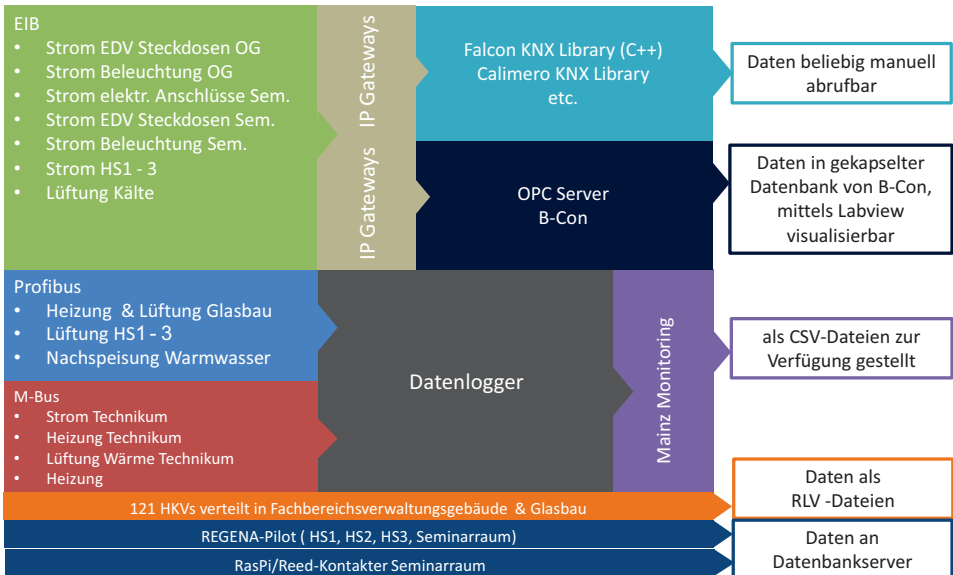


Abbildung 4. Datenquellen und Übertragungswege im untersuchten Gebäude

Um insbesondere in den Hörsälen noch detailliertere Informationen zu erhalten, wurden zusätzliche Multi-Sensoren eigenentwickelt (vgl. Abbildung 3). Diese im Kunststoffgehäuse untergebrachte Kleinstrechner (Pilot) auf Arduino-Basis ermitteln in

kurzen Abständen Raumtemperatur, -feuchtigkeit, CO₂-Konzentration, Lautstärke und Summe flüchtiger Kohlenwasserstoffe. Dies sind nach [GB16] wesentliche und auch ausreichende Werte zur Beurteilung des Raumklima in Hörsälen. Die Werte werden über den Tag auf einer SD-Speicherkarte gesammelt und gegen ein Uhr morgens dann per FTP an eine Datenbank übertragen.

Insgesamt kommen somit aus mehreren Datenquellen Messdaten zusammen, die über unterschiedliche Übertragungswege in die zentrale relationale Datenbank übertragen werden. Abbildung 5 zeigt die Datenflüsse der verschiedenen Quellen.

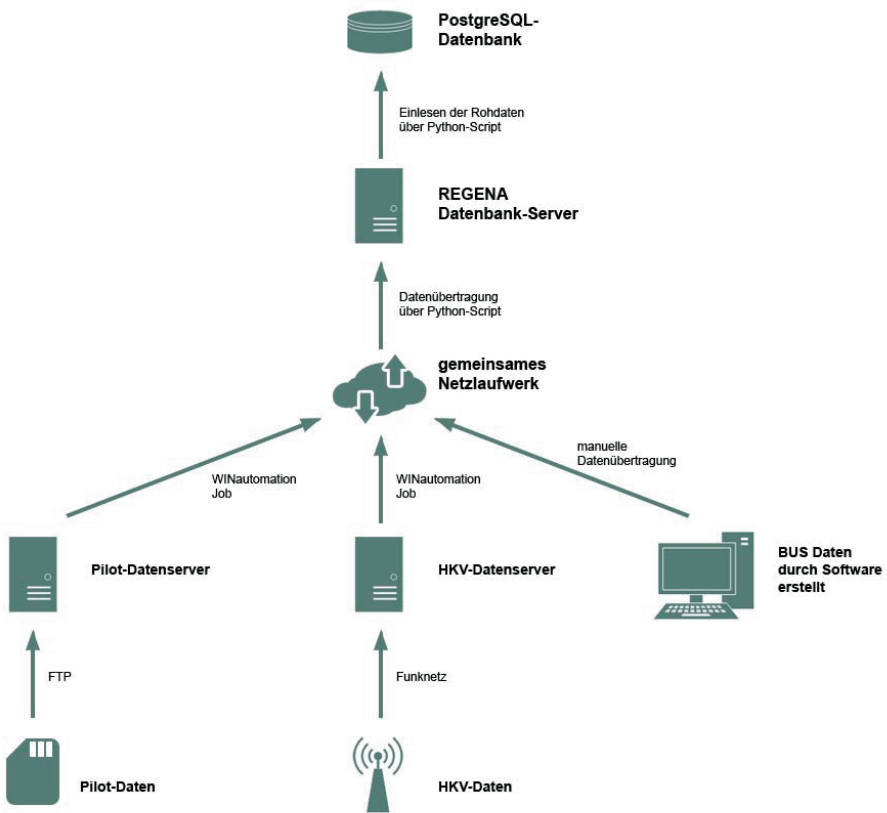


Abbildung 5. Datenflussdiagramm der erhobenen Sensordaten

Dabei laufen die Daten zunächst auf einem ebenfalls zentralen Netzlaufwerk zusammen, von dem aus dann mittels eines Python-Skripts die Vorverarbeitung vorgenommen wird, um schließlich die entsprechenden SQL-Befehle zu generieren. Insgesamt wurden in den

letzten 3 Jahren etwa 35 Millionen Messwerte erfasst. Zur Automatisierung der Messprozesse wurde neben durch Cron-Jobs gesteuerte Skripte die Software WinAutomation (<http://www.winautomation.com/>) genutzt, welche die Abrufe der HKV-Daten automatisiert. Hier war seitens der Herstellerfirma nur eine manuelle Abfrage vorgesehen, was für ein beispielsweise jährliches Auslesen ausreicht, aber für eine hochfrequente tägliche Analyse nicht praktikabel ist.

Abbildung 6 schließlich stellt das Entity-Relationship-Modell der zentralen Datenbank dar. Im Mittelpunkt steht die Tabelle „Messwert“, welche sämtliche Messwerte enthält und diese abhängig von Sensortyp, Messgröße, Sensorort etc. ablegt.

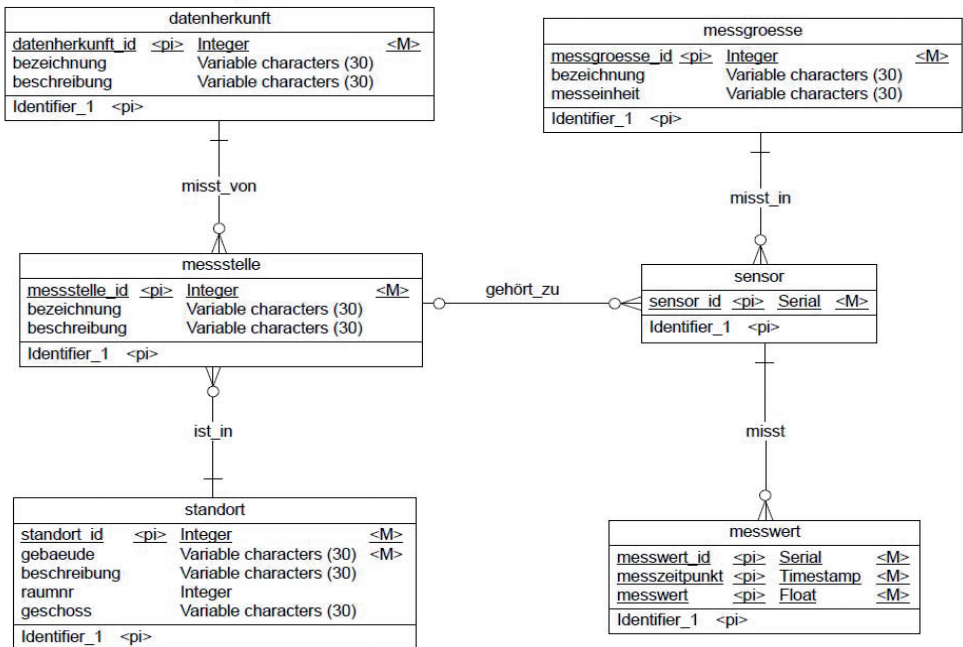


Abbildung 6. Entity Relationship-Modell der zentralen Datenbank

5 Visualisierung der Ergebnisse

Um den Gebäudenutzern Klima- und Verbrauchsdaten als Feedback zur Verfügung zu stellen, wurde neben der Messtechnik auch Software benötigt, mit Hilfe derer sich die gesammelten Messwerte aufbereiten und ansprechend visualisieren lassen. Zu diesem Zweck sowie für eigene Auswertungen wurde die Open-Source Software BIRT, ein Business Intelligence and Reporting Tools, verwendet und angepasst

(<http://www.eclipse.org/birt/>). Das Java-basierte Projekt stellt Werkzeuge zum Aufbereiten und zum Visualisieren von großen Datenmengen zur Verfügung. Dabei wird eine direkte Anbindung an die relationale -PostgreSQL-Datenbank, die alle Messdaten speichert, unterstützt.

Abbildung 7 zeigt exemplarisch den Stromverbrauch der drei untersuchten Hörsäle. Deutlich ist zu erkennen, dass die Verbräuche abhängig von Vorlesungszeiten sind, aber auch, dass der Verbrauch insgesamt zurückgegangen ist. Dies ist unter anderem auf Schulungen zurück zu führen, in denen die Hörsaal-Nutzer (Dozentinnen und Dozenten sowie Studierende) Hinweise zur stromsparenden Raumnutzung erhielten.

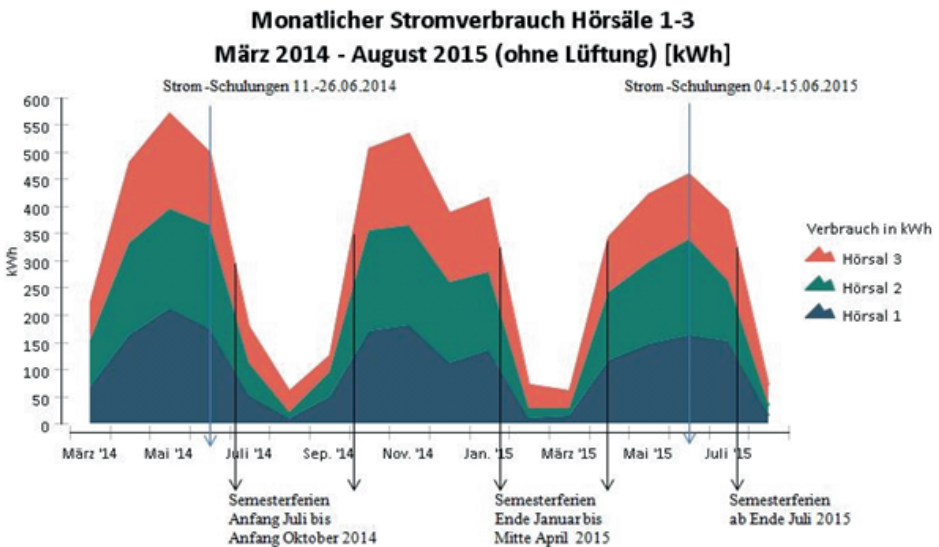


Abbildung 7. Exemplarische Visualisierung von Stromverbräuchen in drei Hörsälen

Eine weitere Rückkopplung an die Hörsaalnutzer ist durch eine Echtzeit-Visualisierung der Messdaten des in Abbildung 3 gezeigten eigenentwickelten Multisensors gegeben. Mittels der Internet-of-Things-Erfassungs- und Visualisierungssoftware Thingspeak (<https://thingspeak.com/>) können in Echtzeit Messergebnisse hochschulintern über das Intranet angezeigt werden. Abbildung 8 verdeutlicht dies am Beispiel des CO₂-Gehaltes im Hörsaal, der gleichzeitig ein Kennzeichen für „verbrauchte Luft“ und damit einhergehender geringerer Konzentrationsfähigkeit ist. Entsprechend kann bei Werten von bspw. über 1.200 ppm stoßgelüftet werden, was deutlich energieeffizienter ist als

ständiges Lüften mit gekippten Fenstern. In Abbildung 9 zeigt einen exemplarischen Messdatenverlauf für die Werte von Luftqualität, Lautstärkepegel und CO₂-Gehalt.

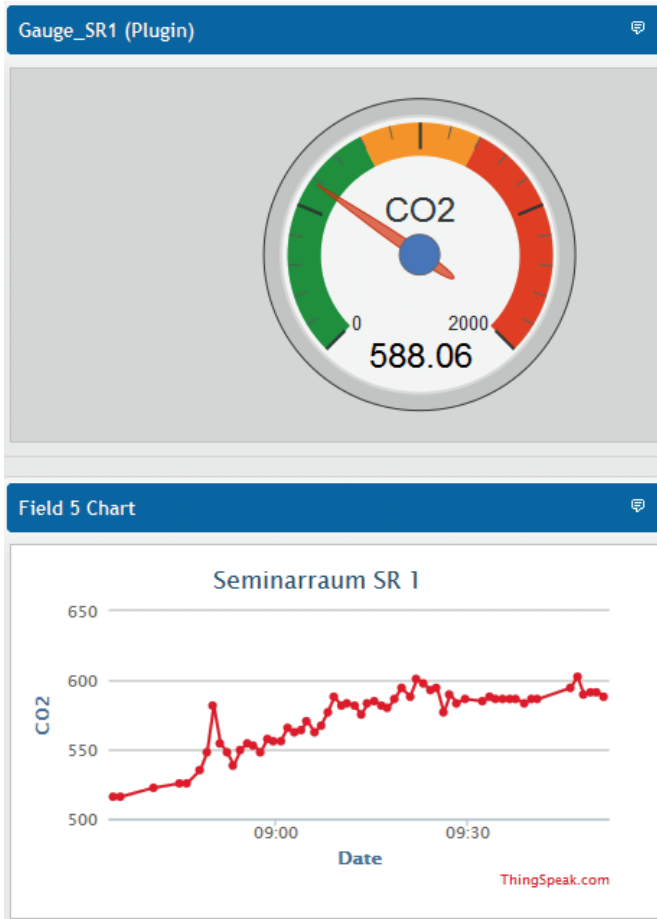


Abbildung 8. Echtzeit-Visualisierung von CO₂-Werten [ppm] in einem Hörsaal mittels Thingspeak

6 Zusammenfassung und Ausblick

Um in komplexen und heterogen genutzten Gebäuden wie Hochschulen Energieeinsparungen zu erzielen, ist eine umfangreiche Datenerhebung notwendig, um die verschiedenen Energieströme aus den Bereichen Wärme, Kälte und Strom zu erfassen und so zu klassifizieren, dass Maßnahmen zur Einsparung vorgenommen werden können. Hierzu sind zum einen zahlreiche Messpunkte zu setzen, und zudem

müssen Datenqualität und auch Lückenlosigkeit fortwährend geprüft werden. In dem vorgestellten Projekt konnte gezeigt werden, dass eine solche umfassende Erhebung Einsparungen ermöglichen kann, wenn Verbräuche den Gebäudenutzern – sei es Haustechnik, Dozentinnen und Dozenten oder anderen Akteuren – gespiegelt werden und sich daraus Maßnahmen ableiten lassen.

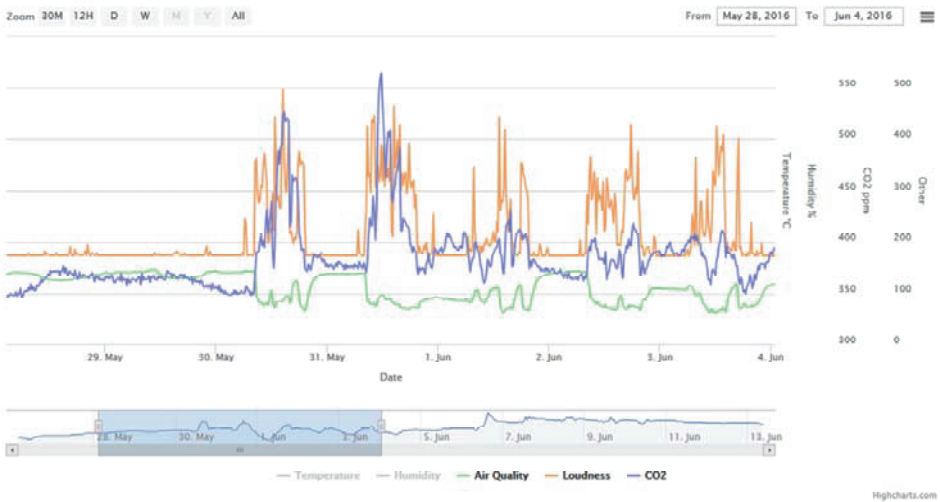


Abbildung 9. Messdatenverlauf und Visualisierung in Detaildarstellung (Script: highcharts.com)

Gleichzeitig hat sich gezeigt, dass es im Zusammenspiel von vorhandener Gebäudetechnik und eigenentwickelter Mess- und Auswertungstechnik erhebliche Schnittstellenprobleme gibt. So sind solche Gebäudeleittechnik-Systeme häufig proprietär, was sowohl die Wartung bei Änderungswünschen erschwert und mangels offener Schnittstellen den Austausch mit anderen Steuerungs- und Auswertungssystemen erschwert. Hier könnten offene Standards Vorteile beim Ermitteln von Einsparpotenzialen und auch bei der Steuerung bieten.

7 Danksagung

Die Autorinnen und Autoren danken dem Bundesministerium für Wirtschaft und Energie für das aufgrund eines Beschlusses des Deutschen Bundestages geförderte Projekt REGENA (Förderkennzeichen 03ET1070), sowie den REGENA-Projektpartnern IZES gGmbH, Saarbrücken, und A.U.G.E.-Institut, Hochschule Krefeld. Die Verantwortung für die Darstellung der Ergebnisse liegt ausschließlich bei den Autoren.

Literaturverzeichnis

- [WRK10] Wicaksono, H.; Rogalski, S.; Kusnady, E.: Knowledge-based Intelligent Energy Management Using Building Automation System, in: Institute of Electrical and Electronics Engineers (ed.): IPEC 2010, The 9th International Power and Energy Conference: 27-29 October 2010, Suntec, Singapore, pp. 1140–1145, 2010
- [KB06] Karatzas, K.; Bassoukos, Anastasios: ISBEEM: Indoor Environment and Energy Management for School Buildings, *EnviroInfo 2006* (Graz), pp. 169-176, 2010
- [Ko13] Kohoun, W.; Christian, A.; Naumann, S.; Gollmer, K.-U.; Ebner, I.; Michels, R.; Koch, P.; Guldner A.; Eigenstetter, M.; Jähn, V., Arns, S.; Groß, B.: Resource Efficiency in Buildings through Automation and User Integration (REGENA). In: Page, B.; Fleischer, A.; Göbel, J.; Wohlgemuth, V. (Eds.): *EnviroInfo2013 – Environmental Informatics and Renewable Energies*. 27th International Conference on Informatics for Environmental Protection, p. 484-491, 2013
- [LOP08] Pérez-Lombard, L.; Ortiz J., Pout, C.: A review on buildings energy consumption information, *Energy and Buildings*, Volume 40, Issue 3, 2008, Pages 394-398, ISSN 0378-7788, <http://dx.doi.org/10.1016/j.enbuild.2007.03.007>, 2008
- [Gr16] Greml A., Blümel, E., Kapferer, R; Leitzinger, W: Technischer Status von Wohnraumlüftungen, *Berichte aus Energie und Umweltforschung* 16, Wien (2004), Online unter http://www.hausderzukunft.at/hdz_pdf/endbericht_greml_id2746.pdf, 12.06.2016

Konzeptionierung einer Webanwendung zur Beurteilung von Stoff- und Energieströmen am Beispiel des Industrie- und Gewerbestandortes Berlin-Schöneeweide

Andrej Matusевич¹ und Volker Wohlgemuth²

Abstract: Die Planung, Entwicklung und das Management eines Industrie- und Gewerbestandortes unter Umwelt- und Nachhaltigkeitsaspekten stellt sowohl für Standortentwickler als auch Unternehmen eine große Herausforderung dar. Eine Unterstützung der Aufgaben durch IKT-gestützte Anwendungen scheint dabei nur logisch. Ausgehend von der Betrachtung von Hauptmerkmalen und Erfolgsfaktoren von Industriesymbiosen stellt dieser Beitrag ein Konzept für eine Webanwendung vor, die Standortentwickler im Voraus bei der Auswahl von Unternehmen für das zu entwickelnde Gebiet unterstützen soll. Unter dem Gesichtspunkt einer ganzheitlichen Analyse von Stoff- und Energieströmen und dem Vergleich mit den Daten von interessierten Unternehmen soll dieses Ziel erreicht werden. Neben Erfassungsmöglichkeiten von Unternehmen oder Materialien kann sich der Nutzer der Software durch den Einsatz von Sankey-Diagrammen und ausgewählten Kennzahlen einen Überblick über den gesamten Standort verschaffen.

Keywords: Industriesymbiosen, Stoffstrommanagement, Webanwendungen, BUIS

1 Einleitung

Das Einführen von geschlossenen Rohstoffkreisläufen zur Erhöhung der Ressourcen- und Materialeffizienz hat für ein relativ rohstoffarmes Land wie Deutschland eine große Bedeutung, um den gestiegenen Bedarf an Rohstoffen zu angemessenen Kosten zu decken. Sowohl in der nationalen Nachhaltigkeitsstrategien, als auch in den konkreten Stadtentwicklungskonzepten Berlins, wie der Smart City-Strategie Berlins, spiegelt sich dieses wichtige Kriterium wider. Eine Schlüsselrolle sollen dabei Informations- und Kommunikationstechnologien in allen Handlungsfeldern, wie z.B. Infrastruktur, Mobilität, Wirtschaft oder Wohnen, übernehmen, um Ressourceneffizienzziele zu erreichen [BU12; SE15].

Die Betrachtung und das Erreichen der oben genannten Ziele verfolgt das Stoffstrommanagement mittels der Analyse sowie des Managements von Stoff- und Energieströmen innerhalb eines bestimmten Systems. Eine spezielle Ebene bildet das vertikale zwischenbetriebliche Stoffstrommanagement, wobei mit vertikal die Zusammenarbeit mehrerer Branchen gemeint ist. Mit steigender Komplexität der untersuchten Systeme ist eine Unterstützung durch Informations- und

¹ HTW Berlin, FB 2, Wilhelminenhofstr. 75A, 12459 Berlin, a.matusевич@gmail.com

² HTW Berlin, FB 2, Wilhelminenhofstr. 75A, 12459 Berlin, volker.wohlgemuth@htw-berlin.de

Kommunikationstechnologien unausweichlich, um Stoffstrommodelle abzubilden oder Kooperationen mehrerer Akteure zu erleichtern [HB02].

Das in diesem Forschungsprojekt betrachtete System in Berlin Schöneeweide hat eine Größe von ca. vier Quadratkilometern und gilt als Standort für Firmen unterschiedlichster Branchen. Aus diesem Grund sollen in diesem Beitrag auch die Forschungsergebnisse der Verwertungsnetzwerke und industriellen Symbiosen mit einbezogen werden, die in der bisherigen Praxis zur Koordinierung von Stoff- und Energieströmen zwischen Unternehmen in Industrie- und Gewerbegebieten dienten.

Dementsprechend soll hier ein Konzept für eine webbasierte Anwendung vorgestellt werden, die unter Berücksichtigung von Stoff- und Energieströmen entscheidungsunterstützend, im Sinne der Auswahl neuer Unternehmen und unter Berücksichtigung der bisherigen Unternehmen, für Standortentwickler fungieren soll. Das Management eines Standortes soll in die Lage versetzt werden, mögliche Kooperationen auf stofflicher oder energetischer Basis aufzudecken und erste Schritte einer Vernetzung einzuleiten.

Durch die Analyse der stofflichen und energetischen Inputs und Outputs der verschiedenen Branchen am Standort soll eine Aussage über die bestmögliche Zusammensetzung der anzusiedelnden Unternehmen unter stofflichen und energetischen Aspekten für das Industrie- und Gewerbegebiet Oberschöneeweide getroffen werden können. Dazu wird die Literatur zur Entwicklung von nachhaltigen Industrie- und Gewerbegebieten vorgestellt, um Hauptmerkmale von Industriesymbiosen herauszuarbeiten und Erfolgsfaktoren zu identifizieren (Kap. 2.1). Ein Status-Quo der Software-Lösungen, die die Planung und Entwicklung von nachhaltigen Gewerbegebieten unterstützen, wird ermittelt (Kap. 2.2).

Im letzten Kapitel wird das Konzept für die Webanwendung, die als Entscheidungsunterstützung bei der Auswahl anzusiedelnder Unternehmen verwendet werden kann, vorgestellt (Kap. 3).

2 Grundüberlegungen zu Industriesymbiosen

Die Analyse und Entwicklung von Industrie- und Gewerbegebieten unter Nachhaltigkeitsaspekten ist ein Hauptforschungsfeld der Industrial Ecology und bildet somit das theoretische Grundgerüst dieser Arbeit [Vh12]. Dieses noch relativ junge Forschungsgebiet hat den Schwerpunkt in der Analyse von industriellen Systemen, den Produktionsprozessen innerhalb der Systeme und der Gestaltung nachhaltiger Produkte sowie der Analyse des Konsumverhaltens der Abnehmer der Produkte [HIZ12]. Besonders charakteristisch für die Ideen aus der Industrial Ecology ist ein systemischer Ansatz. Dies bedeutet, dass industrielle Systeme nicht einzeln betrachtet werden, sondern samt ihrer Interdependenzen mit der Umwelt und anderen Systemen zu erforschen sind. Dazu bedient man sich einer biologischen Analogie, bei der

wirtschaftliche Aktivitäten in industriellen Systemen in Zusammenhang zu natürlichen Ökosystemen gesetzt werden, um auf dieser Basis Austauschbeziehungen mit der Umwelt aufzuzeigen. Dies geschieht vorwiegend durch die Auswertung von Stoff- und Energieströmen in den Systemen mit dem Ziel, Konsequenzen auf die Umwelt zu erarbeiten [Je92;Wh94]. Ebenso sind Lebenszyklusanalysen und dynamische Systemmodellierungen Hauptinstrumente, die bei den Lösungsansätzen der Industrial Ecology eingesetzt werden [LG02]. Neben den ökologischen Auswirkungen untersucht die Industrial Ecology auch wirtschaftliche, politische, gesetzliche und soziale Variablen in industriellen Systemen und unterstreicht damit den interdisziplinären Charakter (z.B. Ingenieurwissenschaften, Wirtschaftswissenschaften, etc.) des Forschungsgebiets [Je92].

2.1 Hauptmerkmale von Industriesymbiosen

Die Planung, Analyse und Weiterentwicklung von Industriesymbiosen ist ein Hauptzweig der Industrial Ecology, wobei das Hauptaugenmerk auf der ganzheitlichen Betrachtung von Stoff- und Energieflüssen auf lokaler oder regionaler Ebene liegt [MJZ14]. In der Literatur tauchen diese Industriesymbiosen unter verschiedenen Bezeichnungen auf. Zu den wichtigsten gehören z.B. so genannte Eco-Industrial-Parks, regionale Recyclingnetzwerke, Zero-Emission-Parks oder nachhaltige Industrie- und Gewerbegebiete.

Bei diesen Gebieten wird darauf Wert gelegt, verschiedene dort ansässige Unternehmen und Dienstleister zu Austausch- und Kooperationsbeziehungen zu bewegen. Vor allem soll dies auf Ebene des Material-, Wasser- und Energieeinsatzes, der Verwertung von Abfällen oder bei der gemeinsamen Nutzung von Infrastruktur realisiert werden, um die jeweiligen Produktionskosten der Akteure zu senken und schädliche Umweltauswirkungen zu reduzieren [Is13; Ch07].

Voraussetzung für ein gutes Ergebnis des Industriestandortes, im Hinblick auf Einsparungen beim Ressourceneinsatz und der Reduktion des Abfalls sowie möglicher Kostensenkungen, ist ein umfangreiches Wissen der Stoff- und Energieströme in dem betrachteten Gebiet bzw. Standort. Solche unternehmensübergreifenden Kooperationen sind oft nicht unproblematisch, da unternehmensinterne Informationen nach außen abgegeben werden, verschiedene Unternehmen unterschiedliche Ziele verfolgen oder Firmen an Eigenständigkeit in der Produktionsplanung verlieren würden. Werden die Daten über Input- und Outputströme weitergegeben, so liegen sie bei verschiedenen Unternehmen in unterschiedlicher Form vor oder können lückenhaft übermittelt werden, was sich negativ auf die Datenverarbeitung niederschlägt [Mü12].

[Is13], [Mü12], [MJZ14] und [Ch07] fassen folgende Hauptfaktoren für eine erfolgreiche Industriesymbiose zusammen:

- Freiwillige Zusammenarbeit
- Gemeinsame Nutzung von Ressourcen, um Synergiepotenziale auszunutzen

- Partnerunternehmen haben ein kritisches Produktionsvolumen
- Keine Konkurrenzängste zwischen den Unternehmen
- Kurze räumliche Distanz
- Ökonomische Vorteile

Hieraus ist ersichtlich, dass der Managementaufwand von Verwertungsnetzwerken sehr hoch ist und ein Anreiz für Unternehmen, sich an speziell diesen Verbänden zu beteiligen, nur dann besteht, wenn im Gegenzug Sicherheit über die Abfallentsorgung, Materialbeschaffung und Verminderung von Umweltschäden besteht sowie Kostenreduktionen möglich sind [Mü12].

Eine übergeordnete Institution, die die Kooperation und Kommunikation zwischen den Unternehmen organisiert, stellt einen wichtigen Schlüsselfaktor dar. Solch ein Organ kann die Stoff- und Energieströme der Firmen erfassen, dokumentieren und mögliche Austauschbeziehungen vorschlagen. Des Weiteren sind Beziehungen zwischen Hochschulen, Technologiezentren und den wirtschaftlichen Partnern für einen effektiven Wissensaustausch förderlich [MJZ14].

Hinsichtlich einer Vernetzung von heterogenen Unternehmen zur Optimierung von Stoff- und Energieflüssen in einem Industrie- und Gewerbegebiet wird als Beispiel immer wieder die Industriesymbiose von Kalundborg (Dänemark) genannt [MJZ14]. So partizipieren in diesem Netzwerk insgesamt neun Akteure, wie die Stadt Kalundborg selbst, Insulin- und Enzymproduzenten, ein Kraftwerk und eine Ö raffinerie, um nur einige Beispiele zu nennen. Die Kooperationen und Austauschbeziehungen zwischen den Unternehmen begannen in den 1960er Jahren aufgrund eines Wassermangels in der Region, der sie dazu zwang, die Nutzung des Grundwassers zu reduzieren und nach gemeinsamen Alternativen zu suchen. Im weiteren Verlauf entwickelte sich eine Kreislaufwirtschaft, die die Verwertung von Abfall und die intelligente Verteilung von Energie – aus Gründen von Kostenvorteilen und der Reduktion von schlechten Umweltauswirkungen - mit einschloss. So werden z.B. Gase, Abwärme, Schlamm oder Gips zwischen den Unternehmen ausgetauscht. Als Haupterfolgskfaktoren der Industriesymbiose werden die offene Kommunikation, Vertrauen unter den Akteuren, die heterogene Struktur und wirtschaftliche Vorteile für die Firmen angesehen [EG97].

Dieses Beispiel steht stellvertretend für viele weitere Ansätze zur Entwicklung von nachhaltigen Industrie- und Gewerbegebieten auf der Welt. [MJZ14] untersuchten in ihrer Studie insgesamt 168 Beispiele für Eco-Industrial-Parks in 27 Ländern, hoben die Bedeutung von informationstechnischen Lösungen zur Unterstützung des Managements von Gewerbegebieten hervor und unterstreichen damit den Forschungsbedarf auf diesem Wissenschaftsgebiet.

2.2 Stand der Technik

Im folgenden Abschnitt wird es darum gehen, bereits bestehende Software-Produkte, die bei der Entwicklung von Industriesymbiosen eingesetzt werden können, zu analysieren. IKT-gestützte Anwendungen sollen einerseits dazu beitragen, den Zugang zu aufbereiteten Informationen zu erleichtern [Is13], andererseits können informationstechnische Lösungen die Akteure eines Standortes bei der Organisation der Austauschbeziehungen (z.B. bei Materialien, Wasser, Energie oder Abfall) unterstützen und dadurch zu einer Erhöhung des Vertrauens im Industrie- und Gewerbegebiet beitragen [IC09].

[Gr10] setzten sich intensiv mit 17 identifizierten Softwareprodukten auseinander. Als Ergebnis der Untersuchung ist zu nennen, dass neun Systeme bereits zu der Zeit nicht mehr verwendet werden. Vier weitere können genutzt werden, sind aber nicht öffentlich zugänglich. Diese Anwendungen befanden sich teilweise im Entwicklungsstadium und werden hauptsächlich in Kanada oder Australien genutzt. Die Einsatzgebiete der Software reichten von der Nutzung als geografische Planungsanwendungen über die Organisation eines Abfallmanagements hin zu komplexeren Datenbankanwendungen innerhalb einer Betrachtung eines ganzen Landes. Auch bei diesen Anwendungen liegen hohe Anforderungen an die Einfachheit in der Bedienung vor. Außerdem müssen Budgetbeschränkungen, der Schulungsaufwand und fachliche Gutachten berücksichtigt werden. Die Autoren konstatierten in ihrer Untersuchung zudem, dass die umgesetzten Softwareansätze bei der Integration sozialer Komponenten, wie dem Austausch von Informationen zwischen den Mitgliedern der Industrie- und Gewerbegebiete zur gegenseitigen Vertrauensbildung, weiterer Bearbeitung bedürfen. Dem gegenüber stehen gute Lösungsansätze im Hinblick auf den Ressourcenaustausch zwischen den beteiligten Akteuren.

Die Erkenntnisse aus der Entwicklung von betrieblichen Umweltinformationssysteme (BUIS), die einen Schnittbereich mit der Industrial Ecology besitzen, können ebenfalls mit einbezogen werden [Is13]. BUIS haben im betrieblichen Umweltmanagement neben den Aufgaben der Informationsbeschaffung, Dokumentation, Bewertung und Entscheidungsunterstützung auch im Bereich der der Steuerung, Planung und Kontrolle von Umweltschutzmaßnahmen Fragestellungen zu lösen und bedienen sich dabei Instrumenten, wie der Erstellung von Input-Output-Bilanzen zur Ökobilanzierung und Stoffstromanalyse. Zur Modellierung und Analyse von Stoff- und Energieströmen zur Ökobilanzierung oder Effizienzsteigerung kann als Beispiel eines BUIS das Software-Werkzeug Umberto oder die auf die Darstellung von Sankey-Diagrammen spezialisierte Anwendung e!Sankey, genannt werden [Hi95;Ra99]. Vergleichbare Lösungsansätze findet man bei dem Software-Produkt GaBi, das unter anderem im Bereich der Produktlebenszyklus-Analysen eingesetzt wird [TH16], dem Stoffstrommanagement Tool Umberto der ifu Hamburg GmbH [IF16] oder in der Open-Source-Alternative der Technischen Universität Wien STAN (subSTance flow ANalysis) [TW12].

Im Hinblick auf eine Erhöhung der Ressourceneffizienz in der verarbeitenden Industrie entwickelte die Effizienz Agentur NRW einen webbasierten Quick-Check von verschiedenen Leistungsmerkmalen. Hauptbestandteil des Quick-Checks ist ein Online-Fragebogen, der Informationen über Verbräuche von Materialien und Energien sammelt sowie durch einen Datenabgleich mit anderen Unternehmen der Branche die Materialeffizienz des eigenen Unternehmens bewertet. Auf dieser Grundlage werden maßnahmenunterstützende Dokumente zur Verfügung gestellt [EF16].

Für die von Darstellung von Sankey-Diagrammen in einer Webanwendung hat sich die Open-Source d3-JavaScript-Bibliothek als eine Alternative erwiesen, die eine direkte Datenmanipulation ermöglicht [Ma15].

3 Konzept der Webanwendung

3.1 Grundidee

Der Verwalter des Standortes hat zum Ziel interessierten Unternehmen aufzuzeigen, welche Vorteile eine Ansiedlung in Schöneweide unter stofflichen und energetischen Gesichtspunkten haben kann und somit die Profilbildung des Industrie- und Gewerbegebiets mit dem Tool als Anreizwerkzeug zu unterstützen. Die zu entwickelnde Anwendung kann bei der Identifikation von Möglichkeiten der Zusammenarbeit eingesetzt werden und einen ersten Hinweis darüber geben, ob ein Unternehmen oder Dienstleister am Standort in Kooperationsbeziehungen treten kann.

Die Vorteile dieser Anwendung lassen sich dabei sowohl auf der Seite des Regionalmanagements als auch auf Unternehmensseite generieren. Für den Standortentwickler kann dieses Tool dazu dienen, den stofflichen und energetischen Ist-Zustand in Schöneweide im Überblick zu haben. Aufgrund der Erfassung der Unternehmensdaten kann die Glaubwürdigkeit und Überzeugungsargument des Regionalmanagements gegenüber anderen Unternehmen gestärkt werden und eine einfachere Kontaktvermittlung zwischen Unternehmen erfolgen. Ein Nebeneffekt wäre der Aufbau einer umfassenden Firmendatenbank für das gesamte Gebiet. Als entscheidungsunterstützende Komponente der Anwendung kann die Tatsache angesehen werden, dass aufgrund ermittelter (oder nicht ermittelter) Übereinstimmungen von Inputs und Outputs die Integrierbarkeit von Unternehmen Industrie- und Gewerbegebiet aus stofflicher und energetischer Sicht im Voraus bewertet werden kann.

Auf der anderen Seite kann das Tool interessierten Unternehmen einen Anreiz geben, eine Stoffstromanalyse durchzuführen, um am Netzwerk aktiv teilzuhaben und von den Vorteilen des Stoff- und Energieaustausches zu profitieren. Hilfestellungen des Regionalmanagements und anderer Unternehmen können in dieser Hinsicht förderlich sein. Darüber hinaus können Dienstleistungsunternehmen ebenso ihren Service als Output anbieten und somit von anderen Unternehmen in Schöneweide genutzt werden.

3.2 Aufbau und Funktionen der Anwendung

Der grundsätzliche Aufbau des Tools wird anhand einer Single Page Application erfolgen. Demnach erfolgt die Umsetzung der grafischen Oberfläche mit Hilfe von HTML5 und CSS sowie des Bootstrap-Frameworks. Die Trennung der grafischen Oberfläche und der Logik wird mit Hilfe des Model View ViewModel-Musters (MVVM) erreicht. Zur Unterstützung der Softwareentwicklung nach dem MVVM-Muster wird das Framework AngularJS eingesetzt.

Die Webanwendung wird einen zweiteiligen Aufbau besitzen, der aus der oberen Navigationsleiste und größeren Arbeitsbereich bestehen wird (vgl. Abb. 1).

The image shows a navigation bar at the top with icons and labels for: Dashboard, Standort, Unternehmen, Rohstoffe, Kooperation, and Kontakt. Below the navigation bar are two forms:

Add new Business

Name des Unternehmens:

Ansprechpartner:

männlich:

weiblich:

Branchen-Typ:

E-Mail:

Add Input

Rohstoff:

Menge:

Einheit:

Bezugsgröße:

Abb. 1: Grundsätzliche Struktur der Anwendung am Beispiel der Erfassung eines Unternehmens

Das Dashboard hat in erster Linie eine Informationsfunktion zur gesamten Situation des Gebiets. Dazu kann der Anwender zwischen einer Sankey-Ansicht mit ein- und ausgehenden Stoff- und Energieflüssen am Standort und einer Input-Output-Bilanz wechseln können. Das Sankey-Diagramm soll dabei als eine Mengen-Input-Output-Bilanz [Sc12] für den gesamten Standort dargestellt werden, wobei mögliche stoffliche oder energetische Rückflüsse aufgrund von Austauschbeziehungen ebenfalls sichtbar werden. Erweitert kann das Dashboard durch Informationen zu bestehenden Kooperationen zwischen Unternehmen oder (Umwelt-)Kennzahlen zum Standort.

Im Menüpunkt „Standort“ sollen auf einer Karte die Positionen der einzelnen Unternehmen erkennbar sein, um z.B. die Entfernung einschätzen zu können. Die Auswahl der anzuzeigenden Firmen kann über Combo-Boxen erreicht werden.

Die Unternehmensverwaltung stellt einen zentralen Bereich der Anwendung dar. Grundsätzlich kann zwischen drei Menüunterpunkten ausgewählt werden. Erfassung eines neuen Unternehmens und der jeweiligen Stoffe und Energien ist in Abb. 1 exemplarisch abgebildet. Diese Abschnitte sollen durch +/- Buttons minimiert bzw. wieder aufgeklappt werden können. Die Bearbeitung erfolgt dann nach demselben Muster. Das Klicken auf den Menüpunkt „Unternehmen“ soll eine Liste der Unternehmen am Standort aktiven Unternehmen, sortiert nach Branchen, dargestellt werden. Über diesen Punkt erfolgt der eigentliche Vergleich der Inputs und Outputs der Unternehmen, in dem ein Unternehmen ausgewählt und mit den Outputs anderer Unternehmen verglichen wird. Im Zuge dessen wird eine neue Seite geöffnet, um die Ergebnisse des Vergleichs übersichtlich darzustellen und abspeichern zu können.

Vergleichbar mit der Unternehmensverwaltung, können in diesem Menü neue Materialien und Stoffe verwaltet werden. Die hier angelegten Stoffe können dann bei der Erfassung der stofflichen und energetischen Ströme ausgewählt werden. Dieser Abschnitt würde an Bedeutung verlieren, wenn eine Anbindung an eine Stoffdatenbank gelingt.

Über den Menüpunkt Kooperationen kann der Nutzer eine Übersicht der Kooperationsbeziehungen erhalten und weitere Details der Zusammenarbeit erfahren. Denkbar sind Aspekte wie die ausgetauschten Stoffe und Mengen oder eine Auskunft über den Kooperationsbeginn.

Im letzten Navigationspunkt (Kontakt) werden sämtliche Kontaktpersonen der Unternehmen in Form einer Liste gesammelt, um dem Standortentwickler, die Kontaktaufnahme zu Firmen und die Kontaktvermittlung zwischen Unternehmen zu verbessern.

Funktionsbereich	Unterfunktion
Nutzerverwaltung	Login/ Rechtemanagement
Anzeige & Visualisierung	Input-Output-Bilanzen Sankey-Diagramm Unternehmen und Kooperationen Standortanzeige
Branchen- und Unternehmensverwaltung	Neu/Bearbeiten/Löschen
Stoffe und Energien	Neu/Bearbeiten/Löschen Anschluss an Stoffdatenbank
Auswertungsfunktion	Vergleich von Bilanzen Bewertung Kooperationsmöglichkeiten
Dokumentation	Berichte (Kooperationen & Bilanzen) Exportfunktion

Tab. 1: Funktionsübersicht

Tab. 1 fasst die angedachten Funktionen der Webanwendung zusammen. Laut diesem Konzept wird das Tool sechs Komponenten beinhalten, die die oben genannten Ziele realisieren sollen. Die Kernpunkte der Anwendung werden in den Funktionsbereichen der Analyse und Visualisierung, der Erfassung von Unternehmen und Stoffen bzw. Energien sowie der Auswertung liegen. Ein weiterer wichtiger Aspekt betrifft die Möglichkeit eines Exports von Übersichten zu Unternehmen, Kooperationen und dem Gesamtzustand des Standortes.

3.3 Datenmodell

Die in dem Datenbankentwurf dargestellten Entitäten repräsentieren alle Daten, die für die vorgestellten Funktionen benötigt werden.

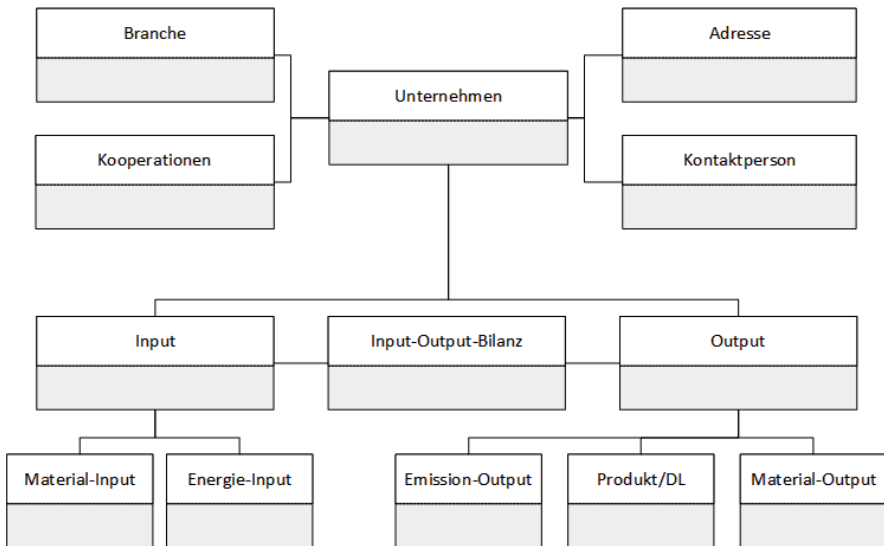


Abb. 2: Datenstruktur der Anwendung

Die Ermittlung der erforderlichen Daten geht insbesondere auf einen erstellten Fragebogen zur Erfassung der Inputs und Outputs von Unternehmen zurück.

Aus der Abbildung 2 wird deutlich, dass die Tabelle Unternehmen viele Daten vereinigen muss. Neben den allgemeinen Informationen zur Branchenzugehörigkeit oder den Kontaktpersonen, die über Fremdschlüssel an die Unternehmenstabelle gekoppelt sind, spielen Daten zu den eingehenden und ausgehenden Stoffen eine besondere Rolle. Die Inputs und Outputs werden zum einen einem bestimmten Unternehmen zugeordnet und finden sich zum anderen in der Input-Outbilanz des relevanten Unternehmens wieder. Um eine gewisse Übersichtlichkeit und Strukturierung in der Bilanz zu haben, sind die Input- und Outputtabellen darüber hinaus untergliedert. Während in der Tabelle

Material-Input Roh-, Hilfs- und Betriebsstoffe erfasst werden, so sind bei der Energie-Input-Tabelle Werte des Strom-, Wärme- und Wasserbedarfs von Interesse.

Ähnlich verhält es sich auf der Outputseite, wobei eine neue Tabelle über die hergestellten Produkte oder die bereitgestellten Dienstleistungen wichtig wird. Unter dem Material-Output werden sämtliche Abfälle verstanden. Demgegenüber stehen die stofflichen Emissionen, wie Abwasser oder Abluft, und die energetischen Emissionen (z.B. Abluft). Die outputseitigen Tabellen haben als Gemeinsamkeit, dass bei jedem Stoff oder Emissionen angegeben werden muss, ob es zur Wiederverwertung oder zum Verkauf bereitsteht. Eine Zeile über den Zustand des Materials und inwieweit es direkt in eine Produktion fließen kann, soll ebenfalls integriert werden. Eine mögliche zwischenbetriebliche Zusammenarbeit zwischen Unternehmen wird in der Kooperationstabelle erfasst.

4 Fazit & Ausblick

Im Laufe der Ausführungen dieses Papers ist klargeworden, welche Faktoren bei der Planung und Entwicklung eines Industrie- und Gewerbestandortes eine Rolle spielen. Dabei gilt es verschiedene Interessenlagen aller Akteure (Unternehmen unterschiedlicher Branchen und Standortentwickler) zu berücksichtigen. Eine Beurteilung von Unternehmen hinsichtlich ihrer Stoff- und Energieströme im Vorfeld der Ansiedlung an einem Standort stellt eine große Herausforderung dar. Zum Lösen dieses vielschichtigen Problems im Rahmen eines zwischenbetrieblichen Stoffstrommanagements soll die hier vorgestellte Anwendung beitragen, in dem ein Vergleich von stofflichen Inputs und Outputs unterschiedlicher Unternehmen erfolgen soll. Die inhaltliche und visuelle Aufbereitung soll mit Hilfe von Sankey-Diagrammen, Kennzahlen und Tabellen erfolgen. Als Ergebnis sollte eine Auskunft darüber gegeben werden, in wieweit ein Unternehmen am Standort basierend auf dem Ist-Zustand mit anderen Unternehmen kooperieren kann. Dies soll die Arbeit des Standortentwicklers auch bei der Überzeugungsarbeit erleichtern.

Ein logischer Schritt nach Fertigstellung des fachlichen Konzeptes der Webanwendung liegt in der kommenden Entwicklung eines Prototyps der Software und den damit verbundenen Aufgaben. Dazu zählen beispielsweise Umsetzung der Oberfläche, die Integration der Funktionalitäten oder das Einbinden einer Stoffdatenbank in die Software. Die prototypische Entwicklung der Anwendung ist zu diesem Zeitpunkt bereits angelaufen. Der Fokus liegt dabei zurzeit in dem Einbinden der Sankey-Diagramme in die Anwendung. Im Verlauf eines sich anschließenden Projektes wird die Erfassung der Stoff- und Energiedaten in Zusammenarbeit mit dem Regionalmanagement Schöneweide und den Unternehmen am Standort im Vordergrund stehen. Gerade für die Erstellung des stofflichen und energetischen Ist-Zustandes ist der Kontakt zu interessierten Firmen herzustellen, die zur Schaffung der Datengrundlage

beitragen sollen, um mögliche Kooperationen sichtbar zu machen und Vergleichsdaten für zukünftige Szenarien zu haben.

Literaturverzeichnis

- [BU12] Die Bundesregierung: Nationale Nachhaltigkeitsstrategie - Fortschrittsbericht 2012 http://www.bundesregierung.de/Content/DE/Publikation/Bestellservice/2012-05-08-fortschrittsbericht-2012.pdf?__blob=publicationFile. Abgerufen am: 10.05.2016.
- [Ch07] Chertow, M.: Uncovering industrial symbiosis. *Journal of Industrial Ecology* 11(1), S. 11-30, 2007.
- [EF16] Effizienzagentur NRW (EFA): Ökobench. <http://www.oekobench.de/index.php/ueber-die-datenbank>. Abgerufen am: 08.05.2016.
- [EG97] Ehrenfeld, J.; Gertler, N.: Industrial Ecology in practice: the evolution of interdependence in Kalundborg. *Journal of Industrial Ecology* 1(1), S. 67-79, 1997.
- [Gr10] Grant, GB., et al.: Information and Communication Technology for Industrial Symbiosis. *Journal for Industrial Ecology* 14(5), S. 740-753, 2010.
- [HB02] Heck, P.; Bemann, U.: *Praxishandbuch Stoffstrommanagement: Strategien – Umsetzung – Anwendung in Unternehmen, Kommunen, Behörden*. Deutscher Wirtschaftsdienst, Köln, 2002.
- [Hi95] Hilty, LM.: Betriebliche und überbetriebliche Umweltinformationssysteme als informationstechnische Infrastruktur für das Stoffstrommanagement. In (Schmidt, M.; Schorb, A., Hrsg.): *Stoffstromanalysen in Ökobilanzen und Öko-Audits*, Springer Verlag, Berlin Heidelberg, S. 193-205, 1995.
- [HIZ12] von Hauff, M.; Isenmann, R.; Müller-Christ, G.: *Industrial Ecology Management – Nachhaltige Entwicklung durch Unternehmensverbände*, Springer Gabler, Wiesbaden, 2012.
- [IC09] Isenmann, R.; Chernykh, K.: Environmental ICT applications for eco-industrial development. In (Wohlgemuth, V.; Page, B.; Voigt, K., Hrsg.): *Environmental Informatics and industrial environmental protection. 23rd international conference on informatics for environmental protection*, Shaker Verlag, Aachen, S. 231–242, 2009.
- [IF16] ifu Hamburg GmbH, Umberto. <http://www.umberto.de/de/>. abgerufen am: 21. 03. 2016.
- [Is13] Isenmann, R.: Beitrag betrieblicher Umweltinformatik für die Industrial Ecology – Analyse von BUIS-Software-Werkzeugen zur Unterstützung von Industriesymbiose. In (Gómez, JM.; Lang, C.; Wohlgemuth, V., Hrsg.): *IT-gestütztes Ressourcen- und Energiemanagement*, Springer Verlag, Berlin Heidelberg, S. 397-407, 2013.
- [Je92] Jelinski, LW. et al.: Industrial Ecology: Concepts and Approaches. *PNAS* 89(3), S. 793–797, 1992.

- [LG02] Lifset, R.; Graedel, TE.: Industrial Ecology: goals and definitions. In (Ayres, RU.; Ayres, L., Hrsg.): A handbook of industrial ecology. Edward Elgar Pub., Cheltenham, UK und Northampton, S. 3–15, 2002.
- [Ma15] Maclean, M.: D3 Tips and Tricks - Interactive Data Visualization in a Web Browser, 2015.
- [MJZ14] Massard, G.; Jacquat, O.; Zürcher, O.: International survey on ecoinnovation parks. Learning from experiences on the spatial dimension of eco-innovation. Federal Office for the Environment and the ERANET ECO-INNOVERA, Bern, 2014.
- [Mü12] Müller-Christ, G.: Vom Industrial Ecology Management zur Entwicklung nachhaltiger Gewerbegebiete. In (von Hauff, M.; Isenmann, R.; Müller-Christ, G., Hrsg.): Industrial Ecology Management – Nachhaltige Entwicklung durch Unternehmensverbände, Springer Gabler, Wiesbaden, S. 57-74, 2012.
- [Ra99] Rautenstrauch, C.: Betriebliche Umweltinformationssysteme - Grundlagen, Konzepte und Systeme, Springer Verlag, Berlin Heidelberg, 1999.
- [Sc12] Schmidt, M.: Visualisierung von Energie- und Stoffströmen. In (von Hauff, M.; Isenmann, R.; Müller-Christ, G., Hrsg.): Industrial Ecology Management – Nachhaltige Entwicklung durch Unternehmensverbände, Springer Gabler, Wiesbaden, S.257-272, 2012.
- [SE15] SenStadtUm Berlin: Smart City-Strategie Berlin. https://www.berlin-partner.de/fileadmin/user_upload/01_chefredaktion/02_pdf/02_navi/21/Strategie_Smart_City_Berlin.pdf. Abgerufen am: 09.05.2016
- [TH16] thinkstep GaBi. <http://www.gabi-software.com/deutsch/software/gabi-software/>. abgerufen am: 21. 03. 2016.
- [TW12] TU Wien. <http://www.stan2web.net/infos/about-stan>. abgerufen am: 21. 03.. 2016.
- [Vh12] Von Hauff, M.: Von der Ökologischen Ökonomie zur Industrial Ecology Science. In (von Hauff, M.; Isenmann, R.; Müller-Christ, G., Hrsg.): Industrial Ecology Management – Nachhaltige Entwicklung durch Unternehmensverbände, Springer Gabler, Wiesbaden, S. 31-42., 2012.
- [Wh94] White, RM.: Preface. In (Allenby, BR.; Richards, DJ, Hrsg): The greening of industrial ecosystems. National Academy Press, Washington D.C., 1994.

MOPS - Mobile Unterstützung eines prozessorientierten Stoffstrommanagements in KMU

Maximilian Schneider¹, Volker Wohlgemuth²

Abstract: Der vorliegende Beitrag beschreibt die Zielsetzung und Konzeption einer mobilen Anwendung zur Unterstützung des prozessorientierten Stoffstrommanagements in kleinen und mittleren Unternehmen. Mit Hilfe von Sankey-Diagrammen sollen Fragen der Ressourceneffizienz in produzierenden Unternehmen durch eine einfache vor Ort Veranschaulichung von Material- und Energieströmen realisiert und des Weiteren die Einstiegshürde für eine Transparenzsteigerung der Produktionsprozesse verringert werden, indem die Vorteile der mobilen Erfassung von Stoff- und Energieströmen an einer Produktionsanlage genutzt werden.

Keywords: Stoffstrommanagement, Stoffstromvisualisierung, Prozessoptimierung, Sankey-Diagramm, KMU, Hybrid-App

1 Motivation und Zielsetzung

Wenn es um umweltschutzbezogene Zukunftsbilder für die entwickelten Industriegesellschaften geht, prägt seit über einem Jahrzehnt vor allem ein Begriff die Diskussion: Sustainable Development. Als eine Methode der konsequenten Umsetzung des Leitbildes der nachhaltigen Entwicklung in der betrieblichen Praxis hat sich das Stoffstrommanagement durchgesetzt, welches Methoden zur Steigerung der Ressourcen- und Energieeffizienz bereithält. Eine Möglichkeit, um die Materialeffizienz zu erhöhen, also die Summe des Inputs im Vergleich zum Output zu verringern, ist neben anderen die Optimierung von Produktionsprozessen [BW09]. Auch in der politischen Ausrichtung der Bundesrepublik Deutschland (Deutsches Ressourceneffizienzprogramm II) ist das Ziel einer ressourcenschonenderen Produktion in den kommenden Jahren als Handlungsfeld verankert [BU16].

Um die Potentiale zur Steigerung der betrieblichen Material- und Ressourceneffizienz in einem Unternehmen zu nutzen, ist es erforderlich, einen Betrieb individuell zu analysieren und daraus unternehmensspezifische Maßnahmen zur Effizienzsteigerung abzuleiten und damit das Stoffstrommanagement im Unternehmen zu verankern. In einem ersten Schritt müssen hierzu Daten über relevante Input- und Outputströme der Produktionsprozesse gesammelt und erfasst werden. Dazu gehören Daten über entsprechende Mengenströme von Roh-, Hilfs- oder Betriebsstoffen und der Bedarf von

¹ HTW Berlin, Fachbereich 2 – Betriebliche Umweltinformatik, Wilhelminenhofstr. 75 A, 12459 Berlin, Maximilian.Schneider@htw-berlin.de

² HTW Berlin, Fachbereich 2 – Betriebliche Umweltinformatik, Wilhelminenhofstr. 75 A, 12459 Berlin, Volker.Wohlgemuth@htw-berlin.de

thermischer oder elektrischer Energie auf der Inputseite der Produktionsprozesse, sowie gewünschter Güter (Produkte) bzw. unerwünschter Güter (Abwasser, Abfälle, Emissionen etc.) auf der Outputseite. Ziel hierbei ist das Zerlegen des Gesamtunternehmens in Prozesse und die verursachungsgerechte Zuordnung der Stoff- und Energieströme zu den Prozessen der Wertschöpfung innerhalb des Betriebes. Durch die Erhebung von Stoff- und Energieflussdaten im Betrieb ändert sich die Betrachtungsweise von einer „Black-Box“ hin zu einer durchschaubaren „White-Box“. Dadurch werden eine größere Transparenz und ein besseres Verständnis über die unternehmerischen Produktionssysteme erlangt [HB02]. Durch die ganzheitliche Betrachtung ist das Stoffstrommanagement daher ideal geeignet, stoffliche und energetische Optimierungspotentiale aufzudecken und umzusetzen und damit gleichzeitig ökonomische Verbesserungen zu erzielen, insbesondere unter Beachtung der aktuell hohen Rohstoff- und Energiekosten [Cr11].

Praktisch genutzte Methoden zur prozessorientierten Unterstützung der Datenerfassung des Stoffstrommanagements sind die Material Stream Mapping (MSM) [Re10] und die Environmental Value Stream Mapping (eVSM)-Methode [EP09]. Beide sehen eine systematische Erfassung von Stoff- und Energieströmen vor, werden aber bisher vornehmlich mit Papier und Bleistift durchgeführt oder aus Betriebsdatenerfassungssystemen gespeist, die bei kleinen und mittleren Unternehmen (KMU) häufig nicht vorhanden sind [Wo10]. Die Prozess- und Datenaufnahme ist komplex, verlangt methodisches Wissen, ist durch den hohen Zeitaufwand und Datenbedarf kostenintensiv und stellt somit eine hohe Einstiegshürde für KMU dar [BB00]. KMU bieten aufgrund ihrer Anzahl insgesamt jedoch ein bedeutendes Potential zur Steigerung der Ressourceneffizienz in Deutschland [VD11]. Neben der fehlenden Transparenz über die Stoffströme ist ein weiteres Problem, dass Stoffstromdaten normalerweise nicht als solche bzw. nicht zentral gesammelt vorliegen.

Praktische Ressourceneffizienz erfordert Maßnahmen und damit Entscheidungen für diese Maßnahmen. Voraussetzung dafür ist das Erkennen und Quantifizieren von Verbesserungspotentialen und die Kommunikation dieser Potentiale an die Entscheidungsträger. Für die Kommunikation komplexer Sachverhalte in Produktionssystemen haben sich Sankey-Diagramme bewährt. Sie können Material- und Energieströme im bestehenden Produktionssystem und die Auswirkung von Maßnahmen proportional visualisieren. Sie eignen sich auch besonders zur Kommunikation von Umweltwirkungen und Kosten in solchen Systemen.

Zentrale Aktivitäten zur Ressourceneffizienzsteigerung in Unternehmen werden vor allem durch Effizienzleitfäden und Handlungsrichtlinien umgesetzt, da es bisher keine Lösungen gibt, die KMU hinsichtlich Fragestellungen zur Ressourceneffizienz in geeigneter Weise zu unterstützen. Laut Erfahrungen der Autoren und der Projektpartner (ifu Hamburg und PIUS-Netzwerk³) setzen solche Effizienzanalysen außerdem eine aufwändige individuelle Beratung voraus. Bestehende Softwarelösungen zu dieser

³ <http://www.pius-netzwerk-deutschland.de/> Abgerufen am 12.05.2016.

Thematik, wie z.B. Umberto oder GaBi, sind für kleinere Unternehmen nicht besonders geeignet. Sie setzen Expertenwissen voraus, das in KMU oft nicht vorhanden ist. Die bisher vorliegenden Softwarelösungen zur Stoffstromerfassung und -modellierung (SSM) unterstützen weder eine dezentrale Datenerfassung noch eine effiziente Darstellung von erfassten Daten und Modellen vor Ort in den Produktionsbereichen. Damit fehlt es an unterstützenden Softwarelösungen für eine mobile Datenaufnahme dort, wo die Stoff- und Energieströme anfallen. Außerdem fehlt es an effektiven Lösungen zur Darstellung der Mengenströme bei Rundgängen in den Produktionsbereichen, um vor Ort die Entwicklung von Prozessoptimierungen zu unterstützen.

In diesem Forschungsvorhaben soll der Frage nachgegangen werden, inwieweit die Interaktionsmöglichkeiten mobiler Apps bei der Datenerfassung hilfreich sein können oder an welchen Stellen eventuell Grenzen sichtbar werden. Ein weiterer Gegenstand der Untersuchung ist die Evaluierung der technologischen Möglichkeiten zur Darstellung von Sankey-Diagrammen in mobilen Anwendungen. Ziel des hier skizzierten Vorhabens ist die Auflösung der bestehenden Einstiegshürden und der fehlenden mobilen Datenerfassung und Visualisierung in Zusammenhang mit Ressourceneffizienzfragestellungen. Deshalb soll ein weiterentwickeltes Konzept zur Datenerfassung und -darstellung von Stoffströmen ausgearbeitet und implementiert werden. Dadurch sollen KMU zukünftig in die Lage versetzt werden, ihre Produktionsprozesse strukturiert zu erfassen und somit betriebliches Stoffstrommanagement umzusetzen.

2 Vision von MOPS

Im Projekt MOPS soll hierzu die mobile Datenerfassung nach der MSM- und der eVSM-Methodik, bzw. einer Mischform (Material Value Stream Mapping, MVSM), umgesetzt werden. Das bedeutet, dass eine mobile Anwendung entwickelt wird, die zu einer Anzahl von Produktionsprozessen Eingabeformulare für die Input- und Output-Daten generiert, welche mit den vorangegangenen und nachfolgenden Produktionsprozessen gekoppelt sind [Wo10]. Durch die Erhebung der Input- und Outputmengen der Produktionsprozesse werden alle Eingabeformulare schrittweise gefüllt und es entsteht die Datengrundlage für das betriebliche Stoffstrommanagement. Diese Erhebung muss nicht in einem einzelnen Durchlauf erfolgen, sondern kann unterbrochen und später fortgesetzt werden, wenn einzelne Daten noch ermittelt werden müssen. Die Eingaben können während der Aufnahme einer Plausibilitätsprüfung unterzogen werden. Eingegebene Daten, die nicht dem erwarteten Eingabemuster entsprechen, werden somit vermieden, was generell die Datenqualität erhöhen soll. Die mobile Erfassung stellt somit insbesondere für KMU eine Vereinfachung dar, für die die Datenerfassung bisher häufig zu komplex ist.

Der Einsatz mobiler Technologien soll ein interaktives, praxisorientiertes Herantasten an die komplexe Datenaufnahme ermöglichen und somit die bisherigen Einstiegshürden senken. Die Anwendung soll so umgesetzt werden, dass die Nutzer angespornt werden, die Stoffstromdaten zu sammeln. Durch das Nutzen der mobilen Anwendung und die schnell sichtbaren Erfolge und Erkenntnisse soll das Interesse am Stoffstrommanagement gesteigert werden. Zu erfassende Daten für die Darstellung der Energie- und Ressourcenflüsse sind z.B. Betriebszustände von Maschinen, Flussmengen, Stoffkonzentrationen, Stoffeigenschaften und Bestände.

Für den Kooperationspartner mit dem Projektpartner ifu Hamburg GmbH soll deren bereits langjährig existierende Desktopanwendung e!Sankey⁴ um ein mobiles Pendant incl. Datenaustauschformat ergänzt werden. Die Erfassung der Prozessdaten kann je nach Verfügbarkeit entweder mobil oder zentral über die Desktopanwendung erfolgen. Somit kommen die Synergien aus der bestehenden Desktopanwendung und den Vorteilen des Mobile-Computings zum Tragen.

Die Projektergebnisse richten sich neben produzierenden KMU auch an beratende Ingenieure, Beratungsstellen und Netzwerke zum Thema Ressourceneffizienz. Durch die Nutzung der Software kann die erste Analyse eines produzierenden Betriebes unterstützt und vereinfacht werden, was eine signifikante Senkung der Kosten und kommunikationsstarke Ergebnisse ermöglicht. Dabei soll vor allem die konkrete softwaretechnische Umsetzung von der initialen Prozesserfassung bis zur Auswertung stoffstromrelevanter Daten unter Nutzung modernster Technologien überzeugen und damit den Einsatz fördern. Dadurch entsteht eine Optimierung und Harmonisierung der ökologischen und ökonomischen Säule der Nachhaltigkeit in der betrieblichen Praxis von KMU. Das Konzept der mobilen Erfassung der Prozessdaten und der Unterstützung von Rundgängen durch mobile Visualisierung dieser Daten kann somit die Einstiegshürden zur Umsetzung des Stoffstrommanagements senken, seine Akzeptanz erhöhen und somit zu mehr Effizienz im Mittelstand beitragen. Die mobile Software soll kostengünstig oder frei verfügbar sein, um die Nutzungshürden zu minimieren.

Durch die Umsetzung einer selbsterklärenden Software, die in praktisch allen produzierenden Unternehmen eingesetzt werden kann, wird erwartet, dass so eine große Breitenwirkung erreicht werden kann.

3 Umsetzung

Um einen Austausch von e!Sankey-Diagrammen zwischen der stationären und der mobilen Prozesserfassung zu realisieren, muss das bestehende .sankey-Format zu einem plattformunabhängigen Datenformat (sog. Rich Sankey Format) erweitert werden, mit dem zusätzliche Daten aus beliebigen Anwendungskontexten integriert und ggf.

⁴ e!Sankey ist das führende Produkt auf dem Markt zur Visualisierung von Material, Energie und Kosten mit Hilfe von Sankey-Diagrammen. <http://www.e-sankey.com/de/> Abgerufen am 12.05.2016.

durchgereicht werden können. Die in diesem Projekt entwickelte Anwendung reicht damit die strukturellen Daten, die als Sankey-Diagramm dargestellt werden, um die im Rahmen der Prozessdatenerfassung gesammelten Daten an. Das in einer Datei vorliegende Rich Sankey Dateiformat soll dann zunächst über ein Datenaustauschverzeichnis kommuniziert werden. Dies kann ein Cloud-Verzeichnis im Internet oder ein über den WebDAV-Standard eingebundenes Firmenverzeichnis sein (siehe Abbildung 1).

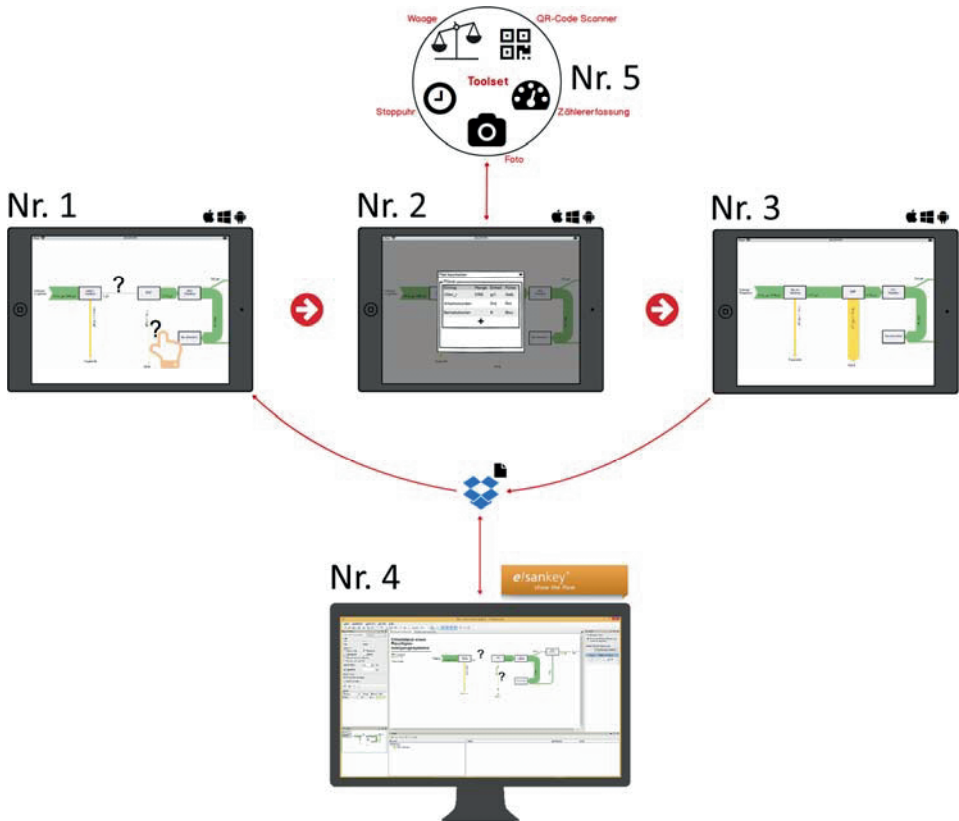


Abbildung 1: Ablaufschema der mobilen Anwendung in Interaktion mit der Desktopanwendung e!Sankey über einen Online-Datenspeicher

Folgende Szenarien zur Datenerfassung sind denkbar, wobei das zweite Beispiel gesondert in Abbildung 1 illustriert ist.

1. Die mobile Anwendung wird genutzt, um eine Stoffstromanalyse zu erstellen. Die Darstellung soll mittels Sankey-Diagrammen erfolgen. Hierbei werden generische

Prozessmodelle zum Einsatz kommen (sog. Templates). Dadurch können die Nutzer der mobilen Komponente sich hauptsächlich auf die Datenerfassung konzentrieren und müssen weniger Aufwand für die Gestaltung des Prozessmodells aufbringen. Dazu wählt der Anwender vor der Datenerfassung ein am Desktop-PC modelliertes Prozessmodell aus und beginnt mit der Sammlung der Daten. Das mobil erstellte Diagramm kann anschließend mit der bestehenden e!Sankey-Anwendung der ifu Hamburg GmbH weiter bearbeitet werden.

2. Eine bestehende Stoffstromerfassung in Form eines Sankey-Diagramms liegt bereits vor, jedoch sind die Daten noch unvollständig oder ungenau. Dieser Umstand soll durch die Fragezeichen angedeutet werden. Die mobile Anwendung wird verwendet, um die nicht vorliegenden Daten vor Ort zu sammeln bzw. zu präzisieren, in dem der entsprechende Materialfluss ausgewählt wird (Nr. 1). Im nächsten Schritt erfolgt dann die tatsächliche Dateneingabe (Nr. 2). Schließlich kann der Nutzer den nächsten Materialfluss ergänzen (Nr. 3). Die Nachbearbeitung erfolgt, wie in Szenario 1 auf der Desktop-Anwendung e!Sankey (Nr. 4). Die Umsetzung einer Rundgangliste mit Notizen zu den geplanten Ablesungen kann den Mitarbeiter in seiner mobilen Tätigkeit zusätzlich unterstützen.
3. Ein bestehendes Modell wird mit Hilfe der mobilen Anwendung auf Korrektheit überprüft oder die darin enthaltenen Daten werden aktualisiert. Im Unterschied zu den vorherigen Szenarien erfolgt nicht zwangsläufig eine neue Erfassung der Daten, sondern eine Validierung mittels einer Plausibilitätsprüfung oder durch Hinzunahme der Tools des mobilen Geräts (s.u.).

Die mobile Anwendung soll von den speziellen technischen Möglichkeiten mobiler Endgeräte Gebrauch machen. Durch GPS bzw. erweiterte Ortungsfunktionen können zum einen Informationen standortbezogen und kontextbezogen (Stichwort: Location-based Services) in der Produktion verortet werden, was später zur vereinfachten Lokalisierung der Datenquelle beitragen könnte.

Die mobile Erfassung soll durch komfortable Funktionen zum Messen der Stoffströme ergänzt werden. Z.B. könnte eine Stoppuhr die Messung der Zeitintervalle in der Fertigung ermöglichen. Ein integrierter Klick-Zähler soll zur Messung von Mengen zum Einsatz kommen. Angedacht ist außerdem die Einbindung von Waagen über die Bluetooth-Schnittstelle der mobilen Endgeräte, um Gewichte in den Produktionsprozessen vor Ort bestimmen zu können. Weitere Möglichkeiten zur Erkennung und Ablesung von Anlagen bei Rundgängen bieten RFID-Chips, QR- sowie Barcodes. Eine weitere Funktion wird die Einbindung der mobilen Kamera zur Anreicherung von Sankey-Modellen mit Fotos von Anlagen und Typenschildern darstellen (Nr. 5).

Somit entsteht ein plattformübergreifendes Erfassungs- und Darstellungssystem von Stoffstromdaten mittels Sankey-Diagrammen, das bisher nicht vorhanden ist. Dieser Ansatz macht die Vorteile der Funktionen der klassischen Desktopanwendung e!Sankey

und von mobilen Diensten für einen vereinfachten Zugang zu Ressourcen-effizienzpotentialen nutzbar. Der Einsatz von Tablets geht dabei zwangsweise mit einer Komplexitätsreduktion einher, da man sich auf wesentliche Daten bei der Erfassung von Stoff- und Energieströmen konzentrieren muss. Auf diese Weise können z.B. bereits vor Ort erste Prozesse identifiziert werden, die Optimierungsmöglichkeiten bieten.

Um ein möglichst breites Spektrum an Anwendern der mobilen Anwendung zu erreichen und dabei gleichzeitig den Arbeitsaufwand bei Entwicklung möglichst gering zu halten, wird eine Lösung, welche die Plattformen iOS, Android und Windows gleichermaßen bedient, angestrebt. Daher soll ein Softwareframework zum Einsatz kommen, das es erlaubt anhand einer Codebasis Apps für die verschiedenen genannten Plattformen umzusetzen. Dabei sind wesentlich entscheidend für die Wahl des Softwareframeworks die Nutzungsmöglichkeiten passender Grafikbibliotheken zur Darstellung von Sankey-Diagrammen. Nach Recherchen und prototypischer Umsetzungen der Autoren hat sich aus den drei Kombinationen von Xamarin⁵ und GoXam⁶ bzw. Ionic⁷ und D3.js⁸ die aus Ionic und GoJS⁹ als am geeignetste Lösung herausgestellt.

Ionic

Das OpenSource Framework Ionic baut auf dem Framework Apache Cordova¹⁰ auf und ermöglicht die Entwicklung mobiler plattformunabhängiger (hybrider) Applikationen. Dabei werden mit JavaScript, HTML5 und CSS entwickelte „Webanwendungen“ lokal in einem gekapselten Webbrowser ausgeführt und dargestellt. Der Programmcode kann dabei über alle Plattformen hinweg vollständig wiederverwendet werden. Die üblichen Defizite beim Look & Feel nicht nativer Apps entgegnet Ionic mit einer Reihe von UI Komponenten, welche speziell für mobile Browser optimiert wurden und sich durch minimale DOM Manipulationen auszeichnen, sodass eine ähnlich hohe Performance wie bei einer nativen App erzielt werden kann. Darüber hinaus integriert Ionic das im Web weit verbreitete Single-Page-Application-Framework AngularJS¹¹, welches ein erprobtes und gut testbares Architekturgerüst der Anwendung vorgibt.

GoJS

GoJS ist eine kommerzielle JavaScript Bibliothek zur Erstellung individueller interaktiver Diagramme und komplexer Visualisierungen auf allen modernen Webbrowser und Plattformen. Mit Hilfe von Knoten, Links und Gruppen sowie anpassbarer Vorlagen und Layouts lassen sich Diagramme komplett im Browser in ein HTML5 Canvas Element oder SVG rendern. Während eine auf Xamarin und GoXam basierende Lösung aufgrund von Inkompatibilität nicht möglich ist, liegt der große

⁵ <https://www.xamarin.com/> Abgerufen am 12.05.2016.

⁶ <http://www.nwoods.com/products/goxam/> Abgerufen am 12.05.2016.

⁷ <http://ionicframework.com/> Abgerufen am 12.05.2016.

⁸ <https://d3js.org/> Abgerufen am 12.05.2016.

⁹ <http://gojs.net/> Abgerufen am 12.05.2016.

¹⁰ <https://cordova.apache.org/> Abgerufen am 12.05.2016.

¹¹ <https://angularjs.org/> Abgerufen am 12.05.2016.

Vorteil von GoJS im Vergleich zu D3.js in der starken Anlehnung der Programmierschnittstelle an der in der e!Sankey-Anwendung verwendeten Bibliothek GoDiagram¹². Dadurch ist eine unkomplizierte Umwandlung und Weiterverarbeitung der erfassten Daten in der mobilen Anwendung auf dem Desktop-PC sichergestellt. Außerdem wird der Entwicklungsaufwand deutlich reduziert, da die Anwendungslogik relativ einfach aus der e!Sankey-Anwendung in die mobile Anwendung portiert werden kann.

4 Ausblick

Das vorgestellte Projekt hat das Ziel zu Ressourceneffizienzsteigerungen in KMU beizutragen. Dieses soll mit Hilfe einer mobilen Anwendung zur Datenerfassung und Visualisierung mit Sankey-Diagrammen an den Produktionsstätten erreicht werden. Zusätzliche Tools zur genaueren und validen Dateneingabe (Zähler, Waage, etc.) sollen ebenso zur Verfügung gestellt werden. Für weitergehende und detaillierte Analysen wird ein plattformunabhängiges Datenformat entwickelt (Rich Sankey Format).

Derzeit befindet sich die mobile Anwendung in einem noch frühen Entwicklungsstadium. Zwar können bereits die Darstellung der Sankey-Diagramme und die Zusammenführung der einzelnen Elemente rudimentär auf einem Tablet erfolgen sowie Material- und Energieflussdaten aus einer auf dem Rich Sankey Format basierende Datei ausgelesen sowie zurückgeschrieben werden, allerdings ist hier noch mit erheblichem Entwicklungsaufwand zu rechnen, da der komplexe Funktionsumfang zur Darstellung von Sankey-Diagrammen in e!Sankey nicht ohne, zum Teil erhebliche Anpassungen der Anwendungslogik an die zugrunde liegende technologische Basis erfolgen kann.

Des Weiteren umfassen die zukünftigen Arbeitspakete die Übertragung der MSM/eVSM-Methode auf die Anwendung und die Umsetzung der zusätzlichen Tools für die Begehung vor Ort. Die Einbindung eines zentralen Servers, der als Datenquelle wie auch -senke für die mobile Anwendung fungiert, ist ebenso geplant wie Usability-Tests im UxLab der HTW Berlin, welche der Optimierung der Gebrauchstauglichkeit der mobilen Anwendung dienen sollen.

Literaturverzeichnis

- [BB00] Bullinger, H., Beucker, S.: Stoffstrommanagement und Betriebliche Umweltinformationssysteme (BUIS) liefern neue Impulse für das Umweltcontrolling. In H. Bullinger, S. Beucker, Stoffstrommanagement Erfolgsfaktor für den betrieblichen Umweltschutz. Fraunhofer IRB-Verlag, Stuttgart, S. 1-18, 2000.

¹² <http://www.nwoods.com/products/godiagram/> Abgerufen am 07.06.2016.

-
- [BU16] Bundesregierung: Programm zur nachhaltigen Nutzung und zum Schutz der natürlichen Ressourcen, <http://dip21.bundestag.de/dip21/btd/18/077/1807777.pdf>, Stand: 06.06.2016.
- [BW09] Brüggemann, A., Wied, T.: Material- und Rohstoffeffizienz in Unternehmen. In Perspektive Zukunftsfähigkeit – Steigerung der Rohstoff- und Materialeffizienz (S. 33ff). KfW Bankengruppe, Frankfurt am Main, S. 41f, 2009.
- [Cr11] Crönertz, O.: Stoffstromorientiertes Kostenmanagement - Evaluierung, Erprobung, und Weiterentwicklung der material- und energieflussorientierten Kostenrechnung. Verlag Dr. Kovac, Hamburg, S. 66ff, 2011.
- [EP09] EPA, U. S.: EPA. Abgerufen am 12.05.2016 von The Environmental Professionals Guide to Lean Six Sigma: <http://www.epa.gov>, 2009.
- [HB02] Heck, P.; Bemann, U.: Praxishandbuch Stoffstrommanagement: Strategien – Umsetzung – Anwendung in Unternehmen, Kommunen, Behörden. Deutscher Wirtschaftsdienst, Köln, 2002.
- [Re10] Reif, W.: Ressourceneffizienz und die mögliche Rolle einer Industrie- und Handelskammer anhand des MESOR-Netzwerkes. UmweltWirtschaftsForum 18, S. 229-236, 2010.
- [VD11] VDI Zentrum Ressourceneffizienz GmbH: Umsetzung von Ressourceneffizienz-Maßnahmen in KMU und ihre Treiber. VDI Zentrum Ressourceneffizienz GmbH, Berlin, 2011.
- [Wo10] Wohlgemuth, V., Krehahn, P., Ziep, T.: Mobile Anwendungen als Datenquelle für das Stoffstrommanagement. In W. Abramowicz, R. Alt, K.-P. Fähnrich, B. Franczyk, & L. Maciaszek, Informatik 2010 - Proceedings. Gesellschaft für Informatik, Bonn, S. 306 – 313, 2010.

Der Eingabestift – Mode oder ein nachhaltiges Bedienkonzept?

Hans-Knud Arndt¹

Abstract: Der Beitrag befasst sich mit einer Diskussion von Mode und Nachhaltigkeit in der Informations- und Kommunikationstechnik (IKT) am Beispiel des Eingabestifts (englisch Stylus oder Touch Pen). Die wechselhafte Entwicklungsgeschichte des Eingabestifts wirft die Frage auf, ob es sich beim Eingabestift um ein nachhaltiges Bedienkonzept der IKT handelt oder ggf. nur um eine (wiederkehrende) Modeerscheinung. In diesem Zusammenhang sind die üblichen Eingabestifte aus Kunststoff von denen mit einer druck- und neigungssensitiven elektronischen Steuerung wie der Apple Pencil zu unterscheiden. Während beim Einsatz der üblichen Eingabestifte aus Kunststoff tendenziell eher von einer (wiederkehrenden) Modeerscheinung gesprochen werden kann, kann für die Eingabestifte mit einer druck- und neigungssensitiven elektronischen Steuerung im Rahmen eines bestimmten Einsatzspektrums eine Nachhaltigkeit prognostiziert werden.

Keywords: Eingabestift, Stylus, Mode, Nachhaltigkeit

1 Konjunktur und Mode

Der Begriff „Konjunktur“ wird definiert als „Mehr oder weniger regelmäßige Schwankungen aller wichtigen gesamtwirtschaftlichen Größen wie z.B. Produktion, Beschäftigung und Preise. Hieraus können zyklische Schwankungen der gesamtwirtschaftlichen Aktivität, gemessen z.B. durch den Grad der Kapazitätsauslastung (Auslastungsgrad), hergeleitet werden. (...) Im allg. Sprachgebrauch wird der Begriff Konjunktur etwas unpräzise als Aufschwung bzw. Aufschwungphase verwendet“ [Gab16].

Als „Mode“ wird „allg. der sich wandelnde Geschmack in den verschiedenen Lebensbereichen“ [MLr93] verstanden. Die Betrachtung der Mode wurde bis Ende der 1960er Jahre lange von der Betrachtung in den Kulturwissenschaften² ausgeschlossen, denn die Mode, „die durch kurzfristige Wandlungsprozesse charakterisiert ist, fügt sich nicht den Prinzipien der Dauer und der Beständigkeit, durch welche man die Volkskultur definiert sah“ [Bau68]. Dagegen ist es heutzutage unstrittig, modische Prozesse auch in den Kulturwissenschaften zu berücksichtigen.

¹ Otto-von-Guericke Universität Magdeburg, Fakultät für Informatik, Arbeitsgruppe Wirtschaftsinformatik – Managementinformationssystem, hans-knud.arndt@iti.cs.uni-magdeburg.de

² Für den Begriff „Kulturwissenschaften“ wurden früher vor allem die Begriffe „Volkskunde“ bzw. „Volkskultur“ verwendet.

Üblicherweise gilt auch für Konzepte in der Informations- und Kommunikationstechnik (IKT), dass sie „Konjunktur haben“ bzw. „in Mode sind“. Dieses Phänomen soll hier am Beispiel des Bedienkonzepts „Eingabestift“ näher diskutiert werden.

2 Der Eingabestift als Bedienkonzept

Als Eingabestift (englisch auch Stylus oder Touch Pen) wird in der IKT ein „dünnes Stäbchen, mit dem auf dem Display von Computern (besonders Smartphones oder Tablets) Texte, Zeichnungen o.Ä. eingegeben werden können“ [Dud16], bezeichnet.

Ein wichtiger Innovationsschritt im Bereich der IKT stellen die sogenannten Quarzuhren dar. Eine Quarzuhr wird definiert als „ein elektro-mechanischer Zeitmesser, dessen Taktgeber nicht wie sonst üblich ein mechanisches Pendel, sondern ein sogenannter ‚Quarzoszillator‘ ist. Dessen Frequenz wird mit Hilfe von Schwingquarz - einem Bauteil zur Erzeugung elektrischer Schwingungen - gehalten. Uhren, die auf diese Weise angetrieben werden, können im Wesentlichen in zwei Kategorien unterteilt werden: Quarzuhren mit digitaler Anzeige (Leuchtdioden oder Flüssigkristalle) sowie Quarzuhren mit analogen Zeigern, die über einen Motor angetrieben werden.“ [WLG12].

Die ersten Quarzuhren wurden in den 1930er Jahren in Deutschland entwickelt, waren in dieser Zeit aber nur ein Nischenprodukt. Für den Massenmarkt wurden die Quarzuhren, vor allem die Quarzarmbanduhren, erst 1969 durch den japanischen Uhrenhersteller Seiko erschlossen. Bis Mitte der 1980er Jahre dominierten Quarzuhren den Uhrenmarkt und führten in der traditionellen Schweizer Uhrenindustrie zu dem Phänomen, welches als „Quarzkrise“ bezeichnet wurde. Der Absatz hochwertiger, mechanischer Uhren brach dramatisch ein [WLG12]. Dieser Trend wurde erst durch den „Swatch-Gründer Nicolas Hayek, der die Kundschaft durch seine Swatch - ebenfalls quarzbetrieben - wieder an Schweizer Produkte heranführte“ [WLG12], gestoppt.

Neben der Tatsache, dass durch die Quarzuhren mit digitaler Anzeige erstmals ein Großteil der allgemeinen Bevölkerung mit einer digitalen Anzeige von Ziffern und Buchstaben vertraut gemacht wurde, boten diese Quarzarmbanduhren den Entwicklern neue technische Möglichkeiten. Ein bevorzugtes Merkmal war in den 1970er Jahren die Integration eines Taschenrechners in die Quarzarmbanduhr. Ein solcher Taschenrechner sollte über miniaturisierte Tasten bedient werden und das Ergebnis einer durchgeführten Berechnung wurde auf der digitalen Anzeige der Quarzarmbanduhr angezeigt. Um die Bedienung der miniaturisierten Tasten zu erleichtern, wurden den Quarzarmbanduhren häufig Eingabestifte beigefügt (siehe Abbildung 1).



Abbildung 1: Die Hewlett Packard HP 01-Rechneruhr aus dem Jahr 1977:
„Der unten liegende Stift zur Bedienung der Tasten des Rechners ist im Armband versteckt.“
(Quelle: [UPU03])

Anfang der 1990er Jahre wurden Eingabestifte in der IKT vor allem im Bereich der sogenannten persönlichen digitalen Assistenten (PDA, Personal Digital Assistant) eingesetzt. Das US-amerikanische Unternehmen Apple Inc. stellte 1993 auf der MAC World in Boston das Newton MessagePad vor. Die Bedienung des Apple Newton erfolgt auf einem monochromen berührungsempfindlichen Bildschirm mit einer Auflösung von 336×240 Pixeln mit Hilfe eines Eingabestifts (siehe Abbildung 2). Eine Besonderheit stellt die Handschrifterkennung mit Hilfe der Rosetta-Engine dar, die in der Praxis allerdings nur unzureichend funktionierte. Nachdem der Apple-Gründer Steve Jobs im September 1997 wieder als vorübergehend Geschäftsführer des Unternehmens eingesetzt wurde, wurde im Februar 1998 das Projekt Apple Newton von ihm beendet. Steve Jobs stand einerseits dem Apple Newton als Projekt seines Vorgängers John Sculley grundsätzlich kritisch gegenüber, andererseits kritisierte er konkret das Bedienkonzept des Eingabestifts [Bor03].



Abbildung 2: Apple Newton MessagePad aus dem Jahr 1993 (Quelle: [TGM14])

Als weiteres Unternehmen, welches sich ganz wesentlich an der Entwicklung stiftbasierter PDAs beteiligte, ist das US-amerikanische Unternehmen Palm Inc. zu nennen. Palm Inc. wurde 1996 vom Unternehmen U.S. Robotics übernommen und brachte im gleichen Jahr den Palm Pilot auf den Markt. Die Bedienung des Palm Pilot erfolgt auf einem monochromen berührungsempfindlichen Bildschirm mit einer Auflösung von 160×160 Pixeln durch einen Eingabestift (siehe Abbildung 3). Auch der Palm Pilot verfügt über eine Handschrifterkennung (Graffiti-Engine), die in der Praxis deutlich besser als beim Apple Newton funktionierte (siehe Abbildung 4). Zum Ende des Jahres 1996 war der Palm Pilot das vorherrschende Produkt auf dem PDA-Markt mit einem Marktanteil von 70 Prozent [Lüd06].



Abbildung 3: US Robotics Palm Pilot aus dem Jahr 1996
(Quelle: [Car15])

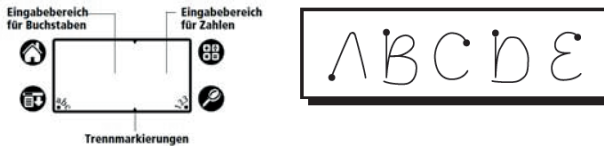


Abbildung 4: Dateneingabe mit der Graffiti-Schrift
(Quelle: [Pal00])

Eine Kombination aus Mobiltelefon und PDA (heutzutage als Smartphone bezeichnet) stellt das Joint Venture aus dem japanischem Unternehmen Sony und dem schwedischen Unternehmen Ericsson, Sony Ericsson Mobile Communications AB, mit dem Sony Ericsson P800 im Jahr 2002 vor. Die Bedienung des Sony Ericsson P800 erfolgt entweder über eine ausklappbare Tastatur ergänzt durch das sogenannte Jogdial (ein kleines Drehrädchen, welches gedrückt sowie nach vorn und hinten geschoben werden kann) an der linken Seite oder über einen Eingabestift auf einem

berührungsempfindlichen Farbbildschirm in der Größe von $6,5 \times 4$ Zentimeter (bei ausgeklappter Tastatur) mit 4096 Farben und einer Auflösung von 208×320 Pixeln (siehe Abbildung 5). Darüber hinaus verfügt das Sony Ericsson P800 Mobiltelefon u.a. über eine Digitalkamera auf der Rückseite mit einer Auflösung von bis zu 640×480 Pixeln sowie über die Möglichkeit, Musikdateien im Audio-Format MP3 abzuspielen [Spe03]. Das Sony Ericsson P800 ermöglicht zudem eine Handschrifterkennung bei Eingabe durch den Eingabestift (Graffiti-Engine). Dabei wird – ähnlich wie beim Palm Pilot, hier nur horizontal getrennt – der Bildschirmbereich in zwei Hälften geteilt, die obere Hälfte dient der Erkennung von handschriftlich geschriebene Zahlen, die untere Hälfte der Erkennung von Buchstaben [ADA03].



Abbildung 5: Sony Ericsson Mobiltelefon aus dem Jahr 2002
(Quelle: [WDD09])

Aber auch im Bereich der Personal Computer (PC) hat das US-amerikanische Unternehmen Microsoft an einer stiftbasierten Bedienung ihres Betriebssystems Windows gearbeitet. Erstmals erschien im November 2002 eine stiftbasierte Version des Betriebssystems Microsoft Windows XP unter der Bezeichnung „Windows XP Tablet PC Edition“. Parallel zur Tablet-Version des Betriebssystems Windows XP wurde von Microsoft eine Microsoft Tablet-PC-Spezifikation veröffentlicht, die zunächst neben der Bedienung durch Tastatur und Maus nur eine stiftbasierte Eingabe ermöglichte. Diese erste Version einer stiftbasierten Bedienung des Betriebssystems Windows erfüllte die hohen Erwartungen nicht. Bis Ende 2004 wurden schätzungsweise weltweit ca. nur eine Million Tablet PCs mit der Windows XP Tablet PC Edition verkauft, dies vor dem Hintergrund eines Gesamtvolumens von rund 200-Millionen verkaufter Windows-basierter PCs allein im Jahr 2004 [FrK05]. Mit der Auslieferung des Microsoft Betriebssystems Windows Vista ab Januar 2007 wurde die Möglichkeit einer stiftbasierten Eingabe fester Bestandteil von fast allen Produktvarianten des Betriebssystems. Eine Handschrifterkennung wird über den sogenannten „Tablet PC-Eingabebereich“ zur Verfügung gestellt. Der Tablet PC-Eingabebereich öffnet sich, wenn der Benutzer den Eingabestift entweder

- (a) auf dem „Symbol **Eingabebereich**, das angezeigt wird, wenn (...) [der Eingabestift] in einem Texteingabebereich wie der Adressleiste in einem Webbrowser oder einem Textfeld in einem Dialogfeld“ oder
- (b) auf der „Registerkarte **Eingabebereich**, die standardmäßig am linken Bildschirmrand angezeigt wird“ oder
- (c) auf dem „Symbol **Tablet PC-Eingabebereich** auf der Taskleiste“ [Mic16]

platziert wird. In den Standardvorgaben von Windows Vista ist es vorgesehen, dass die automatische Handschrifterkennung in dem Tablet-PC-Eingabebereich erst dann startet, wenn der Eingabestift vom Bildschirm wegbewegt oder in einen anderen Bereich verschoben wird (siehe auch Abbildung 6) [Mic16].

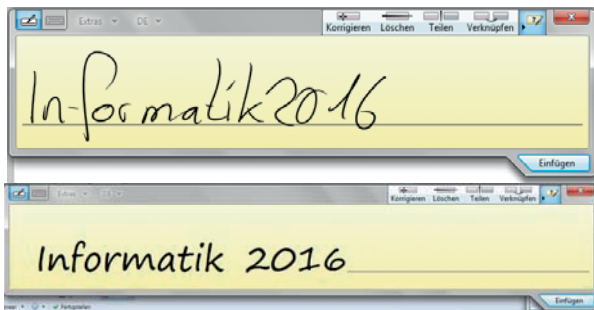


Abbildung 6: Microsoft Windows Vista: Automatische Handschrifterkennung im Tablet-PC-Eingabebereich aus dem Jahr 2007 (Quelle: Eigene)

Dieser (unvollständige) Einblick in die Entwicklungsgeschichte des Eingabestifts wirft die Frage auf, ob es sich beim Eingabestift um ein nachhaltiges Bedienkonzept der IKT handelt oder ggf. nur um eine (wiederkehrende) Modeerscheinung.

3 Mode gegenüber Nachhaltigkeit in der Informations- und Kommunikationstechnik am Beispiel des Eingabestifts

Als Steve Jobs als Geschäftsführer der Apple Inc. im Januar 2007 erstmals das Smartphone-Produkt „iPhone“ auf der Macworld Conference & Expo in San Francisco vorstellte, sprach er den inzwischen legendären Satz aus, dass niemand einen Eingabestift haben möchte („Nobody wants a Stylus“) [Gra15b]. Weiter sagte Steve Jobs damals „, wenn du einen Stylus siehst, haben [die Entwickler] es vermasselt“ [Gra15a]. Seit der Veröffentlichung des ersten iPhones ist der durch Fingerbewegungen zu bedienende berührungsempfindliche Bildschirm das dominierende Bedienkonzept für Smartphones und Tablet Computer geworden. Und es schien schwer vorstellbar, dass der Eingabestift (Stylus) jemals wieder eine Renaissance erfahren würde.

Dennoch wirbt im Jahr 2012 gerade der direkte Konkurrent vom Unternehmen Apple Samsung bei der Vorstellung des Tablet Computer „Galaxy Note Tablet 10.1“ mit dem Eingabestift als das Konzept zur „,innovative Bedienung‘ (..) Mit dem Stift, so die Werbebotschaft, können wir uns endlich vom Papier befreien. Im 21. Jahrhundert sollten die Menschen aufs Display statt auf einen Block kritzeln“ [Erm12]. Damit steht weiterhin die Frage im Raum, ob es sich beim Eingabestift um ein nachhaltiges Bedienkonzept der IKT handelt oder ggf. nur um eine (wiederkehrende) Modeerscheinung.

Und so wird auch der Einsatz des Eingabestifts in der Literatur in die Nähe einer Modeerscheinung gerückt:

„In der Mode, sagt man, ist alles schon einmal da gewesen. Streifenmuster? Kommen alle Jahre wieder. Schlaghosen? Waren nicht nur in der 70ern, sondern auch in den 90ern en vogue. Bomberjacken? Vor ein paar Jahren ausgemustert, nun wieder salonfähig.

Wie erfrischend ist da die Computerbranche. Hier scheinen ständige Querdenker am Werk, die immer neue Geräte mit immer neuen Funktionen entwickeln. Der neueste Hype: Ein Flachrechner mit Stift. (...)

Ganz so innovativ ist das Bedienkonzept allerdings nicht. Schon in den 90er Jahren des vorherigen Jahrhunderts konnte man Minicomputer mit digitalem Griffel kaufen“ [Erm12].

Aber auch das Unternehmen Apple zeigte im Jahr 2015 bei der Vorstellung des Tablet Computers iPad Pro einen Eingabestift mit der Bezeichnung „Apple Pencil“. Haben also die Entwickler von Apple gegen den vom inzwischen verstorbenen Gründer und Geschäftsführer formulierten Grundsatz des Nichteinsatzes eines Eingabestifts verstoßen? Folgt das Unternehmen Apple vielleicht sogar einem Modetrend? Oder irrte Steve Jobs bei seiner Ablehnung eines Eingabestifts als Bedienkonzept und der Eingabestift ist nun doch ein nachhaltiges Bedienkonzept in der IKT?

Bei den üblichen Eingabestiften handelt es sich um Griffel aus Kunststoff, die über keine elektronischen Bauteile verfügen und auf den gängigen berührungsempfindlichen Bildschirmen zum Einsatz kommen können. Diese Art von Eingabestiften sind im „Grunde ein Ersatz für den eigenen Finger“ [Sch16] und kann für filigrane Arbeiten wie das Malen oder Zeichnen besser geeignet sein als der menschliche Finger. Auch für längere Texteingaben wird diese Art von Eingabestift empfohlen (ggf. in Verbindung mit einem Handschrifterkennungssystem) [Sch16], wobei sich hier wiederum die Frage stellt, ob für die Eingabe längerer Texte nicht eine externe Tastatur eher zu empfehlen wäre.

Dagegen stellt der Apple Pencil (siehe Abbildung 7) einen Eingabestift mit einer druck- und neigungssensitiven elektronischen Steuerung dar, die nur auf speziellen Bildschirmen wie beim Apple iPad Pro zum Einsatz kommen können. Der Apple Pencil muss über den Apple Lightning-Anschluss (am Stiftende unter einer Kappe) wiederaufgeladen werden.



Abbildung 7: Apple Pencil für das iPad Pro aus dem Jahr 2015
(Quelle: [App16])

In einer Studie an der Princeton University in den Vereinigten Staaten von Amerika aus dem Jahr 2014 wurde untersucht, ob an Hochschulen bessere Lernergebnisse durch das Mitschreiben in Vorlesungen per Tastatur (an einem Laptop) oder per Stift (auf einem Notizblock) erreicht werden können. Das Ergebnis dieser Studie zeigt: „Ging es um Faktenwissen, so punkteten die Testteilnehmer mit Laptop sogar ähnlich wie jene mit Stift und Papier – doch bei den konzeptionellen oder Transfer-Fragen, die also ein inhaltliches Verstehen abtesten, schnitten die Laptop-Probanden deutlich schlechter ab“ [Saß14]. Eine wichtige Rolle spielt in diesem Zusammenhang, dass auch nicht eine besonders wortwörtliche Mitschrift zu besseren Lernergebnissen führt, sondern eine bereits verarbeitete Form des Lernstoffs mit Hilfe der eigenen Handschrift. Als Schlussfolgerung stellt die Studie deshalb fest, dass in einem solchen Lernumfeld nicht zwangsläufig zum Papier und Notizblock zurückgekehrt werden muss, sondern der Einsatz von Eingabestiften in Verbindung mit Tablet Computern, ggf. auch ergänzt durch eine Handschrifterkennung, zu prüfen sei [Saß14]. Allerdings wurde nicht untersucht, ob die erzielten besseren Lernergebnisse auch tatsächlich durch eine Kombination von Eingabestift und Tablet Computer erreicht werden können.

Um nun vor diesem Hintergrund die Frage beantworten zu können, ob der Eingabestift ein nachhaltiges Bedienkonzept in der IKT ist oder nicht, muss zunächst der Blick auf mögliche Grundsätze einer nachhaltigen Entwicklung innerhalb der IKT gelenkt werden. Dazu sollen die „Zehn Thesen für „Gutes Design“ von Dieter Rams herangezogen werden, da „Umweltschutz bzw. Nachhaltig *untrennbar* mit gutem Design verbunden ist“ [Arn13]. So heißt es in der siebten These „Gutes Design ist langlebig: Es vermeidet, modisch zu sein, und wirkt deshalb nie antiquiert. Im deutlichen Gegensatz zu kurzlebigen Mode-Design überdauert es auch in der heutigen Wegwerfgesellschaft lange Jahre“ [Vit16]. Demnach wären also Modeerscheinungen im Sinne eines nachhaltigen Gestaltens von IKT dringlichst zu vermeiden. Sind also Modeerscheinungen grundsätzlich kontraproduktiv? Sicherlich nicht, denn zurecht wird darauf verwiesen, dass auch uralte, also nachhaltige Bräuche und Technologien „irgendwann doch einmal angefangen und sich in ihrer ersten Phase meistens <modisch> ausgebreitet haben“ [Bau68].

Es gilt also bei der Mode zwischen dem wichtigen Impuls für Innovationen und einer kurzlebigen (Wegwerf-)Orientierung zu unterscheiden. Und vor diesem Hintergrund erscheint (aus heutiger Sicht) der klassische Eingabestift (Stylus) aus Kunststoff als grundsätzliches Bedienkonzept als kurzlebige, modische und damit nicht nachhaltige Erscheinung in der IKT, die in regelmäßigen Abständen eine Renaissance erlebt: „Der Stylus, den wir (...) kannten, wurde dank Apple beinahe vollständig vom Finger ersetzt“ [Gra15b].

Im Gegensatz dazu ist ein Eingabestift mit eigener elektronischer Steuerung wie der Apple Pencil, dessen Einsatz vom Unternehmen Apple selbst unter der Überschrift „Für Kreativität ohne Grenzen“ [App16] zusammengefasst wird, ein „hoch-präzises Instrument für ganz bestimmte Anwendungsgebiete“ [Gra15b]. Mögliche Einsatzgebiete sind [App16]:

- Kunst: Schnelle Skizzen und detaillierte Zeichnungen
- Notizen: Erweiterung durch handgefertigte Skizzen
- E-Mail: Erweiterung durch handgefertigte Skizzen
- Graphikgestaltung: Handgeführte Zeichnungen
- Fotobearbeitung: Bearbeitungsmöglichkeiten für Fotoprofis.

Darüber hinaus ist davon auszugehen, dass *nur* ein Eingabestift mit eigener elektronischer Steuerung wie der Apple Pencil in Verbindung mit einem Tablet Computer zu möglicherweise gleich guten Lernergebnissen im Rahmen von Vorlesungen an Hochschulen wie der Einsatz von Stift und Notizblock führen kann. Denn eine druck- und neigungssensitive elektronische Steuerung ist zwingend erforderlich, um haptisch möglichst nahe an das Arbeiten mit Stift und Notizblock herankommen zu können. Ein klassischer Eingabestift aus Kunststoff stellt in diesem Zusammenhang keine sinnvolle Alternative dar.

Zusammenfassend kann für einen Eingabestift mit einer druck- und neigungssensitiven elektronischen Steuerung im Rahmen eines bestimmten Einsatzspektrums eine Nachhaltigkeit im Sinne der These „Gutes Design ist langlebig“ prognostiziert werden und damit steht ein solcher klar umrissener Einsatz auch nicht im Widerspruch zur These von Steve Jobs in der Ablehnung eines Eingabestifts (aus Kunststoff) als Bedienkonzept.

4 Ausblick

Bei der Frage, ob der Eingabestift Mode oder ein nachhaltiges Bedienkonzept in der IKT darstellt, muss zunächst festgehalten werden, dass auch nachhaltige Technologien und Bedienkonzepte „irgendwann doch einmal anfangen und sich in ihrer ersten Phase meistens <modisch> ausgebreitet haben“ [Bau68]. Daraus folgt, dass erst nach Ablauf einer hinreichenden Einsatzzeit in der IKT die Frage nach (kurzfristiger und wiederkehrender) Mode oder Nachhaltigkeit von Technologien und Bedienkonzepten beantwortet werden kann. Im Jahr 2016 ist bezogen auf den Eingabestift von einer solchen hinreichenden Einsatzzeit auszugehen.

Wichtig im Zusammenhang der Beurteilung des Eingabestift-Bedienkonzepts ist zudem, dass genau beachtet wird, welche Art von Eingabestift beurteilt werden soll. Derzeit

unterschieden werden können die üblichen Eingabestifte aus Kunststoff und die mit einer druck- und neigungssensitiven elektronischen Steuerung wie der Apple Pencil. Während beim Einsatz der üblichen Eingabestifte aus Kunststoff tendenziell eher von einer (wiederkehrenden) Modeerscheinung auszugehen ist, kann für die Eingabestifte mit einer druck- und neigungssensitiven elektronischen Steuerung insofern von einem nachhaltigen Bedienkonzept gesprochen werden, indem solche Eingabestifte nicht als allgemeines Interaktionskonzept für berührungsempfindliche Bildschirme angesehen werden, sondern vielmehr für ganz bestimmte kreative Anwendungsfälle *ergänzend* eingesetzt werden.

Einen äußerst interessanten Fall stellt die Einsatzmöglichkeit bei der Mitschrift von Vorlesungen u.a. an Hochschulen dar. Hier wäre in weiteren Studien zu untersuchen, ob tatsächlich mit einem Eingabestift, der über eine druck- und neigungssensitive elektronischen Steuerung verfügt, ähnlich gute Lernergebnisse erreicht werden, wie beim Einsatz des klassischen Stifts und Papiernotizblocks.

Literaturverzeichnis

- [ADA03] AreaDigital AG (Hrsg.): Sony Ericsson P800, 2003, <http://www.reamobile.de/testberichte/12010-sony-ericsson-p800> [2016-05-10].
- [App16] Apple Inc. (Hrsg.): Apple Pencil für das iPad Pro, 2016, <http://www.apple.com/de/apple-pencil> [2016-05-10].
- [Arn13] Arndt, H.-K.: Umweltinformatik und Design - Eine relevante Fragestellung? In: Horbach, M. (Hrsg.): INFORMATIK 2013: Informatik angepasst an Mensch, Organisation und Umwelt (16.–20. September 2013, Koblenz, Germany), GI-Edition-Lecture Notes in Informatics (LNI), P-220, Gesellschaft für Informatik e.V., Bonn, 2013, S. 931-939.
- [Bau68] Bausinger, H.: Zu den Funktionen der Mode, Ludwig-Uhland-Institut für Empirische Kulturwissenschaft, Wirtschafts- und Sozialwissenschaftliche Fakultät, Eberhard Karls Universität Tübingen, 1968, https://publikationen.uni-tuebingen.de/xmlui/bitstream/handle/10900/47650/pdf/Bausinger_Hermann_Zu_den_Funktionen_der_Mode.pdf?sequence=1&isAllowed=y [2016-05-10].
- [Bor03] Borchers, D.: Apples Newton: Der Schwerkraft getrotzt, doch der Zeit voraus, heise online: Wissen | Hintergrund, 02.08.2003, <http://www.heise.de/ct/artikel/Apples-Newton-Der-Schwerkraft-getrotzt-doch-der-Zeit-voraus-301956.html> [2016-05-10].
- [Car15] Careerplus (Hrsg.): Jubiläumsblog: Disketten formatieren, Windows 95, Palm Pilot – so sah IT in den letzten 20 Jahren aus, 17.04.2015, http://www.careerplus.ch/sites/default/files/palmpilot_img_7025.jpg [2016-05-10].
- [Dud16] Bibliographisches Institut GmbH (Hrsg.): Eingabestift, 2016, <http://www.duden.de/rechtschreibung/Eingabestift> [2016-05-10].
- [Erm12] Ermisch, S.: Tablets mit Stiften: Kritzeln statt tippen, 21.08.2012, <http://www.manager-magazin.de/lifestyle/hardware/a-850921.html> [2016-05-10].

- [FrK05] Fried, I., Kaufmann, J.: Windows Vista: Tablet PC lernt Handschrift des Nutzers, 19.09.2005, <http://www.zdnet.de/39136612/windows-vista-tablet-pc-lernt-handschrift-des-nutzers> [2016-05-10].
- [Gab16] Springer Gabler Verlag (Hrsg.): Gabler Wirtschaftslexikon, Stichwort: Konjunktur, <http://wirtschaftslexikon.gabler.de/Archiv/90190/35/Archiv/90190/konjunktur-v5.html> [2016-05-10].
- [Gra15a] Grabmair, M.: iPad Pro mit Stylus: Ein Verrat an Steve Jobs Prinzipien? – ein Kommentar, 20.01.2015, <http://www.maclife.de/news/ipad-pro-stylus-verrat-steve-jobs-prinzipien-kommentar-10061706.html> [2016-05-10].
- [Gra15b] Grabmair, M.: Apple Pencil: „Wer will schon einen Stylus?“ – Apples Stylus hat die Ansicht von Steve Jobs überholt, 09.09.2015, <http://www.maclife.de/news/apple-pencil-will-schon-einen-stylus-apples-stylus-hat-ansicht-steve-jobs-ueberholt-10069289.html> [2016-05-10].
- [Lüd06] Lüders, D.: Hemdtaschen-kompatibel: Vor zehn Jahren erblickte der Palm Pilot das Licht der Welt, heise online, 2006, <http://www.heise.de/ct/artikel/Vor-zehn-Jahren-erblickte-der-Palm-Pilot-das-Licht-der-Welt-1900059.html> [2016-05-10].
- [Mic16] Microsoft Corporation (Hrsg.): Anpassen des Tablet PC-Eingabebereichs, 2016, <http://windows.microsoft.com/de-de/windows-vista/customizing-tablet-pc-input-panel> [2016-05-10].
- [MLr93] Meyers Lexikonredaktion (Hrsg.): Meyers neues Lexikon in 10 Bänden, Bd. 6, Stichwort: Mode, Meyers Lexikonverlag, Mannheim/Leipzig/Wien/Zürich, 1993, S. 429.
- [Pal00] Palm Inc. (Hrsg.): Handbuch für den Palm™ V-Handheld, Santa Clara, California, 2000, S. 31f.
- [Saß14] Saße, D.: Besser lernen mit Stift statt Tastatur, 04.05.2014 <http://www.heise.de/newsticker/meldung/Besser-lernen-mit-Stift-statt-Tastatur-2182057.html> [2016-05-10].
- [Sch16] Schelhorn, M.: Stifte und geeignete Apps - Die besten Stifte für das iPad, 19.04.2016, <http://www.macwelt.de/ratgeber/Besser-malen-und-schreiben-7893269.html>, [2016-05-10].
- [Spe03] Spehr, M.: Der Alleskönner: Das P800-Handy von Sony Ericsson: ein Traum von Technik, 17.05.2003, <http://www.faz.net/aktuell/technik-motor/audio-video/der-alleskoenner-das-p800-handy-von-sony-ericsson-ein-traum-von-technik-1100270.html> [2016-05-10].
- [TMG14] Telegraph Media Group Limited (Ed.): Ten of the greatest technology flops: The Apple Newton PDA - or MessagePad, 2014, http://i.telegraph.co.uk/multimedia/archive/01470/apple-gallery6_1470383i.jpg [2016-05-10].
- [UPU03] Pambor, U.: Ulf's Space Age LED-Uhren: Die HP-01-Rechneruhr aus dem Jahr 1977, 2003, http://www.pambor.de/Pics/Resize%20of%20HP01_Calc.JPG [2016-05-10].
- [Vit16] Vitsoe Ltd (Hrsg.): Dieter Rams: Zehn Thesen für gutes Design, <https://www.vitsoe.com/de/ueber-vitsoe/gutes-design> [2016-05-10].

- [WDD09] Webdesiger Depot (Ed.): The Evolution of Cell Phone Design Between 1983-2009: Sony Ericsson P800, 22.05.2009, http://netdna.webdesignerdepot.com/uploads/cellphone_design/dkmb86g_436gp75q9cq_b.jpg [2016-05-10].
- [WLG12] World's Luxury Guide (Hrsg.): Die Quarzuhr und ihre Schattenseiten, <http://luxus.welt.de/uhren/die-quarzuhr-und-ihre-schattenseiten> [2016-05-10].

The Development of an Eco-Label for Software Products – a Transdisciplinary Process?

Eva Kern¹

Abstract: Research activities are mainly realized in exclusive academic contexts, excluding social actors. However, some disciplines gain a big social relevance and thus should include non-academics in the research process. One example could be the research in the context of environmental impacts of ICT. Hence, the following paper addresses the question if it stands to reason to apply a transdisciplinary project in the field of Environmental and Sustainability Informatics. To do so, attributes, requirements and a conceptual model of transdisciplinary research are applied to the case example of labelling green software products. The result of this theoretical analysis is that it is possible and promising to apply transdisciplinary ideas to sustainability informatics projects.

Keywords: Environmental Informatics, Transdisciplinary, Green Software Products

1 Introduction

Environmental and sustainability informatics can be basically understood as an interdisciplinary research field bringing together ideas of the chemical, biological, environmental, sustainable and even more disciplines with information and communication technologies (ICT). According to [Hi95] environmental informatics „combines computer science topics such as data base systems, geographic information systems, modeling and simulation, computer graphics, user interface design, knowledge processing, and neural networks, with respect to their application to environmental problems.“ Based on that, sustainability informatics can be seen as an advancement of environmental informatics, dealing with methods, models and information systems supporting a sustainable development [Na08].

In this paper, we differentiate between both – environmental and sustainability informatics – as follows: *Environmental Informatics* deals with e.g. environmental issues of ICT and ICT based applications for environmental and natural sciences. The focus is set on environmental related topic, namely so called “green issues”. Generally, there is a differentiation between “Green in IT” and “Green by IT”. *Sustainability Informatics* extends the activities and comes up with a holistic view of all aspects of sustainability: the environmental, economic and social dimension according to Brundtland. [Un87] Moreover, some approaches add an individual and a technical dimension. [Pe13] Here, one can differ similarly between “Sustainable in ICT” and “Sustainable by ICT”.

¹ Leuphana University Lüneburg, Scharnhorststrasse 1, 21335 Lüneburg, mail@nachhaltige-medien.de

Hence, both research fields deal with aspects that are socially relevant. Moreover, human and social behavior is an issue in applying new strategies, that could be outcomes of research activities, aiming at e.g. reducing environmental impacts of ICT (Green in IT) or at using ICT in order to react to environmental problems by means of ICT (Green by IT). Just to give one example: if the research community comes up with new ideas how to reduce the carbon footprint of using personal computers in offices, these ideas have to be implemented by those working in offices and being able to improve the carbon footprint here. Thus, there seems to be a dependence of scientific approaches and their implementation in practical applications. So far, this connection between research and the practice is just rarely represented in the research activities regarding green and sustainable ICT although it seems to be relevant. New ideas, concepts, models, methods and strategies addressing environmental impacts and sustainability issues of ICT are mainly driven by the science community and do hardly include the societal side.

As a consequence, the idea is to bring the principles of transdisciplinary projects into the sustainability informatics field. The idea of transdisciplinary research in context with green ICT is not new [Hi11]. However, corresponding structures and methods are still missing. Therefore, the following paper describes transdisciplinary research at first (Section 2) and reviews the possibilities of integrating those aspects into environmental and sustainable informatics with the help of this description and a case example. The case example comprises labelling green software products and will be introduced in Section 3. The analysis of the fitting of transdisciplinary characteristics and methods with sustainability informatics will follow in Section 4. So far, the idea of bringing together sustainable ICT and transdisciplinary research by using existing methods is just a theoretical approach. The implementation into practice is still missing. Nevertheless, the presented concept might be seen as a driver to integrate new principles into the research activities regarding sustainable (by) IT, depending on the results of the analysis. The paper closes with a conclusion and outlook (Section 5).

2 Transdisciplinary Research: A short Overview

Although the principles of transdisciplinary research are known for years, there is no commonly accepted definition for transdisciplinary so far [Ja12]. One possible definition is presented by Lang et al.: *„Transdisciplinarity is a reflexive, integrative, method-driven scientific principle aiming at the solution or transition of societal problems and concurrently of related scientific problems by differentiating and integrating knowledge from various scientific and societal bodies of knowledge.“* [La12] In short, Klein et al. [K112] talk about *„Joint Problem Solving among Science, Technology and Society“*. Both definitions point out the importance of including the society into the research activities.

Transdisciplinary projects deal with „real-world problems“, going further than interdisciplinary projects and aim at producing knowledge that is more reflected than

results solely created in scientific processes [Ja12], [K112]. In order to do so, transdisciplinary processes have to „[go] across and through the different disciplines, and beyond each individual discipline“ and to „[transcend] each of their traditional boundaries“ [A111]. By using “community-based, interactive or participatory research approaches” [La12] both the goals of the society and science should be met. Currently, various examples of transdisciplinary research projects can be found in sustainability science, generally known as a problem-driven and solution-oriented field.

By including the society, those stakeholders used to respond and react to the research activities play a bigger role since they become partners in applying these activities and coming up with solutions [Mo10]. At last, the problem definition is oriented towards the needs of non-scientific project participants. The difficulties and goals of a transdisciplinary research project are formulated in real-world contexts independently from scientific theories [A108]. They are then addressed by actors from both areas – scientific as well as social. Thus, while researching the problems, one out of many challenges is to find a suitable way to communicate since all participants “have to become involved as persons who bring distinct interests, roles and practices of communication” [Ja12]. However, the increased effort might be relativized by realizing a learning process for both sides as well as empowering and motivating stakeholders to support the implementation of solutions [La12].

This description of transdisciplinary research just gives an overview of the research field. However, since standardized definitions and methods are missing, none of the descriptions that can be found in current literatures seem to be comprehensive. In the following (Section 4), some more characteristics will be presented and directly combined with the case example, belonging to the field of environmental informatics, that will be presented in the next Section 3.

3 Case Example: Labelling Green Software Products

While the awareness for the environmental impact of software products over their life cycles is addressed with different research activities, the society is still only slightly or not informed about resource consumption and similar effects of software. This is contrary to the fact that those who are developing software products in the majority of cases or constitute the biggest group of software users are non-academic actors.

Hence, the idea is to create a label for software products informing about their green characteristic. There could be a certification of software products in the sense of quality labels and based on standardized criteria to create awareness for the aforementioned topics. In order to reach this aim, the following steps are necessary:

- Finding a commonly accepted definition for the term “green software product”
- Defining criteria to characterize green software products

- Developing methods to measure and proof the greenness of a software product
- Preparing a communication strategy, informing about the green software label
- Nominating an audit committee to award green software products
- In [Kel15] we presented a summary of relevant aspects for the development of a product label for green and sustainable software products. The following aspects were addressed: definition, criteria, form of representation, target groups, and stakeholders. Thus, there are some ideas and approaches for a green software label. One of the findings was that a bold and simple label might be a first step and starting point for advancements. It could be characterized as follows:
- The label in mind should award environmental friendly products, i.e. the underlying criteria evaluate environmental impacts of the product but do not include social or economic aspects.
- It should award the product itself, i.e. without the development and distribution process, the surroundings, and the organization developing the product. The label refers to the whole life cycle of the product but does not include anything that is not part of the product. That means it values e.g. the algorithm and the modules of application software but not the packing of it.
- It should award “green in software” aspects. That means the product itself should be as environmental friendly as possible. Products that support environmental friendly processes or rather make them more environmental friendly (“green by software”) but are not green itself are not included.
- Within this paper, these ideas should be taken as a case example to analyze if the idea of transdisciplinary projects fits into the context of environmental informatics.

4 Analysis: Transdisciplinary Development Process of an Eco-Label for Software Products?

In order to analyze if it stands to reason to apply a transdisciplinary project in the field of environmental and sustainable informatics, we will use different methods in the following section. These include applying transdisciplinary attributes as well as requirements to the case example and the description of possible activities in a labelling development process that could apply the process to a model for transdisciplinary projects. The results of the analysis will be discussed in Section 5.

4.1 Applying Attributes of Transdisciplinarity

As already mentioned, there is no commonly accepted definition for transdisciplinarity

[Ja12], [La12]. Consequently, corresponding quality standards as well as standardized guidelines, methods and characteristics for transdisciplinary research projects are missing so far. However, one can find various approaches for those in the current literature. For example, Kroeze et al. [Kr14] list main attributes of transdisciplinarity. In the following, these will be implied to the case example out of the field of environmental informatics presented in Section 3. The idea is to get an impression if characteristics of transdisciplinarity can be transferred to environmental and sustainable ICT research questions.

The following attributes are extracted from [Kr14] and interpreted in means of labelling green software products:

- *Collaborative nature: Collaboration of researchers from various disciplines, integrating various insights*
In order to find criteria, corresponding measurement methods and guidelines to present a label for green software product, it is necessary to include stakeholders from different disciplines, e.g. computer science, life cycle assessment, communication science, etc. Above, it seems to be promising to include social scientists bringing in knowledge about behaviorism und addressing consumers. Thus, next to the inclusion of society, researchers from various disciplines should participate to find a well-funded solution. Here, just some examples of potential participating fields are given.
- *Goal: To solve problems; to change behavior significantly*
Here, the problem is, among others, the high energy consumption caused by ICT. The consumption should be reduced and further environmental impacts should be addressed in a positive manner by changing the user's behavior while using ICT products. The idea behind the label for green software products is to create awareness for the environmental impacts caused by producing, using and deactivating software. This might be the first step on the way to a more environmental friendly dealing with software products. In the context of transdisciplinarity, Lang et al. [La12] go one step further and talk about aiming at an "intensive learning process" and "motivating stakeholders to contribute more actively to the implementation". This goal could be also persuade with a green software label.
- *Disciplinary nature: Across disciplines, beyond academic disciplines, looking outwardly*
So far, the issue of green and sustainable software is mainly researched in scientific contexts. Here, different disciplines are coming together to find a solution. However, it seems to be necessary to additionally include ideas, requirements and approaches from those who should use the label to be developed.
- *Theoretical nature: Theoretical exercise leading to practical solutions*
The main exercise in labelling green software products is to develop a theoretical concept, namely awarding criteria and a corresponding program. Based on this

theoretical construct, the challenge is to come up with a solution that can be transferred into a practical application field.

- *Source / cause: Stimulated by complex practical problems*
The climate change is one big and complex problem that is related to the real-world. Thus, the motivation for green and sustainable ICT, including the hardware as well as the software side, is based on complex practical problems. Since the source of the challenge of addressing the climate change is not only a scientific issue, it fits with the characteristic of transdisciplinary research. [A108]
- *Methods: Holistic approach to unify knowledge inputs*
In order to create a green software label that will be accepted by scientific and societal actors, it is necessary to combine strategies and methods from different disciplines. It is important that societal actors act as partners in the joint research project [Ja12] to be able to run the project in a transdisciplinary way. Above, it is necessary to find a way of communication everyone can deal with.
- *Result: Better solutions for complex problems*
Analyzing the impact of ICT onto the environment aiming at solutions to decrease is a challenge. Hence, better solutions are required. In this context, “better” means that the environmental impacts of software are realized by more actors. One step to go into this direction is to create a label for green software product in order to create and raise awareness. Therefore, the label should be accepted by many consumers who may promote it. The acceptance for the label should be created by integrating different stakeholders into the label development process as soon as possible to hopefully meet their interests as well as needs in a better way.
- *Scope: Society (broad)*
On the one hand the labelling development project acts a part for the scientific community since it requires a definition of green software products as well as corresponding criteria and reference methods. It might be able to answer current research question. On the other hand the label can be understood as a tool for the society to bring transparency into the activities dealing with environmental impacts of ICT. For us, the second point should be the main target of the project so that the scope should be mainly laid on societal requirements and needs.
- *Properties: Evolving*
Developing a label for green software project is an evolving project since one strategy is to develop a simple and bold label as a first run and start the information process to end users with this. Based on that, the next step could be a more detailed label, presenting more information and / or being based on more criteria. Find more about this strategy in [Ke15].

4.2 Applying Requirements for Transdisciplinary Research

Next to the listed attributes, Lang et al. [La12] come up with requirements that should be

met by transdisciplinary research. Similarly, these requirements will be brought together with the case example in the following:

- *Focusing on societally relevant problems*
 Generally, one of the issues addressed by research activities in environmental and sustainable informatics is the climate change. This can be seen as a societally relevant problem. Indeed, it is not proofed that the greenness of software or rather information technology can also be counted among this kind of problems, neither is the labelling of green software products.
- *Enabling mutual learning processes among researchers from different disciplines (from within academia, other research institutions, and outside academia)*
 Researchers coming from different disciplines and backgrounds are researching on joined projects in the environmental informatics field. Thus, it is possible to start corresponding learning processes. However, in order to meet the requirement of transdisciplinary, non-scientific actors need to be integrated in the activities. This seems to be possible or rather promising in case of labelling green software since different stakeholders from academic and non-academic areas are relevant in transferring to a more sustainable software usage.
- *Aiming at creating knowledge that is solution-oriented, socially robust, and transferable to both the scientific and societal practice*
 The knowledge that is developed in research projects focusing on environmental issues of ICT should be solution-oriented, socially robust, and transferable to practice. This makes an implementation of the generated approaches possible.

4.3 Applying a Conceptual Model for Transdisciplinary Research

Similarly to the characteristics of transdisciplinary research, standardized methods for this kind of research projects are still missing [Ja12]. However, Jahn et al. [Ja12] present an ideal-typical conceptual model for transdisciplinary project that is based on previous research and further developed by Lang et al. [La12]. Both of them apply the model to a finalized research project to outline the different phases. Since our case example is a concept for a research project so far, it is not possible to describe the results of the model phases. Thus, the following Table 1 lists possible activities for each of the phases to analyze if the concept is transferable to the project example.

Table 1 Applying a Conceptual Model for Transdisciplinary Research to the Case Example of Labelling Green Software Products

Phase	Project activities	Possible answers
Phase 1– Problem transformation		
Step 1.1: Framing the societal	- Brainstorming: Who are the stakeholders?	- Stakeholders: end user, software engineers, sustainability activists,

problem	<ul style="list-style-type: none"> - Stakeholder workshops and interviews in order to address the following questions: (a) Is there any knowledge about or rather awareness for the software side of Green IT? If so, what kind of knowledge and on which level? (b) How is the relevance of the issue valued? - Aim: framing the problem and projects, find a starting point and outline project targets 	<p>researchers, purchaser, politicians, etc.</p> <ul style="list-style-type: none"> - Starting point: analyzing the existing research approaches
Step 1.2: Relating the societal problem description to scientific knowledge	<ul style="list-style-type: none"> - Clarifying how to deal with the results of step 1.1. - Analyzing where to find research approaches that can be further developed in a transdisciplinary research team - Finding a common focus for the project - Based on the focus, developing a strategy to frame the project team and to specify the stakeholders that should be involved. 	<ul style="list-style-type: none"> - In order to find research approaches: start with the existing literature reviews - Project focus: finding the aspect of sustainable software that meets the scientific and social relevance in the best way
Step 1.3: Transformation of the societal problem into a boundary object	<ul style="list-style-type: none"> - Defining that project target(s) that should be persuade - Defining criteria to decide if the project was successful or not 	<ul style="list-style-type: none"> - Project targets: creating a label for green software products - Project criteria: the label should be taken into account while buying new software products (empirical evaluation)
Step 1.4: Transformation of the boundary object into epistemic objects	<ul style="list-style-type: none"> - Specifying strategies and methods the project target should be addressed with - Updating the criteria defining the project success (if necessary) 	<ul style="list-style-type: none"> - Transferring existing labels towards software products - Development of suitable criteria, measurement methods - ...
Phase 2 – Interdisciplinary integration		
Step 2.1:	<ul style="list-style-type: none"> - Finding a way to allocate the 	<ul style="list-style-type: none"> - Working groups

<p>Clarification of the roles of researchers and stakeholders</p>	<p>project roles: who is doing what and when?</p> <ul style="list-style-type: none"> - Specifying communication strategies between different stakeholders and outwards - Deciding if it would be helpful to use empirical methods to integrate the social point of view from extern (representatively) 	<p>comprising researchers and software users (being the critical reflection)</p> <ul style="list-style-type: none"> - Communication through regular meetings - Evaluation of the ideas by a representative survey and / or focus groups
<p>Step 2.2: Design of an integration concept</p>	<ul style="list-style-type: none"> - Defining how to work in a collaborative way - Defining the way of dealing with the research question - Defining how to combine the results 	<ul style="list-style-type: none"> - Assembling one working group for each life cycle phase of software products or related to the sustainability aspects - Evaluating the results of each group in a joined process
<p>Step 2.3: Implementation of the integration concept</p>	<ul style="list-style-type: none"> - Applying the different methods to the specific questions and project groups - Defining how to combine the results of different project groups - Finding solutions how to evaluate the results (tried and tested, accepted, verifiable, ...) 	<ul style="list-style-type: none"> - Development of criteria: creativity methods might be helpful - Evaluation by an external steering committee
<p>Phase 3 – Transdisciplinary integration</p>		
<p>Step 3.1: Assessment of integrated results</p>	<ul style="list-style-type: none"> - Retrospection towards the starting question and project definition to evaluate if the results are problem-solving - Finding waysto implement the results and transfer them to the scientific and societal practice ... to present the results ... to extract lessons learned, next steps etc. 	<p>The aspects are mainly depending on the kind of project team and should be addressed in a transdisciplinary way.</p>
<p>Step 3.2: Assembly of products for</p>	<ul style="list-style-type: none"> - Preparing the results in order to make them available for everyone who is interested 	

science and society	- Finding strategies to enable a dissemination and if possible a further development of the results	
---------------------	---	--

4.4 Round-up: Transdisciplinary Projects in Environmental Informatics

Since information and communication technologies are a vital part of the everyday world and today's society, the idea is to integrate social actors into the movements of computer sciences. The view of the society seems to be especially important in context of environmental and sustainable topics since the environment could be understood as the basis of the life of society. Hence, the approach of transdisciplinary research seems to be appropriate.

Overall, the analysis shows that many of the attributes of a transdisciplinary project by Koerze et al. [Kr14] could be transferred to the case example of labelling green software products. By applying the characteristics of a transdisciplinary project to the case example it turned out that they seem to be easily transferable. Environmental informatics is already working together with different disciplines, the collaborative as well as disciplinary nature is given, the research field is solution-oriented by nature and evolving properties are not new. However, aiming at a change behavior and setting the scope onto the broad society might be unprecedented in many cases.

Additionally, the phases of the conceptual model for transdisciplinary research by Jahn et al. [Ja12] or rather Lang et al. [La12], can be transferred to the case example. However, this project strategy covering the described three phases seems to be new for most of the known Green ICT projects. Most of them are mainly research-oriented and implemented in research teams excluding society. Nevertheless, we see the potential to transfer an interdisciplinary project into a transdisciplinary one: there are already projects including stakeholders from different backgrounds, research results are published to a non-specific audience in some cases and teams are working in an interdisciplinary way. That means, it is possible to imagine an implementation of the development process of a green software label by applying the proposed phases of the model. Thus, it seems to be promising to run some projects of environmental informatics as transdisciplinary projects. One example for that is presented in our paper. Possible challenges in doing so will be pointed out in the last Section 5.

Working in an interdisciplinary way and above including the view of society may also drive the innovation of the different disciplinarians forward [Ja12]. Contrary to that, some critics argue that a joint research between scientific and non-scientific actors is not so much research but mutual learning [La12]. Even if this turns out true, we argue that learning from each other and together will lead to new ideas, approaches and innovations. If they are not scientific as it is understood traditionally, these results may influence the research activity addressing the same questions in a positive way.

However, transdisciplinary methods do not fit in all of the cases. Thus, we suggest verifying the matching of the concept in each single case. If possible, it seems to be helpful to analyse the social relevance of the research question before starting a transdisciplinary project. If this relevance is given, the interest and commitment of the society to participate in the project might be higher. In this context, the practice orientation is also an issue since it is assumed that more theoretical questions are less socially relevant at first appearance.

5 Conclusion and Outlook

This paper proposes to apply transdisciplinary projects to research questions from the field of sustainable and environmental informatics. Using the project of labelling green software products as a case example, we come to the conclusion that it is possible and promising to apply transdisciplinary ideas to environmental informatics projects. This conclusion is supported by the application of attributes, requirements and a conceptual model to the case example. We found that many characteristics of transdisciplinary research are compatible with current research challenges addressing environmental aspects of ICT.

Summarizing, we come to the following findings: (1) It sounds promising to address questions regarding environmental and especially regarding sustainable issues of ICT by transdisciplinary projects, i.e. to integrate non-academic persons into the activities. (2) The case example presented in our paper corresponds to transdisciplinary methods and ideas. Generally, we see requirements that have to be fulfilled and / or characteristics a problem should have in order to be researched by a transdisciplinary project. (3) The social relevance and, if possible, also the practice orientation should be given before starting a transdisciplinary project.

Next to these findings we see the following challenges in implementing a transdisciplinary Green ICT project: converting non-scientific stakeholders to join actively in the project, finding a common base of exchanging ideas during the project process, declining a collaborative learning environment, creating a structure including methods of computer science and transdisciplinary research, presentation of project results to non-academic fields and other disciplines.

The next step after the presented theoretical analysis should be a practical application of a transdisciplinary project in environmental informatics. Doing that, the theoretical approach could be reviewed. Next to the review an adaption of the conceptual model into the context of computer sciences might be an outcome of a practical implementation.

Bibliography

- [Al08] Albuquerque, J. P. de et al.: The challenge of transdisciplinarity in information systems research: Towards an integrative platform. In *Information systems research methods, epistemology and applications*, 2008; pp. 88–103.
- [Al11] Alvargonzález, D.: Multidisciplinarity, interdisciplinarity, transdisciplinarity, and the sciences. In *International Studies in the Philosophy of Science*, 2011, 25; pp. 387–403.
- [Hi11] Hilty, L. M.; Lohmann, W.: The Five Most Neglected Issues in “Green IT”. In *CEPIS UPGRADE*, 2011, XII; pp. 11–15.
- [Hi95] Hilty, L. M. et al.: *Environmental informatics as a new discipline of applied computer science: Environmental Informatics*. Springer, 1995; pp. 1–11.
- [Ja12] Jahn, T.; Bergmann, M.; Keil, F.: Transdisciplinarity: Between mainstreaming and marginalization. In *Ecological Economics*, 2012, 79; pp. 1–10.
- [Ke15] Kern, E. et al.: Labelling Sustainable Software Products and Websites: Ideas, Approaches, and Challenges. In (Johannsen, V. K. et al. Eds.): *Proceedings of EnviroInfo and ICT for Sustainability 2015. 29th International Conference on Informatics for Environmental Protection (EnviroInfo 2015) and the 3rd International Conference on ICT for Sustainability (ICT4S 2015)*. Copenhagen, September 7 - 9, 2015. Atlantis Press, Amsterdam, 2015; pp. 82–91.
- [Kl12] Klein, J. T. et al.: *Transdisciplinarity: joint problem solving among science, technology, and society: an effective way for managing complexity*. Birkhäuser, 2012.
- [Kr14] Kroeze, J. H.; van Zyl, I.: *Transdisciplinarity in Information Systems: Extended Reflections: Proceedings of the 20 th Americas Conference on Information Systems*, 2014.
- [La12] Lang, D. J. et al.: Transdisciplinary research in sustainability science: practice, principles, and challenges. In *Sustainability science*, 2012, 7; pp. 25–43.
- [Mo10] Mobjörk, M.: Consulting versus participatory transdisciplinarity: A refined classification of transdisciplinary research. In *Futures*, 2010, 42; pp. 866–873.
- [Na08] Naumann, Stefan: *Sustainability Informatics – A new Subfield of Applied Informatics?* In: Möller, Andreas; Page, Bernd; Schreiber, Martin (Eds.): *EnviroInfo 2008. Environmental Informatics and Industrial Ecology, 22nd International Conference on Environmental Informatics*. Shaker Verlag, Aachen 2008, pp. 384-389
- [Pe13] Penzenstadler, B.; Femmer, H.; Richardson, D.: Who is the advocate? stakeholders for sustainability: Green and Sustainable Software (GREENS), 2013 2nd International Workshop on, 2013; pp. 70–77.
- [Un87] Report of the World Commission on Environment and Development. *Our common future*. UN document no. A/42/427 English, New York, 1987.

IT und Dienstleistungen für die Energiewende und Elektromobilität (IDEE)

IT und Dienstleistungen für die Energiewende und die Elektromobilität (IDEE 2016)

Gerhard Satzger¹, Daniel Beverungen², Martin Matzner³ und Carola Stryja⁴

Liebe Leserinnen und Leser,

der Workshop IT und Dienstleistungen für die Energiewende und die Elektromobilität (IDEE) fand bereits zum vierten Mal im Rahmen der GI-Jahrestagung statt. Der halbtägige Workshop am 30.09.2016 in Klagenfurt/Österreich wurde gemeinsam von Vertretern des Karlsruher Instituts für Technologie, der Universität Paderborn und der Westfälischen Wilhelms-Universität Münster veranstaltet und hatte die Entwicklung innovativer Informationssysteme und IT-gestützter Dienstleistungen für die Energiewende und die Elektromobilität sowie Methoden zur Entwicklung dieser Informationssysteme und Dienstleistungen zum Thema.

Im Rahmen des Workshops wurden Beiträge präsentiert, die die Thematik auf vielfältige und spannende Art und Weise beforschen. Ein Beitrag befasst sich dabei mit der Lastverteilung im Stromnetz und stellt Kriterien zur Bewertung von Datensätzen auf, die eine wichtige Grundlage für Demand Side Management darstellen. Gleich zwei Beiträge behandeln das Thema Geschäftsmodelle u.a. für die Elektromobilität: Ein Paper befasst sich mit der Bewertung eines Erfolgsmodells zur Prognose der E-Carsharing-Nutzungsabsicht indem es auf Kundenentscheidungen, Elektromobilitäts- und Carsharing-Akzeptanz zurückgreift. Der zweite Beitrag zum Thema beschreibt die Thematik, wie die Modellierung von Geschäftsmodellen mittels Software unterstützt werden kann und nähert sich dem Thema mit einem umfassenden Vergleich von elf Modellierungs-Tools sowie dem Vorschlag von Gestaltungsoptionen und Weiterentwicklungsperspektiven. Ein weiterer Beitrag beschäftigt sich damit, wie Elektromobilität im Bereich des Wohnungsbaus besser integriert werden kann. Die Autoren stellen dazu verschiedene Konzepte für ein Mieterticket vor und beschreiben deren Anforderungen sowie die Erwartungen und Bedürfnisse der betroffenen Akteure. Um innovative IT-gestützte Elektromobilitätsdienstleistungen schließlich entwickeln zu können, müssen die Bedürfnisse der Kunden bekannt sein. Ein Beitrag befasst sich daher mit der Identifikation von Nutzerbedürfnissen durch die Analyse von Mikroblog-

¹ Karlsruher Institut für Technologie, Karlsruhe Service Research Institute, Englerstraße 11, 76131 Karlsruhe, gerhard.satzger@kit.edu

² Universität Paderborn, Wirtschaftsinformatik, insb. Betriebliche Informationssysteme, Warburger Straße 100, 33098 Paderborn, daniel.beverungen@uni-paderborn.de

³ Westfälische Wilhelms-Universität Münster, Lehrstuhl für Wirtschaftsinformatik und Informationsmanagement, Leonardo-Campus 3, 48149 Münster, martin.matzner@ercis.uni-muenster.de

⁴ Karlsruher Institut für Technologie, Karlsruhe Service Research Institute, Englerstraße 11, 76131 Karlsruhe, carola.stryja@kit.edu

Daten (Twitter) und vergleicht diese mit Erkenntnissen aus Studien, die auf Umfragen und Interviews basieren.

Wir danken dem Programmkomitee, das dieses Jahr aus neun Mitgliedern bestand, für seine Arbeit im Begutachtungsprozess sowie der Konferenzleitung der GI-Jahrestagung für die stets freundliche und kooperative Zusammenarbeit, die den Workshop so erst möglich gemacht hat.

Mit besten Grüßen

Prof. Dr. Gerhard Satzger

Prof. Dr. Daniel Beverungen

Dr. Martin Matzner

Dipl. Wi.-Ing. Carola Stryja

Datensets für Demand-Side-Management – Literatur-Review-Basierte Analyse und Forschungsagenda

Dennis Behrens¹, Thorsten Schoormann² und Ralf Knackstedt³

Abstract: Demand-Side-Management (DSM) beschreibt die Bestrebung, das Stromnetz bzw. die Lastverteilungen ausbalanciert zu gestalten. Zur Realisierung von DSM existieren verschiedene Ansätzen, die bislang nur unzureichend evaluiert und verglichen wurden. Als Voraussetzung sowohl für das Evaluieren als auch das Vergleichen dieser Ansätze sind jedoch geeignete Inputdaten, sog. Datensets, nötig. Um die Eignung bestehender Datensets zu prüfen, werden zunächst Anforderungen an DSM-Datensets aus existierenden Ansätzen abgeleitet und konzeptualisiert. Die Klassifizierung der Anforderungen stellt die Basis für das anschließende Literatur Review, bei dem insgesamt 17 relevante Datensets identifiziert wurden. Die Analyse der Datensets zeigt, dass Restriktionen hinsichtlich der Nutzerpräferenzen in keinem Datenset berücksichtigt werden. Im Rahmen der Forschungsagenda werden hierfür Lösungsansätze diskutiert.

Keywords: Demand-Side-Management, Datensets, Literatur-Review

1 Motivation

Das DSM-Konzept zur balancierten Verteilung von Lasten in Stromnetzen existiert bereits seit den achtziger Jahren [Ge85] und wird seither immer populärer. Grund hierfür sind u. a. erneuerbare Energien, die in das existierende Stromnetz integriert werden müssen und Elektroautos, die einen zusätzlichen Energiebedarf erzeugen, aber auch Potentiale bieten. Durch die in Deutschland seit 2010 in allen Neubauten vorgeschriebenen Smart Meter, ist es möglich, einzelne Lasten zu analysieren und zu steuern, sodass ein flexibles Lastmanagement realisiert werden kann bzw. der Nutzer weniger (aktiv) eingreifen muss (s. 21b Abs. 3a EnWG). Durch diese Entwicklungen ist DSM neben dem ursprünglichen industriellen Kontext auch im privaten Bereich anwendbar.

Zur Umsetzung von DSM existiert eine Vielzahl von Ansätzen, die durch z. T. starke Unterschiede geprägt sind [Ko13]. Diese Ansätze werden nicht oder nur unzureichend evaluiert, sodass die Effizienz nicht nachgewiesen werden kann [Ge15]. Zur Evaluation werden meist Datensets genutzt. Um der Realität möglichst zu entsprechen, werden

¹Universität Hildesheim, Informationssysteme und Unternehmensmodellierung, 31141 Hildesheim, dennis.behrens@uni-hildesheim.de

²Universität Hildesheim, Informationssysteme und Unternehmensmodellierung, 31141 Hildesheim, thorsten.schoormann@uni-hildesheim.de

³Universität Hildesheim, Informationssysteme und Unternehmensmodellierung, 31141 Hildesheim, ralf.knackstedt@uni-hildesheim.de

vorzugsweise reale Datensets genutzt [Ba14]. Diese unterscheiden sich jedoch hinsichtlich verschiedener Faktoren (z. B. [Mo14]).

Daher ist es nötig, geeignete Datensets zu nutzen, um DSM-Methoden analysieren, evaluieren und miteinander vergleichen zu können. Hier ergibt sich die Frage, welches Datenset für welche Methode und in welchem Kontext genutzt werden kann? Um diese Frage zu beantworten, wird zunächst der State of the Art der bestehenden Datensets analysiert, da diese zum einen sehr unterschiedlich sind und zum anderen nicht alle Anforderungen der DSM-Methoden erfüllen. Es kann daher die folgende Forschungsfrage abgeleitet werden: **Welche DSM-Datensets sind verfügbar und welche Anforderungen von DSM-Methoden werden von ihnen erfüllt?**

Zur Beantwortung werden in einem ersten Schritt bisherige Arbeiten betrachtet (Abschnitt 2). Anschließend wird das methodische Vorgehen vorgestellt (Abschnitt 3), welches eine Konzeptualisierung von DSM-Methoden vorsieht, bei der vorhandene Anforderungen von DSM-Methoden an Datensets identifiziert werden (Abschnitt 4). Darauf aufbauend wird ein Literatur-Review durchgeführt (Abschnitt 5) und die Resultate analysiert und diskutiert (Abschnitt 6). Abschnitt 7 beinhaltet die daraus abgeleitete Forschungsagenda.

2 Related Work

2.1 Datensets im privaten DSM-Kontext

Eine erste Analyse von existierenden DSM-Datensets wurde von [Mo14] durchgeführt. Sie beschreiben 13 Datensets, von denen ein Datenset selbst aufgezeichnet wurde. Dabei wurden die Lokalität und der Zeitraum, Häuser- und Sensorenanzahl, „Datenset-Features“, wie z. B. Blind- und Scheinleistung, sowie die Auflösung der Aufzeichnung – die Abtastrate – erhoben. Viele dieser Datensets sind nicht mit dem Ziel der Evaluation von DSM-Methoden aufgezeichnet worden, sondern finden sich im Non-intrusiv Load Monitoring (NILM) Kontext wieder. NILM bzw. die NILM-Community befassen sich primär mit der Identifikation von einzelnen Lasten, ohne diese separat aufzuzeichnen [Ha92]. Datensets werden hier genutzt, um Methoden zu evaluieren und zu trainieren.

Eine Übersicht bestehender Datensets, speziell für NILM, findet sich z. B. auch auf dem Blog von [Pa16] wieder, auf dem bekannte Datensets aufgelistet und beschrieben werden. Die Analyse zeigt, dass es keinen Aufzeichnungsstandard gibt. Einen ersten Ansatz – auf Typ- bzw. Metaebene einen Standard zu definieren – liefern [KK14]. Diese beschreiben, mit Hilfe eines Entity-Relationship-Models, ein Metamodell zur Vorgabe benötigter Daten und Strukturen eines Datensets.

2.2 Methoden und Benchmarking im privaten DSM-Kontext

[Ge15] haben einen ersten Überblick erstellt, bei dem 55 DSM-Ansätze, geordnet nach genutzten Methoden und Frameworks, analysiert wurden. Die Autoren folgen dabei der gängigen Einteilung in dezentral, autonom und zentral gesteuerte Frameworks [Ko13]. Dabei wird deutlich, dass keine einheitliche Evaluation oder Vergleichsinstrumente für die einzelnen Ansätze existieren. Ansätze (a) *evaluieren nicht*, (b) *beweisen, dass eine optimale Lösung gefunden wird* oder (c) *evaluieren mit einem Datenset*. Dabei unterscheiden sich die genutzten Datensets zwischen realen und künstlich-erzeugten. Problematisch sind Machbarkeit und Vergleichbarkeit bei allen Ansätzen.

Mit dieser Thematik hat sich z. B. die NILM-Community bereits beschäftigt [Ba14]. Hier wurde ein Benchmarking auf der Basis verschiedener realer Datensets entwickelt, bei dem mehrere NILM-Methoden miteinander verglichen werden können. Auf die Nutzung künstlicher Datensets wird verzichtet und stattdessen auf reale Datensets zurückgegriffen, da diese *die Realität besser abbilden* ([Ba14]). Eine weitere Problematik die adressiert wird, ist die Heterogenität der Datensets. Durch unterschiedliche Eigenschaften und Aufzeichnungsweisen müssen die Datensets zunächst standardisiert werden. Dazu nutzen [Ba14] das Metamodell von [KK14].

3 Methodisches Vorgehen

Zur Identifikation und anschließender Analyse verfügbarer DSM-Datensets anhand relevanter Eigenschaften und Anforderungen wird im Rahmen dieses Beitrages ein mehrstufiges Verfahren verfolgt (s. Abb. 1).

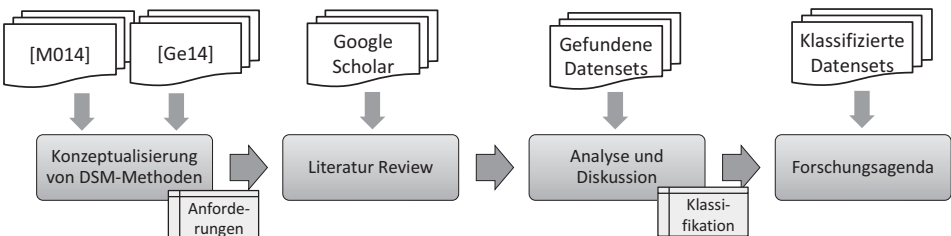


Abb. 1: Übersicht des methodischen Vorgehens

Zunächst werden Anforderungen an DSM-Methoden herausgearbeitet, um darauf aufbauend Dimensionen zur Klassifikation von Datensets abzuleiten. Insbesondere werden dazu die Beiträge von [Mo14] und [Ge15] berücksichtigt, die bereits eine erste Datensetanalyse mit entsprechenden Anforderungen enthalten bzw. eine Übersicht über DSM-Methoden bereitstellen, aus denen weitere Anforderungen identifiziert werden. Das Zwischenfazit der Konzeptualisierung ist die Tabelle „Konzepte“ mit Anforderungsdimensionen und Zuordnung zu DSM-Methoden.

Ziel eines Literatur Reviews ist es, das „große Ganze“ einer Domäne zu verstehen, um irrelevante Forschungsvorhaben zu reduzieren und das bestehende Wissen effektiv nutzen zu können [Br09]. Obwohl das Auswerten der bestehenden Literatur essentiell für die Forschung ist, existieren (a) *teilweise Unklarheiten über die Struktur*, sind (b) *bestehende Vorgehensweisen häufig aus anderen Domänen adaptiert* und (c) *Suchprozesse von gegebenen Guidelines nicht ausführlich beachtet* ([WW02]). Relevant ist vor allem der Suchprozess selbst, weshalb z. B. Literaturfunde, die Suchbegriff-Auswahl und die Evaluation der Ergebnisse präzise dokumentiert werden sollten [Br09]. Aufgrund der methodologischen Rigorosität von Literatur Reviews, folgen wir den fünf Schritten nach [Br09]: (I) *Definition des Forschungsbereichs (Abschnitt 1)*, (II) *Konzeptualisierung von DSM-Methoden (Abschnitt 4)*, (III) *Literatur Review (Abschnitt 5)*, (IV) *Analyse und Diskussion (Abschnitt 6)* sowie (V) *Forschungsagenda (Abschnitt 7)*.

4 Konzeptualisierung von DSM-Methoden

[Mo14] stellen bereits folgende Kriterien von DSM-Methoden vor:

Aufzeichnungsort, Zeitraum und Häuseranzahl. Wo ist das Datenset aufgenommen wurden, wieviele Tage (am Stück) umfasst es und wieviele Wohneinheiten sind enthalten?

Einzelaufzeichnung. Anzahl der einzeln aufgezeichneten Lasten. Entscheidend ist dabei, ob neben dem Gesamtverbrauch die Einzellasten separat aufgezeichnet wurden. [Mo14] geben hier die Anzahl Sensoren und z. T. die Typenanzahl an.

Dimensionierung. Einheit der Lasten. [Mo14] bezeichnen dies als „Features“, da hier u. a. Blind-, Schein- und Wirkleistung getrennt aufgezeichnet werden. Eine eindeutige Angabe in kWh (Wirkleistung) reicht jedoch für die meisten DSM-Methoden aus.

Feingranularität. Integration der Abtastintervalle, d. h. Abstände der Aufzeichnung der Messwerte. Da zwar die Aufwärtskompatibilität jedoch nicht die Abwärtskompatibilität gewährleistet ist, muss eine hohe Abtastrate vorhanden sein. Eine kontinuierliche Optimierung mit DSM ist in der Realität nicht praktikabel und die Taktung kann entsprechend grob sein (z. B. 15 Minuten, wie von vielen DSM-Methoden gefordert).

Bei der Analyse des Datenbedarfs diverser DSM-Methoden, die in [Ge15] aufgeführt sind, konnten die folgenden Anforderungen zusätzlich identifiziert werden:

Erzeugung, Wetterdaten und Speicher. Berücksichtigung der Erzeugung (z. B. PV-Anlagen), um möglichst wenig externen Strom kaufen zu müssen. Kennzeichnung externer Wetterdaten, um z. B. die Auswirkung auf die PV-Anlage(n) zu berechnen oder die Heizleistung abschätzen zu können. Alternativ die Nutzung von Energiespeichern, um z. B. Lastspitzen abzufedern oder Engpässe zu überbrücken.

Flexibilität. Kennzeichnung der zeitlichen Flexibilität von Lasten. Dazu sollten die Lasten getrennt und nicht aggregiert vorliegen. Einige Methoden nehmen unzulässiger Weise pauschal einen Anteil an, der pro Zeiteinheit verschoben werden kann. Eine geeignetere Abbildung wird erzielt, indem die verschiebbaren Lasten getrennt sind und einzeln in das Optimierungsproblem eingehen. Ein Ansatz zur Analyse der Verschiebbarkeit kann bspw. mit einer Referenztabelle erfolgen [KI09].

Trennbarkeit. Kennzeichnung der Trennbarkeit von Lasten. Lasten können zum einen nicht-teilbar sein, d. h. dass die Last während des Auftretens nicht unterbrochen werden kann (z. B. eine Waschmaschine). Andererseits existieren Lasten bei denen dies möglich ist, wie z. B. bei dem Ladevorgang eines Elektroautos. Viele Methoden betrachten daher die Lasten separat und tragen den verschiedenen Lastprofilen Rechnung.

Referenz	Ort	Zeitraum	Anzahl Häuser	Lasten einzeln	Dimensionierung	Feingranularität	Erzeugungs	Flexibilität	Trennbarkeit	Verschiebintervalle	Wetterdaten	Speicher
[Zh13]				X	X	X	X			X	X	
[CC11]				X	X	X	X	X	(X)	X		X
[MM13]		X		X	X	X		X				
[FG12]		X		X	X	X						
[VKM14]		X	X	X	X	X		X		X		
[CYN14]		X	X	X	X	X		X				
[SS14]			X	X	X	X		X		X		
[AYK14]			X	X	X	X		X	X			
[Ma14]				X	X	X		X	X	X		X
[An12b]				X	X	X	X	X				X
[Ji11]				X	X			X		X		
[Du13]			X		X							
[SL14]			X	X	X	X			X	X		
[LF14]				X	X	X	X					
[MHC12]				X	X	X		X		X		
[KG11]				X	X	X		X	X	X		
[MP11]				X	X		X	X	X			
[Al12]				X	X	X		X	X	X		
[Ba14]				X	X	X		X	X			
[Li15]				X	X	X						X
[Mo14b]				X	X							
[HMN12]				X	X	X		X	X	X		
[LPK12]				X	X	X		X				
[Hu14]				X	X	X	X	X				
[Za12]				X	X	X						
[LP13]				X	X	X		X		X		
[BN14]				X	X	X			X			
[Ma12]				X	X	X		X		X		
[AHK13]				X	X	X		X				
[Al13]				X	X	X		X		X		

Tab. 1: Anforderungen der DSM-Methoden

Verschiebungsintervalle. Definition von vordefinierten Anfangs- und Endpunkten zwischen denen eine Last verschoben werden kann. Einige Methoden integrieren daher Nutzerfunktionen. Diese Nutzerfunktionen berücksichtigen zusätzliche Präferenzen der Nutzer, in dem z. B. zusätzliche Kosten bei Missachtung anfallen. Sie enthält Informationen bezüglich des Nutzers und dessen Bedürfnisse, wie z. B. seinen Tagesablauf.

Des Weiteren fokussieren einige Methoden auf Elektroautos oder sog. HVAC-Lasten (Heating, Ventilation and Air-conditioning). Sofern diese keine weiteren Lasten betrachten, wurden die entsprechenden Beiträge nicht betrachtet. In Tab. 1 werden die Resultate dargestellt und den Anforderungen zugeordnet. Die grauen Spalten kennzeichnen neue Anforderungsdimension (vgl. [Mo14]).

5 Literatur Review

Identifikation der Suchbegriffe. Das Ermitteln von Such- und Schlüsselbegriffen im Vokabular der zu analysierenden Domäne ist ein entscheidender Bestandteil des gesamten Suchprozesses [Br09]. Initial wurden daher potenzielle Suchbegriffe verwendet und ausgewertet, um geeignete Items zu identifizieren. In Tab. 2 werden die Begriffe sowie deren Aufkommen in entsprechenden Suchmaschinen dargestellt, graue Zeilen kennzeichnen die ausgewählten Suchbegriffe.

Suchbegriff/-kombination	Google Scholar ⁴	Google	Beschreibung
„Demand-Side-Management“	55.500	466.000	Verknüpfte Items („“).
Demand Side Management	54.200	30.800.000	Auftreten eines Items genügt.
DSM	1.410.000	37.200.000	Abkürz. in anderen Domänen.
DSM „Demand Response“	42.400	423.000	-
NILM	6.960	315.000	-
NIALM	520	12.800	-
„Non-intrusive Load Monitoring“	1.440	20.800	-
“Non-intrusive Appliance Load Monitoring”	529	5.720	-
Dataset „Demand Side Management”	1	32	-
AND “Datensatz”			
“s.o” AND “Datensatz”	61	739	-
“s.o.” AND “Dataset“	2.560	23.200	-
“s.o.” AND “Data Set”	2.840	50.000	-
Kontext "Demand Side Management"	21.400	240.000	-
AND "Residential"			
"s.o." AND "Private"	16.000	215.000	-
"s.o." AND "House"	13.800	173.000	-

Tab. 2: Identifikation der Suchbegriffe

⁴ Zugriff auf Google Scholar (<https://scholar.google.de/>) und Google (<http://www.google.de>) am 31.03.2016.

Durch die Evaluation der Begriffe konnten wir folgenden Ausschlusskriterien bestimmen, die nicht zur Beantwortung unserer Forschungsfrage beitragen: „Gas“, „Water“ und „Traffic“ (weitere DSM-Kontexte). Zudem werden Beiträge eliminiert, die zwar die Verwendung von Datensets beschreiben, allerdings nicht das Dataset oder deren Aufzeichnungselemente definieren (z. B. [Ja00]).

Kriterien, die darüber hinaus zum Ausschluss des Datensets geführt haben sind: (a) *Keine aktuelle Verfügbarkeit des Datensets*, (b) *Keine explizite Benennung der aufgenommenen Lastmerkmale* sowie (c) *das Enthalten mindestens eines der Ausschlusskriterien*.

Auswahl der Suchquelle. Da DSM interdisziplinär ist, sind Quellen nötig, die eine breite Analyse ermöglichen. Um eine Vielzahl geeigneter Beiträge zu erhalten, haben wir die Suchmaschine Google Scholar ausgewählt. Durch das Evaluieren verschiedener Suchbegriffskombinationen haben wir folgende Begriffe hergeleitet: *Demand Side Management, Demand Response, NILM, Dataset, Data Set, residential* und *private*.

6 Analyse und Diskussion

6.1 Übersicht und Klassifikation der analysierten Datensets

Datenset	Referenz	Aufzeichnungsart	Zeitraum [in Tagen]	Häuseranzahl	Einzelanzzeichnung	Flexibilität	Trennbarkeit	Versch.-intervalle
ACS-F1	[Gi13]	Schweiz	1	N/A	X	(X)	(X)	
AMPds	[Ma13]	Vancouver, Kanada	360	1	X	(X)	(X)	
BLUED	[An12]	Pittsburgh, USA	8	1	N/A	N/A	N/A	
GREEND	[Mo14]	Österreich, Italien	90	9	X	(X)	(X)	
HES	[Zi12]	Großbritannien	30	251	X	(X)	(X)	
iAWE	[Ba13]	Indien	75	1	X	(X)	(X)	
IHEPCDS	[Li13]	Frankreich	1440	1	(X)	(X)	(X)	
OCTES	[Oc15]	Finnland, Island, Schottland	120	33	N/A	N/A	N/A	
REDD	[KJ11]	Boston, USA	3	6	X	(X)	(X)	
Sample Dataset	[Pe15]	Austin, USA	7	10	X	(X)	(X)	
Smart*	[Ba12] [Um14]	Western Massachusetts, USA	90	1	X	(X)	(X)	
UK-DALE	[Ke15]	Großbritannien	499	4	X	(X)	(X)	
DRED	[Dr15]	Niederlande	61	1	X	(X)	(X)	
ECO	[Ec15]	Schweiz	244	6	X	(X)	(X)	
Energy	[EH16]	Italien	N/A	10	X	(X)	(X)	

@Home							
REFIT	[Mu15]	Glasgow	300 +	20	X	(X)	(X)
TEALD	[Te16]	Burnaby, Kanada	2+	1	X	(X)	(X)

Tab. 3: Klassifikation bestehender DSM-Datensets

Resultierend aus den Anforderungen werden die identifizierten Datensets analysiert, wobei die Anforderungen der Dimensionierung und der Feingranularität als notwendig erachtet und nur solche Datensets berücksichtigt werden, die diese erfüllen. Die Anforderungen nach Wetterdaten, Erzeugung und Speicher, werden nicht berücksichtigt, da diese meist extern aufgezeichnet werden. Heraus sticht das Datenset PLAID [P116], welches zahlreiche reale Lasten bzw. Lastprofile enthält. Diese sind jedoch nicht über einen Zeitraum aufgezeichnet, sondern werden separat und ohne Kontext zur Verfügung gestellt. Es wird daher nicht den „klassischen Datensets“ zugeordnet.

Insgesamt wurden 1.660 Beiträge gefunden, von denen die ersten 250 Einträge berücksichtigt wurden. Durch das Eliminieren von Duplikaten, sowie einer Vor- und Rückwärtssuche konnten im letzten Schritt 17 relevante Beiträge ermittelt werden, die anhand der Anforderungsdimensionen (Abschnitt 4) klassifiziert wurden (Tab. 3). Die Anforderungen werden in den Spalten abgetragen und die Datensets in den Zeilen. Ein „X“ zeigt die explizite, ein (X) die implizite Erfüllung einer Anforderung (s. Abschnitt 6.2). Grau hinterlegte Datensets kennzeichnen neue Datensets [Mo14].

6.2 Diskussion der Ergebnisse

Ort, Zeitraum und Häuseranzahl. Informationen hinsichtlich Ort, Zeitraum und Wohneinheiten sind in allen Datensets enthalten. Ausnahme sind ACS-F1 und Energy@Home, die die Anzahl der Häuser bzw. den Aufzeichnungszeitraum nicht nennen. Aufgrund zahlreicher Unterschiede ist die Vergleichbarkeit der Datensets nicht gegeben.

Einzelaufzeichnung. Fast alle Datensets zeichnen Lasten separat auf. Lediglich OCTES und BLUED aggregieren die Lasten. IHEPCDS berücksichtigt drei verschiedene Lasttypen, sodass nur bedingt von einer ausreichenden Erfüllung gesprochen werden kann.

Trennbarkeit und Flexibilität. Kein Datenset weist explizit auf die Trenn- und Verschiebbarkeit von Lasten hin. Hier kann zwar z. B. über eine Referenztable [K109] eine Interpretation erfolgen, jedoch wäre eine explizite Kennzeichnung zur Reduzierung des Aufwandes und möglicher Fehler sinnvoll. Bei den Datensets, die die Dimension der Last-Kennzeichnung erfüllen, ist von einer bedingten Erfüllung auszugehen. Bei Lasten, die bereits die vorherige Anforderung nicht erfüllen, kann auch diese Anforderung nicht erfüllt werden, wie z. B. OCTES und BLUED.

Verschiebungsintervall. Keines der Datensets berücksichtigt Verschiebungsintervalle,

wodurch die Angabe von Nutzerpräferenzen wie z. B. in welchem Zeitraum Lasten anfallen dürfen, fehlt. Einige enthalten Zusatzinformationen wie bspw. Bewegungsprofile der Nutzer (z. B. DRED, ECO, iAWE) aus denen sich weitere Daten abgeleitet lassen.

Der Abgleich der bestehenden Datensets mit den Anforderungsdimensionen verdeutlicht, dass Informationsbedarfe bestehen, um Restriktionen aus der realen Welt (besser) berücksichtigen zu können. Vor allem die fehlende Integration von Nutzerpräferenzen in den Datensets stellt eine unzureichend erfüllte Anforderung dar.

6.3 Limitationen

Zunächst wurden im Rahmen dieses Beitrags nur öffentlich-verfügbare Datensets betrachtet, wobei auch nicht-öffentliche Datensets analysiert werden müssen. Diese können jedoch weder für die Evaluation noch für den Vergleich von Methoden genutzt werden, da diese Datensets nicht flächendeckend verfügbar sind. Darüber hinaus ist die geforderte Abbildung der Nutzerpräferenzen sehr aufwendig. Diese Realitätsnähe ist bei einigen Simulationen nicht nötig, da bspw. nur HVAC-Lasten betrachtet werden sollen und somit irrelevant ist, in welchem Intervall Aktionen abgearbeitet werden dürfen. Unabhängig davon sollten die Informationen integriert werden, um entsprechende Funktionen bei der Simulation anzubieten. Die Nutzerakzeptanz, die durch die Nicht-Einhaltung von Nutzerpräferenzen stark sinken würde, ist ein aktuelles Problem des DSM [Pa11].

Das Literatur Review sowie die entsprechenden Suchkonfigurationen basieren auf den Vorgehenshinweisen von [Br09]. Die Auswahl der Suchbegriffe, Suchmaschinen, Ausschlusskriterien sowie die Klassifikation folgen eigenen Entscheidungen, welche Limitationen haben. Es hätten z. B. weitere Schlüsselbegriffe integriert oder weitere Suchmaschinen verwendet werden können, um neue relevante Beiträge zu identifizieren.

7 Forschungsagenda

7.1 Erhebung weiterer Daten

Zur (realitätsnahen) Vervollständigung der Datensets sind neben (I) der Festlegung der Datenbedarfe und Zusatzinformationen, insbesondere (II) die Datenerhebungsmöglichkeiten von großer Bedeutung. Die Erhebungsart muss technische und logistische Details definieren, wie z. B. die Form der Datenaufzeichnung und mögliche Quellen der Daten.

Festlegung weiterer Datenbedarfe (I). Zuerst muss verdeutlicht werden, *welche verschiebbaren Lasten weitere Restriktionen haben.* Hierzu zählen Informationen für die Teil- und Verschiebbarkeit, welche jedoch nachträglich hinzugefügt werden kann.

Schwieriger zu erheben und zu integrieren, ist die *Berücksichtigung der Nutzerpräferenzen*. Es müssen, abhängig von individuellen Präferenzen, die frühesten bzw. spätesten Anfangs- und Endzeitpunkte einer Last dokumentiert werden. Bspw. möchte der Nutzer bei dem Waschvorgang einer Waschmaschine, dass dieser pünktlich zum Ende des Arbeitstages abgeschlossen ist. Daher wäre es sinnvoll, den frühestmöglichen Zeitpunkt als denjenigen zu definieren, an dem die Waschmaschine zu beladen ist. Der spätmöglichste Zeitpunkt wäre entsprechend die geplante Heimkehr abzüglich der Dauer des Waschvorgangs. Allein dieses (einfache) Beispiel zeigt, dass weitere Informationen benötigt werden, wie z. B. Dauer des Waschvorgangs. Eine Möglichkeit wäre, diese Daten sehr detailliert in das Datenset zu integrieren, was jedoch einen Mehraufwand für die Simulation und ein „aufblähen“ des Datensets bedeuten würde. Besser wäre es daher, eine aggregierte Erfassung der Informationen im Datenset, die lediglich die Zeitpunkte enthält.

Festlegung der Datenerhebung (II). Die Form der Aufzeichnung ist weniger schwierig festzulegen und sollte entsprechend den Gegebenheiten angepasst werden. Einige Datensets nutzen hierfür eine Datenbank [Mo14], andere wählen ein simples Dateiformat [Ke15]. Für eine Simulation bietet sich die Bereitstellung der Daten in einfacher Form, z. B. in Textdateien an, da diese vergleichsweise einfach mit Werkzeugen, wie Matlab, eingelesen und verarbeitet werden können. Auch Datenbanken können meist ausgelesen werden, insofern ein entsprechender Auslesevorgang zur Verfügung steht.

Komplexer gestaltet sich die Definition und Auswahl der möglichen Quellen für die Datenerhebung. Hierbei ist insbesondere der Grad der Nutzerintegration entscheidend, um bspw. keine Akzeptanzhemmnisse aufzubauen. Viele Daten sind nötig, um relevante Daten von Anfangs- und Endzeitpunkt generieren zu können. *Ein einfacher und sicherer Weg der Erhebung ist eine direkte Eingabe durch den Nutzer selbst.* Dieser kann via Endgerät, z. B. Smartphone, die benötigten Daten direkt in das System einspeisen. Dies erfordert jedoch einen hohen Aufwand für den Nutzer. Deshalb wäre es von Vorteil, die *Intervalle in denen eine Last verschoben werden kann, durch das System berechnen bzw. prognostizieren zu lassen* und an die individuellen Termine des Nutzers anzupassen. Diese könnten z. B. durch einen Online-Kalender dem System bereitgestellt werden.

Darüber hinaus muss der Nutzer eigene Präferenzen definieren, z. B: Wann trifft der Nutzer zu Hause ein? Wird ein Elektroauto bevorzugt? Neben den Terminen bedarf es zusätzlicher Informationen zur Berechnung. Es wird deutlich, dass ein komplexes Problem und zugleich ein hoher Aufwand zur Integration der Daten bestehen. Da die Simulation dieses nur bedingt leisten kann, entstehen entsprechende Forschungsbedarfe. Um die Simulation mit Daten zu versorgen, bedarf es einer aggregierten Form der Information.

7.2 Generierung von DSM-Datensets

Neben der Möglichkeit, Datensets aufzuzeichnen, können Datensets künstlich erzeugt

werden. Hierbei gibt es verschiedene Vorgehensweisen. Neben der „einfachen“ Möglichkeit, sämtliche Lasten lediglich aneinander zu hängen, wird meist versucht, einem realen Datenset möglichst nahe zu kommen. Dazu sind vor allem die „Verbrauchsprofile“ der einzelnen Lasten wichtig. Zu welchem Zeitpunkt und in welcher Form diese anfallen, bestimmt nicht ein realer Nutzer, sondern der Generator.

Der LoadProfileGenerator bedient sich z. B. einem sog. „desire model“, um das Verhalten von Verbrauchern nachzubilden. Das Modell kann im Generator nahezu beliebig angepasst werden und bietet daher ein weites Nutzer-Spektrum [Pfl16]. Ähnlich gehen auch [Ko14] vor, die mit Hilfe von typischen „load patterns and schedules“ reale Nutzungsprofile nachbilden und zeigen, dass sie (unter bestimmten Bedingungen) ein reales Lastprofil nachbilden können. Auch bei Lastprofilgeneratoren fällt auf, dass wichtige Daten nicht „mitgeloggt“ werden. Selbst wenn es hier ungleich leichter ist dies zu beheben, stellt sich die Frage, woher relevante Daten stammen.

Literaturverzeichnis

- [AHK13] Alam, M. R.; St-Hilaire, M.; Kunz, T.: Cost optimization via rescheduling in smart grids – A linear programming approach. In: Smart Energy Grid Engineering (SEGE), 2013 IEEE International Conference on, 2013.
- [Al12] Ali, S. Q.; Maqbool, S. D.; Imthias Ahamed, T. P.; Malik, N. H.: Pursuit algorithm for optimized load scheduling Power. In: Engineering and Optimization Conference (PEDCO) Melaka, Malaysia, 2012 IEEE International, 2012.
- [Al13] Ali, S. Q.; Maqbool, S. D.; Imthias Ahamed, T. p.; Malik, N. H.: Load scheduling with maximum demand and time of use pricing for microgrids. In: Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS), 2013 IEEE, 2013.
- [An12] Anderson K.; Oceau A.; Benitez D.; Carlson, D.; Rowe, A.; Berges, M.: BLUEd: a fully labeled public dataset for Event-Based Non-Intrusive load monitoring research. In: 2nd KDD Workshop on Data Mining Applications in Sustainability, 2012.
- [An12b] De Angelis, F.; Boaro, M.; Fuselli, D.; Squartini, S.; Piazza, F.; Wei, Q.; Wang, D.: Optimal task and energy scheduling in dynamic residential scenarios. In: 9th International Symposium on Neural Networks, China, 2012.
- [AYK14] Ashfaq, A.; Yingyun, S.; Khan, A. Z.: Optimization of economic dispatch problem integrated with stochastic demand side response. In: Intelligent Energy and Power Systems (IEPS), 2014 IEEE International Conference on, 2014.
- [Ba12] Barker S, Mishra A, Irwin D, Cecchet E, Shenoy P, Albrecht J (2012) Smart*: An open data set and tools for enabling research in sustainable homes. In: SustKDD
- [Ba13] Batra, N.; Gulati, M.; Singh, A.; Srivastava, M. B.: It's Different: Insights into Home Energy Consumption in India. In: Proceedings of the 5th ACM Workshop on Embedded Systems for Energy-Efficient Buildings, BuildSys'13. New York. 2013.
- [Ba14] Barta, N.; Kelly, J.; Parson, O.; Dutta, H.; Knottenbelt, W.; Rogers, A.; Singh, A.; Srivastava, M.: NILMTK: An Open Source Toolkit for Non-intrusive Load Monitoring. In: 5th international Conference on Future Energy Systems. Cambridge,

UK. 2014.

- [Ba14] Bassamzadeha, N.; Ghanem, R.; Lu, S.; Kazemtabarm S. J.: Robust scheduling of smart appliances with uncertain electricity prices in a heterogeneous population. In: *Energy and Buildings* Volume 84, December 2014, S. 537–547, 2014.
- [BN14] Bu, H.; Nygard, K. E.: Adaptive Scheduling of Smart Home Appliances Using Fuzzy Goal Programming. In: *ADAPTIVE 2014: The Sixth International Conference on Adaptive and Self-Adaptive Systems and Applications*, 2014.
- [Br09] vom Brocke, J.; Simons, A.; Niehaves, B.; Reimer, K.; Plattfaut, R.; Cleven, A.: Reconstructing the Giant: on the Importance of Rigour in Documenting the Literature Search Process. In: *Proceedings of the 17 European Conference on Information Systems (ECIS)*. Verona, Italy, 161. 2009.
- [CC11] Castillo-Cagigal, M.; Gutierrez, A.; Monasterio-Huelin, F.; Caamano-Martin, E.; Masa, D.; Jimenez-Leube, J.: A semi-distributed electric demand-side management system with PV generation for self-consumption. In: *Energy Conversion and Management* 52 (2011) 2659–2666. 2011.
- [CYN14] Chavali, P.; Yang, P.; Nehorai, A.: A Distributed Algorithm of Appliance Scheduling for Home Energy Management System. In: *Smart Grid, IEEE Transactions on* (Volume: 5 , Issue: 1), 2014.
- [Dr15] DRED: Dutch Residential Energy Dataset. <http://www.st.ewi.tudelft.nl/~akshay/dred/>. Stand: 27.09.2015.
- [Du13] Dusparic, I.; Harris, C.; Marinescu, A.; Cahill, V.; Clarke, S.: Multi-agent residential demand response based on load forecasting. In: *Technologies for Sustainability (SusTech)*, 2013 1st IEEE Conference on, 2013.
- [Ec15] ECO: Electricity Consumption & Occupancy. <http://www.vs.inf.ethz.ch/res/show.html?what=eco-data>. Stand: 27.09.2015.
- [EH16] Energy@Home: <http://www.energy-home.it/>. Stand: 01.05.2016.
- [FG12] Fathi, M.; Gholami, M.: Localized demand-side management in electric power systems. In: *Smart Grids (ICSG)*, 2012 2nd Iranian Conference on, 2012.
- [Ge15] Gerwig, C.; Behrens, D.; Knackstedt, R.; Lessing, H.: Demand Side Management in Residential Contexts – A Literature Review. In: *Informatik, Cottbus*, 2015.
- [Ge85] Gellings, C. W.: The concept of demand-side management for electric utilities. In: *Proceedings of the IEEE*, 73(10):1468–1470. 1985.
- [Gi13] Gisler, C.; Ridi, A.; Zujferey, D.; Khaled, O.; Hennebert, J.: Appliance consumption signature database and recognition test protocols. In: *8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA)*. 2013.
- [Ha92] Hart, G.W.: Nonintrusive appliance load monitoring; In: *Proceedings of the IEEE*, vol. 80, no. 12, S. 1870-1891, Dec 1992.
- [HMN12] Huang, Y.; Mao, S.; Nelms, R. M.: Smooth electric power scheduling in power distribution networks. In: *Globecom Workshops (GC Wkshps)*, 2012 IEEE, 2012.
- [Hu14] Hu, T.; Chu, X.; Zhang, W.; Liu, Y.: An optimal day-ahead dispatch strategy for deferrable loads. In: *Power System Technology, 2014 International Conference on*, 2014.

- [Ja00] Jardini, J. A.; Tahan, C. M. V.; Ahn, G. S. U.; Figueiredo, F. M.: Daily Load Profiles for Residential, Commercial and Industrial Low Voltage Consumers. In: IEEE TRANSACTIONS ON POWER DELIVERY, VOL. 15, NO. 1, JANUARY 2000.
- [Ji11] Jia, L.; Yu, Z.; Murphy-Hoye, M. C.; Pratt, A.; Piccioli, E. G.; Tong, L.: Multi-scale stochastic optimization for home energy management. In: Computational Advances in Multi-Sensor Adaptive Processing, 2011 4th IEEE International Workshop on, 2011.
- [Ke15] Kelly J (2015) UK Domestic Appliance-Level Electricity (UK-DALE) dataset. <http://www.doc.ic.ac.uk/~dk3810/data/>. Stand: 27.09.2015
- [KG11] Kim, S-J.; Giannakis, G. B.: Efficient and scalable demand response for the smart power grid. In: Computational Advances in Multi-Sensor Adaptive Processing, 2011 4th IEEE International Workshop on, 2011.
- [KJ11] Kolter, J. Z.; Johnson, M. J.: REDD: A Public Data Set for Energy Disaggregation Research. In: SustKDD Workshop on Data Mining Applications in Sustainability, 2011.
- [KK14] Kelly, J.; Knottenbelt, W.: Metadata for Energy Disaggregation. In: 2nd IEEE International Workshop on Consumer Devices and Systems (CDS 2014). Sweden. 2014.
- [KI09] Klobasa, M.: Dynamische Simulation eines Lastmanagements und Integration von Windenergie in ein Elektrizitätsnetz. In: ISI-Schriftenreihe »Innovationspotenziale«, Fraunhofer IRB Verlag. 2009.
- [Ko13] Kosek, A. M.; Costanzo, G. T.; Binder, H. W.; Gehrke, O.: An Overview of Demand Side Management Control Schemes for Buildings in Smart Grids. In: IEEE International Conference on Smart Eergy Grid Engineering. 2013.
- [Ko14] Kong, W.; Dong, Z. Y.; Chen, G.; Joa, Y.: A Rule Based Domestic Load Profile Generator for Future Smart Grid. In: Australasian Universities Power Engineering Conference, 2014.
- [LF14] Liu, P.; Fu, Y.: Construction of multiband uncertainty set for building energy management with uncertain loads and solar power. In: PES General Meeting | Conference & Exposition. 2014.
- [Li13] Lichman, M.: UCI Machine Learning Repository. <http://archive.ics.uci.edu/ml>. Stand: 27.09.2015, 2013.
- [Li15] Liu, G.-R.; Lin, P.; Fang, Y.; Lin, Y.-B.: Optimal Threshold Policy for In-home Smart Grid with Renewable Generation Integration. In: IEEE Transactions on Parallel & Distributed Systems, , no. 1, S. 1, 2014.
- [LP13] Lee, J.; Park, G.-L.: Power load distribution for wireless sensor and actuator networks in smart grid buildings. In: International Journal of Distributed Sensor Networks, 2013.
- [LPK12] Lee, J.; Park, G.-L.; Kim, H.-J.: Multithreaded Power Consumption Scheduler Based on a Genetic Algorithm. In: International Conference, FGCN 2011, Part I, 2012.
- [Ma12] Maqbool, S. D.; Imthias Ahamed, T. P., Ali , S. Q., Pazheri, F. R., malik, N. H.: Comparison of pursuit and ϵ -Greedy algorithm for load scheduling under real time pricing. In: Power and Energy, 2012 IEEE International Conference on, 2012.
- [Ma13] Makonin, S.; Popowich, F.; Bartram, L.; Gill, B.; Bajic, I. V.: AMPds: A Public Dataset for Load Disaggregation and Eco-Feedback Research. In: Proceeding of the

IEEE Conference on Electrical Power and Energy, 2013.

- [Ma14] Manasseh, E.; Ohno, S.; Mvuma, A.; Yamamoto, T.: Distributed Demand-side management with load uncertainty. In: Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, 2014.
- [MHC12] Miao, H.; Huang, X.; Chen, G.: A Genetic Evolutionary Task Scheduling Method for Energy Efficiency in Smart Homes. In: Int. Review of Electrical Engineering, 2012.
- [MM13] McNamara, P.; McLoone, S.: Hierarchical Demand Response using Dantzig-Wolfe decomposition. In: Innovative Smart Grid Technologies Europe (ISGT EUROPE), 2013 4th IEEE/PES. IEEE, S. 1–5, 2013.
- [Mo14] Monacchi, A.; Egarter, D.; Elmenreich, W.; D'Alessandro, S.; Tonello, A. M.: GREEND: An energy consumption dataset of households in Italy and Austria. In: IEEE International Conference on Smart Grid Communications, 2014.
- [Mo14b] Morimoto, N.; Fujita, Y.; Yoshida, M.; Yoshimizu, H.; Takiyama, M.; Akehi, T.; Tanaka, M.: A Power Allocation Management System Using an Algorithm for the Knapsack Problem. In: 38th Annual International Computers, Software and Applications Conference Workshops, 2014.
- [MP11] Missaoui, R.; Ploix, S.: Energy Fluxes optimization for PV integrated Building. In: PowerTech, 2011 IEEE Trondheim, 2011.
- [Mu15] Murray, D.; Coleman, M.; Hauxwell-Baldwin R.; Liao, L.; Kane, T.; Stankovic, L.; Firth, S.; Wilson, C.; Stankovic, V.: REFIT: Smart Homes & Energy Demand Reduction. 2015.
- [Oc15] OCTES: Northern Periphery Programme. <http://www.northernperiphery.eu/en/home/>. Abgerufen am 27.09.2015.
- [Pa11] Paetz, A.-G.; Becker, B.; Fichtner, W.; Schmeck, H.: SHIFTING ELECTRICITY DEMAND WITH SMART HOME TECHNOLOGIES – AN EXPERIMENTAL STUDY ON USER ACCEPTANCE. In: 30th USAEE North American Conference Online Proceedings. 2011.
- [Pa16] Parson, O.: <http://www.oliverparson.co.uk/blog>. Abgerufen om 01.05.2016.
- [Pe15] Pecanstreet (2015) Pecan Street. <http://pecanstreet.org/projects/consortium>. Stand: 27.09.2015.
- [Pf16] Pflugradt, N.: LoadProfileGenerator – LoProGen. <http://www.loadprofilegenerator.de/>. Stand: 01.05.2016.
- [Pl16] PLAID: the Plug Load Appliance Identification Dataset. <http://plaidplug.com/>. Stand: 01.05.2016.
- [SL14] Soliman, H. M.; Leon-Garcia, A.: Game-Theoretic Demand-Side Management With Storage Devices for the Future Smart Grid. In: Smart Grid, IEEE Transactions on (Volume: 5 , Issue: 3), 2014.
- [SS14] Song, L.; van der Schaar, M.: Non-stationary demand side management method for smart grids. In: Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on, 2014.
- [Te16] Makonin, S.: TEALD - The Tautological Energy AnaLog Dataset. <http://teald.org/sample/TEALD.tar.gz>. Stand: 01.05.2016.

- [Um14] UMASS: UMassTraceRepository. <http://traces.cs.umass.edu/index.php/Smart/Smart>. Stand: 27.09.2015.
- [VKM14] Verschae, R.; Kato, T.; Matsuyama, T.: A distributed coordination framework for on-line scheduling and power demand balancing of households communities. In: Control Conference (ECC), 2014 European, 2014.
- [WW02] Webster, J.; Watson, R. T.: Analyzing the past to prepare for the future: Writing a literature review. *Management Information Systems (MIS) Quarterly* 26 (2), 3. 2002.
- [Za12] Zang, C.; Qin, X.; Li, X.; Jin, X.; Che, W.: Simulated annealing based fuzzy Markov game energy management for smart grids. In: Control and Decision Conference, 2013 25th Chinese, 2012.
- [Zh13] Zhao, W.; Cooper, P.; Perez, P.; Ding, L.: Cost-Driven Residential Energy Management for Adaption of Smart Grid and Local Power Generation. 2013.
- [Zi12] Zimmermann, J.-P.; Evans, M.; Griggs, J.; King, N.; Harding, L.; Roberts, P.; Evans, C.: Household Electricity Survey A study of domestic electrical product usage. In: Household Electricity Survey and Household energy, 2012.

Erfolg innovativer E-Carsharing-Geschäftsmodelle – E-Carsharing-Geschäftsmodell-, Elektromobilitäts- und Carsharing-Akzeptanz im öffentlichen Hochschulsektor

Erik Kolek¹, Britta Reinecke² und Olexander Filevych³

Abstract: Die Universität Hildesheim erprobt ein innovatives Geschäftsmodell, welches die Vorzüge von Carsharing und Elektromobilität vereint. Im Rahmen des Tandem-Modells werden Elektroautos gleichzeitig für dienstliche und private Zwecke genutzt. Das Modell stellt somit eine Symbiose von Arbeitgeber- und Arbeitnehmerinteressen dar und kann auch auf andere öffentliche und private Hochschuleinrichtungen und Organisationen übertragen werden. Welche Faktoren dabei den Erfolg – gegeben durch die Nutzung – innovativer E-Carsharing-Geschäftsmodelle beeinflussen, wird anhand der Ergebnisse einer Online-Befragung des Personals der Universität Hildesheim dargestellt. Auf Grundlage einer Stichprobe von 219 vollständig ausgefüllten Fragebögen wurde ein Erfolgsmodell evaluiert, das die Dienstleistungs-Nutzungsabsicht aufgrund der E-Carsharing-Geschäftsmodell-, Elektromobilitäts- und Carsharing-Akzeptanz erklärt. Dieses Erfolgsmodell bildet die Basis für die Imitation und Adaption von tandembasierten Geschäftsmodellen bzw. E-Carsharing-Dienstleistungen – nicht nur im öffentlichen Hochschulsektor.

Keywords: Dienstleistungs-Nutzungsabsicht, E-Carsharing-Geschäftsmodell-Akzeptanz, Elektromobilitäts-Akzeptanz, Carsharing-Akzeptanz, Öffentlicher Hochschulsektor

1 Tandem-Modell ein innovatives E-Carsharing-Geschäftsmodell

Seit 2012 werden vier *Schaufenster Elektromobilität* vom Bund gefördert. Das Projekt *E-Autarke Zukunft* ist ein Teilprojekt des niedersächsischen Schaufensters der Metropolregion Hannover-Braunschweig-Göttingen-Wolfsburg. Im Rahmen dieses Projekts wird an der Universität Hildesheim derzeit das innovative *Tandem-Modell* [Ge14] erprobt. Die Idee ist, dass Beschäftigte ein Elektroauto täglich morgens und abends für den Weg zwischen Arbeit und Wohnort nutzen und dem Arbeitgeber im Rahmen eines Leasingvertrages einen monatlichen Festpreis von 250 Euro zahlen, der alle anfallenden Kosten (Versicherung, Kilometer, Strom, Wartung) deckt. Nach Feierabend und am Wochenende steht den Beschäftigten das Auto zur freien Verfügung.

¹ Universität Hildesheim, Institut für Betriebswirtschaft und Wirtschaftsinformatik, Abteilung Informationssysteme und Unternehmensmodellierung, Universitätsplatz 1, 31141 Hildesheim, Deutschland, erik.kolek@uni-hildesheim.de

² Universität Hildesheim, Institut für Betriebswirtschaft und Wirtschaftsinformatik, Abteilung Betriebswirtschaft und Operations Research, Universitätsplatz 1, 31141 Hildesheim, Deutschland, britta.reinecke@uni-hildesheim.de

³ Universität Hildesheim, Institut für Betriebswirtschaft und Wirtschaftsinformatik, Abteilung Betriebswirtschaft und Operations Research, Universitätsplatz 1, 31141 Hildesheim, Deutschland, olexander.filevych@uni-hildesheim.de

In einem festen Zeitfenster während der Arbeitszeit (09:00 bis 16:00 Uhr) steht das Auto anderen Beschäftigten für Dienstreisen zur Verfügung. Ein Webportal bietet Transparenz über die Belegzeiten und ermöglicht eine bequeme Reservierung. Ein Solarcarport am Hauptcampus der Universität Hildesheim versorgt die Elektroautos mit Strom.

Das innovative E-Carsharing-Geschäftsmodell am Beispiel des Hildesheimer *Tandem-Modells* umfasst geplante Schlüsselpartner, Schlüsselaktivitäten, Schlüsselressourcen, Wertversprechen, Kundenbeziehungen, Kanäle, Kundensegmente, Kosten und Erlöse [Ge14]. Durch die Struktur des innovativen *Tandem-Modells* [Ge14] werden bereits einige Nutzungshemmnisse entkräftet: Beschäftigte müssen keinen vergleichsweise hohen Kaufpreis zahlen, der Leasinggeber trägt im Rahmen des Leasingvertrages das Risiko der Batterielebensdauer und eine geeignete Ladeinfrastruktur wird, wie vielfach von Nutzern von Elektroautos gewünscht, am Arbeitsplatz bereitgestellt. Darüber hinaus wird seitens des Arbeitgebers garantiert, dass die Nutzer nach Feierabend einen Ladezustand vorfinden, der eine unproblematische Heimfahrt garantiert.

Zunächst werden verwandte Arbeiten zum E-Carsharing-Geschäftsmodellerfolg aufgezeigt (Abschnitt 2). Danach erfolgt die Beschreibung der methodischen Vorgehensweise (Abschnitt 3). Danach werden Hypothesen entwickelt und getestet (Abschnitt 4). Der Beitrag schließt mit einer Diskussion (Abschnitt 5).

2 Verwandte Arbeiten zum E-Carsharing-Geschäftsmodellerfolg

Nach vom Brocke et al. [Br09] ist es wichtig geeignete Schlüsselwörter zu identifizieren. Für diesen Zweck wurden relevante Schlüsselwörter in interdisziplinären Datenbanken und Datenbanken mit Wirtschaftsinformatikschwerpunkt [KW06] getestet. Schlüsselwörter mit den größten Suchtreffern wurden beibehalten. Auf Basis der identifizierten Schlüsselwörter wurden Suchphrasen zum E-Carsharing-Geschäftsmodellerfolg definiert und in den Datenbanken JSTOR, Google Scholar, Informs und AISeL angewandt für die Literatursuche (in Anlehnung an [Br09, K114, Sc16]).

Google Scholar/ Informs	("E-Carsharing*" OR "E-Car-Sharing*" OR "Electric-Car-Sharing*") AND ("Business Model*" OR "Business Plan*" OR "Business Concept*") AND ("Success*") AND ("Acceptance*") NOT ("Sustainability*")
AISeL	"E-Carsharing" OR "E-Car-Sharing" OR "Electric-Car-Sharing" AND "Business Model" OR "Business Plan" OR "Business Concept" AND "Success" AND "Acceptance" NOT "Sustainability"
JSTOR	(((((("E-Carsharing*") OR ("E-Car-Sharing*")) OR ("Electric-Car-Sharing*")) AND ("Business Model*")) OR ("Business Plan*")) OR ("Business Concept*")) AND ("Success*")) AND ("Acceptance*"))

Tab. 1: Definition von Suchphrasen für die Literatursuche (in Anlehnung an [Sc16])

In einem ersten Schritt wurden 975 Forschungsartikel mit Hilfe der Anwendung der Suchphrasen für die Literatursuche gefunden. Davon sind 763 Artikel interdisziplinär

(JSTOR: 725; Google Scholar: 38) und 212 Artikel besitzen einen Bezug zur Wirtschaftsinformatik (Informs: 174; AISEL: 38). Durch das Ausschlusskriterium, dass nur englischsprachige Forschungsartikel ohne Bezug zur Nachhaltigkeit relevant sind, wurden die Suchtreffer auf 792 Artikel reduziert. Ebenso wurden Redundanzen aussortiert. Die Evaluation der Forschungsartikel wurde über die Analyse der Titel, Abstracts und Schlüsselwörter sowie über den Volltext realisiert. Zusätzlich wurde – wenn möglich auf Basis der Datenbanken – eine direkte Suche in den Titeln und Abstracts entsprechend der Suchphrasen für die Literatursuche angewandt z. B. in AISEL. Alternativ wurde – zur Überprüfung der Evaluation der Forschungsartikel – eine vereinfachte Suchphrase für die Literatursuche genutzt z. B. (((“E-Carsharing*”) AND (“Business Model*”) AND (“Success*”)) in JSTOR. Forschungsartikel ohne Relevanz hinsichtlich des Forschungshintergrunds wurden ausgeschlossen.

Das Ergebnis der Literatursuche besteht lediglich aus 3 relevanten Forschungsartikeln auf Basis der Anwendung der Suchphrasen für die Literatursuche in den Datenbanken. Daher erfolgte in einem zweiten Schritt die Rückwärtssuche nach relevanten Referenzen in den Forschungsartikeln und in einem dritten Schritt die Vorwärtssuche nach relevanten Referenzen in den Datenbanken. Insgesamt wurden 9 verwandte Arbeiten zum E-Carsharing-Geschäftsmodellerfolg gefunden (vgl. Abb. 1).

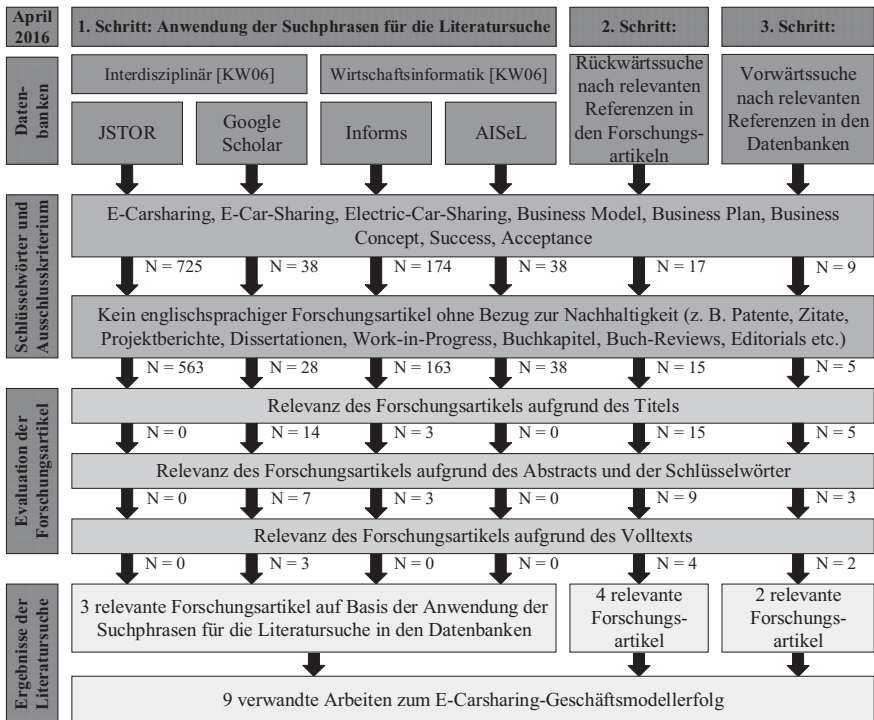


Abb. 1: Ergebnisse der Literatursuche (in Anlehnung an [Sc16])

Die Ergebnisse der Literatursuche werden systematisiert (in Anlehnung an [Sc16]) mit einer Konzeptmatrix, die aufgeliedert ist in fünf Dimensionen (vgl. Tab. 2).

E-Carsharing. Setzt sich der Artikel mit E-Carsharing auseinander?

Geschäftsmodell. Betrachtet der Artikel Carsharing-Geschäftsmodelle?

Erfolg. Wird der Erfolg von Carsharing-Geschäftsmodellen im Artikel untersucht?

Akzeptanz. Wird die Akzeptanz (von Kunden, Konsumenten, Nachfragern etc.) gegenüber der Nutzung von Carsharing-Geschäftsmodellen im Artikel erforscht?

Methode. Welche gängige Forschungsmethode wird im Artikel zur Erzielung der Ergebnisse angewandt z. B. Interviews, Conjoint-Analyse, SGM/PLS-Methode etc.?

Referenz	E-Carsharing	Geschäftsmodell	Erfolg	Akzeptanz	Methode
[Le11]	X		X		SERVQUAL
Ein E-Carsharing-Dienstleistungssystem bestehend aus Buchung und Rückgabe von Autos, externen Kanälen für Zahlungen, den Autos für die Sharing-Dienstleistung und mobilen Nutzern mit Mitgliedschaft in der E-Carsharing-Dienstleistungs-Community.					
[Ri13]			X		Quantitativ
Quantitativer Carsharing-Ansatz bestehend aus einem Optimierungsmodell und Entscheidungsunterstützungssystem. Dieses integriert das Modell und bewirkt so in einer Stadt eine optimale Verteilung zwischen den Carsharing-Standorten und der Nachfrage.					
[SB13]	X	X	X	X	Interviews
Erfolgsfaktoren für das Carsharing mit Elektrofahrzeugen: Akzeptanzgründe wie Eigenschaften der Umwelt, Adoptierenden und Innovation, vorgeschlagene Akzeptanzgründe aus der Literatur wie verfügbare Technologie und Wettbewerb, verschiedene Stakeholder-Perspektiven, Herausforderungen und Möglichkeiten.					
[Wa14]			X		Quantitativ
Die Nutzung von Big Data in einem Entscheidungsunterstützungssystem für standort-basiertes Carsharing und zur Prognose der exakt benötigten Autos in einem bestimmten Stadtgebiet sowie zur Bestimmung der Attraktivität eines bestimmten Standorts.					
[Hi15]		X			Conjoint-Analyse
Eigenschaften eines Carsharing-Angebots in absteigender Wichtigkeit aus Kundensicht: Reservierung (online oder per App), Kooperationsfähigkeit (Kundenkonto für mehrere Carsharing-Angebote in mehreren Städten), Onlinekonto (mit Informationsübersicht über Reisen und Kosten), Autostandort (feste Standorte), Gebühren und Abrechnung (automatische nutzungsbasierte Abrechnung über Zeit und Kilometer), Autozugang (per Smartphone/ Mitgliedschaftskarte) und Anreize (vorsichtiges Fahren wird belohnt).					
[Tr15]		X	X	X	SGM/PLS

Referenz	E-Carsharing	Geschäftsmodell	Erfolg	Akzeptanz	Methode
	Die Nutzungsabsicht des Peer-to-Peer-Carsharing (Privatautos werden geteilt) wird positiv durch den sozialen Einfluss hinsichtlich der Nutzung von P2P-Carsharing als auch durch den wahrgenommenen Nutzen von P2P-Carsharing sowie negativ durch das wahrgenommene Risiko sein eigenes Auto in P2P-Carsharing anzubieten beeinflusst.				
[Wal5a]		X	X		Quantitativ
	Ein Entscheidungsunterstützungssystem für Carsharing-Anbieter zur Bestimmung von nicht lukrativen und lukrativen Gebieten aufgrund der Nachfrage zur Gewinnerhöhung.				
[Wal5b]		X	X	X	Quantitativ
	Ein Entscheidungsunterstützungssystem für Carsharing-Anbieter das durchschnittliche Leerlaufzeiten pro Auto reduziert und die Autonutzung erhöht durch das Angebot von nutzerbasierten Kundenanreizen für den Standortwechsel.				

Tab. 2: Ergebnisse der verwandten Arbeiten zum E-Carsharing-Geschäftsmodellerfolg

Ein Artikel beschäftigte sich inhaltlich zwar mit Elektrofahrzeugen, Geschäftsmodellen und Elektromobilität vernachlässigte jedoch im Vergleich zu den anderen Forschungsartikeln (vgl. Tab. 2) gänzlich das Carsharing mit oder ohne Elektroautos [Br12]. Dieser Artikel wurde daher im Nachhinein nicht als verwandte Arbeit zum E-Carsharing-Geschäftsmodellerfolg betrachtet.

Es existieren zu wenige empirische Arbeiten, die den Erfolg von E-Carsharing-Geschäftsmodellen erkenntnisreich – aufgrund angewandter Methoden – thematisieren. Um diese aktuell existierende Forschungslücke explorativ unter Anwendung der Strukturgleichungsmodellierung (SGM) und Partial Least Squares (PLS)-Methode zu schließen, werden vier Forschungsfragen (F1-F4) abgeleitet, welche bisher nicht ausreichend genug im Kontext des öffentlichen Hochschulsektors beantwortet wurden:

1. *Welchen Einfluss nimmt die E-Carsharing-Geschäftsmodell-Akzeptanz auf die Absicht der Nutzer, die Dienstleistung zu nutzen?*
2. *Wie beeinflusst die Elektromobilitäts-Akzeptanz die Nutzungsabsicht hinsichtlich der E-Carsharing-Dienstleistung?*
3. *Wie wird die Nutzungsabsicht durch die Carsharing-Akzeptanz beeinflusst?*
4. *Welchen Einfluss haben jeweils (F1-F3) die Mobilitätsmerkmale der Nutzer?*

3 Methodische Vorgehensweise

3.1 Datenerhebung

Im Rahmen des Projektes *E-Autarke Zukunft* wurde das Personal der Universität Hildesheim vor der Einführung des Tandem-Modells [Ge14] im Jahr 2013 mit Hilfe einer anonymen Online-Befragung zu ihrer persönlichen Einstellung gegenüber Elektromobilität und Carsharing sowie dem geplanten Dienstleistungsangebot befragt. Für diese Meinungsumfrage hinsichtlich der F1-F4 wurden Aussagen formuliert, die die Teilnehmenden auf einer 5-stufigen Likert-Skala (1 = *stimme voll und ganz zu*; 3 = *neutral*; 5 = *stimme gar nicht zu*) bewerten sollten. Mit den Werten dazwischen konnten die Teilnehmenden ihre Meinung abstufen. Ein entscheidendes Qualitätskriterium einer Meinungsfrage umgesetzt als Personalumfrage ist eine relativ große Stichprobe, die sich aus Personen der verschiedensten Unternehmensbereiche zusammensetzt. Dies konnte im Rahmen der Umfrage realisiert werden: Alle 714 Beschäftigten der Universität Hildesheim (Grundgesamtheit N = 714) wurden per E-Mail angeschrieben und um Teilnahme gebeten, die Rücklaufquote lag bei 30,67% (Stichprobe n = 219). Personen aus den verschiedensten Fachbereichen und Hierarchieebenen haben teilgenommen.

Die Stichprobe besteht etwa zu gleichen Anteilen aus männlichen und weiblichen Teilnehmern. Das Durchschnittsalter der Befragten liegt bei 40 Jahren. 59% der Teilnehmer arbeiten Vollzeit an der Universität Hildesheim, die übrigen in Teilzeit. Die durchschnittliche Haushaltsgröße der Teilnehmer liegt bei 2,5 Personen. Pro Haushalt ist durchschnittlich ein Pkw verfügbar. Die durchschnittliche Distanz zwischen Wohnort und Arbeitsplatz beträgt 35 Kilometer. Dies liegt daran, dass die Beschäftigten der Universität Hildesheim nicht in direkter Nähe wohnen, sondern auch aus größeren Städten in der Umgebung anreisen z. B. aus Hannover. Für den täglichen Weg zur Arbeit werden die Verkehrsmittel in dieser absteigenden Reihenfolge genutzt: Eigener Pkw, Fahrrad, öffentliche Verkehrsmittel, zu Fuß, Motorrad/Roller und Fahrgemeinschaften.

3.2 Datenanalyse

Die Strukturgleichungsmodellierung (SGM) erfährt eine wachsende Wertschätzung in der Wirtschaftsinformatik sowie in angrenzenden Wissenschaftsgebieten [UA10]. In diesem Forschungsbeitrag wird für die Auswertung der Umfragedaten die Partial Least Squares (PLS)-Methode angewandt, um ein Strukturgleichungsmodell und darin enthaltene Messmodelle aufgrund der Daten zu erstellen und zu evaluieren. Die PLS-Methode ermöglicht es, die in der Meinungsumfrage reflektiv gemessenen Indikatoren mit indirekt gemessenen Variablen über deren Messmodelle zu verbinden. Bei reflektiven Indikatoren besteht eine hohe Korrelation mit den verbundenen Variablen, wodurch eine Veränderung der Variablen zu einer Veränderung der Indikatoren führt. Bei formativen Indikatoren ist dies nicht der Fall, da diese die latente Variable verursachen, wodurch die Variable lediglich durch alle Indikatoren gleichzeitig

veränderbar ist, jedoch selbst keinen Einfluss auf die Indikatoren nimmt. Einzelne Indikatoren einer Variablen – unabhängig davon ob die Indikatoren reflektiv oder formativ gemessen wurden – werden als äußere Modelle bzw. Messmodelle bezeichnet. Das innere Modell wird auch als Strukturgleichungsmodell bezeichnet und verbindet die Messmodelle der unabhängigen und abhängigen Variablen [Ch98b].

Die PLS-Ergebnisse wurden mit SmartPLS 2.0.M3 [RWW05] erzeugt und durch das Bestimmtheitsmaß (R^2) sowie durch weitere Qualitätskriterien evaluiert (vgl. Tab. 3). Dazu wurde die PLS- und Bootstrapping-Methode angewandt, um die entwickelten und im Strukturgleichungsmodell abgebildeten Hypothesen auf deren Gültigkeit und Signifikanz hin zu überprüfen [Ch98b]. Im Rahmen der Datenanalyse sind alle vollständig ausgefüllten Fragebögen (Stichprobe $n = 219$) genutzt worden.

4 Erfolgsmodell innovativer E-Carsharing-Geschäftsmodelle

4.1 Hypothesenentwicklung

Die Geschäftsmodell-, Elektromobilitäts- und Carsharing-Akzeptanz können einen Einfluss auf die Nutzung innovativer E-Carsharing-Dienstleistungen nehmen. Es wird angenommen, dass die Nutzungsabsicht potentieller Kunden letztlich auch zur Nutzung innovativer E-Carsharing-Dienstleistungen führt. Daher ist für den Erfolg des innovativen *Tandem-Modells* [Ge14] an der Universität Hildesheim die wahrgenommene Dienstleistungs-Nutzungsabsicht potentieller Kunden bzw. Nutzer relevant. Diese Absicht bildet sich durch die persönliche Wahrnehmung des Dienstleistungsangebots – vor allem durch das von Kunden bzw. Nutzern wahrgenommene Wertversprechen, das als Mobilitätsversprechen formuliert ist. Also ob das im Dienstleistungsangebot beschriebene E-Carsharing-Geschäftsmodell aus Kundensicht z. B. für den Preis bzw. die fixe Monatsrate zu teuer ist, als Pkw-Ersatz, Dienststreiseerleichterung, für die Dienstreisen (z. B. Campuswechsel), Universität Hildesheim und private Zwecke geeignet ist. Die unterschiedliche Wahrnehmung von potentiellen Kunden bzw. Nutzern äußert sich ebenfalls in deren individuell ausgeprägten Mobilitätsbedürfnissen – nicht nur im öffentlichen Hochschulsektor, daher werden verschiedene Aspekte wie Häufigkeit und zurückgelegte Entfernungen beruflicher Dienstreisen, Geschlecht und Arbeitszeit als Nutzer-Mobilitäts-Merkmale (NMM) zusammengefasst betrachtet.

Aufgrund der Ergebnisse der Personalbefragung an der Universität Hildesheim sind sechs *Hypothesen (H1-H6)* abgeleitet und in einem Erfolgsmodell innovativer E-Carsharing-Geschäftsmodelle als einseitige Pfadkorrelationen dargestellt worden (vgl. Abb. 2): *Die E-Carsharing-Geschäftsmodell-Akzeptanz (ECGA) (H1), Elektromobilitäts-Akzeptanz (EMA) (H2) und die Carsharing-Akzeptanz (CSA) (H3) haben jeweils einen Einfluss auf die Dienstleistungs-Nutzungsabsicht (DLNA). Die Nutzer-Mobilitäts-Merkmale (NMM) beeinflussen dabei die Beziehung zwischen ECGA und DLNA (H4), EMA und DLNA (H5) sowie CSA und DLNA (H6).*

4.2 Hypothesentest

Mit SmartPLS 2.0.M3 [RWW05] wurde die Partial Least Squares (PLS)-Methode angewendet. Das Ergebnis ist ein Strukturgleichungsmodell, welches den Erfolg innovativer E-Carsharing-Geschäftsmodelle im öffentlichen Hochschulsektor erklärt. Die Effektgröße nach Cohen [Co88] ist in Abb. 2 durch entsprechend fett markierte Pfeile symbolisiert (vgl. Tab. 4), d. h. ECGA und CSA haben den größten Effekt auf die Dienstleistungs-Nutzungsabsicht (DLNA).

Das Erfolgsmodell innovativer E-Carsharing-Geschäftsmodelle besteht aus einer Moderationsvariablen Nutzer-Mobilitäts-Merkmale (NMM) und dem Zusammenwirken der unabhängigen Variablen E-Carsharing-Geschäftsmodell-Akzeptanz (ECGA), Elektromobilitäts-Akzeptanz (EMA) und Carsharing-Akzeptanz (CSA) unter moderierendem Einfluss von NMM auf die abhängige Variable bzw. Zielvariable Dienstleistungs-Nutzungsabsicht (DLNA). Diese erklärt 48,8% der Modellvarianz aufgrund des Bestimmtheitsmaßes ($R^2 = 0,488$). Letzteres ist bei ECGA, EMA, CSA und NMM nicht verfügbar, da es sich um das äußere Modell des Strukturgleichungsmodells handelt [Ch98b]. Alle Pfadkoeffizienten zwischen den Variablen ECGA, EMA und CSA in Verbindung mit DLNA liegen im Annahmehereich, da diese größer sind als 0,200 und daher als bedeutsam gelten. Dagegen sind alle Pfadkoeffizienten der Moderationsvariablen NMM im Falle der Beziehungen zwischen ECGA und DLNA, EMA und DLNA sowie CSA und DLNA nicht als bedeutsam zu betrachten. Hier liegen die Pfadkoeffizienten im Verwerfungsbereich, da diese weder größer als 0,200 noch kleiner als -0,200 sind [Ch98a]. Der gemessene Pfadkoeffizient in Höhe von -0,176** wurde trotzdem akzeptiert, da hier der Pfadkoeffizient kleiner als -0,100 ist [Lo89] sowie sehr signifikante T-Werte vorliegen (vgl. Tab. 4).

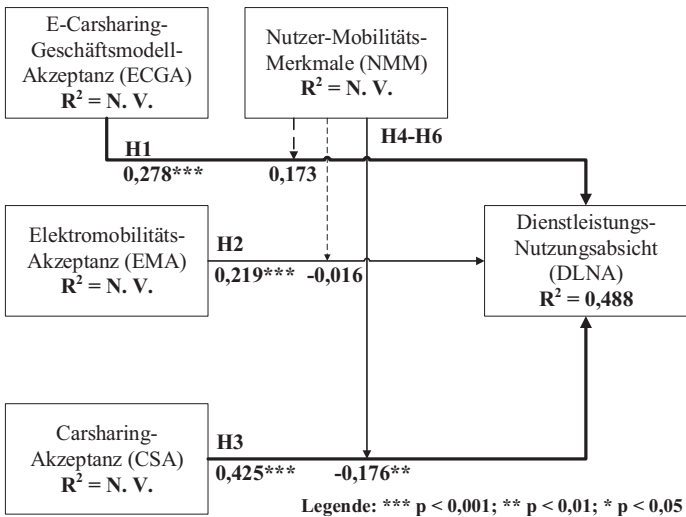


Abb. 2: Erfolgsmodell innovativer E-Carsharing-Geschäftsmodelle

Die Reliabilität und Validität der Messmodelle und des Strukturgleichungsmodells wird aufgrund verschiedener Qualitätskriterien beurteilt (siehe Tab. 3). Die Variable EMA erweist sich aufgrund der Faktor-Reliabilität (REL) und Cronbachs Alpha (CA) [Cr51] noch als akzeptabel für explorative Forschungsergebnisse, da die Werte im Vergleich zu den anderen Variablen nahe 0,6 sind [BY88]. ECGA besitzt das qualitativ hochwertigste Messmodell. Bei den Variablen NMM, DLNA und CSA werden aufgrund der explorativen methodischen Vorgehensweise keine Werte in Höhe von 0,6 zur Bestimmung einer akzeptierbaren Reliabilität erreicht [BY88]. Die durchschnittlich erfasste Varianz (DEV) ist bei allen Variablen kleiner als 0,5. Dagegen ist bei DLNA das R^2 gleich 0,488 und größer als 0,33. Das R^2 von DLNA gilt damit als durchschnittlich. Ein durchschnittliches R^2 meint, dass dieses weder als schwach ($R^2 \geq 0,19$) noch als substantiell ($R^2 \geq 0,67$) zu betrachten ist. Im Falle von DLNA bedeutet dies, dass diese abhängige Variable durchschnittlich durch die unabhängigen Variablen ECGA, EMA und CSA erklärt ist. Beim Erfolgsmodell innovativer E-Carsharing-Geschäftsmodelle handelt es sich demnach um kein schwaches Strukturgleichungsmodell mit geringem Erklärungscharakter. Bei DLNA ist Q^2 ebenfalls größer als 0,0 und daher besitzt das Erfolgsmodell zusätzlich einen prognostizierenden Charakter [Ch98b].

Variablen	Qualitätskriterien der Messmodelle			Qualitätskriterien des Strukturgleichungsmodells		
	Art	REL $\geq 0,6$ [BY88]	CA [Cr51] $\geq 0,6$ [BY88]	DEV $\geq 0,5$ [BY88]	$R^2 \geq 0,19$ [Ch98b]	$Q^2 \geq 0,0$ [Ch98b]
NMM	R	0,162	0,191	0,169	N. V.	N. V.
DLNA	R	0,142	0,255	0,326	0,488	0,087
ECGA	R	0,875	0,853	0,392	N. V.	N. V.
EMA	R	0,662	0,559	0,340	N. V.	N. V.
CSA	R	0,137	0,098	0,350	N. V.	N. V.
R = reflektiv; N. V. = Nicht verfügbar, aufgrund des äußeren Modells [Ch98b]						

Tab. 3: Reliabilität und Validität der Messmodelle und des Strukturgleichungsmodells

Grundsätzlich sollten schlecht ladende Indikatoren (Indikatorladung $> -0,7$ und $< 0,7$) aus den Messmodellen eines Strukturgleichungsmodells entfernt werden [Ch98a, Ch98b]. Da die Qualität des Erfolgsmodells innovativer E-Carsharing-Geschäftsmodelle nur unwesentlich verbessert wird, wenn einzelne Indikatoren einzelner Variablen entfernt werden, wurde entschieden – trotz teilweise nicht ganz ausreichend ladender Indikatoren – diese beizubehalten. Zu beachten ist für dieses explorativ entwickelte Erfolgsmodell, dass akzeptable Indikatoren (vgl. Tab. 5) lediglich einen Wert größer als 0,4 oder kleiner als -0,4 erreichen sollten [Hu99].

Um Aussagen über die Gültigkeit und Signifikanz der Hypothesen (H1-H6) treffen zu können, müssen die Pfadkoeffizienten und T-Werte überprüft werden. Um zuverlässige T-Werte zu bestimmen, wurde die Bootstrapping-Methode mit 219 Fällen und 5.000 gezogenen Stichproben aus den 219 Fällen durchgeführt [Ha11]. Die Pfadkoeffizienten wurden ebenfalls in SmartPLS 2.0.M3 mit 300 Iterationen berechnet [RWW05]. Die

Hypothesen H1-H3 wurden sehr signifikant bestätigt und weisen einen kleinen bis mittleren positiven Einfluss auf DLNA auf, da hier die Pfadkoeffizienten positiv und größer als 0,200 sind [Ch98a] sowie der Effekt klein bei EMA ($f^2 \geq 0,02$) und mittel bei ECGA und CSA ($f^2 \geq 0,35$) bestätigt wurde [Co88]. Die Pfade hinsichtlich der sehr signifikant bestätigten Hypothesen H1-H3 weisen sehr hohe T-Werte auf (***) $p < 0,001$; *** T-Wert $> 3,131$). H4-H5 wurden durch den Test nicht bestätigt aufgrund der in SmartPLS 2.0.M3 [RWW05] berechneten Pfadkoeffizienten und T-Werte. H6 wurde dagegen ebenfalls signifikant bestätigt. Die NMM haben einen kleinen ($f^2 \geq 0,02$) [Co88] negativen Einfluss auf die Beziehung zwischen CSA und DLNA. Die NMM können also diese positive Beziehung etwas verschlechtern. Die Variablen ECGA, EMA und CSA haben demnach jeweils einen Einfluss auf die DLNA. Hierbei ist zu beachten, dass die NMM einen kleinen negativen Einfluss ausüben auf die mittelstarke positive Beziehung zwischen CSA und DLNA. Die beschriebenen Hypothesentestergebnisse sind in Tab. 4 zusammengefasst und werden im Abschnitt 5.2 diskutiert.

Hypothese	Pfadkoeffizient * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$	T-Wert * $> 1,653$; ** $> 2,345$; *** $> 3,131$	Effektgröße [Co88]	Effekt $\geq 0,02$ [Co88]	Validiert
H1	0,278***	4,606***	0,166	Mittel	Ja
H2	0,219***	3,831***	0,090	Klein	Ja
H3	0,425***	5,428***	0,160	Mittel	Ja
H4	0,173	1,163	0,047	Klein	Nein
H5	-0,016	0,098	0,000	Kein Effekt	Nein
H6	-0,176**	2,504**	0,047	Klein	Ja

Tab. 4: Hypothesentestergebnisse

5 Diskussion der Limitationen und Implikationen

5.1 Limitationen der methodischen Vorgehensweise

In diesem Forschungsbeitrag wurde als methodische Vorgehensweise eine empirische, rein quantitative, explorative E-Carsharing-Studie gewählt, um neues Wissen über den Untersuchungsgegenstand „Erfolg innovativer E-Carsharing-Geschäftsmodelle im öffentlichen Hochschulsektor“ zu erzeugen. Über diesen Gegenstand sind bisher nur wenige Erkenntnisse bekannt (vgl. Tab. 2). Darüber hinaus bietet dieser viele wichtige Ansätze für die Gestaltung von tandembasierten Geschäftsmodellen [Ge14] bzw. E-Carsharing-Dienstleistungen – auch in anderen Sektoren wie dem privaten Sektor.

Die Online-Befragung wurde nur mit dem Personal der Universität Hildesheim durchgeführt. Bezogen auf den potentiellen Nutzerkreis stellt dies keinen Nachteil dar, da das E-Carsharing-Geschäftsmodell zunächst nur an der Universität Hildesheim

eingeführt werden sollte. Das *Tandem-Modell* [Ge14] kann von anderen öffentlichen und privaten Hochschuleinrichtungen und Organisationen imitiert und adaptiert werden.

5.2 Implikationen zum Erfolg innovativer E-Carsharing-Geschäftsmodelle

Die Dienstleistungs-Nutzungsabsicht (DLNA) als eine Voraussetzung für den Erfolg innovativer E-Carsharing-Geschäftsmodelle ist durch seine Indikatoren und deren Ladungen beschrieben (vgl. Tab. 5). Das *Tandem-Modell* [Ge14] würde für Dienstreisen und zum Campuswechsel genutzt werden. Die Beschäftigten der Universität Hildesheim sind bereit einen Festpreis zu bezahlen (250 Euro pro Monat), wenn das *Tandem-Modell* [Ge14] den eigenen Pkw ersetzt. Dienstreisen würden durch das *Tandem-Modell* [Ge14] erleichtert werden. Das *Tandem-Modell* [Ge14] wäre genauso für andere Branchen interessant – nicht nur für eine Universität. Die Nutzung des *Tandem-Modells* [Ge14] wird durch Privatpersonen als flexibel genug eingeschätzt.

Die E-Carsharing-Geschäftsmodell-Akzeptanz (ECGA) hat einen mittleren positiven Einfluss auf die Dienstleistungs-Nutzungsabsicht (DLNA) (H1/F1). Besonders relevante Einflussfaktoren sind: längere Batteriereichweite, Ausbau der Ladestationen (mehr Standorte), Lademöglichkeit zu Hause, geringere Akkuladezeit, Privilegien für Elektrofahrzeuge wie z. B. die Nutzung von Busspuren, kostenlose Parkplätze, mehr Wissen über Elektrofahrzeuge, weniger Kosten, jederzeitige, garantierte Verfügbarkeit eines Elektrofahrzeuges für Diensfahrten, mehr Flexibilität für außerplanmäßige Ereignisse im Rahmen der Privatnutzung und kabellose Lademöglichkeit.

Die Nutzungsabsicht von E-Carsharing-Dienstleistungen ist in kleinem Maße positiv durch die Elektromobilitäts-Akzeptanz (EMA) beeinflusst (H2/F2). Elektroautos werden als eine gute Alternative zum traditionellen Pkw angesehen und eignen sich gut für alltägliche Wege und auch für Wege in der Freizeit. Durch die steigenden Ölpreise und dringend erforderlichen Einsparungen von CO₂-Emissionen werden Elektrofahrzeuge zukünftig eine immer größere Rolle spielen. Es besteht persönliches Interesse an dem Thema Elektromobilität.

Die Carsharing-Akzeptanz (CSA) nimmt einen positiven Einfluss auf die Dienstleistungs-Nutzungsabsicht (DLNA) (H3/F3). Carsharing kann für jeden eine gute Alternative zu einem eigenen Auto sein und bietet sich nicht nur für Personen an, die selten ein Auto benötigen. Carsharing ist nicht zu teuer. Carsharing kann Raum für Spontaneität und Flexibilität lassen und wäre nicht nur als Ergänzung zum eigenen Pkw denkbar. Die Buchung eines Carsharing-Autos wird sich nicht als umständlich und zeitaufwendig vorgestellt. Es existiert ein großes Interesse daran, Carsharing mal auszuprobieren.

Die Nutzer-Mobilitäts-Merkmale (NMM) haben keinen bzw. lediglich einen kleinen negativen Einfluss im Falle der Beziehung zwischen der Carsharing-Akzeptanz (CSA) und Dienstleistungs-Nutzungsabsicht (DLNA) (H4-6/F4). Hier spielen die Häufigkeit und der Radius beruflicher Dienstreisen, die Häufigkeit einer Fahrt von mehr als 100

Kilometern (km), das Geschlecht und die Arbeitszeit eine wichtige Rolle.

Die sektoren-spezifische Gültigkeit des Erfolgsmodells innovativer E-Carsharing-Geschäftsmodelle beschränkt sich inhaltlich auf den öffentlichen Hochschulsektor wie z. B. auf Universitäten und Fachhochschulen. Das E-Carsharing-Geschäftsmodell *Tandem-Modell* [Ge14] eignet sich auch für andere öffentliche Hochschuleinrichtungen, da die Personalstrukturen von Hochschulen in Deutschland vergleichbar sind und daher von einer Übertragbarkeit ausgegangen werden kann. Dennoch können ebenfalls private Hochschuleinrichtungen und Unternehmensorganisationen davon profitieren und entsprechende E-Carsharing-Geschäftsmodelle erfolgreich einführen und anpassen.

Literaturverzeichnis

- [BY88] Bagozzi, R. P.; Yi, Y.: On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16/01, S. 74-94, 1988.
- [Br09] vom Brocke, J.; Simons, A.; Niehaves, B.; Reimer, K.; Plattfaut, R.; Cleven, A.: Reconstructing the Giant: on the Importance of Rigour in Documenting the Literature Search Process. In *Proceedings of the 17 European Conference on Information Systems (ECIS)*. Verona, Italy, 161, 2009.
- [Br12] Brandt, T.; Wagner, S.; Neumann, D.: Road to 2020: IS-Supported Business Models for Electric Mobility and Electrical Energy Markets. In *Proceedings of the Thirty Third International Conference on Information Systems*, Orlando, 2012.
- [Ch98a] Chin, W. W.: Issues and Opinion on Structural Equation Modeling. *Management Information Systems Quarterly*, 22/01, S. 7-16, 1998.
- [Ch98b] Chin, W. W.: The partial least squares approach to structural equation modeling. In (Marcoulides, G. A., Hrsg.): *Modern methods for business research*. Quantitative methodology series. Lawrence Erlbaum, Mahwah, NJ, USA, S. 295-336, 1998.
- [Co88] Cohen, J.: *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Routledge Academic, 1988.
- [Cr51] Cronbach, L. J.: Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, S. 297-334, 1951.
- [Ge14] Gerwig, C.; Behrens, D.; Knackstedt, R.; Lessing, H.: Innovatives eCarSharing-Konzept am Beispiel des Hildesheimer Tandem-Modells. In *Informatik*, Stuttgart, S. 1617-1629, 2014.
- [Ha11] Hair, J. F.; Ringle, C. M.; Sarstedt, M.: PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, 19/02, S. 139-151, 2011.
- [Hi15] Hildebrandt, B.; Hanelt, A.; Piccinini, E.; Kolbe, L. M.; Nierobisch, T.: The Value of IS in Business Model Innovation for Sustainable Mobility Services – The Case of Carsharing. In (Thomas, O.; Teuteberg, F., Hrsg.): *Proceedings der 12. Internationalen Tagung Wirtschaftsinformatik (WI 2015)*, Osnabrück, S. 1008-1022, 2015.
- [Hu99] Hulland, J.: Use of partial least squares (PLS) in strategic management research: a

- review of four recent studies. *Strategic Management Journal*, 20/02, S. 195-204, 1999.
- [KW06] Knackstedt, R.; Winkelmann, A.: Online-Litaturdatenbanken im Bereich der Wirtschaftsinformatik. *Wirtschaftsinformatik*, 48/01, S. 47-59, 2006.
- [KI14] Klör, B.; Bräuer, S.; Beverungen, D.; Matzner, M.: IT-basierte Dienstleistungen für die Elektromobilität – Konzeptioneller Rahmen und Literaturanalyse. In (Kundisch, D.; Suhl, L.; Beckmann, L., Hrsg.): *Tagungsband Multikonferenz Wirtschaftsinformatik 2014 (MKWI 2014)*, Paderborn, S. 2047-2066, 2014.
- [Le11] Lee, J.; Nah, J.; Park, Y.; Sugumaran, V.: Electric Car Sharing Service Using Mobile Technology. In *Proceedings of the International Conference on Information Resources Management (CONF-IRM 2011)*, 2011.
- [Lo89] Lohmöller, J.-B.: *Latent Path Modeling with Partial Least Squares*, Heidelberg, 1989.
- [Ri13] Rickenberg, T. A. A.; Gebhardt, A.; Breitner, M. H.: A Decision Support System For The Optimization Of Car Sharing Stations. In *ECIS 2013*, Paper 207, 2013.
- [RWW05] Ringle, C. M.; Wende, S.; Will, A.: *SmartPLS 2.0.M3*. SmartPLS, www.smartpls.de, Hamburg, 2005.
- [SB13] Seign, R.; Bogenberger, K.: Prescriptions for the Successful Diffusion of Carsharing with Electric Vehicles. In *Conference on Future Automotive Technology (CoFAT 2013) with Focus on Electromobility*, München, 2013.
- [Sc16] Schoormann, T.; Behrens, D.; Kolek, E.; Knackstedt, R.: Sustainability in Business Models – A Literature-Review-Based Design-Science-Oriented Research Agenda. In *Proceedings of the 24th European Conference on Information Systems (ECIS)*, Istanbul, Turkey, 2016.
- [Tr15] Trang, S.; Busse, S.; Schmidt, J.; Falk, T.; Marrone, M.: The Danger of Replacing Human Interaction in IS-driven Collaborative Consumption Services. In *Proceedings of the Twenty-Third European Conference on Information Systems (ECIS)*, Münster, Germany, 2015.
- [UA10] Urbach, N.; Ahlemann, F.: Structural Equation Modeling in Information Systems Research Using Partial Least Squares. *Journal of Information Technology Theory and Application*, 11/02, S. 5-40, 2010
- [Wa14] Wagner, S.; Brandt, T.; Kleinknecht, M.; Neumann, D.: In Free-Float: How Decision Analytics Paves the Way for the Carsharing Revolution. In *Proceedings of the Thirty Fifth International Conference on Information Systems*, Auckland, 2014.
- [Wa15a] Wagner, S.; Brandt, T.; Neumann, D.: Data Analytics in Free-Floating Carsharing: Evidence from the City of Berlin. In *Proceedings of the 48th Hawaii International Conference on System Sciences*, S. 897-907, 2015.
- [Wa15b] Wagner, S.; Willing, C.; Brandt, T.; Neumann, D.: Data Analytics for Location-Based Services: Enabling User-Based Relocation of Carsharing Vehicles. In *Proceedings of the Thirty Sixth International Conference on Information Systems*, Fort Worth, 2015.

Anhang

ID	Ladung	Beschreibung des Indikators
DLNA		Das beschriebene Dienstleistungsangebot ...
	0,526	... würde ich für Dienstreisen bzw. Campuswechsel nutzen.
	0,686	... würde ich als Privatperson für die 250 Euro pro Monat benutzen.
	0,720	... könnte für mich einen eigenen Pkw ersetzen.
	0,463	... würde Dienstreisen von Universitätsmitarbeitern erleichtern.
	-0,551	... wäre für andere Branchen interessanter als für eine Universität.
	-0,560	... halte ich, zur Nutzung als Privatperson, für zu unflexibel.
		Welche Veränderungen würden Ihre Entscheidung, das E-Carsharing-Angebot zu nutzen, positiv beeinflussen?
ECGA	0,824	längere Batteriereichweite
	0,850	Ausbau der Ladestationen (mehr Standorte)
	0,730	Lademöglichkeit zu Hause
	0,866	geringere Akkuladezeit
	0,502	Privilegien für Elektrofahrzeuge wie z. B. die Nutzung von Busspuren, kostenlose Parkplätze
	0,402	mehr Wissen über Elektrofahrzeuge
	0,496	weniger Kosten
	0,541	jederzeitige, garantierte Verfügbarkeit eines Elektrofahrzeuges für Dienstfahrten
	0,453	mehr Flexibilität für außerplanmäßige Ereignisse im Rahmen der Privatnutzung
	0,783	kabellose Lademöglichkeit
EMA		Bitte bewerten Sie die folgenden Aussagen zu Elektromobilität
	0,756	Elektroautos sind eine gute Alternative zum traditionellen Pkw.
	0,667	Elektroautos eignen sich gut für alltägliche Wege.
	-0,432	Elektroautos sind nur etwas für Wege in der Freizeit.
	0,725	Durch die steigenden Ölpreise werden Elektrofahrzeuge zukünftig eine immer größere Rolle spielen.
	0,723	Durch die dringend erforderlichen Einsparungen von CO ₂ -Emissionen werden Elektrofahrzeuge in Zukunft eine immer größere Rolle spielen.
	0,735	Ich habe persönliches Interesse an dem Thema Elektromobilität.
CSA		Bitte bewerten Sie die folgenden Aussagen zu Carsharing: Carsharing ...
	0,752	... ist für jeden eine gute Alternative zu einem eigenen Auto.
	-0,403	... bietet sich nur für Personen an, die selten ein Auto benötigen.
	-0,484	... ist teuer.
	-0,735	... lässt keinen Raum für Spontanität und Flexibilität.
	-0,422	... wäre für mich höchstens als Ergänzung zum eigenen Pkw denkbar.

ID	Ladung	Beschreibung des Indikators
	-0,693	Ich stelle mir die Buchung eines Carsharing-Autos umständlich bzw. zeitaufwendig vor.
	0,716	Ich kann mir gut vorstellen, Carsharing mal auszuprobieren.
NMM	Zu Ihrer Person:	
	0,666	Wie häufig sind im Rahmen ihrer beruflichen Tätigkeit Dienstreisen erforderlich? (1 = Sehr häufig; 5 = nie).
	0,564	In welchem Radius bewegen Sie sich in der Regel, wenn Sie eine Dienstreise antreten? (1 = Innerstädtisch; 5 = Deutschlandweit).
	0,410	Wie häufig legen Sie an einem Tag (privat und/oder beruflich) Strecken von mehr als 100 km zurück? (1 = Sehr häufig; 5 = nie).
	0,507	Ihr Geschlecht? (1 = Männlich; 2 = weiblich).
	0,402	Ihre Arbeitszeit? (1 = Vollzeit; 2 = Teilzeit).

Tab. 5: Indikatoren im Erfolgsmodell innovativer E-Carsharing-Geschäftsmodelle

Softwaregestützte Modellierung von Geschäftsmodellen – Vergleich und Weiterentwicklungsperspektiven am Beispiel der Business Model Canvas

Thorsten Schoormann¹, Dennis Behrens² und Ralf Knackstedt³

Abstract: Zur Modellierung von neuen Geschäftsmodellen wie z. B. im Bereich E-Carsharing, der Batterienachnutzung, neuer Ladeinfrastrukturen, der Stromerzeugung oder der Versorgungsdienstleitungen, werden geeignete Werkzeuge benötigt. Diese sollen den Konstruktionsprozess unterstützen und somit die Entwicklung von (kreativen) Lösungen ermöglichen. Dieser Beitrag analysiert daher, welche Funktionen bestehender Softwarewerkzeuge für eine effiziente Konstruktion und flexible Gestaltung von Geschäftsmodellen – am Beispiel der Business Model Canvas – zur Verfügung stehen. Dazu wird ein Toolvergleich durchgeführt, um anhand des Ist-Zustandes zu untersuchen, welche Gestaltungsoptionen und Weiterentwicklungsperspektiven bestehen. Vergleichsmerkmale leiten wir aus dem allgemeinen Konstruktionsprozess sowie weiterer (neuer) Modellierungsansätze her. Insgesamt wurden elf Werkzeuge auf Basis der konzeptualisierten Merkmale klassifiziert. Vor allem Konzepte der Anpassbarkeit der Modellierungsvorlage, der Berücksichtigung potenzieller Relationen zwischen den Elementen, der Kollaboration im Team sowie die Unterstützung des Konstruktionsprozesses durch gezielte Impulse, Beispiele oder Hilfen werden nicht ausreichend adressiert und sind Bestandteil der Forschungsagenda.

Keywords: Softwaregestützte Modellierung, Tool Review, Business Model Canvas

1 Einleitung

Die heutige Zeit ist durch immer kürzere Innovationszyklen und eine Vielzahl neuer Geschäftsideen, z. B. von Start-Ups, geprägt [Ös07]. Aufgrund von u. a. knapper werdender Ressourcen, steigender (Umwelt-)Verschmutzung und der wachsenden Bevölkerungen inklusive derer Bedürfnisse [UN14], sind besonders im Bereich der Energiewende Innovationen nötig. Durch das stärkere Integrieren erneuerbarer Ressourcen entstehen neue Geschäftsmodelle im Bereich der Elektromobilität wie z. B. E-Carsharing [Ge14] oder Ladeinfrastrukturen [K111] sowie in der Energieversorgung wie z. B. für die Stromerzeugung [GB15] oder Smart-Grid-Lösungen für Verteilnetzbetreiber [SJK12].

¹ Universität Hildesheim, Informationssysteme und Unternehmensmodellierung, 31141 Hildesheim, thorsten.schoormann@uni-hildesheim.de

² Universität Hildesheim, Informationssysteme und Unternehmensmodellierung, 31141 Hildesheim, dennis.behrens@uni-hildesheim.de

³ Universität Hildesheim, Informationssysteme und Unternehmensmodellierung, 31141 Hildesheim, ralf.knackstedt@uni-hildesheim.de

Geschäftsmodelle können als Werkzeuge für die Abbildung, Innovation und Evaluation der Geschäftslogik verwendet werden [Ve14]. Um neue Geschäftsmodelle konzipieren, gestalten und umsetzen zu können, werden Methoden und Modelle sowie Werkzeuge benötigt [Wi03]. Zur Reduzierung der Komplexität neuer Geschäftsmodelle sowie zur Erleichterung der Konstruktion sind insbesondere geeignete (IKT-)Werkzeuge gefragt. Softwaregestützte Werkzeuge können Aspekte der Wirtschaftlichkeit, der Anwendung und der Wartung [RSD08] sowie der Editierbarkeit, der Maschinenlesbarkeit, der Syntaxprüfung oder der konsistenten Speicherung und Wiederverwendung von Modellkomponenten ermöglichen [Wi03].

Für die Modellierung entstehen neue Anforderungen durch innovative Geschäftsmodelle, die es durch adäquate Sprachen zu adressieren gilt [Fr14]. In der Literatur existiert bislang kein Konsens über die Form der Modellierung von Geschäftsmodellen (z. B. [DL03] [KJ12] [Wi16]). Unterschieden wird meist zwischen flussorientierter (z. B. [Go02]) und ganzheitlicher (z. B. [OP02] [OP10]) Perspektive der Geschäftslogik. Darüber hinaus existieren weitere Ansätze wie die Beschreibung mit Partialmodellen (z. B. [Wi01] [AZ01]) oder die Integration domänenspezifischer Modellierungskonzepte (z. B. [St15]). Zu den verbreitetsten und häufig adaptierten Ansätzen [ZB14] zählt die *Business Model Canvas (BMC)*. Die BMC beschreibt Geschäftsmodelle in neun Sichten für Partner, Aktivitäten, Ressourcen, Nutzenversprechen, Kundenbeziehungen, Kanäle, Kunden, Einnahmen und Kosten [OP10]. Trotz hoher Verbreitung werden Erweiterungen für die BMC diskutiert (z. B. [LF10] [He12] [Ka12]), wodurch sich *für Werkzeuge die Notwendigkeit der Flexibilität ergibt, um auf Anpassungen und Weiterentwicklungen adäquat reagieren zu können*. Da die BMC als praxisorientiert verstanden wird, design- und innovationsorientiert arbeitet [SEB14] und es als verbreitet gilt, forcieren wir im Rahmen dieser Ausarbeitung die softwaregestützte Konstruktion einer BMC.

Dieser Beitrag analysiert, welche Funktionen bestehender Softwarewerkzeuge für eine effiziente Konstruktion und Gestaltung von Geschäftsmodellen zur Verfügung stehen. Die Forschungsfrage lautet daher: *Inwiefern unterstützen bestehende Werkzeuge die Anforderungen der softwaregestützten Konstruktion der Business Model Canvas?*

Zuerst werden dazu Kriterien und Vorgehensmodelle zur Analyse und zum Vergleichen von Software skizziert (Abschnitt 2). Auf Basis des methodischen Vorgehens eines Tool Reviews (Abschnitt 3) werden Kriterien für BMC-Tools hergeleitet (Abschnitt 4.1), anhand derer geeignete Werkzeuge identifiziert (Abschnitt 4.2) und aufbereitet werden (Abschnitt 4.3). Nach dem Vergleich folgt die Darstellung des Ist-Zustandes der analysierten Werkzeuge (Abschnitt 5), um schließlich Weiterentwicklungsmaßnahmen in einer Forschungsagenda ableiten und darstellen zu können (Abschnitt 6).

2 Theoretischer Hintergrund

Das Schaffen eines transparenten Softwarevergleichsvorgehens ist ein entscheidender Faktor, um die Nachvollziehbarkeit zu erhöhen und z. B. die Überprüfung der ursprünglichen Ziele zu ermöglichen [Ni10]. Die Gestaltung eines transparenten Prozesses kann sich an (a) allgemeinen Vorgehensmodellen der Softwareauswahl sowie an (b) forschungsmethodischen Vorgehenshinweisen von Literatur Reviews orientieren.

Orientierung durch Softwarevergleich/-auswahlvorgehen. Sowohl in der Literatur als auch in der (Beratungs-)Praxis werden verschiedene Ansätze diskutiert [Ni10]. Dabei ähneln sich Vorgehensweisen häufig in den durchzuführenden Phasen. So unterscheidet z. B. [SV03] drei Phasen für die Konstruktion des Auswahlproblems, die Grobanalyse und die Detailanalyse. [Ni09] erarbeitet ein Vorgehensmodell aus Initiierung, Potenzialanalyse, Anforderungen, Vorauswahl, Ausschreibung, Angebotsbewertung, Präsentation, Endauswahl und Vertrag. Ein weiterer Vorschlag bildet Bereiche für das Definieren der Anforderungen, das Vorauswählen, das Erstellen eines Testfahrplans, das Moderieren der Anbieterpräsentationen sowie das Ableiten von Entscheidungsempfehlungen [PB16].

Der Vergleich und die Auswahl sollten durch individuelle Anforderungen geprägt sein [Str15]. Häufig werden dazu Kategorien gebildet, die allgemeine Software-Merkmale (z. B. Preis, Lizenzierung und Plattform), Anbieter-Merkmale (z. B. Services, Marktanteil und Referenzen) und fachspezifische Merkmale (z. B. Funktionalitäten für spezielle Anwendungen) berücksichtigen [La89] [SG16]. Zur Herleitung domänenspezifischer Kriterien wird eine entsprechende Perspektive eingenommen, die es ermöglicht relevante Merkmale zu identifizieren. Da für die Auswahl und Analyse von Werkzeugen zur Konstruktion von Geschäftsmodellen keine einheitlichen Kriterien identifiziert werden konnten, wird zunächst *eine Übersicht relevanter und fachspezifischer Merkmale benötigt*.

Orientierung durch Literatur Review-Vorgehen. Zur Schaffung von Transparenz im gesamten Auswahlprozess bieten Literatur Reviews Unterstützungspotenziale. [Br09] forciert z. B. das kontinuierliche Dokumentieren von Such- und Auswahlprozessen, um die Nachvollziehbarkeit der Resultate zu erhöhen. Dazu erfolgt die Definition des Bereichs, die (domänenspezifische) Konzeptualisierung, das Literatur Review, die Literaturanalyse und -synthese sowie eine Forschungsagenda. [OS10] unterscheidet acht Phasen für den Zweck des Reviews, das Training der Methode, die Suche, das praktische Prüfen, die Qualitätsbewertung, die Datenextraktion, die Synthese sowie das Schreiben des Reviews. [WW02] beschreibt Guidelines und einen grundsätzlichen Aufbau mit Beginn (Topic, Definitionen), Identifikation relevanter Literatur, Zusammenfassung, theoretische Entwicklung, Evaluation sowie Diskussion. Weitere Ansätze ähneln sich häufig im Grundkonzept der Datensammlung, -evaluation, -analyse und -interpretation.

3 Methodisches Vorgehen

Zur nachvollziehbaren Gestaltung des Vorgehens sowie der einzelnen Such-, Auswahl- und Evaluationsphasen verwenden wir die Methode des Tool Reviews [SK16]. Unser Vorgehen umfasst folgende Phasen (Abb. 1): *Definition des Bereichs und Ziels (I)*: Analyse und Vergleich bestehender Tools (vgl. Abschnitt 1). *Konzeptualisierung des Themas (II)*: Herleitung von spezifischen Merkmalen, die bei der Suche und Analyse berücksichtigt werden sollen (vgl. Abschnitt 4.1). *Tool Review (III)*: Ableitung von Schlüsselbegriffen, Konfiguration des Suche, Auswahl von Quellen und Definition von Evaluationskriterien (vgl. Abschnitt 4.2). *Toolanalyse und -aufbereitung (IV)*: Analyse und Klassifikation der Tools mit Hilfe einer Konzept-Matrix. Das Resultat ist der State of the Art der BMC-Werkzeuge (vgl. Abschnitt 4.3). *Forschungsagenda (V)*: Abschließend werden die Resultate interpretiert, um Weiterentwicklungsmaßnahmen für Werkzeuge im Kontext der Konstruktion von Geschäftsmodellen abzuleiten (vgl. Abschnitt 6).

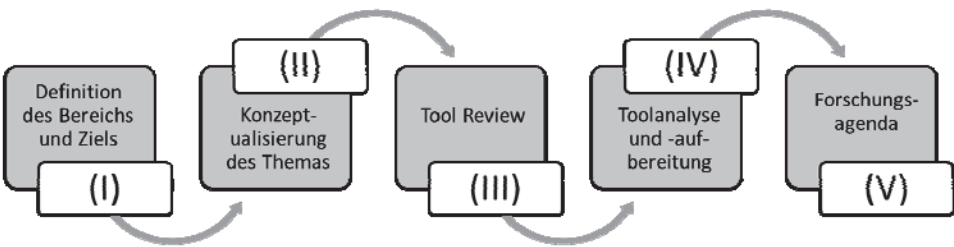


Abb. 1: Methodisches Vorgehen in Anlehnung an [SK16]

Unser Vorgehen adressiert zum einen Aspekte aus dem Softwarevergleich wie z. B. Festlegung von Anforderungen, Vorauswahl und Aufbereitung. Zum anderen werden Phasen aus dem Vorgehen von Literatur Reviews adaptiert wie z. B. Analyse und Synthese relevanter Beiträge sowie das Aufstellen einer Forschungsagenda, um Weiterentwicklungen für den analysierten Gegenstand abzuleiten. Zur Verdeutlichung des Vorgehens werden exemplarisch jeweils drei Vorgehensweisen aus dem Bereich des Softwarevergleichs und aus dem der Literatur Reviews mit unserem Tool Review gegenübergestellt (Tab. 1). Dabei wird deutlich, dass insbesondere das Kombinieren der expliziten Herleitung von Anforderungen (Softwarevergleich) mit dem Skizzieren der Forschungsagenda (Literatur Review) hilfreich ist, um auf Basis des Ist-Zustandes geeignete Weiterentwicklungsmaßnahmen ableiten zu können.

Tool Review	Softwarevergleich			Literatur Review		
[SK16]	[SV03]	[Ni09]	[PB16]	[WW02]	[Br09]	[OS10]
Definition des Bereichs und Ziels	Auswahlproblem	Initiierung/Analyse		Topic, Definitionen	Definition des Bereichs	Zweck
Konzeptualisierung	Grobanalyse	Anforderungen	Anforderungen		Konzeptualisierung	

Tool Review	Softwarevergleich			Literatur Review		
[SK16]	[SV03]	[Ni09]	[PB16]	[WW02]	[Br09]	[OS10]
des Themas	(Anf.)					
Tool Review	Detail- analyse	Vorauswah l/ Bewertung	Vorauswah l/ Fahrplan	Sources/ „back & for“	Literatur Suche	Suche/ Practical Screen
Toolanalyse und -aufbereitung		Präsen- tation/End- auswahl	Anbieterpr äsentation	Relevante Artikel	Literatur Analyse/ Synthese	Bewertung/ Extraktion/ Synthese
Forschungs- agenda				Motivation “close the breach”	Research Agenda	

Tab. 1: Auswahl des methodischen Vorgehens

4 Vergleich von Werkzeugen

4.1 Konzeptualisierung des Themas

Zur Herleitung von Merkmalen für den Vergleich und die Klassifikation der BMC-Werkzeuge untersuchen wir zunächst den (allgemeinen) Prozess der Konstruktion von Geschäftsmodellen, um daraus entsprechende Kriterien abzuleiten.

Kriterien der Geschäftsmodellkonstruktion. Nach [OP10] besteht der Modellierungsvorgang für die BMC aus fünf Phasen: Mobilisieren (Projektziele, Teamzusammenstellung etc.), verstehen (Kontextverständnis etc.), gestalten (Anpassungen des Geschäftsmodells mit Hilfe von Kreativitätstechniken), implementieren (Meilensteine, Finanzierungsplanung etc.) sowie durchführen (kontinuierliche Ausführung des Geschäftsmodells). Folgende Kriterien können dabei aus den Phasen abgeleitet werden:

- *Kollaboration.* In der Phase der Mobilisierung erfolgt die Teamfestlegung. Um das gemeinsame Arbeiten zu ermöglichen sind Kollaborationsfunktionen nötig. Wir unterscheiden zwischen asynchron (keine gleichzeitige gemeinsame Bearbeitung) und synchron (gleichzeitige Bearbeitung z. B. durch Chats).
- *Post-It.* Für das Gestalten werden neben der Vorlage (BMC) meist Post-It-Klebezettel verwendet. Da die Konstruktion durch visuelle Elemente (Visual Thinking) unterstützt werden soll, sind z. B. Text, Abbildungen und Tabellen zu berücksichtigen [SEB14]. Die Klebezettel sollten zudem editierbar sein, z. B. durch Größenanpassung, Textformatierung und Farbanpassung. Ein Vorteil mit Post-Its besteht in der schnellen Editier- und Verschiebbarkeit in einen anderen Block der BMC (Platzierung im Raster).
- *Kreativitätstechnik.* Kreativitätstechniken können den Gestaltungsprozess

unterstützten, wodurch Funktionalitäten gefragt sind, die geeignete Techniken vorschlagen. Empfohlen wird zudem z. B. das Entwickeln mehrerer „Prototypen“ eines Modells, weshalb Varianten- und Versionsmanagement hilfreich sind.

- *Beispiel.* Das zur Verfügung stellen von Beispielen unterstützt das Gestalten von neuen Modellen sowie die Anwendung der BMC [SEB14].
- *Vorgehensassistent.* Da die Gestaltung und Konstruktion durch einen moderierten Prozess erleichtert werden kann [SEB14], sollte die Nutzung durch gezielte Impulse unterstützt werden wie z. B. durch Assistenten oder Vorschlagsfunktionen.
- *Ansicht.* Da bei entsprechenden Detaillierungsgeraden ein Modell unübersichtlich werden kann, sollte das Tool verschiedene Ansichten (z. B. Zoom) unterstützen.
- *Import und Export.* Anzahl möglicher Import- bzw. Exportformate zur Förderung des Austausches/Kollaboration von Dateien.

Kriterien adaptierter Frameworks. Zur Gestaltung von neuen Geschäftsmodellen wie z. B. unter der Berücksichtigung nachhaltiger Aspekte, existieren bereits Vorschläge für Erweiterungen und Änderungen, die im Folgenden ebenfalls integriert werden [Sc16].

- *Vorlageneditierbarkeit.* Zur Modellierung innovativer Ansätze werden Anpassungen an der Vorlage vorgenommen wie z. B. Block-Erweiterung, Block-Teilung, Blockbezeichnung ändern oder die Anordnungsstruktur anpassen.
- *Element-Verknüpfung.* Da sich Potenziale zwischen den Blöcken befinden (z. B. [He12] [Ge14]), sollten Interaktionen und Beziehungen zwischen Elemente modelliert werden können – hier: Block- und Post-It-Verlinkung.
- *Perspektiven und Filter.* Das Legen von Perspektiven über die BMC zur Forcierung eines bestimmten Gegenstandes oder einer Strategie. Alternativ durch Gruppierungen nach den spezifisch gefiltert werden kann wie z. B. über die Farben.
- *Bewertung.* Integration von Bewertungsfunktionen, z. B. durch spezifische Farben der Elemente oder ergänzende Symbole.
- *Detail-Modelle.* Integration von Detail-Modellen in einzelnen Blöcken der BMC, um bspw. weitere Modelle zu hinterlegen, die detaillierte Informationen liefern.
- *Domänenspezifische Modellierung.* Modellierungen, die domänenspezifische Elemente berücksichtigen, um einerseits die Konstruktion gezielt zu unterstützen und andererseits Aspekte der Vergleichbarkeit erhöhen können [St15].

Allgemeine Merkmale (vgl. Abschnitt 2.1) sind weiterhin gültig, stehen hier jedoch nicht im Vordergrund, weshalb sie in der Klassifikation nicht explizit adressiert werden.

4.2 Tool Review

Identifikation von Keywords. Das Festlegen von Suchbegriffen im Vokabular der zu analysierenden Domäne ist ein wichtiger Bestandteil der Suche [Br09]. Um geeignete Begriffen zu ermitteln, werden verschiedene Wörter (BMC, Business Model Canvas, Tool, Software, Applikation, Werkzeug) verwendet und die Resultate analysiert sowie evaluiert. Da zahlreiche Einträge zu „BMC“ nicht die Business Model Canvas adressierten, sondern für andere Abkürzungen stehen, wird dies nicht berücksichtigt. Folgende Begriffe haben wir auf Basis der Analyse für die Suche festgelegt:

Business Model Canvas, Tool, Software und Applikation

Selektion von Datenbanken. Zur Berücksichtigung möglichst vieler verschiedener Werkzeuge für die softwaregestützte Modellierung wird die Google-Suchmaschine ausgewählt. Kombinationen aus den identifizierten Keywords werden schließlich als Suche ausgeführt, um eine Übersicht der bestehenden Tools zu erhalten.

4.3 Toolanalyse und -aufbereitung

Suchresultate. Durch das initiale Evaluieren der Titel und der Beschreibungen können in der ersten Phase 27 relevante Einträge⁴ ermittelt werden. Es werden die ersten 300 Einträge (30 Seiten) von Google analysiert, da die Qualität der Einträge hinsichtlich unserer Forschungsfrage stetig abgenommen hat bzw. als gesättigt empfunden wird. In der nächsten Evaluationsstufe werden die Tools detaillierter analysiert, wodurch sich Ausschlusskriterien ergeben. Ein Werkzeug wird im Folgenden nicht weiter berücksichtigt, insofern es (a) ausschließlich eine Vorlagen (z. B. im PDF-Format) bereitstellt, (b) es ausschließlich Zeichenwerkzeuge – in die Hintergrundbilder wie die BMC eingefügt werden können (z. B. MS Paint) – bereitstellt, (c) es ausschließlich adaptierte Leinwände bereitstellt (z. B. LeanStack.com) oder (d) es von der Anbieterseite nicht mehr zur Verfügung gestellt wird. Unter Berücksichtigung der Ausschlusskriterien werden schließlich 16 Werkzeuge ermittelt, die in Tab. 2 aufgelistet werden.

Werkzeug	Referenz	Beschreibung
Archi	http://www.archimatetool.com/	Client (Windows 7, 8, 10); Open Source
BizCanvas 1.1.3	https://itunes.apple.com/nz/app/bizcanvas-app/id588263801?mt=8#	App; iOS nötig (ab 7.0); erhältlich bei Apple iTunes
BM Canvas	www.bmcanvas.com	Online; Anmeldung nötig
BM Designer	www.bmdesigner.com	Online; Anmeldung nötig
BM Fiddle	www.bmfiddle.com	Online; weitere Funktionen (z. B. Speichern) durch Anmeldung
Business Model Canvas StartUp	www.mobills.com.br/	App; Android nötig (ab 4.0.3); erhältlich im Google-Play-Store
Business Model Toolbox 1.2.2	https://itunes.apple.com/us/app/business-model-toolbox/id431605371?mt=8	App; iOS nötig (ab 5.1); Strategyzer AG; erhältlich bei Apple iTunes; 29\$

⁴ www.google.de, Zugriff am 05.05.2016.

Werkzeug	Referenz	Beschreibung
Canvanizer	www.canvanizer.com	Online; E-Mail Adresse nötig
CanvasBM (Beta)	www.canvasbm.com	Online; Anmeldung möglich
Insight Maker	www.insightmaker.com	Online; Anmeldung nötig; Vorlagen und Simulationsmöglichkeiten der BMC
LeanMonitor	www.leanmonitor.com	Online; Anmeldung nötig; Free Trial; Basis (9\$/Monat); Premium (22\$/Monat)
Lienzo (Free)	www.lienzo.biz	Online; Anmeldung nötig; erweiterte Funktionen (5\$/Monat)
Realtime Board	www. realtimeboard.com	Online; Anmeldung nötig; Teamedition (35\$/Monat)
Strategyzer	www.strategyzer.com	Online; Anmeldung nötig; 299\$/Jahr; Gründer: A. Osterwalder und A. Smith
Tuzzit (Free)	www.tuzzit.com/en/canvas/business_model_canvas	Online; Anmeldung nötig; erweiterte Funktionen (5-49\$/Monat)
Waxidea	http://businessmodelcanvas.waxidea.com/	Online; keine Anmeldung nötig

Tab. 2: Übersicht bestehender Werkzeuge für die BMC

Klassifikation. Da wirtschaftliche Faktoren (z. B. Kosten) ein Hemmnis für die Nutzung sein können, klassifizieren wir im Folgenden zunächst kostenlose BMC-Tools (Tab. 3). Darüber hinaus eliminieren wir Werkzeuge, die nur als Smartphone-App angeboten werden, da diese Restriktionen wie geeignete Endgeräte enthalten. Zur Übersicht sind die Kriterien in die Bereiche *Canvas*, *Post-It* und *Bedienung/Sonstiges* gegliedert.

		Archi	BM Canvas	BM Designer	BM Fiddle	Canvanizer	CanvasBM	Insight Maker	Lienzo (Free)	Realtime Board	Tuzzit (Free)	Waxidea
Canvas	Block-Erweiterung	X	-	-	/	-	-	(X)	-	X	-	-
	Block-Teilung	(X)	-	-	/	-	-	(X)	-	X	-	-
	Blockbezeichnung ändern	X	-	-	X	-	-	-	-	X	-	-
	Anordnungsstruktur	X	-	-	/	-	-	(X)	-	X	-	-
	Block-Verlinkung	X	-	-	/	-	-	(X)	-	-	-	-
Post-It	Post-It + Text	X	X	X	X	X	X	-	X	X	X	X
	Post-It + Abbildungen	X	-	-	X	-	-	-	-	X	X	-
	Post-It + Tabellen	-	-	-	X	-	-	-	-	X	X	-
	Größenanpassung	X	-	-	-	-	-	-	-	X	X	-
	Textformatierung	X	-	X	X	-	-	-	-	X	X	-
	Farbenanpassung (Anzahl)	11	6	X	9	5	1	12	5	X	X	1
	Platzierung im Raster	X	X	X	X	X	X	-	X	X	-	X
	Post-It-Verlinkung	(X)	-	X	-	-	-	-	-	-	Pro	-
Bedienung/Sonstiges	Teamfestlegung	-	-	-	-	X	-	-	X	X	-	-
	Kollaboration (synchron)	-	-	-	-	-	-	-	X	X	X	-

	Archi	BM Canvas	BM Designer	BM Fiddle	Canvanizer	CanvasBM	Insight Maker	Lienzo (Free)	Realtime Board	Tuzzit (Free)	Waxidea
Kollaboration (asynchron)	-	-	-	X	X	-	-	-	X	-	-
Versionsmanagement	X	-	-	-	X	-	-	-	X	-	-
Variantenmanagement	(X)	(X)	-	(X)	(X)	-	-	-	(X)	-	-
Hilfe	X	X	X	X	X	X	X	-	X	-	X
Beispiel	-	-	X	-	-	X	(X)	-	X	-	-
Vorgehensassistent	-	-	-	-	-	-	-	-	-	-	-
Kreativitätstechnik	-	-	-	(X)	X	-	-	-	(X)	(X)	-
Ansicht/Zoom	X	-	X	X	-	-	X	-	X	X	-
Perspektive	-	-	X	-	-	-	-	-	(X)	-	-
Filter	-	X	X	X	-	-	-	-	-	-	-
Detail-Modell	-	-	-	-	-	-	-	-	-	-	-
Bewertung	-	-	-	-	-	-	(X)	-	-	-	-
Domänenspez. Modell	-	-	-	-	-	-	-	-	-	-	-
Import (Anzahl)	2	-	-	-	2	-	-	-	>5	-	-
Export (Anzahl)	4	1	2	-	3	1	-	3	4	1	-

Note: (X) = implizite Umsetzung/Erfüllung des Kriteriums.

Tab. 3: Klassifikation analysierter Werkzeuge

5 Diskussion

5.1 Ist-Zustand der Werkzeuge und deren Funktionalitäten

Im Folgenden werden Besonderheiten, Gemeinsamkeiten und Unterschiede der analysierten Werkzeuge erläutert und diskutiert, um den Ist-Zustand der Tools zu erläutern.

Modellierungsvorlage. Im Bereich der Modellierungsvorlage können insbesondere Archi, Insight Maker und Realtime Board Unterstützungspotenziale liefern. Archi ist ein freies Open Source Modellierungstool zur Konstruktion von Modellen (z. B. ArchiMate-Modelle), mit dem die Vorlage erweitert (neue Blöcke) und bearbeitet (Bezeichnung und Größe) werden kann. Durch die Anpassung der Größe und Erweiterung ist daher implizit auch die Teilung bestehender Blöcke möglich. Allerdings stellt das clientbasierte Archi „nur“ die Möglichkeit zur freien Bearbeitung allgemeiner Objekte (Rechtecke sind nicht explizit für die BMC) zur Verfügung, wie es mit anderen Werkzeugen, z. B. Prezi oder Office Produkten, prinzipiell auch möglich wäre. Der Insight Maker ist ebenfalls ein freies Werkzeug zur Modellierung und Simulation von Modellen. Es können Vorlagen

frei angelegt werden, sodass die BMC implizit erstellt und angepasst werden kann. Dies wäre allerdings wie bei Archi auch mit weiteren Tools möglich. Insight Maker stellt jedoch bereits vordefinierte Templates für die BMC zur Verfügung, in denen modelliert, aber auch simuliert werden können (z. B. Material- und Finanzflüsse). Durch Simulationen können zudem Bewertungen bzw. Kennzahlen integriert werden.

Das Realtime Board bietet in der kostenlosen Ausführung für kleine Projekte eine flexible Plattform zur Kollaboration, die neben weiteren Vorlagen auch die BMC enthält. Die BMC kann, wie bei Archi und Insight Maker, editiert werden. Zudem besteht die Möglichkeit Frames auf die Vorlage zu legen, um bspw. in spezifische Bereiche zu zoomen.

Modellierungselemente. Das Modellieren von Klebezetteln (Post-Its) wird von allen Werkzeugen – außer Insight Maker, das nur freies Zeichnen erlaubt – unterstützt. Darüber hinaus bieten vor allem Archi, BM-Fiddle, Realtime Board und Tuzzit weitere Gestaltungsoptionen an. BM Fiddle ermöglicht das Dokumentieren und das Entwickeln von Ideen anhand vordefinierter Vorlagen wie die BMC. Dazu können neben Texte auch Abbildungen und weitere Dateien verwendet und angepasst werden. Auch Tuzzit bietet verschiedene Anpassungs- und Integrationsmöglichkeiten von Objekten. Verlinkungen zwischen Elementen werden nur durch die Pro-Version unterstützt. Das Verlinken wird von Archi (zwischen Blöcken und Post-Its) und dem BM Designer (zwischen Post-Its) unterstützt.

Bedienung/Sonstiges. Das Arbeiten im Team ist bei der Konstruktion einer BMC von großer Bedeutung, wodurch entsprechende Funktionalitäten gegeben sein müssen. Synchroner Kollaboration (Echtzeit Zusammenarbeit am gleichen Dokument) wird von Insight Maker, Realtime Board sowie Tuzzit und asynchroner Kollaboration vom BM Fiddle, Canvanizer und dem Realtime Board ermöglicht. Zudem kann das Festlegen eines Teams bei dem Canvanizer, Lienzo und dem Realtime Board erfolgen, in dem bspw. Mitglieder via E-Mail eingeladen werden können.

Die Erstellung von neuen Modellen sowie die allgemeine Anwendung der BMC kann durch Beispiele erleichtert werden. Die Umsetzung dieses Merkmals ist bei dem BM Designer, dem Canvas BM und dem Realtime Board explizit und implizit bei dem Insight Maker (abhängig von der Community) gegeben.

Für die Verwaltung der Modelle bietet der Canvanizer eine interaktive Modellhistorie, in der die verschiedenen Stände der BMC chronologisch nachvollzogen und abgespielt werden können. Das Verwalten von Varianten wird nur implizit durch das manuelle Anlegen entsprechender Kopien und Umbenennung ermöglicht (Archi, BM Canvas, BM Fiddle, Canvanizer und Realtime Board).

Ein weiterer Faktor ist die Begleitung durch den Konstruktionsprozess, die durch geeignete Hilfestellungen und Assistenten erfolgen kann. Während die meisten Werkzeuge über eine allgemeine Hilfe (z. B. Tutorials, FAQ oder Dokumentationen)

verfügen, konnten keine integrierten Vorgehensmodelle identifiziert werden, die dem Anwender aufzeigen, welche Schritte durchzuführen sind.

Hinsichtlich der Kreativitätstechniken bietet der Canvanizer einen Bereich für „Brainstorming Notes“ an, dem zu diskutierende Post-Its zugewiesen werden können, die schließlich von den entsprechenden Teammitgliedern in einem Brainstorming erörtert werden. Weitere Tools stellen Vorlagen für z. B. Empathie-Maps, Mindmaps, Personas oder Stroy-Maps zur Verfügung, die das kreative Bearbeiten unterstützen.

5.2 Limitationen

Die Resultate in diesem Beitrag unterliegen insbesondere zwei Einschränkungen. Zum einen beschränken wir uns auf das Framework der Business Model Canvas. Obwohl dieser als ein (weit) verbreiteter Ansatz gilt, können weitere Merkmale für andere Modellierungstechniken von Relevanz sein oder Werkzeuge anderer Modellierungstechniken unsere Anforderung bereits erfüllen. Zum anderen beschränken wir uns aufgrund der Aspekte der Wirtschaftlichkeit in der Klassifikation (vgl. Abschnitt 4.3) auf kostenlose Werkzeuge zur Konstruktion der BMC. Kostenpflichtige Werkzeuge könnten daher höhere Unterstützungspotenziale für die konzeptualisierten Merkmale enthalten.

Das Tool Review sowie die entsprechenden Suchkonfigurationen basieren auf Hinweisen von [Br09]. Die Auswahl der Suchbegriffe, Suchmaschinen, Ausschlusskriterien sowie die Klassifikation folgen eigenen Entscheidungen, welche Limitationen haben. Es hätten z. B. weitere Schlüsselbegriffe integriert oder weitere Suchmaschinen verwendet werden können, um neue relevante Erarbeiten zu identifizieren.

6 Forschungsagenda

Die Klassifikation (vgl. Abschnitt 4.3) verdeutlicht bereits, welche Vor- und Nachteile bei den analysierten Werkzeugen bestehen. Daran orientiert sind entsprechende Weiterentwicklungsmaßnahmen ableitbar, die im Folgenden skizziert werden.

Anpassbarkeit der Modellierungsvorlage. Das Verändern der vorgegebenen Blöcke der BMC sowie deren Bezeichnung ist nur durch Werkzeuge gegeben, die allgemeine Template-Funktionen anbieten, die bspw. auch in anderen Bild- und Textverarbeitungstools umgesetzt werden können. Das Modellieren von innovativen und neuen Geschäftsmodellen kann jedoch durch die Flexibilität der Vorlage selbst profitieren. Die Anpassbarkeit der Vorlage kann dabei z. B. *(a) das Verändern der Blockbezeichnung, (b) das Teilen bestehender Blöcke, (c) das Anpassen der Anordnungsstruktur sowie (d) das Hinzufügen neuer Blöcke* adressieren.

Berücksichtigung potenzieller Relationen. Neben der Anpassbarkeit der Blöcke können sich Potenziale zwischen den verschiedenen Elementen selbst ergeben. Das Verlinken von z. B. (a) *Objekt zu Objekt* (z. B. *Post-It*), (b) *Objekt zu Block* und (c) *Block zu Block* [Sc16] kann durch entsprechende Funktionalitäten im Softwarewerkzeug ermöglicht werden.

Integration verschiedener Perspektiven. Das Integrieren von Perspektiven, die während der Konstruktion eingenommen werden oder spezifische Inhalte in den Vordergrund heben können ist unzureichend durch die Werkzeuge gewährleistet. In Abhängigkeit von Zielgruppe und inhaltlicher Ausrichtung kann dieses bestimmte Anregungen und Hinweise geben. *Perspektiven können über (a) das gesamte Modell, (b) einzelne Blöcke oder (c) das gesamte Modell und einzelne Blöcke gelegt werden* [Sc16].

Orientierung durch Beispiele. Das zur Verfügung stellen von Beispielen kann die Konstruktion neuer Geschäftsmodelle (z. B. in ähnlichen Branchen) sowie den Umgang mit der BMC selbst (z. B. Orientierung an typischen Inhalte bestimmter Blöcke) unterstützen. Beispiele können dabei u. a. (a) *branchenspezifische Best-Practices*, (b) *vereinfachte Modellierungen zum Verständnis* oder (c) *bekannte Geschäftsmodelle zur Orientierung* sein.

Integration von Kollaborationsmöglichkeiten. Da die Konstruktion der BMC insbesondere durch eine geeignete Zusammenstellung eines Teams und deren Kollaboration erfolgen kann, sind Funktionalitäten für (a) *das Integrieren von weiteren Anwendern*, (b) *das Bereitstellen von Chats und Diskussionen* sowie (c) *der gemeinsamen Echtzeitbearbeitung* nötig.

Unterstützung des Konstruktionsprozesses. Der Prozess zur kollaborativen und innovativen Gestaltung kann von gezielten Impulsen und Denkanstößen profitieren. Werkzeuge sollten daher softwarebasierte Assistenten, die (a) *Hinweise zur Vorgehensweise geben* oder (b) *die Vorgehensweise durch entsprechende Funktionen zur phasenweisen Abarbeitung unterstützen* oder (c) *Kreativitätstechniken zur Generierung neuer Ideen vorschlagen*. Initiale Arbeiten nehmen sich der Thematik zur Integration von Kreativitätstechniken in Tools bereits an (z. B. [Ko14]).

Modellierung der Inhalte. Die Inhalte der einzelnen Blöcke können durch verschiedene Medien(-formate) und Gestaltungsoptionen visualisiert werden. Dazu können Softwarewerkzeuge (a) *weitere Formate wie Abbildungen, Diagramme, Listen, Tabellen oder Videos integrieren*, (b) *Detail-Modelle für bestehende Blöcke* wie z. B. *Prozessmodelle für Schlüsselaktivitäten vorschlagen* oder (c) *domänenspezifische Modellierungskonzepte zur Verfügung stellen* (z. B. [St15] [Sc16]).

Bewertung der Inhalte. Modellierte Inhalte (z. B. Antriebsarten, Ressourcen) können zudem anhand vordefinierter Merkmale bewertet werden, um kollaborativ neue Ideen zu verbesserungsfähigen Konzepten zu entwickeln. Eine Bewertung kann grafisch mit Hilfe von Metaphern wie den drei Farben einer Ampel erfolgen. Das Bewerten kann *anhand (a) der gesamten BMC oder (b) einzelner Blöcke* umgesetzt werden [Sc16].

Danksagung

Unterstützt wurde der Forschungsprozess von dem „Seminar Wirtschaftsinformatik“ der Universität Hildesheim. Wir danken insbesondere den Teilnehmern Timo Behrens, Adam Krafczyk und Nils Sommerfeld für ihre Unterstützung.

7 Literaturverzeichnis

- [AZ01] Alt R.; Zimmermann, H. D.: Introduction to special section-business models. In: Electronic Markets-The International Journal 11(1), S. 1019-6781, 2001.
- [Br09] vom Brocke, J.; Simons, A.; Niehaves, B.; Reimer, K.; Plattfaut, R.; Cleven, A.: Reconstructing the Giant: on the Importance of Rigour in Documenting the Literature Search Process. In: Proceedings of the 17th European Conference on Information Systems (ECIS). Verona, Italy, 161, 2009.
- [DL03] Deelmann, T.; Loos, P.: Visuelle Methoden zur Darstellung von Geschäftsmodellen – Methodenvergleich, Anforderungsdefinition und exemplarischer Visualisierungsvorschlag. In: Arbeitsberichte der Johannes Gutenberg-Universität Mainz, Mainz, 2003.
- [Fr14] Frank U.; Strecker, S.; Fettke, P.; vom Brocke, J.; Becker, J.; Sinz, E.: Das Forschungsfeld „Modellierung betrieblicher Informationssysteme“. In: Wirtschaftsinformatik 56(1): 49-54, 2014.
- [GB15] Gsodam, P.; Bachhiesel, U.: Geschäftsmodelle für die Stromerzeugung aus erneuerbaren Energien in Österreich–eine Fallstudienuntersuchung, www.eeg.tuwien.ac.at/eeg.tuwien.ac.at_pages/events/iewt/iewt2015/uploads/abstracts/A_267_Gsodam_Petra_7-Jan-2015_16-51.pdf, Stand: 13.05.2016.
- [Ge14] Gerwig, C.; Behrens, D.; Knackstedt, R.; Lessing, H.: (2014). Innovatives eCarSharing-Konzept am Beispiel des Hildesheimer Tandem-Modells. In: GI-Jahrestagung, S. 1617-1629, 2014.
- [La89] Lang, G.: Auswahl von Standard-Applikations-Software: Organisation und Instrumentarien. Springer-Verlag Berlin Heidelberg, 1989.
- [Go02] Gordijn, J.: Value-based Requirements Engineering – Exploring Innovative e-Commerce Ideas. In Dissertation, Universität Vrije, 2002.
- [He12] Henriksen K.; Bjerre, M.; Maria Almasi, A.; Damgaard-Grann, E.: Green Business Model Innovation-Conceptualization report, 2012.
- [Ka12] Kanshieva, A.: Sustainability-driven entrepreneurship as a tool to increase rural resilience in Russia: framework for transfer of the European experience. In: Dissertation, Universität Padova, 2012.
- [Kl11] Kley, F.: Neue Geschäftsmodelle zur Ladeinfrastruktur. In: No. S5/2011. Working paper sustainability and innovation, 2011.
- [KJ12] Kundisch, D.; John, T.: Geschäftsmodelle als Thema der Wirtschaftsinformatik – State-ments zum Themenfeld „Stand der Forschung“. In (Hess, T. Hrsg.):

Arbeitsberichte des Instituts für Wirtschaftsinformatik und Neue Medien der LMU München 1, S. 9-11, 2012.

- [LF10] Lüdeke-Freund, F.: Towards a conceptual framework of business models for sustainability. In (Wever, R.; Quist, J.; Tukker, A.; Woudstra, J.; Boons, J.; Beute, N.): Proceedings of ERSCP-EMSU Conference, 2010.
- [Ni09] Nissen, V.; Simon, C.: Zur Unternehmensberatung. Kernaufgaben und Vorgehensmodelle in der IV-Beratung. In (Nissen, V. Hrsg.): Forschungsbericht Nr. 2009-01, 2009.
- [Ni10] Nissen, V.: Entstehung und Handhabung von Vorgehensmodellen zur Software-Auswahl und Software-Einführung in der IV-Beratung. In: MKWI, S. 595-610, 2010.
- [OP02] Osterwalder A.; Pigneur, Y.: An e-Business Model Ontology for Modeling e-Business. In Proceedings of the 15th BLED, Slovenien, 2-14, 2002.
- [OP10] Osterwalder, A.; Pigneur, Y.: Business model generation: a handbook for visionaries, game changers, and challengers. John Wiley & Sons, 2010.
- [Ös07] Österle, H.: Business Engineering-Geschäftsmodelle transformieren. Architekturen und Prozesse: Strukturen und Dynamik in Forschung und Unternehmen, S.71-84, 2007.
- [OS10] Okoli, C.; Schabram, K.: A Guide to Conducting a Systematic Literature Review of Information Systems Research. In: Working Papers on Information Systems 10 (26), 2010.
- [PB16] Prof. Becker GmbH 2016, Softwareauswahl und –einführung, <http://www.prof-becker.de/kategorie/softwareauswahl-und-einfuehrung/>, Stand: 13.05.2016.
- [RSD08] Rosemann, M.; Schwegmann, A.; Delfmann, P.: Vorbereitung der Prozessmodellierung. In (Becker, J.; Kugeler, M.; Rosemann, M. Hrsg.): Prozessmanagement. Ein Leitfaden zur prozessorientierten Organisationsgestaltung. 6. Auflage, Berlin et al.: Springer, S. 45-103, 2008.
- [SEB14] Simmert, B.; Ebel, P.; Bretschneider, U.: Empirische Erkenntnisse zur Nutzung des Business Model Canvas. In: Working Paper Series, Nr. 5, Kassel, 2014.
- [SG16] SoftGuide 2016: Kriterien zur Bewertung von Software, www.softguide.de/software-kriterien/, Stand: 13.05.2016..
- [SJK12] Schaeffler, H.; Ullrich C.C.J.; Kossahl, J.: Smart-Grid-Geschäftsmodelle für Verteilnetzbetreiber. In: VDE-Kongress. VDE Verlag GmbH, 2012.
- [Sc16] Schoormann, T.; Behrens, D.; Kolek, E.; Knackstedt, R.: Sustainability in Business Models – A Literature-Review-Based Design-Science-Oriented Research Agenda. In: European Conference on Information Systems (ECIS), Istanbul, 2016.
- [SK16] Schoormann T.; Knackstedt, R.: Rechtsvisualisierung mit Comictools – Vergleich und Weiterentwicklungsperspektiven. In (Schweighofer, E.; Kummer, F.; Hötendorfer, W.; Borges, G. Hrsg.): Internationales Rechtsinformatik Symposium, Salzburg, 2016.
- [Str15] Strahinger, S.: Modellbasierte Einführung Standardsoftware, Enzyklopädie der Wirtschaftsinformatik, www.enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/is-management/, Stand: 13.05.2016.

- [St15] Stryja C.; Schüritz, R.; Kühl, N.; Hottum, P.; Satzger, G.: Entwicklung eines Frameworks zur Beschreibung von Geschäftsmodellen für Elektromobilitätsdienstleistungen. In: 9. Internationale Energiewirtschaft, Wien, 2015.
- [SV03] Schütte, R.; Vering, O.: Erfolgreiche Geschäftsprozesse durch standardisierte Warenwirtschaftssysteme: Marktanalyse, Produktübersicht, Auswahlprozess. Springer-Verlag, 2013.
- [UN14] United Nations: Millenium Development Goals and Beyond 2015, www.un.org/millenniumgoals/bkgd.shtml, Stand: 13.05.2016.
- [Ve14] Veit, D.; Clemons, E.; Benlian, A.; Buxmann, P.; Hess, T.; Kundisch, D.; Leimeister, J. M.; Loos, P.; Spann, M.: Geschäftsmodelle - Eine Forschungsagenda für die Wirtschaftsinformatik. In: Wirtschaftsinformatik, S. 55-64, 2014.
- [Wi01] Wirtz, B. W.: Business Model Management – Design - Instrumente – Erfolgsfaktoren von Geschäftsmodellen. Gabler Verlag, Wiesbaden, 2001.
- [Wi03] Winter, R.: Modelle, Techniken und Werkzeuge im Business Engineering. In: Business Engineering. Springer Berlin Heidelberg. 87-118, 2003.
- [Wi16] Wirtz B. W.; Pistoia, A; Ullrich, S.; Göttel, V.: Business Models: Origin, Development and Future Research. In: Long Range Planning, Vol. 49 Issue 1, S. 36-65, 2016.
- [WW02] Webster, J.; Watson, R. T.: Analyzing the past to prepare for the future: Writing a. In: MIS quarterly, 26(2), 13-23, 2002.
- [ZB14] Zolnowski A.; Böhm T.: Formative Evaluation of Business Model Representations – The Service Business Model Canvas. In: Proceedings of the European Conference on Information Systems (ECIS), Tel Aviv, 2014.

Konzepte für Mietertickets zur Verknüpfung von Wohnen und Verkehr

Franz Lambrecht¹ und Prof. Dr. Carsten Sommer²

Abstract: 90% der ersten Wege eines Tages beginnen an der eigenen Wohnung [Mo08]. Die tagtägliche Entscheidung für das Verkehrsmittel hängt somit wesentlich von der Wohnsituation und dem Wohnumfeld ab. Der öffentliche Verkehr soll bei dieser Entscheidung stärker berücksichtigt werden. Dabei ergibt sich eine Win-Win-Situation für alle Beteiligten: für die Nutzer ein preiswerterer Zugang zum öffentlichen Personennahverkehr (ÖPNV), stabile Einnahmen für den Verkehrsdienstleister, ein besseres Stadtumfeld in der Kommune sowie eine Aufwertung des Kernproduktes und Kostensenkung bei der Stellplatzschaffung für die Wohnungsunternehmen [u.a. BRR08]. Trotz dieser guten Argumente werden entsprechende Angebote bisher nur selten in Mobilitätsstrategien berücksichtigt. Im Forschungsprojekt DieMo RheinMain wurden verschiedene Konzepte für Mietertickets entwickelt (freiwilliges Modell, Solidarmodell und Sharingmodell), die den ÖPNV auf Basis von integrierten Tarifprodukten und Vertriebswegen mit preisreduzierten Abonnements fest in das Angebot von Wohnungsbaugesellschaften integrieren. Bei der Entwicklung wurden alle beteiligten Akteure mit einbezogen, um die Praxistauglichkeit zu gewährleisten.

Keywords: Mieterticket, Verkehrsmanagement, Öffentlicher Verkehr, Multimodalität

1 Einleitung

Vier von fünf Wegen beginnen oder enden am Wohnstandort [Mo08]. Die tagtägliche Entscheidung, sich zu Fuß, mit dem Rad, öffentlichen Verkehrsmitteln oder dem Pkw auf den Weg zu machen, wird am Wohnstandort getroffen und hängt somit wesentlich von der Wohnsituation und dem Wohnumfeld ab [Ve14].

Unter wohnstandortbezogenem Verkehrsmanagement werden Maßnahmen verstanden, die für Bewohner bestimmter Quartiere oder Mieter bestimmter Wohnungsunternehmen in Zusammenarbeit von Wohnungsunternehmen und Verkehrsdienstleistern entwickelt werden und über die herkömmliche Dienstleistung hinaus einen spezifischen Nutzen für die Zielgruppen beinhalten [BK09].

Vielfältige Verkehrsangebote wie wohnstandortbezogenes Carsharing, radverkehrsbezogene Dienstleistungen und Infrastruktur für den Rad- und Fußverkehr, Verbesserung der Erreichbarkeit von ÖPNV-Haltestellen und Mietertickets sollen die Bewohner

¹ Universität Kassel, Fachgebiet Verkehrsplanung und Verkehrssysteme, Mönchebergstr. 7, 34121 Kassel, franz.lambrecht@uni-kassel.de

² Universität Kassel, Fachgebiet Verkehrsplanung und Verkehrssysteme, Mönchebergstr. 7, 34121 Kassel, c.sommer@uni-kassel.de

motivieren, in Zukunft häufiger auf Alternativen zum eigenen Pkw zurückzugreifen. Trotz vieler guter Argumente werden entsprechende Angebote bisher jedoch nur selten in Mobilitätsstrategien berücksichtigt. [Bä09b]

Es handelt sich beim wohnstandortbezogenen Mobilitätsmanagement noch um ein neues, wenig verbreitetes Geschäftsfeld der Wohnungswirtschaft. Wohnungsunternehmen konzentrieren sich vorrangig auf den motorisierten Individualverkehr (MIV), wodurch sich die Frage der Gestaltung von Mobilitätsoptionen bisher überwiegend auf die Frage nach der Anzahl der Pkw-Stellplätze beschränkt. Zudem ist die Erhaltung und Verbesserung der Qualität des Kernproduktes Wohnung das Tagesgeschäft von Wohnungsunternehmen, wodurch Dienstleistungen fern des Kernproduktes nachrangig behandelt werden, solange sie nicht erkennbar zur Vermarktbarkeit der Wohnungsbestände beitragen. [Bä09b]

Das Ziel des vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Projektes „DieMo RheinMain“ war die Entwicklung von umsetzungsfähigen Konzepten für Mietertickets (preisreduzierte Abonnements für den öffentlichen Nahverkehr für Bewohner bestimmter Quartiere oder Mieter bestimmter Wohnungsunternehmen), um den öffentlichen Verkehr fest in das Portfolio von Wohnungsunternehmen zu verankern. Dabei wurden verschiedene Mietertickets für unterschiedliche Wohngebiete und Wohnsituationen entwickelt, die im weiteren Verlauf auch mit anderen wohnstandortbezogenen Verkehrsangeboten, wie Carsharing oder Pedelec-Verleih kombiniert werden können.

2 Methodik

Zunächst wurden umfassende Recherchen zu Verkehrsangeboten für Mieter durchgeführt und dabei insbesondere bestehende oder in der Vergangenheit existierende Angebote analysiert, bewertet und auf Übertragbarkeit hin untersucht (Vgl. Kap. 3). Dabei wurden insbesondere die Erfolgsfaktoren und Hemmnisse oder Einstellungsgründe bei den entsprechenden Angeboten betrachtet. Anhand von Erfolgsfaktoren, Voraussetzungen für die Umsetzungen und rechtlicher Vorgaben wurden erste nutzerspezifische Konzepte entwickelt, die anschließend zusammen mit den beteiligten Akteuren diskutiert wurden.

Dazu wurde ein Workshop mit Vertretern von Wohnungsunternehmen, Kommunen, Verkehrsdienstleistern und Mietern durchgeführt. Das Ziel ist die bessere Nutzung von Potenzialen bei der Umsetzung von mit Wohnen verknüpften Verkehrsangeboten und bei der Zusammenarbeit der Akteure sowie die unterschiedlichen Sichtweisen der beteiligten Akteure und mögliche Barrieren. Dabei diskutierten die jeweiligen Vertreter zunächst in getrennten Gruppen und danach im gemeinsamen Plenum unter anderem über Nutzen und Erwartungen an Mietertickets, sowie Hemmnisse und Rollenverteilungen.

Aus den Ergebnissen wurden Handlungsempfehlungen und verschiedene gebrauchstaugliche Konzepte für Mietertickets entwickelt. Die Konzepte sollen dabei unterschiedliche Rahmenbedingungen in Städten und Gemeinden, sowie bei den Wohnquartieren und der Mieterstruktur berücksichtigen, so dass ein Mieterticket in vielen Regionen eingeführt und mit weiteren Verkehrsangeboten, wie z.B. E-Carsharing, kombiniert werden kann. Die Konzepte wurden in einem zweiten Workshop den Akteuren vorgestellt und auf Praxistauglichkeit untersucht.

3 Mietertickets

Bei Mietertickets handelt es sich um ein wohnungsbezogenes Verkehrsangebot im Bereich des öffentlichen Nahverkehrs, bei dem Mieter eines bestimmten Wohnungsunternehmens oder Bewohner eines bestimmten Wohngebiets preisreduzierte Abonnements des öffentlichen Nahverkehrs als fakultatives Angebot nutzen können. In bestimmten Wohngebieten kann auch ein Siedlungsticket mit obligatorischem Charakter angeboten werden [Go07]. Die einfachsten (und bisher am häufigsten realisierten) Angebote, insbesondere im Wohnungsbestand, beruhen auf dem Prinzip der Freiwilligkeit. Der Vertrieb kann entweder über den Verkehrsdienstleister direkt oder über das Wohnungsunternehmen erfolgen.

Das freiwillige Modell basiert auf dem Prinzip des Großkundenrabattes: Das Wohnungsunternehmen tritt als Vermittler auf und erwirbt beim Verkehrsdienstleister eine größere Menge an Tickets zum Großkundenrabatt, den sie an ihre Mieter weitergibt [Bä09a].

Solidarmodelle sind Modelle, bei denen ein Ticketerwerb für eine festgelegte Nutzergruppe obligatorisch ist. Sie basieren auf einem Semesterticket- oder Jobticketmodell und beinhalten im Regelfall eine Mischkalkulation zwischen Viel- und Wenignutzern des ÖPNVs. Aufgrund größerer und planbarer Nutzerzahlen kann somit eine Jahreskarte mit starker Ermäßigung angeboten werden. Die Kosten des solidarischen Mietertickets können direkt in die monatliche Kaltmiete integriert und darüber erhoben werden, so dass keine zusätzlichen Aufwände für die Mieter entstehen.

3.1 Anforderungen

Mietertickets sind besonders geeignet für Wohngebiete, die gut an das ÖPNV-Netz angeschlossen sind. Abhängig vom individuellen Mobilitätsverhalten und der Frage der Erreichbarkeit von individuellen Zielen mit dem öffentlichen Nahverkehr, sind sie aber auch für andere Standorte geeignet. In Zusammenarbeit mit den örtlichen Verkehrsunternehmen lässt sich auch die ÖPNV-Bedienung (durch Taktverdichtung, Quartiersbusse etc.) ebenso verbessern wie die Erreichbarkeit und Ausstattungsqualität der Haltestellen in den Wohngebieten. Weiterhin sind ein guter Kundenservice mit einfacher Erwerbsmöglichkeit und eine zuverlässige Fahrgastinformation für ein Mieterticketangebot essentiell. [Bä09a]

3.2 Wirkungen

Mit der Schaffung von Mieterticketangeboten kann der Mieter durch entsprechende Angebote in den Genuss preiswerterer ÖPNV-Tickets gelangen. Weiterhin erlangt der Mieter eine größere Wahlmöglichkeit bei der Verkehrsmittelwahl sowie einen besseren Zugang zum Verkehr. Bei der Wohnumfeldgestaltung ergeben sich durch entsprechende Verkehrsangebote Potenziale bei der Verkehrsreduzierung und beim Einsparen von Stellplätzen. Wohnungsunternehmen werten ihr Kernprodukt „Wohnung“ durch die neuen Angebote auf. Für den Verkehrsdienstleister bietet sich ebenfalls die Möglichkeit, sein Produkt aufzuwerten und attraktiver zu gestalten, so dass für beide Seiten die Möglichkeit besteht, neue Kunden zu gewinnen, bestehende Kunden besser und langfristiger zu binden und neue Vertriebsstrukturen aufzubauen. [Go07]; [DR03]. Weiterhin kann der Verkehrsdienstleister durch eine Mindestabnahme von Mietertickets mit festen Einnahmen kalkulieren.

3.3 Umsetzung

Die Umsetzung von Verkehrsangeboten für Mieter erfordert die intensive Kooperation aller am Vorhaben Beteiligten: Wohnungsunternehmen, Verkehrsdienstleister, Mieter und Kommunen. Bei bereits umgesetzten Verkehrsangeboten für Mieter hat sich gezeigt, dass einzelne Maßnahmen Zeit brauchen, um von den Bewohnern angenommen und geschätzt zu werden [Bä09b]. Aus diesem Grund ist auch die stetige Begleitung und Versorgung der potentiellen Nutzer mit Informationen von entscheidender Wichtigkeit.

Eine Akzeptanz umweltfreundlicher Verkehrsmittel als Alternative zum eigenen Pkw ist nur dann erreichbar, wenn diese direkt am Wohnstandort verfügbar und ohne Barrieren zugänglich sind. Weiterhin müssen die Angebote verständlich sein und eine einfache Handhabung aufweisen [Bä09b]; [DR03].

Nicht alle Maßnahmen können von den Wohnungsunternehmen alleine organisiert werden. Ein Engagement der Mieter trägt entscheidend zum Erfolg bei. Die Angebote sollten regelmäßig an die Bedürfnisse der Bewohner angepasst werden. Dies geschieht durch eine enge Zusammenarbeit mit den Mietern. Der notwendige Lernprozess sollte nicht unterschätzt werden. [Mü03]

3.4 Vorgehen

Ein standardisiertes Vorgehen für die Implementierung eines Tickets für Mieter oder Bewohner gibt es bisher nicht und ist aufgrund sehr unterschiedlicher Rahmenbedingungen in Städten und Quartieren kaum möglich. Wenn ein Wohnungsunternehmen seinen Kunden ein solches Produkt anbieten will, müssen in der Regel folgende Punkte geprüft und bearbeitet werden [Bä09a]:

- Es gilt zunächst, den richtigen Ansprechpartner beim Tarifverantwortlichen zu identifizieren.
- Die vorhandenen Ticketangebote inklusive spezieller Konditionen für Schüler, Studenten, Senioren und Großkunden sowie die damit verbundenen Leistungen müssen überprüft werden.
- Auswahl eines für das geplante Mieterticket passenden Tickets; sofern verhandelbar, können zusätzliche Leistungen wie Übertragbarkeit oder die Möglichkeit der Fahrradmitnahme, vereinbart werden.
- Abschluss eines Kooperationsvertrages zwischen Wohnungsunternehmen und Verkehrsdienstleister.
- Gezielte Kommunikationsmaßnahmen an alle angesprochenen Haushalte, die das neue Produkt und seine Vorteile vorstellen. Zusätzlich sollte ein Antragsformular enthalten sein.

3.5 Umsetzungsbeispiele

In Bielefeld wurde von der Bielefelder Gemeinnützige Wohnungsbaugesellschaft (BGW) die Einführung eines Mietertickets in Zusammenarbeit mit der moBiel GmbH (Stadt Bielefeld) initiiert, auch die Wohnungsgesellschaft freie Scholle beteiligte sich an dem Projekt. Seit 2005 ist das Mieterticket für alle Preisstufen und mit weiteren Zusatzleistungen erhältlich. Dabei handelt es sich um das Großkundenangebot von moBiel, dass für eine Mindestabnahme von 100 Tickets einen 10%igen Rabatt vorsieht [Go07]; [BGW16].

Ein ähnliches Mieterticket bietet bereits seit 2003 die VBW Bauen und Wohnen GmbH ihren Mietern in Bochum/Deutschland an. Dabei handelt es sich um ein spezielles Großkundenabonnement der VBW, wodurch den Mietern die Tickets (im Jahresabo, alle Preisstufen) mit einem Preisnachlass von 11,5% angeboten werden. Dabei kann jedes Haushaltsmitglied ein nicht übertragbares Mieterticket direkt beim Verkehrsdienstleister erwerben. Trotz der Preisvorteile muss das Angebot intensiv beworben werden. Im Jahr 2013 nutzten rund 1250 Mieterinnen und Mieter das Angebot. [BS13]; [VB15]

In Zürich bietet die Genossenschaft Zurlinden für die Mieter des neuen Wohnparks Sihlbogen pro Haushalt ein vergünstigtes Abo des Zürcher Verkehrsverbundes an. Dieses ist bereits in der Wohnungsmiete enthalten und wird in Form eines Gutscheins im Wert eines Jahresabonnements für den öffentlichen Verkehr ausgegeben. Als begleitende Maßnahme wurde auch die Zahl der Pkw-Stellplätze auf das gesetzliche Minimum reduziert, Pkw-Besitz ist für die Mieter nur nach Sondergenehmigung erlaubt. Erkenntnisse über die Wirkungen sind noch nicht bekannt [Zu13].

4 Workshops mit Akteuren

Um Potenziale beim wohnstandortbezogenen Mobilitätsmanagement besser nutzen zu können, wurden Akteure der Wohnungswirtschaft, der Kommunen, der Verkehrsdienstleister und der Mieterverbände zum gemeinsamen Informations-, Meinungs- und Ideenaustausch eingeladen. Dabei wurde über folgende Fragen diskutiert:

- Welche Erwartungen haben Sie an mit Wohnen verknüpfte Verkehrsangebote?
- Welchen Nutzen versprechen Sie sich davon?
- Welche Anforderungen haben Sie an solche Angebote?
- Welche Hemmnisse sehen Sie bei der Einführung und Nutzung solcher Angebote?
- Wer sollte mit welchen Aufgaben aktiv werden?

Als Ergebnisse des Workshops erwarten sich die Wohnungsbaugesellschaften Vorteile bei der Vermarktung ihrer Wohnungen, betonten aber auch, dass das Thema Verkehr nicht ihre originäre Aufgabe sei und Fragen der wohnstandortbezogenen Mobilität daher von geringerer Priorität sind. Sie sahen vor allem die Verantwortung bei der (kommunalen) Politik, die die Rahmenbedingungen schaffen und entsprechende Verkehrskonzepte mit Möglichkeiten zur Stellplatzreduktion fördern sollte.

Für Mieterinnen und Mieter können Angebote interessant sein, wenn dadurch bei gleichbleibendem oder gesteigertem Komfort Kosten (z.B. für einen eigenen Stellplatz) gespart werden können. Wichtig sind umfassende Informationen, die dauerhafte und verlässliche Verfügbarkeit der Angebote sowie ein einfacher Zugang mit Möglichkeit zum unkomplizierten Ausprobieren.

Die Kommunen erwarten sich in erster Linie eine Einsparung von Pkw-Stellplätzen und die Vermeidung von Kfz-Verkehr. Dabei sind aber auch Schwierigkeiten mit möglicher geringer Nachfrage sowie Unklarheiten bei der Gegenfinanzierung entsprechender Angebote und der Umsetzungsmöglichkeit in Bestandsgebieten zu erwarten.

Die Verkehrsdienstleister erwarten sich von Mietertickets mehr Nutzer ihrer Dienstleistungen und eine generelle Stärkung des Umweltverbundes. Ein einfacher Zugang durch verknüpfte Angebote mit nur einem Ansprechpartner für alle Angebote und Barrierefreiheit wird dabei als essentiell angesehen. Datenhoheit und Datenschutz, wirtschaftliche Zwänge sowie eine unklare Tarifstruktur können dabei als Probleme bei der Einführung auftreten.

Die Teilnehmer waren sich einig, dass ein vernetztes Angebot, mit mehreren Dienstleistungen (z.B. ÖPNV-Zeitkarte und E-Carsharing) über einen gemeinsamen Zugang, den größten Nutzen erzeugt und daher bei der Konzeptentwicklung von Mietertickets berücksichtigt werden sollte. Sowohl die Wohnungsgesellschaften als auch die Kommunen sehen sich dabei in der Vermittlerrolle zwischen den Mietern und den Verkehrsdienstleistern. Die Entwicklung von Rollenmodellen, in denen die

Verantwortlichkeiten der Akteure und Schnittstellen untereinander dargestellt werden, wird als wichtiger Schritt für eine Umsetzung gesehen, ebenso eine intensive Kooperation aller Beteiligten. Die Unterschiede zwischen dem städtischen und ländlichen Raum und die jeweiligen Rahmenbedingungen müssen berücksichtigt werden. Offen blieb dabei die Frage, ob in ländlichen Räumen überhaupt Mietertickets umgesetzt und wirtschaftlich betrieben werden können.

In einem zweiten Workshop wurden den teilnehmenden Akteuren die vorläufigen Konzeptentwürfe vorgestellt. Dabei wurde u.a. darüber diskutiert, ob die vorgestellten Konzepte (Vgl. Kap. 5) anwendungsfähig sind, die rechtlichen Belange ausreichend berücksichtigt wurden, die Rollen klar verteilt sind und die Finanzierung gesichert ist. Die Wohnungsunternehmen stehen dabei ihre angedachte Rolle als Koordinator kritisch gegenüber, da sie die Aufgabe erhalten, Verantwortlicher und Ansprechpartner für Dienstleistungen zu werden, für die sie bisher nicht zuständig sind und dadurch zusätzliche Mitarbeiter in diesen Kompetenzen ausbilden und bereitstellen müssten. Weiterhin werden die Haftung und das finanzielle Risiko für die Wohnungsunternehmen kritisch gesehen. Neben dem Nutzen für die jeweiligen Akteure standen auch die Kosten im Fokus. Insbesondere die zusätzlichen Personalkosten sollen gering gehalten werden, um das Konzept langfristig tragbar zu gestalten. Ungeklärt blieb hingegen die Frage, wie das Angebot für alle beteiligten Akteure greifbar wird und wer den ersten Schritt dorthin unternimmt.

5 Konzepte für Mietertickets

Basierend auf den Ergebnissen der Bestandsaufnahme, bereits umgesetzter Konzepte und des Workshops mit Akteuren wurden die folgenden Szenarien für Mietertickets entwickelt.

5.1 Freiwilliges Modell

Das freiwillige Modell basiert ähnlich wie beim Jobticket auf dem Prinzip des Großkundenrabattes: Das Wohnungsunternehmen tritt als Vermittler auf und erwirbt beim Verkehrsdienstleister eine größere Menge an Tickets zu einem Großkundenrabatt, den sie an ihre Mieter weitergibt. Die Mieter können das reduzierte Ticket bei Interesse beim Wohnungsunternehmen erwerben. Das freiwillige Mieterticket kann sowohl bei Neubauprojekten als auch bei Bestandsmietern eingeführt werden, ist nicht an ein umfassendes Verkehrskonzept gebunden und kann auch auf längere Sicht aufrechterhalten werden. Es sollte allen Mietern des Wohngebietes oder des Wohnungsunternehmens zur Verfügung stehen. Das freiwillige Mieterticket richtet sich vor allem an Vielnutzer des ÖPNV, die keine Ermäßigungen erhalten oder an Gelegenheitsnutzer, für die sich die Anschaffung einer Monats- oder Jahreskarte zum Normalpreis finanziell nicht rechnet. Die Vorteile des Angebots sind von der Ticketanzahl und einem Mindestwerb von Tickets abhängig. Dadurch eignet sich

dieses Modell nicht für Einzelvermieter oder sehr kleine Wohnungsgesellschaften mit wenigen Wohnungen.

Der Datenschutz ist sichergestellt, da das Wohnungsunternehmen als Vertriebspartner auftritt und ein Datenaustausch nur zwischen Wohnungsunternehmen und Mieter auftritt, wobei keine zusätzlichen zum Mietverhältnis notwendigen Daten erhoben werden. Eine Weitergabe der Daten des Mieters an den Verkehrsdienstleister oder Dritte findet nicht statt. Bei einer Übertragbarkeit des Tickets bestehen rechtliche Bedenken, da hier auch Nichtmieter des Wohnungsunternehmens an ein Ticket zu den vergünstigten Konditionen gelangen können [Be05].

Der Aufbau eines funktionierenden freiwilligen Mietertickets erfordert eine intensive Kooperation zwischen Verkehrsdienstleister und Wohnungsunternehmen, bisweilen auch den Kommunen und Mietern. Die Kommune tritt als koordinierender Partner auf und sollte entsprechende Angebote fördern. Mit den Mietern ist vorrangig die Wirkung und Nutzung des Angebots zu besprechen. Zunächst sind durch das Wohnungsunternehmen und den Verkehrsdienstleister konkrete Vereinbarungen zum Angebot zu treffen. Dies betrifft vor allem die Art des angebotenen Tickets, die Mindestanzahl, Zusatzleistungen und den zu vereinbarenden Großkundenrabatt. Das Wohnungsunternehmen übernimmt den Vertrieb und die Koordination des Angebotes. Jeder interessierte Mieter kann dabei das Mieterticket beim Wohnungsunternehmen erwerben. Anhand der gestellten Anträge erwirbt das Wohnungsunternehmen die Tickets gesammelt beim Verkehrsdienstleister. Das finanzielle Risiko bei Zahlungsausfällen und bei Verfehlung der Mindestabnahmezahl trägt das Wohnungsunternehmen, ebenso fallen hier finanzielle Aufwände im Vertrieb des Tickets und bei der Bewerbung des Angebots an. Ein finanzieller Ausgleich kann hier zum Teil durch die Weitergabe eines geringeren Rabattes an die Mieter erfolgen, als mit dem Tarifverantwortlichen erzielt wurde. Das Angebot muss intensiv beworben werden, sowohl durch Verkehrsdienstleister als auch Wohnungsunternehmen.

Für den Erfolg eines freiwilligen Mietertickets werden begleitende Maßnahmen und eine Einbettung in ein Verkehrskonzept nicht vorausgesetzt, können aber in Kombination mit dem Mieterticket zur gegenseitigen Attraktivitätssteigerung beitragen. Insbesondere können den Mieterticketinhabern Sonderkonditionen (z.B. Erlass der Anmeldegebühr) bei den begleitenden Maßnahmen angeboten werden.

Zu den möglichen Maßnahmen zählt das stationsgebundene Carsharing (mit Elektrofahrzeugen). Hier sollten Wohnungsunternehmen, Carsharing-Unternehmen und die Kommunen intensiv zusammenarbeiten. Bei Neubauvorhaben sollte dies bereits früh im Planungsprozess geschehen, um infrastrukturelle Voraussetzungen zu schaffen. Im Wohnungsbestand müssen Flächen für die Einrichtung des Angebotes gefunden werden. Für die Mieter sollten spezielle Angebotskonditionen verhandelt werden. Die Finanzierung der Stellplätze erfolgt über Ausgleichszahlungen durch das Carsharing-Unternehmen oder über Einsparungen bei den Stellplätzen für die Bewohner.

Als weitere begleitende Maßnahme bietet sich zum Beispiel ein stationsgebundenes

Fahrradverleihsystem mit E-Bikes an, dass exklusiv für die Bewohner der teilnehmenden Wohnquartiere eingerichtet wird. Die Leifahräder sollten in abgeschlossenen Räumlichkeiten angeboten werden. Zugang und Ausleihe können dabei mit einer Mobilitätskarte erfasst werden, die gleichzeitig als Mieterticket fungieren kann. Das Wohnungsunternehmen ist auch für die Bereitstellung der notwendigen Flächen und die Sicherstellung des Zugangs verantwortlich. Die Kommune nimmt begleitend am Planungsprozess teil.

5.2 Solidarmodell

Das Solidarmodell ist ein Modell, bei dem ein ÖPNV-Ticketserwerb für eine festgelegte und abgeschlossene Gruppe von Mietern oder Haushalten obligatorisch ist. Voraussetzung für ein Solidarmodell ist ein einfacher und sehr guter Zugang zum öffentlichen Nahverkehr. Die Angebotskonditionen müssen zwischen Verkehrsdienstleister und Wohnungsunternehmen verhandelt werden und beinhalten eine Mischkalkulation zwischen Viel- und Wenignutzern des ÖPNVs. Die Kosten des Mietertickets werden mit der Monatsmiete erhoben, wodurch für den Mieter keinerlei Aufwände für die Anschaffung des Tickets entstehen. Es ist nur bei Neuvermietungen oder kompletten quartiersbezogenen Neuvermietungen umsetzbar, da ein obligatorisches Mieterticket nach rechtlichem Grundsatz „pacta sunt servanda“ (lat.: Verträge müssen eingehalten werden, Vgl. § 145 BGB) nachträglich nicht einseitig in bestehende Mietverträge eingefügt werden darf. Eine sukzessive Einführung bei Neuvermietungen im Wohnungsbestand ist nicht möglich, da unterschiedliche Voraussetzungen und Verpflichtungen bei Bestands- und Neumieter als Diskriminierung gelten.

Die Integration des Solidartickets in allgemeine Mietverträge besteht über die gemeinsame Abrechnung mit der Grundmiete. Hier kann das Wohnungsunternehmen zur Kaltmiete einen zusätzlichen Beitrag aufschlagen. Im Gegenzug wird dem Mietvertragsnehmer das Mieterticket kostenlos überlassen. Dies wird nicht als zusätzliche Klausel in den Mietvertrag übernommen werden, sondern als freiwillige Leistung des Wohnungsunternehmens angeboten. Dadurch besteht rechtssicher eine einseitige Kündigungsmöglichkeit, falls das Angebot beendet werden soll. Eine Ticketübertragung soll nur auf Haushaltsmitglieder möglich sein, da sonst auch einzelne Nichtmieter des Wohnungsunternehmens an das Ticket gelangen können, wodurch Diskriminierungsgefahr besteht [Be05].

Der Aufbau eines funktionierenden Solidartickets erfordert eine intensive Kooperation zwischen Verkehrsdienstleister, Wohnungsunternehmen und Kommunen, bisweilen auch den Mietern. Da die Einrichtung eines solidarischen Mietertickets im Rahmen von neuen Wohnprojekten entsteht, ist es frühzeitig in die Planungen des Wohnquartiers einzubeziehen. Dabei sind durch das Wohnungsunternehmen und den Verkehrsdienstleister konkrete Vereinbarungen zum Tarifprodukt zu treffen. Zusatzleistungen, wie z.B. Fahrradmitnahme, sind sinnvoll. Das Wohnungsunternehmen übernimmt den Vertrieb und die Koordination des Angebotes. Der Mieter bekommt das

Mieterticket vom Wohnungsunternehmen ausgehändigt und gibt es nach Auszug an das Wohnungsunternehmen zurück. Das Wohnungsunternehmen erwirbt die Tickets im Vorfeld gesammelt beim Verkehrsdienstleister. Das Mieterticket ist vor allem für Wohngebiete attraktiv, bei denen eine hohe Affinität zum öffentlichen Verkehr vorherrscht und der Großteil der zukünftigen Bewohner nicht bereits andere, ermäßigte Tickets besitzt.

Die Einbindung des solidarischen Mietertickets in ein wohnstandortbezogenes Verkehrskonzept kann erheblich zum Erfolg des Mietertickets und des Verkehrskonzeptes beitragen. Hier ist die Kommune für verkehrliche und bauordnungsrechtliche Rahmenbedingungen verantwortlich. Dazu muss eine Möglichkeit zur Reduktion von Stellplätzen, sowie eine sehr gute Anbindung an den öffentlichen Nahverkehr geschaffen werden.

Als erfolgversprechende Begleitmaßnahmen zählen zum Beispiel wie beim freiwilligen Modell (Vgl. Kap. 5.1) die Einrichtung von stationsgebundenem Carsharing (mit Elektrofahrzeugen) sowie ein stationsgebundener Fahrradverleih mit E-Bikes.

5.3 Sharingmodell

Ticketsharing im ÖPNV orientiert sich an bestehenden Carsharing-Modellen. Dabei werden festgelegten Nutzergruppen (z.B. Bewohner eines Wohnquartiers oder auch einzelner Wohngebäude) übertragbare Netzkarten für den ÖPNV zur Verfügung gestellt, die von den Bewohnern nach Bedarf ausgeliehen werden können und mit allen Nutzern geteilt werden. Der Zugang ist auf die Bewohner der teilnehmenden Wohngebäude beschränkt und erfolgt über eine Zugangsmöglichkeit im Gebäude. Beim Ticketsharing können die Preisvorteile einer vergünstigten Netzkarte durch die Kostenverteilung auf alle Nutzer auch bei nur gelegentlichen Fahrten genutzt und somit Kosten gespart werden. Das Ticketsharing-Modell richtet sich vor allem an Nicht- oder Wenignutzer des ÖPNV, für die sich die Anschaffung einer eigenen Monatskarte finanziell nicht rechnen würde. Es ist für die Mieter dann attraktiv, wenn eine hohe Ticketverfügbarkeit vorherrscht. Der Erfolg und die Vorteile des Angebots sind nicht von der Gesamtticketanzahl oder einem Mindesterverb von Tickets abhängig. Dadurch eignet sich dieses Modell besonders bei einer gemischten Mieterstruktur sowie auch für einzelne Wohngebäude und Einzelvermieter. Es ist relativ einfach umsetzbar, sowohl im Bestand als auch bei Neuvermietungen. Die Einrichtung und Betreibung dieses Ticketmodells kann ohne Beteiligung der Kommune oder des Verkehrsdienstleisters erfolgen.

Aus rechtlichen Gesichtspunkten gibt es nur geringe bis keine Hürden bei der Einführung des Ticketsharing. Für alle Mieter der am Angebot beteiligten Häuser ist ein diskriminierungsfreier und sicherer Zugang sicher zu stellen. Weiterhin ist eine Abrechnung der Ticketnutzung durch das Wohnungsunternehmen zu gewährleisten, hierbei ist sowohl eine nutzungsabhängige Kostenverteilung als auch eine gleichmäßige Kostenverteilung auf alle teilnehmenden Mieter eines Hauses möglich. Eine Verpflichtung der Mieter zur Teilnahme am Ticketsharing ist nicht möglich (Vgl.

Solidarmodell). Aus diesem Grund ist für die teilnehmenden Bewohner ein exklusiver Zugang zu gewährleisten.

Zunächst muss im Wohngebäude durch den Vermieter eine Zugangsmöglichkeit installiert und betrieben werden, die den Zugang sowie die Ausleihe und Erfassung des Nutzers regelt. Je nach zu erwartender Ticketnutzung und Anzahl der teilnehmenden Haushalte muss eine entsprechende Anzahl an übertragbaren Netzkarten vom Vermieter erworben und zur Verfügung gestellt werden. Bei der Tarifgestaltung für die Mieter ist darauf zu achten, dass einerseits der Preis für die Dauernutzung unattraktiv hoch ist, um einen Missbrauch des Tickets als Monatskartenersatz und Dauernutzung durch einzelne Mieter zu verhindern und andererseits die Preise pro Ausgang geringer sind als die Preise für Einzelfahrkarten. Somit bietet sich hier ähnlich zum Carsharing ein zeitabhängiger Tarif mit oder ohne Grundgebühr an. Der Vermieter hat dabei die Möglichkeiten, die Kosten für die Anschaffung und Betrieb des Angebots auf die Nutzer umzulegen. Damit sich das Angebot für den Vermieter und für die Mieter lohnt, ist eine gewisse Mindestanzahl als teilnehmenden Haushalten pro Wohngebäude nötig. Für Wohnungsunternehmen mit mehr teilnehmenden Wohnhäusern gibt es keine Mindestanzahl an Teilnehmern pro Wohngebäude, da hier Wohngebäude mit mehr Teilnehmern solche mit weniger Teilnehmern ausgleichen und sich hier mehrere Wohngebäude ein Ticket teilen können.

Für den Erfolg des Ticketsharings werden begleitende Maßnahmen und eine Einbettung in ein Verkehrskonzept nicht vorausgesetzt, können aber in Kombination mit dem Mieterticket zur gegenseitigen Attraktivitätssteigerung beitragen. Insbesondere können den Mieterticketinhabern Sonderkonditionen (z.B. exklusiver Zugang, Erlass der Anmeldegebühr) bei den begleitenden Maßnahmen angeboten werden.

Als wesentliche begleitende Maßnahme bietet sich das quartiersbezogene Carsharing an: Das Wohnungsunternehmen stellt dem Carsharing-Unternehmen einen Stellplatz für ein Fahrzeug zur Verfügung, das Carsharing-Unternehmen sorgt dabei für einen exklusiven Zugang und spezielle Angebotskonditionen für die Mieter. Das Angebot eignet sich vor allem bei Neubauvorhaben, ist aber auch im Bestand realisierbar. Zum Aufbau des quartiersbezogenen Carsharing muss eine Kooperation zwischen Wohnungsunternehmen und einem Carsharing-Unternehmen getroffen werden. Bei Neubauprojekten und einer Einbettung in ein Verkehrskonzept sind dabei auch die Kommunen früh in die Planungen einzubeziehen, die eine Möglichkeit zur Reduktion der Stellplätze schaffen sollen. Dadurch können Finanzmittel im Rahmen der Stellplatzschaffung eingespart werden. Weiterhin kann zum Beispiel auch mit quartiersbezogenem Fahrradverleih (Vgl. Kap. 5.1) die Grundlage für ein multimodales Sharingangebot (Ticketsharing, Carsharing und Bikesharing) mittels einer gemeinsam genutzten Zugangskarte geschaffen werden.

6 Fazit

Mit der Schaffung von Mietertickets ergibt sich eine Win-Win-Situation für alle Beteiligten, wird aber gegenwärtig trotz entsprechender Vorteile kaum in Mobilitätsstrategien betrachtet. Die entwickelten Modelle für Mietertickets, freiwilliges Modell, Solidarmodell und Sharingmodell, berücksichtigen unterschiedliche Gegebenheiten in unterschiedlichen Wohngebieten, Bestand oder Neubau sowie verschiedene weitere Einflussfaktoren. Damit bietet sich die Möglichkeit, ein Mieterticket in vielen Wohngebieten einzuführen. Für die Akteure ist insbesondere wichtig, dass ein konkreter Nutzen für alle Beteiligten entsteht, die Kosten im laufenden Betrieb gering gehalten werden und die Zuständigkeiten unter den jeweiligen Akteuren klar verteilt sind.

Modell	Freiwilliges Modell	Solidarmodell	Sharingmodell
Funktionsweise	Mindestabnahme an Tickets und direkte Rabattweitergabe optional	Mischkalkulation zwischen Viel- und Wenignutzern obligatorisch	Teilen einer übertragbaren Netzkarte durch Nutzergruppe optional
Ticketerwerb	gering	hoch	mittel bis sehr hoch
Rabatt gegenüber Jahresabopreis	alle Mieter des Wohnungsunternehmens	alle Mieter des Wohnungsunternehmens oder Wohngebietes	Mieter einzelner Wohngruppen,
Zielgruppe	Bestand und Neubau stationsgebundenes Carsharing mit offenem Zugang	nur Neubau stationsgebundenes Carsharing mit offenem Zugang	Bestand und Neubau Wohnstandortbezogenes Carsharing
Einführung	offener Pedelec-Verleih	Wohnstandortbezogenes Carsharing	offener Pedelec-Verleih
Weitere Maßnahmen (Auswahl)		offener Pedelec-Verleih	Wohnstandortbezogener Pedelec-Verleih

Tab. 1: Übersicht über die verschiedenen Mieterticketmodelle

Die Mietertickets bilden das Grundgerüst bei der Verknüpfung von Wohnen und Verkehr. Sie lassen sich gut mit weiteren Angeboten wie Carsharing mit Elektrofahrzeugen und Pedelec-Verleih kombinieren, über einen gemeinsamen Zugang (z.B. Mobilitätskarte) können Synergieeffekte genutzt und der Vertrieb sowie der Nutzerzugang vereinfacht werden. Ein großes Potenzial und die größtmögliche Flexibilität kann dabei ein multimodales Sharingangebot für ÖPNV, MIV und Rad bieten.

Die existierenden Beispiele und Initiativen in der Vergangenheit haben gezeigt, dass Mietertickets sich gut als wohnungsbezogenes Mobilitätsangebot anbieten und positive Effekte für alle Beteiligten erzielen, wenn notwendige Rahmenbedingungen wie ein guter und einfacher Zugang zum ÖPNV geschaffen werden. Die größten

Erfolgsaussichten versprechen sich die Autoren bei Einbettung in ein integriertes Verkehrskonzept, das beispielsweise auch infrastrukturelle Maßnahmen (wie die Anbindung ans regionale und überregionale Radwegenetz) oder ergänzende Services (u.a. Fahrgastinformationen in Echtzeit, Fahrplanaushänge, Fahrplan- und Verfügbarkeitsapps) enthält und dadurch für eine Attraktivitätssteigerung des Mietertickets sorgt. Aber es müssen die Anforderungen und Hemmnisse beachtet werden, um einen langfristigen Erfolg zu erzielen. Mit einer guten und intensiven Kooperation zwischen allen Akteuren sind Mietertickets eine vielversprechende Maßnahme des Mobilitätsmanagements.

Literaturverzeichnis

- [Bä03] Bäumer, D. (2003). Mobilitätsservice für die VBW Bauen und Wohnen GmbH. In: U. Reutter & R. Penczek (Hrsg.), Mieterticket & Co. – Erfolgsfaktoren siedlungsbezogener Mobilitätsdienstleistungen, 8–14. Dortmund: Institut für Landes- und Stadtentwicklungsforschung und Bauwesen des Landes Nordrhein-Westfalen.
- [Bä09a] Bäumer, D. (2009). Preiswerte ÖPNV-Tickets für Mieter und Bewohner bestimmter Quartiere oder Wohnungsbestände. http://add-home.eu/docs/ILS_ticket_for_tenants_guidelines_DE.pdf, Abruf: 13.03.2015.
- [Bä09b] Bäumer, D. et.al. (2009). Mobility meets housing. http://add-home.eu/docs/ADDDHOME_brochure_mobility_meets_housing_germany_final.pdf, Abruf: 16.03.2015.
- [Be05] Berliner Mieterverein (2005). Wohnumfeld und Autos. <http://www.berliner-mieterverein.de/magazin/online/mm0305/030514.htm>, Abruf: 06.05.2015.
- [BGB15] BGB (2015). Bürgerliches Gesetzbuch. Berlin: Bundesministerium für Justiz und Verbraucherschutz (Hrsg.). Letzte Änderung: 20.11.2015.
- [BGW16] BGW Bielefeld – Mieterticket. http://www.bgw-bielefeld.de/fileadmin/downloads/Mieterticket_neu.pdf. Abruf: 13.02.2016
- [BK09] Bäumer, D. & Köllinger, C. (2009). Einführung – Das Projekt Add-Home. In: Reutter, U., et.al. (Hrsg.), Wohnstandortbezogenes Mobilitätsmanagement – Projektdokumentation des EU-Projektes Add Home, 2–3. Dortmund: Institut für Landes- und Stadtentwicklungsforschung und Bauwesen des Landes Nordrhein-Westfalen.
- [BRR08] Bäumer, D., Reutter, U. & Reudenbach, L. (2008). Mobilität trifft Wohnen – Eine aussichtsreiche Begegnung. Dortmund: Institut für Landes- und Stadtentwicklungsforschung und Bauwesen des Landes Nordrhein-Westfalen.
- [BS13] Bäumer, D. & Stiewe, M. (2013). Klimaverträglich mobil in Zeichen des demografischen Wandels – wie wohnen Mobilität bestimmt. In: Proceedings REAL CORP 2013, 475 – 484. Rom.
- [DR03] Dittrich-Wesbuer, A. & Reutter, U. (2003). Von der Idee zur Umsetzung – Erfolgsfaktoren siedlungsbezogener Mobilitätsdienstleistungen. In: U. Reutter & R.

- Penczek (Hrsg.), Mieterticket & Co. – Erfolgsfaktoren siedlungsbezogener Mobilitätsdienstleistungen, 8–14. Dortmund: Institut für Landes- und Stadtentwicklungsforschung und Bauwesen des Landes Nordrhein-Westfalen.
- [Go07] Gogol, A. (2007). Evaluation zielgruppenspezifischer Mobilitätsdienstleistungen von Wohnungsunternehmen. Dortmund: Institut für Landes- und Stadtentwicklungsforschung und Bauwesen des Landes Nordrhein-Westfalen.
- [MB15] moBiel GmbH & Bielefelder Gemeinnützige Wohnungsgesellschaft mbH (2015). Bitte einsteigen – das BGW-Mieterticket für Bus und Bahn. <http://www.bgw-bielefeld.de/mieten-wohnen/lexikon-mieten-wohnen/m/mieterticket-fuer-bus-und-bahn.html>, Abruf: 13.11.2015.
- [Mo08] Mobilität in Deutschland (MID) (2008). Datenbasis der MID 2008. www.mobilitaet-in-deutschland.de.
- [Mü03] Müller, G. (2003). Von der Reiseauskunft zum Mieterticket...Ideen und Anwendungsfelder des siedlungsbezogenen Mobilitätsmanagements. In: U. Reutter & R. Penczek (Hrsg.), Mieterticket & Co. – Erfolgsfaktoren siedlungsbezogener Mobilitätsdienstleistungen, 23–30. Dortmund: Institut für Landes- und Stadtentwicklungsforschung und Bauwesen des Landes Nordrhein-Westfalen.
- [VB15] VBW (2015). Das Mieterticket der VBW Bauen und Wohnen in Bochum. <http://www.vbw-bochum.de/service/rund-ums-wohnen>, Abruf: 13.11.2015.
- [Ve14] Verkehrsclub Österreich VCÖ – Mobilität mit Zukunft (2014). Wohnen bewegt Menschen. In: VCÖ-Magazin. – 6 (2014), 5, 1. Wien.
- [Zu13] Zurlinden (2013). Der Sihlbogen in Zürich-Leimbach. http://www.bgzurlinden.ch/projekte/fileadmin/user_upload/Sihlbogen/Medien/Zuerich_2_14112013.pdf, Abruf: 06.05.2015.

”Need Tweets”: New Insights about Customer Needs from Micro Blog Data in the Field of E-mobility

Niklas Kühl¹ Marc Goutier²

Abstract: In order to design new services in e-mobility, the knowledge of customer needs is crucial. In this paper, we compare the results from a literature review about customer needs with the results from harvesting a modern data source: micro blog data.

Keywords: Customer Needs, E-Mobility, Literature Review, Micro Blog Data, Twitter

1 Introduction

The correct identification of customer needs is one of the main challenges for service providers [AP02]. As part of research, it plays an important role in different areas, namely *Needs Elicitation* [HD03], *Customer Requirement Analysis* [BL95], *Service Engineering* [BS06] and *Service Design Thinking* [St11]. While there is a large amount of publications on customer needs in time-tested domains, e.g. tourism [AH98] or software (engineering) [RJ00], the field of electric mobility (e-mobility³) is a rather new, emerging phenomenon which is still under development with high dynamics [BPS10]. As stated in [K114, St15], there is a lack of e-mobility services to accelerate the successful implementation and adoption of the new technology. In order to support the development of such services, it is important to identify customer needs in the field of e-mobility to translate them into new or improved services at a later point.

In the work at hand we aim at identifying customer needs in the field of e-mobility by leveraging a modern data source: micro blog data (e.g. Twitter). Recent studies show that people using social media tend to talk about technology-related topics [MLF15]. Therefore, we assume that a certain share of (potential) e-mobility customers have an affinity for social media which implies that they share their opinions and needs in social media networks. These statements enable us to identify new customer needs. Although this data is not representative, it allows to regard first-hand postings of (potential) customers, who expressed a need in an emotional situation like dissatisfaction or joy. This is an advantage compared to data from traditional surveys, where non-response bias is a major challenge [Gr06].

¹ Karlsruhe Institute of Technology (KIT), Karlsruhe Service Research Institute (KSRI), Engelstr. 11, 76131 Karlsruhe, Germany, kuehl@kit.edu

² Karlsruhe Institute of Technology (KIT), Karlsruhe Service Research Institute (KSRI), Engelstr. 11, 76131 Karlsruhe, Germany, marc.goutier@student.kit.edu

³ In order to allow for a uniform comprehension, we define *e-mobility* as a ”highly connective industry which focuses on serving mobility needs under the aspect of sustainability with a vehicle using a portable energy source and an electric drive that can vary in the degree of electrification.”([Sc15, p.9]).

The contribution of this work is twofold. First, it gives an overview about the current state of research of customer needs in the field of e-mobility. We propose a clustering pattern to structure the needs in four major need categories. Second, we use relevant micro blog data in the field of e-mobility in the German-speaking area, expressed over a time span of six months in 2015. These data provide us deep insights about customer needs on the subject of e-mobility. We compare the expressed customer needs with the current state of research and discuss differences.

2 Prerequisites

Before we are able to focus on the methodology, the study and the results, we have to outline the term *customer need* and relate to the used dataset as a prerequisite in order to lay a foundation for the remaining work.

Customer Needs Traditionally, the term *customer need* originates from the discipline of Marketing. Kotler and Armstrong suggest to split the term into three different categories: *need*, *want* and *demand* [KA01]. *Needs* cover—amongst others—the basic human requirements in life like security or food. Every need is necessary for a decent existence of a human being. A *want* is the form a need takes. It is shaped by culture and personality. In case a want is backed by the power to buy something, the want is called *demand*. When we regard our domain of e-mobility, we could imagine a human being who has the need of mobility. The need can take the form of an electric vehicle (EV), so someone wants an EV and demands (exemplary) a Tesla Model S. *Needs*, *wants* and *demands* can be translated into every other requirement. A *need* can be translated into a *demand* similar to the translation of a *want* into a *need*. Additionally, Harding et al. outline that a customer need can also be expressed as a *requirement* of a product or service [Ha01]. *Requirements* are often expressed by humans when a service or a product lacks on a certain attribute. Referring to our previous example, the person can express that he⁴ *requires* an EV with more airbags. In case there is an EV with more airbags than currently on the market, this person would buy that car. Therefore, the person has the *requirement* for an airbag, but his *requirement* is actually referred to the *need* of safety. *Requirements* can also be translated into *needs*, *wants* and *demands*—and since the terms can be easily translated, humans express their requirements in every of the four categories. Referring to our aim to quantify customer needs, there is no necessity to differentiate between the different characterizations. For simplicity, we use the terms *needs* or *customer needs* as thesauri for all four mentioned types *need*, *want* and *demand* as well as *requirement*.

Micro Blog Dataset The retrieval of our relevant micro blog data set is not part of this work and is only explained briefly. A more detailed description can be found in Kuehl et al. 2016, which illustrates an approach to automatically detect micro blog instances containing customer needs [KSS16]. As stated before, the domain for our dataset is e-mobility.

⁴ To ensure a steady reading flow in this work, we use only one gender. Based on the outcome of a coin-flip, we use male pronouns (he, his, him) when necessary. This always includes the female gender as well.

In addition to the domain, we also define a geographical area as well as the language. The first is important because we need a comparable area in the field of legislation, design of economy and living standard of the population to be not biased towards one specific group of people in the area. The second is important because different languages tend to have different semantic structures which would distort our analysis. With our decision to use only German tweets, we cover both of these additional conditions: German is only spoken in Germany, Austria and in parts of Switzerland which have comparable socio-economic preconditions.

As a source for micro blog instances we use Twitter since it is the most popular micro blogging service provider [De16]. We conduct the retrieval of relevant micro blog data by using the Twitter Streaming API. We collect every instance (tweet) which contains at least one word of a predefined keyword list. The list is reasoned on the opinion of professionals and popular EVs in Germany. It consists of eight German⁵ and five English⁶ generic terms which are supplemented by ten EVs vehicles⁷. From March 3rd to August 25th 2015, tweets were collected. To cover the time before March 3rd, IBM Insights for Twitter was used to retrieve older instances. 645,226 instances were acquired in total. The database consists of 86.3% instance from the six month period and only the remainder is from the time before these six months.

Based on the language information of Twitter, all non German instances are sorted out which reduces our dataset to 39,739 instances. Thereafter, *Descriptive Coding* reveals that there is only one conspicuous correlation: Tweets containing an URL also contain a need with the probability of only 3.64%. Since the remaining dataset comprises 91.5% instances with URLs, one achieves a great reduction and coincidentally loosing a significantly low share of relevant needs. After removing duplications in the dataset which result either in the multiple collection with the tools or the collection of re-tweets, the dataset has the amount of 2,396 possibly relevant instances. The last step is the labeling of all remaining tweets on whether or not they contain a customer need, which is done in a lab session by at least three participants per tweet. Finally, we end up with 332 remaining instances containing needs, which were only identified as such if at least 2 out of 3 labelers agreed on the tweet containing a need. This resembles the dataset of the work at hand, called "need tweets" or "instances". An exemplary tweet is shown in figure 1.



Fig. 1: Exemplary "Need tweet" in the field of e-mobility

⁵ e-tankstelle, eauto, elektroauto, elektrofahrzeug, elektromobilitaet, elektromobilität, ladesaeule, ladesäule

⁶ ecar, electric mobility, EV vehicle, e-mobility, emobility

⁷ bmw i3, egolf, eup, fortwo electric drive, miev, nissan leaf, opel ampera, peugeot ion, renault zoe, tesla model s

3 Methodology and Study

To identify customer needs in recent research, we have to conduct a literature review. We select relevant publications about needs in the field of e-mobility, determine their identified customer needs and cluster them in a bottom-up approach to summarize major need categories (3.1). In a second step, we assign every micro blog instance to one or more major need categories and use this allocation as a basis to analyze our given dataset by splitting the major categories into smaller clusters for a more detailed analysis (3.2). The comparison of the identified needs in the literature and the allocation of our dataset represents the third step of our results (4).

3.1 Identification and Clustering of Customer Needs from Literature

The identification of customer needs in recent research starts with the selection of relevant publications. There is no systematized approach in current need research which depicts every aspect of customer needs in every possible field of interest. Therefore, we limit our focus only on our domain of interest, e-mobility, to depict the current state of need research in this field. Our main database for the literature is *Google Scholar*. We use the application *Paperfinder* to search, filter and download literature [Sc15]. According to the principles of Randolph, we start with a keyword search based on the generic words "e-mobility" and "need" [Ra09]. The relevance of the found literature is assessed mainly by its content but we also evaluate the number of citations and the authors. Afterwards, we repeat this step with "electric mobility" and its German translation ("Elektromobilität"). The next step is to use the keywords and titles of the found literature as the basis for a new keyword search. Additionally, the bibliographies in the identified literature are also a source for the detection of relevant publications. We continue to send search requests with the keywords and parts of the title of the already found literature until we reach saturation, which means that we only receive already identified literature.

Moreover, we discuss our found literature with experts in the field of e-mobility. They help us to enrich our list of literature with publications which we did not find e.g. because the publication is in progress and does not appear in online databases. In total we determine a number of 38 papers, field studies and computer simulations. We look at each publication separately and roughly ascertain their identified needs. Afterwards, we exclude five publications from the time period 1981-2000 because we consider these papers as too old to reflect current research. We also exclude 15 papers about EV adoption to be not affected by a specific research method. Since we roughly ascertained the identified needs in every publication, we ensure that the excluded papers do not contain any fundamentally different needs compared to the needs in the remaining literature. Finally, our set consists of 18 publications⁸ from the period between 2010 and 2016, representing all kinds of studies, e.g. surveys, observations or computer simulations. The next step is to analyze the papers more precisely. We are especially interested in a clustering of needs by the researchers and which gradation of needs (e.g. needs, wants and demands) they use in their research.

⁸ [Hi11, EL12, Fr12, CGK10, CFM10, ABH12, PjL11, Gh13, Pr13, P114, FA12, Gö11, De12, Du13, Wi13, PG13, Fr15, SJF16]

We find neither information about the gradation of needs nor a uniform clustering pattern to summarize them. Based on our research, none of the selected publications mind the different aspects of needs⁹. Due to our best knowledge, we know of no publication clustering needs in e-mobility. Since a more global clustering of needs is crucial to find a pattern of e-mobility needs in research, we have to develop a clustering. We start clustering with the condensation of different expressions for one and the same need. We explain how this condensation works, e.g. on the need *retail price*: Based on the literature, we learn that the initial price a user has to pay to buy an EV is very relevant for him. When we look at the needs which are described in literature, some state that the customers demand a lower sales tax (e.g. [CGK10]), some found a request of subsidies of the government (e.g. [Hi11]) or they ask for lower sales prices by the selling company (e.g. [CFM10]). Although the wording and the method to fulfill the need differs from publication to publication, the need remains the same. The customer asks for a lower retail price, independently if this lower retail price is the result of a subsidy, a lower sales price by the company or other measures. Therefore, we sum up every need in the literature under the name *retail price*—as long as it is related to the retail price in some way. We conduct this kind of condensation of differently expressed but semantically identical needs into one general need expression for every need in our literature. Afterwards, we create an overview which can be found as a concept matrix in table 1. Every need in this table is one of the general needs we received due to the described condensation of need expressions.

With the basis of table 1, the next step is to build major categories out of the identified needs. It is important that the emerging clusters of needs are disjoint but intrinsically similar. There will always be a trade-off between broadly defined clusters and the number of clusters—which can not be solved perfectly. Most importantly, the clustering must be unambiguous and every need is represented in one category. Subsequently, we assign every need from the literature review to exactly one major category. In some cases needs do not fit in any of the categories, meaning the categorical system is not entire and has to be revised. A method to create a clustering pattern, to which every need can be assigned, is to start with the single needs and merge them continuously with the most similar needs or cluster of needs until only a low number of major need categories remain. We start with needs which have the closest distance to each other. In our case, the closest distance to each other are all needs which have to do with the engine of a car, like its sound (e.g. [Gö11]) or performance (e.g. [CFM10]). We merge them to an *engine* cluster. We search again for needs or clusters which have the closest distance to each other after the execution of the first step. In this second step, we merge almost every need excluding retail price and environment, because the distance to other clusters or need expressions is still too wide. Besides these two categories, retail price and environment, we find a category which covers every cost which is dependent on the driving usage of the car, a cluster with essentials of an EV like its engine and safety, a category which covers all extras a car can have like the size of the luggage space, a cluster with all needs referring charging and a cluster which covers all personal needs like the demand to drive a popular brand.

⁹ We tried to ascertain the used aspects on our own but due to the fact that the gradations are not part of their research, our findings are very subjective. The letters in the up right corner of the cells in table 1 represent our assessment which gradation of need is used (N=Need, W=Want, D=Demand).

Publication	Type of Study	Citations from 01.-10-16	cost-related needs			car-related needs			needs referring charging			individual and social needs	
			single costs	multiple costs	engine	essentials	extras	charging time	charging stations availability	driving range	environment	environmental	personal
Hilde et al. 2011 - Willingness to pay for electric vehicles and ...	stated preference survey	203	D	D	D	D	D	D	D	D	D	D	D
Ehne and Long 2012 - Barriers to widespread adoption of electric vehicles	internet-based survey	162	D	D	D	D	D	D	D	D	D	D	D
Frank et al. 2012 - Experience made in an electric vehicle	6-month field study	109											
Chandra et al. 2010 - Green observations	empirical study with about 38,000 drivers or free riders?	98	D	D	D	D	D	D	D	D	D	D	D
Canfield et al. 2010 - Examining individuals' preferences for hybrid ...	questionnaire / stated preference experiments	66	D	D	D	D	D	D	D	D	D	D	D
Achtnich et al. 2012 - The impact of fuel availability on demand ...	computer-assisted personal interview with choice experiment	37	D	D	D	D	D	D	D	D	D	D	D
Pierre et al. 2011 - Driving an electric vehicle: A sociological	social science study / interviews	34	D	D	D	D	D	D	D	D	D	D	D
Gierm et al. 2013 - Forecasting the demand for electric vehicles	stated preference survey	26	D	D	D	D	D	D	D	D	D	D	D
Poppe et al. 2013 - Market penetration analysis of electric vehicles in ...	scenario-based computer simulation	22	D	D	D	D	D	D	D	D	D	D	D
Pfizer et al. 2014 - Who will buy e-cars? vehicles? Identifying early adopters ...	Paper-and-pencil and online questionnaire	19	D	D	D	D	D	D	D	D	D	D	D
Flium and Agrawal 2012 - Constraints to green vehicle ownership.	recorded focus group discussions / focus group study	11	D	D	D	D	D	D	D	D	D	D	D
Goh et al. 2011 - Attractivität und Akzeptanz von Elektroautos	focus group study and component-analysis	8	D	D	D	D	D	D	D	D	D	D	D
Defner et al. 2012 - Elektrolieferung in Betriebsleistung ...	field test observation	5	D	D	D	D	D	D	D	D	D	D	D
Dredenhöfer 2013 - Why electric vehicles linked.	comprehensive test drives / experiment	2	D	D	D	D	D	D	D	D	D	D	D
Wunder 2013 - The impact of policy measures on the demand ...	scenario-based constraints analysis	2	D	D	D	D	D	D	D	D	D	D	D
Pfizer and Guann 2013 - Who should buy electric vehicles?	public driving survey	2	D	D	D	D	D	D	D	D	D	D	D
Frenzel et al. 2015 - Bestraume von Elektrofahrzeugen ...	online questionnaire	1	D	D	D	D	D	D	D	D	D	D	D
Schahle et al. 2016 - Cross-border Mobility for Electric Vehicles	cross-border field test observation	0	D	D	D	D	D	D	D	D	D	D	D

Tab. 1: Overview about the current state of research in the field of e-mobility. The publications are listed in the bibliography. (N=Need, W=Want, D=Demand)

overhead = (cost to run, operating cost, repair and maintenance, warranty) often used in TCO
 charging stations availability = charging infrastructure financing = leasing price

Afterwards, we merge these categories again. The retail price is similar to the variable costs, so we decide to build a category for all *cost-related* needs. We also combine the essential and the extra car needs to one larger *car-related* needs cluster. The charging cluster is too distant to every other cluster, so it is not merged in this step. It remains a cluster which covers all the *charging-related* needs. However, we can unite the environment cluster and the personal cluster which covers all the *social and individual* needs. Now, we consider the merging as completed. Every further combination would lead to very heterogeneous clusters. Therefore, we define the following clusters as our four major need categories in the field of e-mobility; *Cost-related* needs, *car-related* needs, *charging-related* needs as well as *social and individual* needs.

3.2 Clustering of Customer Needs from Tweets

To compare the identified customer needs from our literature review with our dataset, we use the previously described categories. We start with the assignment of the *need tweets* to one or more major need category. Therefore, we look at every single instance and decide to which major category or categories it belongs. To determine its affiliation, we use the content of the instance, our interpretation what the author of the tweet meant and our definitions of the four major need categories from the literature review. It can occur that an instance is assigned to a major need category because the content refers to this category—although the need expression or the need itself was not part of the category creation process previously. We are also able to assign an instance to a fifth category named *other*, if it contains a need which is not covered by any of our major four need categories from the literature review. The allocation is depicted in figure 2.

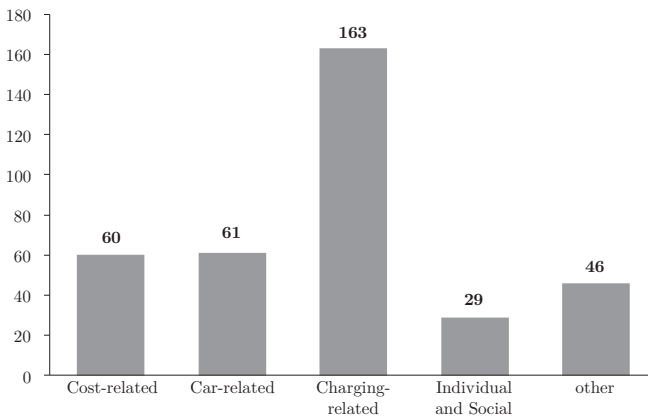


Fig. 2: Number of assigned *need tweets* to the different major need categories

During our assignment of instances to their need categories, we notice there are instances in which the concrete need is almost impossible to detect or it is hard to determine if the need fits in a major need category or is rather a need for the *other* category. Nevertheless, we consider our assignment as reasonable and scientifically reproducible as possible and it is checked by both researchers. Having assigned every instance to one or more of the

four major need categories (*cost-related*, *car-related*, *charging-related* and *social and individual* needs) and the *other* category, we proceed top-down when we use this allocation to build smaller clusters, derived from the major need categories. This allows an analysis in more detail and we can compare if we find the same sub-categories we discovered in our literature review.

We start with every instance which is assigned to the major need category *cost-related*. The instances of this category contain four tweets which refer to the electricity price, so we pool them together to a sub-category. All the other instances do not refer to a specific price because the people mix up retail prices and costs for a vehicle in general. Hence, we decide to group every instance, which does not contain an *electricity-price* need, to one large retail price and cost cluster. In the next step, we try to analyze our sub-categories in more detail to determine if there are smaller clusters we can build out of these instances. In this case a further reduction is not possible. Therefore, the set of *cost-related* instances is split and every instance is assigned to one of the two sub-categories. The number of how many instances are assigned to one sub-category stands for the quantity of this need in our whole dataset.

We do the same proceeding with every instance which is assigned to the major need categories or to the *other* category. We try to find sub-categories in which we can subdivide our dataset. Sometimes, an instance can contain two needs of the same major need category. In this case we duplicate the instance and assign one instance to the one sub-need category and one to the other. Therefore, and because an instance can also be assigned to more than one major need category, we found 361 needs in 332 instances. Consequently, we receive a tree of major needs and sub-categories. This tree is depicted in figure 3. The numbers on the leaves represent the number of instances which are assigned to this sub-category and therefore contain this specific need. There are also some nodes which contain a number. In these cases, the instance contains a need which is expressed so widely that we are only able to say it refers to the node but do not recognize to which sub-category of this node.

4 Discussion

Since our dataset mainly covers the timespan from March 3rd to August 25th 2015 and we only have instances with content in German, we are also only able to give insights for this specific timespan and only for the German-speaking area. Events in the meantime like the reveal of the Volkswagen emissions scandal in September 2015 or the presentation of the Tesla Model 3 in March 2016 could have changed the current needs of people in the field of e-mobility—but is not reflected in our dataset. As mentioned before, the following statements do not claim to be representative but to deliver new insights and stimulus.

The dataset contains 60 *cost-related* needs, 61 *car-related* needs, 163 *charging-related* needs, 30 *social and individual* needs and 46 needs which are not represented by one of these categories. The *cost-related* needs can be split into needs which refer to the electricity price and in needs which refer to the retail price and the costs for an EV. In contrast to the needs from our literature review, the needs referring to the electricity price are underrepresented. The electricity price is a need in over half of the used literature whereas

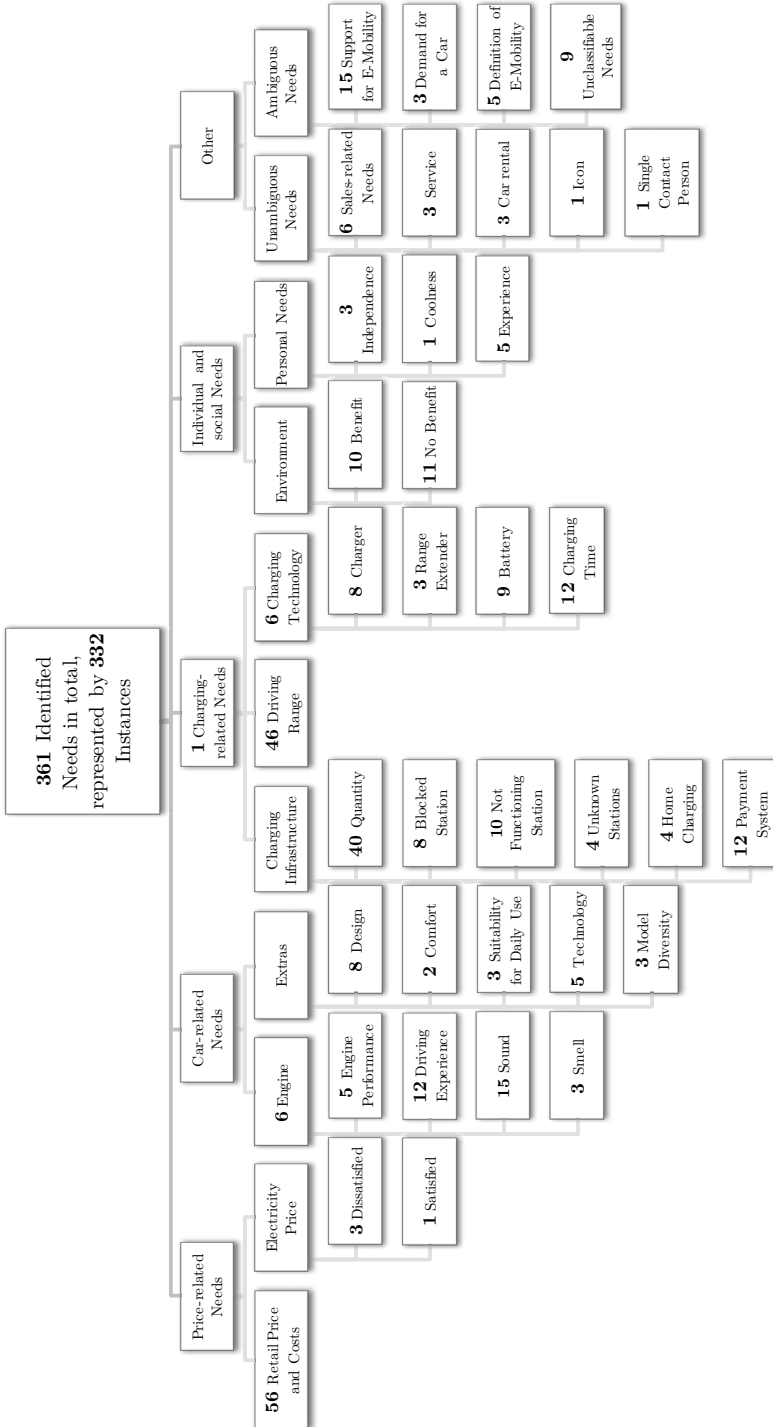


Fig. 3: Manual clustering tree of needs in the dataset

only 6.7% of the *cost-related* tweets contain a need in this domain. The remaining 93.3% instances contain needs which refer to the retail price in particular or to the costs of EVs in general. In many tweets, one is not able to say if the author of the tweet refers just to the retail price or to the total costs of the car, a distinction which is clearly made in the literature. Only two instance have positive sentiments towards the current costs whereas 54 request lower retail prices or costs.

Compared to the previous category, the *car-related* needs show more diversity. The driving experience is lauded by every of the 12 instances whereas the opinion about the engine performance is balanced. Moreover, the sound of an EV is discussed very controversial: Four people like the silence of EVs but five people miss the sound of a fossil fuel engine during driving. Additionally, six people believe that the absence of a car sound is dangerous for traffic participants and pedestrians. The current design of EVs is mainly rejected and the comfort and suitability for daily use is disbelieved by 13 instances. Compared to the literature review, no tweet questions the reliability or safety of an EV. We assume the (potential) customers do not expect any differences in these topics between a fossil fuel vehicle and an EV.

Our dataset is dominated by various needs which refer to *charging*. 38 potential consumers and users of EVs fault the low quantity of charging stations and only two are satisfied with it. A great nuisance for current users is when they are not able to use existing charging infrastructure (see figure 1), a problem which can not be found in any of the used publications. Eight instances refer to blocked charging stations by non EVs or fully charged cars, ten users report not-functioning stations and four tweets describe situations where the user did not know of an existing charging station. Moreover, twelve people have problems with the different payment systems and often with the charging on a station of a provider, divergent from those they usually use.

The driving range of EVs, which is significantly lower to the range of fossil fuel vehicles, is an issue in 46 instances. There are two people which are satisfied with the current driving ranges, but the vast majority criticizes it. In addition, the charging technology is another topic of *charging-related* needs. The technology itself, the used charger and the battery are represented in 23 instances. The greatest single need in technology is the charging time with twelve mentions but in comparison to the other needs of the category *charging* it is only a minor need. This is also a difference to the findings of our literature review in which a fast charging time is a frequently mentioned need. Moreover, a fact that is represented by almost every instance is that the (potential) consumers are not willing to change their driving habits and demand the same comfort referring to the driving range like a fossil fuel vehicle has.

The *social and individual* needs are dominated by environmental needs. In contrast to the literature, a majority of people think that e-mobility is not a benefit for the environment. The most criticized issues are the origin of electricity, which is mainly produced by coal-fired power plants in Germany, and the emissions which are caused during the production. Other needs are the demand for test drives to get experience with EVs, to use e-mobility to be "cool" or to become independent from oil companies. However, compared to our

dataset in total, the *social and individual* needs, especially the environment, only make up 8.3% of the total needs—and prove to be less frequent than current research assumes.

The last category *other* contains every need which is not covered by the major need categories and therefore also not represented in the current state of research. The instances are divided in needs which can be named and needs which are ambiguous and we do not have the ability to name the need unambiguously. The unambiguous needs refer to better sales and distribution by the companies, a better service and more EVs available to rent. There is also respectively one person who demands an icon for e-mobility and a single contact person for the topic e-mobility. Furthermore, the ambiguous needs include instances which contain the desire of a (better) support for e-mobility without specifying precisely how this support should look like, if it is financial support for the buyers, the companies, legislative changes or something totally different. These instances have in common that they are mainly addressed to the government or politicians. Additionally, other ambiguous needs are the criticism of the current definition of e-mobility by the public and the demand for a car without specifying why. We also find nine instances for which we are not able to determine the need.

When we look at the instances in general, we notice two issues which do not refer to a specific major needs category. Most instances are written by people who have no experience with EVs¹⁰ and never did a test drive. They are the vast majority of people who criticize the costs as well as the driving range of cars. However, when we only look at the tweets of people who have experience with EVs, the sentiment towards e-mobility becomes much more positive e.g. every of the twelve opinions about the great driving experience is expressed by them. The other observation is that e-mobility is equaled with personal EVs. Other characteristics like new mobility concepts besides owning an own car are hardly ever represented in our dataset.

5 Conclusion and Outlook

We motivated the work at hand with the importance of the identification of customer needs, which can be of support in the design of market-driven services. By comparing needs in the field of e-mobility from literature with needs from micro blog posts, we get valuable insights about customer needs expressed in social media.

We gave an overview of current literature about customer needs in e-mobility. Afterwards, we analyzed the customer needs in the literature and assembled a representative set of 18 publications. Based on this set we learned that the detected needs can be pooled in four major categories (*cost-related* needs, *car-related* needs, *charging-related* needs and *social and individual* needs). We were able to assign every customer need in current research of e-mobility to exactly one category. To our knowledge there is no such pattern to cluster customer needs in the field of e-mobility research.

The received (unrepresentative) Twitter dataset gave us useful insights in the field of e-mobility. *Charging-related* needs are the mostly expressed needs in our dataset. Besides

¹⁰ As far as we can judge such statements.

the concerns about the quantity of charging stations and driving range, which can also be found in literature, we learned that a large set of instances refer to situations in which the infrastructure was available but could not be used. For instance, the case that a charging station is blocked by another car is not covered by current research yet—but often occurred in our dataset. Additionally, the needs which refer to the costs of an electric vehicle are also frequently expressed. Their overwhelmingly share demand lower prices. When we look at the environmental benefit of e-mobility, the opinion is split because the origin of the used electricity is predominantly based on fossil energy sources. The analysis of customer needs, which do not fit in our major categories, reveals the need for better interaction of the selling company with the customer, when it comes to the sale of a electric vehicle or services which are offered by the company. Moreover, we learned that customer experience in the field of e-mobility leads to a better sentiment towards e-mobility—which is also a key finding of Dudenhöffer [Du13].

The work at hand has three major limitations. First, we used data which is collected by keywords. This could lead to the case that relevant instances are excluded from the dataset, because they did not match any of the keywords. The same could happen when we try to eliminate all non-relevant instances from the retrieved dataset. Second, subjective assessments can not be ruled out. Although we tried to work as impartial as possible, we can not guarantee that some decisions are influenced by personal circumstances as part of the clustering of needs to find our major need categories. In this case, we decided to merge clusters based on our opinion of homogeneity of the resulting cluster. Same applies for assignment of the micro blog posts to the categories; Some contain texts which allow interpretation about the underlying need or reply to an unknown tweet of another user which could lead to false interpretations. Third, the work only shows a static snap-shot of tweets in 2015 and does not consider dynamics of customer needs.

Nonetheless, the work at hand shows how valuable an in-depth analysis of first-hand customer expressions about a certain topic (in this case e-mobility) can be. We were able to detect new, previously unknown needs which were not named in the state of the art literature. This can be of help in the development of future, customer-centered e-mobility services. Future work will concentrate on ways to automate the described process of extracting needs from micro blog data within an automated tool [Ku16, KSS16]—an interesting field of work lies ahead.

References

- [ABH12] Achtnicht, Martin; Bühler, Georg; Hermeling, Claudia: The impact of fuel availability on demand for alternative-fuel vehicles. *Transportation Research Part D: Transport and Environment*, 17(3):262–269, 2012.
- [AH98] Augustyn, Marcjanna; Ho, Samuel K: Service quality and tourism. *Journal of Travel Research*, 37(1):71–75, 1998.
- [AP02] Alam, Ian; Perry, Chad: A customer-oriented new service development process. *Journal of Services Marketing*, 16(6):515–534, 2002.
- [BL95] Bailetti, Antonio J; Litva, Paul F: Integrating customer requirements into product designs. *Journal of Product Innovation Management*, 12(1):3–15, 1995.

-
- [BPS10] Brown, Stephen; Pyke, David; Steenhof, Paul: Electric vehicles: The role and importance of standards in an emerging market. *Energy Policy*, 38(7):3797–3806, 2010.
- [BS06] Bullinger, Hans-Jörg; Scheer, August-Wilhelm: *Service engineering - Entwicklung und Gestaltung innovativer Dienstleistungen*. Springer, 2006.
- [CFM10] Caulfield, Brian; Farrell, Séona; McMahon, Brian: Examining individuals preferences for hybrid electric and alternatively fuelled vehicles. *Transport Policy*, 17(6):381–387, 2010.
- [CGK10] Chandra, Ambarish; Gulati, Sumeet; Kandlikar, Milind: Green drivers or free riders? An analysis of tax rebates for hybrid vehicles. *Journal of Environmental Economics and management*, 60(2):78–93, 2010.
- [De12] Deffner, Jutta; Birzle-Harder, Barbara; Hefter, Tomas; Götz, Konrad et al.: *Elektrofahrzeuge in betrieblichen Fahrzeugflotten—Akzeptanz, Attraktivität und Nutzungsverhalten. Ergebnisbericht im Rahmen des Projekts Future Fleet*. Frankfurt am Main. ISOE-Institut für sozial-ökologische Forschung, 2012.
- [De16] Deutsch, Alison L.: *4 Ways Sina Weibo and Twitter Compare*, 2016.
- [Du13] Dudenhöffer, Kathrin: Why electric vehicles failed. *Journal of Management Control*, 24(2):95–124, 2013.
- [EL12] Egbue, Ona; Long, Suzanna: Barriers to widespread adoption of electric vehicles: An analysis of consumer attitudes and perceptions. *Energy policy*, 48:717–729, 2012.
- [FA12] Flamm, Bradley J; Agrawal, Asha Weinstein: Constraints to green vehicle ownership: A focus group study. *Transportation Research Part D: Transport and Environment*, 17(2):108–115, 2012.
- [Fr12] Franke, Thomas; Neumann, Isabel; Bühler, Franziska; Cocron, Peter; Krems, Josef F: Experiencing range in an electric vehicle: Understanding psychological barriers. *Applied Psychology*, 61(3):368–391, 2012.
- [Fr15] Frenzel, Ina; Jarass, Julia; Trommer, Stefan; Lenz, Barbara: *Erstnutzer von Elektrofahrzeugen in Deutschland. Nutzerprofile, Anschaffung, Fahrzeugnutzung*. Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Berlin, Germany, 2015.
- [Gl13] Glerum, Aurélie; Stankovikj, Lidija; Thémans, Michaël; Bierlaire, Michel: Forecasting the demand for electric vehicles: accounting for attitudes and perceptions. *Transportation Science*, 48(4):483–499, 2013.
- [Gö11] Götz, Konrad; Sunderer, Georg; Birzle-Harder, Barbara; Deffner, Jutta; Berlin, Büro: *Attraktivität und Akzeptanz von Elektroautos*. Institut für sozial-ökologische Forschung, S. 8, 2011.
- [Gr06] Groves, Robert M: Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, 70(5):646–675, 2006.
- [Ha01] Harding, Jennifer A; Popplewell, K; Fung, Richard YK; Omar, Abdul R: An intelligent information framework relating customer requirements and product characteristics. *Computers in Industry*, 44(1):51–65, 2001.
- [HD03] Hickey, Ann M; Davis, Alan M: Elicitation technique selection: how do experts do it? In: *Requirements Engineering Conference, 2003. Proceedings. 11th IEEE International. IEEE*, S. 169–178, 2003.

- [Hi11] Hidrue, Michael K; Parsons, George R; Kempton, Willett; Gardner, Meryl P: Willingness to pay for electric vehicles and their attributes. *Resource and Energy Economics*, 33(3):686–705, 2011.
- [KA01] Kotler, Philip; Armstrong, Gary: *Principles of marketing*. pearson education, 2001.
- [KI14] Klör, Benjamin; Bräuer, Sebastian; Beverungen, Daniel; Matzner, Martin: *IT-basierte Dienstleistungen für die Elektromobilität-Konzeptioneller Rahmen und Literaturanalyse*. 2014.
- [KSS16] Kuehl, Niklas; Scheurenbrand, Jan; Satzger, Gerhard: *Needmining: Identifying Micro Blog Data Containing Customer Needs*. In: *24th European Conference of Information Systems*. 2016.
- [Ku16] Kuehl, Niklas: *Needmining: Towards Analytical Support for Service Design*. In: *International Conference on Exploring Service Science*. Jgg. 6. Springer, 2016.
- [MLF15] Marshall, Tara C; Lefringhausen, Katharina; Ferenczi, Nelli: The Big Five, self-esteem, and narcissism as predictors of the topics people write about in Facebook status updates. *Personality and Individual Differences*, 85:35–40, 2015.
- [PG13] Plötz, Patrick; Gnann, Till: *Who should buy electric vehicles?—The potential early adopter from an economical perspective*. *Proceedings of the 2013 ECEEE summer study*, Hyeres, France, 2013.
- [PJL11] Pierre, Magali; Jemelin, Christophe; Louvet, Nicolas: *Driving an electric vehicle. A sociological analysis on pioneer users*. *Energy Efficiency*, 4(4):511–522, 2011.
- [PI14] Plötz, Patrick; Schneider, Uta; Globisch, Joachim; Dütschke, Elisabeth: *Who will buy electric vehicles? Identifying early adopters in Germany*. *Transportation Research Part A: Policy and Practice*, 67:96–109, 2014.
- [Pr13] Propfe, Bernd; Kreyenberg, Danny; Wind, Joerg; Schmid, Stephan: *Market penetration analysis of electric vehicles in the German passenger car market towards 2030*. *International Journal of Hydrogen Energy*, 38(13):5201–5208, 2013.
- [Ra09] Randolph, Justus J: *A guide to writing the dissertation literature review*. *Practical Assessment, Research & Evaluation*, 14(13):1–13, 2009.
- [RJ00] Rising, Linda; Janoff, Norman S: *The Scrum software development process for small teams*. *IEEE software*, 17(4):26, 2000.
- [Sc15] Scheurenbrand, Jan; Engel, Christian; Peters, Florin; Kuehl, Niklas: *Holistically Defining E-Mobility: A Modern Approach to Systematic Literature Reviews*. In: *Proceedings of the First Karlsruhe Service Summit Workshop*. Karlsruhe, Germany, S. 17–27, 2015.
- [SJF16] Schäuble, Johannes; Jochem, Patrick; Fichtner, Wolf: *Cross-border Mobility for Electric Vehicles - Selected results from one of the first cross-border field tests in Europe*. KIT Scientific Publishing, 2016.
- [St11] Stickdorn, Marc; Schneider, Jakob; Andrews, Kate; Lawrence, Adam: *This is service design thinking: Basics, tools, cases*. Wiley Hoboken, NJ, 2011.
- [St15] Stryja, Carola; Fromm, Hansjörg; Ried, Sabrina; Jochem, Patrick; Fichtner, Wolf: *On the Necessity and Nature of E-Mobility Services—Towards a Service Description Framework*. In: *Exploring services science*, S. 109–122. Springer, 2015.
- [Wi13] Windisch, Elisabeth: *The impact of policy measures on the demand for privately-held electric vehicles: a socio-economic analysis for France*. In: *13th World Congress on Transportation Research*, Rio de Janeiro, Brazil, July 15th-18th. 2013.

ICT-based Interventions for Water and Heat Energy Conservation

ICT-based Interventions for Water and Heat Energy Conservation

Thorsten Staake¹, Anna Kupfer²

The accomplishment of the ambitious goals of the energy transition heavily depends on the use of modern information and communication technology (ICT). In fact, the use of ICT to support a sustainable development already today comprises virtually all fields of our daily lives: It spans from automation and control systems in industry, the mobility sector, and households to system that motivate and enable citizens to make decision and form habits in favour of a low environmental footprint. The majority of applications focuses on electricity, transportation systems, and space heating and include issues related to smart grids, e-mobility, and home automation. Recently, the use of ICT to promote water conservation has also gained considerable attention. This development is driven by the eminent water stress in many parts of the world as well as by the considerable amount of energy that is used for water heating in countries where water is abundant.

These workshop proceedings capture applications of ICT to promote (hot) water conservation among citizens and address the design of intelligent water networks. Topics include – in short – techniques to retrieve data on water use, descriptive statistics on water consumption behaviour, the use and impact of “low-tech” flow restrictors as a benchmark for saving technologies, and the adoption of technologies that engage users in water conservation. The contributions are briefly summarized below.

In their work on non-intrusive disaggregation of household water use, Ilya Kozlovskiy, Samuel Schöb, and Mariya Sodenkamp take high-resolution water consumption data that was collected on a household-level as input to separate consumption events into different categories. They use clustering methods, based on level set trees, to identify groups of events that are similar to each other.

Based on a large field study, Liliane Ableitner, Samuel Schöb, and Verena Tiefenbeck provide a descriptive analysis of shower data. Doing so, they help to quantify the impact of consumer behaviour and to understand what drives the variance in individuals’ energy consumption. They analyze a smart meter data set comprising 5,610 individual showers and find a large variance in consumption values within and between households. This suggests that a large potential for energy savings exist. Furthermore, the authors investigate correlations between energy use, socio-demographic factors, and attitudes

¹ Otto-Friedrich Universität, Lehrstuhl Energieeffiziente Systeme, An der Weberei 5, 96047 Bamberg, thorsten.staake@uni-bamberg.de

² Otto-Friedrich Universität, Lehrstuhl Energieeffiziente Systeme, An der Weberei 5, 96047 Bamberg, anna.kupfer@uni-bamberg.de

towards the environment of the test households.

Also based on a field experiment with a large number of sensor nodes, Liliane Ableitner, Samuel Schöb, and Verena Tiefenbeck shed light on the potential of mechanical flow restrictors in the shower. The authors assess the real-world influence of the flow rate on energy and water use and compare these numbers to the 45% reduction anticipated by engineering calculations. They find that users of low-flow showerheads do, indeed, take longer showers; nevertheless, they consume only 38% (1.0 kWh) less energy per shower compared to the baseline level of 2.6 kWh due to extended shower durations.

In their work “Adoption of Behavioral Feedback Devices” Anja Peters, Daniel Hanss, and Christian Sartorius investigate how consumers can be motivated to adopt feedback technologies and consequently intensify their water conservation efforts. Their study addresses an important gap in the related work by discussing several psychological factors that are believed to influence consumer adoption of feedback devices.

All submissions have been subject to a double-blind review process. The program committee consisted of Spiros Athanasiou, Athena Research Center, Anna Kupfer, Otto-Friedrich Universität Bamberg, Anja Peters, Fraunhofer ISI, Thorsten Staake, Otto-Friedrich Universität Bamberg, Thomas Stiefmeier, Amphiro AG and Verena Tiefenbeck, ETH Zurich. The program committee members herewith express their gratitude to the external reviewers.

Non-intrusive disaggregation of water consumption data in a residential household

Ilya Kozlovskiy¹, Samuel Schöb², and Mariya Sodenkamp³

Abstract: The water conservation campaigns in residential households are hindered by the poor understanding of residents of how much water they use. For the better designed interventions new tools are necessary to educate the consumers on the water usage of different consumption events. In this paper we use the fine grained (0.5 Hz) water consumption data that was collected non-intrusively in a household over the period of 21 days to develop such tools. We examine the collected data and disaggregate the consumption events into three different categories: short events (e.g., toilet flush), long regular events (e.g., washing machine) and long irregular events (e.g., showers). To achieve this, we use clustering methods, based on level set trees, to identify groups of events that are similar to each other.

Keywords: Water conservation, Disaggregation, Water consumption, Smart meter

1 Introduction

An escalating demand on potable water resources resulting from increasing populations, droughts and unpredictable weather patterns due to climate change is commonplace in many parts of the world [Ba08]. As a result, the sustainable management of urban water has become imperative, particularly for countries prone to severe droughts [Wi11]. In the USA domestic (residential) water use was the third largest water use category after thermoelectric power generation and irrigation [Bu16]. The residential sector is the largest urban water use sector, and it offers the largest volume of potential savings compared with other urban sectors [Gl03]. But most individuals have poor understanding of how much water they use [At14]. They only get the annual, or in best case monthly bills, and lack the tools to monitor and improve their water consumption [Ta15]. The problem of residents not getting the immediate feedback also occurs in the energy sector with the electricity consumption. In this case multiple non-intrusive load disaggregation methods were developed to better educate the residents on their consumption [Ha92, Zo12, Do13].

A study that tried to educate the residents on their water expenditure with the help of in home displays has only reached a short term benefit [Fi13]. On the other hand, studies displaying the water consumption of individual showers have produced substantial (~22%) reduction in short [Ti13] and long term [Ta15] water consumption. But this

¹ Universität Bamberg, Kapuzinerstraße 16, 96047 Bamberg, ilya.kozlovskiy@uni-bamberg.de

² Universität Bamberg, Kapuzinerstraße 16, 96047 Bamberg, samuel.schoeb@uni-bamberg.de

³ Universität Bamberg, Kapuzinerstraße 16, 96047 Bamberg, mariya.sodenkamp@uni-bamberg.de

method required the installation of a metering device directly into the shower. To create intervention campaigns more suitable for the mass market we therefore look into possibility of *non-intrusive disaggregation of water consumption*. Our aim is to show that it is possible based on the overall water consumption of the households to provide the residents with the information on how much water they consume in individual categories (e.g., showers, toilets, washing machines, or faucets).

2 Data collection methodology

To collect the water consumption data, we have mounted two sensors directly to the water intake pipe of multiple households. In this study we only consider the data from one household. The first sensor measures the flow speed in the pipe in m/s and with the measured interior diameter of the pipe we could calculate the volumetric flow rate also in l/min. The second sensor measured the temperature. Both measurements were collected simultaneously on average once every two seconds. The measurements were collected over the period of 21 days from 17.04 to 08.05.2016. Both of the sensors produce the measurements with a small error. The values for the flowrate are only influenced, by the water consumption in the household, but the values for the temperature also change during the day.

3 Three Step Disaggregation of Water Consumption Data

In order to identify the different classes of consumption patterns we have developed a three step methodology for analysis of water consumption. As a first step we identify the times where the active water withdrawal takes place (water consumption events) by determining the times with no consumption. In the second step we identify and describe the events that occur often (e.g., toilet flushes have similar consumption pattern). And in the last step, we identify and describe the events that are more rare, but use more water (i.e., events with longer duration like showers).

The sensor data for the water flow is the primary evidence to determine if the water is consumed. Since the sensor data is noisy, it is important to detect the level during which water withdrawal is highly probable. To do this, we compute the kernel density estimation over all sensor values for velocity and take the first local minimum as the boundary. We use this method since we assume, that the error in sensor measurements is normally distributed. The kernel density estimation is represented in the Fig. 1. Here we set the boundary value to 0.13 l/min. Every value for velocity that lies below this boundary is then set to zero. With this, we can define individual water extraction events, as time segments with the positive flow velocity. In the considered household we detect 1807 individual water extraction events during the 21-day period. Most of the events have only a short duration (median duration is 18 s) and nearly constant (up to measurement error) water flowrate. During most of the events there is also a slight drop

(average 0.05°) in the temperature. Here we also make the assumption that all the events are singular and not a superposition of multiple withdrawals. Since most of the withdrawal events are short, this assumption should be satisfied in nearly cases.

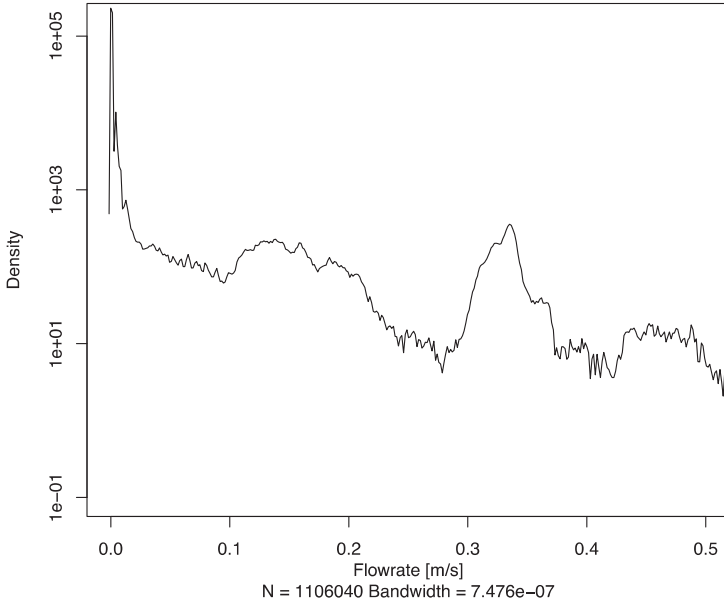


Fig. 1: The kernel density estimation of the water velocity sensor values in one household in 21-day period

Since most of the events have a similar pattern to each other we need to define measures that describe the events more closely. The definition of these features is an engineering task, in this work we use the following 5 features: duration of the event, average flowrate during the event, time of the day (in seconds), time with water consumption during the 20-minute window surrounding the event, and the temperature change during the event. The temperature change is calculated as the difference between the mean temperature during ten measurements before and after the event. Our goal is to use the computed features to find groups of similar events that occur often (i.e., the the density is higher for these events). Therefore, we use the density-based clustering algorithm that relies upon the level set trees [Ha75, St11, Ke13]. In this algorithm the individual clusters are the simple connected components at the given density levels. The resulting clusters are then presented in a tree hierarchy allowing for easy interpretation and visualization of the results. In our approach, we use the implementation in the R-package TDA [Br15], with the Gaussian kernel density estimator, the bandwidth h of 0.1 and $k=100$ for neighborhood estimations. The resulting hierarchy tree is presented in the Fig. 2. The individual clusters are the colored vertical lines and their length represents for how long the clusters survives until it connects into other clusters. We find 2 main clusters in our

dataset, each with two sub-clusters. For better visualization we plot the event duration vs. the average flowrate with the color corresponding to the cluster in the Fig. 3.

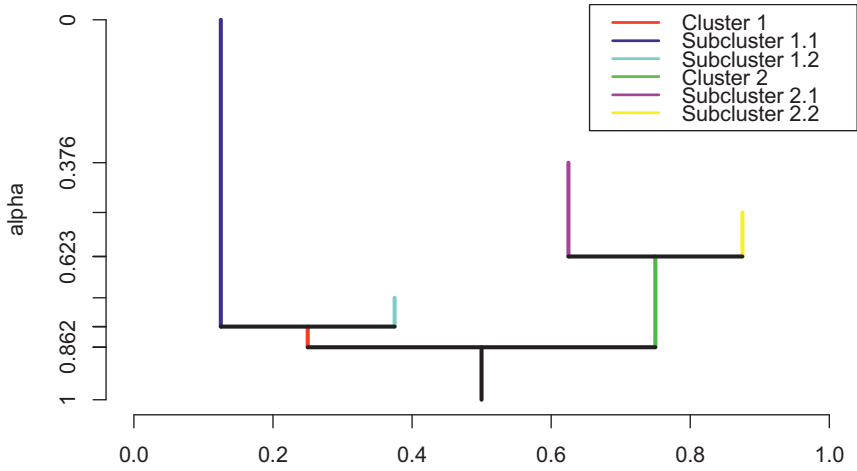


Fig. 2: The sublevel tree with different clusters (vertical lines) as result of the density clustering

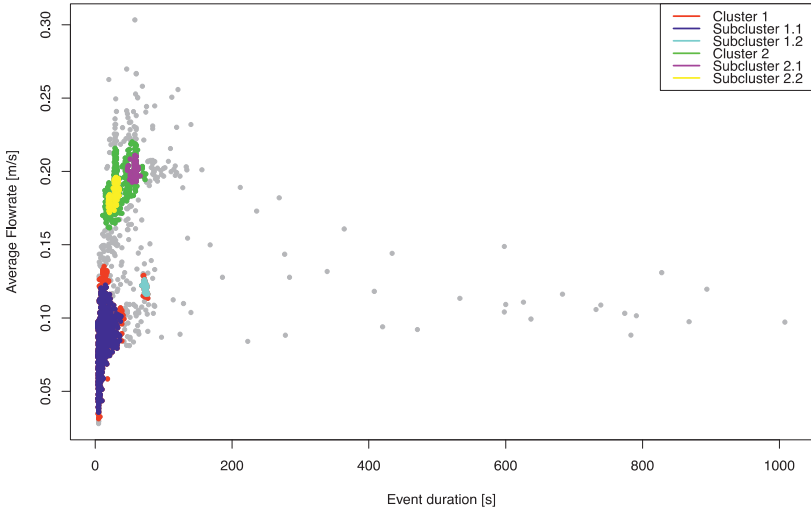


Fig. 3: The duration and average flowrate for the individual clusters

The two main clusters correspond to events with either a high (green) or a low (red)

average flow rate. Each of these clusters is then further split up into events with a high and low duration. Especially interesting is the magenta cluster, since nearly all the events in this cluster have a duration of 60 seconds and constant flow during this duration. These events are most likely produced by an appliance. Similarly the cyan cluster has events with duration mostly in the interval of 70-76 seconds and is also most likely produced by an appliance. The events in the both the cyan and magenta cluster also have mostly constant flowrate. The blue and yellow clusters on the other hand are the short events with a wide spread in duration and average flow and therefore most likely events produced by the residents directly.

With the level tree set method, the events with longer duration are not assigned to any clusters, because they occur less often. Amongst these longer events we expect to find the water withdrawals from showers and appliances. These events are more likely to not have a constant flow rate. In this step we only consider the events with duration. Therefore, we use a different approach to cluster these events, based on the underlying flowrate time series. We compute the dtw (dynamic time wrap) distance [MR81] between all different pairs of events. This metric tries to make the two time series resemble each other, by stretching and compressing them locally. The distance is computed after stretching by summing the distances of the aligned segments. For the computation we use the implementation in the R-package “dtw” [Gi09] with the Euclidean distance and “symmetric2” step pattern as parameters. Only velocity is used, because the temperature does not have a clear baseline and changes during the day even without any water withdrawal. Having calculated the distance matrix we perform the agglomerative hierarchical clustering [Jo67] with the package “flashClust” [LH12] with the complete linkage method. The resulting dendrogram is shown in the Fig. 4. There is a clear separation into two main clusters. The cluster on the left side contain the events with more irregular consumption levels and are more likely to be the shower events.

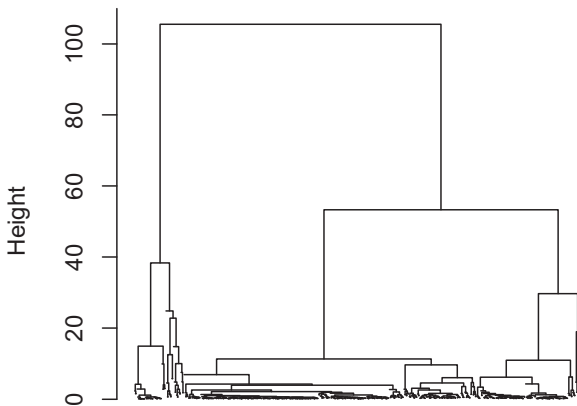


Fig. 4: The result of the hierarchical clustering of the events longer than 30 seconds

4 Conclusions

In this paper we have presented the methodology on disaggregation of the consumption events based on the detailed water consumption data. We can determine three kinds of events: Short events (with average duration less than 40s), regular long events (e.g., appliances), irregular long events (e.g., showers). By categorizing all the consumption events in one of three categories it is possible to provide the residents with a more detailed information on their water consumption (e.g., by showing the cost that occurred due to showers, or the usage of washing machines). The short consumption events are hard to categorize in more details, since they mostly have a constant flowrate.

The presented work can be extended in multiple aspects. A more detailed analysis of the resulting clusters can lead to better insights about the consumption origin. Further in this paper we consider events as separated by zero consumption. But some events are actually multi-part events (e.g., multiple cycles of the washing machine). To recognize such events, we would need first to aggregate nearby events to larger events. During distance calculations with the dtw metric, subpattern matchings (e.g., by calculating the dtw distance with open ends) can then be performed. If there are multiple matches for the sub-patterns, the events can be considered separately. Additionally, the temperature sensor data could be included in the consumption disaggregation to distinguish between the patterns.

References

- [At14] Attari, S.Z.: Perceptions of water use. *PNAS* 111, pp. 5129–5134, 2014.
- [Ba08] Bates, B. C.; Kundzewicz, Z. W.; Wu, S.; Palutikof, J. P.: *Climate Change and Water - IPCC Technical Paper VI*. Intergovernmental Panel of Climate Change (IPCC) Secretariat, Geneva, 2009.
- [Br15] Fasy, B. T.; Kim, J.; Lecci, F.; Maria, C.; Rouvreau, V.: The included GUDHI is authored by Clement Maria, Dionysus by Dmitriy Morozov, PHAT by Ulrich Bauer, Michael Kerber and Jan Reininghaus. *TDA: Statistical Tools for Topological Data Analysis*. R package version 1.4.1, 2015.
- [Bu16] *Building Energy Data Book*. Chapter 8: Water, <http://buildingsdatabook.eren.doe.gov/ChapterIntro8.aspx>, accessed on 30.05.16
- [Do13] Dong, M.; Meira, P. C.; Xu, W.; Chung, C. Y.: Non-intrusive signature extraction for major residential loads. *IEEE Transactions on Smart Grid* 4/3, pp. 1421-1430, 2013.
- [Fi13] Fielding, K.S.; Spinks, A.; Russell, S.; McCrea, R.; Stewart, R.; Gardner, J.: An experimental test of voluntary strategies to promote urban water demand management. *Journal of Environmental Management* 114, pp. 343–351, 2013.
- [Gi09] Giorgino T.: Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package. *Journal of Statistical Software* 31/7, pp. 1-24, 2009.

-
- [Gl03] Gleick, P. H.; Wolff, G.H.; Cushing K. K.: Waste not, want not: the potential for urban water conservation in California. Oakland, CA: Pacific Institute for Studies in Development, Environment, and Security, 2003.
- [Ha75] Hartigan, J. A.: Clustering Algorithms, 1975.
- [Ha92] Hart, G. W.: Nonintrusive appliance load monitoring, In: Proceedings of the IEEE 80/12, pp. 1870-1891, 1992.
- [Jo67] Johnson, S. C.: Hierarchical clustering schemes, *Psychometrika* 32/3: 241-254, 1967.
- [Ke13] Kent, B. P.; Rinaldo, A.; Verstynen, T.: DeBaCl: A Python package for interactive DEnsity-BASed CLustering. arXiv preprint arXiv:1307.8136, 2013.
- [LH12] Langfelder, P.; Horvath, S.: Fast R Functions for Robust Correlations and Hierarchical Clustering. *Journal of Statistical Software* 46/11, pp. 1-17, 2012.
- [MR81] Myers, C. S.; Rabiner, L. R.: A Comparative Study of Several Dynamic Time-Warping Algorithms for Connected-Word Recognition. *Bell System Technical Journal* 60/7, pp. 1389-1409, 1981.
- [St11] Steinwart, I.: Adaptive Density Level Set Clustering. COLT, 2011.
- [Ta15] Tasic, V.; Tiefenbeck, V.; Schöb, S.; Staake, T.: Short-term Spark or Sustained Impact? Investigating the Long-term Effect of Real-time Feedback" ECIS 2015 Research-in-Progress Papers, 2015.
- [Ti13] Tiefenbeck, V.; Tasic, V.; Schoeb, S.; Staake, T.: Mechatronics to drive environmental sustainability: Measuring, visualizing and transforming consumer patterns on a large scale, in: Proceedings of the IEEE IECON 2013, Special Section on Energy Informatics, Energy Informatics, 2013.
- [To16] Tomorrow & Beyond, Water Sense, https://www3.epa.gov/watersense/our_water/tomorrow_beyond.html, accessed on 30.05.16.
- [Wi11] Willis, R. M.; Stewart, R. A.; Panuwatwanich, K.; Williams, P. R.; Hollingsworth, A. L.: Quantifying the influence of environmental and water conservation attitudes on household end use water consumption. *Journal of environmental management* 92/8, 1996-2009, 2011.
- [Zo12] Zoha, A.; Gluhak, A.; Imran, M. A.; Rajasegarar, S.: Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. *Sensors* 12/12, pp. 16838-16866, 2012.

Digitalization of Consumer Behavior – A Descriptive Analysis of Energy Use in the Shower

Liliane Ableitner¹, Samuel Schöb², and Verena Tiefenbeck³

Abstract: In order to develop effective technology- or behavior-based energy conservation measures, it is necessary to quantify the impact of consumer behavior and to understand what drives the variance in individuals' energy consumption. While sensors and communication technologies are becoming more and more ubiquitous, surprisingly little data is still available on many aspects of consumer behavior in the energy sector. Water heating is the second largest energy use in households – yet, data on hot water consumption and on factors influencing the variance in that domain is particularly scarce. We analyze a smart meter data set comprising 5,610 individual showers. We find that the average shower consumes 43.9 liters of water and 2.60 kWh heat energy. The high observed variance in consumption values within and between households suggests a large potential for energy savings. Furthermore, we investigate correlations between energy use and consumer characteristics (sociodemographic factors and environmental attitudes).

Keywords: Shower data, descriptive analysis, energy consumption, hot water, smart meter

1 Motivation and Context

In modern societies, a life without energy is unimaginable. Today's energy demand is mainly covered by fossil fuels. This drives environmental (e.g., carbon emissions) and geopolitical problems (e.g., energy security). Domestic emissions in the US and the UK accounts for about third of national emissions [BN08, GS08]; in Switzerland households consume 26,5% of total energy use [Sc14]. The international community of governments, companies and other organizations therefore pursue energy efficiency targets. Despite enormous investments in smart grid and smart metering infrastructure, most consumers still have a poor understanding of effective strategies for energy conservation [At14].

A particularly large domestic energy use is hot water consumption: With 18% of energy demand of private US households and 13% for European households, respectively

¹ ETH Zürich, M-TEC, Weinbergstrasse 56/58, CH-8092 Zürich, lableitner@ethz.ch

² Otto-Friedrich-Universität Bamberg, WIAI, An der Weberei 5, D-96047 Bamberg, samuel.schoeb@uni-bamberg.de

³ ETH Zürich, M-TEC, Weinbergstrasse 56/58, CH-8092 Zürich, vtiefenbeck@ethz.ch

[Re13, De13], it is the second largest domestic energy use after space heating. Yet, the majority of consumers are unaware of the large amount of energy contained in hot water. In fact, most people do not associate showering with energy consumption. Consequently, we expect a high saving potential for hot water consumption. Yet, before we evaluate energy efficiency means from technical and behavioral viewpoints, we first need to quantify the energy consumption associated with that behavior, gauge the variance in that behavior, and understand how humans consume hot water.

With decreasing prices for sensors measuring the full set of circumstances (e.g., GPS location, altimeter, aerial carbon dioxide) tracking one's life gets easier. Recently, a Swiss startup developed a smart meter for the shower that tracks hot water consumption with the help of a temperature and a flowrate sensor. In that way, hot water consumption data is gathered on a disaggregated level and allows a better understanding of the yet unexplored energy end use showering.

This paper analyzes such a shower data set comprising of 5,610 data points collected in 636 Swiss households in 2012. We study resource consumption in the shower by examining the consumption variables water volume, water temperature, flowrate, shower time and consumed energy. We observe variance patterns within and between households and analyze consumption for differences in age, gender, and attitude towards the environment.

Giving insights into human shower behavior based on the worldwide first large-scaled shower data set accesses one more domestic resource use which could not have been explained so far. Yet, we have to stress that all findings are based on our particular data and that this paper is in every respect descriptive. Please note that other articles based on the same data have been published so far; parts of the methodology section therefore correspond to our previous work [Ti16]. In the following we describe our methodology and present our results before we discuss their implications.

2 Study Design, Data Collection and Sample

We conducted a large-scale field study in cooperation with a local utility that gave the study device (smart shower meter) as a gift (unconditional of participating in our study) to 5,000 customers. Among these, 1- and 2-person households who agreed to answer two surveys and to make their shower data available to the researchers were eligible for the energy efficiency study with a limitation of 700 participating households (due to budget and logistic reasons). The purpose of the surveys was to get demographical data as well as the participant's attitude towards the environment. For the estimation of the latter, we used a scale provided in previous literature [Di07] 10 items (measured on 5-point Likert scales) assess individuals' level of environmental concern. In the following two months of winter 2012/ 2013, study participants installed the smart meters in their showers. The technology tracks all water extractions with the help of a flowrate and a temperature sensor and stores the data on the internal memory. At the end of the study, participants

sent in their devices in order to have them manually read out before they were returned to the participants for good. The course of study was split into two periods: the baseline period which tracks shower behavior without of any contrived researcher interference (first ten tracked showers on the device) and the intervention period in which participants are provided with real-time information on their resource use (from shower eleven on). During the baseline period the smart meter displayed only temperature; we therefore limited the baseline period to ten showers per household in order to not bore the study participants. Later on in the intervention phase water and energy consumption, temperature and an energy efficiency class were shown. This paper first provides a descriptive analysis of the baseline data in order to give insights into human shower behavior⁴ and then tried to explain variance in shower behavior with the help of demographics.

Overall, we collected survey and shower data from 636 Swiss households located in the region of Zurich. In total, we analyze shower behavior of 975 individuals, 50% of which are female. The mean age of study participants is 46.3 years with a standard derivation of 15.6 years. Regarding attitude toward the environment, our sample reflects well the Swiss society with a mean of 3.61 (SD=.46) compared to the mean score provided by Diekmann et al. (2009) of 3.64. The smart meter data comprises flowrate, temperature, shower time and volume of a total of more than 46'000 showers. Energy consumption is calculated using the standard engineering formula for heat energy ($E = m * cp * \Delta T / \eta$, with heat energy E, mass of water m, heat capacity cp, ΔT the difference between the measured water temperature and cold water temperature, and η the coefficient of energy efficiency. The latter depends on the type and age of the heating system. A study for the Swiss Office of Energy [De13] gives a detailed breakdown of residential water heating systems for Switzerland. The vast majority use fossil fuels (40% oil, 25% electric resistance heaters, 21% gas), with an average conversion efficiency of 65%. As mentioned in the previous section, baseline showers only (up to ten showers) are considered in this paper. Water extractions of below 4.5 liters are not considered a shower. With this we ensure that water extractions for cleaning or watering the plants will not distort average consumption values. Data points with more than 200 liters per shower or an average temperature of above 47°C were excluded as extreme outliers or measurement errors (85 showers). Furthermore, we discarded the first data point of every dataset, as its temperature and volume distribution strongly deviated from all other showers recorded. We assume that in many cases, the first water extraction was not an actual shower; instead, participants who had just completed the installation turned on the water for several seconds to see if the device worked and what information it displays. In all, this leaves us with up to nine data point per household for our data analysis (N = 5,610).

⁴ We will not further discuss capabilities of the technology and the study period intervention. The energy savings – the main purpose of the study – are not subject of this paper but are discussed in other articles [Ti16].

3 Results

In a first step, we examine frequency distributions of the five dependent variables average temperature of a shower (in C°), water consumption (in liters), energy use (in kWh), flowrate (in liters per minute) and shower time (in minutes). Fig. 1 presents a visualization.

As can be seen in Fig. 1 the frequency distributions of temperature and flowrate are close to the normal distribution whereas resource consumption (water and energy) and shower time are skewed to the right. This pattern indicates that more people take quick showers than long excessive showers. Tab. 1 provides the descriptive statistics mean, standard derivation, minimum and maximum. In addition, we provide descriptive data of the best (lower 20%) and least performing participant groups (upper 20%) such as variances of variables within and between households.

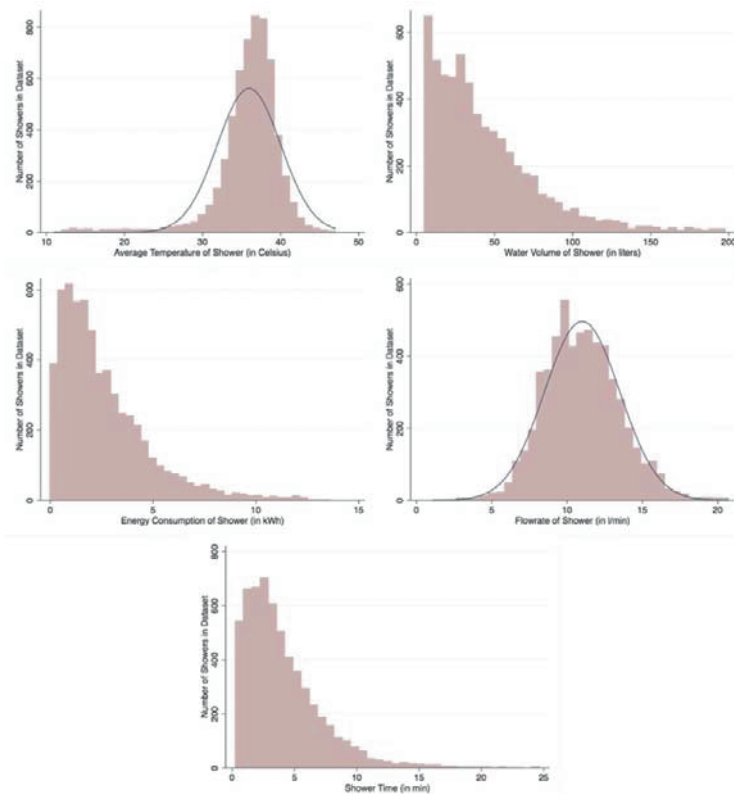


Fig. 1: Frequencies of Showers regarding Flowrate, Average Temperature, Water Consumption, Energy Consumption, and Shower Time

Tab. 1 shows that the average individual consumes 43.9 liters of hot water at a temperature of 36 C° when showering. This corresponds to 4 minutes showering at a flowrate of 11 liters per minute and to 2.60 kWh of energy to heat the water. We find a high variance in shower behavior: the bottom quintile of our sample use approximately 10 liters per shower, whereas the top quintile (upper 20%) use 100 liters, respectively. Tab. 1 further provides information on variances in shower behavior between and within households. N refers to the total number of considered showers and n to the number of households; T-bar is the average number of showers per household. For non-parametric variables we observe negative values as the within number refers to the deviation from each household's average. While between and within values for behavior related variables (water, energy, temperature and shower time) do not vary much; we find for the technical variable flowrate that variance within household is remarkably lower than between households. This shows that individuals are exposed to a rather predefined flowrate by their shower and are free to adapt the shower behavior (water consumption and temperature) to their individual comfort needs.

		<i>Mean</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>Observations</i>
<i>Average Temperature (in C°)</i>	Overall	36.0	4.09	11.0	47.0	N = 5610
	Lower 20%	30.0	4.94	11.0	33.5	N = 1075
	Upper 20%	39.9	1.47	38.5	47.0	N = 1442
	Between		2.68	18.6	43.1	n = 636
	Within		3.11	10.4	50.3	T-bar = 8.82
<i>Water Consumption (in liters)</i>	Overall	43.9	33.7	5.00	198	N = 5610
	Lower 20%	9.92	3.29	5.00	15.0	N = 1060
	Upper 20%	97.6	30.1	66.0	198	N = 1140
	Between		24.9	8.33	152	n = 636
	Within		23.3	-82.1	178	T-bar = 8.82
<i>Energy Consumption (in kWh)</i>	Overall	2.60	2.15	0.00	13.6	N = 5610
	Lower 20%	0.50	0.23	0.00	0.87	N = 1122
	Upper 20%	6.05	1.98	3.98	13.6	N = 1123
	Between		1.60	0.16	9.23	n = 636
	Within		1.48	-5.16	12.1	T-bar = 8.82
<i>Flowrate (in l/min)</i>	Overall	11.0	2.52	1.07	20.7	N = 5583
	Lower 20%	7.67	1.14	1.07	8.88	N = 1108

	Upper 20%	14.7	1.41	13.1	20.7	N = 1118
	Between		2.34	3.19	20.0	n = 633
	Within		0.98	4.56	16.9	T-bar = 8.82
	Overall	4.04	3.04	0.27	24.7	N = 5583
<i>Shower Time (in minutes)</i>	Lower 20%	0.98	0.36	0.27	1.58	N = 1111
	Upper 20%	8.84	2.82	6.03	24.7	N = 1124
	Between		2.16	0.87	14.0	n = 633
	Within		2.16	-6.75	18.5	T-bar = 8.82

Tab. 1: Descriptive Shower Data and Analysis of Variance Within and Between Households

In a next step, we investigate the correlation of the five shower variables (temperature, water consumption, energy consumption, flowrate and shower time) with the explanatory variables age, gender, and attitude toward environment of study participants. Tab. 2 presents pairwise correlation coefficients along with their significance level. As energy use is a product of water consumption, shower time and temperature, these values are obviously highly correlated. Apart from this, we find all correlations to be under the threshold of .7, which makes multicollinearity issues in the later introduced regression model very unlikely [Do13]. We conjecture that discriminant validity is assured meaning that all variables measure different contents.

In a next analysis step we investigate whether demographics and attitude towards the environment are able to predict shower behavior. Tab. 3 presents regression estimation for all five dependent variables. For each, we list regression coefficients, their significances, the explanatory potential R², and the total number of showers considered N. The independent variables are coded as follows: age in years, fraction of females in household with 0 for men only and 1 for women only, and attitude towards the environment from 1= low to 5=high. Independent variables were mean-centered for interpreting regression output correctly.

	<i>Temp.</i>	<i>Water</i>	<i>Energy</i>	<i>Flowr.</i>	<i>Time</i>	<i>Age</i>	<i>Fem. Env.</i>
<i>Temperature</i>	1						
<i>Water</i>	.37***	1					
<i>Energy</i>	.44***	.99***	1				
<i>Flowrate</i>	-.02	.22***	.20***	1			
<i>Shower Time</i>	.40***	.93***	.93***	-.08***	1		
<i>Age</i>	-.22***	-.24***	-.26***	.05*	-.26***	1	

<i>Females</i>	.03*	-.04**	-.03**	.00	-.04**	.08***	1
<i>Environment</i>	-.03**	-.13***	-.13***	-.11***	-.09***	.14***	.04**

*** p<.01 ** p<.05 * p<.1

Tab. 2: Pairwise Correlations of Dependent and Independent Variables

		<i>DV</i>				
		<i>Temperature</i>	<i>Water</i>	<i>Energy</i>	<i>Flowrate</i>	<i>Shower Time</i>
<i>IV</i>	<i>Age</i>	-0.702***	-5.868***	-0.406***	0.124***	-0.598***
	<i>Females</i>	0.452***	-1.204	-0.043	-0.027	-0.158
	<i>Environment</i>	-0.025	-7.795***	-0.479***	-0.668***	-0.411***
	<i>Constant</i>	35.983***	43.784***	2.588***	10.998***	4.037***
	<i>R²</i>	.049	.068	.076	.017	.074
<i>N</i>	5117	5117	5117	5099	5099	

*** p<0.01

Tab. 3: Regression Coefficients of Shower Data Explained by Age, Fraction of Females in Households, and Attitude Toward Environment

Age significantly determines shower behavior: Except for flowrate, age impacts all outcome variables negatively. Even though young peoples' flowrates are significantly lower, they shower longer and warmer and hence need more resources. On average, a 25-year old individual uses 3.54 kWh per shower while a 57-year old individual only needs 2.10 kWh per shower. In general, women shower significantly warmer than men. Yet, this does not influence shower consumption concerning used water and energy significantly. Attitude toward the environment does not impact the temperature of the water used, but, in fact, has an effect on shower time and consumed resources. The higher it is, the shorter are showers and less water and energy is consumed. Despite of the fact that these independent variables significantly explain variance of the outcome variables, the R^2 values indicate that their explanatory potential is relatively low: only up to 7.6% of variance (here: energy consumption) can be explained with the help of age, gender, and attitude towards the environment. We conjecture that yet-to-be-determined other factors explain shower behavior to a better degree.

4 Discussion

This paper presents the first analysis of shower data collected in a large-scaled field study. On average, a shower in our study needs 43.9 liters per shower accounting for 2.60 kWh. Increasing age and attitude toward the environment positively influence resource use in the shower: An increase by age of 10 years reduces the energy use by 0.41 kWh and an increase in attitude toward environment by 1 point (measured on a 5-point Likert scale) cuts energy use by 0.48 kWh. Analyzing the data regarding gender differences does not impact energy consumption. From the variance in measured variables, we conjecture that shower behavior is quite diverse and certainly leaves a huge potential for energy saving. Having confirmed this, research can now identify effective means to decrease energy consumption in the shower. This can be either realized by efficiency means (e.g., energy-saving shower heads) or curtailment (e.g., intervening in human behavior with real-time feedback).

The findings presented in this work are subject to two limitations. First, we have to take a Hawthorne effect into account, which might come with the smart meters being installed in the participants' showers. People might feel observed and might have therefore adapted their shower behavior accordingly because they knew that someone will analyze the collected data afterwards [MT13, Sc13]. When analyzing energy savings, the study design should always include a control group for the reason of Hawthorne effects. One can assume that consumption might be even higher than presented in our data. Second, most parts of this paper are descriptive in every respect and all findings are based on an opt-in sample of Swiss residents. The generalizability and scalability of the results have yet to be determined.

References

- [At10] Attari, S. Z.; DeKay, M. L.; Davidson, C. I.; De Bruin, W. B.: Public perceptions of energy consumption and savings. *Proceedings of the National Academy of sciences* 107/37, pp. 16054-16059, 2010.
- [BN08] Burgess, J.; Nye, M.: Re-materialising energy use through transparent monitoring systems. *Energy Policy* 36/12, pp. 4454-4459, 2008.
- [De13] Prognos AG, *Der Energieverbrauch der Privaten Haushalte 2000 - 2012, 2013*, www.bfe.admin.ch/php/modules/publikationen/stream.php?extlang=de&name=de_850806043.pdf, accessed on: 01.04.2016.
- [Di07] Diekmann, A.; Meyer, R.; Mühlemann, C.; Diem, A.: *Schweizer Umweltsurvey 2007: Codebuch ohne Prozessvariablen*, 2007, www.ethz.ch/content/dam/ethz/special-interest/gess/chair-of-sociology/dam/documents/research/umweltsurvey/2007/documentation/uws07CodebuchOhneProzessvariablen.pdf, accessed on: 05.03.2016.
- [Do13] Dormann, C. F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J. R. G.; Gruber, B.; Lafourcade, B.; Leitão, P. J.; Münkemüller, T.; McClean, C.; Osborne, P.

- E.; Reineking, B.; Schröder, B.; Skidmore, A. K.; Zurell, D.; Lautenbach, S.: Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36/1, pp. 27-46, 2013.
- [GS08] Gardner, G. T.; Stern, P. C.: The Short List: The Most Effective Actions U.S. Households Can Take to Curb Climate Change. *Environment: Science and Policy for Sustainable Development* 50/5, pp. 12-25, 2008.
- [MT13] McKerracher, C.; Torriti, J.: Energy consumption feedback in perspective: integrating Australian data to meta-analyses on in-home displays. *Energy Efficiency* 6/2, pp. 387-405, 2013.
- [Re13] EIA, Residential Energy Consumption Survey, 2013, www.eia.gov/consumption/residential/; Stand: 04.05.2016.
- [Sc13] Schwartz, D.; Fischhoff, B.; Krishnamurti, T.; Sowell, F.: The Hawthorne effect and energy awareness. *Proceedings of the National Academy of Sciences* 110/38, pp. 15242-15246, 2013.
- [Sc14] Bundesamt für Energie, Schweizerische Gesamtenergiestatistik 2014, 2014, www.connaissance-des-energies.org/sites/default/files/pdf-pt-vue/ges-stat_2014.pdf, accessed on: 04.05.2016.
- [Ti16] Tiefenbeck, V.; Goette, L.; Degen, K.; Tasic, V.; Fleisch, E.; Lalive, R.; Staake, T.: Overcoming Salience Bias: How Real-Time Feedback Fosters Resource Conservation. In AEA ASSA, Presented at the Annual Meeting, 2016.

Quantifying the Potential of Voluntary Energy Efficiency Measures: The Case of Flow Restrictors

Liliane Ableitner¹, Samuel Schöb², and Verena Tiefenbeck³

Abstract: Aside from regulatory measures and price-based instruments, many governments use voluntary energy efficiency measures as a means to reduce greenhouse gas emissions, both in the commercial and residential sector. In the domain of water heating -- the second largest domestic energy end use -- flow restrictors in showers are being contemplated as a dual-purpose policy instrument, to conserve energy and to tackle water scarcity. While the implementation of regulatory measures often faces fierce resistance by consumers and industry, also in the context of voluntary measures it is still unclear whether individuals do not simply compensate for low flow rates by taking longer showers (rebound effects). ICT can help to create more realistic estimates of the savings potential of these instruments. Based on a dataset of 5,610 individual showers, we assess the real-world influence of the flow rate on energy and water use and compare these numbers to the 45% reduction anticipated by engineering calculations. We find that users of low-flow shower heads do, indeed, take longer showers; nevertheless, they consume 38% (1.0 kWh) less energy per shower compared to the baseline level of 2.6 kWh. While more research on acceptance issues and potential biases is needed to assess the savings potential in case of mandatory measures, the findings have first important implications for policy makers and industry.

Keywords: Regulation, voluntary efficiency standards, rebound effect, energy consumption, water heating, flowrate, flow restrictors, energy efficiency

1 Motivation and Context

Today's life highly depends on energy of which demand is mostly covered by fossil fuels. The environmental (e.g., carbon emissions) and geopolitical problems (e.g., energy security) related with that are well known. The international community therefore strives for effective energy efficiency measures. Private households in Switzerland, UK and the US account for around a third of carbon emissions associated with energy production [Sc14, BN08, GS08]. Although smart meters can provide a much better data basis to

¹ ETH Zürich, M-TEC, Weinbergstrasse 56/58, CH-8092 Zürich, lableitner@ethz.ch

² Otto-Friedrich-Universität Bamberg, WIAI, An der Weberei 5, D-96047 Bamberg, samuel.schoeb@uni-bamberg.de

³ ETH Zürich, M-TEC, Weinbergstrasse 56/58, CH-8092 Zürich, vtiefenbeck@ethz.ch

determine effective efficiency measures, surprisingly little field data is still available in many domains.

In general, there are two different approaches to promote energy efficiency: technological upgrades (e.g., replacing light bulbs by energy-saving bulbs), further referred to as *efficiency* actions on the one hand; and *curtailment* actions on the other hand, which focus on prompting human behavior toward an energy reduction (e.g., turning off lights when leaving a room) [GS08]. The authors state that “efficiency-improving actions generally save more energy than curtailing use of intrinsically inefficient equipment” (p. 22).

In the residential sector, a promising candidate for energy efficiency measures is water heating: With 18% of energy demand of private US households and 13% for European households, respectively [Re13, De13], it is the second largest domestic energy use after space heating. In water-stressed regions, reductions in hot water consumption even serve the dual purpose of energy conservation and tackling water scarcity.

In many countries and several U.S. states, legislators have therefore enacted or are currently considering bills requiring plumbing fixtures sold to meet certain water efficiency standards [Wa15]. Given the large share of hot water that is consumed in the shower, shower heads with an integrated flow restrictor have moved into the focus of policy makers as a technical solution to simultaneously meet energy efficiency goals and address water scarcity. Some regions impose stringent regulatory measures, while others opt for voluntary industry standards (eco-labeling schemes). In Europe, shower heads that curb the flowrate to a maximum of 6 liters per minute are entitled to carry the label “eco” or “energy-saving” [Th15]. Yet in many regions, low-flow showerheads face severe resistance by consumers and industry [Ba09, Po10] as a paternalistic intrusion in the individual’s scope of action. Economists also criticize that that kind of imposed behavior change does not take into account individual differences in the costs of changing behavior [see, e.g., Fr92]. Yet, it is still unclear whether an installation of flow restrictors results in the desired resource savings.

One of the key arguments that is often brought forward against flow restrictors is that individuals will simply compensate by taking longer showers. Different reasons are put forward for that argument: First, from a practical point of view, certain actions like rinsing one’s hair simply require a certain quantity of water; a reduced flow rate thus merely results in a longer shower duration. Second, reduced flow rates tend to reduce perceived comfort; individuals might compensate for that by extending the duration of their shower. Third, from an economics point of view, flow restrictors may generate rebound effects, as has been shown for the purchase of efficient vehicles (“Buy a more fuel-efficient car, drive more.” – [GRW16, p. 68]): as the cost per minute of showering decreases, the individual may respond by taking longer showers.

Therefore, it is important to quantify the actual energy (and water) saving potential of flow restrictors. Aside from acceptance issues, the key question is what fraction of the savings anticipated by engineering calculations (i.e., reduction in flow rate translates entirely into

energy savings) is actually realized in practice and what fraction is compensated by increases in shower duration. The ubiquity of ICT and the “seamless integration of the physical and digital worlds through networked sensors, actuators, embedded hardware” [De15, p.4] increasingly makes it possible to collect this kind of data for better-informed policy decisions.

This paper aims to shed light on these questions based on a large set of individual shower data comprising 5,610 individual showers. The dataset was collected by 636 smart shower meters in Swiss households. We analyze the data regarding the interplay of flowrate, shower time and energy consumption. Based on this, we run a simulation with a fictive flowrate restriction. After presenting methodology and results, we discuss implications for politics and industry. In this context, the portability of the results from a voluntary setting into a regulatory setting is discussed in detail. Please note that other articles based on the same data set have been published so far; parts of the context, methodology and limitations sections therefore correspond to our previous work [Ti16].

2 Study Design, Data Collection and Sample

We conducted a large-scale field study in cooperation with a local utility that gave the study device (smart shower meter) as a gift (unconditional of participating in our study) to 5000 customers. Among these, 1- and 2-person households who agreed to answer two surveys and to make their shower data available to the researchers were eligible for the energy efficiency study with a limitation of 700 participating households (due to budget and logistics reasons). In the following 2-months of winter 2012/ 2013, study participants installed the smart meter in their showers. The technology tracks all water extractions with the help of a flowrate and a temperature sensor and stores the data on the internal memory. At the end of the study, participants sent in their devices in order to have them manually read out before they were returned to the participants for good. The course of study was split into two periods: the baseline period, which tracked shower behavior and during which the devices displayed only water temperature (first ten tracked showers on the device) and the intervention period, during which participants were provided with real-time information on their resource use (from shower eleven on). This paper focuses on the baseline data, which are not affected by the subsequent manipulations in the intervention period. Our goal is to investigate the interplay of flowrate, shower time, and energy consumption in the blindly measured baseline data⁴.

Overall, we collected shower data from 636 Swiss households located in the region of Zurich. In total, we analyze shower behavior of 975 individuals, 50% of which are female. The mean age of study participants is 46.3 years, with a standard derivation of 15.6 years. The smart meter dataset comprises shower data of a total of more than 46'000 showers.

⁴ We will not further discuss capabilities of the technology and the study period intervention. The energy savings – the main purpose of the study – are not subject of this paper but are discussed in other articles (blinded for review).

As mentioned in the previous section, baseline showers only (up to ten showers) are considered in this paper. Water extractions of below 4.5 liters were not considered a shower. With this we ensure that water extractions for cleaning or watering the plants will not distort average consumption values. Data points with more than 200 liters per shower or an average temperature of above 47°C were excluded as extreme outliers or measurement errors (85 showers). We discarded the first data point of every dataset, as its temperature and volume distribution strongly deviated from all other showers recorded. We assume that in many cases, the first water extraction was not an actual shower; instead, participants who had just completed the installation turned on the water for several seconds to see if the device worked and what information it displays. This leaves us with up to nine data point per household for our data analysis (N = 5,610) with the variables flowrate in liters per minute (M = 11.0, SD = 2.52), temperature in degree C (M = 36.0, SD = 4.09), shower time in minutes (M = 4.04, SD = 3.04) and water volume in liters (M = 43.9, SD = 33.7). Energy consumption is calculated using the standard engineering formula for heat energy ($E = m * cp * \Delta T / \eta$, with heat energy E, mass of water m, heat capacity cp, ΔT the difference between the measured water temperature and cold water temperature, and η the coefficient of energy efficiency. The latter depends on the type and age of the heating system. A study for the Swiss Office of Energy [De13] gives a detailed breakdown of residential water heating systems for Switzerland. The vast majority use fossil fuels (40% oil, 25% electric resistance heaters, 21% gas), with an average conversion efficiency of 65%. The resulting energy use per shower is (M = 2.60, SD = 2.15). To the best of our knowledge, this is the largest shower data set worldwide to date.

3 Results

In a first data analysis step, we examine whether flowrate is a predictor for shower time and energy consumption. Tab. 2 shows the regression estimation for both independent variables; Fig. 1 helps to interpret the results with a visualization.

	<i>DV: Shower Time</i>	<i>DV: Energy Consumption</i>
<i>IV: Flowrate</i>	-0.0946***	0.1681***
<i>Constant</i>	4.0440***	2.6000***
<i>R-sq</i>	.0062	.0387
<i>N</i>	5583	5583

DV: Dependent variable, IV: Independent variable, *** p<0.001

Tab. 1: Regression Output for Flowrate as Predictor of Shower Time and Energy Consumption

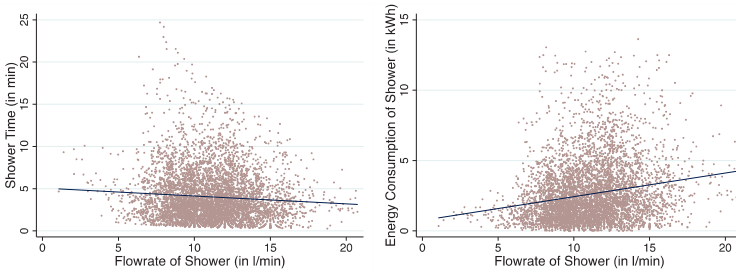


Fig. 1: Regression Plots with Trends for Shower Time and Energy Consumption

As can be seen from Tab. 1, flowrate has indeed a highly significant influence both, on shower time (negative) and energy consumption per shower (positive). Our data indicates that increasing the flowrate by 1 liter per minute decreases shower time by 0.09 minutes and increases energy consumption by 0.17 kWh. Thus, we find that individuals who have a shower head with a lower flowrate do indeed consume substantially less energy, yet they do take somewhat longer showers. Yet, the effect of combining both factors has to be determined.

To quantitatively derive the energy saving potential of low-flow shower heads, we run a simulation in which we estimate the effect of flow restrictors based on our data set. We set the flowrate to 6 liters per minute for all showers that exceed this threshold (upper margin for low-flow shower heads). First, we conduct the “engineering calculation” (i.e., assessing the impact of the technically imposed threshold, without considering rebound effects on shower duration). Thus, we calculate a new water volume with the help of shower time and the fictive flowrate and get a new energy value ((2) in Tab. 2). Taking the above presented regression results into account, however, we also have to consider that individuals adapt their shower behavior to the flowrate (i.e, they would take longer showers). Hence, in a next step, we adjust shower time according to the regression coefficient. This results in a time surcharge (water flow reduction in liters per minute multiplied by 0.09 minutes, the beta-coefficient from Tab. 1). Again, water volume is calculated with the new flowrate and – this time – the new shower time. The resulting values for energy consumption are presented in (3) in Tab. 2. Tab. 2 presents descriptive data of the original energy value (1) and the newly derived energy values (2) and (3).

<i>Energy consumption</i>	<i>Mean</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
(1) Actually consumed energy present in data set	2.60	2.15	0	13.6
(2) Simulation with low-consuming shower head	1.43	1.17	0	9.58
(3) Simulation with low-consuming shower head and adjusted shower time	1.60	1.18	0	9.67

Tab. 2: Energy Values per Shower based on different Simulations

Based on the engineering calculations (2) in Tab. 2, the replacement of inefficient shower

heads results in energy savings of 45%. The average shower now consumes 1.43 kWh instead of 2.60 kWh. However, convenience needs of humans should not be neglected: if being exposed to a lower water pressure, humans tend to take longer showers to maintain their feeling of well-being. Taking this into account in our simulation model, energy consumption of the average shower is estimated for 1.60 kWh. Conclusively, human behavior harms the potential energy savings achieved by efficiency means by 15%. Yet, the final energy reduction in the magnitude of 1 kWh per shower, which corresponds to energy savings of 38% compared to the baseline energy use, would be a great achievement. Furthermore, we find that replacing inefficient shower heads by low-energy-consuming ones would cut down variances in energy values by around 45%. As a result, flowrate is one of the main predictors of original variance in energy consumption.

4 Discussion

This paper simulates the energy saving potential of shower heads with integrated flow restrictors based on 5,610 data points of individual showers in a voluntary setting. Analyzing existing shower data, we find that decreasing the flowrate by 1 liter per minute, increases shower time by approximately 0.09 minutes and decreases energy consumption of the shower by around 0.17 kWh. Based on these interdependencies, one could expect that replacing inefficient shower heads with low-flow models would generate energy savings of 38%. These numbers take into account that 15% of absolute savings are compensated by an increased shower time due to the new flowrate. This results in an average reduction of 1 kWh per shower (or per person if we assume one shower per person per day), which corresponds to the total amount of energy spent on daily lighting in an European household [LPS14].

This finding is of high importance for governments and industry. In fact, several governments have already started to reduce hot water consumption either by specifying labeling of shower heads (voluntary eco-labeling measures) or regulatory measures that ban the sale of showerhead above a certain flow rate threshold. Our results indicate that low-flow showerhead do indeed have a substantial potential for energy and water conservation in a voluntary setting. This can be meaningful particularly for regions affected by draughts and water scarcity. The sanitary industry also benefits from this finding. The quantitative energy saving potential certainly is a value proposition for the replacement of inefficient old showerheads that can be used for marketing purposes.

The results of this study provide a first estimate of the fraction of the (engineering calculation) savings that are likely to be offset, as individuals respond to increase their shower time. Nevertheless, caution is warranted with the generalization of the results, in particular when it comes to the savings potential of regulatory (mandatory) measures. The results are not based on an exogenous allocation of showerheads with different flow rates, but on the pre-existing heterogeneity in flow rates, energy consumption, and shower time we measured in an opt-in sample of 636 Swiss households. While these regression-based

estimates are clearly informative, more research is necessary to evaluate in particular acceptance issues. After all, those individuals in our sample who had showerheads with lower flow rates had not been forced to install them, but chosen to do so (or at least accepted the existing infrastructure as it was). This raises two questions: First of all, are the individuals in our sample who had a low(er)-flow showerhead systematically different for instance in their environmental attitudes than the general population? And would the general population exhibit a similarly low degree of compensation behavior (i.e., only slightly increase their shower time), in particular in response to flow rate regulations imposed by legislators? Based on the discussion presented by [Ba09, Po10], a non-legible fraction of the population would clearly oppose that kind of regulation or try to tamper with the infrastructure if that kind of policy measure is enforced.

Another key question regards welfare analysis. Policy makers do not only need to consider what works and what does not, but also how those policies affect public welfare. Our data indicate that flow restrictors would clearly generate considerable energy savings, but the effect on overall welfare is less clear. Many individuals might perceive at least mandatory measures as an infringement of their personal freedom [Ba09, Po10].

In spite of these concerns, the results based on the dataset at hand already represent a first good estimate of the magnitude of rebound effects in that domain. The study shows that ICT can provide policy makers with large datasets from real households, which can serve as a valuable empirical foundation for better-informed policy decisions. We consider the results as a first step toward the quantification of the energy saving potential of flow restrictors. To assure generalizability and scalability of the results in a voluntary setting the same simulation should be run with other shower data sets from other regions and cultures. To go beyond self-initiation and to quantify the potential of flow restrictors in a regulatory setting, ideally, a large randomized controlled field experiment should be conducted in which the flow rate is varied exogenously.

References

- [Ba09] Ball, J.: Under Pressure: Bathers Duck Weak Shower Heads, 2009, www.wsj.com/articles/SB125807041772846273, accessed on: 01.05.2016.
- [BN08] Burgess, J.; Nye, M.: Re-materialising energy use through transparent monitoring systems. *Energy Policy* 36/12, pp. 4454–4459, 2008.
- [De13] Prognos AG, Der Energieverbrauch der Privaten Haushalte 2000 - 2012, 2013, www.bfe.admin.ch/php/modules/publikationen/stream.php?extlang=de&name=de_850806043.pdf, accessed on: 01.04.2016.
- [De15] World Economic Forum, Deep Shift: Technology Tipping Points and Societal Impact, 2015, www.wefo-rum.org/reports/-deep-shift-technology-tipping-points-and-societal-impact/, accessed on: 01.05.2016.
- [Fr92] Frank, R. H.: *Microeconomics and behavior* (4. ed.). McGraw-Hill, New York, 1992.
- [GRW16] Gillingham, K.; Rapson, D.; Wagner, G.: *The Rebound Effect and Energy Efficiency*

- Policy. *Review of Environmental Economics and Policy* 10/1, pp. 68-88, 2016.
- [GS08] Gardner, G. T.; Stern, P. C.: The Short List: The Most Effective Actions U.S. Households Can Take to Curb Climate Change. *Environment: Science and Policy for Sustainable Development* 50/5, pp. 12-25, 2008.
- [LPS14] Lapillonne, B.; Pollier, K.; Samci, N.: Energy efficiency trends for households in the EU, 2014, www.odysseemure.eu/publications/efficiency-by-sector/household/household-eu.pdf, accessed on: 01.05.2016.
- [Po10] Power, S.: A Water Fight Over Luxury Showers, 2010, www.wsj.com/articles/SB1000142405274870491330457537146-2611463490, accessed on: 07.05.2016.
- [Re13] EIA, Residential Energy Consumption Survey, 2013, www.eia.gov/consumption/residential/; accessed on: 04.05.2016
- [Sc14] Bundesamt für Energie, Schweizerische Gesamtenergiestatistik 2014, 2014, www.connaissance-des-energies.org/sites/default/files/pdf-pt-vue/ges-stat_2014.pdf, accessed on: 04.05.2016.
- [Th15] Water Label Company Limited, The European Water Label Industry Scheme, 2015, www.europeanwaterlabel.eu/pdf/schemeandtesting.pdf, accessed: 01.05.2016.
- [Ti16] Tiefenbeck, V.; Goette, L.; Degen, K.; Tasic, V.; Fleisch, E.; Lalive, R.; Staake, T.: Overcoming Salience Bias: How Real-Time Feedback Fosters Resource Conservation. In AEA ASSA, Presented at the Annual Meeting, 2016.
- [Wa15] National Conference of State Legislatures, Water-Efficient Plumbing Fixtures, 2015, www.ncsl.org/research/environment-and-natural-resources/water-efficient-plumbing-fixtures635433474.aspx, accessed on: 02.05.2016.

Determinants of adopting technological feedback devices and water conservation behaviors

Anja Peters¹, Daniel Hanss² und Christian Sartorius³

Abstract

Background and Aim

With regard to water scarcity in many parts of the world as well as the energy used for water heating in countries where water is abundant, the use of ICT for providing feedback and promoting (hot) water conservation has gained increasing attention. Insights into how consumers can be motivated to adopt technological feedback devices and intensify their water conservation efforts at home are crucial for a successful and effective distribution of such ICT devices. To the best of our knowledge, there is no substantial psychological literature on the adoption of technological water consumption feedback devices. The present study addresses this gap in the literature by investigating psychological factors that may influence consumer adoption of feedback devices as well as current water conservation behaviors.

Theory

Based on a literature review, we chose the norm-activation model (NAM) [Sc77], which is one of the dominant theoretical models in psychological research to explain environmental behaviors. With regard to its predictive power for different types of environmental behaviors, the NAM has been shown to be particularly helpful in explaining behaviors that involve low individual costs. Buying a feedback device can be considered a low-cost environmental behavior, as it may pay off over time through reduced water and energy bills. Following to the NAM-paradigm, the predictive qualities of the following explanatory variables were investigated: *personal norm* (here: feeling morally obliged to conserve water), *problem awareness* (being aware of environmental problems related to water consumption), *ascription of responsibility* (feeling personally responsible for water conservation and the environmental consequences of water consumption), *response efficacy* (being aware of means to effectively monitor and

¹ Fraunhofer ISI, Nachhaltigkeit und Infrastruktursysteme, Breslauer Str. 48, 76139 Karlsruhe, Germany, anja.peters@isi.fraunhofer.de

² Hochschule Darmstadt - University of Applied Sciences, Department of Social and Cultural Sciences and Social Work, Haardtring 100, 64295 Darmstadt, daniel.hanss@h-da.de

³ Fraunhofer ISI, Nachhaltigkeit und Infrastruktursysteme, Breslauer Str. 48, 76139 Karlsruhe, Germany, christian.sartorius@isi.fraunhofer.de

reduce water consumption at home), *self-efficacy* (believing that one can personally monitor water consumption and behave in ways to effectively conserve water in one's home), and *social norm* (believing that important others such as friends and family members expect one to conserve water). As outcome variables, consumers' *intentions to adopt feedback devices as well as current water conservation* were investigated.

Method

A questionnaire was developed which consisted of separate sections, one for each of the explanatory and outcome variables mentioned above. In addition, one section included general instructions on how to fill out the questionnaire as well as questions to assess socio-demographic information, including participants' age, gender, level of education, the household's total monthly income and size. Another section provided information on smart shower meters, their functions and how to install them. This information was presented after participants had provided their socio-demographic information and before they completed the sections measuring the NAM variables.

The questionnaire was sent out to a total of $N = 1000$ individuals aged 18 to 65 years who were randomly selected from the population registry of Darmstadt, a medium-sized German university city. The total sample of participants, after one reminder had been sent, added up to $n = 243$ individuals, resulting in a response rate of 25.18%.

Findings

Correlation and regression analyses were conducted to analyse the data. A general finding of this study is that all of the explanatory variables were positively associated with at least one of the outcome variables: adoption intention or water conservation. However, according to the results of the bivariate and the multivariate analyses, the relative importance of the explanatory variables differed depending on the outcome variable. In connection with the intention to adopt feedback devices, social norm showed the strongest association, followed by response efficacy and personal norm. In connection with water conservation, personal norm showed the strongest association, followed by self-efficacy.

Problem awareness and ascription of responsibility were positively associated with both the intention to adopt feedback devices and current water conservation in bivariate correlation analyses. However, when adjusting for common variance in multivariate regression analyses, these explanatory variables did not turn out to be significant predictors.

In sum, the regression model for adoption intention explained approximately 29% of the observed variance, while the regression model for water conservation explained 38% of the observed variance of the outcome variable.

Conclusions

Efforts to promote the adoption of smart shower meters and the conservation of warm water should target those variables that turned out to be closely related to the outcome variables in the present study.

In connection with the intention to adopt feedback devices, the findings suggest that initiatives to strengthen social norms should have strong potential to change consumers' intention. One way to increase social norms is by informing consumers about other people's interest in adopting feedback devices or actual adoption of such devices. Response efficacy could be strengthened by communicating the results of empirical studies that illustrate the effects of feedback devices (or other feedback strategies) on household water and energy conservation.

In connection with water conservation, the findings suggest that initiatives to enhance self-efficacy should have strong potential to reduce water consumption. One way how self-efficacy could be strengthened is by improving the public's knowledge of feedback technologies and how these can help monitor and reduce daily water and energy consumption.

Campaigns aimed at strengthening personal norms could target problem awareness (e.g., through information about unsustainable levels of water consumption) or ascription of responsibility (e.g., through prompts or commitment strategies). Strengthened personal norms should help promote adoption intention and water conservation behaviors.

Acknowledgements

This study was conducted within the research project DAIAD which is funded by the European Commission's 7th Framework Programme. We are very grateful to a team of students from the Hochschule Darmstadt - University of Applied Sciences, consisting of Susanna Dörr, Ralf Kaufmann, Simon Koj, Mari Schwarz, Nelli Sioud and Yohanna Weiß, for their extensive contribution during the entire research process.

Literature

- [Sc77] Schwartz, S. H.: Normative influences on altruism. *Advances in Experimental Social Science* 10, p. 221-279, 1977.

Industrie 4.0 –
Wünschenswertes, Machbares
und Grenzen für Technik,
Wirtschaft und Mensch

Industrie 4.0 – Wünschenswertes, Machbares und Grenzen für Technik, Wirtschaft und Mensch

Detlev Buchholz¹

1 Vorwort

Im Wort genommen bedeutet eine Umsetzung der Ziele von Industrie 4.0, dass man durch den Einsatz von Informationstechnologie Prozesse maximal vernetzt und automatisiert. In die Vernetzung einbezogen werden dabei alle vertikal wie horizontal vor- und nachgelagerten Prozesse: Produktionsprozesse, Logistikprozesse, Vertriebsprozesse, Unternehmensprozesse (neben Verwaltung auch Strategie, Planung und Steuerung). Alle Systeme und Prozesse sollen im Idealfall also Informationen austauschen, Prozesse planen und anstoßen sowie sich möglichst selbst steuern. Es wird somit ein IT-Integrationsgrad angestrebt, der nach jetzigem Stand der Erkenntnis nicht übertroffen werden kann.

Die Informatik ist im Besonderen aufgefordert, sich zu dem Begriff Industrie 4.0 zu verhalten. Sie hat sich im Kontext von Industrie 4.0 in sensibler Weise als übergreifend handelnde Schlüsseldisziplin zu verstehen, die die technische Realisierung der betreffenden Anforderungen nicht isoliert betrachten kann. Die Informatik ist in diesem Zusammenhang beispielweise der Frage nach der Organisation und Gestaltung der Übergänge zwischen Mensch und Maschine verpflichtet. Hierbei geht es um den Menschen in der Arbeitswelt wie um den Menschen als Konsument.

Zusätzlich wirft die mit Industrie 4.0 verbundene „Renaissance“ länger bekannter KI-Techniken, beispielsweise durch semantische Web-Services oder automatisches Planen, klassische Fragen der Technikfolgenabschätzung auf. Das sind Fragen zu den Grenzen der Machbarkeit, der Komplexität, aber auch ethische Fragen sowie Fragen nach tatsächlichen Mehrwerten.

Schließlich stellt sich auch die Frage nach der Auswirkung auf unser Mensch-Sein. Die in der Literatur vorherrschende, rein marktorientierte Sichtweise blendet die Frage nach den tatsächlichen Bedürfnissen des Menschen aus. Den Nutzen in der Arbeitswelt mag man mindestens anhand singulärer Werkzeuge in einzelnen Zusammenhängen noch einsehen. Die Frage nach den tatsächlichen Bedürfnissen aber zielt auf die Schwachstelle des im Kontext von Industrie 4.0 immer wieder beschworenen Prinzips, dass man mittels Smart Production eben immer konkreter und immer filigraner auf die individuellen Wünsche des Kunden (= des Menschen) eingehen könne. Nämlich jene Schwachstelle, dass damit nicht geklärt ist, was dem Menschen bei der Ermittlung oder bei der Artikula-

¹ Euro-FH Hamburg, buchholz.euro-fh@mail.de

tion seiner Bedürfnisse hilft oder was ihn dazu befähigt, im Sinne eines aufgeklärten, mündigen Menschen, seine Bedarfe selbstständig in vernünftiger Weise und seinen wahrhaften Bedürfnissen gemäß zu bestimmen und zu formulieren.

Auf die Informatik kommt damit eine deutlich größere Verantwortung zu, als bei herkömmlichen Entwicklungen. Industrie 4.0 hat Auswirkung auf die Arbeitswelt, auf den einzelnen Menschen, auf unsere Gesellschaften.

2 Beiträge und Begutachtung

Im Rahmen dieser allgemeinen Fragstellung behandelt der Workshop folgende Einzelthemen:

- Wirtschafts- und Sozialphilosophische Kritik
- Technikfolgenabschätzung anhand der Digitalisierung in der Bauindustrie
- Industrie-4.0-Readiness
- Datenhoheit und Datenschutz
- Reifegradmodelle

Für den Workshop wurden vier Artikel in der Langfassung und ein Artikel als Positionspapier angenommen. Alle Beiträge wurden von je zwei Mitgliedern des Programmkomitees unabhängig begutachtet.

3 Das Programmkomitee

Dem Programmkomitee gehören – in alphabetischer Reihenfolge – an:

- Dr.-Ing. Detlev Buchholz, Euro-FH Hamburg
- Prof. Kai Hofmann, Euro-FH Hamburg
- Dipl.-Logist. Christian Prasse, Fraunhofer Institut für Materialfluss und Logistik (IML) Dortmund
- Prof. Dr.-Ing. Isabel Schaller, Euro-FH Hamburg
- Dr.-Ing. Frank Schulze, Institut für Technische Logistik und Arbeitssysteme, Technische Universität Dresden
- Prof. Justus Theinert, Fachbereich Gestaltung, Hochschule Darmstadt

Allen Mitgliedern ganz herzlichen Dank für die Mitwirkung sowie die sehr hilfreichen inhaltlichen Impulse. Besonderer Dank gilt den KollegInnen Schaller und Hofmann für die koordinative Unterstützung innerhalb der Euro-FH.

Den Teilnehmenden und LeserInnen wünsche ich, auch im Namen des Komitees, eine einblicksreiche und anregende Lektüre sowie spannende Diskussionen.

Industrie 4.0 und der Mensch

Detlev Buchholz¹

Abstract: Eine für Unternehmensentscheidungen wie für die betriebliche IT-Realität hilfreiche Betrachtung von Industrie 4.0 orientiert sich üblicherweise am Nutzen, an dem Fortschritt für das einzelne Unternehmen. Doch wie sieht es mit dem Nutzen für eine gesamte Gesellschaft, oder weiter gefasst: mit dem Nutzen für den Menschen an sich aus? Aus der Betrachtung des Vorgehens bei einer betrieblichen Industrie 4.0.-Migration, seiner technischen, wirtschaftlichen wie ethischen Vor- und Nachteile aus Unternehmenssicht sollen Grundsatzfragen zum menschlichen Nutzen und damit zu Bedürfnissen des Menschen an sich entwickelt werden. Dazu wird sowohl eine wirtschaftsphilosophische als auch eine sozialphilosophische Sichtweise herangezogen. Die aufgeworfenen Aspekte sollen IT-Verantwortlichen wie Unternehmensentscheidern als Anregung und erste Anleitung zu einer ernsthaften Folgenbetrachtung transformativer technischer Entwicklungen dienen.

Keywords: Industrie 4.0-Readiness, Migration, Vorgehen, Technikfolgenabschätzung, Ethik, Sinn des Lebens.

1 Einleitung

In den Betrachtungen über Industrie 4.0 begegnet uns zuweilen der Befund, dass die Komponenten, die die technologische Basis von Industrie 4.0 bilden, zum Teil schon seit Jahrzehnten existieren. Im Kern bedeutet Industrie 4.0 also hauptsächlich, diese Technologien konsequent miteinander zu verknüpfen.

Warum ist man auf diese Verknüpfungsmöglichkeit nicht bereits früher gekommen? Möglicherweise stieß der Gedanke, eine Fabrik organisieren sich hauptsächlich vollautomatisch, in der Unternehmensführung auf Vorbehalte. Das können *ethische* Vorbehalte sein oder auch eine intuitive Skepsis gegenüber *Machbarkeit* und *Mehrwert* der Automation. *Technische* Vorbehalte könnten darin liegen, dass hohe Laufzeiten und erforderliche teure Hardware lange Zeit keinen ernsthaften Einsatz beispielsweise von KI-Technik erlaubten.

Was hat sich inzwischen geändert? Rechenleistung ist heutzutage selbst zu E-Business-Ware geworden ist, die in prinzipiell beliebiger Mächtigkeit und mit beliebig kurzen Dauern und daher mit geringem Investitionsrisiko gemietet werden kann. Damit erledigen sich technische Hindernisse zumindest zum Teil. Vorbehalte in der Unternehmensführung wiederum hängen nur wenig mit unterschiedlichen Generationen zusammen: Auch unsere Väter und Großväter haben innovativ gedacht, vielleicht sogar innovativer

¹ Euro-FH Hamburg, buchholz.euro-fh@mail.de

und gleichzeitig klüger als wir. Vielmehr haben sich in einer Welt des prinzipiell entfernungs- und grenzenlosen Marktes die Vorbehalte durch die Konkurrenz mit Kulturen bzw. Nationen, in denen der Gedanke der Automation auf deutlich geringere Hemmnisse trifft (Ostasien), auch für uns relativiert – darin ist sich zumindest die einschlägige Literatur einig [Ba2014, S. 5ff; PFV2014, S. 160ff; We2014, S. 346; Kag2014, S. 603ff].

So erinnert man sich nun an Innovationen, die das Forschungs- bzw. Simulationsstadium bisher nicht verlassen haben oder nach ihrer Beforschung mangels Interesse zunächst nicht weiterverfolgt wurden.

2 Die Technologie

Die Technologie an sich ist, wie gesagt, an sich nichts Neues, dennoch passiert etwas Revolutionäres: Ihre Nutzung erfährt eine bisher nicht dagewesene Breite und entwickelt sich zu einem primär relevanten Faktor der Konkurrenzfähigkeit.

2.1 Industrie-4.0-Readiness

Wie weit ist mein Betrieb bereit zum Übergang zur Industrie 4.0? Betrachten wir dazu anhand von sieben grundlegenden Eckpunkten das Wesen von Industrie 4.0:

1. Gerätschaften, Transportmedien und Produkte (der dinglichen Welt) werden mit eingebetteten Systemen ausgestattet, es kann ihnen somit ein bestimmtes Verhalten einprogrammiert werden. Zu den programmierten Fähigkeiten gehört – mittels Methoden der Künstlichen Intelligenz bzw. der Logik-Programmierung – das selbständige Entscheiden und Planen von Prozessen und Wegen.
2. Die Gerätschaften können über Sensoren Informationen über ihre Umwelt/ihr Umfeld aufnehmen und als Eingabe für die unter Punkt 1 genannten Programme verwenden.
3. Die Gerätschaften haben Aktuatoren, die wiederum Umwelt bzw. Umfeld verändern.
4. Die Gerätschaften und Produkte sind mit anderen Gerätschaften und Produkten vernetzt. Entscheiden und Planen kann somit als kooperativer Vorgang mehrerer Gerätschaften als verteiltes System realisiert werden.
5. Die Gerätschaften haben Zugang zum Internet (auch als semantisches Internet) und können somit Informationen als Input (sowohl als T-Box-Wissen als auch als A-Box-Wissen) verwenden. Sie sind damit selbst Bestandteile des Internets. Das Internet wird damit zum Internet der Dinge.

6. Der Mensch (als Beschäftigter) ist über „intelligente“ Assistenzsysteme an den betreffenden Prozessen, Entscheidungen und Planungen beteiligt. Er programmiert, kalibriert, überwacht und korrigiert.
7. Die Beschäftigten sind mittels sozialer Medien miteinander vernetzt.

Bei jedem einzelnen Punkt kann man also sagen: Das gibt es längst, das ist längst (zumindest partiell und lokal) realisiert. Man spricht in der Literatur auch von den sogenannten „Enabling Technologies“ [Bi2014]. Das sind jene Technologien, die den Gedanken der Industrie 4.0 befördern, unter anderem schnellere Rechner, größere Rechner, Rechnerverbünde durch die starke Verbreitung von Grid- und Cloud-Computing oder die bessere Vernetzung von Diensten durch einheitliche Modellierungsstandards im Web (v.a. W3C-Standards) und durch Web-Services. Auch Auto-ID, eingebettete Systeme, ethernetbasierte Feldbusse, OPC-UA und Soft-SPSen oder breitbandige, kabellose Netzwerke sind zwischenzeitlich verlässlich zu nutzen und werden stetig weiterentwickelt.

2.2 Entscheidungsfindung

Die Industrie 4.0-Readiness beginnt somit nicht bei der technischen Bereitschaft, sondern sie beginnt „im Kopf“. Wie weit ist die Unternehmensführung bereit, sich auf Industrie 4.0 einzulassen? Was ist notwendig, um konkurrenzfähig zu sein? Was kann man aus Industrie 4.0. herausziehen, um sich einen Vorsprung am Markt zu verschaffen?

Hierbei kommt auf die Informatik bereits eine deutlich größere Verantwortung zu als bei herkömmlichen strategischen Entscheidungen eines Unternehmens: Wir müssen nicht nur die technische Voraussetzungen des Betriebs sorgfältig einschätzen und die zu erwartenden Investitionskosten in einem besonderen Maße realistisch vorhersehen. Entscheidend ist unsere Einschätzung zur Machbarkeit – gerade wegen der steigenden Realisierbarkeit der erforderlichen Technologie im Einzelnen.

Welche Fähigkeiten der Selbstorganisation von Maschinen sind denn wirklich für den Betrieb angemessen? Welche davon sind gegenüber dem Einsatz von menschlicher Handlungskompetenz tatsächlich im Vorteil – sei es im Zeitvorteil oder weil Maschinen besser den Überblick behalten, wenn die Zahl von Handlungsverzweigungen die menschliche Kognitionsfähigkeit übersteigt? Die Erörterung dieser Fragen alleine kann je nach Größe des Betriebs zu einer Forschungsarbeit ausarten. Die Vorbereitung entsprechender Entscheidungen können Monate in Anspruch nehmen, wenn diese sorgfältig getroffen werden sollen. Daraus ergeben sich weitere – eher menschliche – Fragen: wie hält man solange das Top-Management geduldig? Was bedeutet eine sich Richtung Industrie 4.0-Migration abzeichnende Entscheidung für die Führungsarbeit gegenüber den Beschäftigten, die nicht wissen, was auf sie zukommt? Wie werden diese verständlich und dennoch treffend und gleichzeitig einführend informiert?

Eine vollumfassende Migration in einem Schritt ist sowohl in einem großen Betrieb (aufgrund der Komplexität) als auch in einem KMU-Betrieb (aufgrund der hohen Investition) nicht ratsam. Ein Ausnahme mag die Neugründung sein, bei der ein Industrie 4.0-Betrieb von vornherein strategisches Planungsziel sein kann. Im übrigen aber wird die „4. industrielle Revolution“ eher eine Evolution sein.

2.3 Überlegungen zum Vorgehen

[Bi2014] stellt ein Vorgehensmodell für die Industrie 4.0-Migration vor, das sowohl die Vorsicht und Verantwortung der Entscheider und Entwickler und das Prinzip, Industrie 4.0-Technologien nach und nach einzuführen, gut und schlüssig miteinander verbindet. Dieses Modell sieht die folgenden sieben Stufen vor:

1. Herstellung von Verständnis, Erkenntnis und Beteiligung: Erarbeitung von Fachinformationen, Diskurse mit Führungskräften, Besuch von Referenzfirmen, Austausch auf der Plattform I4.0 (<http://www.plattform-i40.de>)
2. Entwicklung einer umfassenden Liste von Use Cases (zur Wahrung einer einheitlichen Strategie), dazu dezentrale Entwicklung von Use Cases mit dem Fokus Echtzeit und Nutzung von Software Services. Erste grobe, aber ehrliche Kosten-Nutzen-Analyse.
3. Auswahl einer engeren Wahl an Use Cases im Diskurs mit Führungskräften (Kriterien: bestes Kosten-Nutzen-Verhältnis und geringe Umsetzungsrisiken); Aufstellen eines Projektprogramms
4. Beteiligung von Beschäftigten, Betriebsrat, Gleichstellungs-, Sicherheits-, Gesundheits- und Behinderten-Beauftragten sowie von Kunden und Lieferanten
5. Umsetzung der unter 3. getroffenen, engeren Wahl als Pilotprojekte („80/20“-Regel), Evaluation von Nutzen, Kosten und Dokumentation der Umsetzungserfahrungen
6. Plan zur Ausweitung der erfolgreich evaluierten Use Cases auf das gesamte Unternehmen, Präzision des Projektprogramms
7. Fortsetzung der Umstellung unter kontinuierlicher Evaluation; Hinzunahme weiterer Use Cases.

Aus der Stufe 4 können wir deutlich herauslesen, dass Industrie 4.0 kein Selbstzweck sein kann. Auch Kunden müssen von den Vorteilen überzeugt sein (beispielsweise eine schnellere und individuellere Umsetzung von Kundenanforderungen auch bestätigen können). Und die Lieferanten müssen mitspielen. Sie dürfen von neuen Anforderungen an die Zulieferung oder auch an die intrabetriebliche digitale Vernetzung nicht überrascht werden. Die Zulieferung ist schließlich ein sensibles Bindeglied in der Prozesskette.

Das 7-Stufen-Modell ist nachdrücklich auf ein schrittweises Vorgehen angelegt. Prozesse bzw. Use Cases werden gemäß einer vorher definierten Priorität nach und nach umgestellt. Somit kann für die jeweils als nächstes zu erfolgenden Umstellungen die Erfahrung aus dem früheren Vorgehen berücksichtigt werden. Das verringert Investitions- und Ausfall-Risiken.

Bis hier hin sei zunächst der aktuelle Stand wiedergegeben, inwieweit der Mensch in die Entscheidungen der Industrie 4.0 üblicherweise einzubeziehen ist. In der folgenden Debatte möchte ich das eingangs dargelegte Ausmaß an Konsequenzen der Industrie 4.0 aufgreifen, um nicht das Machbare oder das unreflektiert Wünschenswerte, sondern den Menschen in den Mittelpunkt stellen. Diese Debatte schließt einerseits an die bisherigen Ausführungen an, stellt aber bezogen auf den Blickwinkel ein Gegenentwurf zu den üblichen Überlegungen zur Industrie 4.0-Migration dar. Sie wirft dadurch tiefergehende Fragen auf – darunter auch volkswirtschaftliche, gesellschaftliche wie politische Fragen.

3 Der Mensch

Industrie 4.0 führt man nicht ein wie irgendeine spezifische Neuerung oder eine neue Software. Bei Umstellung auf Industrie 4.0 handelt es sich eben nicht um ein einzelnes Projekt, sondern um eine grundsätzliche Transformation. Die damit verbundene, besondere Verantwortung wahrzunehmen bedeutet, sich über die Konsequenzen im klaren zu sein, insbesondere wenn man innerhalb dieser Entwicklung eine gestaltende, vielleicht sogar steuernde Rolle ausfüllt. Das erfordert auch, dass man sich selbst und sein eigenes Leben im Rahmen dieser Entwicklung ernsthaft betrachtet.

In diesem Sinne werde ich die Konsequenzen von Industrie 4.0 für den Menschen von einfachen praktischen Annehmlichkeiten bis hin zu der Frage nach dem Sinn des Lebens beleuchten.

3.1 Unmittelbar einsichtige Vorteile der Industrie 4.0

Dem einfachen Skeptiker, der behauptet, bei Industrie 4.0 ginge es nur um Spielerei oder um Fortrationalisierung menschlicher Tätigkeiten, sei entgegenghalten, dass allein der Abbau von Medienbrüchen für den Menschen unmittelbare Vorteile hat. Als Hauptargument für die Einführung von Industrie 4.0 wird in der einschlägigen Literatur jedoch ein besseres Eingehen auf Bedarfe des Kunden angeführt [u.a. Ba2014; Da2014; We2014]. Tatsächlich ermöglicht Industrie 4.0 eine deutlich filigranere Auffächerung von Gestaltungsoptionen der Produkte. Produkte können damit noch stärker nach individuellen Bedarfen modelliert und dem individuellen Geschmack angepasst werden, ohne kostenintensive, spezifische Gerätschaften vorzuhalten. Die Gerätschaften kalibrieren sich und die Prozesse, in die sie eingebunden sind, weitgehend selbsttätig. Die dazu notwendigen digitalen Ressourcen (Hardware wie Software) können als Services angemietet werden, was weitere Kostensenkungen erlaubt.

Die Präzision der Maschinen einer smarten Fabrik geht in eine neue Qualität über. Die Anlagen reagieren auf Schwankungen der Ausgangssituation deutlich flexibler, in Echtzeit und mit einer sehr hohen Reproduzierbarkeit und Präzision.

Damit sind mehrere Vorteile, die sich üblicherweise wechselseitig entgegenstehen, in Einklang gebracht: Berücksichtigung individueller Kundenwünsche, geringe Produktionszeit, niedrige Produktionskosten und hohe Produktqualität.

Ein weiterer wichtiger Vorteil auf der Unternehmensseite ist, dass cyberphysische Systeme noch stärker als die bisherige Automation zur körperlichen Entlastung der Beschäftigten beitragen. Hebehilfen und z.B. Exoskelette, die über das Internet konfiguriert werden und sich während des Einsatzes automatisch adaptieren bzw. nachkalibrieren, unterstützen Ergonomie am Arbeitsplatz und reduzieren körperlichen Verschleiß. Ubiquitous-Computing Komponenten passen sich an die Fähigkeiten ihrer Bediener an – z.B. an Sehbeeinträchtigungen oder an etwaige Fehlerängste.

Zudem wird die Arbeit des Menschen von ermüdender, als unnötig empfundener Routine befreit. Gleichzeitig wird die Verfügbarkeit von Information erhöht. An dieser Stelle ist der Mensch dann stärker gefordert: Durch die erhöhte Informationsdichte und die Verlagerung von Routine- zu Entscheidungsaufgaben erhält seine Arbeit deutlich mehr Sinngehalt. Dies ist natürlich nur dann ein Gewinn, wenn er den Umgang mit der Information und seiner neuen Verantwortung auch meistert, doch darauf komme ich später zu sprechen.

Entscheidungskompetenzen – und das beinhaltet nicht nur die Entscheidungspflicht, sondern auch das Entscheidungs*recht* – werden so in die der Führungsebene nachgelagerten Bereiche hineingetragen. Die Hierarchien werden flacher. An der operativen Basis einer Produktion entsteht also mehr Arbeitsinhalt, mehr Sinngehalt und letztlich auch die Chance auf eine höhere Entlohnung der Arbeit.

Zudem erlaubt die digitale Vernetzung, die mit einer Verlagerung körperlicher zu geistigen Tätigkeiten einhergeht, Arbeitszeiten flexibler zu gestalten. Sowohl Zeit als auch Ort der Aufgabenerfüllung können durch den Einzelnen stärker mitbestimmt werden und hängen mehr von erforderlicher Kompetenz und Fähigkeit sowie vom aktuell gefragten Einsatz ab als von einer Stechuhr.

Spätestens beim letzten Argument wird deutlich, dass dieser Vorteil nicht mehr für jeden unmittelbar einsichtig ist: Es gibt sowohl Führungskräfte, die sich von dem Gedanken der starren Zeiterfassung nur schwer lösen können, als auch Beschäftigte an der produktiven Basis, denen ein klar strukturierter Arbeitsalltag lieber ist als gestalterische Freiheit. Doch das betrifft wieder den Sinn des Lebens, und dieser Frage will ich mich im kommenden Abschnitt wieder etwas weiter annähern.

3.2 Wofür wird in der Industrie 4.0 der Mensch gebraucht?

Wir haben im vorangegangenen Abschnitt die neue Entscheidungskompetenz in Betrieben und ihre bisher nicht dagewesene Verbreitung angesprochen. Wir sehen im Zusammenhang mit Industrie 4.0 aber auch Methoden des automatisierten Entscheidens und Planens. Diese Methoden werden in singulären Produktions- und Managementzusammenhängen durchaus erfolgreich eingesetzt. Die *allumfassende* Smart Factory hat aber das Forschungsstadium weiterhin nicht verlassen. Dies trifft auf viele entsprechende Konzepte seit Anbeginn der KI-Forschung Ende der 1950er Jahre zu.

Wolfgang Wahlster wird dazu in [SGG2013, S. 130] folgendermaßen zitiert:

„[...] nach über 35 Jahren Forschung im Bereich künstliche Intelligenz bin ich überzeugt, dass jeder Grundschüler selbst den besten intelligenten Computersystemen bezüglich seiner Alltagsintelligenz überlegen ist. □ Bei Entscheidungen unter Unsicherheit, die nicht einem klaren Muster folgen, und im sensor-motorischen Bereich haben Menschen daher einen klaren Vorteil selbst gegenüber den leistungsfähigsten Maschinen aus dem KI-Bereich.“

Das bedeutet, wir fangen nach Jahrzehnten der Technikgläubigkeit bzw. des Szientismus [Kam1979, S. 301] wieder an, Intelligenz, Kreativität, Einfühlungsvermögen und Feinmotorik als rein menschliche Fähigkeiten anzusehen, deren Komplexität (in formalen Sinne gebraucht) eine technische Modellierung als kaum sinnvoll erscheinen lässt.

Die komplette Automatisierung von Entscheidung und Planung ist zudem sehr heikel. Nehmen wir eine auf Business Intelligence-Methoden oder Machine Learning basierende Selektion von Erfahrungswissen: Wenn dieselben Maschinen ihr „Erfahrungswissen“ wieder ins Internet der Dinge und der Dienste zurückspeigeln und daraus wieder neues Erfahrungswissen automatisch abgeleitet wird, dann dreht sich das System irgendwann im Kreis: Maschinen lernen von Maschinen, ohne dass neue Intuitionen einfließen können. Gleichwohl: Mittels Zufallskomponenten können Maschinen zu Lösungsvorschlägen kommen, an die ein Mensch noch nicht gedacht hat. Darunter mögen durchaus anwendbare und sinnvolle, ja als „innovativ“ zu bezeichnende Vorschläge sein. Die Stärke des Rechners ist hier, dass er trotz notwendiger Einschränkung durch Heuristiken innerhalb kürzester Zeit immer noch eine immens hohe Zahl an Entscheidungs- bzw. Handlungsoptionen bewerten kann. Eine Zahl, die durch den Menschen nicht zu bewältigen ist. Zufallskomponenten können die Heuristiken in entlegene „Ecken“ des Entscheidungsraums führen (das zeigen beispielsweise automatische Kompositionssysteme, s. [Zi2001]). Ob maschinell hervorgebrachte Lösungen auch sinnvoll sind, kann aber nur der Mensch entscheiden. Dann jedoch liegt keine *vollständige* Automatisierung mehr vor.

Menschliche Assoziationsfähigkeit und Intuition sind für Innovationsprozesse unentbehrlich, egal ob rechnergestützt oder nicht. Rechnerstützung ist dennoch hilfreich, um Komplexität zu reduzieren und dem Menschen zu neuen Inspirationen verhelfen, nur eben nicht in einer verlässlichen Weise. Die Maschine ist und bleibt also Werkzeug, ein

Werkzeug, das zwei grundlegende Fragen in einem bisher nicht gekanntem Ausmaß aufwirft:

1. Bis zu welchem Grad (von der rudimentären Nutzung bis hin zur fast vollständigen Automation) kann ich mich auf das Werkzeug verlassen?
2. Welchen effektiven Nutzen hat das Werkzeug für mich?

3.3 Unbeantwortete Fragen

Diese beiden Fragen will ich im folgenden anhand konkreter Einzelfragen beleuchten. Sie sind kaum nur im Rahmen einer Vorgehensplanung zu beantworten. Wie derart komplexe Systeme, um die es hier geht, letztendlich auf jene wirken, die damit täglich umgehen müssen, kann nur erfahren werden, indem man die Konsequenzen *erlebt*. Und dabei wird es intersubjektiv erhebliche Unterschiede geben. Schon einige Male ist unsere Geschichte technischer Entwicklungen so verlaufen: Ist ein System erst einmal in der Welt, müssen wir uns damit abfinden, müssen wir uns fügen. Eine Rückentwicklung oder den Versuch dazu erlebt man nur in besonders drastischen Fällen (Beispiel: Atomenergie). Um so wichtiger ist eine besonders umsichtige Vorausschau. Aus der Bewertung der technischen Machbarkeit und vielleicht noch der unmittelbaren Auswirkungen auf die Arbeitsstätte können aus Gründen der Komplexität der Systeme keine verlässlichen Antworten erwartet werden.

Vielleicht hilft eine Veränderung des Blickwinkels. Ab welchem Grad ist Automation schädlich für das Mensch-Sein, für das Funktionieren unserer Gesellschaften? Banales Beispiel: Die weiter oben angesprochene Befreiung von stupider Routine und vor allem von einseitiger, auf Dauer gesundheitsschädlicher körperlicher Arbeit, führt bereits seit einiger Zeit zur Diskussion über das Renteneintrittsalter. Damit bringt uns die Bewertung der weiteren Entwicklung von Automation in ganz typische Zielkonflikte, in Situationen, in denen negative und positive Effekte am besten im Vorhinein erkannt werden sollten. Aber allein eine beispielhaft aufgeführte Frage (nach dem Renteneintrittsalter) ist nicht zu beantworten, ohne eine tatsächliche entsprechende Entwicklung bzw. ohne den Ausgang dann eintretender Diskussionen – einschließlich des Zustandekommens und Verhaltens parlamentarischer Zufalls-Mehrheiten – zu verfolgen; und das ist – nach Erfahrung mit vergleichbaren Entwicklungen wie der Atomenergie – ein sehr langfristiger Prozess, zumindest gemessen an der Dauer eines Menschenlebens.

Ganz gleich, wie Sie diese Diskussionen bzw. Prozesse bewerten, es gibt genügend Menschen, die sich anhand solcher und weiterer Konsequenzen fragen: Ist der geschaffene Nutzen wirklich insgesamt lohnend? Und das führt zu der eigentlich interessanten, aber in unserer Gesellschaft in der Breite dann wiederum doch verdrängten Frage: Was ist wirklich von Nutzen? Was ist überhaupt ein Nutzen?

3.4 Letzte Fragen

In der einschlägigen Literatur über Industrie 4.0. läuft die Antwort nach dem Nutzen auf eine Orientierung am Markt hinaus. Als ein Beispiel möchte ich hier Bauernhansl [Ba2014, S. 30] anführen:

„Im Zusammenhang mit Industrie 4.0 wird fast ausschließlich über Technologien und cyber-physische Systeme gesprochen. Die Einführung in die Industrie soll technologiegetrieben (Technology Push) erfolgen. Die Industrie braucht aber einen nutzenorientierten Migrationsansatz, der als Change-Management-Prozess zu verstehen ist. Also den Nutzen für ‚Kopf‘ – Rationale [sic!] Ebene und ‚Bauch‘- Emotionale [sic!] Ebene nachvollziehbar macht. Der Nutzen aber entsteht im Markt, im eigentlichen Geschäftsmodell. Daher muss die Argumentation über die Nutzenpotenziale, über den Markt kommen. Damit kann man den beteiligten Firmen transparent machen, dass es sinnvoll ist, in diese Veränderung zu investieren. Das ist nicht nur eine Investition in neue Technologie, sondern auch eine Investition in neue Geschäftsmodelle, in neue Organisationsformen, in die Mitarbeiter.“

Wirtschaftsphilosophische Kritik

Solchen Überlegungen hält bereits Kambartel [Kam1979] eine methodische Kritik der gängigen quantitativ-ökonomischen Prinzipien entgegen. Nach diesen Prinzipien stelle sich der „Markt“ als eine „ständige dynamische Volksabstimmung“ [ebd., S. 311] dar. Konkret führt Kambartel dazu aus:

„Selbst wenn es möglich wäre, auf der Basis des faktischen Bedarfs zeitabhängige, jedoch im Zeitquerschnitt intersubjektiv aussagbare Nutzenquanta zu konstituieren, so wäre doch eben ein rationaler Zusammenhang der darauf gestützten quantitativen Aussagen mit der bedürfnisbezogenen Aufgabe der Ökonomie methodisch [...] nicht herzustellen.“

Kambartels Gedankengang beginnt damit, dass er die Nachfrage auf dem Markt (Bedarf) darauf hin befragt, ob sie tatsächliche Bedürfnisse befriedige, d.h. wirklichen menschlichen Nutzen schaffe. Schließlich ergebe sich ein großer Teil des Güter- und Dienstleistungsbedarfs lediglich aus den *Voraussetzungen*, um eigentliche Bedürfnisse zu befriedigen, und nicht aus diesen Bedürfnissen selbst: Wer z.B. ein Auto benutze, um damit zur Arbeit zu fahren, dem sei diese Fahrt eben kein Bedürfnis. Es handele sich vielmehr um einen Bedarf, der daraus erwachse, einen anderen Zweck zu erreichen (nämlich zur Arbeit zu kommen). Auch dieses Ziel, auf der Arbeit zu sein, müsse wiederum kein Bedürfnis sein, sondern sei womöglich selbst wieder Mittel für weitere Zwecke (Geld verdienen etc.). Eine rationale Nachfrage liege nur dann vor, wenn am Ende solcher Mittel-Zweck-Ketten echte Bedürfnisse ständen, was nicht unbedingt der Fall sein müsse.

Diese Unterscheidung zwischen unseren wahren Bedürfnissen und unserem „faktischen Begehren und Wollen“ besitzt für Kambartel eine unhintergehbare „lebensweltliche Evidenz“ [Kam1979, S. 308] und betreffe „ein wesentliches Problem unserer Lebensori-

entierung“ [ebd., S. 310]. Auf unseren Zusammenhang bezogen, bedeutet das: Die Möglichkeiten der smarten Produktion geben aus sich heraus noch keine Antwort auf die Frage, wann sich menschliches Kaufverhalten (Nachfrage) mit tatsächlichen Bedürfnissen deckt.

Wer sich diese Frage nicht stellt, umgeht letztlich die Frage nach dem Sinn des Lebens. Diesen Zusammenhang behandelt ganz explizit Buchholz [Bu2009; Bu2016, S. 26ff]. Auch Buchholz macht darauf aufmerksam, dass das liberale Wirtschaftsmodell auf einem unhinterfragten und nicht näher begründeten Menschenbild beruhe, dessen Prinzipien „zunehmend auch die Wertvorstellungen des Privatlebens infiziert“ [Bu2009, S. 131]. An die Kant'sche Definition von Aufklärung erinnernd, weist er zudem nach, dass der liberale Marktbegriff mündige Konsumenten voraussetzt, die sich ihres Verstandes „ohne Leitung eines anderen“ bedienen – was nicht ohne weiteres faktisch der Fall sei [Bu2009, S. 131; Kan1923, S. 35]. Das zielt genau auf die Schwachstelle des im Kontext von Industrie 4.0 immer wieder beschworenen Prinzips, dass man mittels Smart Production eben immer konkreter und immer filigraner auf die individuellen Bedürfnisse bzw. Wünsche des Kunden (= des Menschen) eingehen könne. Nämlich jene Schwachstelle, dass damit nicht geklärt ist, was dem Menschen bei der Ermittlung oder bei der Artikulation seiner Bedürfnisse hilft oder was ihn dazu befähigt, im Sinne eines aufgeklärten, mündigen Menschen, seine Bedarfe selbstständig in vernünftiger Weise und seinen wahrhaften Bedürfnissen gemäß zu bestimmen und zu formulieren. So kann beispielsweise der „Geschmack des Kunden“ (Abschnitt 3.1), an den sich eine smarte Produktion fast individuell anpassen lässt, auch eine Geschmacklosigkeit bedeuten, die zur Entästhetisierung unserer Welt und zur Entkultivierung unseres Lebens beiträgt.

Sozialphilosophische Kritik

Zu einer ähnlichen Bewertung der rein quantitativ-ökonomischen Begründung von Automation kommt Hannah Arendt in ihrem bereits 1958 erschienenen Werk *Vita Activa* [Ar1961]. Sie betrachtet die am Ende von Abschnitt 3.2 aufgeworfene zweite Frage nicht, wie Kambartel oder Buchholz, aus Sicht des Kunden bzw. Konsumenten, sondern aus Sicht der Arbeitswelt.

Zunächst bringt Arendt die Vertracktheit unserer Frage nach dem effektiven Nutzen der Automation in folgender Weise auf den Punkt [Ar1961, S. 32]:

„Die häufigen Klagen, die wir über die Verkehrung der Mittel in Zwecke und umgekehrt der Zwecke in Mittel in der modernen Gesellschaft hören: daß die Mittel sich als stärker als die Zwecke erweisen und daß der Mensch der Knecht der Maschinen wird, die er selbst erfunden hat, daß er sich ihren Erfordernissen anpaßt, anstatt sie als bloße Mittel für menschliche Zwecke und Bedürfnisse zu nutzen – haben ihre Wurzel in der tatsächlichen Situation des Arbeitens. Denn für das Arbeiten, das ja primär in einer Präparierung von Gütern für den Konsum besteht, ergibt die für die Herstellung so außerordentliche Unterscheidung zwischen Zweck und Mitteln einfach keinen Sinn weil in ihm Zweck und Mittel gar nicht getrennt genug auftreten, um überhaupt scharf auseinandergehalten und geschieden werden zu können. Daher verlieren die vom Homo faber erfundenen

Instrumente und Werkzeuge, mit denen er dem Animal laborans bei seiner Arbeit zu Hilfe gekommen ist, sofort ihren instrumental Charakter, wenn sie erst einmal in den Arbeitsprozeß eingegangen sind. So ist es auch müßig, an das Leben und den Lebensprozeß, von dem die Arbeit einen integrierenden Teil bildet und den sie als solchen niemals übersteigt, Fragen zu stellen, die die Zweck-Mittel-Kategorie voraussetzen, also z.B. zu fragen, ob der Mensch lebt und seine Bedürfnisse stillt, um die Kraft zur Arbeit zu haben, oder ob umgekehrt er nur arbeitet, um dann auch seine Bedürfnisse stillen zu können“ [Hervorhebungen vom Autor]

Um die Situation in ihrer Vertracktheit leichter begreifbar zu machen, teilt Arendt die Sicht auf den Menschen der Arbeitswelt in einerseits einen Homo faber und ein Animal laborans. Homo faber verfügt über eine gewisse Intelligenz, die ihm die Gestaltung von Automation ermöglicht. Animal laborans ist hingegen nur der „Knecht“ des Homo faber, also eigentlich ein Knecht von sich selbst, dessen Lebensinhalt sich nur auf die Arbeit zum Zwecke der Existenzsicherung reduziert.

Der zuerst im Zitat hervorgehobene Befund der Vermischung von Mittel und Zweck unterstützt die von Kambartel bzw. Buchholz aufgeworfene Debatte über den Unterschied von (erdachtem, eingebildetem) Bedarf und (wahren, vernünftigen) Bedürfnissen. Die zweite Hervorhebung beinhaltet den weiter oben von mir angeführten Gedanken, dass ein zu hoher Grad an Automation schwer zu revidieren ist, und dass deren Auswirkungen einen unmittelbaren – und zwar einschränkenden, ggf. schmerzlich einschränkenden – Einfluss auf unsere Lebensgestaltung haben.

Hannah Arendt konnte 1958 noch nichts von Industrie 4.0 wissen, aber auf einer allgemeineren Ebene gibt sie einen weitsichtigen Ausblick auf die betreffenden Konsequenzen: Im selben Kapitel stellt sie die ersten beiden industriellen Revolutionen vor, um dann das für ihre Argumentation wichtige „Endstadium“ zu schildern. [Ar1961, S. 134]:

„Der Historiker weiß nur zu gut, daß der Sinn geschichtlicher Abläufe meist erst zum Vorschein kommt, wenn sie ihren Abschluß erreicht haben, niemals aber zu erkennen ist, bevor die Entwicklung auf ihren Höhepunkt gekommen ist.“

Diesen „Höhepunkt“ beschreibt sie als „letzte Phase in dieser Entwicklung“ [ebd., S. 135], eine „Automation, die nun ... den Gipfelpunkt dieser Entwicklung“ bilde. Die Verschwimmung von Fabrikationsprozeß und Fabrikat ist dabei integraler Bestandteil. Und das sind ganz klar konstituierende Merkmale von Industrie 4.0. Arendt wagt – nur nebenbei erwähnt – die Vorhersage, dass man in dieser letzte Phase „nur noch Elektrizität“ benötige [ebd., S. 136]. Das ist bemerkenswert, weil mit der Energiewende und der immer weiteren Vollendung der Automation tatsächlich eine gewisse Tendenz zu verzeichnen ist, dass wir immer weniger bereit und in der Lage sind, über den Sinn des Lebens ernsthaft nachzudenken.

Arendt macht die von Kambartel und Buchholz aufgeworfene Sorge um die vernünftige Bewertung wahrer Bedürfnisse durch den Menschen explizit [Ar1961, S. 300]:

„Die große Revolution der Denkungsart, welche die Neuzeit mit sich brachte, hat ... zwar die Fähigkeiten von Homo faber ungeheuer erweitert, hat ihn gelehrt, Apparate herzustellen und Instrumente zu erfinden, mit denen man das unendlich Kleine und das unendlich Große messen und handhaben kann, sie hat ihn aber zugleich der festen Maßstäbe beraubt, die ihrerseits, weil sie jenseits des Herstellungsprozesses selbst liegen, ihm einen echten, aus der Tätigkeit selbst stammenden Zugang zu etwas Absolutem und unbedingt Verlässlichen verschafften. So viel dürfte gewiß sein, daß keine der Tätigkeiten der Vita activa so unmittelbar von dem Verlust der Anschauung und der Kontemplation als eines sinngebenden, menschlichen Vermögens betroffen war wie gerade das Herstellen.“

Arendt wird wenig später noch deutlicher [ebd., S. 300]:

„Daß Homo faber schließlich an den Bedingungen der Neuzeit scheitern sollte, daß es ihm nicht gelingen würde, in der modernen Welt seine Maßstäbe und Wertungen zur Geltung zu bringen, war eigentlich entschieden, als das Nützlichkeitsprinzip, der wirkliche Inbegriff des herstellenden Denkens, von dem Glücksgefühl abgelöst wurde.“

Sehr treffend sagt Arendt voraus, was sich in der Arbeitswelt seit einiger Zeit tatsächlich abzeichnet und durch Industrie 4.0 auf eine neue Spitze getrieben wird [ebd., S. 301]:

„Grundsätzlich ereignet sich dieser radikale Werteverlust [...] selbst bereits, wenn er anfängt, sich primär nicht mehr als den Hersteller von Dingen und den Ersteller eines Gebildes von Menschenhand zu betrachten, sondern als einen Werkzeugmacher und zwar vor allem als einen Fabrizierer von ‚Werkzeugen, die ihrerseits Werkzeuge erzeugen‘, der nur nebenbei auch noch Dinge hervorbringt. Will man in diesem Zusammenhang das Nützlichkeitsprinzip überhaupt noch zur Anwendung bringen, so wird es sich primär nicht auf Gebrauchsgegenstände und ihren unmittelbaren Nutzen beziehen, sondern auf den Produktionsprozeß als solchen. Was immer die Produktivität überhaupt steigert, bzw. die für die notwendige Anstrengung herabsetzt, gilt als nützlich.“

Und in dieser Verlagerung der Aufgabe liegt auch der Verlust an Erfahrung und an Werten, die die Tätigkeit des Denkens einschließt. Auch hierzu konstatiert Arendt [ebd., S. 314]:

„Vergleicht man die moderne Welt mit den Welten, die wir aus der Vergangenheit kennen, so drängt sich vor allem der enorme Erfahrungsschwund auf, der dieser Entwicklung inhärent ist. Nicht nur, daß die anschauende Kontemplation keine Stelle mehr hat in der Weite spezifisch menschlicher und sinnvoller Erfahrungen, auch das Denken, sofern es im Schlußfolgern besteht, ist zu einer Gehirnfunktion degradiert, welche die elektronischen Rechenmaschinen erheblich besser, schneller und reibungsloser vollziehen als das menschliche Gehirn. Das Handeln [im Sinne eines motivegeleiteten, zielgerichteten Wirken des Menschen durch Wort, Tat oder Unterlassung] wiederum, das erst mit dem Herstellen gleichgesetzt wird, sinkt schließlich auf das Niveau des Arbeitens.“

Daraus könnte man ableiten, worin die immense Verantwortung für jene liegt, die an der

Weiterentwicklung von Industrie 4.0 aktiv beteiligt sind. Dass die Forschenden selbst eine Verantwortung für die spätere Verwendung ihrer Ergebnisse tragen, wissen wir in drastischer Weise spätestens seit Entdeckung der nuklearen Kettenreaktion.

Der Erfolg freiheitseinschränkender Politik, die in (für mich) scheinbar unerklärlicher Weise immer mal wieder Mehrheiten findet – selbst in Staaten, denen man dies lange nicht mehr zugetraut hätte – könnte tatsächlich von dem Verlust des Vermögens einiger Menschen herrühren, sich eigenverantwortlich begründete Entscheidungen und Haltungen zu erarbeiten. Dieses setzt eben einige Mühen um glaubwürdige Information und einiges an Vermögen zum wohlüberlegten Denken voraus, die wir uns in unserer mittels Technisierung erworbener Bequemlichkeit nicht mehr zumuten können und/oder wollen.

Wie Arendt es ausdrückt, ist es „erheblich leichter, unter den Bedingungen tyrannischer Herrschaft zu handeln als zu denken.“ [Ar1961, S. 317] Und das hat möglicherweise auch mit dem Grad an Automation in unserer Arbeitswelt zu tun: Wenn ein Staat uns das freie Entscheiden abnimmt, dann ist das für viele Menschen offensichtlich erst einmal bequem – und es ist vielleicht deswegen (wieder) möglich, weil entsprechende Veränderungen in der Arbeitswelt diesen Trend perpetuiert haben.

Literaturverzeichnis

- [Ar1961] Arendt, H.: *Vita Activa oder Vom tätigen Leben*. München: Piper, 1985 [1961].
- [Ba2014] Bauernhansl, Th.: *Die Vierte Industrielle Revolution. Der Weg in ein wertschaffendes Produktionsparadigma*. In (Th. Bauernhansl, M. ten Hompel, B. Vogel-Heuser Hrsg.): *Industrie 4.0 in Produktion, Automatisierung und Logistik*, S. 5-36, Wiesbaden: Springer, 2014.
- [Bi2014] Bildstein, A.: *Industrie 4.0-Readiness: Migration zur Industrie 4.0-Fertigung*. In (Th. Bauernhansl, M. ten Hompel, B. Vogel-Heuser Hrsg.): *Industrie 4.0 in Produktion, Automatisierung und Logistik*, S. 581-598, Wiesbaden: Springer, 2014.
- [Bu2009] Buchholz, K.: *Der mündige Konsument*. In (G. Böhme Hrsg.): *Der mündige Mensch*, Darmstadt: Wissenschaftliche Buchgesellschaft, 2009.
- [Bu2016] Buchholz, K.: *Design und Philosophie*. In (J.-C. Dissel Hrsg.): *Design & Philosophie. Schnittstelle und Wahlverwandschaften*. Bielefeld: transcript 2016.
- [Da2014] Dais, S.: *Industrie 4.0 – Anstoß, Vision, Vorgehen*. In (Th. Bauernhansl, M. ten Hompel, B. Vogel-Heuser Hrsg.): *Industrie 4.0 in Produktion, Automatisierung und Logistik*, S. 625-634, Wiesbaden: Springer, 2014.
- [Kag2014] Kagermann, H.: *Chancen von Industrie 4.0 nutzen*. In (Th. Bauernhansl, M. ten Hompel, B. Vogel-Heuser Hrsg.): *Industrie 4.0 in Produktion, Automatisierung und Logistik*, S. 603-614, Wiesbaden: Springer, 2014.
- [Kam1979] Kambartel, F.: *Ist rationale Ökonomie als empirisch-quantitative Wissenschaft möglich?* In (J. Mittelstraß Hrsg.): *Methodenprobleme der Wissenschaften vom gesellschaftlichen Handeln*. Frankfurt am Main: Suhrkamp, 1979.

- [Kan1923] Kant, E.: Beantwortung der Frage: Was ist Aufklärung? In (ders.): Kant's Werke, Bd. 8, Berlin, Leipzig: de Gruyter, 1923.
- [SGG2013] Spath, D (Hrsg.), Gangschar, O., Gerlach, St., Hämmerle, M., Krause, T., Schlund, S.: Produktionsarbeit der Zukunft – Industrie 4.0, Stuttgart: Fraunhofer Verlag, <https://www.iao.fraunhofer.de/images/iao-news/produktionsarbeit-der-zukunft.pdf>, 2013.
- [PFV2014] Pötter, Th., Volmer, J., Vogel-Heuser, B.: Enabling Industry 4.0 – Chancen und Nutzen für die Prozessindustrie, In (Th. Bauernhansl, M. ten Hompel, B. Vogel-Heuser Hrsg.): Industrie 4.0 in Produktion, Automatisierung und Logistik, S. 159-172, Wiesbaden: Springer, 2014.
- [We2014] Wegener, D.: Industrie 4.0 – Chancen und Herausforderungen für einen Global Player, In (Th. Bauernhansl, M. ten Hompel, B. Vogel-Heuser Hrsg.): Industrie 4.0 in Produktion, Automatisierung und Logistik, S. 343-358, Wiesbaden: Springer, 2014.
- [Zi2001] Zimmermann, D.: Zweckorientierte Automatische Musikkomposition. Musikalische Grundlagen, Geschichte und Modellierung, Wiesbaden: Deutscher Universitätsverlag, 2001.

Chancen und Risiken der Digitalisierung in der Bauindustrie im Kontext von Industrie 4.0 – Situationsanalyse und Zieldefinition im Zuge einer Technikfolgenabschätzung

Thuy Duong Oesterreich¹, Frank Teuteberg²

Abstract: Der Begriff Industrie 4.0 hat sich in den vergangenen Jahren auch in der Bauindustrie zum Synonym für die Automatisierung, Digitalisierung und Vernetzung der Produktionsprozesse entwickelt. Trotz der propagierten Vorteile, die mit der Anwendung dieses Konzeptes verbunden werden, existieren große Vorbehalte bei den Unternehmen der Bauindustrie. Dies liegt darin begründet, dass die mit der Anwendung dieses Konzeptes verbundenen Implikationen für Wirtschaft und Gesellschaft noch weitgehend unbekannt sind. Vor diesem Hintergrund ist es das primäre Ziel dieser Arbeit, im Rahmen einer Situationsanalyse die derzeit mit dem Industrie 4.0 Begriff verbundenen Erwartungen, Chancen und Risiken aus verschiedenen Perspektiven aufzuzeigen. Mithilfe einer quantitativen und qualitativen Inhaltsanalyse der aktuellen Literatur wird zunächst eine fundierte Definition des Industrie 4.0 Begriffs für die Bauindustrie aufgestellt und die Erwartungen, Chancen und Risiken aufgezeigt. Anschließend werden Ziele und konkrete Ansätze für die weitere Vorgehensweise im Rahmen einer Technikfolgenabschätzung abgeleitet.

Keywords: Industrie 4.0, Digitalisierung, Situationsanalyse, Technikfolgenabschätzung, Literaturrecherche, Inhaltsanalyse, Frequenzanalyse, Bauindustrie

1 Problemstellung und Zielsetzung

Das Schlagwort „Industrie 4.0“ wird derzeit im wissenschaftlichen und wirtschaftlichen Kontext als Synonym für die zunehmende Digitalisierung und Automatisierung von Produktionssystemen aus Menschen und Maschinen sowie deren Vernetzung untereinander verwendet [KWH13]. Im Fokus steht dabei die das Potenzial zur Kostensenkung und eine damit einhergehende Effizienzsteigerung. Bei näherer Betrachtung kann jedoch festgehalten werden, dass die mit der Anwendung von Industrie 4.0 Technologien verbundenen Implikationen für Menschen, Prozesse, Unternehmenskultur und Umwelt sowie deren ökonomischer Nutzen für die Unternehmen noch weitgehend unbekannt sind. Insbesondere für die Bauindustrie ist Industrie 4.0 mit heutigem Stand ein inhaltlich unklarer Begriff. Aktuelle Studien zufolge sind bspw. Unklarheit über den wirtschaftlichen Nutzen, fehlende Fachkräfte, fehlende Normen und Standards sowie mangelnde Datensicherheit

¹ Universität Osnabrück, Fachgebiet Unternehmensrechnung und Wirtschaftsinformatik, Katharinenstraße 1 49074 Osnabrück, toesterreich@uni-osnabrueck.de

² Universität Osnabrück, Fachgebiet Unternehmensrechnung und Wirtschaftsinformatik, Katharinenstraße 1 49074 Osnabrück, frank.teuteberg@uni-osnabrueck.de

wesentliche Gründe, weshalb die Unternehmen bei Investitionen in Industrie 4.0 Technologien immer noch zurückhaltend sind [Li15].

Auf dem Weg zur Beseitigung der Hemmnisse ist somit der erste sinnvolle Schritt der, die mit dem Industrie 4.0 Konzept verbundenen Technologien, aber auch Erwartungshaltungen, Chancen und Risiken als mögliche Folgeauswirkung im Sinne einer Technikfolgenabschätzung (TA) aufzuzeigen. Vor diesem Hintergrund stellen sich folgende Kernfragen, die im Rahmen einer Situationsanalyse zu beantworten sind:

1. Welche Technologien und Konzepte werden aktuell im spezifischen Umfeld der Bauindustrie mit dem Begriff Industrie 4.0 assoziiert?
2. Welche Erwartungshaltungen, Chancen und Risiken werden in der Bauindustrie mit dem Industrie 4.0 Begriff verbunden?
3. Kann auf Basis der Situationsanalyse die Notwendigkeit einer TA bestätigt werden? Welche Fragen und Ziele sind für die weitere Vorgehensweise relevant?

Im Abschnitt 2 gehen wir zunächst auf die Besonderheiten der Bauindustrie ein und skizzieren deren Stand der Digitalisierung. Anschließend stellen wir in Abschnitt 3 unsere methodische Vorgehensweise zur Beantwortung der Forschungsfragen vor. Auf Basis der Ergebnisse unserer qualitativen und quantitativen Inhaltsanalyse wird im Abschnitt 4 eine Definition des Industrie 4.0 Begriffs unter Berücksichtigung der Besonderheiten der Bauindustrie abgeleitet und die damit verbundenen Erwartungshaltungen, Chancen und Risiken aufgezeigt. Im Abschnitt 5 werden weiterführende Fragen, die mögliche Zielsetzung und konkrete Handlungsempfehlungen für die weitere TA erörtert.

2 Digitalisierung in der Bauindustrie

Einer Studie aus dem Jahr 2014 zufolge befindet sich die Bauindustrie im Vergleich mit anderen Wirtschaftszweigen auf einem der hinteren Plätzen [Ac14]. Diese Tatsache ist angesichts ihrer volkswirtschaftlichen Bedeutung als einer der Schlüsselbranchen sehr verwunderlich. Die Ursachen für diesen niedrigen Digitalisierungsgrad sind den Besonderheiten der Baubranche geschuldet und werden auch als Grund für die abnehmende Arbeitsproduktivität in den letzten Jahrzehnten gesehen, während diese sich in den anderen Industriesektoren verdoppelt hat [Te13].

Geringe Investitionen in Forschung und Entwicklung: Ein Zahlenbeispiel verdeutlicht diese Schieflage: Im Jahre 2012 wurde 37,3% der gesamten F&E-Ausgaben in der Automobilbranche investiert, aber nur 0,1% entfielen auf die Baubranche [BM15].

Komplexität und Unsicherheit: Bauprojekte gehören aufgrund ihrer äußerst differenzierten Wertschöpfungskette zu den komplexesten Wirtschaftsgegenständen. Sie sind Unikate, die in stark arbeitsteiligen „wandernden Fabriken“ in einem ortsgebundenen, spezifischen Umfeld und in einem begrenzten Zeitrahmen hergestellt werden [BR09], [DG02]. Die Tatsache, dass es sich bei Bauprojekten um stark individualisierte Einzelfertigung

handelt und sie in hohem Maße von Witterungsflüssen abhängig sind, macht zudem eine genaue Vorhersage des Produktionsverlaufs nahezu unmöglich [DG02].

Fragmentierte Wertschöpfungskette: In die stark arbeitsteiligen Prozesse und Subprozesse ist eine hohe Anzahl an kleinen und mittelständischen Unternehmen (KMU) eingebunden, die nur begrenzt in neue Technologien investieren können [KHS14].

Fehlende Innovationskraft und konservative Branchenkultur: Die Zeitbegrenzung der Bauprojekt auf wenige Monate bis zu wenigen Jahren mit jeweils wechselnden Projektbeteiligten führt dazu, dass kurzfristige Problemlösung begünstigt und nachhaltige Innovationen erschwert werden [DG02]. Des Weiteren ist die Bauindustrie für ihre traditionelle Kultur und ihre geringe Bereitschaft für Veränderungen bekannt [SK15].

3 Methodik

Für die Untersuchung der mit dem Industrie 4.0 Konzept verbundenen Erwartungen, Chancen und Risiken eignet sich insbesondere die Methode der Technikfolgenabschätzung (TA), weshalb wir sie als methodischen Rahmen auswählen (Abb. 1). Der Vorteil dieser Methode liegt darin begründet, dass die Analyse der Auswirkungen einer neuen Technologie über die rein technische Betrachtungsweise hinausgeht und auch ökonomische, ökologische, rechtliche, soziale und ethische Fragestellungen mit einbezieht.

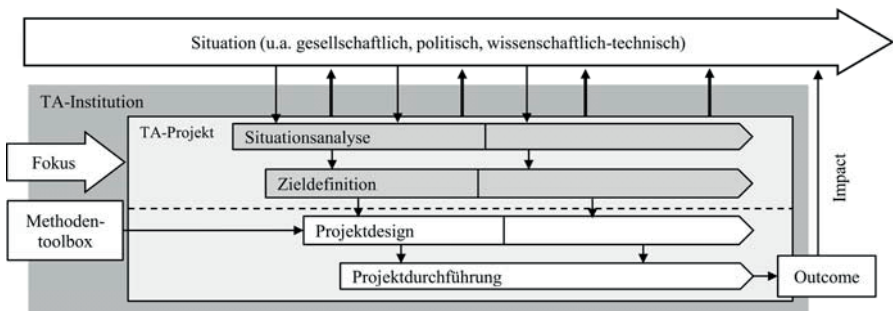


Abb. 1: Struktur einer TA (Angelehnt an [13])

Den zentralen Ausgangspunkt einer TA bildet die Situationsanalyse, in der die öffentliche Resonanz bzgl. der Technikdebatte analysiert und die Zieldefinition für das TA-Projekt festgelegt wird [DS15]. Der Fokus dieser Arbeit richtet sich auf die ersten beiden Schritte der TA und soll die Grundlage für die weitergehenden Phasen schaffen (Abb. 1).

3.1 Qualitative und quantitative Textanalyse

Zur Beantwortung der Forschungsfragen ist die Anwendung einer Kombination aus qua-

litativen und quantitativen Methoden erforderlich. Um eine fundierte Definition des Industrie 4.0 Begriffs herleiten zu können, wird zunächst im Rahmen einer Literaturrecherche nach Publikationen aus dem bauwirtschaftlichen Umfeld gesucht. Dabei orientieren wir uns methodisch an dem Vorgehensschema von [WW02]. Aufgrund der Tatsache, dass es sich bei Industrie 4.0 im Baubereich um ein relativ junges Forschungsgebiet handelt, existieren kaum wissenschaftliche Publikationen zu diesem Thema. Aus diesem Grund werden bei der Literaturrecherche auch nicht wissenschaftliche Publikationen wie Zeitungen, Magazine, Produktbroschüren, Positionspapiere, so genannte „graue Literatur“ berücksichtigt. Dieser Auswahlrahmen ist auch deshalb sinnvoll, weil es bei einer Situationsanalyse vor allem darum geht, die öffentliche Wahrnehmung zu einer technischen Neuentwicklung zu untersuchen [DS15].

Unsere breit gestaltete Literaturrecherche bei der Literaturdatenbank Google Scholar und der Suchmaschine Google mit den deutschen Suchwörtern „Industrie 4.0“ in Kombination mit Bauindustrie, Bauwirtschaft oder Baubranche sowie den englischen Suchwörtern „Industry 4.0“ in Kombination mit „Construction Industry“ ergeben rund 40 relevante Publikationen, die wir im Rahmen der Analyse vertiefend untersuchen. Auf eine weitere Ergebniseingrenzung (z.B. nach Datum, Suche im Titel, etc.) haben wir bewusst verzichtet, um möglichst viele Ergebnisse zu erhalten. Wesentlich für die Auswahl der Publikationen ist der klare inhaltliche Bezug zum Thema Industrie 4.0. Um die Aussagekraft der Ergebnisse zu maximieren, wenden wir bei der Ergebnisanalyse sowohl die qualitative Inhaltsanalyse [Ma10] als auch die einfache Frequenzanalyse [Be52] an. Bei der Datenaufbereitung für die Frequenzanalyse folgen wir einem bewährten Ablaufschema [Si08] und verwenden des Weiteren zur technischen Unterstützung die Software QDA Miner [Pr11] sowie das integrierte Textanalysemodul WordStat [Pr14]. Die Ergebnisse dieser Untersuchung werden im Abschnitt 4 präsentiert und im Abschnitt 5 diskutiert.

3.2 Verwandte Arbeiten

Im Rahmen einer Literaturrecherche bei der Suchmaschine Google mit den Suchworten <intitle:Technikfolgenabschätzung "Industrie 4.0"> können wir lediglich 2 verwandte Arbeiten identifizieren. Nähere Details zu den verwandten Arbeiten können auf http://bit.ly/Info2016_Anhang eingesehen werden (Anhang 1). Es handelt sich hierbei um Publikationen zum Thema TA von Industrie 4.0 ohne spezifischen Branchenbezug. Aufgrund der im Abschnitt 2 genannten Spezifika der Bauindustrie ist es jedoch erforderlich, eine TA von Industrie 4.0 mit Bezug zur Bauindustrie zu durchleuchten, um auf diese Weise den speziellen Branchenanforderungen gerecht zu werden.

4 Situationsanalyse

In der Bauindustrie ist Industrie 4.0 inzwischen ein zwar geläufiger, aber inhaltlich unklarer Begriff. Um zu eruieren, welche Technologien und Teilkonzepte darunter subsumiert werden, untersuchen wir im nächsten Schritt die selektierten Publikationen mithilfe einer

quantitativen und qualitativen Inhaltsanalyse (Abb. 2). Bei diesen Publikationen handelt es sich größtenteils um mediale Beiträge wie Zeitungsartikel, Zeitschriftenartikel, Magazinartikel und Blog-Posts, wissenschaftliche Beiträge sind dagegen in der Unterzahl. Eine Eingruppierung der untersuchten Publikationen nach der Perspektive der Veröffentlichung bringt die interessante Erkenntnis hervor, dass das Thema Industrie 4.0 mit 67,5% derzeit eine große mediale Aufmerksamkeit erfährt, in der Wirtschaft, Politik und in der Wissenschaft dagegen ist die Resonanz geringer (Abb. 4).

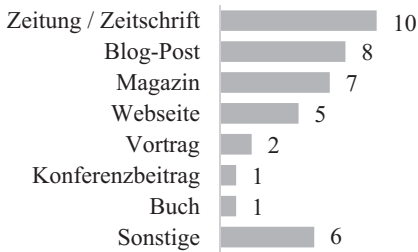


Abb. 2: Art der Publikationen

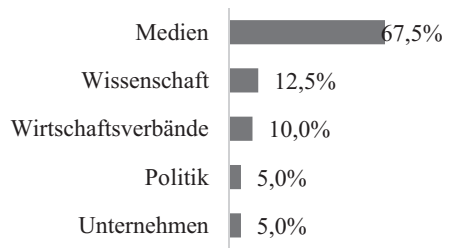


Abb. 3: Initiator der Publikationen

Diese Erkenntnis unterstützt die allgemein bekannte These, dass das Thema Industrie 4.0 in den Unternehmen der Bauindustrie noch nicht angekommen ist und auch im wissenschaftlichen Bereich in den Kinderschuhen steckt. Eine detaillierte Übersicht aller untersuchten Publikationen können dem Anhang 2 auf http://bit.ly/Info2016_Anhang entnommen werden.

4.1 Begriffsdefinition Industrie 4.0 in der Bauindustrie

Das Ergebnis unserer Textanalyse in Tab. 2 soll im nächsten Schritt Informationen darüber liefern, welche Basistechnologien und Komponenten mit dem Industrie 4.0 Begriff in Verbindung gebracht werden. Dabei gilt es zu untersuchen, in wie vielen Publikationen welche Ausprägungen genannt werden und inwiefern sich die bauspezifische Definition des Industrie 4.0 Begriffs von der „klassischen“ Definition unterscheidet (Tab. 3).

Zusammenfassend lässt sich feststellen, dass die in den bauspezifischen Publikationen und in den Top 10 meist zitierten Industrie 4.0 Publikationen genannten Technologien und Konzepte sich weitgehend decken. Die Resultate zeigen zudem, dass Building Information Modeling (BIM) als das zentrale Element des Industrie 4.0 Konzeptes in der Bauindustrie betrachtet wird. Bei BIM handelt es sich um eine Methode zur Abbildung von digitalen, parametrisierten Planungsmodellen von Bauwerken inkl. der physikalischen und funktionalen Eigenschaften eines Bauwerks und anderen relevanten Informationen wie z.B. Kosten und Zeit [Eg13]. Da BIM alle Phasen innerhalb des Produktlebenszyklus von der Planung über die Bauausführung und Wartung bis hin zur Demontage eines Bauwerkes abdeckt, ist auch die Forderung nach einem „Product Lifecycle Management“ zur nachhaltigen Produktgestaltung erfüllt. Das Konzept der Simulation und Modellierung als Me-

thode zur Beherrschung komplexer Zusammenhänge und zur Unterstützung der interdisziplinären Zusammenarbeit [KWH13] deckt BIM damit ab. In einigen Publikationen wird BIM als „eine Art Industrie 4.0 für die Bauwirtschaft“ [Ma15] bezeichnet, das die Baubranche revolutionieren kann [Bi15a]. Es gibt allerdings auch kritische Stimmen, die vor überzogenen Erwartungen an die BIM-Methode warnen, da diese keine vollständige Digitalisierung und Automatisierung der Prozesse ermöglicht, sondern nur die Komplexität innerhalb der Wertschöpfungskette reduziert [SK15].

Industrie 4.0 Komponenten und Basistechnologien	Anzahl Publikationen	Industrie 4.0 Komponenten und Basistechnologien	Anzahl Publikationen
Building Information Modelling (BIM)	28	Cyber-Physical Systems (CPS) / Eingebettete Systeme	10
Internet der Dinge (IoT) / Internet der Dienste (IoS)	12	Internet der Dinge (IoT) / Internet der Dienste (IoS)	8
Product-Lifecycle-Management (PLM)	11	Sensortechnik/ RFID	8
		Autonome Robotics	6
Cloud Computing	10	Big Data	6
Mobile Computing	10	Cloud Computing	6
Augmented Reality (AR)/Virtual Reality (VR)/ Mixed Reality (MR)	9	Simulation und Modellierung	6
		Smart Factory	6
Autonome Robotics	9	Mensch-Technik-Interaktion (MTI)	5
Sensortechnik/ RFID	8	Modularisierung	5
Big Data	7	Product-Lifecycle-Management (PLM)	5
Additive Manufacturing	5		
Smart Factory	4	Augmented Reality (AR)/Virtual Reality (VR)/ Mixed Reality (MR)	4
Mensch-Technik-Interaktion (MTI)	3	Additive Manufacturing	3
Modularisierung	3	Mobile Computing	3
Cyber-Physical Systems (CPS) / Eingebettete Systeme	2		
		Social Media	3

Tab. 2: Übersicht analysierte Publikationen (n=40)

Tab. 3: Übersicht Top 10 meist zitierte Industrie 4.0 Publikationen (n=10)

Vielmehr muss die Anwendung von BIM mit der Digitalisierung und Vernetzung relevanter Datenflüsse im Unternehmen kombiniert werden [Sc16]. Erst durch die Nutzung von Basistechnologien wie Cloud Computing und Big Data in Verbindung mit mobilen Endgeräten [Bo14], [Sc16] kann BIM dazu beitragen, eine enge Vernetzung aller am Bauprozess beteiligten Akteure zu realisieren [Do16]. Als Technologie dient die Cloud aber auch als Plattform für andere Apps und Anwendungen oder als Speicherplatz zur Archivierung digitaler Baustellendokumente [In15]. Die Nutzung mobiler Endgeräte ermöglicht dabei den zeit- und ortsunabhängigen Zugriff auf BIM-Modelle, ERP-System und digitale Baustellendokumente als Basis für eine interdisziplinäre Zusammenarbeit.

Eine weitere Komponente des Industrie 4.0 Konzeptes sind das Internet der Dinge (IoT) und das Internet der Dienste (IoS), die sowohl in den Top 10 meist zitierten Industrie 4.0 Publikationen als auch in den untersuchten bauspezifischen Publikationen die zweithäufigste Aufmerksamkeit findet (Tab. 2 und Tab. 3). Mit diesen Begriffen wird die intelligente Vernetzung von Gegenständen, Maschinen und Personen verstanden [Do16], die in

Echtzeit miteinander kommunizieren und in einen automatisierten Datenaustausch treten [Sc16]. In dieser vernetzten Umgebung besteht die große Herausforderung darin, die wachsenden Datenmengen mittels Big Data sinnvoll zusammenzuführen und zielgerichtet zu nutzen [He16], wie bspw. für Predictive Maintenance. Als weiteres wichtiges Hilfsmittel zur Verbesserung der Arbeitssicherheit und der Kommunikation werden Augmented Reality, Virtual Reality und Mixed Reality-Konzepte zur Schaffung von mobilen Assistenzsystemen vorgeschlagen, z.B. in Kombination mit mobilen Endgeräten oder Datenbrillen [Bi15a], [Si14], [Sc16]. Auch der Einsatz von autonomen Robotern und Drohnen zur Ausführung gefährlicher oder repetitiver Arbeiten wird genannt [Ri16]. Die Nutzung von RFID-Technik zur Lokalisierung und Vernetzung von Objekten, der 3D-Druck und die Modularisierung als Teil des Fertigungskonzeptes, oder auch die Smart Factory werden mehrfach genannt.

Auffällig ist, dass Cyber-Physical-Systems (CPS) als zentrales Element des klassischen Industrie 4.0 Konzeptes wenig Beachtung in den Publikationen aus dem Baubereich findet (Tab. 2 und Tab. 3). Social Media wird als weiterer Punkt in Tab. 3 genannt, in den untersuchten Publikationen findet es keine Beachtung. Viel Beachtung findet dagegen in beiden Publikationsgruppen das Thema Mensch-Technik-Interaktion (MTI), welches im Zuge der technologischen Weiterentwicklungen für die Gestaltung der Mensch-Technik-Schnittstelle von zentraler Bedeutung ist. Mit Blick auf die genannten Technologien und Konzepte kann zusätzlich festgehalten werden, dass es sich bei den meisten von ihnen um technisch ausgereifte Technologien handelt, die jederzeit einsetzbar sind (bspw. Cloud Computing, BIM, RFID, Augmented Reality, etc.).

Aufgrund der branchenspezifischen Besonderheiten müssen zwischen der Anwendung des Industrie 4.0 Konzeptes in der stationären Industrie und der Anwendung in der Bauindustrie einige Unterschiede beachtet werden. Bei Bauprojekten handelt es sich bspw. um einen arbeitsteiligen Prozess, an dem der Kunde maßgeblich beteiligt ist, während ein Automobilhersteller dem Kunden lediglich ein fertiges Produkt liefert, welches in einem strukturierten Produktionsprozess hergestellt wird [Ma15]. Die Anwendung von Industrie 4.0 muss also an die Komplexität des jeweiligen Produktionssystems angepasst werden. Die folgenden Charakteristika von Industrie 4.0 [KWH13] lassen sich dennoch auf die Anwendung in der Bauindustrie beziehen (Abb. 5).

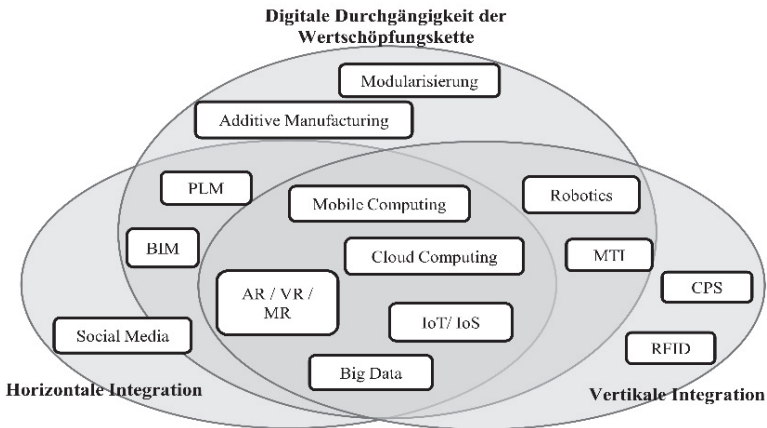


Abb. 5: Übersicht Konzeptmatrix Industrie 4.0

Um ein umfassendes Bild der bauspezifischen Definition von Industrie 4.0 zu skizzieren, haben wir die identifizierten Technologien und Konzepte in die genannten Charakteristika eingeordnet. Da viele Technologien sich zu mehreren Charakteristika zuordnen lassen, gibt es einige Schnittmengen, die sich in Abb. 5 entsprechend widerspiegeln.

Mit der *horizontalen Integration über Wertschöpfungsnetzwerke* wird eine Integration von IT-Systemen, Prozessen und Informationsflüssen zwischen den Geschäftspartnern über die Unternehmensgrenzen hinweg angestrebt. Auf die Bauindustrie bezogen wäre dies die enge Zusammenarbeit von allen Projektbeteiligten wie Bauherr, Lieferanten, Nachunternehmer etc. auf einer gemeinsamen Datenbasis. Die Schaffung einer *digitalen Wertschöpfungskette* setzt eine digitale Durchgängigkeit aller Prozessschritte voraus, bei der die Modellierung zur Beherrschung der zunehmenden Komplexität als Hilfsmittel genannt wird. In Bauprojekten müssen somit alle Prozessschritte von der Planung über die Kalkulation bis hin zur Bauausführung, Inbetriebnahme und Wartung digital vernetzt werden. Die *vertikale Integration und Vernetzung der Produktionssysteme* zielt auf eine flexible, selbststeuernde und selbstoptimierende Gestaltung der Produktionssysteme und die Integration der unternehmensinternen IT-Systemen, Prozessen und Informationsflüssen ab. Aufgrund der Komplexität der bauprojektbezogenen Produktion und der hohen Anzahl der Projektbeteiligten ist dies nur zum Teil auf die Bauindustrie übertragbar.

4.2 Erwartungen, Chancen und Risiken

Nach der Begriffsdefinition ist es ein weiteres Ziel, die mit Industrie 4.0 verbundenen Erwartungshaltungen, Chancen und Risiken aufzugreifen. Aufgrund der Tatsache, dass 67,5% aller Publikationen medialen Ursprungs sind (Abb. 4), spiegeln die Ergebnisse hauptsächlich die öffentliche Debatte um die Industrie 4.0 Thematik wieder. Zur Kategorisierung der Ausprägungen eignet sich die PESTEL-Analyse [KN08], da sich damit die verschiedenen Dimensionen der Situationsanalyse strukturiert abbilden lassen. Die Ergeb-

nisse aus den untersuchten Publikationen werden in Tab. 4 daher in 6 Dimensionen (politische, ökonomische, soziale, technologische, ökologische und rechtliche Dimension) zusammengefasst. In den Spalten E, C, R und H werden die Ausprägungen als Erwartungen (E), Chancen (C), Risiken (R) und Herausforderungen (H) kategorisiert.

Beschreibung		E	C	R	H
P	<ul style="list-style-type: none"> ▪ Wenig Forschung und Entwicklung im Baubereich und geringe Bereitschaft zur Anwendung von Industrie 4.0 Technologien ▪ Politik in der Rolle des Motivators und Initiators, staatliche Impulse zur Beseitigung der Adoptionsbarrieren gefordert 	x			x
	<ul style="list-style-type: none"> ▪ Komplexe Planungs- und Prozessabläufe, veraltete Arbeitsweisen ▪ Reduzierung der Komplexität zur Optimierung der Bauprozesse ▪ Erhöhung der Transparenz für eine bessere Entscheidungsgrundlage ▪ Fehlerreduzierung, Erhöhung der Planungs- und Ausführungsqualität ▪ Steigerung der Effizienz und Wirtschaftlichkeit (Niedrigere Produktionskosten, Schnelligkeit) ▪ Erhöhung der Kosten- und Termsicherheit 		x x x x		x
E	<ul style="list-style-type: none"> ▪ Verbesserung der interdisziplinären und unternehmensübergreifenden Zusammenarbeit (Kollaboration, Kommunikation) ▪ Kundenindividualisierung, Verbesserung der Kundenbeziehung ▪ Schaffung neuer innovativer Geschäftsmodelle ▪ Erhöhung der Wettbewerbsfähigkeit ▪ Notwendigkeit zur Änderung interner Prozesse ▪ Hohe Investitionen bei unklaren Nutzen 		x x x	x x	
	<ul style="list-style-type: none"> ▪ Geringe Akzeptanz bzw. Bereitschaft für Veränderungen ▪ Vertrauen als Basis für die Zusammenarbeit ▪ Bewältigung des demographischen Wandels ▪ Erhöhung der Arbeitssicherheit ▪ Steigerung der Arbeitgeberattraktivität ▪ Verbesserung des Branchenimages ▪ Neue Arbeitsanforderungen bei Fachkräftemangel und fehlender Qualifikation des Personals ▪ Angst vor Jobverlust (Ersatz Mensch durch Technik) ▪ Erhöhung der psychischen Arbeitsbelastung (Stress, Leistungsdruck, höhere Anforderungen) 		x x x x	x x	x x
T	<ul style="list-style-type: none"> ▪ Ausbau der Breitbandinfrastruktur ▪ Fehlende Kompatibilität und Standardisierung ▪ Datensicherheit und Datenschutz (Angst vor Datenmissbrauch) ▪ Wachsende Datenmengen und sinnvolle Datennutzung 	x		x x	x
	<ul style="list-style-type: none"> ▪ Verbesserung von Energieeffizienz und Nachhaltigkeit ▪ CO²-Reduktion und Energieeinsparung (z.B. durch weniger Papierverbrauch und durch energieeffizientes Bauen) 		x x		

L	▪ Unzuverlässige rechtliche Rahmenbedingungen (z.B. strikte Vorgaben bzgl. Datenschutz und Datenaustausch)				x
	▪ Fehlende Regularien, Vorschriften, Normen (BIM, VOB, HOAI, etc.)				x
	▪ Unklare Vertragsausgestaltung bzgl. Urheberrechte, Haftung, Risikovertelung. Beispiel: Wer ist Urheber eines BIM-Modells? Wer haftet für planungsbedingte Konstruktionsfehler?				x

Tab. 4: PESTEL-Analyse der Erwartungen, Chancen, Risiken und Herausforderungen

Politische Dimension (P): Die Themen Industrie 4.0, BIM und Digitalisierung in der Bauindustrie sind derzeit Gegenstand der politischen Diskussion. Seit einigen Jahren werden von der Bundesregierung Anstrengungen unternommen, die Methode BIM bis 2020 verpflichtend für alle öffentlichen Infrastrukturprojekte einzuführen [BV15]. In den untersuchten Beiträgen ist die Grundhaltung daher klar erkennbar: Der Staat befindet sich in der Rolle des Motivators, der unter Nennung der wirtschaftlichen Vorteile die Unternehmen zur Umsetzung des Industrie 4.0 Konzeptes motiviert [Do16]. Im Gegenzug ist auch die Erwartungshaltung der Wirtschaft und Gesellschaft an die Politik deutlich. In den Beiträgen wird oft die geringe Forschung und Entwicklung bemängelt und staatliche Impulse in Form von Initiativen und Förderung gefordert [Pö14].

Ökonomische Dimension (E): Mit dem Industrie 4.0 Begriff werden in der öffentlichen Diskussion viele wirtschaftliche Nutzenerwartungen verbunden. Dazu zählen u.a. die Erhöhung der Effizienz [Si14], [SK15], der Transparenz [Pö14], [Bo14] und der Kosten- und Terminalsicherheit [Sc15]. Des Weiteren ist die Verbesserung der Kundenbeziehung [He16] sowie der interdisziplinären und unternehmensübergreifenden Zusammenarbeit ein weiterer oft genannter Punkt. Die Nutzung neuer Technologien eröffnet zudem Möglichkeiten zur Schaffung neuer Geschäftsmodelle [He16], [Sc16]. Mit der Nutzenrealisierung wird auch eine höhere Wettbewerbsfähigkeit erwartet, die sich auf den internationalen Wettbewerb auswirkt. Die hohen Investitionen bei unklaren Nutzenerwartungen und die Notwendigkeit zur internen Prozessänderung werden im Gegenzug als Risiken genannt, die sich derzeit als Adoptionsbarriere erweist.

Soziale Dimension (S): Wie bereits im Abschnitt 2 dargelegt, ist die Wertschöpfungskette Bau geprägt von stark arbeitsteiligen Prozessen und einer großen Anzahl der beteiligten Akteure [BR09], [DG02]. Im Mittelpunkt jedes Bauvorhabens steht somit der Faktor Mensch und deren Interaktion untereinander. Für eine erfolgreiche Zusammenarbeit wird daher als Grundvoraussetzung „gegenseitiges Vertrauen, Fairness und Offenheit zwischen den Projektbeteiligten“ genannt [Sc15] sowie eine „Kultur der Kooperation und nicht des Gegeneinanders“ gefordert [Bi15a]. Als große Herausforderung wird in den untersuchten Beiträgen die fehlende Bereitschaft für Veränderungen bzw. fehlende Akzeptanz gegenüber neuen Technologien genannt [Bi15a], [SK15]. Die sozialen Chancen von Industrie 4.0 werden u.a. in einer Verbesserung der Arbeitsbedingungen wie z.B. Arbeitssicherheit, der Verbesserung des Branchenimages und eine damit einhergehende Steigerung der Attraktivität für Fachkräfte [Ri16], [Sc15] gesehen. Als soziale Risiken eines hohen Digitalisierungsgrades werden zum einen die steigenden fachlichen Anforderungen an das Personal und eine damit verbundene Erhöhung der psychischen Arbeitsbelastung (Stress,

Leistungsdruck, höhere Anforderungen) sowie die allgemeine Angst vor Jobverlust (Ersatz Mensch durch Technik) genannt [Sc16].

Technologische Dimension (T): Grundvoraussetzung für eine durchgängige Digitalisierung und Vernetzung der Wertschöpfungskette Bau ist der Ausbau der Breitbandinfrastruktur [Sc16]. Große Risiken sind ebenfalls die fehlende Kompatibilität und Standardisierung der Systeme (z.B. BIM) [Pö14], [Sc15], [SK15] sowie das Thema Datensicherheit und Datenschutz [Ri16], [Mü15]. Vor dem Hintergrund der wachsenden Datenmengen ist zudem eine sinnvolle Datennutzung von wesentlicher Bedeutung [Mü15], [Sc15].

Ökologische Dimension (E): Mit der Anwendung von BIM sowie der durchgängigen Digitalisierung der Prozesse werden die Chancen eines energieeffizienten Bauens sowie eine CO²-Reduktion und Energieeinsparung verbunden [Si14], [Pö14], [Sc15]. Zum einen ermöglicht die umfassende Lebenszyklusbetrachtung (PLM) eines Bauwerkes eine energieoptimierte und ressourcenschonende Wiederverwendung der ursprünglich eingesetzten Materialien auch nach der Inbetriebnahme und Demontage [Sc15]. Zum anderen sind digitale Prozesse nachhaltiger, da sie zu weniger Papierverbrauch führen.

Rechtliche Dimension (L): Auf rechtlicher Ebene werden ebenfalls erhebliche Risiken gesehen, die zum einen die unzuverlässigen rechtlichen Rahmenbedingungen betreffen, wie z.B. strikte Vorgaben bzgl. Datenschutz und Datenaustausch auf nationaler Ebene [Mü15], fehlende Regularien, Vorschriften und Normen [Ri16], [Sc15] sowie unbeantwortete Fragen in Bezug auf die Vertragsausgestaltung [Ma15], [Mü15].

4.3 Ergebnisse der quantitativen Textanalyse

Ergänzend zu den qualitativen Ergebnissen möchten wir mithilfe des quantitativen textanalytischen Ansatzes eruieren, ob die aufgezeigten Technologien, Erwartungshaltungen, Chancen und Risiken sich bestätigen lassen. Zu diesem Zweck ermitteln wir mithilfe einer Frequenzanalyse die häufigsten Wortstämme (Tab. 5) sowie die häufigsten zusammenhängenden 2- und 3-Wort-Phrasen (Tab. 6).

#	Wortstamm	Häufig-
1	BIM	472
2	PLAN	287
3	INDUSTRI	251
4	TECHNI	199
5	DIGITAL	166
6	CLOUD	159
7	SYSTEM	158
8	DATEN	153
9	MODEL	147
10	UNTERNEHMEN	141
11	PROZESS	139

#	Phrasen mit 2 und 3 Wörtern	Häufig-
1	INDUSTRIE 4.0	94
2	BUILDING INFORMATION	51
3	INDUSTRY 4.0	40
4	CLOUD COMPUTING	25
5	BAUEN 4.0	14
6	WERTSCHÖPFUNGSKETTE BAU	14
7	BIG DATA	13
8	FÜR KMU	13
9	INTERNET DER DINGE	12
10	ÖFFENTLICHE HAND	11
11	EINFÜHRUNG VON BIM	10

12	ROBOT	138
13	ENTWICK	134
14	DIGITALISIER	131
15	ARBEIT	129
16	BAUEN	127
17	PROJEKT	113
18	INFORM	110
19	TECHNOLOGI	109
20	BETEILIGT	106

Tab. 5: Top 20 Worthäufigkeiten

12	GESAMTEN LEBENSZYKLUS	9
13	FÜR ALLE BETEILIGTEN	9
14	FÜR WIRTSCHAFT	9
15	INTERNET OF THINGS	8
16	SMART FACTORY	8
17	BIG BIM	8
18	FÜR DEN BETRIEB	8
19	AUSFÜHRUNG UND BEWIRTSCHAFTUNG	7
20	DIGITALES BAUEN	7

Tab. 6: Top 20 Phrasen mit 2 und 3 Wörtern

Die Ergebnisse bestätigen die hohe Bedeutung der BIM-Methode, die in vielen Wortstämmen (BIM, PLAN, MODEL) und Phrasen (BUILDING INFORMATION MODELING, EINFÜHRUNG VON BIM, BIG BIM) wiederzufinden sind. Die Tatsache, dass die Häufigkeiten der Wortstämme und Phrasen INDUSTRI, INDUSTRIE 4.0 und INDUSTRY 4.0 ebenfalls die Ranglisten anführen ist nicht weiter verwunderlich, da es sich um das Hauptthema der untersuchten Publikationen handelt. Bei den Basistechnologien spiegelt sich die Relevanz von Cloud Computing, Big Data, PLM, Internet der Dinge, Smart Factory und Robotertechnik wieder. Auch die Bedeutung der Zusammenarbeit innerhalb der Wertschöpfungskette Bau werden in den Begriffen ARBEIT, PROJEKT, INFORM, PROZESS, BETEILIGT, WERTSCHÖPFUNGSKETTE BAU und FÜR ALLE BETEILIGTEN zum Ausdruck gebracht. Die politische Dimension lässt sich in den Phrasen ÖFFENTLICHE HAND, FÜR KMU, FÜR WIRTSCHAFT, FÜR DEN BETRIEB erkennen. Andere Wortstämme wie TECHNI, DIGITAL, SYSTEM, DATEN, ENTWICK, UNTERNEHMEN und TECHNOLOGI sind weitere Eigenschaften, die mit dem Industrie 4.0 in Verbindung gebracht werden. Dahinter verbergen sich Begriffe wie digitale Modelle, Datenschutz, Datensicherheit oder Digitalisierung.

5 Diskussion und Ausblick

Die Situationsanalyse auf Basis der abgeleiteten Begriffsdefinition sowie der damit verbundenen Erwartungshaltungen, Chancen und Risiken zeigt auf, dass die Bauindustrie sich auf dem Weg zur Umsetzung des Industrie 4.0 Konzeptes noch ganz am Anfang befindet. Grundsätzlich werden damit große Potenziale verbunden, die weit über die wirtschaftlichen Vorteile hinausgehen (Tab. 4). Gleichwohl existieren diverse Hemmnisse, die teilweise den Besonderheiten der Baubranche geschuldet sind (Abschnitt 2), aber auch in der „klassischen“ Industrie 4.0 Literatur genannt werden. Angesichts der vielen ungeklärten Fragen kann die Notwendigkeit einer weiterführenden TA eindeutig bejaht werden. Eine TA kann dazu beitragen, die mit dem Industrie 4.0 Begriff verbundenen Folgewirkungen zu plausibilisieren und greifbarer zu machen. Da die Bauindustrie im Vergleich mit anderen Industriezweigen zu den Nachzüglern gehört und damit erheblichen Nachholbedarf hat [Ac14], sollte es bei der TA aber nicht um die grundsätzliche Frage gehen, ob die Einführung von Industrie 4.0 für die Unternehmen sinnvoll ist, sondern eher um die Frage, wie ein gangbarer Weg zur Anwendung aussehen kann.

5.1 Offene Fragen und Handlungsempfehlungen

Mit Synonymen wie Bauen 4.0, Building 4.0 oder Handwerk 4.0 wird das Thema Industrie 4.0 in der Bauindustrie derzeit medienwirksam diskutiert und propagiert. Bei näherer Betrachtung lässt sich jedoch feststellen, dass viele unbeantwortete Fragen bleiben, wie bspw. die nach praxisnahen Anwendungsszenarien, neuen Geschäftsmodellen, Methoden zur Kosten-Nutzen-Abwägung etc. Auf dem Weg zu einer vertiefenden TA sind daher folgende weiterführenden Fragen zu beantworten:

1. Wie können mögliche Anwendungsszenarien in der Bauindustrie aussehen? Dabei können unterschiedliche Digitalisierungsgrade angenommen werden, um die Folgewirkungen in verschiedenen Abstufungen umfassend beurteilen zu können.
2. Welche strukturellen Unterschiede müssen bei den potenziellen Anwendern berücksichtigt werden? Diese Frage ist insbesondere für die stark fragmentierte Baubranche wichtig, da es sich bei den am Bauprozess beteiligten Unternehmen sowohl um große Baukonzerne als auch um kleine und mittelständische Unternehmen handelt.
3. Was sind praktikable Methoden zur Kosten-Nutzen-Quantifizierung? Dies ist insbesondere wichtig, um die im Rahmen der Situationsanalyse aufgezeigten wirtschaftlichen Chancen und Risiken objektiv beurteilen zu können.
4. Welche konkreten Geschäftsmodelle können aus der Nutzung von Industrie 4.0 Technologien abgeleitet werden? Um die Chancen aus der Entstehung neuer innovativer Geschäftsmodelle beurteilen zu können, muss im ersten Schritt zunächst einmal feststehen, um welche Geschäftsmodelle es sich handeln kann.

Basierend auf die Zieldefinition, die Folgewirkungen einer Industrie 4.0 Anwendung in der Bauindustrie und einen gangbaren Weg zur Umsetzung aufzuzeigen, ist der nächste Schritt im Rahmen einer TA die Festlegung des Projektdesigns (Abb. 4).

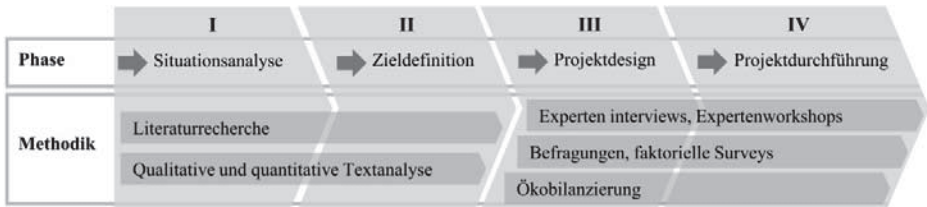


Abb. 4: Phasen und geeignete Methoden der TA

Methodisch eignet sich hierfür eine Kombination aus quantitativen und qualitativen Ansätzen sowie eine klar strukturierte Vorgehensweise [DS15]. Zur Betrachtung politischer, ökonomischer, technologischer und rechtlicher Aspekte der Fragestellung eignet sich eine expertenorientierte Vorgehensweise auf Basis von Experteninterviews und Expertenworkshops, um mittels einer Aggregation und Validierung von Expertenwissen zu fundierten Antworten zu gelangen [DS15]. Bei den sozialen Aspekten der Fragestellungen

wie bspw. Akzeptanz, Angst vor Jobverlust etc. eignen sich Befragungen wie z.B. faktorielle Surveys bzw. Vignettenanalysen, da sie eine Simulation von komplexen Beurteilungs- und Entscheidungsproblemen sozialwissenschaftlicher Natur erlauben [HAL09]. Bei ökologischen Aspekten eignet sich als ergänzende Methode die Ökobilanzierung [DS15], da sie eine genaue Quantifizierung der Umwelteffekte ermöglicht.

5.2 Limitationen und Ausblick

Die Ergebnisse unserer Untersuchung basieren auf einer beschränkten Anzahl an Publikationen, die größtenteils nicht wissenschaftlicher Natur sind. Es handelt sich hierbei um mediale Publikationen, die keinem Review-Verfahren durchlaufen sind und daher nicht dem Anspruch einer wissenschaftlichen Fundierung genügen. Dies haben wir in unserem Beitrag jedoch akzeptiert, da die aktuelle Forschung in diesem Bereich sich noch in den Anfängen befindet und somit kaum wissenschaftliche Literatur hierzu existiert. Im weiteren Schritt gilt es, diese Forschungslücke weiter zu schließen und die aufgezeigte Vorgehensweise zur vertiefenden TA aufzugreifen.

Literaturverzeichnis

Eine Übersicht aller 40 Artikel der Inhaltsanalyse sind auf http://bit.ly/Info2016_Anhang abrufbar.

- [Ac14] Accenture: Neue Geschäfte, neue Wettbewerber. Die Top500 vor der digitalen Herausforderung, 2014.
- [Be52] Berelson, B.: Content Analysis in Communication Research. Free Press, 1952.
- [BM15] BMBF: Bundesbericht Forschung und Innovation 2014. <http://www.datenportal.bmbf.de/portal/de/Tabelle-1.5.1.html>, 2015-11-24.
- [BR09] Butzin, A.; Rehfeld, D.: Innovationsbiographien in der Bauwirtschaft, Stuttgart, 2009.
- [BV15] BMVI: Building Information Modeling (BIM) wird bis 2020 stufenweise eingeführt.
- [DG02] Dubois, A.; Gadde, L.-E.: The construction industry as a loosely coupled system: implications for productivity and innovation. In *Construction Management and Economics*, 2002, 20; S. 621–631.
- [DS15] Decker, M.; Schippl, J.: Technikfolgenabschätzung. In (Kaltschmitt, M.; Schebek, L. Hrsg.): *Umweltbewertung für Ingenieure*. Springer Berlin Heidelberg, 2015; S. 403–437.
- [Eg13] Egger, M. et al.: *BIM-Leitfaden für Deutschland*, 2013.
- [HAL09] Hinz, T.; Auspurg, K.; Liebig, S.: Komplexität von Vignetten, Lerneffekte und Plausibilität im Faktoriellen Survey. In *MDA – Methoden, Daten, Analysen*, 2009, 3.
- [KHS14] Kraatz, J. A.; Hampson, K. D.; Sanchez, A. X.: *The global construction industry and R&D: R&D Investment and Impact in the global construction industry*. Routledge, 2014; S. 4–23.
- [KN08] Kaplan, R. S.; Norton, D. P.: *The Execution Premium: Linking Strategy to Operations*

for Competitive Advantage. Harvard Business Review Press, Boston, Mass, 2008.

- [KWH13] Kagermann, H.; Wahlster, W.; Helbig, J.: Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0, Frankfurt/Main, 2013.
- [Li15] Lichtblau, K. et al.: Industrie 4.0-Readiness, 2015.
- [Ma10] Mayring, P.: Qualitative Inhaltsanalyse. In (Mey, G.; Mruck, K. Hrsg.): Handbuch Qualitative Forschung in der Psychologie. VS Verlag für Sozialwissenschaften, 2010; S. 601–613.
- [Pr11] Provalis Research: QDA Miner 4.0. <http://provalisresearch.com/Documents/QDAMiner40.pdf>, 2016-05-14.
- [Pr14] Provalis Research: WordStat 7. <http://provalisresearch.com/Documents/WordStat7.pdf>, 2016-05-14.
- [Si08] Sidorova, A. et al.: Uncovering the Intellectual Core of the Information Systems Discipline. In MIS Quarterly, 2008, 32; S. 467–482.
- [Te13] Teicholz, P. M.: Labor-Productivity Declines in the Construction Industry: Causes and Remedies (a second look). In AECbytes, 2013, 1.
- [WW02] Webster, J.; Watson, R.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. In Management Information Systems Quarterly, 2002, 26.

Industrie 4.0-Readiness: Eine empirische Fallstudie in der Automobilproduktion

Kathrin Schwinghammer¹, Ricardo Buettner²

Zusammenfassung: Einer systematische tiefeninterviewgeführte Fallstudie eines DAX30 Automobilproduzenten zeigt interessante Ergebnisse zum Verständnis der Industrie 4.0 Herausforderungen und der Bereitschaft, diese umzusetzen.

Keywords: Industrie 4.0 Readiness, Smart Factory, Digitalisierung, Leadership 4.0, HR 4.0.

1 Einleitung

Aufgrund der Neuartigkeit und der nationalen Prägung des Begriffes „Industrie 4.0“ (vgl. Call Workshop „Industrie 4.0 – Wünschenswertes, Machbares und Grenzen für Technik, Wirtschaft und Mensch“ der INFORMATIK 2016 Tagung), liegen aus der Industrie kaum empirische Ergebnisse zum Begriffsverständnis und zur Umsetzungsbereitschaft vor.

Der vorliegende Beitrag liefert deshalb entsprechende Ergebnisse aus einer tiefeninterview-analysierten Fallstudie eines „Old-Economy“ Automobil-OEMs. Diese Ergebnisse sind für Wissenschaftler und Praktiker relevant. Insbesondere liefert die Arbeit erste Antworten auf relevante Leitfragen zur Entscheidung des Übergangs zu Industrie 4.0-Systemen im betrieblichen Kontext, zur Bereitschaft eines Betriebs zur Einführung von Industrie 4.0 und allgemeine und spezifische technische Grenzen.

2 Methodik

Gemäß der Empfehlungen von Benbasat et al. [BGM87] sowie Dube & Pare [DP03] wurden 27 Personen eines DAX30 Automobil-OEMs, die aus einer systematischen Stakeholderanalyse ermittelt wurden, befragt. Die Personen stammen aus den Unternehmensbereichen Produktion, Entwicklung, Strategie und Betriebsrat; von Zulieferern im Bereich Automatisierungstechnik; von der Gewerkschaft IG Metall sowie aus externer Forschung und Unternehmensberatung. Alle Tiefeninterviews wurden aufgezeichnet und gemäß den Empfehlungen von Myers & Newman [MN07] sowie Mayring [PM10] transkribiert und

¹ FOM Hochschule, Institute of Management & Information Systems (mis), Hopfenstraße 4, 80335 München, kathrinschwinghammer@me.com

² FOM Hochschule, Institute of Management & Information Systems (mis), Hopfenstraße 4, 80335 München, ricardo.buettner@fom.de

kategorisiert. In der Leitfragenerstellung, der Interviewdurchführung sowie der Datenauswertung wurden die Empfehlungen von Nederhof [N85] zur weitgehenden Vermeidung von Verzerrungen berücksichtigt.

3 Auszug aus den Ergebnissen

3.1 Einschätzungen zur Industrie 4.0-Umsetzungsbereitschaft in der OEM-Produktion

Nach Einschätzung der Interviewpartner werden selbstorganisierende, intelligente Produktionssysteme [B10] die kontinuierlich getriebene Hochautomatisierung in der Fertigung in seiner Struktur verändern. Die Mitarbeiter, insbesondere diejenigen, die in der Fertigung im Bereich Materialbestückung, Intralogistik, Instandhaltung und Serienplanung tätig sind, stehen vor einem grundlegenden Wandel. Der Mensch als Problemlöser und die Maschine als produktiver Akteur bzw. als Assistenzsystem ist nach Einschätzung der Befragten das wahrscheinlichste Szenario für die nahe Zukunft. Sensoren und Aktoren an und in den Robotern brauchen jedoch nach wie vor die menschliche Problemlösefähigkeit und ganzheitliche Betrachtungskompetenz. Diese Veränderungen bringen Chancen, bergen aber auch Herausforderungen. Ein Wechsel zu mehr Partizipation der Beschäftigten durch neue Führungsmethoden ist genauso notwendig, wie die Entlastung der Beschäftigten bei körperlich schweren und hoch repetitiven Tätigkeiten. Auf der anderen Seite sind noch viele Fragen zu den Themen Datenschutz, Sicherheit und zur Ethik offen, die dringend und umfassend beantwortet werden sollen. Dabei ist es notwendig, klar zwischen der Automatisierung im Sinne der „Industrie 3.0“ [B06a, B06b] und der Vernetzung der Systeme im Sinn von „Industrie 4.0“ [B10, B15] abzugrenzen.

3.2 Einschätzungen zu Veränderungen für den Menschen

Die Befragten denken übereinstimmend, dass der technologische und der organisatorische Wandel dazu führen, dass sich auch der Mensch als Teil des Arbeitssystems zukünftig in einer veränderten Position wiederfinden wird. Neue Stellenprofile und Organisationsformen [B15] werden sich herausbilden und das Verständnis und das Wissen von Informationstechnologien werden sich ändern. Der Mitarbeiter der Zukunft wird in einem einzigen umfassenden cyber-physischen Netzwerk interagieren [[B15]. Der stetige und permanente Druck, sich fort- und weiterzubilden, um diese neuen Systeme zu beherrschen, wird zunehmen. Die Arbeit der Mitarbeiter in einer Industrie-4.0-Umgebung wird sich nach Einschätzung der Befragten in vielen Punkten ändern, dazu zählen:

1. Veränderung der Aufgabeninhalte (Integration neuer IT-, Multimedia-, Cloud-Technologien, Assistenzsysteme),
2. Virtualisierung von Arbeitsschritten/Arbeitsprozessen (beständiger Wechsel virtueller und realer Schreibtische),

3. Verstärkte, umfassende Flexibilisierung von Arbeits- und Lerninhalten und der begleitenden Rahmenbedingungen (Arbeitszeit).

Hierauf muss nach Wertung der Interviewpartner reagiert werden. Folgende Empfehlungen haben sich auf Basis der Tiefeninterviews herauskristallisiert:

1. Anpassung der Weiterbildungswege und -möglichkeiten auf die Lerntypen und das Alter der Mitarbeiter,
2. Stärkung der Transparenz und der Kommunikation bis zum Shop-Floor über Perspektiven und Möglichkeiten für die einzelnen Mitarbeiter und deren Mehrwert in der zukünftigen Produktion (intrinsische Motivation und unternehmerisches Denken fördern),
3. Umsetzung eines Führungsleitbildes, welches dem Industrie 4.0-Begriffsverständnis hinsichtlich Innovation und Vernetzung gerecht wird und fehlertolerant die Entfaltung intrinsischer Mitarbeitermotivation ermöglicht,
4. Abbau des Silodenkens und stärkere abteilungsübergreifende Vernetzung als wesentlicher Schritt zu dauerhafter bereichs- und unternehmensübergreifenden Zusammenarbeit, um den Industrie 4.0 Vernetzungsgedanken zu ermöglichen,
5. Aufschließen der anderen Unternehmensbereiche an die Taktfrequenz und die IT-Systemausstattung der Produktion.

Die Befragten sehen hohe Potenziale für die Fertigung durch die Industrie 4.0 in der Instandhaltung. Dazu zählen beispielsweise:

- nahtlose Kommunikation zwischen allen Teilen der Produktion inkl. Personal,
- Transparenz der Maschinenzustände durch miniaturistische kabellose Sensoren,
- einfache Austauschbarkeit und Wandlungsfähigkeit der Produktionsanlagen,
- Überwachung durch übergreifende Datenerfassung und -auswertung,
- vorhersagbares Systemverhalten durch hohe Sensordichte und Simulation.

4 Zusammenfassung

Die Ergebnisse aus einer systematische tiefeninterviewgeführte Fallstudie eines DAX30 Automobilproduzenten zeigen interessante Ergebnisse zum Verständnis der Industrie 4.0 Herausforderungen und der Bereitschaft, diese umzusetzen.

Das Verständnis des Begriffes „Industrie 4.0“ bezieht sich dabei im Wesentlichen auf zwei Ebenen: Makroperspektivisch erfolgt die komplexe Vernetzung autonomer Systeme und menschlicher Akteure mittels intelligenter Koordinationsmechanismen [B06a, B06b, B10, B15]. Mikroperspektivisch verbessern die autonomen Systeme ihr eigenes intelligentes Verhalten vor allem durch die Nutzung neuer Data-Mining Methoden bei der systematischen Auswertung großer Datenmengen von Sensoren, Maschinen und Nutzern [B14, B16a, B16b]. Schließlich erlauben die zukünftigen neuroadaptiven Mensch-Maschine In-

teraktionssysteme [B13] die Herstellung des Mikro-Makro Links. Diese drei Ansatzpunkte könnten in zukünftigen Studien systematisch genutzt werden, um ein einheitliches Begriffsverständnis herzustellen.

Trotz der Einhaltung der wesentlichen methodischen Richtlinien [BGM87, DP03, MN07, N85, PM10] und einer diversen breiten Auswahl von Interviewpartnern sind die Ergebnisse nicht einschränkungslos auf die Allgemeinheit übertragbar. Sie bieten dennoch wertvolle Einsichten in die Bereitschaft zur Industrie 4.0-Umsetzung in produzierenden Großkonzernen.

Literaturverzeichnis

- [BGM87] Benbasat, I.; Goldstein, D.K.; Mead, M.: The Case Research Strategy in Studies of Information Systems. *MIS Quarterly* 11(3), S. 369-386, 1987.
- [B06a] Buettner, R.: A Classification Structure for Automated Negotiations. In *WI-IAT Proceedings*, S. 523-530, 2006.
- [B06b] Buettner, R.: The State of the Art in Automated Negotiation Models of the Behavior and Information Perspective. *ITSSA* 1(4):351-356, 2006.
- [B10] Buettner, R.: Automatisierte Verhandlungen in Multi-Agenten-Systemen: Entwurf eines argumentationsbasierten Mechanismus für nur imperfekt beschreibbare Verhandlungsgegenstände, Gabler-Verlag, 2010.
- [B13] Buettner, R.: Cognitive Workload of Humans Using Artificial Intelligence Systems: Towards Objective Measurement Applying Eye-Tracking Technology. In *KI 2013 Proceedings*, S. 37-48, 2013.
- [B14] Buettner, R.: A Framework for Recommender Systems in Online Social Network Recruiting: An Interdisciplinary Call to Arms. In *HICSS Proc.*, S. 1415-1424, 2014.
- [B15] Buettner, R.: A Systematic Literature Review of Crowdsourcing Research from a Human Resource Management Perspective. In *HICSS-48 Proc.*, S. 4609-4618, 2016.
- [B16a] Buettner, R.: Personality as a predictor of business social media usage: An empirical investigation of XING usage patterns. In *PACIS Proceedings, Taiwan, 2016*, In Press.
- [B16b] Buettner, R.: Innovative Personality-based Digital Services. In *PACIS Proceedings, Taiwan, 2016*, In Press.
- [DP03] Dube, L.; Pare, G.: Rigor in Information Systems Positivist Case Research: Current Practices, Trends, and Recommendations. *MIS Quarterly* 27(4), S. 597-635, 2003.
- [MN07] Myers, M.D.; Newman, M.: The qualitative interview in IS research: Examining the craft. *Information and Organization* 17(1), S. 2-26, 2007.
- [N85] Nederhof, A.J.: Methods of coping with social desirability bias: A review. *European Journal of Social Psychology* 15(3), S. 263-280, 1985.
- [PM10] Mayring, P.: *Qualitative Inhaltsanalyse*. Beltz Verlag, 2010. 11. Auflage.

5 Anhang



Abbildung 1: Fragestellungen der Interviews

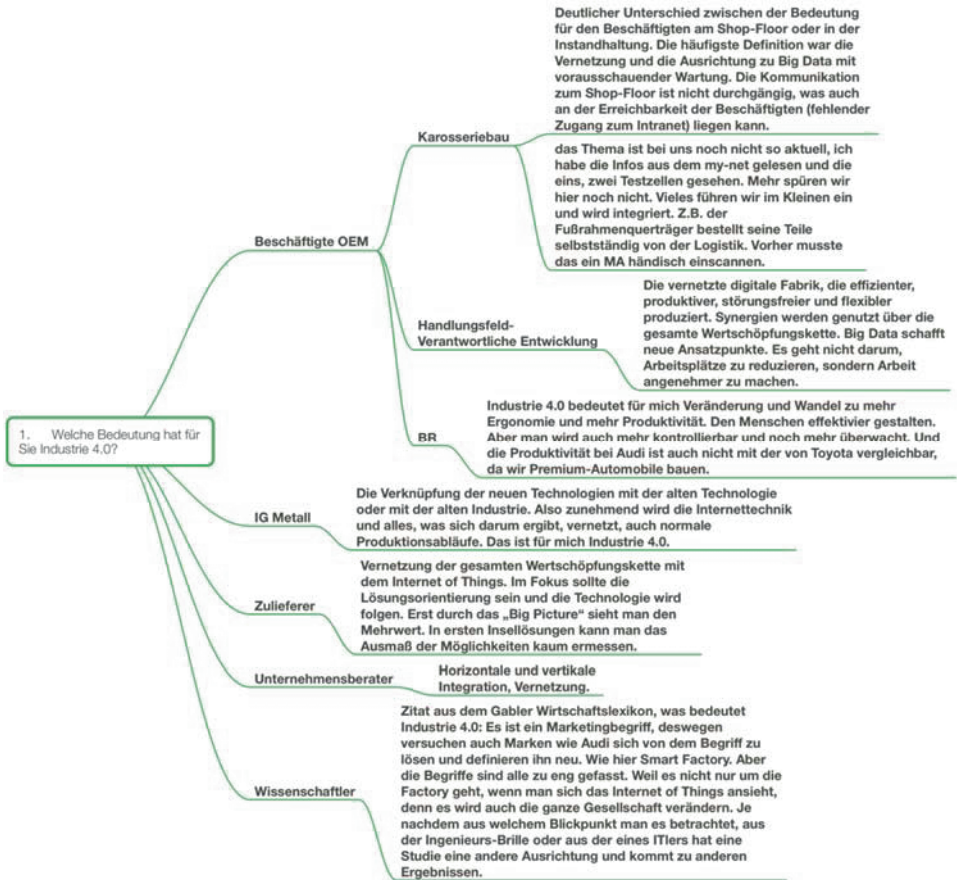


Abbildung 2: Bedeutung und Definition von Industrie 4.0

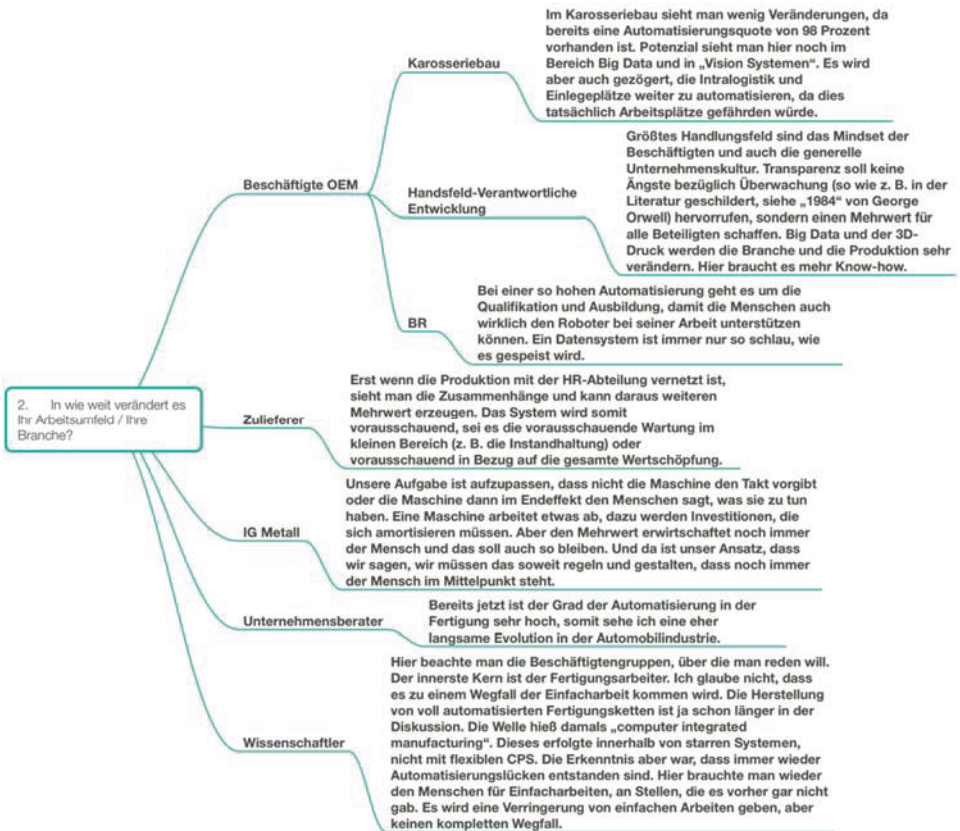


Abbildung 3: Veränderungen des Arbeitsumfeldes / der Branche durch Industrie 4.0

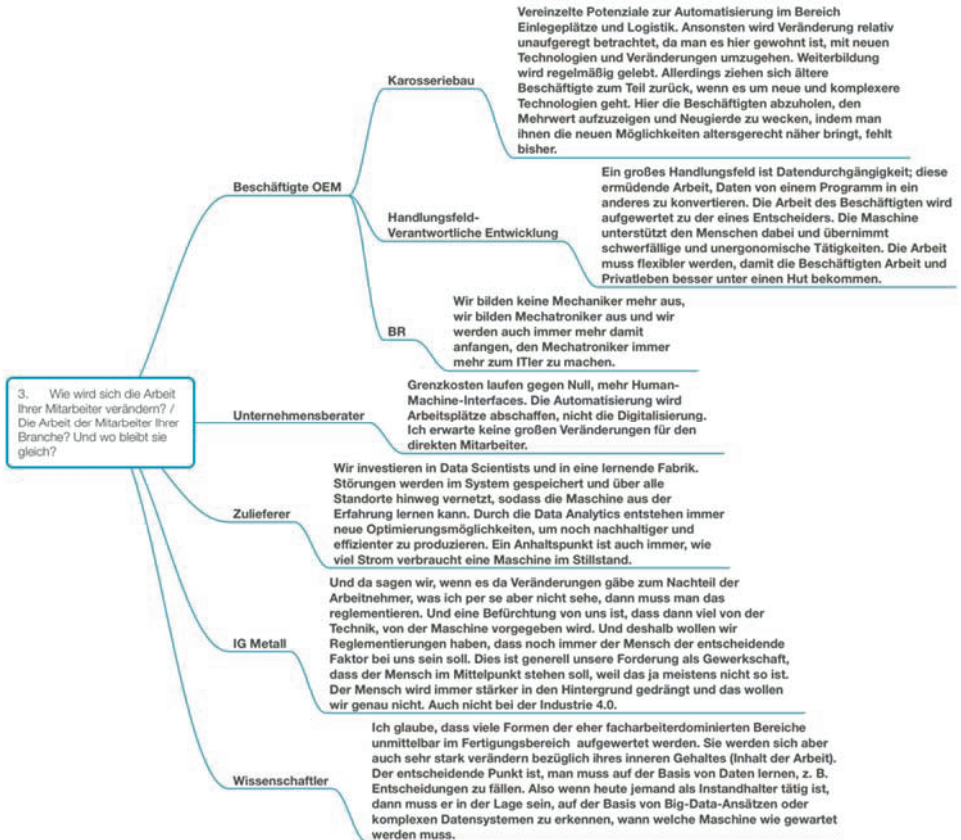


Abbildung 4: Veränderung der Tätigkeit des Mitarbeiters

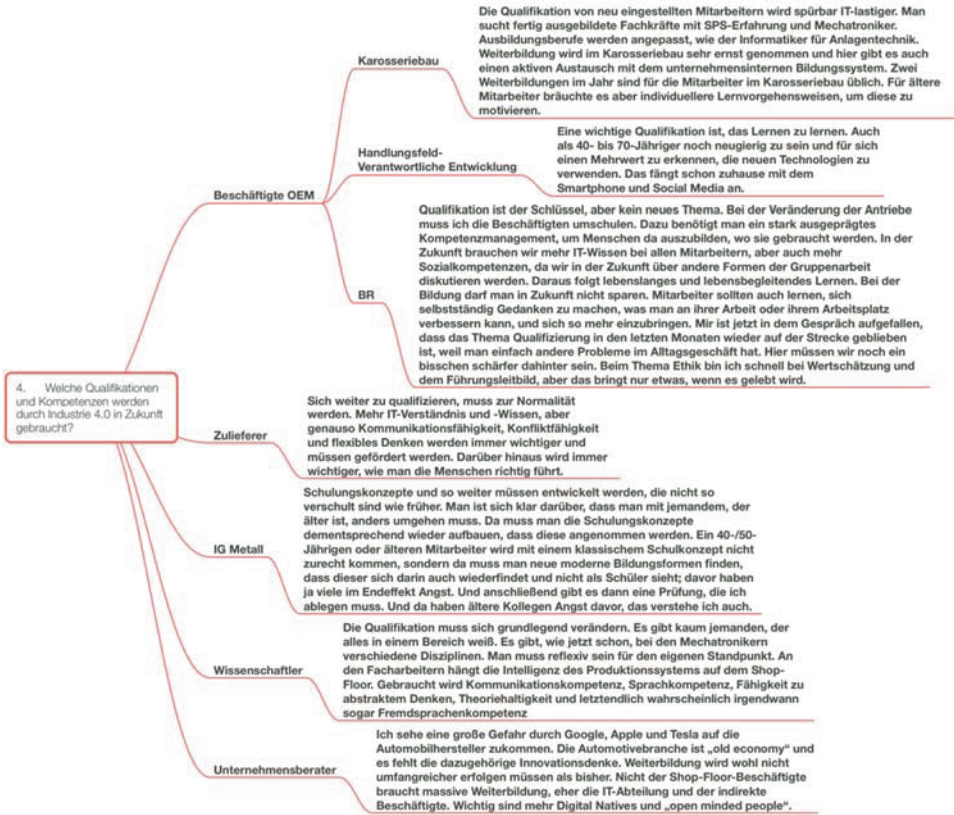


Abbildung 5: Anspruch an die Qualifikation und Kompetenz des Mitarbeiters durch Industrie 4.0

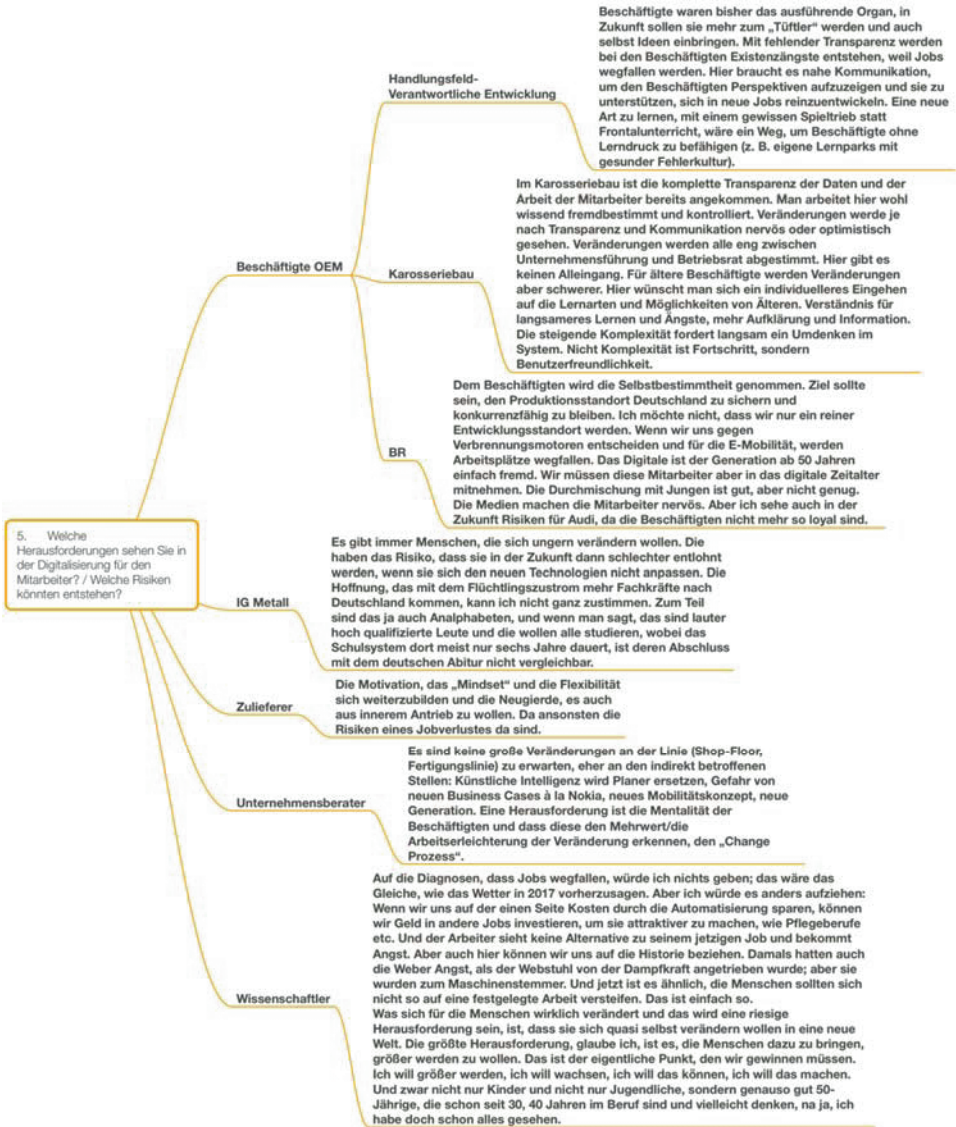


Abbildung 6: Herausforderungen und Risiken für den Mitarbeiter durch Industrie 4.0

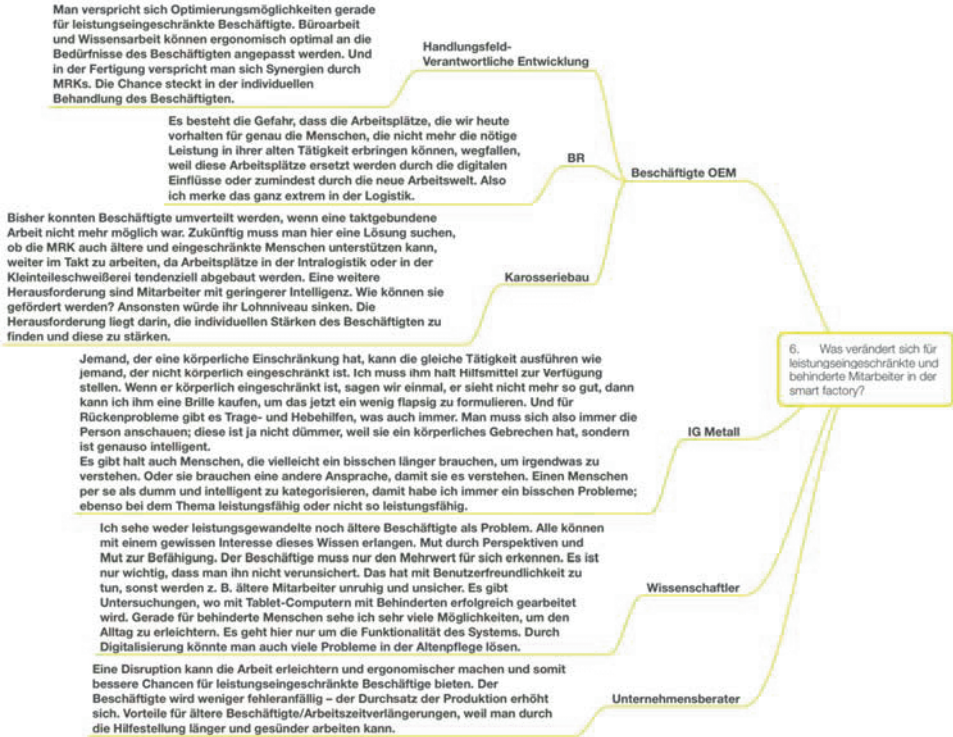


Abbildung 7: Veränderungen für leistungseingeschränkte Mitarbeiter

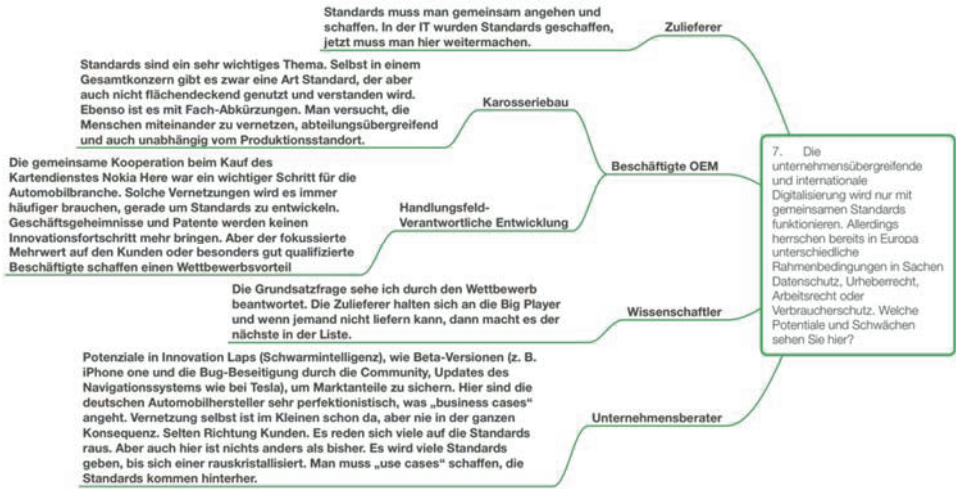


Abbildung 8: Rechtliche Rahmenbedingungen und Standards

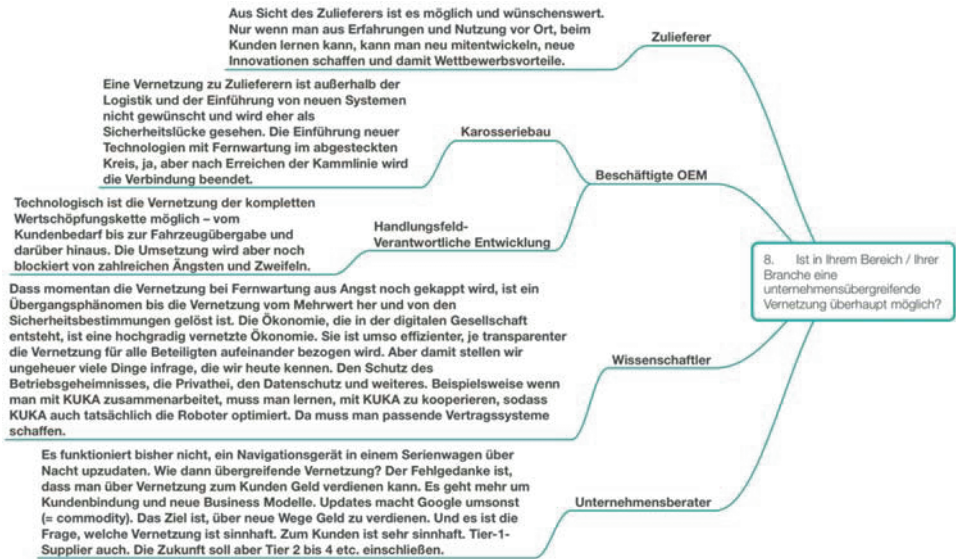


Abbildung 9: Vernetzung über Unternehmens- und Branchengrenzen hinweg

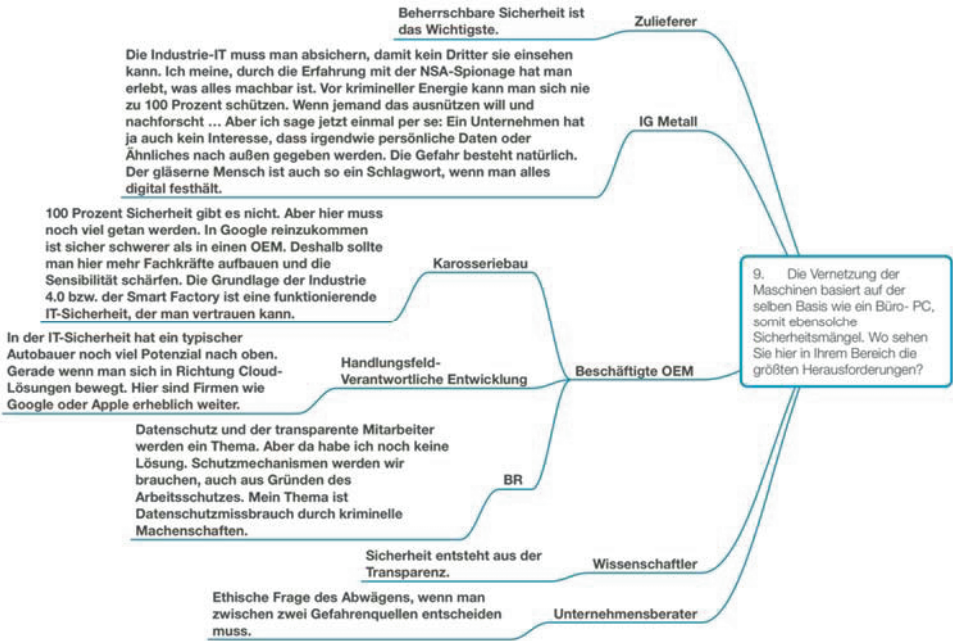


Abbildung 10: Cybercrime und Cybersecurity

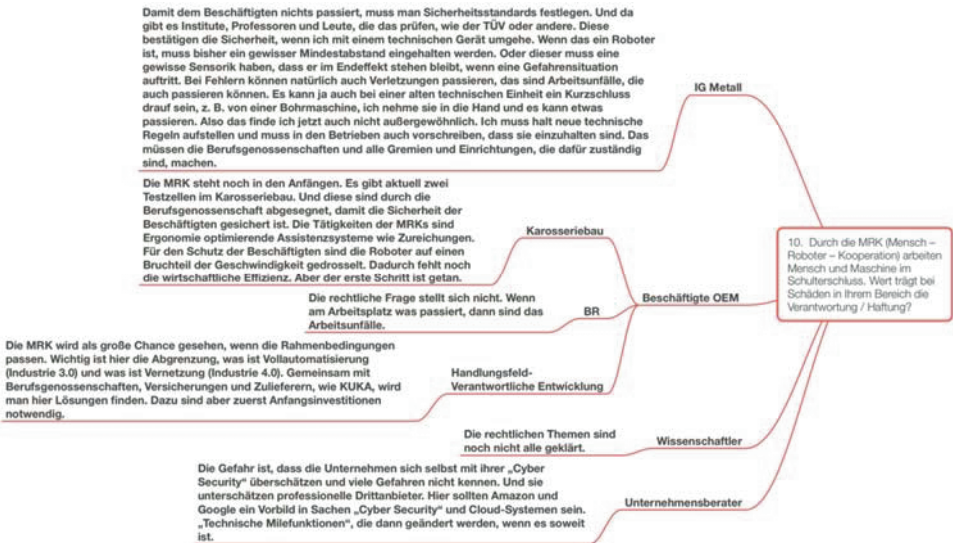


Abbildung 10: Die Haftungsfrage bei Mensch-Roboter-Kollaboration Lösungen

Industrie 4.0 – Datenhoheit und Datenschutz

Univers.-Prof. Dr. Dr. Jürgen Ensthaler¹ und Dr. Martin S. Haase²

Abstract: Daten werden im Zusammenhang mit vernetzten „intelligenten“ Maschinen, mit Gerätschaften, die vom gewerblich tätigen oder privaten Konsumenten benutzt werden und Informationen über Nutzungsart, Funktionsfähigkeit oder Störungen liefern, zu einem Wirtschaftsfaktor werden. Die europäische Kommission ist in ihrer Strategie für einen digitalen Binnenmarkt der Auffassung, dass Daten zum eigenen Produktionsfaktor und Wirtschaftsgut werden.³ Mit dieser Ansicht verbunden ist dann die Frage nach der „Datenhoheit“ – wem gehören die Daten? Darüber hinaus kann die Erhebung und Verarbeitung von Daten, die Aussagen über einen einzelnen Menschen beinhalten, dessen Persönlichkeitsrecht berühren. Soweit Industrie 4.0 zu neuen „Arten“ von Datenverarbeitungsprozessen führt, wird der (personenbezogene) Datenschutz vor Herausforderungen gestellt, die im Ansatz bereits teilweise diskutiert wurden und teilweise neu bedacht werden müssen. Außerdem sind in die Anwendung und Auslegung des Datenschutzrechts die neuen technischen Entwicklungen einzubeziehen.

Keywords: Datenhoheit, Datenschutz, personenbezogene Daten, datenschutzrechtliche Zulässigkeit.

1 Berechtigung an Daten – „Datenhoheit“

1.1 Welche Daten sind gemeint?

Unter Daten werden in dem ersten Abschnitt dieses Beitrags Informationen verstanden, die nicht oder nicht ausreichend derart bearbeitet wurden, dass sie den Schutzbereich von Immaterialgüterrechten, wie insbesondere das Patent- oder Urheberrecht erreichen. Umgangssprachlich formuliert handelt es sich bei den Daten um virtuelle Rohmaterialien, die noch keiner oder keiner für einen bislang anerkannten Schutzbereich hinreichenden Bearbeitung unterzogen wurden.

Solche Informationen sind dem immaterialgüterrechtlichen Sprachgebrauch folgend gemeinfrei, d.h., sie können von jedermann benutzt werden. Den Informationen fehlt eine Leistung erfinderischer, origineller und auch leistungsschutzrechtlicher Art, durch die der

¹ TU Berlin, Fakultät VII, Wirtschafts-, Unternehmens- und Technikrecht, Sekr. H- 41, Straße des 17. Juni 135, 10623 Berlin, juergen.ensthaler@tu-berlin.de.

² TU Berlin, Fakultät VII, Zivil-, Handels-, Gesellschafts- und Innovationsrecht, Sekr. H- 41, Straße des 17. Juni 135, 10623 Berlin, m.haase@tu-berlin.de.

³ Commission Staff Working Document, A Digital Single Market Strategy for Europe – Analysis and Evidence, SWD (2015) 100 final, 59; nach einem Bericht in der Zeitschrift FOCUS vom 30.04.2016, S. 106, wurden 90 % der heutigen Daten in den vergangenen zwei Jahren generiert. Weltweit soll danach das Datenvolumen um 50 % pro Jahr steigen.

Schutz derart eingebundener Daten verdient werden kann. Auch das Datenbankrecht des § 87a UrhG, ein Leistungsschutzrecht, wurde bislang für die Schutzgewährung eher abgelehnt. Dies soll im Folgenden näher untersucht werden; der Europäische Gerichtshof (EuGH) hat jedenfalls in mittlerweile zwei Entscheidungen einen Leistungsschutz für erst zu generierende Daten abgelehnt; geschützt werden sollen nur Investitionen in bereits vorhandene Daten.⁴ [Leistner in De09, S. 427, 438; Ho04, S.34, 35; Ze15, S. 1151, 1157f.]

Um diese Situation geht es aber gerade im Zusammenhang mit den bedeutsamen Daten für den Bereich von Industrie 4.0. Es geht bei der Diskussion um die Datenhoheit darum, dass Unternehmen an ihren Maschinen oder Maschinenteilen Hard- und Softwareelemente anbringen, die beim Nutzer Daten über die Art und Weise der Nutzung sowie über die Funktionalität, die Anfälligkeit und die Abnutzung der Aggregate liefern, die Auskunft über Prozessverträglichkeiten, Nutzungsarten u.ä. generieren und übermitteln. Der Hersteller derartiger Maschinen benötigt die Daten für zahlreiche Zwecke, u.a. zur Optimierung seiner Produkte, zur Erfassung der Kundenwünsche, zur Marktbeobachtung und zur Einschätzung von Gewährleistungs- und Schadensersatzansprüchen.

Die Interessen der kommerziellen oder privaten Nutzer an der Weiterleitung der durch ihre Nutzungen generierten Daten werden recht unterschiedlich sein. Man kann aber unterstellen, dass zumindest die kommerziellen Nutzer bei Kenntnis über die Bedeutung und den wirtschaftlichen Wert der Daten diese Daten nicht ohne Gegenleistung herausgeben bzw. nicht generieren werden.

1.2 Zuordnung über das Vertragsrecht?

Die Frage nach der Nutzungsmöglichkeit und der dafür zu erbringenden Gegenleistung wird künftig wohl zunehmend in Verträgen geregelt werden. Damit ist aber nicht die Frage beantwortet, wem die Daten ursprünglich zugeordnet sind, wem sie gehören. Für die Vertragsgestaltung hat diese Frage zweifache Bedeutung. Es wird nur derjenige eine Gegenleistung erbringen, der etwas erhält, was ihm bislang nicht gehört. Bei den Vertragsregelungen wird es sich zudem regelmäßig um allgemeine Geschäftsbedingungen handeln und für eine Inhaltskontrolle ist die ursprüngliche Zuordnung ein wichtiges Kriterium.

1.3 Berechtigte Interessen an der Zuweisung

Bei der Frage nach der Berechtigung an den Daten gibt es zwei unterschiedliche Interessen zu berücksichtigen. Die Interessen des Unternehmens, das das mit einem Datenerfassungsgerät versehene Produkt liefert und das Unternehmen oder auch die privat nutzende Person, die die Daten durch Nutzung des Produkts generiert, und zwar in der eigenen betrieblichen oder privaten Sphäre.

Beim Herstellerunternehmen wird zu berücksichtigen sein, dass die Datenerhebung durch

⁴ EuGH, GRUR 2005, 244 – BHB; EuGH, GRUR 2005, 254 – Fixtures Marketing III.

deren Arbeit vorbereitet wurde. Hinsichtlich des nutzenden Unternehmens bzw. der nutzenden Privatperson wird zu berücksichtigen sein, dass von ihnen die Daten unmittelbar generiert werden.

1.4 Schutzmöglichkeiten de lege lata

1.4.1 Urheberrechtlicher Datenschutz

aa) Die Vorbereitung der Datenerhebung, d.h. die Entwicklung einer Struktur für die Einordnung der Daten, könnte nach dem Datenbankrecht, § 87 a UrhG, geschützt sein. Dieser Schutz bezieht sich aber nicht auf die zu erhebenden und entsprechend einzuordnenden Daten. Deren Schutz ist gerade nicht in den Schutz der Datenbankstruktur einbezogen; der EuGH hat dies gerade auf der Grundlage eines Vorlagebeschlusses des BGH bestätigt.⁵ Die europäische Richtlinie 96/9 bestimmt auch ausdrücklich, dass der Schutz nicht auf die Inhalte, die eingebrachten Daten, bezogen ist. Dies ist im Hinblick auf die Begründung des Schutzes selbstverständlich. Es soll durch dieses Leistungsschutzrecht nicht erreicht werden, dass die in eine Sammlung aufgenommenen Elemente selbst einen Schutz erfahren, nur weil sie für die jeweils gegenständliche Sammlung tauglich waren. Selbstverständlich ist auch, dass die einzelnen in der Sammlung befindlichen Elemente nicht ihren evtl. bestehenden Schutz verlieren, nur weil sie Aufnahme gefunden haben.

Mit diesen durch die Richtlinie und den jeweils nationalen Ausführungsgesetzen vorgegebenen Regelungen ist aber für die Bestimmung der genauen Schutzvoraussetzungen noch nicht viel gewonnen.

bb) Damit aus urheberrechtlicher Sicht überhaupt eine Sammlung, eine Datenbank besteht, muss es sich nach Art. 1 Abs. 2 der Richtlinie 96/9/EG um eine Sammlung voneinander unabhängiger Elemente handeln. Dies ist im Hinblick auf die urheberrechtliche Ordnung eine sich selbst erklärende Anforderung. Soweit die einzelnen Elemente nur zusammenhängend einen Sinn ergeben und anders nicht verständlich, nicht irgendwie sinnvoll einzuordnen wären, läge keine Sammlung von Daten vor, sondern ein zusammengehöriges Element, evtl. schon ein Werk oder eben ein schutzloser „Gegenstand“.

Der EuGH hat auf der Grundlage eines Vorlagebeschlusses des BGH jüngst entschieden, dass diese Unabhängigkeit großzügig zu bestimmen ist. Es soll nicht allein auf die Zweckbestimmung der Sammlung abgestellt werden, sondern auch auf alle irgendwie möglichen Verwendungsmöglichkeiten. Gibt es danach weitere Verwendungsmöglichkeiten der einzelnen Elemente einer Sammlung, ist deren Unabhängigkeit gegeben. Dem ist zuzustimmen. Käme es für die Feststellung der Unabhängigkeit allein auf die Zweckbestimmung für die konkrete Sammlung an, könnte die Unabhängigkeit nicht mehr nachgewiesen werden; jede Datenbank würde die Schutzfähigkeit verlieren, weil schon keine voneinander unabhängigen Daten aufgenommen wären.⁶

⁵ EuGH, GRUR 2015, 1187 ff.; BGH, GRUR 2014, 1197.

⁶ EuGH, GRUR, 2015, 1187 ff.

cc) Damit steht nun aber die Frage an, ob die schutzbegründende Zusammenfassung an sich unabhängiger Elemente in einer Datenbank irgendeine Qualität oder auch nur Besonderheit haben muss, um durch die Zusammenfassung von grundsätzlich ungeschützten und voneinander unabhängigen Elementen einen Leistungsschutz zu erlangen. Richtliniengeber und Gesetzgeber haben nur geregelt, dass Schutzvoraussetzung die Existenz einer Sammlung sein muss, und diese muss wiederum durch erhebliche Investitionen geschaffen worden sein.

Wesen der Leistungsschutzrechte ist es, die Amortisation gewerblich geschaffener Leistungen zu schützen. Die Verwendung kostengünstiger Übernahmetechniken soll nicht dazu führen, dass der ErsthHersteller die Chance auf Amortisation verliert. Von daher betrachtet ist die Aufnahme des Investitionserfordernisses in § 87 a UrhG sicher nicht falsch, der Normzweck wird benannt. Fraglich ist nur, ob dies reicht. Die Frage gewinnt an Bedeutung, wenn in der Literatur nun auch gefolgert wird, dass es keine irgendwie ins Gewicht fallenden Anforderungen an die Systematik der Sammlung, an ihre Ordnungselemente, zu geben braucht; es soll schon genügen, dass die Daten wieder auffindbar sind [*Dreier* in DSS15, §87a Rdnr.7; *Thum/Hermes* in WB14, §87a, Rdnr.21; *Ze15*, S. 1151, 1157 Fn.51].

Dem ist wohl zuzustimmen, weil es beim Datenbankschutz gerade nicht um den Schutz eines Werkes (im urheberrechtlichen Sinne) geht; insofern ist es richtig, auch einfache Ordnungsstrukturen genügen zu lassen. Dann gewinnt aber die Schutzvoraussetzung „erhebliche Investitionen“ in § 87 a UrhG an Bedeutung. Wenn die semantische Ebene, die jeweilige Ordnungsstruktur, zu vernachlässigen ist, muss sich die Anforderung auf das Sichten und Sammeln beziehen. Es wird dann aber fraglich, wem die Daten bzw. die Sammlung von Daten zuzuordnen sind – dem die Daten generierenden Nutzer oder dem die Erhebung vorbereitenden Unternehmen.

Der EuGH hat vielleicht wegen dieser Schwierigkeit den Schutz auf Investitionen in vorhandene Daten beschränkt. Damit wird dann aber auch die Frage nach dem Schutz erst zu generierender Daten aus dem Datenbankrecht herausgelöst.⁷

1.4.2 Schutz der Daten als Betriebsgeheimnis

Die Daten sind auch nicht als Geschäftsgeheimnis für den Hersteller der Erfassungseinrichtungen geschützt. Der Geheimnisschutz nach § 17 UWG wie auch der Geheimnisschutz nach dem europäischen Richtlinienentwurf hat zur Voraussetzung, dass die entsprechenden Informationen dem Unternehmen zuzuordnen sind. Im Richtlinienentwurf steht, dass der „Inhaber des Geschäftsgeheimnisses“ geschützt wird [*Ze15*, S. 1151, 1155].⁸

1.4.3 Weitere Leistungsschutzrechte

Leistungsschutzrechte, die Informationen als solche schützen, sind nicht vorhanden. Auch

⁷ EuGH, GRUR 2005, 244; EuGH, GRUR 2005, 254.

⁸ Art. 2 I lit. C des Richtlinienentwurfs.

das neu ins Urheberrecht eingefügte Schutzrecht für die Presseverleger hat keine Vorbildfunktion, weil die einfachen Daten, also Daten, die nicht irgendwie redaktionell aufgearbeitet wurden, gerade nicht geschützt sind.

Wenn nun davon auszugehen ist, dass weder die Vorbereitung der Datenerhebung Einfluss auf den möglichen Schutzbereich hat, die Daten auch nicht über den Geheimnisschutz geschützt sind, aber auch die Art und Weise der Nutzung insofern ohne Bedeutung sind, wird es schwierig, Zuordnungskriterien zu entwickeln.

1.5 Rechtsdogmatische Anknüpfungspunkte für eine Regelung zum Schutz der Daten

1.5.1 Übernahme sachenrechtlicher Grundsätze

Im Zusammenhang mit dem Schutz der Informationen ist in der Literatur häufig vom „Dateneigentum“ die Rede [Ze15, S. 1151, 1153]. Es bereitet zunächst Schwierigkeiten, die Frage nach einer Berechtigung an Daten bzw. der sog. Datenhoheit vom Sacheigentum abzuleiten.

Die Ableitung ist aufgrund der bürgerlich-rechtlichen Konzeption des Eigentums aus dem 19. Jahrhundert kaum möglich. Das Sacheigentum wurde als eine nahezu allumfassende Berechtigung angesehen (§ 903 BGB).

Dieser Grundsatz ist im Hinblick auf die vielfältigen Nutzungsbeschränkungen zwar ziemlich ausgehöhlt, die Einschränkungen folgen aber weniger aus der Natur der Sache, sondern sind sozial, wettbewerbspolitisch etc. orientiert.

Die Kritik am Immaterialgüterrecht richtet sich dann auch vornehmlich gegen eine zu starke Orientierung am Sacheigentum. So war es über Jahrzehnte üblich, eine Annäherung des Immaterialgüterrechts an das Sacheigentum zu verlangen; das Immaterialgüterrecht sollte nicht mindere Rechte zuteilen, sondern eine möglichst gleiche Qualität haben.⁹

Die Kritik wird in jüngster Zeit zumindest in grundlegenden wissenschaftlichen Veröffentlichungen mehr institutionenökonomisch geführt [Go07, S. 505 ff., 525, 527].¹⁰ Zumindest in der modernen Privatrechtstheorie hat sich mittlerweile die Konzeption eines einheitlichen Eigentums gewandelt, und zwar hin zu einem Verständnis vom Eigentum als „Rechtebündel“. Es wird dahin argumentiert, dass es einen ungeteilten Eigentumsbegriff nicht geben kann, weil der Ausgleich von Individuellem unterschiedliche „Mischverhältnisse“ ergeben muss. Es wird dahin argumentiert, dass, wenn es ein absolutes Eigentumsrecht gäbe, der Gesetzgeber auch nichts mehr zu gestalten bzw. zwischen widerstreitenden Interessen abzuwägen hätte [Go07, S.549; Le02, S.69].

⁹ Die wohl einflussreichsten Lehren im deutschsprachigen Raum zum Immaterialgüterrecht (Ulmer, Troller) zielten auf einen möglichst umfassenden, dem Sacheigentum angenäherten Schutz mit nur konkret benannten Sozialschranken als Ausnahmesituationen hin.

¹⁰ Zur Rechtsprechung siehe auch die Handelsvertreterentscheidung des BVerfG aus 1990, BVerfGE 81, 242.

Im Immaterialgüterrecht gibt es seit einiger Zeit eine entsprechende Entwicklung. Galt es noch bis in die achtziger Jahre hinein z.B. im Urheberrecht die Schutzbereiche allein aus der Interpretation des Rechtsbegriffs „geistig persönliche Schöpfung“ zu gewinnen, wird heute der Schutzbereich auch durch eine Abwägung zwischen den Interessen des Berechtigten und dem Freihaltungsinteresse der Allgemeinheit festgelegt.

So hat auch im Hinblick auf das Urheberrecht der BGH die Schranken als nicht mehr eng zu begrenzende Ausnahmen angesehen, sondern als durchaus auslegungsfähige Regelungsbereiche, die das Interesse der Allgemeinheit an Freihaltung auch widerspiegeln sollen. Im Patentrecht zeigt sich diese Rechtsprechung im Zusammenhang mit dem Spannungsverhältnis zwischen patentrechtlichem Schutz und Standardisierung. Man kann durchaus auch der Ansicht sein, dass die wissenschaftliche Diskussion um den patentrechtlichen Schutz der Computersoftware oder um den patentrechtlichen Schutz auf dem Gebiet der Genexpressionen einzig um die Frage nach dem Freihaltungsinteresse kreist, wenn auch die Ansatzpunkte sich unterscheiden.¹¹

1.5.2 Übernahme von Regelungen bezüglich des Sacheigentums

Für die anstehende Frage nach einem Schutz für Informationen hat diese Gleichstellung zumindest wesentlicher Grundlagen der beiden Rechtsgebiete dort Bedeutung, wo es um die Übernahme von Regelungsbereichen geht. In Betracht kommt hier § 950 BGB, der die Rechtsfolgen der Bearbeitung eines Rohstoffes bzw. die der Weiterbearbeitung eines Produkts regelt. Der Bearbeiter erwirbt das Eigentum, soweit nicht der Wert der Bearbeitung geringer als der des Stoffes ist. Die Regelung erscheint passend. Bei den gegenständlichen Informationen handelt es sich um „Rohmaterialien“, um unbearbeitete Informationen, Rohinformationen [Ze15, S. 1151 ff.] (mit denen etwas geschieht, die – wenn auch nicht im Sinne von Patent- und Urheberrecht – bearbeitet werden.

Der Normzweck von § 950 BGB ist auf den Leistungsschutz gerichtet und zwar in besonderer Art. Es wird dem Bearbeiter bei entsprechendem Wert der Bearbeitung nicht nur ein schuldrechtlicher Ausgleichsanspruch gegen den Eigentümer der Sache eingeräumt, sondern der bisherige Eigentümer verliert sein Eigentum und erhält seinerseits einen Ausgleichsanspruch zugewiesen. Die Regelung könnte dahin kritisiert werden, dass nur der Wert der Bearbeitung ausgeglichen werden sollte und nicht der Übergang der Verfügungsbefugnis, der Übergang des Eigentums.

Der Grund dafür liegt darin, dass mit einer Bearbeitung die Sache einer neuen bzw. einer bestimmten Zweckbestimmung zugeführt wird. Soweit dies nicht unerlaubterweise geschieht, ist es sachgerecht, nun auch dem Bearbeiter die Verfügungsmöglichkeit über die derart von ihm bearbeitete Sache zuzuordnen. Das ist interessengerecht, soweit man nachvollziehen will – und wohl auch sollte –, dass eine ins Gewicht fallende, für eine Nutzung

¹¹ Das US-amerikanische Urheberrecht kann dies noch mehr verdeutlichen. Im Unterschied zum Sacheigentum gibt es dort keine konkret benannten Sozialschranken, sondern eine allgemeine „fair-use-Regel“. Einfach gependet: Wenn die Nutzung nicht die Interessen des Rechtsinhabers berühren, kann genutzt werden.

des Ausgangsstoffes bedeutsame Bearbeitung wohl demjenigen am besten nützt, der entsprechend bearbeitet hat bzw. der Bearbeitung entsprechend nutzen will. Verlangt ist nach § 950 BGB, dass eine neue Sache entsteht.

Die Frage, wer Bearbeiter und wer Lieferant des Ausgangsstoffes ist, lässt sich gut nachvollziehbar bearbeiten. Bearbeiter ist das Unternehmen, das die technischen Vorrichtungen zur Erfassung und Übermittlung der Daten an der jeweiligen Maschine anbringt; der Nutzer, der die Daten generiert, liefert das Rohmaterial. Regelmäßig wird es sich so verhalten, dass die Datengenerierung ohne gesonderten Aufwand erfolgt, und weiterhin wird man regelmäßig dahin urteilen können, dass der Wert der Rohdaten erst durch die vorbereitende Bearbeitung wertvoll wird. Die vorbereitende Bearbeitung schafft erst die semantische Ebene.

1.6 Neues Leistungsschutzrecht

Der rechtlichen Einordnung nach würde es sich bei einem dem Normzweck von § 950 BGB nachgeordneten Recht um ein Leistungsschutzrecht handeln. Belohnt wird nicht die neue technische Idee oder eine geistig persönliche Schöpfung, sondern der mit der Erhebung der Daten erforderliche Aufwand im gewerblichen Bereich. Geschaffen würde auch nicht ein Ausschließlichkeitsrecht an Daten der entsprechenden Art. Jedermann dürfte entsprechende Daten erheben und verwerten; verboten wäre die Entnahme der bzw. einzelner Daten aus der Sammlung. Von einem bereits vorhandenen Leistungsschutzrecht, dem urheberrechtlichen Datenbankrecht, würde sich das neue Leistungsschutzrecht unterscheiden, weil es hier um den Schutz der Daten selbst geht und nicht um den Schutz der Struktur, nach der die Daten geordnet sind. Begrenzt wäre dieses Recht um die Voraussetzung, dass die Informationsgenerierung vom Hersteller der benutzten technischen Einrichtung vorbereitet wurde.

Die durch ein solches Leistungsschutzrecht berührte Interessenkollision zwischen dem Hersteller der Gerätschaften und den die Daten durch Nutzung der Gerätschaften generierenden Personen würde gegen die Nutzer aufgelöst werden; dies wäre rechtspolitisch im Hinblick auf § 950 BGB sehr gut zu begründen. Derjenige, der berechtigterweise die (Roh-) Datenerhebung seinen Verwendungswünschen entsprechend vorbereitet, hat das Sachinteresse an derart eingeordneten Daten und derjenige, der die Rohdaten zur Verfügung stellt, der sie erhebt, erhält für den Verlust einen Wertausgleich.

Das letztlich zu lösende Problem ist dann die Bestimmung des Wertausgleichs. Nicht richtig wäre es, den Wert der Daten danach zu bestimmen, was sie dem Hersteller für seine Planungen von Wert sein könnten; ebenso falsch wäre es, den Wert danach zu bestimmen, welche Gegenleistung auf dem Markt – soweit er besteht – zu erlangen wäre. Solche Berechnungen sind für das Sacheigentum von Bedeutung, wegen der durch die Körperlichkeit gegebenen Exklusivität. Bei den Leistungsschutzrechten wird diese Exklusivität gerade nicht geschaffen. Die Daten könnten von jedermann erhoben werden, jeder könnte sie entsprechend strukturiert erheben. Dies hat dann auch für den gegenständlichen Ausgleichsanspruch Bedeutung. Der Anspruch ist am Aufwand im Zusammenhang mit der

Datenerhebung zu messen und dürfte gering ausfallen, weil die Daten regelmäßig „nebenbei“, bei zweckentsprechender Nutzung der Maschine erhoben werden. Eine sachgerechte Preisfindung wäre über einen Vergleich zu den Kosten für die Speicherung von Datenmengen bei Dritten angezeigt.

2 Der Schutz des Persönlichkeitsrechts durch das Datenschutzrecht

Die beschriebene Zunahme der Erhebung und Verarbeitung von Daten in der Industrie durch vernetzte „intelligente“ Maschinen oder Gerätschaften ist rechtlich über die bereits angesprochenen Rechtsbereiche hinaus unter den Aspekten des Schutzes personenbezogener Daten zu beurteilen. Wie bereits erwähnt, erfolgt die Verarbeitung¹² zunehmend über Hard- und Softwareelemente, die an Maschinen, Maschinenteilen, Gerätschaften oder sonstigen Gegenständen angebracht werden. Häufig können über die Auswertung solcher Daten auch Aussagen über einzelne Menschen abgeleitet werden (z.B. Art und Weise der Nutzung einer Maschine¹³).

Ist eine Zuordnung von Daten [zu den Begriffen „Angabe“, „Datum“ und „Information“, siehe: Ha15, S.116 ff.] zu einer einzelnen natürlichen Person möglich, führt dies i.d.R. zur sachlichen Anwendbarkeit des Datenschutzrechts, wenn durch die Zuordenbarkeit ein Personenbezug entsteht. Die Konsequenz ist, dass die strengen datenschutzrechtlichen Vorschriften zum Schutz personenbezogener Daten eingreifen, und somit die für die Verarbeitung verantwortliche Stelle verpflichtet wird, zahlreiche datenschutzrechtliche Grundsätze und Vorschriften zu beachten.

Das Datenschutzrecht dient – wie die anderen oben bereits erwähnten Rechtsbereiche – ebenfalls dem Ausgleich von unterschiedlichen Interessen. Auf der einen Seite stehen die Interessen der verarbeitenden Stelle(n) an der unbeschränkten Verarbeitung und Nutzung der Daten. Auf der anderen Seite werden durch die Vorschriften und Grundsätze des Datenschutzrechts die Interessen des sog. Betroffenen gewahrt, indem sie das Persönlichkeitsrecht des Betroffenen schützen.¹⁴ Das Datenschutzrecht ermöglicht der betroffenen Person und ggf. der zuständigen Behörde, unzulässige Erhebungs- und Verarbeitungsprozesse rechtlich zu verhindern bzw. zu verbieten.

Im Zusammenhang mit dem Phänomen „Industrie 4.0“ muss hinsichtlich jedes Datenverarbeitungsprozesses geprüft werden, ob und inwieweit das (personenbezogene) Datenschutzrecht anwendbar ist. Darüber hinaus ist festzulegen, welche Stelle für den Verarbeitungsvorgang verantwortlich ist. Soweit keine gesetzlichen Erlaubnisnormen eingreifen, die eine Verarbeitung von personenbezogenen Daten rechtfertigen, ist die verantwortliche

¹² Obwohl das deutsche Datenschutzrecht zwischen der „Erhebung“, „Verarbeitung“ und „Nutzung“ von personenbezogenen Daten differenziert, wird in diesem Beitrag zur Förderung der Übersichtlichkeit nur der Begriff „Verarbeitung“ verwendet.

¹³ Konkretes Beispiel: Der Arbeitnehmer A benutzt ein elektronisch kontrolliertes Werkzeug, dessen Betrieb und Nutzung analysiert werden.

¹⁴ Vgl. BVerfGE 65, 1 („Volkszählungsurteil“); § 1 Abs. 1 BDSG.

Stelle auf die Einholung einer wirksamen Einwilligung angewiesen.¹⁵

2.1 Der sachliche Anwendungsbereich des Datenschutzrechts

EU-weit setzt der sachliche Anwendungsbereich des Datenschutzrechts das Vorliegen personenbezogener Daten voraus.¹⁶ Nach Art. 2 a) EG-Datenschutzrichtlinie¹⁷ sind "personenbezogene Daten" alle Informationen über eine bestimmte oder bestimmbare natürliche Person ("betroffene Person").¹⁸ Die Voraussetzungen des Merkmals „personenbezogene Daten“ – insbesondere das Vorliegen von „Informationen“, der „Bezug zu einer natürlichen Person“ und die „Bestimmbarkeit“ – sind hoch umstritten und werden seit Jahren in Rechtsprechung und Literatur umfassend diskutiert [Ha15].

Das Phänomen „Industrie 4.0“ stellt zusätzliche, neue Anforderungen an die rechtliche Auslegung der Voraussetzungen des Merkmals „personenbezogene Daten“. Durch die neuen Maschinen, die zunehmend mit Sensoren ausgestattet und untereinander vernetzt werden, findet eine massenhafte Generierung und Verarbeitung von Daten statt, die in erster Linie zwar von den Maschinen „stammen“ aber vielfach auch Aussagen über einzelne Menschen enthalten.

Das Merkmal „Information“ wird in der datenschutzrechtlichen Literatur und Rechtsprechung weit ausgelegt. Hierunter fallen Informationen „unabhängig von der jeweiligen Semantik, Sigmantik, Pragmatik, Darstellungsart, Darstellungsform und Herkunft.“ [Ha15, S. 453] In dieser Hinsicht werden im Rahmen der „Industrie 4.0“ unzählige neue Datenverarbeitungsprozesse entstehen, durch die diese Kriterien erfüllt sind und damit potenzielle Anknüpfungspunkte für das Datenschutzrecht geboten werden.

Personenbezogene Daten liegen allerdings nur dann vor, wenn sich die Informationen (zumindest auch) auf eine natürliche Person beziehen. Vordergründig sind Informationen, die in der Industrie erhoben werden, vielfach Informationen über Gegenstände (z.B. Maschinen). Teilweise beziehen sich die Informationen jedoch nicht nur auf Gegenstände, sondern auch auf deren Nutzer. Auch Informationen, die sich in erster Linie auf einen Gegenstand beziehen, zusätzlich jedoch Aussagen über eine bestimmbare natürliche Person

¹⁵ Anmerkung: Neben dem Schutz personenbezogener Daten durch das Datenschutzrecht bestehen einzelne Gesetzesvorschriften zum Schutz des sog. „Know-hows“ (Bsp.: § 17 UWG, § 202a ff. StGB), welche vorliegend jedoch nicht näher betrachtet werden.

¹⁶ Vgl. Art. 1 Richtlinie 95/46/EG des Europäischen Parlaments und des Rates vom 24. Oktober 1995 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr, Amtsblatt Nr. L 281 vom 23/11/1995, S. 0031 ff. (EG-Datenschutzrichtlinie 95/46/EG; ab dem 25.05.2018 wird die EG-Datenschutzrichtlinie 95/46/EG durch die neue Datenschutz-Grundverordnung ersetzt werden).

¹⁷ Richtlinie 95/46/EG des Europäischen Parlaments und des Rates vom 24. Oktober 1995 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr, Amtsblatt Nr. L 281 vom 23/11/1995, S. 0031 ff.

¹⁸ Eine ähnliche Definition findet sich in Art. 4 Nr. 1 der Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung), Amtsblatt Nr. L 119 vom 04/05/2016, S. 1 ff. (Inkrafttreten: 25.05.2018; Anwendbar ab: 25.05.2018).

enthalten, sind als personenbezogen anzusehen [Ha15, S. 144 f., 174 f.].

Als bestimmbar wird eine Person angesehen, die direkt oder indirekt identifiziert werden kann, insbesondere durch Zuordnung zu einer Kennnummer oder zu einem oder mehreren spezifischen Elementen, die Ausdruck ihrer physischen, physiologischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität sind (Art. 2 a) HS 2 EG-Datenschutzrichtlinie 95/46/EG).¹⁹ In Rechtsprechung und Literatur ist nach wie vor umstritten, ob und inwieweit Zusatzwissen – insbesondere dritter Stellen – mit in die Beurteilung der Bestimmbarkeit einbezogen werden [Ha15, S. 290 ff.]. Die angesprochenen vernetzten, „intelligenten“ Maschinen verarbeiten über eine Person i.d.R. nicht die „klassischen“ personenbezogenen Daten wie Name, Vorname, Anschrift und Geburtsdatum. Vielmehr handelt es sich um Daten, die die Nutzung einer solchen Maschine durch einen Menschen automatisch mit sich bringen. Hierzu zählt z.B. das Arbeitsverhalten (Wie wird eine Maschine durch eine Person genutzt?) oder der Standort (Wo befindet sich die Person, die eine Maschine nutzt?). Häufig können diese Daten nicht direkt einer Person zugeordnet werden. Eine direkte Zuordnung ist auch nicht erforderlich. Denn bei der Frage der Bestimmbarkeit sind zusätzliche Möglichkeiten der verantwortlichen Stelle zu beachten.

Nach dem objektiven Ansatz sind unabhängig von den individuellen Möglichkeiten der verantwortlichen Stelle zusätzlich alle „objektiv“ zur Verfügung stehenden Zusatzmöglichkeiten einzubeziehen [Ge13, S.478 ff. (479); Pa08, S.34 ff. (38 f.)]. Nach dem relativen Ansatz kommt es in erster Linie vielmehr auf die individuellen und konkreten Möglichkeiten der jeweils verarbeitenden Stelle an [Ec11, S. 339 (342); GS15, §3, Rdnr.10].²⁰ Sowohl die streng objektive Ansicht als auch die streng relative Ansicht sind abzulehnen. [AFK06, S. 700 ff. (704)] Nach einer vermittelnden Ansicht sind „dritte Stellen in die Beurteilung mit einzubeziehen (objektives Element), wobei diese Einbeziehung über die Merkmale „unverhältnismäßiger Aufwand“ (Art. 2 a) i.V.m. EG 26 S. 2 EG-Datenschutzrichtlinie 95/46/EG) einzuschränken ist.“ [Ha15, S.320] Nach Erwägungsgrund 26 S. 2 sollten bei der Entscheidung, ob eine Person bestimmbar ist, alle Mittel berücksichtigt werden, die vernünftigerweise entweder von dem Verantwortlichen für die Verarbeitung oder von einem Dritten eingesetzt werden könnten, um die betreffende Person zu bestimmen.

2.2 Verantwortliche Stelle

Im Rahmen der Vernetzung von Gegenständen und Maschinen wird es zunehmend zu einer „Vermischung“ von verantwortlichen Stellen kommen. Nach Art. 2 d) EG-Datenschutzrichtlinie 95/46/EG ist der "für die Verarbeitung Verantwortliche" die

¹⁹ Eine ähnliche Definition findet sich in Art. 4 Nr. 1 der Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung), Amtsblatt Nr. L 119 vom 04/05/2016, S. 1 ff.

²⁰ AG München, Urteil vom 30.09.2008, Az.: 133 C 5677/08, ZUM-RR 2009, 413 ff. (414).

natürliche oder juristische Person, Behörde, Einrichtung oder jede andere Stelle, die allein oder gemeinsam mit anderen über die Zwecke und Mittel der Verarbeitung von personenbezogenen Daten entscheidet.²¹ Im Rahmen der „Industrie 4.0“ kommt es zu einer Vermischung der Entscheidungsbefugnisse und -möglichkeiten. Über Informationen die mittels Hard- und Softwareelementen erhoben und verwendet werden, die an einer Maschine oder Maschinenteilen angebracht sind, entscheidet häufig sowohl der Hersteller, ggf. der Anbieter als auch der Nutzer.

So überwachen heutzutage beispielsweise Hersteller von Fahrstühlen über verschiedene Sensoren die verschiedenen Abläufe des Fahrstuhlbetriebes. Nichtsdestotrotz findet der Vorgang oft in der Sphäre des Gebäudeeigentümers statt. Nutzer kann schließlich ein Mieter sein, der mit dem Fahrstuhl direkt in seine Wohnung fährt. Darüber hinaus finden Verkettungen von Daten unabhängig von menschlichen Einflüssen automatisch bei den Geräten oder Maschinen untereinander statt. Gerade bei Alltagsgegenständen liegt wiederum die direkte Kontrolle eines vernetzten Produktes (Zahnbürste, Kühlschrank, etc.) in der Regel bei dem Nutzer, der das Produkt häufig als Privatperson nutzt.

Sowohl das deutsche als auch das europäische Datenschutzrecht basieren auf der Vorstellung, dass jeder Erhebungs-, Verarbeitungs- und Nutzungsvorgang einer sog. verantwortlichen Stelle zuzuordnen ist (vgl. §§ 1, Abs. 2, 2, 3 Abs. 7 BDSG; Art. 2 d) S. 1 Datenschutzrichtlinie 95/46/EG). In der deutschen Literatur und Rechtsprechung finden sich verschiedene Konkretisierungen. Teilweise wird in Bezug auf das BDSG vertreten, dass es bezogen auf einen Datenvorgang auch mehrere verantwortliche Stellen geben könne [Dammann in Si14, §3, Rdnr.226]. Der Umgang mit personenbezogenen Daten könne auch durch „mehrere natürliche oder juristische Personen (...) in gemeinsamer Verantwortung“ erfolgen [Dammann in Si14, §3, Rdnr.226]. Dem ist zuzustimmen, da ein Datensatz von mehreren Stellen gleichzeitig erhoben, verarbeitet oder genutzt werden kann (vgl. § 3 Abs. 3 – 5 BDSG) und jede dieser Erhebungen, Verarbeitungen oder Nutzungen eine Verantwortlichkeit begründet.²²

Der Begriff „verantwortliche Stelle“ wurde mit dem BDSG 2001 in das deutsche Datenschutzrecht eingeführt. Ersetzt wurde dadurch das Wort „speichernde Stelle“. Nach § 3 Abs. 7 BDSG ist verantwortliche Stelle jede Person oder Stelle, die personenbezogene Daten für sich selbst erhebt, verarbeitet oder nutzt oder dies durch andere im Auftrag vornehmen lässt. Verantwortlich sei, „wer objektiv über die Daten bestimmen kann, wer die Entscheidungsgewalt über den Zweck und die Mittel der Datenverarbeitung hat“.²³ [Wei-

²¹ Nach Art. 4 Nr. 7 der Datenschutz-Grundverordnung (FN. 12) ist: „Verantwortlicher“ die natürliche oder juristische Person, Behörde, Einrichtung oder andere Stelle, die allein oder gemeinsam mit anderen über die Zwecke und Mittel der Verarbeitung von personenbezogenen Daten entscheidet; sind die Zwecke und Mittel dieser Verarbeitung durch das Unionsrecht oder das Recht der Mitgliedstaaten vorgegeben, so kann der Verantwortliche beziehungsweise können die bestimmten Kriterien seiner Benennung nach dem Unionsrecht oder dem Recht der Mitgliedstaaten vorgesehen werden“.

²² vgl. § 3 Abs. 7 BDSG; eine ähnliche gemeinsame Verantwortung existiert auch im allgemeinen Zivilrecht z.B. bei der gesamtschuldnerischen Haftung.

²³ Auch die EG-Datenschutzrichtlinie 95/46/EG stellt darauf ab, wer über die Zwecke und Mittel der

chert in Dä14, §3 Rdnr. 54] Für die Begründung einer datenschutzrechtlichen Verantwortung i.S.d. BDSG würde ausreichen, „dass die Verarbeitungstätigkeit im eigenen Tätigkeits- und Haftungsbereich stattfindet und die Möglichkeit besteht, in tatsächlicher Hinsicht auf den Verarbeitungsvorgang einzuwirken.“ [Buchner in TG13, §3, Rdnr.52] Bei der Abgrenzung komme es in erster Linie auf eine „juristische“ und nicht auf eine „funktionale“ Betrachtung an [Weichert in Dä14, §3 Rdnr. 54].

Die Verfügungs- und Entscheidungsgewalt sind geeignete Anknüpfungspunkte für die Festlegung der verantwortlichen Stelle, denn diese sind unmittelbar ausschlaggebend für die Missbrauchsfahr und das Gefährdungspotenzial im Hinblick auf das informationelle Selbstbestimmungsrecht. Auch die reine Einwirkungsmöglichkeit kann ausreichen, denn nach § 3 Abs. 4 Nr. 3 b) BDSG ist Übermitteln auch das Bekanntgeben gespeicherter oder durch Datenverarbeitung gewonnener personenbezogener Daten an einen Dritten in der Weise, dass der Dritte zur Einsicht oder zum Abruf bereitgehaltene Daten einsieht oder abrufen. Durch eine solche Übermittlung kann also eine weitere Verantwortlichkeit entstehen. Dieser Vorgang führt i.d.R. jedoch nicht dazu, dass die ursprüngliche Stelle die Verantwortung verliert. In erster Linie müssen die objektiven und tatsächlichen Umstände herangezogen werden.²⁴

Die datenschutzrechtliche Verantwortlichkeit setzt weder den „Besitz der Daten“ noch die „physische Herrschaft über den Verarbeitungsprozess“ voraus [Dammann in Si14, §3, Rdnr. 225]. Das bedeutet, dass das Eigentum an der oder die Verantwortung für die Datenverarbeitungsanlage und die Verantwortung für die Datenerhebung, -verarbeitung oder -nutzung auseinanderfallen können [Dammann in Si14, §3, Rdnr. 224]. Dies ergibt sich aus den Regelungen zur Auftragsdatenverarbeitung (vgl. § 11 BDSG). Auch kommt es nicht auf die „Belegenheit der Datenverarbeitungsanlage“ an. [Weichert in Dä14, Rdnr. 56]

In den vergangenen Jahren wurde die datenschutzrechtliche Verantwortung in der Bundesrepublik Deutschland unter anderem im Zusammenhang mit der Nutzung sozialer Netzwerke und dem Anbieten von Suchmaschinen diskutiert. Biete ein soziales Netzwerk seinen Nutzern die Möglichkeit, eine sog. „Fanpage“ einzurichten, läge die datenschutzrechtliche Verantwortung bei dem Anbieter des sozialen Netzwerks, wenn dieser über das Netzwerk unmittelbar die IP-Adresse erhalte und Cookies anlege.²⁵ Die datenschutzrechtliche Verantwortlichkeit sei untrennbar mit der Datenerhebung und -verarbeitung verbunden.²⁶ Außerdem komme es darauf an, wer die Entscheidung über die Erhebungen und Verarbeitungen treffe und welche Zwecke verfolgt würden.²⁷ Auch sei relevant, wer die „Mittel zur Erreichung des Zweckes“ sowie „Art und Weise“ der Zielerreichung festlege.²⁸

Verarbeitung entscheidet (vgl. Art. 2 d) S. 1).

²⁴ Allerdings kann über einen Willen zur Einrichtung einer Auftragsdatenverarbeitung nach § 11 BDSG die Verantwortlichkeit auch subjektiv beeinflusst werden. Schließlich müssen in diesem Fall mit Schaffung einer entsprechenden („juristischen“) Vereinbarung wiederum tatsächliche Umstände geschaffen werden.

²⁵ OVG für das Land Schleswig-Holstein, Urteil vom 04.09.2014, Az.: 4 LB 20/13, juris, Rdnr. 77.

²⁶ OVG für das Land Schleswig-Holstein, Urteil vom 04.09.2014, Az.: 4 LB 20/13, juris, Rdnr. 78.

²⁷ OVG für das Land Schleswig-Holstein, Urteil vom 04.09.2014, Az.: 4 LB 20/13, juris, Rdnr. 78.

²⁸ OVG für das Land Schleswig-Holstein, Urteil vom 04.09.2014, Az.: 4 LB 20/13, juris, Rdnr. 79.

Dieser Bewertung stehe auch nicht entgegen, dass derjenige, der die „Fanpage“ eingerichtet hat (sog. „Fanpage-Betreiber“), ein eigenes Interesse an der Datenverarbeitung haben könnte und von dieser (mit)profitiert.²⁹ Diese Kriterien lassen sich auf typische Vorgänge im Rahmen der Industrie 4.0 übertragen. Pauschal gesagt, können stets auch die Personen und Stellen in die Verantwortung genommen werden, die bildlich gesprochen im Hintergrund stehen.

2.3 Einwilligung

Innerhalb der Europäischen Union müssen verantwortliche Stellen jede Erhebung und Verarbeitung von personenbezogenen Daten rechtfertigen (vgl. Art. 6 ff. EG-Datenschutzrichtlinie 95/46/EG, Art. 5 ff. Datenschutz-Grundverordnung, § 4 Abs. 1 BDSG). Soweit keine Erlaubnisnorm vorliegt, ist eine Einwilligung des Betroffenen erforderlich. Die datenschutzrechtliche Einwilligung ist basierend auf dem informationellen Selbstbestimmungsrecht an besondere Voraussetzungen gebunden (Bestimmtheit, Informiertheit, Freiwilligkeit, i.d.R. Schriftform, vgl. Art. 2 h) EG-Datenschutzrichtlinie 95/46/EG, Art. 7 Datenschutz-Grundverordnung, § 4a BDSG). Die Informationspflicht stellt dort eine rechtliche Herausforderung dar, wo die Informationsverarbeitung auf komplexen oder geheimhaltungsbedürftigen technischen Vorgängen basiert. Die Schriftform könnte insbesondere im Widerspruch zur steigenden Automatisierung stehen. Die Auslegung dieser Voraussetzungen hat die gesellschaftliche und technische Entwicklung in den Blick zu nehmen.

Literaturverzeichnis

- [AFK06] Arning, Marian; Forgó, Nikolaus; Krügel, Tina: Datenschutzrechtliche Aspekte der Forschung mit genetischen Daten, *Datenschutz und Datensicherheit (DuD)* 2006, S. 700 – 705.
- [Dä14] Däubler, Wolfgang; Klebe, Thomas; Wedde, Peter; Weichert, Thilo: *Bundesdatenschutzgesetz – Kompaktkommentar zum BDSG*, 4. Auflage, Bund Verlag, Frankfurt am Main, 2014.
- [DSS15] Dreier, Thomas; Schulze, Gernot; Specht, Louisa: *Urheberrechtsgesetz*, 5. Auflage, C.H.BECK, München, 2015.
- [De09] Derclaye, Estelle: *Research Handbook on the Future of EU Copyright*, Edward Elgar, Cheltenham, 2009.
- [Ec11] Eckhardt, Jens: IP-Adresse als personenbezogenes Datum – neues Öl ins Feuer, *Personenbezug im Datenschutzrecht, Computer und Recht (CR)* 2011, S. 339 – 344.
- [Ge13] Gerlach, Carsten: *Personenbezug von IP-Adressen, Computer und Recht (CR)* 2013, S. 478 – 484.

²⁹ OVG für das Land Schleswig-Holstein, Urteil vom 04.09.2014, Az.: 4 LB 20/13, juris, Rdnr. 78.

- [Go07] Godt, Christine: Eigentum an Information: Patentschutz und allgemeine Eigentumstheorie am Beispiel genetischer Information, Mohr Siebeck, München, 2007.
- [GS15] Gola, Peter; Schomerus, Rudolf: BDSG Bundesdatenschutzgesetz Kommentar, 12. Auflage, C.H.BECK, München, 2015.
- [Ha15] Haase, Martin Sebastian, Datenschutzrechtliche Fragen des Personenbezugs, Mohr Siebeck, Tübingen, 2015.
- [Ho04] Hoeren, Thomas: Anmerkung zu EuGH Urteil vom 09.11.2004 – C-203/02 (Datenbank-schutz) MultiMedia und Recht (MMR) 2005, Heft 1, 34 – 36.
- [Le02] Lepsius, Oliver: Besitz und Sachherrschaft im öffentlichen Recht, Mohr Siebeck, Tübingen, 2002.
- [Pa08] Pahlen-Brandt, Ingrid: Datenschutzrecht braucht scharfe Instrumente, Datenschutz und Datensicherheit (DuD) 2008, S. 34 - 40.
- [Si14] Simitis, Spiros: Bundesdatenschutzgesetz – Kommentar, 8. Auflage, Nomos, Baden-Baden, 2014.
- [TG13] Taeger/Gabel, Kommentar zum BDSG und den Datenschutzvorschriften des TKG und TMG, 2. Auflage, Fachmedien Recht und Wirtschaft, Frankfurt a.M., 2013.
- [WB14] Wandtke, Artur-Axel; Bullinger, Winfried; Praxiskommentar zum Urheberrecht, 4. Auflage, C.H.BECK, München, 2014.
- [Ze15] Zech, Herbert: “Industrie 4.0” – Rechtsrahmen für eine Datenwirtschaft im digitalen Binnenmarkt, Gewerblicher Rechtsschutz und Urheberrecht (GRUR) 2015, Heft 12, 1151 - 1160.

Reifegradmodell Industrie 4.0 - Ein Vorgehensmodell zur Identifikation von Industrie 4.0 Potentialen

Herbert Jodlbauer¹ und Michael Schagerl²

Abstract: Zahlreiche Kongresse, Tagungen und Symposien werden unter dem Begriff „Industrie 4.0“ abgehalten. Für eine breite Anwendung in Unternehmen fehlt es oft an Umsetzungsideen mit entsprechender Bewertungsmöglichkeit. Das Reifegradmodell Industrie 4.0 ist ein strategiegeleitetes Vorgehensmodell für Unternehmen, das in den Dimensionen Daten, Intelligenz und Digitale Transformation den Industrie 4.0 Reifegrad bestimmt. Nach Feststellung des IST-Reifegrades lässt sich strategiegeleitet der angestrebte SOLL-Reifegrad bestimmen. Konkrete Projektvorschläge zur Erreichung dieses Sollzustandes lassen sich ebenfalls aus dem Reifegradmodell für die teilnehmenden Unternehmen ableiten.

Keywords: Reifegradmodell Industrie 4.0, Digitale Transformation, Big Data, Maschinelle Intelligenz

1 Einleitung und Problemstellung

Der Begriff Industrie 4.0 ist in der Öffentlichkeit zurzeit weit verbreitet. Industrie 4.0 verspricht eine neue Zukunft, die mit innovativen Technologien, Vernetzung und neuen Geschäftsmodelle frischen Wind in die heimischen Unternehmen bringen soll. Das Reifegradmodell Industrie 4.0 leistet einen Beitrag zur Dissemination von Industrie 4.0 und liefert Unternehmen ein breiteres Verständnis und Umsetzungsvorschläge zu Industrie 4.0.

Eine wichtige Rolle kommt unter Industrie 4.0 der Informationsverarbeitung zu. Neue Entwicklungen in der IT ermöglichen Lösungen, die bis vor wenigen Jahren unvorstellbar gewesen wären. Ein mit der Entwicklung einhergehender Preisverfall erweitert zusätzlich den Handlungsspielraum für Unternehmen [Jo16]. Kaufentscheidungen des Kunden können besser verstanden werden. Produkte werden individueller gestaltet. Prozesse werden in Echtzeit dargestellt. Die IT wird als Enabler für Industrie 4.0 gesehen. Die IT ermöglicht aufgrund des Preisverfalls und der Leistungssteigerung die Substitution von etablierten Geschäftsmodellen und Produkten. Innovative Geschäftsmodelle werden innerhalb kurzer Zeit neue Märkte erobern und bestehende Produkte verdrängen [Br15, PH14].

¹ Institut für Intelligente Produktion, Fachhochschule OÖ, Campus Steyr, Wehrgrabengasse 1 – 3, A-4400 Steyr, Herbert.Jodlbauer@fh-steyr.at

² Institut für Intelligente Produktion, Fachhochschule OÖ, Campus Steyr, Wehrgrabengasse 1 – 3, A-4400 Steyr, Michael.Schagerl@fh-steyr.at

Bei bestehenden Unternehmen entsteht unter Industrie 4.0 die Notwendigkeit, Prozesse auf ihre Wirtschaftlichkeit hin zu überprüfen. Neue IT-Entwicklungen sind in der Lage unternehmerische Prozesse effizienter zu gestalten oder die Menschen in der Arbeitswelt in einem größeren Umfang als bisher zu unterstützen. Dadurch werden Unternehmen vor folgende Fragen gestellt: Durch welche Industrie 4.0 Technologien kann ein Unternehmen für seine Kunden einen höheren Wert generieren oder seine eigenen Kosten reduzieren? Wie kann mittels Internet of Things (IoT) zukünftig mit Produkten und Diensten Geld verdient werden?

In Unternehmen gilt es, Aufwände zu reduzieren oder neue Erlöse zu generieren. Dafür benötigt es ein Erkennen und Realisieren von Potentialen. Im Zuge der Reifegradbewertung Industrie 4.0 zeigen Experten neue Sichtweisen auf und helfen Potentiale zu identifizieren. Vor allem bei gut etablierten Prozessen kann eine Unterstützung hilfreich sein. Der Fokus liegt auf Unternehmensprozessen, die maßgeblich zur Wertsteigerung beitragen oder die hohen Aufwände verantworten. Potentiale, die diese Bereiche betreffen, werden den benötigten Aufwand zur Realisierung schnell decken.

2 Reifegradmodell Industrie 4.0

Das Reifegradmodell Industrie 4.0 ist ein Vorgehensmodell, das Unternehmen unterstützt Industrie 4.0 besser zu verstehen und zielorientiert anzuwenden. Mit diesem Vorgehensmodell wird der Nutzen von Industrie 4.0 spezifisch für ein Unternehmen aufgezeigt. Durch Bestimmung des IST-Reifegrades erfolgt eine Dokumentation und Bewertung der Ausgangssituation. Diese Analyse der aktuellen Situation unterstützt, Potentiale, entsprechend der Strategie und der Unternehmensziele, zu identifizieren. Konkrete Maßnahmen für ein Unternehmen werden durch die Anwendung des Reifegradmodells vorgeschlagen, um den festgestellten IST-Reifegrad zum anzustrebenden SOLL-Reifegrad zu entwickeln. Die Ergebnisse der Unternehmensbewertungen fließen in eine Benchmark-Datenbank, wodurch sich aktuelle Marktsituationen in den Branchen identifizieren lassen. Ein anonymisierter Vergleich wird ermöglicht, der den eigenen Fortschritt jenem der Branche gegenüberstellt.

2.1 Dimensionen und Kriterien der Reifegradbewertung

Die Reifegradbewertung basiert auf den drei Dimensionen Daten, Intelligenz und Digitale Transformation (siehe Abb. 1). Diese drei Dimensionen sind nicht voneinander unabhängig zu sehen, sie bauen aufeinander auf. Daten in ausreichender Qualität bilden die Grundlage, damit maschinelle Intelligenz überhaupt ermöglicht wird. Ist ausreichend maschinelle Intelligenz implementiert, so können erst dann Prozesse, Aufgaben, Produkte und Dienstleistungen digital transformiert werden. Eine Skala von 0 bis 10 zeigt den Reifegrad je Dimension an. Je höher eine Bewertungszahl ist, desto mehr

Aspekte von Industrie 4.0 sind im Unternehmen umgesetzt. Zur Bemessung der Dimensionen werden sie in Kriterien unterteilt, und diese wiederum in Subkriterien. Die Dimension Daten wird zur Veranschaulichung in diesem Beitrag detailliert beschrieben.

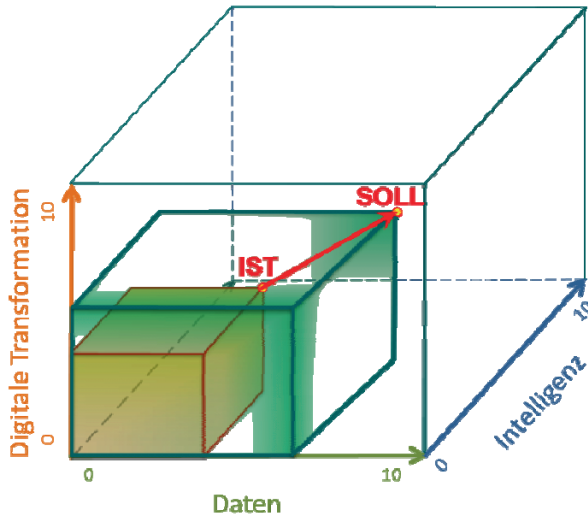


Abb. 1: Dimensionen der Reifegradbewertung Industrie 4.0

Daten bilden die Grundlage und die Hauptressource für Industrie 4.0 und werden als Öl der Zukunft dargestellt [Be14, Jo16]. *Big Data* mit den 5Vs (*Volume, Velocity, Variety, Veracity, Visualization*) beschreibt die Fähigkeit große Datenmengen zu erfassen, zu bearbeiten, zu korrigieren, zu analysieren, zu bewerten, und zuletzt sinnvoll zu nutzen. Weiters kommt in einer vernetzten Welt den Kriterien *Open* und *Security*, eine hohe Bedeutung zu, da vorhandene Daten idealerweise von mehreren genutzt werden und daher auch vor Missbrauch geschützt werden müssen. Die Bewertungsstruktur der Dimension Daten mit ihren Kriterien und Subkriterien ist in folgender Aufzählung zusammengefasst.

- Big Data
 - Volume
 - Velocity
 - Variety
 - Veracity
 - Visualization
- Open
- Security

Unter dem Kriterium *Big Data* befinden sich die 5Vs als Subkriterien. Sie beschreiben die Fähigkeit der Datennutzung von unterschiedlichen Quantitäten und Qualitäten. McAfee und Brynjolfsson schlussfolgern: „*Using big data leads to better predictions, and better predictions yield better decisions.*“ [MB12]

Das Subkriterium *Volume* bewertet die Fähigkeit enorme Datenmengen zu speichern und zu verarbeiten [MB12]. Eine große Menge an Daten zu besitzen und verarbeiten zu können führt jedoch nicht notwendigerweise zu richtiger Informationsgenerierung [RBQ13]. Wichtiger als die Menge an Daten sind Anwendungswissen bzw. Kontextwissen, mit welchem sich die Gefahr reduziert aus Daten etwas Falsches herauszulesen [Jo16].

Unter *Velocity* wird die Geschwindigkeit der Datenerfassung, -übermittlung, -verarbeitung und -bereitstellung verstanden. Die Zeitdauer von der Entstehung der Daten durch ein reales Ereignis bis zur Nutzung kann bewertet werden. Industrie 4.0 Produkte und Geschäftsmodelle funktionieren vielfach auf Basis von Echtzeitdaten. Aus diesem Grund bildet eine echtzeitfähige Erfassung, Aufbereitung und Zurverfügungstellung von Daten die Grundlage für neuartige Produkte und Dienste [MB12].

Variety beschreibt die Mannigfaltigkeit der Daten. Kann ein Unternehmen mit einer großen Datenvielfalt umgehen? Können nur eigene Daten verwendet werden, oder ist das System fähig, unstrukturierte Daten aus Fremdsystemen zu nutzen? Dahinter liegt das Ziel, die Vernetzung verschiedenster Dinge, Systeme und Dienste zu ermöglichen. Benötigte Funktionen sind oft nicht vollständig in einem System integriert. Zum Beispiel kann eine Erfassung der Umwelt (z.B. Feuchtigkeit, Temperatur, Personen im Gefahrenbereich) für eine Anlage von wesentlicher Bedeutung sein. Liefern bereits andere Systeme Daten in ausreichender Qualität, so entfällt der Bedarf an diversen Sensoren für diese Anlage. Die Fähigkeit Daten aus Fremdsystemen richtig zu interpretieren ist für die Anlage eine notwendige Voraussetzung [JS16].

Das Subkriterium *Veracity* bewertet den Wahrheitsgehalt und die Richtigkeit der Daten. Ist dieser bekannt, so können auch unsichere Daten in geeigneter Form weiterverwendet werden. Beispielsweise kann durch Wissen über Unsicherheiten in Forecasts und zur Kostenstruktur eine qualitativ hochwertigere Entscheidung getroffen werden [SG09]. Richtige Daten bzw. das Wissen über den Wahrheitsgehalt ermöglichen neue datenbasierte Produkte oder Dienste. Ist der Wahrheitsgehalt der Daten nicht bekannt, so können darauf basierende Entscheidungen und Geschäftsmodelle nur eingeschränkt genutzt werden [JS16].

Gesammelte Daten sollen zur Generierung von Informationen und Wissen dienen. Wie weit es gelingt, die Wissensgenerierung durch geeignete Darstellung der Daten zu unterstützen, bewertet das Subkriterium *Visualization*. Eine automatische, übersichtliche, personenorientierte und zielorientierte Darstellung der Daten unterstützt den Menschen in der Interpretation der Daten. Darauf aufbauend können datenbasiert Entscheidungen getroffen werden und für ein Unternehmen zu besseren Resultaten führen [MB12].

Das Kriterium *Open* beurteilt einerseits die Offenheit der Daten gegenüber internen und externen Stakeholdern und andererseits die Nutzung von frei zur Verfügung stehenden Daten und Services. Dieses Kriterium basiert auf der Annahme, dass letztendlich nur offene Dienste und Daten eine vernetzte Welt ermöglichen. Bei einer weit verbreiteten hohen Open-Reife kann eine herstellerunabhängige Datendurchgängigkeit erreicht werden. Dies würde zu einer Reduktion von derzeitigen IT-Aufwänden für automatisierten Datenaustausch mit proprietären Systemen führen.

Das Kriterium *Security* bewertet die Informationssicherheitsmaßnahmen und gibt an, inwiefern Daten vor Manipulation und unberechtigten Zugriff geschützt sind. Je bewusster sich ein Unternehmen mit *Security* auseinandersetzt und sich immer näher an das ISO/IEC 27001 Informationsmanagementsystem annähert, desto reifer wird das Unternehmen angesehen [IS13].

Die Dimension **Intelligenz** bewertet Dinge oder Systeme (*Enabler*), welche auf Basis von Daten maschinell intelligent handeln (*Nutzung maschineller Intelligenz*). Mittels maschineller Intelligenz verringert sich die Notwendigkeit des menschlichen Eingreifens in der Abwicklung von Aufgaben. Auf der einen Seite benötigt es *Enabler*, welche durch ihre Eigenschaften maschinell intelligentes Handeln ermöglichen. In einer Produktion können beispielsweise Transportbehälter, Produkte, Maschinen solche *Enabler* darstellen. Nur wenn ausreichend *Enabler* in einem Unternehmen vorhanden sind, kann maschinell intelligentes Handeln durch diese realisiert werden. Das Kriterium *Nutzung maschineller Intelligenz* betrachtet dieses maschinell-intelligente Handeln. Dabei wird unterschieden, ob *Enabler* lediglich eingesetzt werden, um Standardaufgaben abzarbeiten oder ob Systeme vorhanden sind, die mit Menschen als auch mit anderen Systemen kooperieren, selbstständig arbeiten, reagieren und lernen [JSB16]. Die Bewertungsstruktur der Dimension Intelligenz mit ihren Kriterien und Subkriterien ist in folgender Aufzählung zusammengefasst.

- Enabler
 - Identifizierbarkeit
 - Lokalisierbarkeit
 - Connectivity-Grad
 - Speicherfähigkeit
 - Sensorausstattung
 - Ausstattung Aktoren
 - Rechenfähigkeit
- Nutzung maschineller Intelligenz
 - Selbstständigkeit
 - Reaktionsfähigkeit

- Anpassungsfähigkeit
- Kooperationsfähigkeit
- Mensch-Maschine Symbiose

Für eine hohe Reife der dritten Dimension **Digitale Transformation** werden ausreichend Daten und maschinelle Intelligenz benötigt, um eine *Transformation* herbeizuführen. Ziel ist es, Materielles durch Digitales zu ersetzen oder möglichst viel Wertschöpfung in der virtuellen Welt zu erreichen. Um digitale Transformation im Unternehmen zu ermöglichen, benötigt es Menschen, die an den Dimensionen Daten, Intelligenz und Digitale Transformation aktiv arbeiten und damit den Fortschritt ermöglichen [Br15, JS16]. Die Bewertungsstruktur der Dimension Digitale Transformation mit ihren Kriterien und Subkriterien ist in folgender Aufzählung zusammengefasst.

- Mitarbeiter
 - Mitarbeiter (Können & Wollen)
 - Führung (Dürfen)
- Transformation
 - Durchgehende digitale Modellbildung
 - Simulation & Optimierung
 - Ersetzen Materielles durch Digitales

Intelligenz.Enabler.Identifizierbarkeit			
Langfristig, technologieunabhängig	Maschinell automatisch identifizierbar	10	RFID, Bluetooth, 3D Scanner
		9	
		8	
		7	
		6	
	Maschinell lesbare Identität mit menschlicher Aktivität	5	Barcode, QR-Code, OCR
		4	
		3	
		2	
	Eindeutige Identitätsmerkmale, welche nicht maschinell gelesen werden	1	Identnummer
Keine eindeutige Identität	0	Keine Identnummer	

Abb. 2: Referenztablelle Intelligenz/Enabler/Identifizierbarkeit

Jedes Subkriterium wird mit Hilfe einer Referenztablelle von 0 bis 10 bewertet. In Abb. 2 ist beispielhaft die Referenztablelle für die Dimension Intelligenz, Kriterium Enabler und

Subkriterium Identifizierbarkeit dargestellt. Dieses Subkriterium behandelt die maschinell eindeutige Identität des Enablers. Taucht dieses Gerät im betrachteten System nur einmal auf und kann es immer wieder unter diesem Merkmal von einem Menschen identifiziert werden, so wird 1 Punkt vergeben. Erfolgt eine vollautomatische maschinelle Identifizierung des Gerätes ohne menschliches Zutun, so werden 10 Punkte vergeben.

2.2 Die Anwendung des Reifegradmodells

Vor Applikation des Reifegradmodells, erfolgt eine strategiegeleitete Auswahl an Applikationsfeldern, die den Untersuchungsgegenstand definieren. Ein Applikationsfeld könnte zum Beispiel ein Fertigungsbereich, die externe Werkzeug- und Anlagenbeschaffung oder auch der Kundenauftragsabwicklungsprozess sein. In einem Startworkshop werden auf breiter Basis Strategie, Visionen und Ziele des Unternehmens besprochen. Fragen zum Unternehmen ermöglichen die Einschränkung auf wesentliche Applikationsfelder. Wie und in welchen Bereichen kann Industrie 4.0 die derzeitige Organisation stärken? Wodurch können Wettbewerbsvorteile erzielt werden? Welche Prozesse, Produkte und Dienstleistungen sind für den Unternehmenserfolg wichtig? Eine Diskussion über Ziele in jedem Applikationsfeld fokussiert die Anwendung der Reifegradbewertung.

Für jedes Applikationsfelder werden nun Träger definiert. In Bezug auf obige Bewertungskriterien ist ein Träger die größte Einheit, die einer eindeutigen Bewertung zugeführt werden kann. Ein Träger kann ein Prozess, eine Aufgabe oder ein Ding sein. Träger stellen wesentliche Funktionen des Applikationsfeldes dar. Im Startworkshop definierte Ziele werden durch die Träger wesentlich beeinflusst. Mögliche sinnvolle Träger des Applikationsfeldes Fertigung könnten die Maschinenbelegungsplanung, die Personaleinsatzplanung, Transportbehälter oder Maschinen sein. Die ausgewählten Träger werden mittels der obigen Kriterien in allen drei Dimensionen bewertet.

Die zur Bewertung notwendigen Informationen werden in strukturierten Interviews eingeholt. Die Interviewpartner beschreiben den aktuellen Zustand des Trägers und die Auswertung von Prozessbeschreibungen, Daten und Beobachtungen liefern zusätzliche Informationen. Potentiale können nun identifiziert und dokumentiert werden. Auf Basis dieser Analyse erfolgt die IST-Reifegradbestimmung je Träger. Die endgültige Bewertung auf der Skala von 0 bis 10 wird durch einen Vergleich mit der Referenztafel je Subkriterium unterstützt (siehe Abb. 2). Eine mit der Punktevergabe einhergehende Begründung wird ebenfalls je Subkriterium dokumentiert. Das Bewertungsteam erarbeitet aufgrund der Strategie, der definierten Ziele und der IST-Analyse mit den festgestellten Potentialen eine SOLL-Reife, die für das Unternehmen erstrebenswert ist. Wirtschaftliche als auch technische Restriktionen werden in der SOLL-Reife berücksichtigt. Ein SOLL-Reifegrad von 10 wird selten angestrebt werden. Grund dafür ist, dass zur Erreichung der Unternehmensstrategie Ressourcen bestmöglich einzusetzen sind. Ein Reifegrad von 10 stellt jedoch das Höchstmaß an technologischen

Fortschritt zu Industrie 4.0 dar. Bei großer Differenz zwischen SOLL und IST wird die erforderliche Investition nicht mehr rentabel sein.

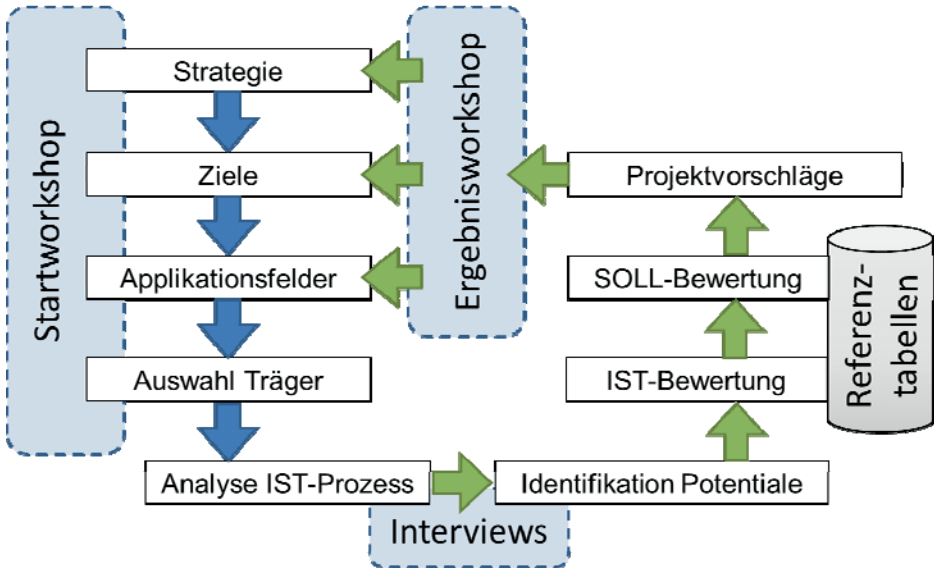


Abb. 3: Vorgehensmodell der Reifegradbewertung

Zur Erreichung der SOLL-Reife, werden Projektvorschläge erarbeitet. Diese Projektvorschläge basieren einerseits auf identifizierten Potentialen und andererseits auf den anzustrebenden SOLL-Reifegrad. Bewertungen und Projektvorschläge werden abschließend im Ergebnisworkshop mit dem Unternehmen diskutiert und konkrete Umsetzungsschritte fixiert. Das gesamte Vorgehensmodell ist in Abb. 3 dargestellt. Die Anwendung des Reifegradmodells Industrie 4.0 kann auch iterativ erfolgen. Nach Abschluss der ersten Bewertungen können weitere Applikationsfelder der Untersuchung unterzogen werden. Eine Bewertung aller Applikationsfelder und der Einsatz von Industrie 4.0 im gesamten Unternehmen werden damit ermöglicht. Ebenso besteht die Möglichkeit die Erreichung der SOLL-Positionen zu späteren Zeitpunkten zu überprüfen.

2.3 Vergleich und Abgrenzung von Reifegradbewertungen

Für die Messung von Industrie 4.0 Anwendungen existieren auch andere Vorgehensweisen. In diesem Kapitel wird das vorgestellte Reifegradmodell mit einigen ausgewählten Methoden und Werkzeugen verglichen und es werden die Unterschiede in den verschiedenen Ansätzen herausgearbeitet.

VDMA Werkzeugkasten

Im Leitfaden Industrie 4.0 der VDMA wird ein Werkzeugkasten Industrie 4.0 für die Umsetzung in Unternehmen vorgestellt. Der Leitfaden soll Unternehmen des deutschen Maschinen- und Anlagenbaus unterstützen, kreative Ideen für neue Geschäftsmodelle zu entwickeln. Das Vorgehen ist in einen mehrstufigen Prozess unterteilt. Mittels Analyse wird der Ist-Stand zu Kompetenzen des Unternehmens und zur Marktpositionierung erhoben. Auf Basis dessen werden mittels Kreativitätstechniken und mit Hilfe des Werkzeugkastens Industrie 4.0 neue Ideen generiert und darauf basierend Geschäftsmodelle konzeptioniert. Diese gilt es zu bewerten und umzusetzen. Der Werkzeugkasten Industrie 4.0 dient dabei als Unterstützung bei der Ideengenerierung. Der Werkzeugkasten soll *„Visionen und grundlegende Technologien auf Entwicklungsstufen reduzieren, die eine schrittweise Umsetzung innovativer Ideen [...] ermöglicht.“* [Le15] Die Applikation des Werkzeugkastens kann für die Bereiche Produkte und Produktion erfolgen. Eine Unterscheidung verschiedener Visionen und Technologien erfolgt durch die Anwendungsbereiche. Eine Einstufung von eins bis fünf stellt die Entwicklung je Anwendungsbereich dar. Je höher die Stufe, desto mehr Industrie 4.0 Visionen und Technologien werden im Unternehmen umgesetzt. Durch diese Einstufung wird die Reife eines Unternehmens je Anwendungsbereich dargestellt. Beteiligte Personen bekommen durch die Einstufung ein Verständnis, wie sich Industrie 4.0 auf ihre Produkte und die Produktion im Maschinen- und Anlagenbau auswirken kann. Davon ausgehend wird überlegt, wie höhere Stufen zu erreichen sind und welche Geschäftsmodelle davon abgeleitet werden können. [Le15]

Wesentliche Unterschiede zum Reifegradmodell Industrie 4.0 sind:

- **Einschränkung in der Applikation:** Der VDMA Werkzeugkasten fokussiert sich auf Unternehmen des Maschinen- und Anlagenbaus. Dabei sind Werkzeugkästen für Produkte und für die Produktion vorhanden. Das Reifegradmodell beinhaltet diese Einschränkungen nicht. Diese Breite an Anwendungsmöglichkeiten erfordert jedoch tieferes Wissen zu den Kriterien des Reifegradmodells.
- **Visuelles Werkzeug zur Ideenfindung:** Der Werkzeugkasten eignet sich zur Ideenfindung für neue Geschäftsmodelle. Über eine zuvor durchgeführte Kompetenz- und Marktanalyse werden mittels Werkzeugkasten Ideen generiert, die zu neuen Geschäftsmodellen und neuen Strategien führen. Das Reifegradmodell wird zur Umsetzung vorhandener Strategien als auch zur Findung neuer Strategien oder Geschäftsmodelle eingesetzt. Vorhandene Kompetenzen werden dabei als Subkriterium betrachtet. Der Aufbau des Werkzeugkastens hat Ähnlichkeiten zu den Referenztabellen des Reifegradmodells. Beide stellen in ihrer tabellarischen Form je Kriterien Reifegrade dar. Je besser ein Unternehmen die Kriterien erfüllt, desto höher ist die Industrie 4.0 Reife.

VDMA Impuls Studie

Von der VDMA wurde eine Industrie 4.0 Readiness-Studie durchgeführt. Ziele der Studie waren, Verständnis für Industrie 4.0 zu schaffen, den Reifegrad für die Unternehmen zu erheben und notwendigen Veränderungsbedarf in den unternehmerischen Rahmenbedingungen aufzuzeigen. In der Studie wurde mit Hilfe von sechs Dimensionen (Strategie und Organisation, Smart Factory, Smart Operations, Smart Products, Data-driven Services und Mitarbeiter) die Reife eines Unternehmens bestimmt. Die Unternehmensbewertung erfolgte über Online-Selbsttests. Teilnehmende Unternehmen gaben in einem Online-Fragebogen Antworten zu gestellten Fragen. Davon ließ sich der Reifegrad je Dimension ableiten. Die Bewertung erfolgte in sechs Stufen, die durch Mindestanforderungen je Stufe definiert wurden. Je Dimension kann ein Unternehmen urteilen, welche Schritte notwendig sind, um eine höhere Industrie 4.0-Reife zu erreichen [Li15].

Folgender Unterschied zeigt sich zum Reifegradmodell Industrie 4.0:

- Erhebungsmethode: Ein Onlinefragebogen bietet die Möglichkeit, viele Unternehmen mit geringem Aufwand zu erreichen und zu bewerten. Die Ergebnisse stellen Selbsteinschätzungen der Unternehmen dar. Das Resultat ist ein generelles Branchenverständnis. Tipps zur Verbesserung und zur spezifischen Anwendung von Industrie 4.0 sind allgemein formuliert. Das Reifegradmodell bewertet individualisiert und liefert strategiegeleitete Verbesserungsvorschläge. Den individualisierten und strategiegeleiteten Ergebnissen stehen höhere Aufwände in der Erhebung gegenüber. Inhaltlich decken sich zu weiten Teilen die Bewertungskriterien der VDMA Impuls Studie mit den Subkriterien des Reifegradmodells.

Readiness Modell nach Bildstein und Seidelmann

Bildstein und Seidelmann beschreiben ein weiteres Industrie 4.0 Readiness Modell. Mittels Analyse relevanter Geschäftsprozesse wird der Ist-Stand erhoben und Verbesserungspotentiale werden identifiziert. *„Im Kern der Überlegungen steht hierbei die Frage, ob Industrie 4.0-Konzepte, wie beispielsweise Vernetzung in Echtzeit, smarte Objekte, Software Services in der Cloud oder cyber-physische Produktionssysteme dabei helfen können, diese Prozesse zu verbessern.“* [BS14] Die zuerst analysierten Ist-Prozesse werden mit Industrie 4.0 Standardanwendungsfällen verglichen. Standardanwendungsfälle stellen allgemeine Unternehmensprozesse dar, in denen Industrie 4.0-Konzepte bereits erfolgreich umgesetzt wurden. Durch Vergleich der Ist-Prozesse mit Standardanwendungsfällen lassen sich Soll-Prozesse und Umsetzungsvoraussetzungen erarbeiten. Durch die Differenz zwischen Ist und Soll wird der benötigte Aufwand zur Industrie 4.0 Implementierung bemessen. Ein zu erarbeitendes Kosten-Nutzen-Verhältnis und iteratives Vorgehen bilden die Basis für die Umsetzungsplanung und eine unternehmensweite Migration. [BS14]

Ein Unterschied zum Reifegradmodell Industrie 4.0 liegt in der Prozessorientierung:

- Die Anwendung des Readiness Modells orientiert sich am Prozessmanagement. Wesentliche Prozesse werden ausgewählt und diese mit Best Practices zu Industrie 4.0 („Standardanwendungsfälle“) gegenübergestellt. Eine Prozessoptimierung mittels neuen Lösungsansätzen aus Industrie 4.0 ist das Resultat. Das Reifegradmodell bewertet Prozesse, Abläufe und Dinge auf Ebene der Träger. In der Bewertungsstruktur der Referenztablelle werden je Subkriterium die technologischen Reifegrade aufgelistet. Es erfolgt keine Fokussierung auf Best Practices, sondern auf technologische Möglichkeiten. Der SOLL-Reifegrad zeigt den wirtschaftlich sinnvollen Reifegrad innerhalb der technologischen Möglichkeiten auf.

2.4 Praxisbeispiel

Das Reifegradmodell Industrie 4.0 wurde bereits in einer Pilotphase getestet. Ein Praxisbeispiel aus der Pilotphase dient zur Veranschaulichung, wie das Reifegradmodell in Unternehmen eingesetzt wird. Es stellt allerdings nur einen Auszug aus einer umfassenden Reifegradbewertung dar.

Bei diesem Beispiel handelt es sich um einen Tier 1 Automobilzulieferanten. Ziel der Reifegradanwendung war es, die Planung der Halbfabrikats-Fertigung zu optimieren und den Halbfabrikat-Bestand zu reduzieren. Als Applikationsfeld wurde die Produktion einer Baugruppe ausgewählt. Unter mehreren identifizierten Trägern werden in diesem Beitrag die Träger Produktionsplanung-Halbfabrikate, Rüsten und Halbfabrikats-Fertigung beschrieben.



Abb. 4: Fertigungsablauf einer Produktgruppe aus einem Unternehmensbeispiel

Abb. 4 zeigt den Fertigungsablauf einer Produktgruppe vom Rohmaterial bis zur Auslieferung. In der betrachteten Produktgruppe gibt es vier Varianten an Rohmaterialien. Die Presse fertigt in Wochenlosgröße. Da für jede Produktgruppe ein eigenes Pressenwerkzeug existiert, ist beim Wechsel der Produktgruppe auch ein Werkzeugwechsel erforderlich. Vier Rohmaterialvarianten führen zu vier verschiedenen Varianten im Lager der Halbfabrikate. Die Finalisierung erfolgt auftragsbezogen und die Kunden erwarten sich eine Anlieferung in Just in Sequenz (JIS). Das Unternehmen finalisiert und verpackt in der geforderten Reihenfolge, wobei das Unternehmen die

Informationen dazu drei Tage vor Auslieferung erhält. Entsprechend der Information wird in Losgröße 1 die entsprechende Halbfabrikat-Variante aus dem Lager entnommen, finalisiert und für die Auslieferung verpackt.

Die drei ausgewählten Träger wurden wie folgt analysiert:

- **Produktionsplanung-Halbfabrikate:** Auf Basis des Verbrauches werden Losgrößen für die Fertigung definiert. Jede Woche wird ein Los einmal aufgelegt. Die Fertigungsreihenfolge der Produktgruppen wird von der Produktionsplanung definiert, die Reihenfolge innerhalb der Produktgruppen (Rohmaterialvarianten) wird vom Produktionspersonal bestimmt. Produktionsrückmeldungen erfolgen retrograd über ein ERP-System bei vollen Behältern oder durch Eingabe der Mitarbeiter (Rüstzeiten, Fertigungszeit, Gutteile, Ausschuss).
- **Rüsten:** Werkzeugwechsel erfolgen durch das Produktionspersonal manuell. Bei elektrischer Verbindung des Werkzeuges identifiziert die Presse das Werkzeug selbstständig. Vordefinierte Anlagenparameter werden auf Basis des Werkzeuges automatisch geladen. Materialwechsel erfolgen durch das Produktionspersonal ebenfalls manuell. In der Anlage wird dabei das Rohmaterial durch ein neues ersetzt. Feinjustierungen, die einen wesentlichen Zeitanteil bei den Rüstvorgängen darstellen, erfolgen in Verantwortung des Produktionspersonals. Aufgrund unterschiedlicher Eigenschaften des Rohmaterials (Materialvariante, Materialzusammensetzung, Lagerung) und der Umgebung (Jahreszeit, Temperatur, Luftfeuchtigkeit) ist diese Feinjustierung erforderlich. Die Feinjustierung erfolgt mittels Veränderung bestimmter Anlagenparameter, die durch das Produktionspersonal nicht dokumentiert werden muss. Eine laufende Überprüfung der Halbfabrikate und Nachjustierung der Anlagenparameter ist notwendig, um Gutteile zu produzieren. Ein wesentlicher Anteil an Ausschuss wird durch die unterschiedlichen Eigenschaften im Rohmaterial und die Umgebung verursacht.
- **Halbfabrikats-Fertigung:** Die Presse wird vom Produktionspersonal mit Rohmaterial versorgt. Entsprechend der Anlagenparameter erfolgt die Fertigung. Sensoren überwachen die Anlage und das Werkzeug. Die Steuerungseinheit kontrolliert die Anlage und dient als Schnittstelle zum Personal. Ein Hubzähler erfasst Anzahl und Zeitpunkte der Hübe innerhalb der Presse. Daten zum Auftrag, zur Umgebung, zum Rohmaterial, zum Prozess (Maschine und Werkzeug z.B. Einstellparameter, Sensordaten ausgenommen Hubzähler) werden nur teilweise erfasst und nicht dokumentiert. Daten zur Halbfabrikat-Qualität sind nicht systematisch vorhanden.

Im Rahmen der IST-Bewertung wurden für jeden Träger die Kriterien des Reifegradmodells Industrie 4.0 angewandt. Beispielhaft findet sich hier ein Auszug davon: Der Träger Halbfabrikats-Fertigung erhält beim Subkriterium Volume (Dimension Daten, Kriterium Big Data) 2 Punkte, die sich durch den Vergleich mit der Referenztablelle ergeben. Die Begründung dafür ist, dass wenige Daten (Größenordnung

Megabyte) elektronisch vorhanden sind und verarbeitet werden. Für die Bewertung des Subkriteriums Identifizierbarkeit (vergleiche Abb. 2) benötigt es Enabler. In dem vorgestellten Fertigungsablauf stellen die Anlage (Presse) und das Werkzeug solche Enabler dar. Das Werkzeug ist maschinell automatisch identifizierbar. Aus diesem Grund erhält das Werkzeug 10 Punkte beim Subkriterium Identifizierbarkeit. Die Anlage hat eindeutige Identitätsmerkmale aufgrund einer eindeutigen Bezeichnung im Unternehmen. Die Anlage ist jedoch nicht maschinell identifizierbar. Aus diesem Grund erhält die Anlage 1 Punkt beim Subkriterium Identifizierbarkeit.

In der SOLL-Bewertung wurden die Optimierung der Halbfabrikat-Planung und die Reduktion des Halbfabrikat-Bestandes als Ziele des Unternehmens berücksichtigt. Eine Möglichkeit zur Zielerreichung ist die Reduktion der Rüstzeiten. Dadurch werden kleinere Produktionslose ermöglicht, die einen geringeren Bestand im oben dargestellten Halbfabrikat-Lager zur Folge haben. Ebenso kann eine Reduktion der Schwankungen in den Rüstzeiten zu besseren Planungsergebnissen führen. Da in diesem Unternehmen die Feinjustierung einen wesentlichen Anteil der Rüstzeit darstellt und Schwankungen der Rüstzeiten verursacht, trägt eine Automatisierung der Feinjustierung zum Erreichen der unternehmerischen Ziele bei. Eine Identifikation wesentlicher Parameter und ein entsprechender Algorithmus sind für die automatisierte Feinjustierung erforderlich. Die Speicherung und Verarbeitung von Daten zum Auftrag, zur Umgebung, zum Rohmaterial, zum Prozess und zur Halbfabrikat-Qualität bilden das Fundament dafür. Aus wirtschaftlichen Gründen ist die Erhebung aller Daten zum derzeitigen Zeitpunkt jedoch noch nicht sinnvoll. Die Speicherung und Verarbeitung ausgewählter Material-, Umgebungs- und Prozessdaten in Verbindung mit Daten zur Halbfabrikat-Qualität scheinen aus Unternehmenssicht das größte Potential zu haben. Die SOLL-Bewertung zum Subkriterium Volume liegt in der Folge entsprechend der Referenztafel bei 5 Punkten.

Projektvorschlag: Der Projektvorschlag für das Unternehmen ist die Erhebung und die Analyse ausgewählter Daten und Parameter. Daten und Parameter, die die Feinjustierung wesentlich beeinflussen, sollen identifiziert werden. Ein daraus zu entwickelnder Algorithmus soll eine Automatisierung der Feinjustierung ermöglichen.

Wird dieser Projektvorschlag im Unternehmen umgesetzt, so können im nächsten Schritt weitere Daten gespeichert und verarbeitet werden, um weitere Optimierungen zu erreichen. Eine iterative Anwendung des Reifegradmodells und laufende Verbesserungen ermöglichen ebenso eine auftragsbezogene Halbfabrikats-Fertigung. Das Entfallen der Produktionsplanung-Halbfabrikate und die Reduktion von Bestand wären die Folgen.

3 Zusammenfassung

Das vorgestellte Industrie 4.0 Reifegradmodell unterstützt Unternehmen in der Bestimmung des IST-Reifegrades und des SOLL-Reifegrades. Die resultierenden

Verbesserungsmaßnahmen versetzen das Unternehmen in die Lage, sich Richtung SOLL-Reifegrad strategiegeleitet zu entwickeln. Zur Reifegradbestimmung werden die Dimensionen Daten, Intelligenz und Digitale Transformation herangezogen. Die Dimension Daten beschreibt dabei die Fähigkeit Daten richtig sowie zielorientiert zu verwenden. Unter Intelligenz werden die Fähigkeit und die Nutzung von maschinell-intelligenten Handlungen bewertet. Die Digitale Transformation betrachtet den Anteil an Wertschöpfung, der in der digitalen Welt erfolgt.

Ein Vergleich zu anderen Reifegradbewertungen zeigt, dass das vorgestellte Industrie 4.0 Reifegradmodell sich nicht auf bestimmte Branchen oder Anwendungsgebiete spezialisiert. Strukturiertes Vorgehen und die Bewertung in den Subkriterien der drei Dimensionen ermöglicht ein breites Anwendungsgebiet. Um die Bewertungsstruktur anzuwenden, wird jedoch Fachwissen zu jedem einzelnen Kriterium benötigt. Als Resultat erhalten teilnehmende Unternehmen applikationsbezogene und strategiegeleitete Projektvorschläge zur Umsetzung von Industrie 4.0.

Literaturverzeichnis

- [AL05] Allmendinger, G.; Lombreglia, R.: Four Strategies for the Age of Smart Services. Harvard Business Review. Oct2005, Vol. 83 Issue 10, p131-145, 2005.
- [Be14] Berinato, S.: With Big Data Comes Big Responsibility. Harvard Business Review. Nov2014, Vol. 92 Issue 11, p100-104, 2014.
- [BS14] Bildstein, A.; Seidelman, J.: Industrie 4.0-Readiness: Migration zur Industrie 4.0-Fertigung. In (Bauernhansl, T.; ten Hompel, M.; Vogel-Heuer, B., Hrsg.): Industrie 4.0 in Produktion, Automatisierung und Logistik, Springer Fachmedien Wiesbaden, S. 581-597, 2014.
- [Br15] Brauckmann, O.: Smart Production, Springer-Verlag Berlin Heidelberg, 2015.
- [IS13] ISO/IEC 27001 - Information security management, <http://www.iso.org>, Stand: 01.10.2013.
- [Jo16] Jodlbauer, H.: Die Datenspinne. Im Netz von Google, Amazon, Alibaba und Co, Leykam, Graz, 2016.
- [JS16] Jodlbauer, H.; Strasser, S.: Geschäftsmodellinnovation basierend auf Industrie 4.0 sichern den zukünftigen Erfolg der Unternehmen. In (Gleich, R.; Losbichler, H.; Zierhofer, R., Hrsg.): Der Controlling Berater. Band 43, Haufe Gruppe, S. 109-122, 2016.
- [JSB16] Jodlbauer, H.; Schagerl, M.; Brunner, M.: Industrie 4.0 versus Automatisierung. ZWF 04/2016, S. 222-224, 2016.
- [Le15] Leitfaden Industrie 4.0. Orientierungshilfe zur Einführung in den Mittelstand, VDMA Forum Industrie, Frankfurt a. M., 2015.
- [Li15] Lichtblau, K. et. al.: Industrie 4.0-Readiness, IMPULS/VDMA, Aachen & Köln, 2015

- [MB12] McAfee, A; Brynjolfsson, E.: Big Data: The Management Revolution. Harvard Business Review. Oct2012, Vol. 90 Issue 10, p60-68, 2012.
- [PH14] Porter, M. E.; Heppelmann, J. E.: How Smart, Connected Products Are Transforming Competition. Harvard Business Review. Nov2014, Vol. 92 Issue 11, S. 64-88, 2014.
- [RBQ13] Ross, J. W.; Beath, C. M.; Quaadgras, A.: You May Not Need Big Data After All. Harvard Business Review. Dec2013, Vol. 91 Issue 12, p90-98, 2013.
- [SG09] Sanders, N.; Graman, G.: Quantifying costs of forecast errors: A case study of the warehouse environment. Omega 2009; Vol. 37; p116-125, 2009.

14. Workshop Automotive Software Engineering

14. Workshop Automotive Software Engineering

Marcel Baunach¹ und Leandro B. Ribeiro²

Der 14. Workshop “Automotive Software Engineering” fand am 27. September 2016 im Rahmen der INFORMATIK, der Jahrestagung der Gesellschaft für Informatik, in Klagenfurt, Österreich statt. Im Vordergrund stand die Diskussion aktueller Herausforderungen und Lösungsansätze im Wandel zur Smart Mobility mit vernetzten und autonomen Fahrzeugen. Teilnehmer aus Industrie und Wissenschaft beschäftigten sich mit Methoden, Techniken und Werkzeugen zur Entwicklung eingebetteter Systeme für moderne und zukünftige Fahrzeuge.

Neben dem Ziel der GI Fachgruppe “Automotive Software Engineering”, mit diesem Fachgruppentreffen eine Plattform für den inspirierenden Gedankenaustausch anzubieten, wurden im Vorfeld zusätzlich 11 Fachbeiträge aus Theorie und Praxis eingereicht. Nach intensiver Begutachtung durch Experten aus dem Programmkomitee konnten letztlich 8 Beiträge für die Publikation im Tagungsband sowie für Fachvorträge auf dem Workshop angenommen werden (Akzeptanzrate: 73%).

Diese Beiträge stehen repräsentativ für die brennenden Themen der Automobilindustrie: Zukünftige Fahrerassistenzsysteme (ADAS), Fahrzeugnetzwerke, Cloud-Integration, individualisierte Produktion sowie Navigation und Betriebsstrategien für gänzlich autonomes Fahren erfordern hochzuverlässige Softwaresysteme und Entwicklungsprozesse. Safety, Security, Echtzeiteigenschaften sowie Wartbarkeit sind daher zentrale Aspekte in sämtlichen Beiträgen.

Mit der AVL List GmbH als Unternehmen für die Entwicklung von Antriebssystemen sowie zugehöriger Software, Simulations- und Prüftechnik für Fahrzeuge konnte zudem ein kompetenter Industriepartner für den Workshop gefunden werden, der die besondere Relevanz dieser Themen durch die Keynote weiter verdeutlicht:

“Accelerated lifecycle validation of complex networked automotive control units”
(Michael Paulweber, AVL List GmbH)

Unser besonderer Dank gilt den Teilnehmern unseres Workshops, insbesondere den Autoren der Beiträge, dem Programmkomitee für die Erstellung der Gutachten, der AVL List GmbH als Industriepartner, der Fachgruppe Automotive Software Engineering, sowie dem Organisationsteam der INFORMATIK 2016.

September 2016

Marcel Baunach
Leandro Ribeiro

¹ TU Graz, Institut für Technische Informatik, Inffeldgasse 16, A-8010 Graz, baunach@tugraz.at

² TU Graz, Institut für Technische Informatik, Inffeldgasse 16, A-8010 Graz, lbataribeiro@tugraz.at

Organisation

Die Organisation des Workshops erfolgte durch die Embedded Automotive Systems Group (EAS) des Instituts für Technische Informatik der Technischen Universität Graz in enger Abstimmung mit der GI Fachgruppe Automotive Software Engineering (ASE) sowie unserem Industriepartner:

- Embedded Automotive Systems Group (EAS):
Prof. Dr. Marcel Baunach, TU Graz (baunach@tugraz.at)
Dipl.-Ing. Leandro B. Ribeiro, TU Graz (lbatistariibeiro@tugraz.at)
Homepage: <http://www.iti.tugraz.at/eas/>
- Fachgruppe Automotive Software Engineering (ASE):
Homepage: <http://fg-ase.gi.de>
- Industriepartner 2016:
AVL List GmbH, Graz
Homepage: <http://www.avl.com>

Programmkomitee

- Dr. Christian Allmann (Audi AG)
- Prof. Dr. Marcel Baunach (Technische Universität Graz)
- Prof. Dr. Dr. h.c. Manfred Broy (Technische Universität München)
- Dr. Stefan Bunzel (Continental AG)
- Dr. Mirko Conrad (samoconsult GmbH)
- Prof. Dr. Werner Damm (Universität Oldenburg)
- Dr. Heiko Dörr (Model Engineering Solutions GmbH)
- Bernd Frielingsdorf (Ford Werke GmbH)
- Prof. Dr. Steffen Helke (Technische Universität Cottbus)
- Prof. Dr. Bernhard Hohlfeld (Technische Universität Dresden)
- Prof. Dr. Stefan Jähnichen (Technische Universität Berlin)
- Ralf Kalmar (Fraunhofer IESE)
- Prof. Dr. Stefan Kowalewski (RWTH Aachen)
- Prof. Dr. Reiner Kriesten (Hochschule Karlsruhe)
- Prof. Dr. Thomas Kropf (Robert Bosch GmbH)
- Peter Manhart (Daimler AG)
- Prof. Dr. Peter Pickel (John Deere European Technology Innovation Center)
- Prof. Dr. Klaus Pohl (Universität Duisburg-Essen)
- Prof. Dr. Wolfgang Pree (Universität Salzburg)

- Peter Priller (AVL List GmbH)
- Dr. Alexandre Saad (BMW Group)
- Prof. Dr. Ina Schaefer (Technische Universität Braunschweig)
- Prof. Dr. Jörn Schneider (Hochschule Trier)
- Dr. Sebastian Siegl (Audi Electronics Venture GmbH)
- Claus Stellwag (Elektrobit Automotive GmbH)
- Dr. Christoph Stoermer (ETAS GmbH)
- Dr. Reinhard Stolle (BMW Car IT GmbH)
- Prof. Dr. Michael Uelschen (Fachhochschule Osnabrück)
- Dr. Marcel Wille (Volkswagen AG)
- Dr. Fabian Wolf (Volkswagen AG)
- Dr. Dirk Ziegenbein (Robert Bosch GmbH)

Towards Identifying Spurious Paths in Combined Simulink/Stateflow Models¹

Marcus Mikulcak², Thomas Göthel³, Paula Herber² and Sabine Glesner²

Abstract:

MATLAB/Simulink and its state machine design toolbox Stateflow are widely-used industrial tools for the development of complex embedded systems. Due to the dynamic as well as heterogeneous nature of models that contain both Simulink and Stateflow components, their analysis poses a difficult challenge. This paper outlines an approach to relate the semantics of both Simulink and Stateflow and how to use it to perform an information flow analysis on a combined Simulink and Stateflow model. In the first step, we analyze the Stateflow automata and generate *timed output traces* for arbitrary inputs. In the second step, we use an existing *timed path condition* extraction algorithm for the Simulink components to identify conditions for information flow on paths of interest. Finally, we analyze whether the compiled sets of timed path conditions are contained in the timed output traces that we derive by using a novel trace notation for Stateflow automata. This approach makes it possible to safely rule out the existence of information flow on specific paths through a model. Further, it presents a starting point to reason about non-interference between model parts, compliance with security policies as well as the generation of feasible, efficient test cases.

Keywords: Embedded Systems, MATLAB, Simulink, Stateflow, Path Conditions, Information Flow

1 Introduction

In the area of safety-critical embedded software, such as in the automotive and aerospace domain, programming errors can lead to disastrous and, if occurring, often fatal accidents. At the same time, the complexity of such systems has increased dramatically over recent years. To cope with the steadily increasing complexity, current design processes rely more and more on models. One of the most widely-used tools for model-based design is MATLAB/Simulink [Ma15] by MathWorks, which supports the graphical design and simulation of time-continuous as well as time-discrete systems using block diagrams. To additionally support the design of state machine-based embedded controllers and model them in conjunction with these dynamical systems, Stateflow [Ma14], an extension to Simulink, is widely used in industrial contexts.

However, due to the complexity and the dynamic character of the developed models, the analysis of a given model is a difficult challenge, in particular if timing aspects are

¹ Funded by the German Federal Ministry of Education and Research as part of the research project CISM0

² Technische Universität Berlin, Software and Embedded Systems Engineering Group, Ernst-Reuter-Platz 7, 10587 Berlin, Germany, marcus.mikulcak@tu-berlin.de

³ Service and Software Engineering Group, Universität Potsdam, August-Bebel-Straße 89, 14482 Potsdam, thomas.goethel@uni-potsdam.de

considered. At the same time, knowledge about the existence of certain paths, the conditions under which they are executed and how an embedded Stateflow controller influences their behavior is highly desirable.

Using knowledge about the behavior of Stateflow controllers and their influence on conditionally executed components of the model makes it possible to identify interference and non-interference between model parts and, thus, to reason about compliance with security policies. Furthermore, knowledge about the existence of paths provides a basis for generating feasible, efficient test cases for quality assurance and for calculating more precise *Worst Case Execution Time* (WCET) bounds for Simulink/Stateflow models.

Additionally, the information gathered using our approach can be used as a first step to identifying areas of low dynamic coupling in Simulink/Stateflow models and to subsequently extract abstract boundaries between these areas. This makes it possible to safely break down large models into smaller components which in turn can be analyzed using existing quality assurance methods, such as model checking, thereby avoiding the state space explosion problem. Together with our industrial partner, *Model Engineering Solutions GmbH* (MES) [ME16], we identified a number of models from the automotive sector that match the modeling style our approach is able to analyze and that therefore benefit from such a separation, such as electronic gearbox system controllers and interior and exterior lighting controllers. These models are developed using one or multiple Stateflow automata controlling data flow sections designed in Simulink and due to their size and dynamic complexity, are difficult to analyze without prior separation.

In this paper, we illustrate our three-step approach to identify spurious paths in combined Simulink/Stateflow models using a small running example. First, we extract *timed output traces* from an embedded Stateflow controller. These extracted traces form an over-approximation of the automaton behavior as we aim to support arbitrary input signals. We then use a previously published algorithm to extract timed path conditions from dynamic data-flow components developed in Simulink and prepare them for analysis. Finally, we present a comparison algorithm for timed output traces and timed path conditions to identify overlap between them. If there is an overlap, the conditionally executed paths under analysis exist. If not, they are identified as infeasible and will never be executed in the model.

2 Preliminaries

2.1 Path Conditions

In general, *path conditions* [Ki76] describe necessary conditions for paths to be executed. In [HSS08], path conditions are used to capture all paths where information might flow from a source to a target. Consider the example given in Listing 1. There, a variable `src` is assigned to an element of the array `a`. Inside the `if` statement, the variable `tgt` is assigned with an element of the array. A static analysis of the possible paths of this program would detect potential information flow from `src` to `tgt`. However, a more precise computation of path conditions takes data and control dependencies into account.

```

1 a[i + 3] = src;
2 if ((i > 10) && (j < 5))
3     tgt = a[2 * j - 42];

```

Listing 1: Path condition example

When traversing the path of information flow between lines 1 and 3, the necessary conditions are derived from the array indices and the `if`-condition and can be expressed as:

$$PC(1 \rightarrow 3) = \exists i, j ((i > 10) \wedge (j < 5) \wedge (i + 3 = 2j - 42))$$

The result is that there is no assignment to `i` and `j` such that the equation holds. In contrast to the result from a static analysis, this means that no information flow is possible. This simple example demonstrates that a path condition-based analysis is able to offer more precise answers about which conditions have to hold for information to flow.

2.2 Information Flow Analysis

The protection of confidentiality of information inside a software system is a long-standing and increasingly important problem in the areas of general computing as well as embedded systems. Protecting not only the data itself but also the integrity of the functionality that produces and handles data is a goal of software non-interference policies [GM82]. Such policies, based on the assignment of security levels to data elements, describe rules between which levels information flow is allowed or forbidden [SM03]. When aiming at assuring *confidentiality*, data is prohibited to flow *to* inappropriate locations, while in the context of *integrity*, data is prohibited to flow *from* inappropriate sources. As non-interference refers to the absence of information flow, it ensures both confidentiality and integrity.

2.3 MATLAB/Simulink

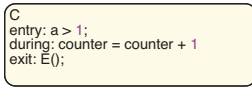
MATLAB/Simulink [Ma15] is an add-on to the MATLAB IDE by MathWorks that enables graphical modeling and simulation of reactive systems. In its data-flow oriented notation, Simulink employs *blocks* which are connected using *signals*. Additionally, each block and signal is assigned a set of *parameters*.

Simulation of Simulink models is performed using *solvers*, which compute the output of each block according to its semantics. *Variable step* solvers aim at automatically finding a simulation step size for each block in the model to achieve a level of precision set by the model developer. *Fixed step* solvers use a fixed simulation step size at the expense of precision while increasing performance. The former class of solvers is commonly used for hybrid or purely time-continuous systems, while the latter is used for time-discrete models.

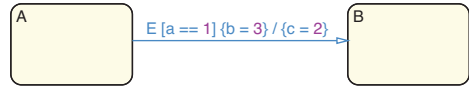
2.4 Stateflow

Stateflow [Ma14] is a further add-on to the MATLAB IDE, especially to Simulink, and gives the designer the possibility to integrate decision logic based on state machines and flow

charts into a Simulink model. Stateflow provides complex modeling styles incorporating multiple state, event and transition types as well as an execution semantic not only dependent on the structure and annotations of the model, but also on its layout. The underlying semantics resemble that of statecharts [Ha87] as they utilize hierarchical and concurrent state structures, junctions splitting transitions and actions as part of state definitions.



(a) Example of a State in Stateflow



(b) Example of a Transition in Stateflow

2.4.1 States

States form the basic building block of the decision logic implemented in Stateflow. An example of a state is shown in Figure 1a. When the execution enters a state, a set of *actions* modeled by the designer takes place. *When* these actions are performed is determined by the *action type*: entry, during and exit. While the sets of entry and exit actions occur only once every time the state is active, the during actions are performed with every simulation step and are therefore dependent on the selected solver of the Simulink and Stateflow model.

2.4.2 Transitions

To design the state change logic of an automaton, Stateflow states are connected via *transitions*. Similar to states, it is possible to add guards, trigger events and actions to transitions. Figure 1b shows an example transition containing all three mentioned semantical elements. After receiving event *E* as the *exit* action of state *C*, it is evaluated if the current transition is valid, which in turn evaluates the guard condition $a == 1$ and, if true, perform the transition. As soon as this guard is evaluated to true, the *guard action* $b = 3$ is executed. Before finally marking state *B* as active, however, the *transition action* $c = 2$ is executed.

2.5 Timed Path Conditions in MATLAB/Simulink

We have previously presented our approach to compute *timed path conditions* in MATLAB/Simulink [Mi15], which we extend upon in this paper. The main idea of timed path conditions is to transfer the concept of path conditions from sequential programming languages like C to the Simulink modeling language. The main challenges are to take both data and control dependencies into account and to cope with timing dependencies. Data dependencies are simply resolved by following signal lines where each connection corresponds to a direct dependency. Control dependencies are more difficult to compute as they introduce conditional dependencies which are locally resolved. To overcome this problem, control flow dependencies are propagated backwards through the model to the input signals. With that, it is possible to decide whether a certain path actually exists on a

very fine-grained level. For both data and control dependencies, timing dependencies are taken into account. An output might only depend on an input at certain points of time, and routing policies might even take advantage of timing delays to make sure that two signals can never interfere. Timing dependencies are taken into account by introducing *time slices* and timed path conditions are expressed with respect to relative time slices.

3 Path Existence in Simulink/Stateflow Models

In the following section, we describe the main contribution of this paper. First, we present an overview of our approach and an example model to illustrate the problem we aim at solving. Subsequently, we introduce the assumptions we impose on the models that our approach in its current proof-of-concept form is able to analyze. Finally, using our running example, we present our timed trace format and how to extract them from Stateflow automata, the generation of timed path conditions and the comparison algorithm for both trace sets.

3.1 General Approach

The heterogeneous nature of models containing both Simulink and Stateflow parts makes their analysis hard. The main challenge is to reconcile the inherently different semantics of Simulink and Stateflow, and in particular their timing. The semantics of Simulink is defined by the simulation semantics of the solver, where the functionality and timing depend on the simulation step size. The semantics of Stateflow is defined by evaluation rules that determine which transition fires in each step, whereby a step is made whenever one of the input signals changes. The main idea of our approach for identifying spurious paths in combined Simulink/Stateflow models is to relate a Stateflow Controller with its surrounding Simulink model using the concept of timed traces. As the identification of spurious paths in Simulink models [Mi15] is achieved by a backwards analysis through the model, the incorporation of Stateflow components requires us to analyze the effects of a Stateflow component on its (signal and variable) outputs for arbitrary inputs. Note that a

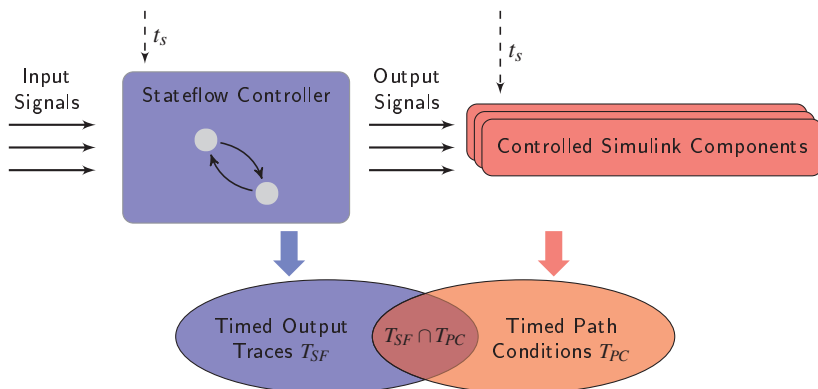


Figure 2: Approach to identify spurious paths in combined Simulink/Stateflow models

Stateflow state machine is connected to the surrounding model via Simulink signals that form the variables used inside guards. Variables modified in state or transition actions inside Stateflow state machines form their output signals and are routed to the Simulink model. The evaluation of a Stateflow automaton is performed whenever one of the input signals to the automaton changes its value. Then, its state is reevaluated and a single possible transition is taken. We can therefore define a minimal time interval between every change in the output of a Stateflow automaton. Under the assumption of a uniform sample time throughout the model, it is equal to the simulation step size t_s . This relation between the discretely-timed solver of the Simulink model and the evaluation of the Stateflow automaton makes it possible to relate both semantics. As shown in Figure 2, we assume that a Stateflow Controller is embedded into a Simulink model and has an effect on some of its components. Our approach to identify spurious paths in combined Simulink /Stateflow is threefold: First, we analyze the behavior of an embedded Stateflow automaton under functional and timing aspects. Specifically, we extract timed traces for outputs signals influencing the control flow of the surrounding Simulink model by unrolling the automaton until a stable state is reached. The output of this step is a set of possible traces T_{SF} for an output signal of interest that forms the basis for analysis of path existence in the combined model. Second, we utilize an existing algorithm Section 2.5 to extract timed path conditions for paths of interest from the Simulink model. Along these paths that carry, e.g., safety-critical information whose flow needs to be analyzed, the conditions for information flow as well as their timing are gathered and expressed as traces T_{PC} . Third, we compare the traces derived from Simulink path conditions with the extracted timed output traces of the Stateflow automaton. In this final step, we analyze whether the sets of timed path conditions can be generated by the Stateflow automaton, i. e., if they overlap with the timed output traces ($T_{SF} \cap T_{PC}$). We explain these three steps using a running example, presented in the following section.

Running Example To illustrate our approach, we use a model of a shared buffer originally presented in [WHW10], shown in Figure 3. There, information of the security levels *public* and *confidential* is fed into a shared buffer, implemented as a Mem block. According to the operation mode, confidential (mode 1) or public (mode 2), information is saved in the buffer and passed to the corresponding output. The mode of operation to access the buffer is determined by the Stateflow model in Figure 4. Write access to the buffer on the corresponding security level is requested using the `write_request` signal if the state machine is in its `waiting` state. To return to the initial state, a read request with the corresponding level is used. Data output from the automaton to the surrounding Simulink components is realized using the *entry actions* inside the states. Whenever the automaton *enters* a state, the action is performed. In our running example, the output signal `current_state` is written. While the automaton is in the `public` or `confidential` mode, the Simulink switches are set to only allow for input and output of public or confidential data, respectively.

Figures 5a to 5c illustrate that due to the timing behavior of the Simulink components in the system, information flow from the confidential input to the public output is indeed possible. Shown there, a change in operation mode from `confidential` to `public` at time $t = 5$ results in a leak of confidential information stored in the buffer to the public data output.

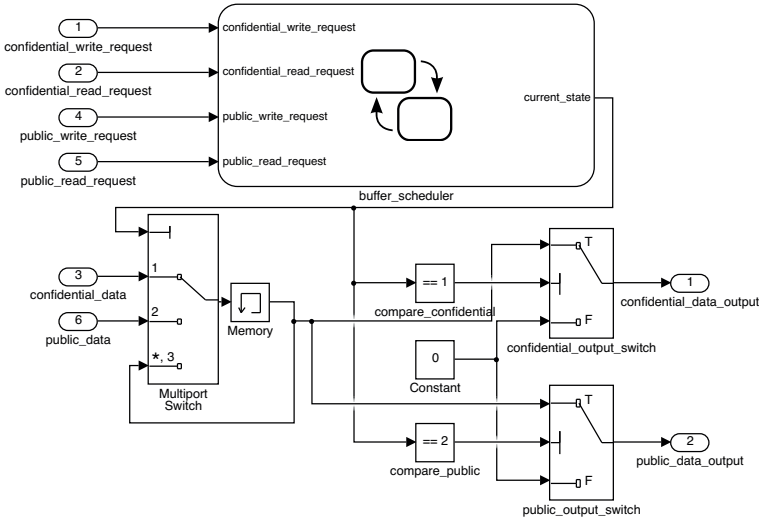


Figure 3: Model of a shared buffer in Simulink

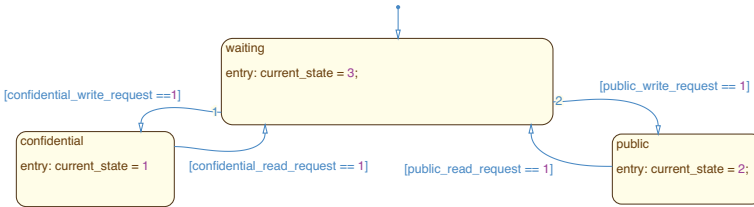


Figure 4: The Stateflow automaton of our running example

Timed Traces As a basis for comparison of the outputs of a Stateflow automaton and the sets of timed path conditions extracted using the algorithm described in Section 2.5, we define a data format to capture all required information. We call this a *timed trace*. For a given signal s_1 we capture its valuations v_1, v_2, \dots, v_n and the number of simulation steps that pass between the occurrences of these valuations. The general formats of the traces are:

$$s_1 = \{v_1 \xrightarrow{nt_s} v_2 \xrightarrow{kt_s} \dots\}, \quad s_2 = \{v_1 \xrightarrow{\lfloor nt_s \rfloor} v_2 \xrightarrow{\lfloor kt_s \rfloor} \dots\}$$

There, s_1 and s_2 denote signals in the Simulink model or outputs of the Simulink automaton while n and k are factors of the simulation step size t_s . Shown in s_2 is the notation of *at least* n and k simulation steps passing between the two valuations of s_2 .

Assumptions As we present a proof-of-concept implementation for the evaluation of path existence in combined Simulink/Stateflow models, both the Stateflow as well as Simulink components have to fulfill certain requirements. For the latter, the requirements are listed in [Mi15]. For the former, we as of now assume a basic form of Stateflow automaton consisting of only states and transitions without any hierarchical grouping or the use of

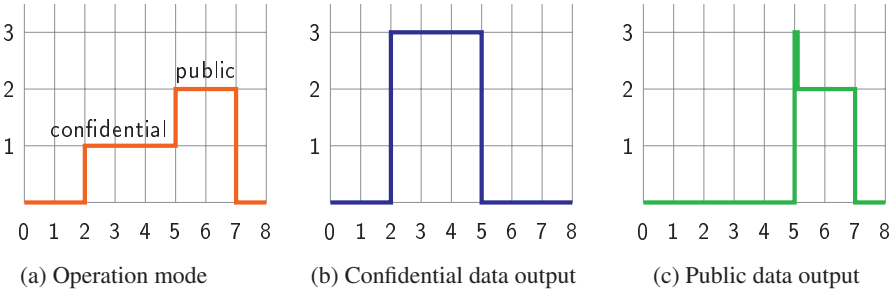


Figure 5: Timing of the shared buffer signals

junctions. However, we are confident that the presented approach can be extended to support a wide range of Stateflow modeling features and styles which we describe in Section 6.

3.2 Generating Timed Traces from Timed Path Conditions

In the first step of our approach, we use the algorithm summarized in Section 2.5 to generate timed path conditions from the dynamic data-flow components. The solved constraints in our example model result in the following path conditions:

$$C(p(o_2^t, i_3^{t-1})) = \{(s_{\text{control}}^t == 2), s_{\text{control}}^{t-1} == 1\}$$

This set shows the conditions for information to flow through the path starting at data input i_3 for confidential data and leading to the data output port i_2 for public data. As can be seen there, two conditions are imposed on the signal denoting the current state of the system (either `confidential` or `public`). The conditions denote that information flows from the public input to the confidential output whenever s_{control} is first set to 1 (at an arbitrary time $t - 1$) and then set to 2 in the next time step (at time t). Note that $t - 1$ in the algorithm denotes a time that is a single simulation step earlier than t , i. e., time $t - t_s$. We can therefore rearrange the condition set into a timed trace $\tau \in T_{\text{PC}}$ as follows:

$$\tau(s_{\text{control}}) = \{1 \xrightarrow{t_s} 2\}$$

In our example, the trace $\tau(s_{\text{control}})$ defines the conditions for a security policy violation to occur in the Simulink components, i. e., confidential data is led to the public data output. Therefore, if we are able to identify $\tau(s_{\text{control}})$ in the output traces of the Stateflow automaton, the security violation occurs in the combined system and vice versa.

3.3 Establishing a Control Relation

To elevate the timed path conditions and to generate a global statement about the existence of traces, a connection between the signals controlling the execution of the path and the output signals A of the Stateflow automaton needs to be identified. The timed path condition

extraction algorithm aims to unify the signals controlling all control flow elements such as `Switch` and `If-Action-Subsystem`. Using this information, our approach follows these signals backwards through the Simulink model until it encounters a Stateflow automaton.

Running Example In our example, this relation is established by following the signal path such that:

$$s_{\text{control}} = o_{\text{current_state}}, \quad \text{with } o_{\text{current_state}} \in O$$

If a connection is found, we extract output traces from the Stateflow automaton, specifically for the control signals of the paths. This process is illustrated in the following section.

3.4 Extracting Timed Stateflow Automata Traces

In this section, we present our approach to extract timed output traces from Stateflow automata. As explained above, a prerequisite to this analysis is an established connection between the control signals of the path through the Simulink components, i. e., the path under analysis, and the Stateflow controller setting these signals. The inputs to this step are the traces generated from the timed path conditions and their connection to the automaton.

In this step, we especially consider the possibilities of modeling states and transitions in Stateflow. When a state is entered, a set of *actions* takes place, such as the modification of output signals of the automaton or the triggering of *events*. When these actions are performed is determined by the *action type*: `entry`, `during` and `exit`. Depending on this choice by the programmer, the timing behavior and the frequency of the modifications and triggers changes. While the sets of `entry` and `exit` actions occur only once every time the state is active, the `during` actions are performed with every simulation step and are therefore dependent on the selected solver of the Simulink and Stateflow model. To design the state change logic of an automaton, Stateflow states are connected via *transitions*. Similar to states, it is possible to add guards, trigger events and actions to transitions.

Our analysis, seen in Algorithm 1, computes an over-approximation of all traces that can be produced by a given Stateflow model by performing a depth-first search on its state space. It starts with the initial node of a system and a set of empty traces $T(o)$ for all output signals of the Stateflow automaton. When processing a node, the function `nodeAnalysis` first analyzes the set of actions assigned to the current state. If any of the actions contains an assignment to o_{control} , the value of this assignment is appended to $\tau(o_{\text{control}})$. Additionally, our algorithm internally saves the origin of the assignment, i. e., the state and the action type containing the assignment. If the trace was not previously empty, the time step calculation, which we present below, is initiated. After the time interval to the previously appended value is calculated and added to the timed trace, or if the trace was previously empty, the algorithm continues by adding all outgoing transitions to a work list. For every transition on the work list, a similar analysis is performed by the function `transitionAnalysis`. As we assume arbitrary inputs to the system, we do not need to include guards into our analysis and consider them to be *true* in all cases. Similarly, we assume that transition trigger events are always active. When taking a transition from the worklist, its condition and transition

Function extractTraces

```

Input : output signal of interest  $o_{\text{control}}$ 
         worklist of transitions  $t_{\text{work}}$ 
         visited transitions  $t_{\text{visited}}$ 
         state  $v$ 

// Does current state contain assignment to control signal?
// If so, write them into  $\tau(o_{\text{control}})$ 
nodeAnalysis ( $v, o_{\text{control}}, \tau(o_{\text{control}})$ );
// If state does not have outgoing transitions, end recursion
if  $|t_{\text{out}}(v)| == 0$  then return;
for  $t : t_{\text{out}}(v)$  do
    // return when encountering previously visited transitions
    if  $t \in t_{\text{visited}}$  then return;
    else  $t_{\text{work}}.add(t)$ ;
end
 $t_{\text{next}} = \text{pop}(t_{\text{work}})$ ;
// Search next transition for occurrence of control signal assignment
transitionAnalysis ( $t_{\text{next}}, o_{\text{control}}, \tau(o_{\text{control}})$ );
 $v_{\text{next}} = \text{dst}(t_{\text{next}})$ ;
// work on next state
extractTraces ( $o_{\text{control}}, t_{\text{work}}, t_{\text{visited}}, v_{\text{next}}$ );

```

Algorithm 1: Extracting traces from Stateflow automata

actions are searched for assignments to o_{control} . If none is found, the algorithm continues by visiting the next state and performing the analyses explained above. For every outgoing transition on the initial node, the algorithm is run recursively while maintaining the worklist of unvisited transitions, possibly creating an over-approximation for states included in loops. However, this behavior is deliberate to facilitate a trace comparison across loops. Whenever a previously visited transition is encountered or when a state does not have any outgoing transitions, the base case of the recursion is reached and a set of traces for outputs of interest is returned. Internally, an ordered list of visited states is maintained to facilitate the time step calculation. This way, starting with the initial state, all states are visited and all transitions are taken while traces for all variables of interest are constructed.

Time Step Calculation As explained above, we aim at relating the timing of changes to the outputs of a Stateflow automaton to each other. We are interested in the *minimal* time interval between such changes, as these define the influence of the automaton on the surrounding model when analyzing information flow. This calculation is triggered whenever an assignment to the control signals under analysis is recognized in the actions of a state or a transition. In this case, the current state and the list of visited states, i. e., the path of the current recursion are used to perform a backwards analysis of the minimal time interval t_{diff} between the control signal assignments. The calculation is performed by iterating backwards over the path of visited states and transitions while observing the transition guards. If a guard does not use temporal operators, t_{diff} is increased by $1t_s$ as the reevaluation of Stateflow automata is triggered every simulation step of the surrounding Simulink model

with an interval of t_s . If it does use temporal condition operators, further analyses are necessary. These operators take one of two forms: *event-based* and *absolute-time*. Event-based operators, such as *after*, *at*, *every* and *before* appear in the form of, e. g., *after*(5, E). In this example, the operator evaluates to true after at least 5 occurrences of event E have taken place since activation of the Stateflow chart. Absolute-time operators work in a similar fashion. However, they do not count the number of occurrences of an event but evaluate the simulation time of the model. For our implementation, we currently only support the extraction of the event-based notation *after*(n , E) from which the number of event occurrences n is extracted and added to t_{diff} . As we are extracting a minimal time interval, we do not need to evaluate the precise timing of event occurrences. Therefore, due to the Stateflow automaton evaluation interval, n occurrences of an event take at least nt_s .

Running Example In the case of our motivating example, the (over-approximated) set of traces T_{SF} extracted by our algorithm is the following:

$$T(o_{current_state}) = \left\{ \{3 \xrightarrow{[t_s]} 1 \xrightarrow{[t_s]} 3 \xrightarrow{[t_s]} 2 \xrightarrow{[t_s]} 3\}, \{3 \xrightarrow{[t_s]} 2 \xrightarrow{[t_s]} 3 \xrightarrow{[t_s]} 1 \xrightarrow{[t_s]} 3\} \right\}$$

3.5 Comparing Trace Sets

In the last step, we aim to establish a relation between the two sets of traces, i. e., the set of path conditions expressed as timed traces and the set of possible output traces of the Stateflow automaton. In order to show that information flow on the path under analysis can never occur, the relation $T_{SF} \cap T_{PC} = \emptyset$ has to hold. If it does not, then at least one of the path condition traces is contained in the set of possible output traces and information flow can therefore occur. Note that the *contains* relation does not only look for matching values in traces but also analyzes timing similarities. The comparison analyzes each pair of the cross product of both trace sets. In the first step, the values of τ_{PC} are compared in order with the values of its corresponding pair element τ_{SF} . Note that the values of τ_{PC} only need to appear in τ_{SF} in their correct order. The latter can contain additional assignments. If all values in the correct order are found, the second step identifies the timing relation of both traces. Starting with the initial value in τ_{PC} (1^\diamond) and its corresponding value in τ_{SF} (1^*), all time steps in τ_{SF} are added until the next value of τ_{PC} (2^\mp) is encountered (2^*). If this number is smaller or equal than the time step in τ_{PC} , the timing of the value update matches and the analysis continues until each time step is matched. If the values in their correct order cannot be found or their respective timing relations do not match, τ_{SF} does not contain τ_{PC} .

Running Example For our running example, our approach identifies the following sets:

$$\begin{aligned}
 T_{SF} &= \{ \tau(o_{current_state})_1, \tau(o_{current_state})_2 \} \\
 &= \left\{ \{3 \xrightarrow{[t_s]} 1^* \xrightarrow{[t_s]} 3 \xrightarrow{[t_s]} 2^* \xrightarrow{[t_s]} 3\}, \{3 \xrightarrow{[t_s]} 2 \xrightarrow{[t_s]} 3 \xrightarrow{[t_s]} 1 \xrightarrow{[t_s]} 3\} \right\} \\
 T_{PC} = \tau_{PC} &= \left\{ \{1^\diamond \xrightarrow{t_s} 2^\mp\} \right\}
 \end{aligned}$$

When analyzing the value and timing relation of these sets $T_{SF} \cap T_{PC}$ using our algorithm, the result is that their intersection is empty. This is due to the fact that although the values of τ_{PC} appear in order in $\tau(o_{\text{current_state}})_1$, their timing does not match. In the controller, a change of its output `current_state` from 1 to 2 takes place only with $t_{\text{diff}} = 2$. The violation of the security policy that is present in the timing behavior of the Simulink components of the model can therefore never occur in the combined Simulink/Stateflow model.

4 Evaluation

To evaluate our approach, we have implemented it in Java. It uses a modified version of the path condition extraction architecture described in [Mi15]. We use the Java-based constraint solving system JaCoP [KS13] instead of the external Gecode solver to increase performance and to facilitate its integration into our tool. We have extended it with a parser for Stateflow automata and the analysis algorithms described above. Our implementation is accessible via an Eclipse plug-in. Table 1 shows the results of our analysis of the running

Automaton Size	Time		
	Extraction T_{PC}	Extraction T_{SF}	Comparison
3	58 ms	14 ms	4 ms
10		39 ms	12 ms
100		258 ms	113 ms
1,000		2,183 ms	946 ms

Table 1: Evaluation results

example as well as of generated Stateflow automata. Note that we only adapted the size of the controller by randomly adding states and unguarded transitions while maintaining the size of the Simulink components of the system as well as the size of T_{PC} . In the first row, the results of with an automaton size of 3 states can be seen. The following rows show the performance of our algorithm for automaton sizes of 10, 100 and 1,000 as well as the algorithm run times necessary for a complete analysis. The extraction of T_{PC} includes both the extraction of sets of timed path conditions as well as the constraint solving process by JaCoP. As can be seen in the table, the extraction of both sets as well as the comparison in the case of our motivating example is performed in under 100ms. For the generated Stateflow automata, the extraction and the comparison scale linearly.

5 Related Work

Extensive work has been done in the area of translating subsets of Simulink/Stateflow models into formals language with well-defined semantics, especially Lustre and the graphical modeling suite SCADE, in order to perform model checking on the translated systems [Sc04, Mi05, WHW10]. However, as these approaches rely on a translation of models into a target language using functional and timing semantics different to those of Simulink and Stateflow, properties of the original systems are lost and the timing of models

cannot be analyzed precisely. Further, the translation process for industrial-sized models poses strong restrictions on their design and is therefore often not applicable [WL15].

Only few authors have addressed the problem of formalizing the behavior of Stateflow automata. Although we also do not provide a complete formal semantics, timed traces partially formalize some of the Stateflow semantics. In [HR04] and [Ha05], operational and denotational semantics for a subset of Stateflow are presented. While they represent a wide range of specific functionality of the controller semantics, they do not consider the timing of automata and the connection with surrounding Simulink models. Due to the similarities between Stateflow and Statecharts, we analyze previous formalization efforts for such models, most notably [Ha87]. However, these similarities are merely superficial, as the underlying solver for Stateflow automata works in a purely sequential fashion, and their semantic differences make an elevation of the this approach infeasible.

6 Conclusion

In this paper, we have presented an approach to extract *timed output traces* from Stateflow automata and compare them to *timed path conditions* extracted from the Simulink parts of a combined Simulink/Stateflow model. The resulting conditions can be used to, e.g., reason about the existence of paths. We have argued that for embedded software models only a combined analysis of both the behavior of the controller (in Stateflow), and the dynamic data-flow components (in Simulink) can cope with the complex timing behavior. Using the example of a shared buffer for confidential and public data using an embedded controller, we have demonstrated the usability of our approach in an *Information Flow Analysis* (IFA). Thereby, we have shown that although timed path conditions alone detect a security policy violation, it is recognized as spurious when also analyzing the controller component.

To increase the precision of our approach, we aim to extend the support for further Stateflow design constructs, such as hierarchical states, events, junctions, further state action types and temporal condition operators. We plan to incorporate these elements into a Stateflow formalization to generate timed traces for single Stateflow output signals and timing relations between multiple output signals to support non-unified control signals as explained in Section 3.3. Furthermore, we aim at supporting additional design patterns for the integration of Stateflow controllers into Simulink models, such as scheduling of individual Simulink components using an automaton or the integration of Simulink functions into controllers. Finally, an interesting artifact in the quality assurance process of embedded software is the trace of input data to the Stateflow automaton that leads to the output of a certain trace, which in turn executes a path in the Simulink components of the model. We plan to extend our algorithm to support the generation of such traces.

References

- [GM82] Goguen, Joseph A; Meseguer, José: Security Policies and Security Models. In: IEEE Symposium on Security and Privacy. 1982.

- [Ha87] Harel, David: Statecharts: A Visual Formalism for Complex Systems. In: Science of Computer Programming. volume 8. Elsevier, 1987.
- [Ha05] Hamon, Grégoire: A Denotational Semantics for Stateflow. In: ACM International Conference on Embedded Software. ACM, 2005.
- [HR04] Hamon, Grégoire; Rushby, John: An Operational Semantics for Stateflow. In: Fundamental Approaches to Software Engineering. Springer, 2004.
- [HSS08] Hammer, Christian; Schaade, Rüdiger; Snelting, Gregor: Static Path Conditions for Java. In: ACM SIGPLAN Workshop on Programming Languages and Analysis for Security. ACM, 2008.
- [Ki76] King, James C: Symbolic Execution and Program Testing. Communications of the ACM, 1976.
- [KS13] Kuchinski, Krzysztof; Szymanek, Radoslaw: Jacop - Java Constraint Programming Solver. In: International Conference on Principles and Practice of Constraint Programming. unpublished, 2013.
- [Ma14] MathWorks Stateflow. www.mathworks.com/products/stateflow/, Accessed: 06/2016.
- [Ma15] MathWorks MATLAB/Simulink. www.mathworks.com/products/simulink, Accessed: 06/2016.
- [ME16] Model Engineering Solutions GmbH. www.model-engineers.com, Accessed: 06/2016.
- [Mi05] Miller, Steven; Anderson, Elise; Wagner, Lucas; Whalen, Michael; Heimdahl, Matts: Formal verification of flight critical software. In: Proceedings of the AIAA Guidance, Navigation and Control Conference and Exhibit. 2005.
- [Mi15] Mikulcak, Marcus; Herber, Paula; Göthel, Thomas; Glesner, Sabine: Timed Path Conditions in MATLAB/Simulink. In: International Embedded Systems Symposium. Springer, 2015.
- [Sc04] Scaife, Norman; Sofronis, Christos; Caspi, Paul; Tripakis, Stavros; Maraninchi, Florence: Defining and Translating a Safe Subset of Simulink/Stateflow into Lustre. In: International Conference on Embedded Software. ACM, 2004.
- [SM03] Sabelfeld, Andrei; Myers, Andrew C: Language-based Information-Flow Security. IEEE Journal on Selected Areas in Communications, 2003.
- [WHW10] Whalen, Michael W.; Hardin, David; Wagner, Lucas G.: Model Checking Information Flow. In: Design and Verification of Microprocessor Systems for High-Assurance Applications. Springer US, 2010.
- [WL15] Walde, Georg; Luckner, Robert: Automatic Translation of Complex Flight Control Systems from Simulink/Stateflow to SCADE - An Experience Report. Technical report, Deutsches Zentrum für Luft- und Raumfahrt, 2015.

Real World Evaluation of a Novel Security Testing Environment for Vehicular Control Units via CAN Networks

Jürgen Wurzinger¹, Peter Priller¹, Aleš Kolar¹ and Markus Nager¹

Abstract: Due to the introduction of networked entertainment and safety features in modern cars, vehicular communication systems are no closed networks anymore. As research demonstrated, based on these new attack surfaces existing safety and functional tests cannot fully satisfy security needs of modern cars. With this work we introduce an application called CAN Communication Tester (CAN-CT) which addresses this problem for the Controller Area Network (CAN).

CAN-CT is an application that makes use of hacking attacks to systematically inject, replay and invalidate messages. We show that we can successfully spoof messages, suppress all communication on the bus and transition electronic control units (ECUs) into error states. CAN-CT is capable of learning from the traffic on the bus and executes targeted attacks. By attacking an ECU the same way a hacker would, we are able to examine the implementation of protection mechanisms in an intuitive yet effective way. Furthermore this approach allows us to test the proper working of possible attack detection techniques as well.

Using Hardware-in-the-Loop Systems (HiLs) we tested CAN-CT with actual state of the art ECUs. We revealed vulnerabilities like lacks in the plausibility checking of messages or completely omitted examinations of the frame structure. Those flaws opened up serious threats to hackers. In exploiting those vulnerabilities we showed that we were able to take over message IDs owned by another ECU and hence achieve targeted manipulations.

Revealing weaknesses in real-world ECUs allowed us to demonstrate the applicability and impact of applications like CAN-CT for automotive systems. Our results highlight the need for security tests to complement traditional testing environments.

Keywords: CAN; Controller Area Network; Security Testing; Automotive; Car Hacking; Test Framework; Hardware in the Loop; ECU; Electronic Control Unit; Penetration Testing;

1 Introduction

Modern cars consist of up to 100 electronic control units (ECUs) communicating over various vehicular networks [Ch09]. The upcoming interconnection of state of the art vehicles with their environment introduced advances regarding safety and entertainment, while simultaneously opening up a large number of new attack surfaces as well. The prevalent security by isolation approach is broken and the internal communication systems are no closed networks anymore [Ko14].

Even though manufacturers use strong security measures to protect a vehicle's I/O channels from hackers, there is no 100 percent secure system. This fact was demonstrated by

¹ AVL LIST GmbH | Hans-List-Platz 1, A-8020 Graz

many researchers, who were able to exploit almost every interface of a car with its outside world [Ch11, VM15]. This circumstance stresses the need that besides the protection of those interfaces also the communication on internal vehicular networks has to be secured. This is especially true for the Controller Area Network (CAN), as it is the most widely used communication system for automotive applications [La13] and connects the most critical components in a vehicle.

For previous generations of vehicles, security measures and possible hacking attacks were hardly a concern. CAN communication was just tested for safety and functional requirements but not for malicious behavior. Nowadays, stronger emphasis is and will be put on more secure CAN communication [NSL13]. Yet due to the growing complexity of modern ECUs and in-vehicle communications it is very likely that flaws in the implementation of such safety and security measures exist. This implies that we need a testing environment to verify the proper working of these techniques. An important way to test implemented security measures is to simulate hacking attacks to complement traditional safety and functional tests.

With this work we introduce a tool named CAN Communication Tester (CAN-CT) that addresses this requirement for the CAN protocol in its current version 2.0. Making use of various hacking attack types allows us to detect implementation flaws of the CAN protocol's and partially also of the application layer's communication mechanisms. We are able to falsify messages, suppress all communication on the bus and transition ECUs into error states. Thus, the contribution of this work is two-fold. On the one hand we introduce an intuitive security testing approach with novel CAN specific attacks (see Section 4.2) and on the other hand we provide a detailed evaluation of the applicability of such a testing environment and its impact for real-world scenarios and commercial ECUs.

2 Related Work

In the last few years extensive academic as well as non-academic research was done to determine the vulnerability of vehicles against hacking attacks. Koscher et al. [Ko10] were one of the first to test this vulnerability comprehensively in a real-world scenario. Similar to Koscher et al. Hoppe, Kiltz, and Dittmann [HKD11] did a thorough examination of security threats to CAN buses. Besides a theoretical analysis of vulnerabilities and possible countermeasures, various attacks on the CAN bus were executed in a self-made testing environment. Yet, unlike our approach, no systematic bus off, spoofing of recurring messages, or Denial of Service (DoS) attacks (see Section 4.2) were carried out.

Other than these approaches we are not focusing on the exploitation of protection mechanisms or developing countermeasures; instead we are using CAN attacks as the basis for our systematical tests to detect possible vulnerabilities of the CAN communication.

Our attack-based testing approach differs from previously introduced functional or model-based testing frameworks by its *systematic and comprehensive execution of hacking attacks*. Even though model-based testing approaches like Marinescu's et al. [Ma14] validate

internal functions in detail, they still cannot sufficiently simulate the malicious intelligence of a hacker. We believe that this is also true for other traditional testing methodologies.

Bayer et al. [Ba14] as well as Talebi [Ta14] outlined concepts similar to our's on a more theoretical level. While Bayer et al. discussed the topic of a systematical execution of automotive security evaluations and penetration tests, Talebi studied the field of security evaluations by looking at the CAN bus as a fully simulated environment using the simulation software CANoe². However, to the best of our knowledge, a fully working prototype of such a testing environment has not been presented so far.

In October 2015, Jenik presented a black box testing approach based on a concept called exhaustive fuzzing [Je15]. Similar to this approach also CAN-CT can be used for black box testing; however, instead of fuzzing we learn from the traffic on the bus and execute sophisticated and targeted attacks. This allows us to examine not only the correct communication of ECUs but also to test potential attack detection mechanisms.

As a consequence, CAN-CT can be understood as a tool that helps to verify secure implementation of CAN communication. In a final version security testing environments like the CAN-CT shall be used in combination with security evaluation methodologies like the security framework for the automotive domain of Glas et al. [GGV14] or the "automotive security evaluation assurance levels" (ASEAL) introduced by Bayer et al. [Ba14].

3 Current CAN Communication Protection Mechanisms

Vehicular communication networks followed a security by isolation approach. As, however, this isolation is not existent anymore, new security techniques are needed. So far no generally accepted security measures for CAN communication exist. As car hacking research demonstrated, security through obscurity remains the prevailing approach to protect cars [VM15, MV13, Ko14].

Authentication mechanisms are only implemented for diagnostic purposes, like flashing the ECU [Ko14, MV13, VM15]. Koscher et al. state, in their examinations, that these mechanisms use weak keys, a fixed challenge (seed) and are both just 16 bits long [Ko10]. Miller and Valasek [MV13] showed that normal CAN communication is not protected at all. Besides the nondisclosure of the message IDs and the encoding of the content, some manufacturers use additional application layer cyclic redundancy checks (CRCs), alive counters and time out windows; however, no actual security measures are implemented.

Application layer CRCs use a simple calculation over data bytes and the CAN message ID to assure integrity of the message. *Alive counters* serve as a sequence number which increases with every sent message. This allows the recipient to determine whether the messages arrived in the right order or if a message was lost. Most vehicular ECUs rely on periodically incoming status messages with a fixed cycle. A higher bus load for instance may cause the transmitting ECU to not exactly adhere to this defined cycle time. This

² see http://vector.com/vi_canoe_en.html

circumstance makes it necessary for ECUs to also accept messages deviating from the expected time of receipt. This time tolerance to determine the validity of a cyclic message is called *time out window*.

As a consequence, the current ways to protect the CAN bus against unauthorized manipulation are the nondisclosure of message IDs and the proprietary encoding of data, the usage of mechanisms like application layer CRCs, alive counters and time out windows as well as a (weak) authentication for diagnostic applications [TAA14]. Besides those techniques, ECUs use plausibility checks to determine the correctness of a received message. These plausibility checks verify the validity of a received CAN signal by comparing it to associated signals generated by any sensor, sent by another ECU or calculated from related other values.

4 CAN Communication Tester (CAN-CT)

In this paper we introduce a software application prototype named CAN Communication Tester (CAN-CT) to efficiently test vehicular ECUs for security (and potentially safety) vulnerabilities.

CAN-CT executes hacking attacks on the CAN bus. Various kinds of attacks, like replay, spoofing or Denial of Service (DoS) attacks, are implemented, can be parameterized, combined and will then be sent on the bus. Our aim is to cause unexpected behavior of the targeted ECU and in further consequence to unearth flawed implementations of the CAN communication or of the application itself.

We use a PCAN-USB³ as interface for CAN-CT to the CAN bus of the device under test. This adapter in combination with the *basic* library provided by PEAK-System⁴ allowed us to read and write CAN messages from any PC.

For CAN-CT's architecture we made use of known software design patterns to avoid dependencies on the used hardware and to clearly separate logic from view and data. The resulted modular design allows for an easy extensibility regarding further attacking types, specific higher layer protocols and possible future protection techniques. We are able to achieve a timing accuracy of our attacks of less than 100 μ s. Furthermore CAN-CT is capable of analyzing the CAN traffic in-depth and thus, provides the user with information about protection mechanisms in use and interesting targets on the bus.

4.1 Attack Scenario

The underlying attack scenario of CAN-CT is based on recent car hacks like of Valasek and Miller as well as of Checkoway et. al [VM15, Ch11]. It is best described by understanding CAN-CT as a compromised ECU. It assumes that we, as an attacker, already got full

³ see <http://www.peak-system.com/PCAN-USB.199.0.html?&L=1>

⁴ see <http://www.peak-system.com/PCAN-Basic.239.0.html?&L=1>

access to the CAN bus, by, for instance, exploiting a vulnerability in an ECU with both CAN access and an external interface (e.g. Bluetooth), thus acting as gateway. This means CAN-CT is capable of reading and writing arbitrary CAN messages. As the CAN bus is a broadcast medium, deleting messages between two nodes is not possible. Furthermore as we are not reprogramming the CAN controller of the compromised ECU, all attacks have to comply with the CAN protocol.

4.2 Attack Features

To provide a complete security testing environment, we focused on four main tasks: *monitoring the CAN bus*, *analyzing the traffic*, *attacking ECUs* and *simulating CAN nodes*.

Besides the possibility of monitoring the messages and events on the CAN bus, CAN-CT can also be used to simulate other ECUs. By connecting additional CAN controllers, we can specify details of the CAN communication of each node. Based on this description CAN-CT then sends the CAN messages accordingly. Thus, it is possible to build customized CAN networks, mainly needed for basic testing purposes.

CAN-CT further offers an in-depth analysis of received messages. Each message will be examined for the presence of specific protection mechanisms like application layer CRCs or alive counters. By comparing the details of the currently analyzed message with its preceding messages, we can determine its cycle time and figure out if and in which data bytes such protection measures are used.

The gathered intelligence is of great importance for further attacks, as it offers a profound knowledge about possible targets. This information can then be used by CAN-CT to serve as basis for arbitrary attacks. Besides a C#-scripting engine CAN-CT offers attack templates for

- replay and spoofing attacks,
- bus off attacks, and
- suppression of CAN communication attacks.

While replay and spoofing attacks can be understood as masquerade attacks with the aim to manipulate an ECU by impersonating the original sender of a message, bus off attacks represent a special form of a Denial of Service (DoS) attack (see Section 5.3.2). With this attack we can make an ECU unavailable while preserving all other communication on the bus. In contrast to this targeted DoS attack, we further implemented another form of a DoS attack which suppresses all or part of the CAN communication (see Section 5.3.1).

By specifying different parameters and attack options, we can execute sophisticated attacks with CAN-CT without the need of writing any line of code. Furthermore CAN-CT supports the validation of protection mechanisms like sequence numbers, application layer checksums, time outs, plausibility checks and error or failure handling mechanisms. We

can use triggers to determine the time or the event when an attack should be executed (e.g. after the successful execution of another attack, after a specified time, upon the receipt of a specific message etc.). All these attack definitions can be specified in CAN-CT's attack definition user interface depicted in Figure 1.

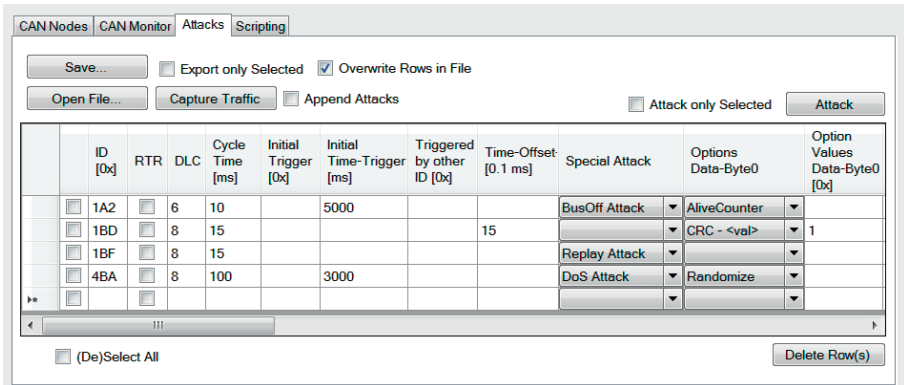


Fig. 1: CAN-CT's plan of attack definition

4.3 Limitations

As we made use of a third party USB to CAN adapter (PCAN-USB) all our actions had to conform to the CAN protocol (layer 1 and 2). A delay of around 1.3 ms caused by the used hardware did not provide us with the accuracy needed for our tests. We circumvented this limitation by making use of time-triggered attacks. These attacks anticipate the next time of receipt of a target message and react accordingly. This way we are able to assure an accuracy of less than 100 μ s

Due to the large number of different higher layer protocols we decided to not implement any of those protocols. Yet based on CAN-CT's generic design, it is possible to define attacks in a way to meet most requirements of higher layer protocols as well as to extend it, to support future protection mechanisms and attacks. However, complex bidirectional communications are not directly supported for now. Thus, the built-in support of attacks to verify the implementation of diagnostic protocols is subject of future work.

5 Evaluation of Real-World Applicability and Impact

To validate the general functioning of CAN-CT as well as to evaluate its real-world applicability and impact, we examined it in multiple setups.

After proving the general functioning of CAN-CT in a laboratory setup we tested it with actual vehicular ECUs. For this purpose we decided to make use of Hardware-in-the-Loop Systems (HiLs). A HiL can be understood as a test framework where parts of the system are simulated while other parts are actual hardware devices. This approach allows to test

a hardware in a true-to-nature test condition without the need of the actual system to be built [RSW07]. In the automotive domain the real hardware is commonly an ECU. The HiL acts as an interface to the simulated environment (e.g. engine), which provides all necessary electrical inputs as well as measures back all ECU electrical outputs. Hence, the hardware ECU under test gets all the information from the HiL as if it was coming from actual vehicle components; providing a test environment which is nearly similar to a real-life vehicle test.



(a) Photo of a heavy duty vehicle Hardware-in-the-Loop System setup (b) Connecting CAN-CT via the PCAN-USB adapter

Fig. 2: A heavy duty vehicle Hardware-in-the-Loop System setup

We used two different HiL setups at the *Virtual Test Field* (HiL laboratory) of the AVL LIST GmbH⁵ to verify CAN-CT. One setup represented a heavy duty vehicle and the other one a passenger car. The passenger car setup consisted of one real hardware ECU: the engine control unit; whereas the heavy duty setup consisted of two ECUs: engine and aftertreatment control unit (see Figure 2). To the best of our knowledge these ECUs represent quite typical vehicular control units, used in current-generation vehicles.

5.1 Overview of Evaluation Results

Using the HiL environment we were able to test the impact and the applicability of CAN-CT in close to real-world scenarios. Here we revealed significant weaknesses and unexpected behaviors of the ECUs under test.

These flaws encompassed serious lacks in the plausibility checking of messages, a too loose time out window for incoming messages, a not strict enough alive counter checking

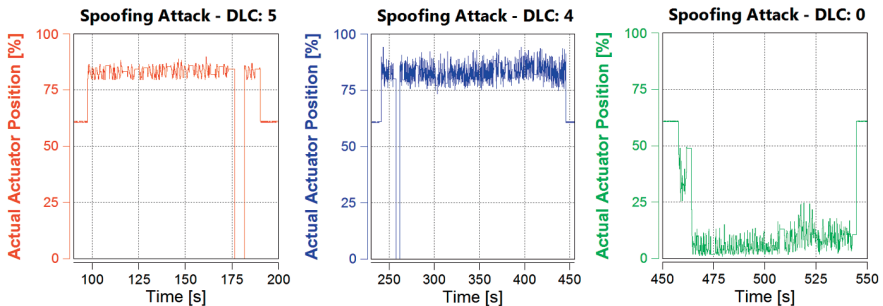
⁵ see <https://www.avl.com/>

procedure and a completely omitted examination of the correct frame structure. Furthermore we demonstrated that all tested ECUs are susceptible to our DoS attacks. In the following sections we are going to detail selected parts of these results.

5.2 Spoofing Attack

Being able to spoof messages on the CAN bus serves as the key task in many hacking attacks. With CAN-CT we were able to successfully impersonate other ECUs. We used information gained by analyzing the CAN traffic. Based on the knowledge if, and which message parts contained additional protection mechanisms we were able to draft our spoofed messages to meet all validation criteria at the receiving ECU. By sending our message right after the original message of the sender, we were often able to overwrite the original message in the recipients receive buffer in both HiL setups. This way our message prevails and is able to manipulate an ECU. Besides spoofing recurring messages we can further send CAN commands to cause a requested behavior.

Based on this approach we carried out various attacks to examine the security level of the target's protection mechanisms. One test focused on the target's validation of the received message frame. We discovered that the targeted ECU of the heavy duty vehicle HiL setup did not carry out any validation checks of the frame structure. A message that was defined in its specification to have a fixed length of eight data bytes, was accepted by the receiving ECU even when bytes were missing or the message was sent without any data at all. Missing values were just treated as zeros.



(a) Spoofing attack with a data length of 5 bytes (b) Spoofing attack with a data length of 4 bytes (c) Spoofing attack with a data length of 0 bytes

Fig. 3: Overview of different frame structures with regard to their data length (DLC) and their impact on the acceptance by the targeted ECU in the heavy duty vehicle HiL system setup. The noise, especially in Figure 3b, is caused by a not perfectly successful spoofing attack. The original transmitter of the message was able to place some messages in between our spoofed messages. This caused a minor adaption of the actuator position by the manipulated ECU.

This behavior is demonstrated in Figure 3. The picture shows spoofing attacks with different payload lengths, indicated by the data length code (DLC). The actuator position, we aimed to manipulate to correspond with our spoofed target position, changed for each plot due to the shorter data length of the message. For the first plot (data length of five bytes

(DLC: 5)) it was set to 80 percent (hex 0320). The second attack with four bytes of data already missed the last part of the target position. Instead of the hex value 0320 we just sent 03 which equals 76.8 percent when completed with zeros to the value 0300. The last graphic shows the attack without any data (DLC: 0). Instead of detecting this message as faulty it is again completed with all zeros which led to a target position of zero percent.

As a message without any payload does not contain any information, it should under no circumstances be complemented with zeros and treated as valid by the receiver. This implies that the ECU under test lacks a tight enough verification of incoming messages.

Similar to the above described shortcoming we validated the correctness of the target's time out window. A message that was sent every 15 ms, was still accepted even with an additional delay of 5 ms. Although we observed a deviation between the supposed and the actual time of receipt of up to 1.5 ms in both HiL setups, we believe that a tolerance of more than 5 ms is larger than required.

We were further able to detect vulnerabilities with regard to the alive counter checking procedure and the plausibility checks in the passenger car HiL setup. We demonstrated that the tested ECUs did not check whether the received signals are plausible. For instance we could show that a rise in the signaled vehicle speed from 0 to 250 km/h within one message was accepted by the ECU.

All these shortcomings ease potential hacking attacks as no knowledge of the underlying system is needed. As the common security principle for normal CAN communication in a vehicle is still best described by the security through obscurity approach, these flaws weaken this protection mechanism and offer susceptibilities for a number of attacks.

5.3 Denial of Service Attacks

Besides the manipulation of an ECU the second key task of a CAN hacking tool is the suppression of communication as well as making resources unavailable. For this purpose we implemented two different attacking types. The ECUs under test of both HiL setups did not differ significantly from each other regarding their reaction and behavior during our attacks. Thus, the following paragraphs detail the results of the heavy duty vehicle HiL setup.

5.3.1 Suppressing Communication

The first attack aims to suppress all lower priority messages. Due to the arbitration process of the CAN protocol and the dominant state of a binary zero, only messages with a lower ID (higher priority) than the one, that is flooding the bus, can prevail.

This behavior is illustrated in Figure 4. Here we can see, that, when the attack started at around 90 seconds, no more messages were received (no fluctuations anymore). Just in the

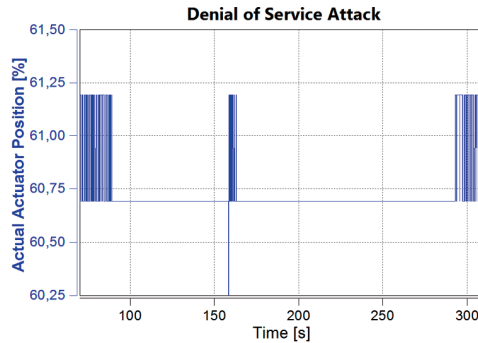


Fig. 4: DoS attack with ID 1BEh in the heavy duty vehicle HiL system setup. The fluctuations at the beginning and at the end of the figure show the original signal. The steady line indicates that no messages were received during this period of time.

middle of the attack the targeted ECU was able to prevail for a short period of time due to a collision on the bus.

During the attack we were not able to observe any error states or failures of the monitored devices. This implies that the ECUs continue their work with the lastly received values. However, when all communication on the bus is blocked, we would have expected the ECU to enter a "safe" mode or signal a warning. In a real-world hacking attack this could cause serious threats to the driver as no communication over this CAN bus is possible anymore.

5.3.2 Bus Off Attack

The second DoS attack we implemented does not block the CAN communication per se, but it continuously causes collisions on the bus, whenever the targeted ECU tries to transmit a message. Sending a message at the exactly same time as the sender, where the attack message differs from the original message only after the arbitration field, allows us to cause collisions without the target having any way to prevent it. This way we create the impression that the target's CAN controller is faulty.

Due to the CAN protocol's error handling mechanisms, the target has to increase its internal error counter, every time a collision occurs. As soon as this error counter reaches a value of 255, it enters the bus off state. In this state the ECU is not allowed to send nor to receive messages anymore. However, as we observed, the targeted ECUs regularly reset their internal error counters, allowing them to participate on the bus again as soon as possible.

This behavior is depicted in Figure 5. Here we aimed to send a falsified message content along, to directly take over this message ID (original actuator value 35 percent, spoofed value 30 percent). Shortly after we started our attack (bold, red line), we were able to make the target unavailable for the first time. However, a bus error was detected - the signal fell

to a value of zero - which caused our spoofed message to be declined by other recipients as well. At second 760 the targeted ECU reset its error counter, allowing it to send the actual message again. Soon afterwards we were able to successfully spoof our actuator position target value between seconds 780 to 810. At second 810 neither we nor the actual ECU could transmit the target value anymore. Due to the continuous reset of the target's error counter, we had to continue to cause collisions, not allowing us to manipulate other ECUs simultaneously. However, we were able to make the target unavailable while preserving all other CAN communication.

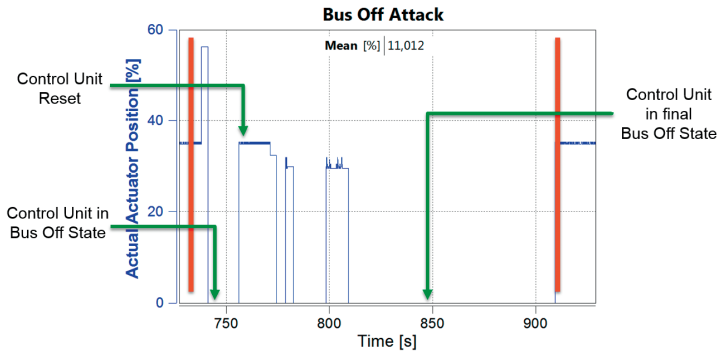


Fig. 5: Bus off attack in the heavy duty vehicle HiL system setup (the bold, red lines indicate the start and the end time of the attack)

6 Outlook and Conclusion

As car hacking becomes a serious threat, it is getting increasingly important to ensure a correct and secure behavior of every single ECU. Protection mechanisms are just as good as their implementation. We believe that a very efficient method to find security vulnerabilities in ECUs is, to attack them the same way as a hacker would.

Currently almost no security measures are implemented for normal communication on vehicular CAN buses. Thus, it is even more important to assure the proper working of existing protection techniques. Therefore, with CAN-CT we implemented an application that is capable of attacking ECUs while simultaneously tracking whether undesired or unspecified behavior was detected. This allows us to test the correct implementation of protection mechanisms on the one hand, while on the other hand attack detection techniques can be examined this way as well.

In this work we focused on the proof of concept of CAN-CT. The next step will be a successful integration into security evaluation methodologies and safety and security standards. Here we will have to develop a suitable concept for the systematic generation of test cases according to defined security goals.

The focus of CAN-CT in its current version is on the examination of normal CAN communication. Sophisticated bidirectional communications or diagnostic protocols are not implemented so far and are subject of future work. Using a specifically designed CAN

controller would further allow for not having to conform to the CAN protocol. This way additional attack scenarios like the physical connection of a malicious device to the CAN bus can be supported more realistically.

With the use of Hardware-in-the-Loop Systems we were able to prove CAN-CT working properly in a real-world scenario. We revealed significant shortcomings in commercial ECUs and thus, demonstrated its applicability and impact for current automotive systems. For that reason we want to highlight the importance of a security testing framework to complement existing safety and functional testing environments.

References

- [Ba14] Bayer, Stephanie; Enderle, Thomas; Oka, Dennis-Kengo; Wolf, Marko: Automotive - Safety & Security 2014 (2015), Sicherheit und Zuverlässigkeit für automobile Informationstechnik, Tagung, 21.-22.04.2015, Stuttgart, Germany. volume 240 of LNI. GI, 2014.
- [Ch09] Charette, Robert N.: This car runs on code. In: IEEE Spectrum. volume 46, p. 3, 2009.
- [Ch11] Checkoway, Stephen; McCoy, Damon; Kantor, Brian; Anderson, Danny; Shacham, Hovav; Savage, Stefan; Koscher, Karl; Czeskis, Alexei; Roesner, Franziska; Kohno, Tadayoshi: Comprehensive Experimental Analyses of Automotive Attack Surfaces. In: 20th USENIX Security Symposium, San Francisco, CA, USA, August 8-12, 2011, Proceedings. 2011.
- [GGV14] Glas, Benjamin; Gramm, Jens; Vembar, Priyamvadha: Automotive - Safety & Security 2014 (2015), Sicherheit und Zuverlässigkeit für automobile Informationstechnik, Tagung, 21.-22.04.2015, Stuttgart, Germany. volume 240 of LNI. GI, 2014.
- [HKD11] Hoppe, Tobias; Kiltz, Stefan; Dittmann, Jana: Security threats to automotive CAN networks - Practical examples and selected short-term countermeasures. Rel. Eng. & Sys. Safety, 96(1):11–25, 2011.
- [Je15] Jenik, Aviram: Increasing resilience by finding unknown vulnerabilities. In: iCC 2015 Proceedings - 15th international CAN Conference, Vienna, Austria, October 27-28, 2015. pp. 06–1–06–5, 2015.
- [Ko10] Koscher, Karl; Czeskis, Alexei; Roesner, Franziska; Patel, Shwetak; Kohno, Tadayoshi; Checkoway, Stephen; McCoy, Damon; Kantor, Brian; Anderson, Danny; Shacham, Hovav; Savage, Stefan: Experimental Security Analysis of a Modern Automobile. In: 31st IEEE Symposium on Security and Privacy, S and P 2010, 16-19 May 2010, Berkeley/Oakland, California, USA. pp. 447–462, 2010.
- [Ko14] Koscher, Karl: Securing Embedded Systems: Analyses of Modern Automotive Systems and Enabling Near-Real Time Dynamic Analysis. PhD thesis, University of Washington, 2014.
- [La13] Lawrenz, Wolfhard: CAN System Engineering. Springer-Verlag London, 2013.
- [Ma14] Marinescu, Raluca; Saadatmand, Mehrdad; Bucaioni, Alessio; Seceleanu, Cristina; Pettersson, Paul: A Model-Based Testing Framework for Automotive Embedded Systems. In: 2014 40th EUROMICRO Conference on Software Engineering and Advanced Applications. IEEE, pp. 38–47, August 2014.

- [MV13] Miller, Charlie; Valasek, Chris: , *Adventures in Automotive Networks and Control Units*, 2013.
- [NSL13] Navet, Nicolas; Simonot-Lion, Françoise: *In-vehicle communication networks - a historical perspective and review*. In (Zurawski, Richard, ed.): *Industrial Communication Technology Handbook*, Second Edition. CRC Press Taylor&Francis, 2013.
- [RSW07] Ren, Wei; Steurer, Michael; Woodruff, Steve: *Accuracy evaluation in power hardware-in-the-loop (PHIL) simulation center for advanced power systems*. In: *Proceedings of the 2007 Summer Computer Simulation Conference, SCSC 2007*, San Diego, California, USA, July 16-19, 2007. pp. 489–493, 2007.
- [Ta14] Talebi, Sareh: *A Security Evaluation and Internal Penetration Testing Of the CAN-bus*. Master of science thesis, Chalmers University of Technology, 2014.
- [TAA14] Tao, Zhang; Antunes, Helder; Aggarwal, Siddhartha: *Defending Connected Vehicles Against Malware: Challenges and a Solution Framework*. *IEEE Internet of Things Journal*, 1(1):10–21, 2014.
- [VM15] Valasek, Chris; Miller, Charlie: , *Remote Exploitation of an Unaltered Passenger Vehicle*, 2015.

Automatische Duplikateliminierung in Aktivitätsdiagrammen von Fahrzeugfunktionen

Martin Beckmann¹ und Aaron Schlutter²

Abstract: Die Spezifikation von Fahrzeugfunktionen ist eine komplexe Aufgabe. Zum Umgang mit dieser Komplexität werden zur Beschreibung der Funktionen grafische Modellierungssprachen wie UML verwendet. Bei der Modellierung können Duplikate entstehen, welche Ausgangspunkt für Fehler und Inkonsistenzen in der weiteren Entwicklung sind. Dieser Beitrag widmet sich der Eliminierung von Duplikaten, welche bei der Spezifikation von Fahrzeugfunktionen mittels UML Aktivitätsdiagrammen auftreten. Es wird dargestellt, wie in UML Aktivitätsdiagrammen identifizierte Duplikate automatisiert eliminiert werden, ohne die ursprüngliche Funktionalität zu verändern. Mehrfach auftretende Elemente werden zusammengefasst und durch das Einfügen von weiteren Elementen und Verbindungen zusammengesetzt.

Keywords: Automatische Duplikateliminierung, UML Aktivitätsdiagramme, Modellbasierte Spezifikation

1 Einführung

Zur Beherrschung der zunehmenden Komplexität bei der Entwicklung von Fahrzeugsoftware wird vermehrt modellbasierte Softwareentwicklung verwendet [Br06]. Verhaltensmodelle wie das UML *Aktivitätsdiagramm* [UM15, S. 371] können dazu genutzt werden, um Funktionsabläufe zu beschreiben und grafisch darzustellen. Die Erstellung eines solchen Modells im Rahmen der Spezifikation kann daher frühzeitig dazu beitragen, Fehler, Widersprüche und Inkonsistenzen aufzuzeigen [AD97]. Eine mögliche Anwendung des Modells im weiteren Entwicklungsprozess ist die automatische Generierung von Bestandteilen einer textuellen Spezifikation. Diese Generierung ist eines unserer weiteren aktuellen Forschungsvorhaben.

Während der Erstellung der Funktionsmodelle können innerhalb des Modells Redundanzen auftreten. Diese entstehen oft unabsichtlich, da bei deren Erstellung viele Personen beteiligt sind und trotz des Einsatzes von grafischen Werkzeugen der Überblick schwierig ist. Die Redundanzen entstehen jedoch auch absichtlich [LEM02]. Beispielsweise werden redundante Elemente eingefügt, um die Modelle besser zu gruppieren, wenn ein einzelnes Element in mehreren Zusammenhängen innerhalb der Fahrzeugfunktion auftritt. Die dabei entstehenden Duplikate können jedoch Schwierigkeiten bei der Wahrung der Konsistenz verursachen, wenn die Diagramme in weiteren Entwicklungsschritten verwendet werden. Duplikate sollten daher entfernt werden.

¹ Technische Universität Berlin, Daimler Center for Automotive IT Innovations, Ernst-Reuter-Platz 7, 10587 Berlin, martin.beckmann@tu-berlin.de

² Technische Universität Berlin, Daimler Center for Automotive IT Innovations, Ernst-Reuter-Platz 7, 10587 Berlin, aaron.schlutter@tu-berlin.de

Dieser Beitrag widmet sich der Eliminierung identifizierter Duplikate in UML *Aktivitätsdiagrammen*. Bei der Eliminierung werden mehrfach auftretende Elemente zusammengefasst. Bestehende Verbindungen der zusammengefassten Elemente werden übernommen und durch zusätzliche Elemente und Verbindungen so zusammengesetzt, dass die Funktionalität des ursprünglichen *Aktivitätsdiagramms* erhalten bleibt.

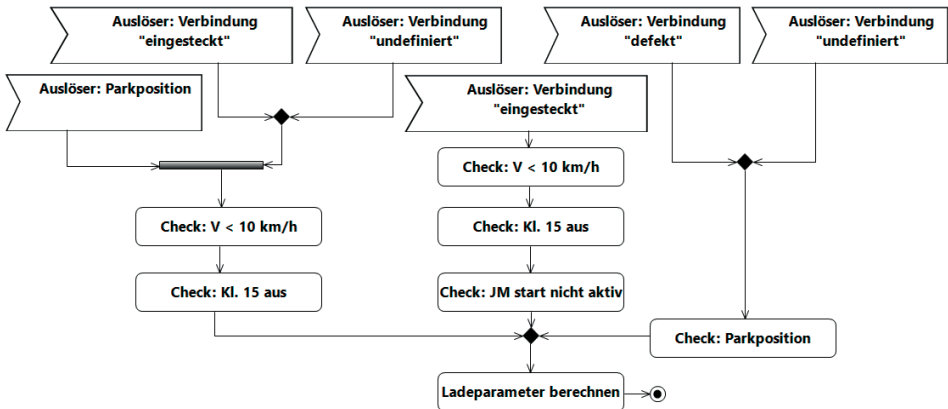
Dieser Beitrag ist wie folgt aufgebaut. Im nächsten Kapitel werden zunächst Grundlagen erläutert und anschließend werden anhand eines einführenden Beispiels Annahmen zum Verständnis dieses Beitrags dargelegt. Eine Übersicht über verwandte Arbeiten ist im dritten Kapitel enthalten. In Kapitel 4 wird das Konzept zur Eliminierung der Duplikate vorgestellt und im fünften Kapitel wird das Konzept auf das einführende Beispiel angewendet. Der Beitrag wird mit einer Zusammenfassung und einem Ausblick in Kapitel 6 abgeschlossen.

2 Grundlagen

Im Rahmen dieser Arbeit werden UML *Aktivitätsdiagramme* zur Beschreibung von Fahrzeugfunktionen betrachtet. *Aktivitätsdiagramme* sind Petri-Netz ähnliche Graphen [UM15, S. 283]. Mithilfe von Tokens (Marken) wird die Ausführungsreihenfolge von Elementen und damit der mögliche Ablauf des *Aktivitätsdiagramms* modelliert. Für die Ausführung eines Knotens muss sich ein Token in diesem befinden. Das Token kann bei Abschluss der Ausführung des Knotens an weitere Knoten weitergegeben werden. Das Verhalten eines Petri-Netzes beziehungsweise *Aktivitätsdiagramms* kann durch einen Erreichbarkeitsgraph dargestellt werden. Ein Erreichbarkeitsgraph ist ein gerichteter Graph, der alle möglichen Reihenfolgen (Pfade) zur Ausführung der Knoten zeigt. Ein Knoten des Erreichbarkeitsgraphs stellt einen möglichen Markierungszustand aller Knoten des Petri-Netzes dar.

In Abbildung 1 ist ein *Aktivitätsdiagramm* der Funktion *Ladeparameter berechnen* dargestellt. Verschiedene Knoten (*ActivityNodes*) sind über Kanten (*ActivityEdges*) miteinander verbunden. Die Funktion hat verschiedene Auslöser, welche als *AcceptEventAction* Elemente dargestellt sind. Neben den Auslösern werden bei der Ausführung der Funktion auch *Actions* durchlaufen. Die Zusammenhänge dieser Elemente wird durch *ControlNodes* gesteuert. Auslöser und *Actions* können mit *JoinNodes* UND-verknüpft oder mit *MergeNodes* ODER-verknüpft werden. Weitere in diesem Beitrag genutzte *ControlNodes* sind *InitialNodes*, *DecisionNodes* und *ForkNodes*. Es werden nur diese Elemente verwendet. Der Grund dafür ist, dass die von einem Industriepartner bereitgestellten Fahrzeugfunktionen ausschließlich mithilfe dieser Elemente modelliert sind.

Die *Actions* mit dem Präfix *Check* stellen Überprüfungen von Bedingungen dar. Schlägt diese Überprüfung fehl, so endet der Fluss in diesem Knoten. Der Fluss eines Tokens von einem Auslöser zu einer Überprüfung bildet daher eine UND-Verknüpfung von Auslöser und Überprüfung ab. Eine Möglichkeit, zur Ausführung der *Action Ladeparameter berechnen* zu gelangen, besteht, wenn der Auslöser *Verbindung "eingesteckt"* aktiviert wird und die darauffolgenden Überprüfungen *V < 10 km/h*, *Kl. 15 aus* und *JM start nicht aktiv* zu


 Abb. 1: Funktionsablauf der Fahrzeugfunktion *Ladeparameter berechnen*

wahr ausgewertet werden. Besitzt eine *Action* oder *AcceptEventAction* mehrere ausgehende Kanten, so wird an allen ausgehenden Kanten ein Token angeboten [UM15, S. 401]. Dies entspricht der Anwendung von *ForkNodes*. Treten bei anderen Elementen wie zum Beispiel *DecisionNodes* mehrere ausgehende Kanten auf, so wird nur ein einzelnes Token an einen nachfolgenden Knoten gegeben, welcher das Token akzeptiert [UM15, S. 373, S. 387f].

In der in Abbildung 1 dargestellten Fahrzeugfunktion sind folgende Duplikate enthalten: die Auslöser *Verbindung "eingesteckt"* und *Verbindung "undefiniert"* sowie die Überprüfungen *V < 10 km/h* und *Kl.15 aus*. Im Rahmen dieses Beitrags besteht ein Duplikat aus Knoten, welche dieselbe Bedeutung besitzen. Beispielsweise bilden die beiden Überprüfungen mit dem Text *V < 10 km/h* ein Duplikat. Es wird davon ausgegangen, dass die Duplikate bereits identifiziert sind. Die Identifikation von Duplikaten bei der Spezifikation ist beispielsweise mit textuellen Vergleichen möglich [Da01]. Es ist jedoch zu beachten, dass innerhalb einer Aktivität auch ähnliche Elemente den gleichen Namen haben können [UM15, S. 372] und trotzdem kein Duplikat darstellen. Dies ist beispielsweise der Fall, wenn von den Elementen mit gleichen Namen verschiedene Aktivitäten aufgerufen werden.

Für die in dieser Arbeit betrachteten Fahrzeugfunktionen und deren *Aktivitätsdiagramme* ist dabei die exakte Reihenfolge zur Ausführung der Knoten nicht relevant. Vielmehr ist es notwendig, dass jeweils alle Vorgänger ausgeführt worden sind und die Gesamtanzahl der Ausführungen nicht verändert wird. Unter Umständen kann durch das Umformungen des in Kapitel 4 vorgestellten Konzeptes die Ausführungsreihenfolge der einzelnen Knoten verändert werden. Dies erfolgt durch die mehrfache, zeitversetzte Ausführung von Auslösern. Der Sachverhalt wird in Kapitel 4 genauer erläutert.

3 Verwandte Arbeiten

In [CST02] werden verschiedene Arten von Redundanzen identifiziert, die in UML Modellen auftreten können. Dabei findet eine Beschränkung auf die Diagrammtypen *Anwendungsfalldiagramm*, *Zustandsdiagramm* und *Sequenzdiagramm* bei der Verwendung während der objektorientierten Analyse statt. Betrachtet werden Redundanzen, die zwischen verschiedenen Modellen auftreten. Außerdem werden Vorteile aufgezeigt, die durch die Beseitigung der Redundanzen entstehen wie zum Beispiel die Wahrung der Konsistenz bei Modifikationen und die erleichterte Nutzung in späteren Phasen der Entwicklung.

Die Umformung von UML Modellen unter Beibehaltung des Verhaltens ist das Thema von [Su01]. Es werden Operationen vorgestellt, die genutzt werden können, um Mängel im Design zu beseitigen. Ebenso wird begründet, wieso das Verhalten der Modelle bei der Durchführung der Operationen erhalten bleibt. Der Beitrag stellt Operationen für *Klassendiagramme* und *Zustandsdiagramme* vor. Auf den Aspekt der Eliminierung von Redundanzen zur Verbesserung der Modelle wird nicht eingegangen.

Zwei Operationen, die auf *Aktivitätsdiagrammen* Anwendung finden, sind in [BSF03] beschrieben. Die beiden vorgestellten Operationen behandeln die Reihenfolge der Ausführung von *Actions* und ob diese parallelisiert oder sequenziert auszuführen sind. Zum Thema der Umformung von UML Modellen existieren eine Reihe weiterer Beiträge. Nach Kenntnis der Autoren widmet sich jedoch keine Arbeit der Eliminierung von Duplikaten.

4 Konzept zum Eliminieren von Duplikaten

Das im Folgenden vorgestellte Konzept zum Eliminieren von Duplikaten in *Aktivitätsdiagrammen* besteht aus zwei Teilen. Zum einen werden redundante Elemente zusammengefasst und zum anderen muss die Funktionalität vom *Aktivitätsdiagramm* sichergestellt werden. Dies wird erreicht, indem ein- und ausgehende Kanten der zusammengefassten Knoten weitergeführt werden oder unter Umständen indem neue *ControlNodes* eingefügt und mit dem zusammengefassten Knoten verbunden werden. Nach der Zusammenfassung der Knoten kann zusätzlich nach nicht mehr notwendigen Kanten und *ControlNodes* gesucht werden. Diese Elemente zeichnen sich dadurch aus, dass nach deren Entfernen aus dem *Aktivitätsdiagramm* dessen Funktionalität nicht verändert wird.

In Abbildung 2 sind beispielhaft verschiedene Ausschnitte aus *Aktivitätsdiagrammen* dargestellt. Die Knoten, deren Text mit *T* beginnt, stellen Auslöser dar. Die anderen Knoten sind *Actions*. Ein Hochkomma im Text des Knotens bezeichnet einen redundanten Knoten.

In der ersten Situation in Abbildung 2a sind die Knoten des Duplikats die Auslöser der Funktion. Die Auslöser besitzen keine eingehenden Kanten. In Abbildung 3a ist die duplikatfreie Darstellung zu sehen. Die Auslöser sind in einem Knoten zusammengefasst und alle ausgehenden Kanten haben als neue Quelle den zusammengefassten Knoten und als Ziel ein vor der *Action* befindliches *DecisionNode*. Die Kanten können nicht entfernt oder zusammengeführt werden, da vor der Eliminierung durch beide Auslöser zwei Tokens er-

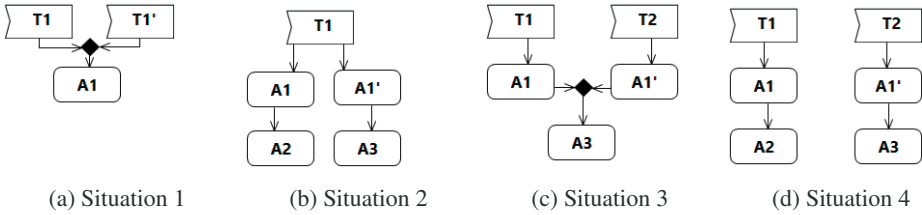


Abb. 2: Unterscheidung verschiedener Situationen

stellt werden, sodass die *Action A1* zweimal ausgeführt wird. Durch die Umformung wird beim Auslösen von *T1* die *Action A1* ebenfalls zweimal ausgeführt.

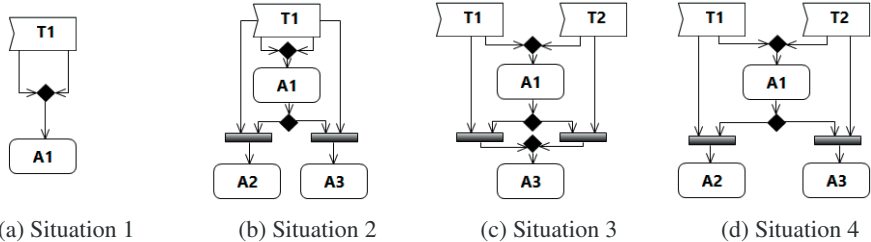


Abb. 3: Situationen ohne Duplikate

In den anderen Situationen in Abbildung 2b, 2c und 2d besitzen die redundanten Knoten eingehende Kanten. In der zweiten Situation ist der Vorgänger identisch und die Nachfolger unterschiedlich. In der dritten Situation sind die Vorgänger unterschiedlich und der Nachfolger ist identisch. In der vierten Situation sind sowohl Vorgänger als auch Nachfolger unterschiedlich. In diesen Situationen werden zusätzliche *ControlNodes* und Kanten eingefügt, um die Funktionalität beizubehalten. Diese zusätzlichen *ControlNodes* und Kanten sind im vierten Fall notwendig, damit durch den Auslöser *T1* nach der Beendigung von *Action A1* lediglich *Action A2* aktiviert werden kann. Der *JoinNode* stellt sicher, dass Auslöser *T1* und *Action A1* zuvor ausgeführt worden sind. Der *MergeNode* dient in der dritten und vierten Situation dazu, dass *Action A1* durch ein einzelnes Token von *T1* oder *T2* ausgeführt werden kann. Der *DecisionNode* leitet das Token am Ausgang von *Action A1*, an den entsprechenden *JoinNode* weiter. In der zweiten und dritten Situation werden dabei *ControlNodes* und Kanten eingefügt, welche nicht notwendig sind und wieder entfernt werden können. Um ein einheitliches Konzept für diese Situationen zu ermöglichen, werden diese Situationen jedoch gleich behandelt. Eine Vereinfachung ist im Anschluss möglich.

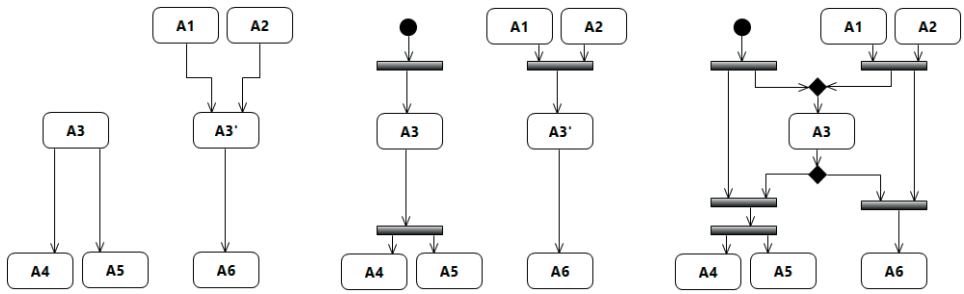
Für die Umformung der redundanten Knoten werden zwei Fälle unterschieden:

1. Keiner der Knoten eines Duplikats hat eingehende Kanten.
2. Mindestens einer der Knoten des Duplikats hat mindestens eine eingehende Kante.

Für den ersten Fall werden die betroffenen Knoten zu einem Knoten zusammengefasst und alle ausgehenden Kanten der ursprünglichen Knoten haben als Quelle den zusammengefassten Knoten. Im zweiten Fall werden ebenfalls die redundanten Knoten in einen Knoten zusammengefasst. Alle eingehenden Kanten werden über einen neuen *MergeNode* gebündelt zu dem zusammengefassten Knoten geführt. Aus dem zusammengefassten Knoten geht eine Kante zu einem *DecisionNode*. Von dort aus gehen Kanten zu *JoinNodes*. Zu den *JoinNodes* führen zudem eingehende Kanten von den Vorgängern der redundanten Knoten. Handelt es sich bei den Vorgängern nicht um *Actions*, werden zusätzlich noch *ForkNodes* hinter diesen eingefügt. Diese dienen stattdessen als Vorgänger, um die Funktionalität von mehreren ausgehenden Kanten bei *Actions* nachzustellen. Die ausgehende Kante des *JoinNodes* führt zu dem dazugehörigen Nachfolger des redundanten Knotens. Sind keine Nachfolger vorhanden, entfallen die entsprechenden Elemente und Verbindungen für die ausgehende Kante.

Besonderheiten

Die in Abbildung 2 dargestellten Situationen decken nicht alle Möglichkeiten der beiden Fälle ab. Es ist denkbar, dass ein Duplikat existiert, welches Knoten besitzt, von denen ein Knoten einen Vorgänger hat, ein anderer Knoten jedoch keinen Vorgänger besitzt. Des Weiteren ist es möglich, dass Knoten von Duplikaten mehr als einen Vorgänger oder Nachfolger besitzen. Ein Beispiel, in denen solche Konstellationen auftreten, ist in Abbildung 4 dargestellt.



(a) Ausgangssituation

(b) Explizite Verbindungen

(c) Duplikatfrei

Abb. 4: Sonderfall

In Abbildung 4a besitzt die *Action* A3 keinen Vorgänger und zwei Nachfolger. Die *Action* A3' hat zwei Vorgänger und einen Nachfolger. Für den Knoten ohne Vorgänger ist ein *InitialNode* als Vorgänger zu ergänzen. Um mit den Situationen mit mehreren Vorgängern und Nachfolgern umzugehen, müssen zunächst die impliziten Verbindungen ([UM15, S. 401]) explizit dargestellt werden. Bei *Action* A3 wird durch die ausgehenden Kanten eine implizite Parallelisierung vorgenommen (vgl. *ForkNode*). Bei *Action* A3' wird durch die mehrfachen Vorgänger eine implizite UND-Verknüpfung erzeugt. In Abbildung 4b sind das Einfügen eines *InitialNodes* und die expliziten Verbindungen dargestellt. Der

zusätzliche *ForkNode* nach dem *InitialNode* ist notwendig, da der Knoten des Duplikats Nachfolger haben kann. In Abbildung 4c ist die Eliminierung nach dem vorgestellten Konzept dargestellt. Als Vorgänger und Nachfolger für die Umformung dienen an den entsprechenden Stellen die eingefügten Elemente der expliziten Verbindungen.

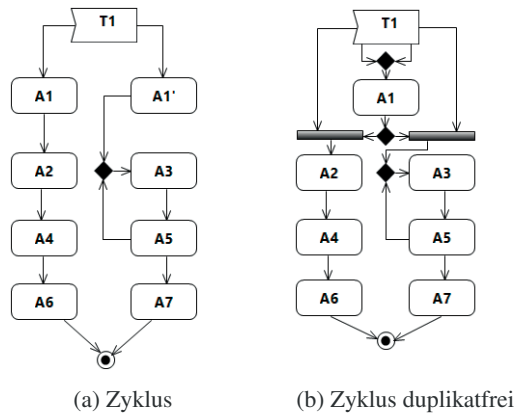


Abb. 5: Zyklus im Aktivitätsdiagramm

Eine weitere Besonderheit stellen Zyklen in den *Aktivitätsdiagrammen* dar. Ein solcher Zyklus und dessen Lösung nach dem vorgestellten Konzept ist in Abbildung 5a und in in Abbildung 5b dargestellt. Das Konzept lässt sich korrekt auf eine Reihe von möglichen Zyklen anwenden. Es ist jedoch noch nicht abschließend geklärt, ob dies immer dazu führt, dass die Funktionalität erhalten bleibt.

Vergleich der Erreichbarkeitsgraphen

Allgemein ist für den Erhalt der Funktionalität eines *Aktivitätsdiagramms* bei der Zusammenführung von redundanten Knoten darauf zu achten, dass alle möglichen Pfade des *Aktivitätsdiagramms* erhalten bleiben. Die möglichen Pfade können aus dem Erreichbarkeitsgraphen des *Aktivitätsdiagramms* ermittelt werden. In Abbildung 6 ist der Erreichbarkeitsgraph des *Aktivitätsdiagramms* mit der Startmarkierung $(1,1,0,0,0,0)$ für das Aktivitätsdiagramm aus Abbildung 2d und in Abbildung 7 der Erreichbarkeitsgraph $(T1, T2, A1, A2, A3)$ mit der Startmarkierung $(1,1,0,0,0)$ für das Aktivitätsdiagramm aus Abbildung 3d dargestellt. Für den zweiten Erreichbarkeitsgraphen wird angenommen, dass die neu eingefügten *ControlNodes* direkt ausgeführt werden. Diese Elemente sind daher zur besseren Vergleichbarkeit nicht separat in der Markierung aufgeführt.

Die möglichen Pfade des Erreichbarkeitsgraphen in Abbildung 6 für das *Aktivitätsdiagramm* aus Abbildung 2d lauten:

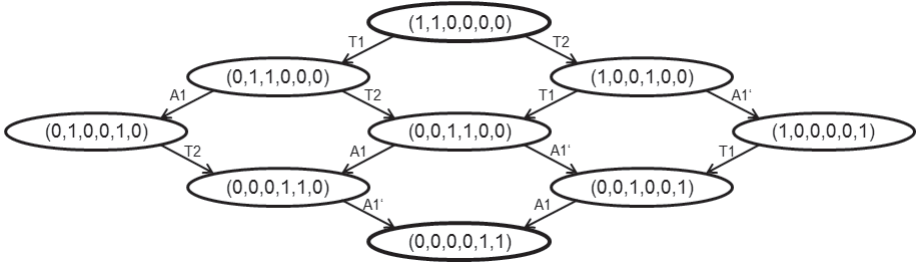


Abb. 6: Erreichbarkeitsgraph ($T1, T2, A1, A1', A2, A3$) des Aktivitätsdiagramms aus Abbildung 2d

- | | |
|----------------------|----------------------|
| 1. $T1, A1, T2, A1'$ | 4. $T2, T1, A1, A1'$ |
| 2. $T1, T2, A1, A1'$ | 5. $T2, T1, A1', A1$ |
| 3. $T1, T2, A1', A1$ | 6. $T2, A1', T1, A1$ |

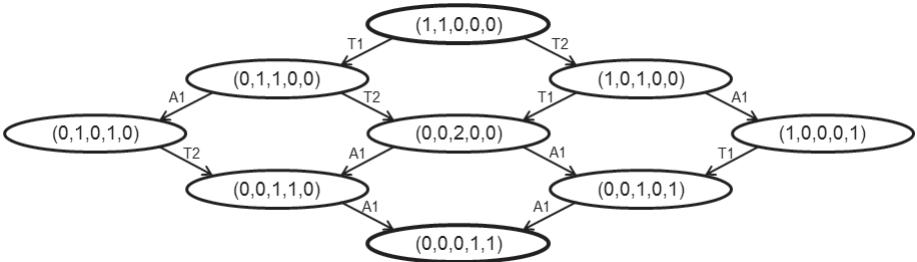


Abb. 7: Erreichbarkeitsgraph ($T1, T2, A1, A2, A3$) des Aktivitätsdiagramms aus Abbildung 3d

Die möglichen Pfade des Erreichbarkeitsgraphen in Abbildung 7 für das Aktivitätsdiagramm aus Abbildung 3d lauten:

- | | |
|---------------------|---------------------|
| 1. $T1, A1, T2, A1$ | 3. $T2, T1, A1, A1$ |
| 2. $T1, T2, A1, A1$ | 4. $T2, A1, T1, A1$ |

Unter der Voraussetzung $A1=A1'$ sind die möglichen Pfade der Erreichbarkeitsgraphen aus Abbildung 6 und 7 äquivalent. Die Pfade Nr. 1 und 6 aus Abbildung 6 sind äquivalent zu den Pfaden Nr. 1 und 4 aus Abbildung 7. Die Pfade Nr. 2 und 3 aus Abbildung 6 sind äquivalent zum Pfad Nr. 2 aus Abbildung 7. Das Gleiche gilt für die Pfade Nr. 4 und 5 aus Abbildung 6 und den Pfad Nr. 3 aus Abbildung 7.

Im Vergleich der beiden Erreichbarkeitsgraphen ist zudem zu erkennen, dass die Anzahl und Verteilung der Token in jedem Knoten (d.h. Zustand im Aktivitätsdiagramm) immer gleich sind, unter der Bedingung $A1=A1'$ und dass die Anzahl der Token von $A1$ und $A1'$ aufsummiert werden.

Der Erreichbarkeitsgraph ist nur bedingt geeignet, den vollständigen Ablauf eines Aktivitätsdiagramms darzustellen. Zum einen gilt ein Erreichbarkeitsgraph immer nur für eine bestimmte Startmarkierung von Token. Für eine andere Startmarkierung als $(1,1,0,0,0)$ in Abbildung 6 ergibt sich ein neuer Erreichbarkeitsgraph. Zum anderen gibt es keine vollständige Semantik von Aktivitätsdiagrammen. Im Gegensatz zu Knoten in Petri-Netzen muss eine *Action* nicht zwangsläufig in jedem Schaltschritt vollständig ausgeführt werden in dem Sinne, dass eingehende Token angenommen und nach der Aktivierung innerhalb des Zeitschritts direkt weitergegeben werden. Ein weiterer Unterschied besteht darin, dass ein Auslöser jederzeit ausgeführt werden kann und einen neuen Token produziert.

Durch diese Abweichungen kann es vorkommen, dass unter bestimmten Umständen der zuvor bestehende Ablauf einzelner Fälle nicht exakt identisch nach der Umformung besteht. Im vierten Fall aus Abbildung 3d kann durch Auslöser $T1$ die *Action* $A1$ ausgeführt werden. Anschließend liegt an dem *DecisionNode* nach $A1$ ein Token. Der Auslöser $T2$ kann nun direkt $A3$ ausführen, ohne dass direkt zuvor $A1$ ein weiteres Mal ausgeführt worden ist. Der Pfad entspricht dann (auszugsweise) $T1, A1, T2, A3$.

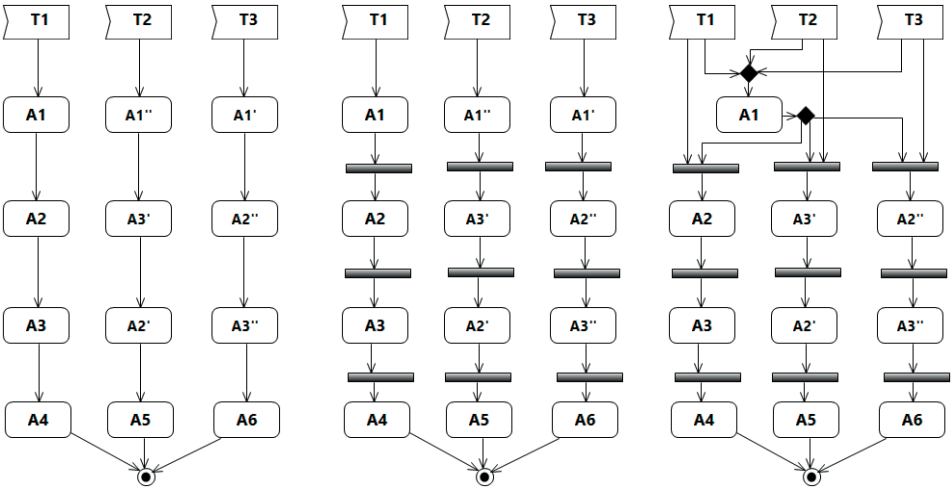
Der entsprechende Erreichbarkeitsgraph $(T1, T2, A1, A2, A3)$ hat die Markierung $(0,0,1,0,1)$, wobei zusätzlich noch ein Token am *JoinNode* zwischen $T1$ und $A2$ am Eingang von $T1$ vorhanden ist. In diesem Fall wird durch die Ausführung von $A1$ anschließend auch $A2$ ausgeführt. Der fortgeführte Pfad lautet $T1, A1, T2, A3, A1, A2$. Im Vergleich dazu ist ein möglicher Pfad für die parallele Ausführung von beiden Auslösern $T1, T2, A1, A3, A1, A2$. In beiden Fällen werden die gleichen *Actions* ausgeführt. Diese Eigenschaft ist für die in dieser Arbeit betrachteten *Aktivitätsdiagramme* ausreichend.

Alternativ kann auch mithilfe von speziellen, wertbehafteten *ObjectTokens* erreicht werden, dass diese nur von bestimmten *Actions* akzeptiert werden. Auf diese Weise kann die zuvor beschriebene Konstellation vermieden werden. Als weitere Möglichkeit können Bedingungen an den Kanten verwendet werden, die sicherstellen, woher ein Token stammt.

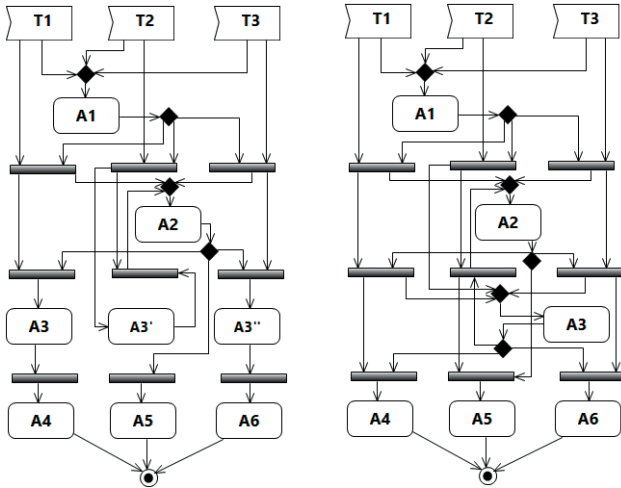
Algorithmisches Vorgehen

Die Fälle sollen nachfolgend an der Umformung in Abbildung 8 verdeutlicht werden. Dabei werden schrittweise alle Duplikate ausgehend vom Startknoten entfernt und zudem wird gezeigt, dass das vorgestellte Konzept auch für mehrere Duplikate hintereinander angewendet werden kann, auch wenn diese nicht in gleicher Abfolge vorliegen. Aus Darstellungsgründen wird auf *ForkNodes* verzichtet, die notwendig sind, wenn eingefügte *JoinNodes* als Vorgänger verwendet werden.

In der Ausgangssituation in Abbildung 8a sind drei unterschiedliche Auslöser $T1, T2$ und $T3$ als Startknoten und drei unterschiedliche *Actions* $A4, A5$ und $A6$ als Endknoten vorhanden. Im ersten Schritt in Abbildung 8b werden zunächst nach allen Knoten der identifizierten Duplikate *JoinNodes* eingefügt. Die *JoinNodes* ändern den Ablauf des *Aktivitätsdiagramms* nicht, sondern dienen als die direkten Nachfolger der redundanten Knoten.



(a) Ausgangssituation (b) Einfügen von *JoinNodes* (c) Duplikat A1 entfernt



(d) Duplikat A2 entfernt (e) Duplikat A3 entfernt

Abb. 8: Vorgehen der Duplikateliminierung

Beim Entfernen des ersten Duplikats bestehend aus den Knoten A1, A1' und A1'' wird vor dem zusammengefassten Knoten ein *MergeNode* und hinter dem Knoten ein *DecisionNode* eingefügt. Alle Vorgänger werden mit dem *MergeNode* verbunden. Der *DecisionNode* erhält ausgehende Kanten zu den folgenden *JoinNodes*. Anschließend werden die Vorgänger und Nachfolger der ursprünglichen redundanten Knoten miteinander verbunden.

Die gleiche Umformung wird im Folgenden für die Duplikate A2, A2' und A2'' in Abbildung 8d sowie für A3, A3' und A3'' in Abbildung 8e angewendet. Hierbei ist erkennbar,

dass die Umformung auch angewendet werden kann, wenn sich die Abfolge der Duplikate unterscheidet. Zudem ist es unerheblich, in welcher Reihenfolge die Duplikate eliminiert werden.

5 Anwendung am Beispiel

Zur Eliminierung aller identifizierten Duplikate in einem *Aktivitätsdiagramm* wird das in Kapitel 4 präsentierte Konzept für alle Duplikate angewendet. Abgeschlossen wird es mit der Minimierung der Anzahl der Kanten und der eingefügten Elemente.

Bei der Anwendung des Konzepts auf das Beispiel aus Kapitel 2 müssen die vier dort hergestellten Duplikate betrachtet werden. Das Endergebnis ist in Abbildung 9 dargestellt.

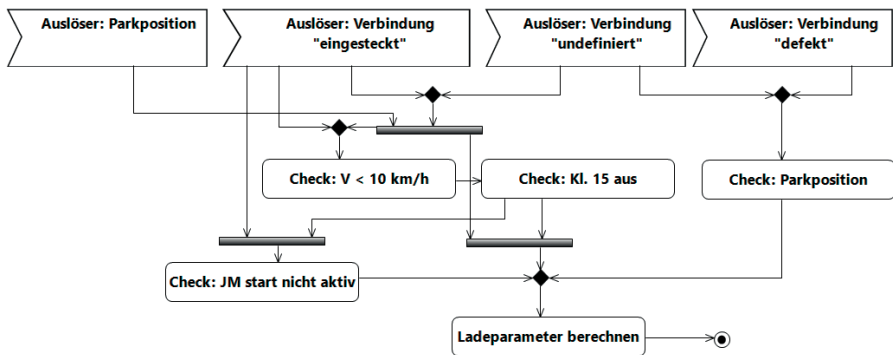


Abb. 9: Funktionsablauf der Fahrzeugfunktion *Ladeparameter berechnen* duplikatfrei

Die redundanten Knoten der Duplikate von Auslöser *Verbindung "eingesteckt"* und von Auslöser *Verbindung "undefiniert"* sind zusammengeführt. Die ausgehenden Kanten sind den einzelnen verbleibenden Knoten zugeordnet. Eingehende Kanten sind nicht vorhanden. Ebenfalls sind die Knoten der Duplikate von den Überprüfungen *V < 10 km/h* und *Kl. 15 aus* zusammengeführt. Zum Erhalt der Funktionalität sind zwei *JoinNodes* und ein *MergeNode* im Diagramm hinzugekommen. Alle weiteren eingefügten *ControlNodes* und Kanten, die während des Vorgehens entstanden sind, werden durch Vereinfachungen wieder entfernt.

Im vorliegenden Beispiel sind vier Duplikate eliminiert worden, welche aus jeweils zwei identischen Elementen bestehen. Das Vorgehen ist jedoch unabhängig von der Anzahl der identischen Knoten anwendbar. Ebenso gelten keine Beschränkungen für die Anzahl der Knoten aus verschiedenen Duplikaten, die hintereinander im Ablauf stehen. Dies liegt darin begründet, dass das Vorgehen schrittweise angewendet wird und somit immer vergleichbare Situationen auftreten.

Ein Nachteil der Eliminierung der Duplikate besteht laut Meinung von Anwendern darin, dass die Darstellung an Übersichtlichkeit verlieren kann, da die zuvor bereits bekannte Anordnung nicht mehr sofort erkennbar ist. Diagramme sind jedoch ab einer bestimmten

Größe nur noch eingeschränkt für Menschen sofort verständlich [Mi56]. Darüber hinaus ist die Übersichtlichkeit als Basis für weitere Schritte in der Entwicklung nicht das primäre Kriterium. Unabhängig davon kann eine übersichtlichere Darstellung mit Duplikaten als bereits bekannte Darstellung für den Anwender weiterhin verwendet werden. In diesem Fall ist eine Synchronisation bei Änderungen der beiden Diagramme beispielsweise durch eine entsprechende Verlinkung anzustreben.

6 Zusammenfassung und Ausblick

Im vorliegenden Beitrag wird erstmalig ein Konzept präsentiert, um in UML *Aktivitätsdiagrammen* auftretende Duplikate automatisch zu eliminieren. Bei dem vorgestellten Konzept bleibt die Funktionalität erhalten. Dies ermöglicht eine bessere Verwendung des Diagramms in darauffolgenden Schritten der Entwicklung. Es sind zwei Fälle dargestellt, welche unterschieden werden. Je nach Fall wird eine andere Umformung angewendet. Das resultierende Ergebnis kann anschließend auf nicht mehr notwendige *ControlNodes* und Kanten untersucht werden, die durch die Duplikateliminierung entstehen. Diese *ControlNodes* und Kanten sind wieder zu entfernen. Zusätzlich erfolgt eine Betrachtung von auftretenden Besonderheiten der vorgestellten Fälle und von Zyklen.

Es handelt sich dabei um einen ersten Schritt zur Eliminierung von Duplikaten unter Beibehaltung der Funktionalität. In zukünftigen Arbeiten wird adressiert, inwiefern die hier dargestellten Fälle vollständig sind. Damit einhergehend wird an dem formalen Nachweis gearbeitet, dass die Nutzung des vorgestellten Konzepts das Verhalten der UML *Aktivitätsdiagramme* erhält. Darüber hinaus beschäftigen sich die Autoren mit weiteren Elementen, die in *Aktivitätsdiagrammen* vorkommen und hier noch nicht betrachtet worden sind.

Angrenzend an diese Vorhaben stellt die Identifikation der redundanten Elemente einen weiteren Aspekt dar, dessen Automatisierung zu erheblichen Zeitersparnissen führt und hilft die Qualität der Diagramme zu verbessern.

Literaturverzeichnis

- [AD97] Apfelbaum, Larry; Doyle, John: Model Based Testing. In: Software Quality Week Conference. S. 296–300, 1997.
- [Br06] Broy, Manfred: Challenges in Automotive Software Engineering. In: Proceedings of the 28th international conference on Software engineering. ACM, S. 33–42, 2006.
- [BSF03] Boger, Marko; Sturm, Thorsten; Fragemann, Per: Refactoring Browser for UML. In (Aksit, Mehmet; Mezini, Mira; Unland, Rainer, Hrsg.): Objects, Components, Architectures, Services, and Applications for a Networked World: International Conference NetObject-Days, NODe 2002 Erfurt, Germany, October 7–10, 2002 Revised Papers. Springer Berlin Heidelberg, S. 366–377, 2003.
- [CST02] Costal, Dolores; Sancho, Maria-Ribera; Teniente, Ernest: Understanding Redundancy in UML Models for Object-Oriented Analysis. In (Pidduck, Anne Banks; Ozsü, M. Tamer;

Mylopoulos, John; Woo, Carson C., Hrsg.): *Advanced Information Systems Engineering: 14th International Conference, CAiSE 2002 Toronto, Canada, May 27–31, 2002 Proceedings*. Springer Berlin Heidelberg, S. 659–674, 2002.

- [Da01] och Dag, Johan Natt; Regnell, Björn; Carlshamre, Pär; Andersson, Michael; Karlsson, Joachim: *Evaluating Automated Support for Requirements Similarity Analysis in Market-Driven Development*. In: *Proc. 7th Int. Workshop on Requirements Engineering: Foundation for Software Quality (REFSQ'01)*. 2001.
- [LEM02] Liu, WenQian; Easterbrook, Steve; Mylopoulos, John: *Rule-Based Detection of Inconsistency in UML Models*. In: *Workshop on Consistency Problems in UML-Based Software Development*. Jgg. 5, 2002.
- [Mi56] Miller, George A: *The magical number seven, plus or minus two: some limits on our capacity for processing information*. *Psychological review*, 63(2):81, 1956.
- [Su01] Sunyé, Gerson; Pollet, Damien; Traon, Yves; Jézéquel, Jean-Marc: *Refactoring UML Models*. In (Gogolla, Martin; Kobryn, Cris, Hrsg.): *UML 2001 — The Unified Modeling Language. Modeling Languages, Concepts, and Tools: 4th International Conference Toronto, Canada, October 1–5, 2001 Proceedings*. Springer Berlin Heidelberg, S. 134–148, 2001.
- [UM15] UML, OMG: *UML 2.5 Specification*. URL <http://www.omg.org/spec/UML/2.5/>, 2015.

Automatische Klassifikation von Anforderungen zur Unterstützung von Qualitätssicherungsprozessen

Jonas Winkler¹

Abstract: Während des Requirements-Engineering werden Anforderungen an zu entwickelnde Systeme und Komponenten in Form von Lastenheften dokumentiert. Diese Lastenhefte enthalten neben rechtlich relevanten Anforderungen weitere Inhalte wie Erklärungen, Zusammenfassungen und Abbildungen. Um zwischen Anforderungen und solchen Zusatzinformationen unterscheiden zu können, werden alle Lastenheftinhalte manuell als Anforderung oder Information eingestuft. Analysen haben ergeben, dass diese Klassifikation nicht konsequent durchgeführt wird. Daher liegt es nahe, diese Klassifikation zu automatisieren. In diesem Beitrag wird ein Ansatz vorgestellt, der diese Klassifikation mithilfe von Techniken aus dem Bereich Data Mining und Machine Learning automatisch durchführen kann.

Keywords: Anforderungsmanagement, Qualitätssicherung, Data Mining, Machine Learning, Textklassifikation, Text-Clustering, Neuronale Netzwerke

1 Motivation

Im Rahmen des Entwicklungsprozesses von Fahrzeugen werden Anforderungen an neu entwickelte Fahrzeugkomponenten in Form von Lastenheften dokumentiert. Diese Dokumente dienen dem Zweck, die Verbindlichkeiten zwischen Auftraggeber (Automobilhersteller) und Auftragnehmer (Lieferant) festzulegen. Vor der Übergabe eines Lastenhefts an einen Lieferanten durchläuft dieses einen intensiven Qualitätssicherungsprozess. Dabei wird unter anderem darauf geachtet, dass die Anforderungen vollständig und eindeutig formuliert sind.

Neben Anforderungen enthalten Lastenhefte weitere Informationen wie beispielsweise Erklärungen, Zusammenfassungen, Beispiele, Abbildungen und Verweise auf andere Dokumente. Diese Elemente sind keine Anforderungen, die der Lieferant in seinem Produkt umsetzen muss. Um zwischen solchen Zusatzinformationen und rechtlich verbindlichen Anforderungen zu unterscheiden, werden alle Elemente eines Dokuments mit speziellen Markierungen versehen, die ein Objekt entweder als Anforderung oder als Information kennzeichnen.

Durch die Analyse mehrerer realer Lastenhefte wurde festgestellt, dass diese Markierung nur selten konsequent durchgeführt wird. Bei der Lastenhefterstellung wird das Attribut oftmals nicht beachtet und die nachträgliche Bearbeitung stellt einen erheblichen Aufwand dar. Dieser Umstand erschwert die Weiterverwendung eines Dokuments in späteren Entwicklungsschritten, beispielsweise bei der Erstellung der Testspezifikation zum Lastenheft.

¹ Technische Universität Berlin, DCAITI, Ernst-Reuter-Platz 7, 10587 Berlin, jonas.winkler@dcaiti.com

In diesem Beitrag wird untersucht, inwieweit es möglich ist, die Klassifizierung von Anforderungen und Informationen mithilfe computergestützter Werkzeuge automatisch durchzuführen. Mithilfe eines solchen Werkzeugs wäre es möglich, sowohl nicht klassifizierte Dokumente zu klassifizieren, als auch in bereits klassifizierten Dokumenten den Autor auf möglicherweise falsch klassifizierte Dokumententeile hinzuweisen.

Dazu wurde zunächst eine Feldstudie an mehreren Anforderungsdokumenten durchgeführt, um die Machbarkeit und mögliche Grenzen besser einschätzen zu können und um die Entwicklung eines Ansatzes zur automatischen Klassifizierung planen zu können. Diese Studie, ein daraufhin entwickelter Ansatz und erste Evaluationsergebnisse werden in diesem Beitrag vorgestellt.

2 Grundlagen

In diesem Beitrag wird mit Dokumenten gearbeitet, welche in der Dokumentendatenbank IBM Rational DOORS² gepflegt werden. In dieser Datenbank besteht jedes *Dokument* aus mehreren *Objekten*. Jedes Objekt eines Dokuments enthält jeweils einen atomaren Dokumentenbaustein wie beispielsweise einen Satz oder eine Abbildung.

Jedes Objekt eines Dokuments besitzt *Attribute*. Diese können von Autoren frei definiert werden. Häufig verwendete und für diesen Beitrag relevante Attribute sind die folgenden:

- **Text:** In diesem Attribut wird der Textkörper des Objekts abgelegt.
- **Schlüssel:** Der Schlüssel identifiziert ein Objekt innerhalb eines Dokuments eindeutig. Wird ein Dokument kopiert, bleiben die Schlüssel erhalten. So kann ermittelt werden, welche Objekte eines Dokuments neu erstellt wurden und welche aus einem anderen Dokument stammen.
- **Typ:** In diesem Attribut wird der Typ des Objekts gespeichert. Für diesen Beitrag sind folgende Typen relevant:
 - **Anforderung:** Der Text des Objekts ist eine verbindliche Anforderung und vom Lieferanten umzusetzen.
 - **Information:** Das Objekt enthält lediglich ergänzende, nicht verbindliche Zusatzinformationen.

3 Studie

Untersucht wurden mehrere deutsche Lastenhefte eines OEMs von Komponenten aus dem Fahrzeuginnenraum. Die Art des Inhalts ist bei allen Dokumenten identisch, geschrieben wurden die Lastenhefte allerdings von unabhängigen Personen aus verschiedenen Abteilungen. Anhand dieser Lastenhefte wurden folgende Beobachtungen gemacht:

² <http://www.ibm.com/software/products/de/ratidoor>

1. Viele Lastenheftbestandteile wie beispielsweise die Kapitelstruktur, Erklärungen zu den Lastenheftinhalten und komponentenunabhängige Anforderungen sind in allen Dokumenten nahezu identisch. Dies liegt daran, dass Dokumente stets auf Basis einer gemeinsamen Vorlage erstellt werden. Für alle aus einer Vorlage stammenden Objekte steht die Klassifizierung fest und darf nicht geändert werden.
2. Die meisten komponentenspezifischen Lastenheftinhalte werden in natürlicher Sprache verfasst. Es wurde beobachtet, dass für Anforderungen generell eher präzisere Formulierungen verwendet werden, als für Zusatzinformationen. Weiterhin werden bestimmte Inhalte wie Verweise auf andere Dokumente und Erläuterungen zu einem Kapitel grundsätzlich immer als Information klassifiziert.
3. Einige Lastenheftinhalte werden nicht ausformuliert, sondern in Form von Schlüssel-Wert-Paaren, Aufzählungen, Tabellen und Abbildungen dokumentiert. Einige Inhalte wie Signalspezifikationen und Kontaktflächenspezifikationen werden stets nach dem gleichen Muster erstellt und immer gleich klassifiziert.
4. In vielen Dokumenten ließen sich wiederholt Formulierungen und Formatierungen finden, die für das jeweilige Dokument einzigartig sind. Dies lässt sich auf die unterschiedlichen Autoren der Dokumente zurückführen.
5. Die Klassifizierung wurde in einigen Lastenheften nicht konsequent durchgeführt. Teilweise wurden Hinweise, die offensichtlich keine Anforderungen sind, nicht als Information markiert, teilweise wurde die Klassifizierung gar nicht vorgenommen.

Zur Veranschaulichung werden im Folgenden einige markante Beispiele gegeben. Beispiele für die in (2) genannten natürlichsprachlichen Objekte sind die folgenden Sätze:

- (Anforderung) Es muss ein Resetzähler implementiert werden.
- (Anforderung) Das Relais darf bei einem Spannungseinbruch nicht abfallen.
- (Information) Weitere Informationen sind dem Kapitel „A“ zu entnehmen.
- (Information) Das Schaltbild zeigt schematisch die Bauteile dieser Komponente.

Anforderungen können meist an präzisen Formulierungen („muss“, „darf nicht“) erkannt werden. Informationen werden meist mit anderen Formulierungen geschrieben („Weitere Informationen“, „zeigt schematisch“).

Beispiele für nicht natürlichsprachliche Objekte in (3) sind Pinnings (Spezifikation der Schnittstelle einer elektronischen Komponente) oder auch Ansprechpartner:

Pin Nr: 3	Ansprechpartner
Pin Bezeichnung: Kontrollsignal	Name, Vorname: Mustermann, Max
Nenn-Spannung: 12 V	Abteilung: ABC12
Nenn-Strom: 0,1 A	Telefon: +49 123 456789
Max-Strom: 0,2 A	E-Mail: max@mustermann.de

Alle Objekte, in denen ein Pin spezifiziert wird, werden grundsätzlich als Anforderung klassifiziert. Analog dazu werden Ansprechpartner grundsätzlich als Information klassifiziert. Für jedes dieser Muster lässt sich meist eindeutig eine Klassifizierung bestimmen.

Da die Lastenheftinhalte in sehr unterschiedlichen Formaten vorliegen und die notwendigen Merkmale zur Diskriminierung zwischen Anforderung und Information sehr unterschiedlich sind, kann ein einziger Klassifizierungsalgorithmus keine guten Ergebnisse liefern. Vielmehr muss zwischen den unterschiedlichen Formaten unterschieden werden und je nach Format ein spezialisiertes Klassifizierungsverfahren eingesetzt werden. Innerhalb der betrachteten Dokumente wurden vorrangig folgende Formate identifiziert:

- **Überschrift.** Das Objekt ist eine Überschrift.
- **Satz.** Das Objekt enthält einen oder mehrere wohlgeformte, natürlichsprachliche Sätze.
- **Aufzählung.** Das Objekt enthält eine Aufzählung oder Nummerierung verschiedener Begriffe.
- **Definition.** Eine Definition umfasst eine oder mehrere Zeilen, die jeweils dem Muster „<Schlüssel>: <Wert>“ folgen.
- **Abbildung.** Das Objekt enthält ein Foto, eine schematische Darstellung oder sonstige Abbildung.
- **Tabelle.** Das Objekt enthält eine Tabelle.

Diese Typisierung wird in den weiteren Abschnitten *Strukturtyp* genannt. Auf Basis der getroffenen Beobachtungen wird im Folgenden ein Ansatz zur automatischen Klassifizierung vorgestellt.

4 Ansatz zur automatischen Klassifikation

Alle Objekte eines Dokuments werden einzeln klassifiziert. Jedes Objekt durchläuft während der Klassifizierung mehrere Klassifizierungsschritte. Der Ablauf ist in Abbildung 1 schematisch dargestellt. Jeder Schritt (in der Abbildung grau hinterlegt) untersucht ein Objekt und versucht, mithilfe geeigneter Techniken eine Klassifizierung durchzuführen. Ein Schritt kann ein Objekt entweder in eine Klasse einordnen oder keine Aussage über die Klassifizierung treffen. Wird ein Objekt durch einen Schritt erfolgreich klassifiziert, endet die Klassifizierung mit diesem Schritt. Ist durch einen Schritt keine Aussage über die Klassifikation möglich, wird der nächste Schritt ausgeführt.

Schritte, die sich nur auf wenige Objekte anwenden lassen, dafür allerdings eine sehr hohe Genauigkeit haben, werden zuerst ausgeführt. Schritte, die sich auf sehr viele Objekte anwenden lassen und eine geringere Genauigkeit haben, werden später ausgeführt. Dies hat zur Folge, dass einfach zu klassifizierende Objekte durch einfache Klassifizierungstechniken richtig klassifiziert werden und nicht durch komplizierte Techniken möglicherweise falsch klassifiziert werden.

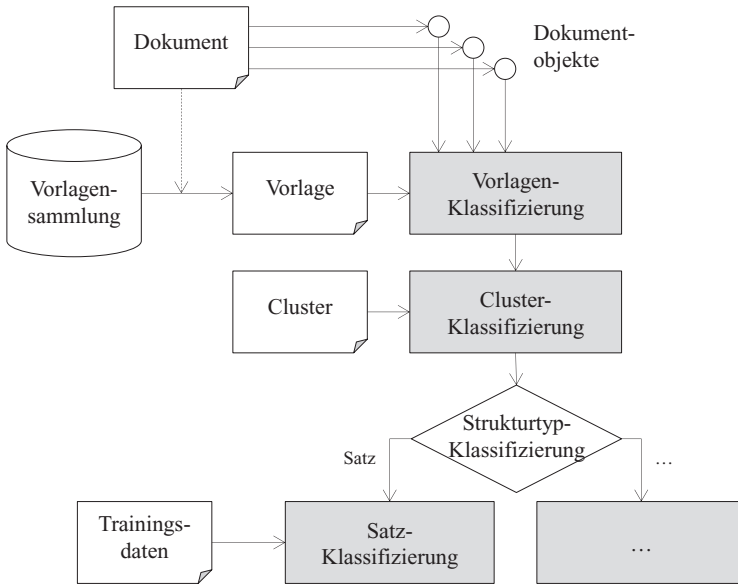


Abb. 1: Übersicht über Klassifizierungsansatz

Die Vorlagen-Klassifizierung überprüft Objekte, die aus einer Vorlage kopiert wurden und klassifiziert entsprechend der Vorlage. Die Cluster-Klassifizierung sucht nach textuell ähnlichen und identisch klassifizierten Objekten und benutzt die daraus entstehenden Cluster zur Klassifizierung häufig wiederkehrender Muster.

Danach wird abhängig vom Strukturtyp eines Objekts ein speziell auf den Strukturtyp ausgerichtetes Klassifizierungsverfahren angewendet. Objekte, deren Strukturtyp nicht bestimmt werden kann, werden nicht weiter klassifiziert. Dadurch werden schwierig zu klassifizierende Objekte und Objekte mit nicht eindeutiger Klassifikation übersprungen. Derzeit existiert lediglich für den Strukturtyp Satz ein Klassifizierungsverfahren.

4.1 Klassifizierung nach Vorlage

In diesem Schritt werden Objekte, welche aus einer Vorlage stammen, identifiziert und klassifiziert. Dazu wird zuerst die dem Dokument zugrundeliegende Vorlage identifiziert und aus einer Vorlagensammlung geholt. Diese Vorlage enthält von allen Objekten der Vorlage deren Text, Typ und Schlüssel.

Anhand des Schlüssels und des Texts eines zu klassifizierenden Objekts wird ermittelt, ob es aus der Vorlage kopiert wurde. Dies ist möglich, da der Schlüssel bei der Kopie eines Objekts nicht verändert wird. Ist dies der Fall, wird das Objekt gemäß der Vorlage klassifiziert. Alle Objekte, welche nicht aus einer Vorlage kopiert wurden, werden durch diesen Schritt nicht klassifiziert.

4.2 Klassifikation durch Clustering

Während der Studie in Abschnitt 3 wurde festgestellt, dass viele Arten von Inhalten (bspw. Pinning, Ansprechpartner) meist nach einem einheitlichen Muster dokumentiert werden. Weiterhin werden alle Objekte einer Art immer gleich klassifiziert. Es liegt also nahe, wiederkehrende Muster zu erkennen und entsprechend der für das jeweilige Muster üblichen Klassifizierung zu klassifizieren.

Dazu wurde zunächst ein hierarchischer Complete-Link-Clusterer auf eine sehr große Anzahl bereits klassifizierter Objekte aus vielen Dokumenten angewendet. Als Textdistanzmaß wurde die Levenshtein-Distanz verwendet. Dieses Verfahren bildet aus einer Menge von Zeichenketten Cluster, wobei die Distanz jeweils aller Paare zweier Zeichenketten innerhalb eines Clusters nicht größer ist als ein festgelegter Schwellenwert. Die Elemente eines Clusters sehen sich daher sehr ähnlich. Details zu den verwendeten Verfahren werden in [AZ12b] beschrieben.

Ein geeigneter Schwellenwert für das Textdistanzmaß wurde empirisch ermittelt.

Die erzeugten Cluster enthielten sowohl die erwarteten Cluster für Muster wie Pinning und Signalspezifikation, allerdings auch Cluster ähnlicher, aber sehr unterschiedlich klassifizierter Sätze. Weiterhin enthielten einige Cluster Objekte aus lediglich einem oder wenigen Dokumenten. Solche Cluster sind zur Erkennung von Mustern in anderen Dokumenten nicht hilfreich. Daher wurden die erkannten Cluster mithilfe zweier Metriken gefiltert.

Sei C ein Cluster. Sei $c_{types}(C)$ eine Liste, wobei diese Liste so viele Elemente hat, wie in C einzigartige Klassifizierungen vorkommen. Jedes Element stellt die Anzahl der Vorkommen einer Klassifizierung dar. Bei einem Cluster C , in dem 10 Objekte als Anforderung, 2 Objekte als Information und 1 Objekt nicht klassifiziert sind, ist beispielsweise $c_{types}(C) = [10, 2, 1]$. Analog dazu sei $c_{sources}(C)$ eine Liste, wobei jedes Element die Anzahl der Objekte enthält, welche aus demselben Dokument stammen. Bei einem Cluster C mit 4 Objekten aus Dokument A und 6 Objekten aus Dokument B wäre somit $c_{sources}(C) = [4, 6]$. Mithilfe der Funktion

$$f(L) = \frac{\sum L}{\max L}$$

kann ein Cluster C nun wie folgt bewertet werden:

- Ist $f(c_{types}(C)) = 1$, wird für alle Objekte die gleiche Klassifizierung verwendet. Je größer der Wert von f , desto mehr unterschiedliche Klassifizierungen werden verwendet.
- Ist $f(c_{sources}(C)) = 1$, stammen alle Objekte eines Clusters aus demselben Dokument. Je größer der Wert von f , aus desto mehr unterschiedlichen Dokumenten stammen die Objekte eines Clusters.

Wünschenswert sind Cluster mit kleinem $f(c_{types}(C))$ und hohem $f(c_{sources}(C))$. Schwellenwerte wurden wiederum empirisch ermittelt. Einige der erkannten Cluster sind in Tabelle 1 dargestellt. Neben den erwarteten Clustern für Signalspezifikationen, Pinning und Ansprechpartner wurden auch viele Cluster gefunden, die ähnlich formulierte Sätze zusammenfassen.

Clustername	Anzahl	$f(c_{types}(C))$	$f(c_{sources}(C))$	Klassifikation
Ansprechpartner	177	1.208	4.750	Information
Signalspezifikation	84	1.201	3.448	Anforderung
Pinning	130	1.127	4.057	Anforderung
Kontaktierung	60	1.089	4.357	Anforderung
Verweis auf anderes Dokument	59	1.594	2.204	Information
Identifikationsnummer	198	1.200	2.250	Information

Tab. 1: Erkannte Cluster

Auf Basis dieser Cluster kann nun ein einfaches und effektives Klassifizierungsverfahren konstruiert werden. Soll ein Objekt klassifiziert werden, wird die Distanz des Texts des Objekts zu jedem Cluster ermittelt. Dabei wird dasselbe Textdistanzmaß eingesetzt, welches auch zur Erstellung der Cluster verwendet wurde. Das Objekt wird demjenigen Cluster zugeordnet, zu dem die Distanz am kleinsten ist. Dabei werden allerdings nur Cluster betrachtet, zu denen die Distanz kleiner ist, als der während der Cluster-Erstellung festgelegte Distanz-Schwellwert. Wird ein Cluster gefunden, wird das Objekt entsprechend der im Cluster am häufigsten vorkommenden Klassifizierung klassifiziert. Wird kein passendes Cluster gefunden, trifft dieser Klassifizierungsschritt keine Aussage über die Klassifizierung.

4.3 Strukturtyp-Klassifizierung

Zur weiteren Klassifikation muss der Strukturtyp eines Objekts ermittelt werden. Dazu wurde ein einfaches Verfahren entwickelt, welches ein Objekt schrittweise auf bestimmte Merkmale untersucht:

1. Besitzt das Objekt eine Kapitelnummer, wird es als Überschrift eingestuft.
2. Enthält das Objekt eine Grafik, so wird es als Abbildung eingestuft.
3. Enthält das Objekt eine Tabelle, so wird es als Tabelle eingestuft.
4. Umfasst das Objekt mehrere Zeilen und beginnt ein Großteil der Zeilen mit demselben Zeichen oder beginnt ein Großteil der Zeilen mit den für Nummerierungen typischen Zeichen (1, 2, 3, ...; a, b, c,...), so ist es eine Aufzählung.
5. Umfasst das Objekt eine oder mehrere Zeilen, die dem Muster „<Schlüssel>: <Wert>“ folgen, so ist das Objekt eine Definition.

Alle weiteren Objekte sind potentiell Sätze oder schwer zu klassifizierende Sonderformate. Zur Erkennung von Sätzen wird der Text der übriggebliebenen Anforderungen mithilfe des

Stanford-Parsers [KM03] analysiert. Dieser ordnet jedem Wort ein sogenanntes Part-of-Speech-Tag zu und erkennt unter anderem Sätze, Nominalphrasen und Adverbialphrasen. Alle Objekte, deren Text durch den Stanford-Parser als vollständiger Satz eingestuft wird (erkennbar an den Markierungen „S“ und „CS“ für *Sentence* und *Coordinated Sentence*), werden im weiteren Verlauf als Satz behandelt.

Alle durch diese Schritte nicht erfassten Objekte werden derzeit nicht betrachtet.

4.4 Satz-Klassifikation

Zur Klassifikation von Sätzen werden Convolutional Neural Networks (CNN) eingesetzt [Ki14]. Diese bieten gegenüber etablierten Textklassifizierungsverfahren wie beispielsweise Naive Bayes [AZ12a] und Support-Vector-Machines [Kw98] die folgenden Vorteile:

- Andere Verfahren nutzen meist die Bag-of-Words-Technik. Bei dieser gehen Informationen über die Reihenfolge der Wörter verloren. Dies ist bei CNNs nicht der Fall.
- Durch Verwendung eines Wortvektorisierungsverfahrens (hier: word2vec [Mi13]) können auch Muster erkannt werden, bei denen die verwendeten Wörter leicht unterschiedlich sind.
- Im Vergleich zu anderen Verfahren hat diese Art von neuronalen Netzwerken bei diversen Textklassifizierungsaufgaben bessere Ergebnisse erzielt [KGB14].

Das Netzwerk wurde gemäß den in [ZW15] gegebenen Hinweisen erstellt. Die verwendete Netzwerkarchitektur ist mit der dort vorgestellten Architektur identisch. Die Vorbereitung der Trainingsdaten, die generelle Funktionsweise und der Trainingsprozess des Netzwerks werden im Folgenden kurz beschrieben. Für detaillierte Erklärungen zur Funktionsweise von CNNs sollte [Ki14] und [ZW15] konsultiert werden.

Ausgangsbasis sind alle Objekte aus sämtlichen zur Verfügung stehenden Anforderungsdokumenten, die durch die Strukturtyp-Klassifizierung als Satz eingestuft wurden und entweder als Anforderung oder Information klassifiziert wurden. Die Trainingsmenge umfasst etwa 10000 Sätze aus 89 Dokumenten. Alle Sätze werden vorverarbeitet, um das Training des Netzwerks zu vereinfachen: Klammersausdrücke und Sonderzeichen werden entfernt, Großbuchstaben werden in Kleinbuchstaben umgewandelt. Die daraus resultierende Datenmenge wird zur Erstellung einer Word-Vektor-Abbildung gemäß [Mi13] genutzt. Nach diesem Schritt besitzt jedes Wort aus der Datenmenge eine Vektordarstellung und damit jeder Satz eine Matrixdarstellung. Diese Matrizen dienen als Eingabe für das neurale Netzwerk.

Kernbestandteil des Netzwerks sind sogenannte Filter: Diese haben eine festgelegte Länge und werden wie ein Fenster über die Matrix eines Satzes geschoben. Diese Filter ermöglichen es dem Netzwerk, Muster in Wortfolgen entsprechend der Länge der Filter zu erlernen. Für diese Arbeit wurden die Filterlängen 2, 3, 4, 5 und jeweils 64 Filter pro Filterlänge verwendet.

Das Netzwerk besitzt für die zwei Klassen Information und Anforderung zwei Ausgangsneuronen. Diese sind jeweils vollständig mit allen Filtern verbunden. Über diese Neuronen zeigt das Netzwerk die Klassifizierung einer Eingabe an.

Zum Training des Netzwerks wurde wiederum die vorverarbeitete Datenmenge verwendet. Diese wurde in 90% Trainingsdaten und 10% Testdaten geteilt. Das Netzwerk wurde mithilfe der Trainingsdaten solange optimiert, bis es die Trainingsdaten zu 100% korrekt abbilden konnte.

Die Leistung des Netzwerks zur Klassifikation von Sätzen wurde mithilfe der Testdaten evaluiert. Dazu ist in Abbildung 2 die Entwicklung der Genauigkeit der Klassifikation der Trainingsdaten (blau) und Testdaten (rot) während des Trainingsprozesses dargestellt. Nach etwa 1000 Trainingsschritten ist das Netzwerk in der Lage, die Trainingsdaten zu 100% korrekt zu klassifizieren. Die dabei aus den Trainingsdaten erlernten Muster ermöglichen es dem Netzwerk, insgesamt ca. 81% der Testdaten richtig zu klassifizieren.

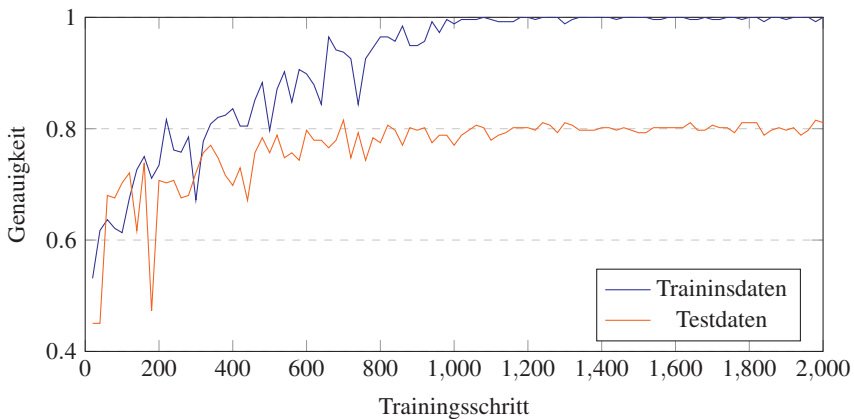


Abb. 2: Evaluation Lernprozess CNN

Zur Klassifizierung eines Objekts mithilfe des trainierten Netzwerks wird dieses genauso wie die Trainingsdaten vorverarbeitet, mithilfe der Wort-Vektor-Abbildung in eine Matrix konvertiert und als Eingabe für das Netzwerk verwendet. Das Netzwerk zeigt an den Ausgangsneuronen die Klassifizierung an.

5 Evaluation

Zur Evaluation wurde das Verfahren auf ein ausgewähltes Lastenheft angewendet. Dieses wurde weder zum Training verwendet, noch waren die Inhalte während der Konstruktion des Ansatzes bekannt. Die Ergebnisse jedes einzelnen Klassifizierungsschritts werden im Folgenden vorgestellt.

Das Lastenheft enthält insgesamt 747 Objekte. 20,7% dieser Objekte sind Überschriften. Insgesamt wurden 3,6% durch die Vorlagen-Klassifizierung, 11,6% durch die Cluster-

Klassifizierung und 41,5% durch die Satz-Klassifizierung abgedeckt. Die verbleibenden 22,6% konnten nicht klassifiziert werden.

Zur Evaluation der Strukturtyp-Klassifizierung wurde der Strukturtyp aller Objekte des Lastenhefts manuell gemäß der in Abschnitt 3 definierten Strukturtypen bestimmt und mit der automatischen Klassifizierung verglichen. Die Ergebnisse sind in Tabelle 2 dargestellt. Die Klassifizierung von Überschriften ist erwartungsgemäß fehlerfrei. Die Genauigkeit der Erkennung von Sätzen ist akzeptabel, könnte durch den Einsatz anderer Verfahren ggf. noch verbessert werden. Die beiden Strukturtypen Aufzählung und Definition konnten nur mit mäßiger Genauigkeit erkannt werden, dies liegt vor allem an inkonsistenten und teilweise sehr unterschiedlichen Formatierungen im Lastenheft. Für die restlichen 136 von insgesamt 747 Objekten konnte kein Strukturtyp ermittelt werden.

Strukturtyp	Anzahl	Recall	Precision	f1-Score
Überschrift	155	1.0	1.0	1.0
Satz	383	0.937	0.970	0.954
Aufzählung	13	1.0	0.813	0.897
Definition	60	0.817	0.645	0.721
Unbekannt	136			

Tab. 2: Evaluation Strukturtyp-Erkennung

Zur Evaluation der Genauigkeit der einzelnen Klassifizierungsschritte wurde die von den Schritten ermittelte Klassifizierung mit der tatsächlichen Klassifizierung verglichen. Jeder Schritt wurde dabei anhand der Objekte evaluiert, die durch den jeweiligen Schritt klassifiziert wurden. Die Ergebnisse sind in Tabelle 3 dargestellt.

Klassifizierer	Typ	Recall	Precision	f1-Score
Vorlage	Anforderung	1.0	1.0	1.0
	Information	1.0	1.0	1.0
Cluster	Anforderung	0.982	0.964	0.973
	Information	0.938	0.968	0.952
CNN	Anforderung	0.828	0.951	0.885
	Information	0.870	0.626	0.728
Gesamt	Anforderung	0.862	0.955	0.906
	Information	0.906	0.737	0.813

Tab. 3: Evaluation Klassifizierungsschritte

Die Genauigkeit der Vorlagenklassifizierung ist erwartungsgemäß perfekt. Von den 747 Objekten werden allerdings nur 27 durch diesen Schritt abgedeckt. Die Cluster-Klassifizierung hat im Lastenheft einige Muster und viele Sätze erkannt, die häufig in anderen Dokumenten in ähnlicher Form vorkommen und hat die Klassifizierung übernommen. Dies funktioniert erstaunlich gut. Die Satz-Klassifikation mittels CNNs hat 83% der klassifizierten Objekte richtig klassifiziert. Dies deckt sich mit der in Abschnitt 4.4 an den Testdaten ermittelten Genauigkeit. Ein Problem ist die niedrige Präzision von Informationen: Viele Anforderungen wurden durch das Netzwerk als Information klassifiziert. Die drei Klassifizierungsschritte konnten insgesamt 87.5% der Objekte richtig klassifizieren.

Die Evaluation hat gezeigt, dass die Klassifikation von Sätzen mittels CNNs nur mäßig funktioniert. Um die Stärken und Schwächen der Klassifizierung von Sätzen durch CNNs einschätzen zu können, werden in Tabelle 4 einige prägnante Beispiele gegeben, bei denen die Klassifikation besonders gut oder schlecht funktioniert.

Satz	Typ	Korrekt?	Ausgabe Inf / Anf
Der Wandler muss entsprechend der Strategie in den Selbstschutz gehen.	Anf	ja	-1.655 / 0.986
Der Strom des Heizelements muss mit einer Auflösung von ... gemessen werden .	Anf	ja	-3.860 / 1.483
Für die LED ist das Signal ... zu verwenden.	Anf	ja	-1.701 / 0.343
Diese Funktion ist in der Systembeschreibung ... im Detail beschrieben.	Inf	ja	3.386 / -4.359
Kurzschlüsse im Geräteinnern durch mutwilligen Eingriff dürfen zur Zerstörung führen.	Inf	nein	-0.919 / -0.551
Die Funktionsanteile für die Komponente sind im Systemlastenheft ... beschrieben.	Inf	nein	-0.898 / 0.302

Tab. 4: Beispiele

Die Tabelle enthält zu jedem Satz die jeweils richtige Klassifizierung und die Information, ob das CNN den Satz richtig klassifizieren konnte. Die letzte Spalte enthält die Ausgabe des Netzwerks für die beiden Klassen Information und Anforderung. Der jeweils größere Wert ist ausschlaggebend für die Ausgabe des Netzwerks. Je größer / kleiner eine Ausgabe, desto sicherer ist sich das Netzwerk, dass der Satz der jeweiligen Klasse angehört / nicht angehört.

Beispiel 1 und 2 konnten durch das CNN richtig klassifiziert werden. Anhaltspunkte für die richtige Klassifizierung dieser Beispiele sind möglicherweise die Wörter und Wortgruppen „muss“, „gemessen werden“ oder auch „entsprechend der Strategie“. Die Klassifizierung durch das Netzwerk ist bei diesen beiden Beispielen glaubwürdig, da die Ausgabewerte hinreichend groß bzw. klein sind.

Beispiel 3 wurde ebenfalls richtig klassifiziert. Die Ausgabe des Netzwerks für die Klasse Anforderung ist bei diesem Beispiel allerdings deutlich kleiner. Ursache dafür ist möglicherweise das Nichtvorhandensein ausschlaggebender Formulierungen im Satz. Die Glaubwürdigkeit dieses Ergebnisses muss daher hinterfragt werden.

Beispiel 4 ist eine Zusatzinformation, die durch das CNN mit großer Sicherheit als solche erkannt wurde. Angezeigt wird dies durch die sehr großen bzw. kleinen Ausgabewerte.

Beispiel 5 ist ebenfalls eine Zusatzinformation, die durch das Netzwerk fälschlicherweise als Anforderung eingestuft wird. Bei Betrachtung der Ausgabewerte fällt allerdings auf, dass keine der beiden Klassen auf diesen Satz zutreffen. Dieses Ergebnis deutet auf ungünstig gewählte Formulierungen des Satzes hin, oder auch auf eine zu kleine Trainingsdatenmenge.

Der in Beispiel 6 gegebene Satz wurde ebenfalls falsch klassifiziert, obwohl er große Ähnlichkeit zu dem sehr gut klassifizierbaren Beispiel 4 besitzt. Die Gründe dafür müssen weiter untersucht werden.

Anhand dieser Beispiele lässt sich erkennen, dass bei der Klassifikation durch CNNs in vielen Fällen erkennbar ist, ob das Ergebnis der Klassifikation glaubwürdig ist. Allerdings ist es unklar, welche Teile der Sätze für die jeweilige Klassifikation ausschlaggebend sind, da das Netzwerk hierüber keine Informationen preisgibt. Ob und wie solche Aussagen anhand des Netzwerks getroffen werden können, muss in weiteren Arbeiten genauer untersucht werden.

Die von keinem Schritt klassifizierten Objekte (22.6%) konnten nicht klassifiziert werden, da entweder der Strukturtyp nicht erkannt wurde, oder für den erkannten Strukturtyp kein Klassifizierungsschritt zur Verfügung steht.

6 Verwandte Arbeiten

Die Autoren von [HLR14] stellen ein graphisches Werkzeug zur Analyse von Anforderungen vor und haben für dieses Werkzeug mehrere Komponenten zur Klassifikation von Anforderungen entwickelt. Das Werkzeug kann unter anderem temporale Anforderungen von nicht temporalen Anforderungen unterscheiden. Auf Basis von Fallstudien zeigen die Autoren, dass Techniken aus dem Bereich Machine Learning zur Lösung von Problemen des Requirements-Engineering grundsätzlich anwendbar sind.

In [Ot13] wird ebenfalls ein Ansatz zur Klassifizierung von Anforderungen vorgestellt. In diesem Beitrag werden Naive-Bayes und Support Vector Machines zur Bestimmung des Themas von Anforderungen eingesetzt, um Reviewprozesse zu unterstützen.

7 Zusammenfassung und Ausblick

In diesem Beitrag wurde ein Ansatz zur Klassifikation der Objekte von Lastenheften in Anforderung und Information vorgestellt. Der Ansatz kombiniert unterschiedliche Klassifizierungstechniken, um eine möglichst genaue Klassifizierung zu gewährleisten.

Die vorgestellte Cluster-Klassifizierung sucht nach textuell ähnlichen, identisch klassifizierten Objekten und benutzt die daraus entstehenden Cluster zur Klassifizierung. Die erzielten Ergebnisse sind bereits sehr gut, ggf. können allerdings mit anderen Cluster-Techniken oder Textdistanzmaßen noch bessere Ergebnisse erzielt werden.

Ein weiterer Bestandteil des Ansatzes ist die Unterscheidung von Strukturtypen: Abhängig vom Format der in einem Objekt vorkommenden Inhalte wird ein geeigneter Klassifizierungsschritt ausgewählt. Dies wurde für den Strukturtyp Satz gezeigt. Es muss weiter untersucht werden, wie andere Strukturtypen klassifiziert werden können und wie die Erkennung unterschiedlicher Strukturtypen ggf. durch den Einsatz neuraler Netzwerke verbessert werden kann, um noch mehr Objekte klassifizieren zu können.

Die Klassifizierung von Sätzen mittels Convolutional Neural Networks liefert akzeptable Ergebnisse. Größtes Verbesserungspotential liegt hier in der Trainingsdatenmenge: Durch weitere Vorverarbeitung und Filterung der Datenmenge kann das Netzwerk möglicherweise besser lernen.

Neben der Verbesserung der verwendeten Techniken ist der nächste Schritt die Integration in ein GUI-Werkzeug zur Benutzerunterstützung. Dieses soll den Benutzer auf falsch klassifizierte Objekte hinweisen. Für den Nutzer ist dabei neben einer hohen Genauigkeit wichtig, dass gegebene Hinweise zu falsch klassifizierten Objekten begründet werden. Insbesondere bei neuronalen Netzwerken ist eine derartige Begründung schwierig, da der Entscheidungsprozess innerhalb des Netzwerks sehr komplex ist. Im Rahmen einer Werkzeugintegration muss dieses Problem weiter untersucht werden.

Literaturverzeichnis

- [AZ12a] Aggarwal, Charu C.; Zhai, ChengXiang: A Survey of Text Classification Algorithms. In (Aggarwal, Charu C.; Zhai, ChengXiang, Hrsg.): Mining Text Data, S. 163–222. Springer US, 2012.
- [AZ12b] Aggarwal, Charu C.; Zhai, ChengXiang: A Survey of Text Clustering Algorithms. In (Aggarwal, Charu C.; Zhai, ChengXiang, Hrsg.): Mining Text Data, S. 77–128. Springer US, 2012.
- [HLR14] Hayes, Jane H.; Li, Wenbin; Rahimi, Mona: Weka meets TraceLab: Toward Convenient Classification: Machine Learning for Requirements Engineering Problems: A Position Paper. In: 1st IEEE International Workshop on Artificial Intelligence for Requirements Engineering. AIRE, S. 9–12, 2014.
- [KGB14] Kalchbrenner, Nal; Grefenstette, Edward; Blunsom, Phil: A Convolutional Neural Network for Modelling Sentences. arXiv preprint, abs/1404.2188, 2014.
- [Ki14] Kim, Yoon: Convolutional Neural Networks for Sentence Classification. arXiv preprint, abs/1408.5882, 2014.
- [KM03] Klein, Dan; Manning, Christopher D.: Accurate Unlexicalized Parsing. In: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1. ACL '03, Association for Computational Linguistics, Stroudsburg, PA, USA, S. 423–430, 2003.
- [Kw98] Kwok, James Tin-yau: Automated Text Categorization Using Support Vector Machine. In: In Proceedings of the International Conference on Neural Information Processing (ICONIP. S. 347–351, 1998.
- [Mi13] Mikolov, Tomas; Chen, Kai; Corrado, Greg; Dean, Jeffrey: Efficient Estimation of Word Representations in Vector Space. arXiv preprint, abs/1301.3781, 2013.
- [Ot13] Ott, Daniel: Automatic Requirement Categorization of Large Natural Language Specifications at Mercedes-benz for Review Improvements. In: Proceedings of the 19th International Conference on Requirements Engineering: Foundation for Software Quality. REFSQ, Springer-Verlag, Berlin, Heidelberg, S. 50–64, 2013.
- [ZW15] Zhang, Ye; Wallace, Byron C.: A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification. arXiv preprint, abs/1510.03820, 2015.

Cloud-Basierte Optimierung von Fahrzeugbetriebsstrategien durch Clustering mit Genetischen Algorithmen

Natalia Ogulenko¹, Sören Frey², Jens Nahm³ und Manfred Rössle⁴

Abstract: Die zunehmende Onlinevernetzung von Fahrzeugen eröffnet Automobilherstellern neue Wege, ihren rechtlichen Verpflichtungen zur Senkung der CO₂-Emissionen nachkommen und entsprechende Wettbewerbsvorteile realisieren zu können. Die für viele Fahrten übermittelten Telemetrie- und Streckendaten können analysiert und zur Senkung des Flottenverbrauchs verwendet werden. Wir stellen einen Ansatz vor, der diese Daten zur Optimierung von Fahrzeugbetriebsstrategien verwendet. In einer Cloud-Anwendung werden die übermittelten Daten anonymisiert gespeichert. Um die für bestimmte Charakteristiken, wie z.B. Motorleistung und Wetterbedingungen, am besten geeignete Betriebsstrategie aus den großen Datenmengen effizient bestimmen zu können, schränkt unser Ansatz den Suchraum mit einer Clusteranalyse ein. Wir verwenden hierzu ein Verfahren auf Basis von genetischen Algorithmen. Ein Cluster enthält eine Menge von möglichst ähnlichen Telemetrie- und Streckendaten. Die für eine bevorstehende Fahrt am besten geeignete Betriebsstrategie wird dann aus Daten des ähnlichsten Clusters bestimmt. Zukünftig könnten die so gefundenen optimalen Betriebsstrategien dynamisch vor Fahrtantritt zum jeweiligen Fahrzeug übertragen werden.

Keywords: Fahrzeugbetriebsstrategien, Optimierung, Clusteranalyse, Genetische Algorithmen

1 Einleitung

Durch die Onlinevernetzung von Fahrzeugen und die Verwendung von übermittelten Daten in Cloud-basierten Big Data Analysen ergeben sich vielfältige Potenziale in weiten Teilen der Automobilindustrie [KW14]. Dieses Papier beschreibt unseren Ansatz [Og16] zur Bestimmung einer optimalen Fahrzeugbetriebsstrategie pro bevorstehender Fahrt aus historischen Fahrten durch Clustering mit genetischen Algorithmen (GAs) in einer Cloud-Anwendung.

¹ Daimler TSS GmbH, Wilhelm-Runge-Straße 11, 89081 Ulm, natalia.ogulenko@daimler.com

² Daimler TSS GmbH, Wilhelm-Runge-Straße 11, 89081 Ulm, soeren.frey@daimler.com

³ Daimler TSS GmbH, Wilhelm-Runge-Straße 11, 89081 Ulm, jens.nahm@daimler.com

⁴ Hochschule Aalen - Technik und Wirtschaft, Wirtschaftsinformatik, Anton-Huber-Straße 25, 73430 Aalen, manfred.roessle@htw-aalen.de

1.1 Kontext

Viele Jahre waren Automobile fahrende "Daten-Inseln", d.h. sie waren in sich geschlossene Systeme. Seit ca. 20 Jahren werden die elektronisch gesteuerten Funktionseinheiten bei Werkstattbesuchen ausgelesen und die somit gesammelten Nutzungs- und Betriebsdaten weiterverwendet. Die seit ca. 10 Jahren eingesetzte Onlinevernetzung der Fahrzeuge wird als disruptiv bezeichnet, weil sie ganz neue Ansätze in vielen verschiedenen Dimensionen eröffnet.

Es steht beispielsweise eine viel größerer Umfang an Datenpunkten mit einer höheren Auflösung als bisher zur Verfügung. Die Erkenntnisse aus der Analyse dieser Daten fließen auch in die Entwicklung und Optimierung der nächsten Fahrzeuggeneration mit ein (Produktentwicklungszyklus ca. fünf Jahre). Darüber hinaus können Verbesserungen und die Beseitigung von Fehlern im Backend erfolgen und stehen dem Endnutzer sofort zur Verfügung (Produktentwicklungszyklus im Bereich von wenigen Tagen bis Wochen möglich).

In das Backend übermittelte Daten können etwa mit Methoden des maschinellen Lernens ausgewertet und mit aktuellen Kontextdaten wie z.B. zu Verkehr-, Wetter- und Straßenverhältnissen oder dem Luftschadstoffgehalt situativ per Onlineabruf aus dem Backend in das Fahrzeug übertragen werden.

Die rechen- und datenintensive Vorverarbeitung sprengt alle heute in Fahrzeugen verfügbaren Möglichkeiten. Die Onlinevernetzung der Fahrzeuge erlaubt verteilte Systemarchitekturen, bei denen die Speicherung der Daten und ein Großteil der Rechenlast in Cloud-basierte Backendsysteme verlagert wird. Durch die hohen Stückzahlen der produzierten Fahrzeuge zahlen sich selbst kleine Kostenoptimierungen pro Fahrzeug schnell aus. Wenn also anstatt eines sehr leistungsfähigen Rechners im Auto eine kleinere Variante ausreicht, die bei Bedarf mit Rechenleistung im Backend angereichert wird, können Kostenvorteile entstehen.

Ausgelagerte Datenanalysen im Backend eröffnen nun auch prinzipiell die Möglichkeit zur zentralen Optimierung von Fahrzeugbetriebsstrategien.

1.2 Problemstellung

Fahrzeugbetriebsstrategien zielen auf den effizienten Betrieb von Fahrzeugen ab (siehe Abschnitt 2.1) und werden bisher hauptsächlich statisch für eine Menge von Fahrzeugen, z.B. für bestimmte Baureihen, während deren Entwicklung optimiert. Das Potential zur dynamischen Optimierung einer Betriebsstrategie pro spezifischer, bevorstehender Fahrt durch eine zentrale Datenanalyse auf Basis von vielen historischen, ähnlichen Fahrten wird meistens noch nicht genutzt.

Eine entsprechende Datenanalyse könnte es z.B. ermöglichen, die energieeffizienteste Betriebsstrategie dynamisch für eine bevorstehende Fahrtroute für das jeweilige

Fahrzeug- und Motormodell und den gewählten Fahrstil (z.B. „Sport“), sowie unter Berücksichtigung der aktuellen Wetter- und Stauprognosen passend aus historischen Fahrtdaten mit ähnlichen Charakteristiken bestimmen zu können.

Formal stellt sich das adressierte Problem und das Ziel einer solchen Datenanalyse als die Minimierung einer Menge von Kriterien K dar. Im Kontext dieses Papiers sollen für eine geplante Fahrt mit einem Fahrzeug F von der Startlokation S (zum Zeitpunkt t_S) zur Ziellokation Z die Kriterien Kraftstoffverbrauch k_{KV} und Fahrzeit k_{FZ} ($K = \{k_{KV}, k_{FZ}\}$) minimiert werden. Hierbei soll eine beliebige Gewichtung von k_{KV} und k_{FZ} verwendbar sein. Für die Datenanalyse stehen hierzu folgende Informationen zu einer Menge von historischen Fahrten, sowie weitere Informationen zum jeweiligen Fahrkontext zur Verfügung:

- **Telemetriedaten:** Werden mehrfach pro Fahrt von Messinstrumenten im Fahrzeug aufgezeichnet und können an eine Cloud-Anwendung übermittelt werden. Z.B. Geschwindigkeit, Beschleunigung, Fahrstil (z.B. „Sport“), Geoposition.
- **Streckendaten:** Werden einer Cloud-Anwendung von externen Datenquellen bereitgestellt. Z.B. Geschwindigkeitsbegrenzungen oder Temperatur und Luftdruck bei einer bestimmten Geoposition und Uhrzeit.

Aufgrund der potenziell großen Menge an Telemetrie- und Streckendaten für viele verschiedene Fahrten muss ein geeignetes Verfahren zur Datenanalyse effizient berechenbar und skalierbar sein.

Viele allgemeine Optimierungsverfahren (z.B. Gradientenverfahren) setzen zur Minimierung von k_{KV} und k_{FZ} die Beschreibung des Zusammenhangs zwischen den Datenpunkten der Telemetrie-/Streckendaten und k_{KV} und k_{FZ} für bestimmte Fahrzeug- bzw. Motormodelle in Form von analytischen Modellen voraus. Diese Modelle sind jedoch z.B. für neue Datenpunkte der Telemetrie- und Streckendaten häufig nicht bekannt und müssen erst aufwendig erstellt werden. Um eine flexiblere Erweiterbarkeit um neue Fahrzeug- bzw. Motormodelle sowie neue Datenpunkte der Telemetrie- und Streckendaten zu erreichen, soll eine Datenanalyse zur Optimierung der Betriebsstrategie ohne entsprechende analytische Lösungsverfahren bzw. Simulationsmodelle verwendet werden können.

2 Hintergrund

2.1 Fahrzeugbetriebsstrategien

Fahrzeugbetriebsstrategien haben zum Ziel, ein Fahrzeug effizient betreiben zu können, z.B. hinsichtlich des Kraftstoffverbrauchs. Der Begriff *Fahrzeugbetriebsstrategie* (kurz *Betriebsstrategie*) wird zwar häufig im Kontext von Hybrid- und Elektrofahrzeugen verwendet, beschreibt aber allgemein die Auswahl von Betriebspunkten eines Fahrzeugs

(z.B. Gang bei Automatikgetrieben) [Ro14]. Die Auswahl erfolgt nach Kriterien wie z.B. der Beeinflussung der Energieeffizienz oder der Bauteilbelastung. Neben der direkten Steuerung von Kontrollsystemen im Fahrzeug können die berechneten Betriebspunkte einer Betriebsstrategie auch dem Fahrer zur manuellen Beeinflussung oder Kenntnisnahme signalisiert werden, indem z.B. eine Anzeige im Kombiinstrument oder ein haptisches Feedback am Gaspedal erfolgt.

Auch die Wahl einer Geschwindigkeit kann einen Betriebspunkt einer Fahrzeugbetriebsstrategie darstellen. Der Kraftstoffverbrauch kann durch die Wahl von geeigneten Geschwindigkeiten wesentlich beeinflusst werden [Le14]. Mit der Wahl einer adäquaten Geschwindigkeit kann z.B. der Einfluss des Luftwiderstands verringert und ein Verbrennungsmotor effizienter betrieben werden. Typischerweise steigt der Verbrauch bei geringen wie auch bei höheren Geschwindigkeiten ab einer bestimmten Grenze wieder an [Le14].

2.2 Genetische Algorithmen

Genetische Algorithmen (GAs) sind etablierte stochastische Suchverfahren zur Lösung von komplexen Optimierungsproblemen. Aufgrund ihrer Flexibilität kommen sie auch häufig bei Problemstellungen des Software Engineerings zum Einsatz (z.B. bei der Optimierung des Deployments einer Anwendung in einer Cloud Umgebung [FFH13]).

Die Funktionsweise von GAs ist dem Evolutionsprozess von Lebewesen nachempfunden [Po00]. GAs können auch Probleme adressieren, für deren Zielfunktion keine analytischen Modelle bekannt sind (im Gegensatz zu z.B. Gradientenverfahren, die differenzierbare Funktionen voraussetzen). GAs betrachten mehrere potenzielle Lösungskandidaten gleichzeitig. Diese Menge der Lösungen wird *Population* genannt. Die potenziellen Lösungen selbst werden als *Individuen* bezeichnet. Die einzelnen Bestandteile (z.B. Attribute) von Individuen werden *Gene* genannt. Alle Gene eines Individuums werden als *Chromosom* bezeichnet.

GAs arbeiten iterativ. In jeder Iteration wird eine neue *Generation* der Individuen erzeugt. Das Erzeugen von neuen Generationen wird mit Hilfe der Operatoren *Selektion*, *Rekombination* und *Mutation* durchgeführt. Diese Operatoren sind auch an natürliche Prozesse angelehnt und werden häufig spezifisch für bestimmte Problemdomänen erstellt.

Zuerst wird eine Menge an Individuen mit Hilfe des Selektions-Operators anhand der Bewertung mit einer Zielfunktion (sog. *Fitnessfunktion*) ausgewählt. Daraufhin wird der Rekombinations-Operator (auch *Crossover* genannt) zur Erzeugung von Nachkommen-Individuen verwendet, indem er Teile der Chromosome von Eltern-Individuen entsprechend bestimmter Regeln austauscht. Danach wird der Mutations-Operator angewandt, bei dem Teile eines Individuums zufällig verändert werden.

Parameter eines GAs, wie z.B. die Größe der Populationen oder die Anzahl an Generationen, werden *Kontrollparameter* genannt und können für ein gegebenes

Optimierungsproblem geeignet konfiguriert werden. GAs verursachen zwar oftmals einen erhöhten Rechenaufwand. Dieser Aspekt wird allerdings durch ihre sehr gute Parallelisierbarkeit abgeschwächt [Ha99].

2.3 Clustering mit Genetischen Algorithmen

Für die Bestimmung von optimalen Fahrzeugbetriebsstrategien verwenden wir Clustering auf Basis von GAs. *Clustering* (bzw. auch *Clusteranalyse* oder *Partitionierung*) bezeichnet hierbei Verfahren zur Aufteilung von Objekten in Klassen. Dabei sind die Kriterien für die Aufteilung vorab nicht bekannt. Somit gehört das Clustering zu der Kategorie des nicht-überwachten Lernens. Beim Clustering wird versucht, die Objekte so aufzuteilen, dass sie innerhalb eines Clusters möglichst ähnlich und in verschiedenen Clustern möglichst unterschiedlich sind [JMF99].

Häufig eingesetzte Verfahren wie K-Means haben oft den Nachteil, dass bei ihnen die Clusteranzahl der Ziel-Partitionierung vorgegeben werden muss. Diese Anzahl ist allerdings oftmals, wie auch in unserer Problemdomäne, nicht a priori bekannt. Diesem Nachteil wird häufig begegnet, indem das Clustering mehrfach mit unterschiedlichen Clusteranzahlen durchgeführt und die Verfahren somit kalibriert werden. Bei einer veränderlichen Datenbasis ist dabei eine regelmäßige Nachkalibrierung notwendig.

Clustering auf Basis von GAs [Ga11] adressiert die Vorgehensweise der Exploration des Lösungsraums mit verschiedenen Clusteranzahlen (in der von uns eingesetzten Variante) inhärent. Die Lösungskandidaten (Individuen) kodieren eine Menge von Clusterzentren, wobei die Clusteranzahl der Ziel-Partitionierung nicht vorgegeben werden muss. Da die Länge der Individuen nicht festgelegt ist, können während der Optimierung verschiedene Clusteranzahlen untersucht werden. Die Änderung der untersuchten Anzahl und Ausprägung der Clusterzentren wird durch die genetischen Operatoren Crossover und Mutation über mehrere Generationen hinweg durchgeführt.

Zur Bewertung der Lösungskandidaten, d.h. der Qualität der Partitionierung, verwenden wir in der Fitnessfunktion den etablierten *Davies-Bouldin-Index (DB-Index)* [Ar13]. Zur Bewertung der resultierenden endgültigen Clustering-Lösung wird von uns zusätzlich der *Silhouette-Index* [Ar13] eingesetzt, da er einen normalisierten Wert liefert, der auch allein (d.h. nicht nur im Vergleich zu anderen Partitionierungen) Auskunft über die Qualität eines Clusterings gibt. Da der DB-Index allerdings weniger Rechenressourcen als der Silhouette-Index bei gleichzeitig guten Ergebnissen liefert, ist die Anwendung des DB-Index in der Fitnessfunktion des GAs vorteilhaft.

3 Cloud-Basierte Optimierung von Betriebsstrategien

Das generelle Vorgehen zur Cloud-basierten Optimierung von Fahrzeugbetriebsstrategien wird in Abschnitt 3.1 beschrieben. Die Verwendung von Clustering zur Bestimmung von optimalen Betriebsstrategien in der Cloud erläutert

Abschnitt 3.2. Die speziellen Aspekte der GA-basierten Clusteranalyse beschreibt danach Abschnitt 3.3 im Detail.

3.1 Generelles Vorgehen

Die Optimierung der Betriebsstrategie für eine bevorstehende Fahrt macht sich die Erfahrungen aus vorhergehenden ähnlichen Fahrten unterschiedlicher Fahrer automatisiert zunutze. Abb. 1 zeigt einen Überblick über das generelle Vorgehen zur Cloud-basierten Optimierung von Fahrzeugbetriebsstrategien. Der Fahrer gibt im Fahrzeug das gewünschte Reiseziel ein (z.B. in dessen Infotainmentsystem) und stellt darüber hinaus den präferierten Fahrstil ein bzw. verwendet eine Voreinstellung. Daraufhin wird die Route berechnet und zusammen mit weiteren Informationen, wie z.B. zum gewählten Fahrstil oder dem verwendeten Fahrzeug(modell), zur Cloud-Anwendung übertragen. Dort findet kontinuierlich eine Aggregation der Telemetriedaten aller übermittelten Fahrten, sowie der zugehörigen Streckendaten statt (weitere Details zu dieser Datenaggregation werden in Abschnitt 3.2 erläutert).

Mittels eines Clusterings dieser Daten auf Basis eines GAs wird der Suchraum für Fahrten mit ähnlichen Charakteristiken stark eingeschränkt. Somit lässt sich für eine bevorstehende Fahrt der Cluster mit den ähnlichsten Telemetrie- und Streckendaten auswählen und die Suche nach der jeweiligen optimalen Betriebsstrategie auf diesen Cluster begrenzen. Die Realisierung der Clusteranalyse mit Hilfe eines GAs unterstützt hierbei auch das Ziel einer guten Skalierbarkeit (vgl. Abschnitt 1.2).

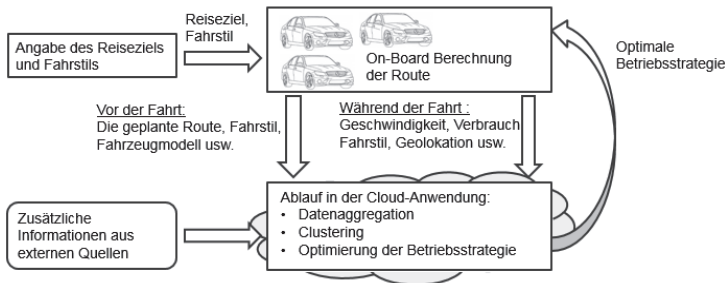


Abb. 1: Generelles Vorgehen zur Cloud-basierten Optimierung von Betriebsstrategien

Durch eine vorausschauende Fahrstrategie soll eine Minimierung des Kraftstoffverbrauchs und der Fahrzeit (k_{KV} und k_{FZ} , siehe Abschnitt 1.2) erreicht werden. Die Regulierung von k_{KV} und k_{FZ} erfolgt über die Berechnung einer idealen Geschwindigkeit (vgl. Abschnitt 2.1) für einzelne Streckenabschnitte in einem gegebenen Kontext (z.B. Verkehrsaufkommen, Wetterbedingungen, Motorleistung).

Eine optimale Betriebsstrategie für eine Route setzt sich aus den optimalen berechneten Geschwindigkeiten für die einzelnen Streckenabschnitte der Route zusammen. Nachdem die optimale Betriebsstrategie für eine Route in der Cloud-Anwendung berechnet wurde,

kann diese in das Fahrzeug übermittelt und dort angewandt werden. Während der Fahrt werden aktuelle Telemetrie- und Streckendaten kontinuierlich zur Cloud-Anwendung übertragen. Somit kann der Vorgang zur Bestimmung und Übermittlung der optimalen Betriebsstrategie während der Fahrt wiederholt werden, um beispielsweise auf veränderte Stau- oder Wetterprognosen reagieren zu können.

3.2 Bestimmung von Optimalen Betriebsstrategien mittels Clustering

Der Ansatz zur Bestimmung von optimalen Betriebsstrategien in einer Cloud-Anwendung unter Verwendung von Clustering ist in Abb. 2 dargestellt. In der Cloud-Anwendung findet die Analyse der übermittelten Daten mittels Clustering statt. Hier wird zuerst die Reiseroute und ihr Kontext, wie z.B. die Antriebsart des Fahrzeugs, Temperatur und für die Tageszeit charakteristische Stauprognose, in Abschnitte von 100 Metern aufgeteilt (Schritt 1). Hierbei werden die durchschnittlichen Werte pro Abschnitt berechnet. Für jeden Abschnitt wird der Cluster mit den ähnlichsten Strecken bestimmt (gemessen am euklidischen Abstand eines Abschnitts zu einem Clusterzentrum) und der jeweilige Abschnitt wird diesem Cluster zugeordnet (Schritt 2). Für jeden dieser Cluster wird das erste Prozent der Abschnitte aus den historischen Daten bestimmt, die die höchste Ähnlichkeit (auch gemäß dem euklidischen Abstand) mit dem zu analysierenden Abschnitt aufweisen (Schritt 3). Aus der Menge der ähnlichen Abschnitte werden diejenigen ausgewählt, die die am besten geeignete Betriebsstrategie aufweisen (Schritt 4).

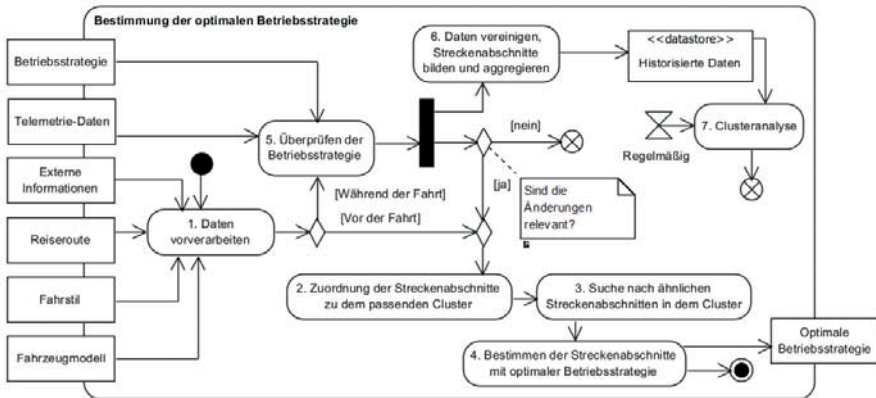


Abb. 2: Ansatz zur Bestimmung der optimalen Betriebsstrategie mittels Clustering

Hierbei wird der Abschnitt mit der geringsten normalisierten Summe von Kraftstoff- bzw. Energieverbrauch und Fahrzeit, entsprechend ihren Gewichten, ausgewählt. Da jedem Streckenabschnitt eine durchschnittliche Geschwindigkeit zugeordnet ist, wird auf diese Weise die optimale Geschwindigkeit für jeden geplanten Streckenabschnitt ermittelt. Anschließend kann die gesamte optimale Betriebsstrategie für die ganze

bevorstehende Strecke in das Fahrzeug übermittelt werden. Eine Priorisierung von Zeit und Kraftstoff- bzw. Energieverbrauch kann vom Fahrer gemäß seiner Präferenzen variiert werden.

Wenn sich die Bedingungen während der Fahrt ändern, wird die alte Betriebsstrategie überprüft (Schritt 5). Falls die Änderungen relevant sind, z.B. bei veränderter Stauprognose oder bei der Änderung der Reiseroute, wird eine neue optimale Betriebsstrategie durch erneutes Ausführen der Schritte 2-4 bestimmt. Die neue Betriebsstrategie ersetzt in diesem Fall die alte.

Die Datenbasis für die Optimierungen wird aus historischen Daten geschaffen. Während der Fahrt werden vom Fahrzeug erzeugte Telemetrie-Daten kontinuierlich in die Cloud-Anwendung übertragen. Im nächsten Schritt werden die Fahrzeugdaten in 100 Meter langen Streckenabschnitten aggregiert und in der Cloud-Anwendung abgelegt (Schritt 6). Im Unterschied zu Schritt 1 werden hierbei auch für jeden Streckenabschnitt die durchschnittlichen Werte der dynamischen Sensormesswerte (u.a. auch der Geschwindigkeit) berechnet. Die in Abschnitten von 100 Metern aggregierten Daten werden mit zusätzlichen Informationen, z.B. zu Fahrzeugcharakteristiken, Wetterbedingungen, Verkehrsaufkommen und dem eingestellten Fahrstil, ergänzt. Anschließend wird die Clusteranalyse durchgeführt (Schritt 7), d.h. die Aufteilung von Streckenabschnitten mit ähnlichen Eigenschaften in Cluster zur Einschränkung des Suchraums. Da die neuen Fahrzeug- und Streckendaten (ohne Bezug zu Fahrerinformationen) kontinuierlich zu der Datenbasis hinzugefügt werden, wird auch das Clustering regelmäßig durchgeführt. Die GA-basierte Clusteranalyse wird in Abschnitt 3.3 detailliert beschrieben. Generell können die Länge der Streckenabschnitte und der Anteil der ähnlichsten Abschnitte in Abhängigkeit von den zur Verfügung gestellten Rechenkapazitäten variiert werden.

3.3 GA-Basierte Clusteranalyse

Bei der Clusteranalyse mit GA sind die potenziellen Lösungen (eine Lösung ist eine Aufteilung in Cluster) wie folgt aufgebaut: Die Gene der Chromosome kodieren Clusterzentren. Bei n Attributen der Telemetrie- und Streckendaten repräsentiert jedes Clusterzentrum pro Streckenabschnitt einen Punkt im n -dimensionalen Raum (Abb. 3). Durch diese Repräsentation lassen sich auch neue Datenattribute der Telemetrie- und Streckendaten leicht in die Clusteranalyse integrieren.

Die Länge der Chromosome variiert im Intervall zwischen minimaler und maximaler Anzahl der Clusterzentren. Das Intervall wird für den Kontext der experimentellen Evaluation (vgl. Abschnitt 4) von 2 bis 100 Clusterzentren empirisch festgelegt, es sind jedoch auch Verfahren zur dynamischen Adaption des Intervalls denkbar. Bei der Erzeugung der ersten Population wird die Anzahl der Clusterzentren für jedes Individuum (d.h. jedes Chromosom) zufällig aus diesem Intervall ausgewählt.

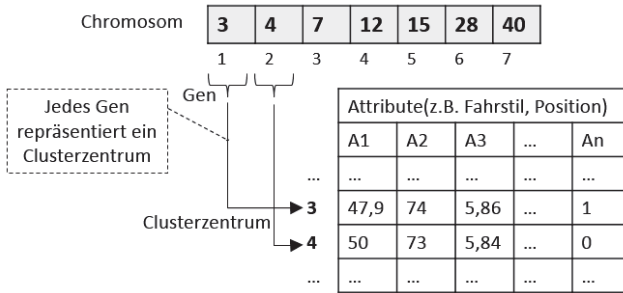


Abb. 3: Ein exemplarisches Chromosom

Die Berechnung des Fitnesswerts von Individuen wird mit einer Fitnessfunktion realisiert, bei der zuerst alle Streckenabschnitte der historischen Daten anhand des niedrigsten euklidischen Abstands zu bestehenden Clusterzentren des Individuums zugeordnet werden. Daraufhin wird die Aufteilung mit dem DB-Index (vgl. Abschnitt 2.3) bewertet. Als *Selektions*-Operator wird eine Fitness-proportionale Selektion mit linearer Skalierung angewendet. Die Selektion erfolgt auf Basis von Auftrittswahrscheinlichkeiten, die den Individuen entsprechend ihrer Fitnesswerte zugeordnet werden. Um eine vorzeitige Konvergenz auf ein lokales Optimum zu vermeiden, wird bei der Selektion ein Rauschen [P00] hinzugefügt. Hierdurch werden auch ein Prozent der Individuen mit unterdurchschnittlichen Fitnesswerten für die Erzeugung der nächsten Generation verwendet.

Die Reihenfolge der Clusterzentren in Individuen ist bei der *Rekombination (Crossover)* unwichtig. Von Bedeutung ist die Abwesenheit von Duplikaten in jedem Nachkommen-Individuum (Kind). Deswegen werden die einzigartigen Gene, d.h. die Clusterzentren, der zwei Eltern-Individuen bei der Rekombination in einen Pool abgelegt. Aus diesen Genen werden zwei Nachkommen erzeugt. Die Anzahl ihrer Clusterzentren wird wie folgt aus den Clusterzentren ihrer Eltern berechnet. Wenn $k1$ und $k2$ die Länge der zwei Eltern-Individuen und $h1 (=ceiling(k1/2))$ und $h2 (=floor(k2/2))$ Kreuzungspunkte (bzw. die Länge der Eltern-Individuen bis zu diesen Kreuzungspunkten) bezeichnen, wird die Länge der Nachkommen-Individuen als $h1+k2-h2$ und $h2+k1-h1$ festgelegt. Nach der Erzeugung des ersten Kindes (durch das zufällige Entnehmen von Genen aus dem Pool) werden die in dem Genpool noch nicht verwendeten Gene für die Erzeugung des zweiten Kindes zufällig entnommen. Falls diese verbleibenden Gene nicht ausreichen, werden für die Erzeugung des restlichen Teils des zweiten Kindes gleiche (zufällig ausgewählte) Gene wie bei dem erstem Kind verwendet. Abb. 4 zeigt dieses Vorgehen für den Rekombinations-Operator an einem Beispiel, bei dem das sechste Gen des ersten und fünfte Gen des zweiten Eltern- Individuums gleich sind. Somit wird das Gen nur einmal in den Genpool übernommen.

Bei der *Mutation* wird ein Clusterzentrum in einem Individuum durch ein anderes zufällig ersetzt und die Clusteranzahl zufällig vergrößert oder verkleinert.

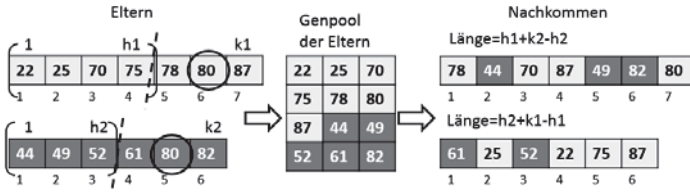


Abb. 4: Rekombination und Vorgehen bei doppelten Clusterzentren (markiert mit Kreis)

Hierbei werden die Individuen, die weniger als sechs Clusterzentren enthalten, nicht verkleinert. Diese empirisch bestimmte Grenze verhindert effektiv die Konvergenz auf nur ein Clusterzentrum. Die Anzahl der Clusterzentren, die abgeschnitten oder hinzugefügt werden, wird zufällig aus dem Intervall von Null bis zur Hälfte der Anzahl der Clusterzentren des ursprünglichen Individuums ausgewählt. Bei der Mutation wird darauf geachtet, dass zu einem Individuum nur Clusterzentren hinzugefügt werden, die es noch nicht enthält. In Abb. 5 wird das Vorgehen bei der Mutation illustriert. Das zu mutierende Individuum umfasst sieben Gene. Im ersten Schritt wird zufällig das Gen an fünfter Position mit einem anderem (dunkelgrau markiert) ersetzt. Im zweiten Schritt wird entschieden, ob das Individuum verkürzt oder verlängert werden muss. In diesem Beispiel werden die letzten beiden Gene abgetrennt.

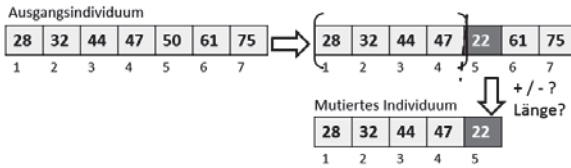


Abb. 5: Mutations-Operator

4 Experimentelle Evaluation

Als Datenbasis werden aggregierte historische Daten aus 62 aufgezeichneten Fahrten verwendet, die mit drei unterschiedlichen Fahrzeugen mit verschiedenen Antriebsarten und in unterschiedlichen Regionen für die Evaluation durchgeführt wurden. Nach der Aggregation der Daten zu Abschnitten von 100 Metern entstanden hierbei 14.449 Datensätze. Bei der Datenanalyse werden für die Evaluation 17 Datenattribute verwendet, u.a. die Steigung und Kurvigkeit der Streckenabschnitte, die Luftfeuchtigkeit und Umgebungstemperatur, der eingestellte Fahrmodus, sowie das vom Hersteller angegebene maximale Drehmoment und die Leistung und Beschleunigung eines Fahrzeug- bzw. Motormodells. Die Abtastrate der Telemetriedaten beträgt 2 Hz, d.h. alle 0,5 Sekunden werden die Telemetriedaten ausgelesen und gespeichert.

4.1 Validierung des Clusterings

Um die Kontrollparameter des GA empirisch zu bestimmen, die in unserem Kontext die besten Fitnesswerte ermöglichen, werden Versuche mit variierenden Kontrollparametern durchgeführt. Die folgenden Parameter erweisen sich in Bezug auf den DB-Index und den Silhouette-Index als am besten geeignet [Og16]. Die Populationsgröße beträgt lediglich 100 Individuen. Deswegen wird eine große Anzahl von Generationen (5.000) festgelegt. Der Algorithmus terminiert, wenn der beste Fitnesswert während 300 Iterationen konstant bleibt. Eine hohe Crossoverrate (80%) und eine relativ kleine Mutationsrate (10%) werden in Kombination mit Elitismus [Po00] (4%) pro Generation angewandt.

Mit dieser Konfiguration findet der GA in den Fahrtdaten drei Cluster. Zur Bewertung der Ergebnisse des Clusterings wird der Silhouette-Index verwendet, dieser liegt generell im Intervall $[-1;1]$, wobei größere Werte bessere Ergebnisse repräsentieren und positive Werte dem Fall entsprechen, dass die Abstände zu den anderen Clustern größer sind als die Distanz innerhalb der Cluster. Der Silhouette-Index der berechneten Clusterlösung beträgt $\sim 0,48$. Dies stellt somit ein gutes Ergebnis dar.

In Abb. 6 werden exemplarisch die Verteilungen der Attribute „Pedalposition“ (links, wie weit ist ein Gaspedal gedrückt) und „Lufttemperatur“ (rechts) pro Streckenabschnitt in jedem Cluster dargestellt. Cluster 1 separiert z.B. deutlich die Abschnitte, bei denen mehr Gas gegeben wurde und solche mit niedrigen Temperaturen.

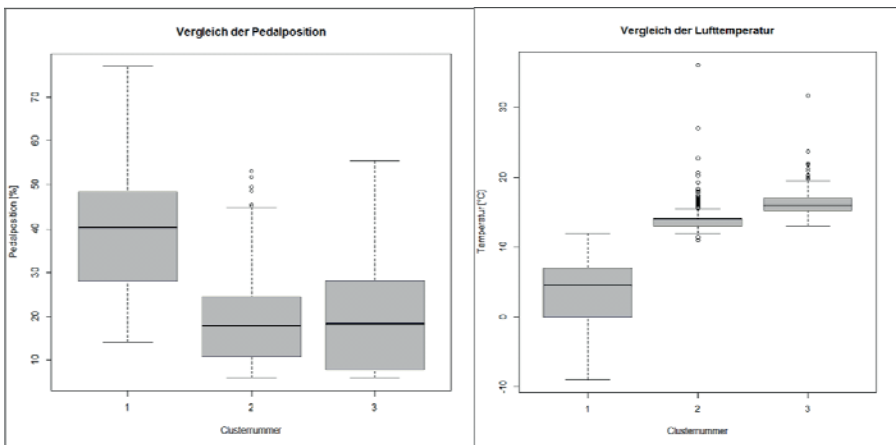


Abb. 6: Verteilung von Pedalposition (links) und Lufttemperatur (rechts) in den Clustern

Um den Unterschied in den Ergebnissen des Clusterings in Abhängigkeit von der Größe der Datenbasis festzustellen, wird der GA mit verschiedenen großen Anteilen der Datenbasis evaluiert. Hierbei wird auch die sequenzielle Ausführung des GAs mit der parallelen Ausführung verglichen (16 GB RAM, CPU mit 4 Kernen, Intel® Core(TM) i5-2400 CPU @ 3.10 GHz). Die Ergebnisse sind in Abb. 7 dargestellt.

Die linke Grafik in Abb. 7 veranschaulicht die ansteigende Rechenzeit des Algorithmus pro Iteration mit dem Anstieg des Datenvolumens. Jedoch bringt die parallele Berechnung des Algorithmus einen deutlichen Performancegewinn mit sich. Die Rechenzeit verringert sich über alle Datenvolumina durchschnittlich um ca. die Hälfte. Dies entspricht den Erwartungen, da bei paralleler Verarbeitung der Performancegewinn nicht linear mit der Zahl der Cores skaliert [PKG16]. Die rechte Grafik in Abb. 7 zeigt die Abhängigkeit des Fitnesswertes vom Datenvolumen. Der maximale Fitnesswert wird mit 100% der Daten erreicht und unterscheidet sich bei der sequenziellen und parallelen Ausführung des GAs fast nicht.

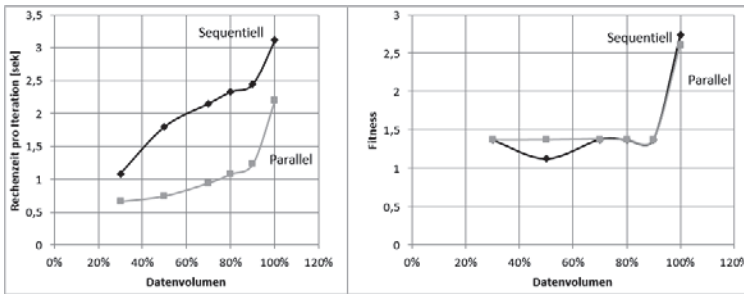


Abb. 7: Vergleich der sequenziellen mit der parallelen Ausführung des GAs

4.2 Bestimmung der optimalen Betriebsstrategien

Für die Validierung der Bestimmung der optimalen Betriebsstrategien werden mit unserem Ansatz die optimalen Geschwindigkeiten für zwei Strecken berechnet und mit den tatsächlichen Geschwindigkeiten während der Fahrt verglichen. Hierbei werden die Fahrzeit, der Spritverbrauch und die CO₂-Emissionen pro Strecke für unterschiedliche Optimierungskriterien berechnet, d.h. für unterschiedliche Gewichtungen der Zeit und des Spritverbrauchs. Den Berechnungen zufolge sinkt der Kraftstoffverbrauch der Fahrzeuge für beide Strecken mit der berechneten optimalen Betriebsstrategie. Falls das Optimierungskriterium zu 100% den Spritverbrauch priorisiert, sind, wie erwartet, die CO₂-Emissionen und der Spritverbrauch pro Strecke am geringsten. Falls das Optimierungskriterium zu 100% die Zeit priorisiert, ist die Zeit pro Strecke am kürzesten, die CO₂-Emissionen und der Spritverbrauch erhöhen sich jedoch.

5 Verwandte Arbeiten

Terwen [Te09] beschäftigt sich mit der vorausschauenden Längsregelung schwerer Lastkraftwagen und Radke [Ra13] mit der energieoptimalen Längsführung von Personenkraftwagen. Beide Arbeiten zielen auf die Reduktion des Kraftstoffverbrauchs durch den Einsatz vorausschauender Fahrstrategien. Die Optimierung wird bei Terwen [Te09] mit Hilfe der Quadratischen Programmierung für modellbasierte prädiktive

Regelung und bei Radke mittels Dynamischer Programmierung in Echtzeit im Fahrzeug durchgeführt. Der Gang und wie bei unserem Ansatz die Fahrgeschwindigkeit sind die Basis für die Bestimmung der optimalen Betriebsstrategie [Ra13]. Genaue Angaben zum Umfang der Betriebspunktendatenbanken wie beispielsweise die Zahl der gespeicherten Streckenabschnitte finden sich nicht.

Unter Verwendung der Dynamischen Programmierung wurde von Roth [Ro14] ein Optimierungsansatz auf Basis einer Betriebspunktendatenbank vorgestellt. Hierbei werden die per Simulation erzeugten Fahrzeug-Betriebspunkte auf Streckenabschnitten von fünf Metern aggregiert und in einer Datenbank gespeichert.

Die Arbeit von Salmasi [Sa07] beschäftigt sich mit Regelungsstrategien für Hybrid-Fahrzeuge. Er verwendet zur Darstellung die vier Klassen „Deterministische“ und „Fuzzy“ regelbasierte Ansätze sowie „Globale“ und „Echtzeit“ Optimierungsansätze und stellt die jeweiligen Verfahren mit ihren mathematischen Grundlagen dar. Ein konkreter Anwendungsfall findet sich jedoch nicht.

Zusammenfassend bleibt festzuhalten, dass alle hier referierten Arbeiten im Wesentlichen dezentrale, fahrzeugbasierte Ansätze verfolgen. Unser Ansatz nutzt genetische Algorithmen wie sie auch in [Sa07] dargestellt sind, geht jedoch durch die Nutzung einer Cloudlösung einen wichtigen Schritt hin zur fahrzeugunabhängigen Nutzung der Streckendaten und Optimierungsansätze.

6 Zusammenfassung und Ausblick

Die Optimierung von Fahrzeugbetriebsstrategien erfolgte bisher hauptsächlich statisch für eine Menge von Fahrzeugen, z.B. für bestimmte Baureihen, während deren Entwicklung. Dieses Papier beschreibt einen Ansatz, der sich die zunehmende Onlinevernetzung von Fahrzeugen zunutze macht und Telemetrie- und Streckendaten von vielen unterschiedlichen Fahrten vieler unterschiedlicher Fahrzeuge in eine Cloud-Anwendung überträgt. Mittels eines Clusteringverfahrens auf Basis von genetischen Algorithmen können daraufhin für bevorstehende Fahrten optimale Betriebsstrategien berechnet werden, die für vorliegende spezifische Charakteristika (z.B. Verkehrsaufkommen, Temperatur, Motorleistung) den Kraftstoffverbrauch und/oder die Fahrtzeit minimieren.

Für zukünftige Erweiterungen unseres Ansatzes sind besonders die Einbeziehung von weiteren Attributen der Telemetrie- und Streckendaten relevant. Des Weiteren könnten durch die Variation der Auflösung der übermittelten Daten, sowie der Länge der Streckenabschnitte, die für die Aggregation dieser Daten verwendet werden, weitere Qualitäts- und Performanceverbesserungen erzielt werden.

Literaturverzeichnis

- [Ar13] Arbelaitz, O. et. al.: An extensive comparative study of cluster validity indices. In: Pattern Recognition, Vol. 46, Issue 1, Elsevier, S. 243–256, 2013.
- [FFH13] Frey, S.; Fittkau, F.; Hasselbring, W.: Search-Based Genetic Optimization for Deployment and Reconfiguration of Software in the Cloud. In: Proceedings of the 35th International Conference on Software Engineering (ICSE 2013), IEEE Press, S. 512–521, 2013.
- [Ga11] Gajawada, S. et. al.: Optimal Clustering Method Based on Genetic Algorithm. In: Proceedings of the International Conference on Soft Computing for Problem Solving (SocProS 2011), 2011.
- [Ha99] Hansohm, J.: Clusteranalyse mit Genetischen Algorithmen. In: Mathematische Methoden der Wirtschaftswissenschaften, Springer, S. 57–66, 1999.
- [JMF99] Jain, A. K.; Murty, M. N.; Flynn, P. J.: Data Clustering: A Review, ACM Computing Surveys, Vol. 31, Issue 3, S. 264–323, 1999.
- [KW14] Köhler, T. R.; Wollschläger, D.: Die digitale Transformation des Automobils: 5 Mega-Trends verändern die Branche. Media-Manufaktur, 2014.
- [Le14] Lederer, M.: Energieeffizientes Fahren. In: Energiemanagement im Kraftfahrzeug. Optimierung von CO₂-Emissionen und Verbrauch konventioneller und elektrifizierter Automobile, Springer Vieweg, S. 307–321, 2014.
- [Og16] Ogulenko, N.: Cloud-basierte Optimierung von Fahrzeugbetriebsstrategien auf Basis historischer Telemetrie- und Streckendaten, Masterthesis, Hochschule Aachen - Technik und Wirtschaft, 2016.
- [PKG16] Pei, S.; Kim, M.-S.; Gaudiot, J.-L.: Extending Amdahl's Law for Heterogeneous Multicore Processor with Consideration of the Overhead of Data Preparation. In: IEEE Embedded Systems Letters, Vol. 8, No. 1, S. 26–29, 2016.
- [Po00] Pohlheim, H.: Evolutionäre Algorithmen - Verfahren, Operatoren und Hinweise für die Praxis, Springer, 2000.
- [Ra13] Radke, T.: Energieoptimale Längsführung von Kraftfahrzeugen durch Einsatz vorausschauender Fahrstrategien. Karlsruher Institut für Technologie (KIT) Fakultät für Maschinenbau: KIT Scientific Publishing, Dissertation, 2013.
- [Ro14] Roth, M.: Betriebsstrategie. In: Energiemanagement im Kraftfahrzeug. Optimierung von CO₂-Emissionen und Verbrauch konventioneller und elektrifizierter Automobile, Springer Vieweg, S. 323–365, 2014.
- [Sa07] Salmasi, F.: Control Strategies for Hybrid Electric Vehicles: Evolution, Classification, Comparison, and Future Trends. In: IEEE Transactions on Vehicular Technology, S. 2393–2404, 2007.
- [Te09] Terwen, S.: Vorausschauende Längsregelung schwerer Lastkraftwagen. Schriften des Instituts für Regelungs- und Steuerungssysteme Karlsruher Institut für Technologie: KIT Scientific Publishing, Dissertation, 2009.

Einsatz von dynamisch rekonfigurierbaren FPGAs in Fahrzeugen

Peter Brungs¹ und Marcel Baunach²

Abstract: Die Anforderungen zukünftiger Fahrerassistenzsysteme (ADAS) an Konnektivität, Flexibilität und Verarbeitungsleistung werden immer höher, sodass aktuelle Prozessoren (MCU/DSP) durch immer aufwendigere Algorithmen an ihre Grenzen stoßen. Die vorgestellte Arbeit zeigt ein Konzept, das durch eine flexible FPGA-Architektur eine rekonfigurierbare Hardwarebeschleunigung erlaubt. Statt komplexe Algorithmen weiterhin als Software für MCUs oder DSPs zu implementieren, werden partiell rekonfigurierbare Hardwarekomponenten in FPGAs realisiert. In verteilten Systemen dieser Art werden Daten über Kommunikationsschnittstellen mit anderen Hardwarekomponenten ausgetauscht, sodass eine flexible und skalierbare Architektur entsteht. Zusätzlich gestattet das vorgestellte Konzept Spezialisierung und Redundanz kritischer Verarbeitungseinheiten hinsichtlich Performanz und Ausfallsicherheit. Die FPGA-Implementierung ist modular aufgebaut, sodass die Komponenten beliebig austauschbar und erweiterbar sind. Dies ermöglicht u.a. den Wechsel der Kommunikationsschnittstelle (z.B. von Ethernet nach CAN-Bus) oder die Erweiterung des Netzwerkprotokolls für zukünftige Anforderungen.

Keywords: Dynamische Rekonfiguration, Verteilte Systeme, FPGA, ADAS

1 Einleitung

Aufgrund wachsender Anforderungen an die Leistungsfähigkeit und Konnektivität eingebetteter Systeme bzw. elektronischer Steuergeräte (ECUs) in Fahrzeugen wurden mit AUTOSAR4 Multi-Core- und Ethernetfähigkeiten eingeführt [We13]. Zusammen mit zusätzlicher Sensorik und Aktuatorik erzeugen diese jedoch große Datenmengen, die gefiltert, aggregiert, analysiert und aufbereitet werden müssen. Hier kann rekonfigurierbare Hardware (FPGAs) aufgrund ihrer flexiblen und konfigurierbaren Architektur zur applikationsspezifischen Verarbeitung großer Datenmengen eingesetzt werden. Verglichen mit reinen Softwarelösungen bietet dies zusätzliche Möglichkeiten, wie Hardwareparallelität und Time-Multiplexing von Algorithmen, für die Ausführung hart echtzeitkritischer Berechnungen [FF12]. Zudem können durch den Einsatz von FPGAs in ECUs erkannte Defekte und Implementierungsfehler der Hardware durch erneute Konfiguration behoben werden; ggf. sogar so, dass sich das gesamte verteilte System eines Fahrzeugs selbständig reparieren kann (Configuration Scrubbing). Ein anderer Aspekt ist die Stabilität im Fehlerfall, wenn ganze ECUs ausfallen. Es besteht hier die Möglichkeit, Funktionen automatisch aus einem FPGA in ein anderes FPGA zu verlagern, sodass defekte Steuergeräte im Fahrzeugnetzwerk dynamisch ersetzt werden können. Hier will das vorliegende Papier einen neuen Ansatz für eine verteilte FPGA Architektur in Fahrzeugen aufzeigen.

¹ Universität Würzburg, Lehrstuhl für Technische Informatik, mail@peter-brungs.eu

² Technische Universität Graz, Institut für Technische Informatik, baunach@tugraz.at

Die Arbeit

- will eine FPGA Architektur vorstellen, die es erlaubt, heterogene sowie flexibel skalierbare FPGA-Netzwerke aufzubauen.
- zeigt exemplarisch eine modulare FPGA-Implementierung, die den Austausch und die Erweiterung von Hardwarekomponenten sowie deren Schnittstellen ermöglicht.
- zeigt, wie sich zusätzliche, rekonfigurierbare Hardwarekomponenten über ein neues Netzwerkprotokoll dynamisch in ein FPGA programmieren lassen.

Dieses Papier ist wie folgt strukturiert: Kapitel 2 gibt eine Übersicht zu vergleichbaren Architekturen, die dynamisch partielle Rekonfiguration einsetzen. In Kapitel 3 wird die Struktur der FPGA-Architektur beschrieben und in Kapitel 4 wird das eigene Konzept der netzwerkkontrollierten FPGA-Rekonfiguration vorgestellt. In Kapitel 5 wird der Ressourcenverbrauch auf dem FPGA diskutiert, und in Kapitel 6 wird mit einem Ausblick auf zukünftige Möglichkeiten die Arbeit abgeschlossen.

2 Architekturen im Vergleich

Von der Vielzahl an Ansätzen, rekonfigurierbare FPGA-Systeme zu realisieren, sollen hier einige exemplarisch aufgezeigt werden. Die Erlangen Slot Machine (ESM) [An10] ist ein Beispiel, in dem ein PowerPC Prozessor die Konfiguration steuert. Diese CPU ist außerhalb des eingesetzten Virtex II FPGAs (MotherBoard) angeordnet und steuert über eine Crossbar mit Memory Mapped I/O den Rekonfigurationsmanager. Der Rekonfigurationsmanager ist ebenfalls ein FPGA (Spartan II), der die Bitstream Informationen aus seinem Flash Speicher über eine Crossbar in die rekonfigurierbaren Slots des FPGAs lädt. Die Slots sind wiederum unterteilt in MicroSlots, die dann die kleinsten rekonfigurierbaren Einheiten der Erlangen Slot Machine bilden. In dieser Architektur ist der Ressourcenaufwand an elektronischen Komponenten sehr hoch, da zwei FPGA Bausteine sowie eine separate CPU verwendet werden. Die Anbindung zur Außenwelt ist über USB, Ethernet, IEEE1394 und PCMCIA Schnittstellen möglich.

Der Prozessor für die FPGA-Steuerung muss jedoch nicht zwingend außerhalb des FPGAs untergebracht sein. Heutige FPGAs sind ausreichend leistungsfähig, um Softcore-Prozessoren (z.B. PowerPC oder MicroBlaze) direkt integrieren zu können. Die in [Hü10] vorgestellte Lösung synthetisiert neben einem Prozessor, einen ProcessorBus, einen Speichercontroller und einen Hardware Controller für den Internal Configuration Access Port (ICAP) direkt für den FPGA-Baustein. Der Speichercontroller ist mit einem externen Speicher für die Konfigurationsdaten verbunden. Diese Lösung ist im Vergleich zur ESM hinsichtlich der Systemarchitektur effizienter, da die wichtigsten Elemente in einem einzigen FPGA zusammengefasst wurden. Ersatzweise werden nun allerdings Ressourcen in Form von Logikzellen auf dem FPGA benötigt.

Die in [BY08] gezeigte Architektur kommt komplett ohne dedizierten Prozessor aus. Dieser wurde durch einen Parallel Configuration Access Port (PCAP) ersetzt, welcher die

partielle Rekonfiguration durch den SelectMap Port, einem externen ICAP Port, steuert. Die Konfiguration-BitStreams werden im internen BlockRAM des FPGAs gespeichert. Diese Lösung ist sehr ressourcensparend und nur die nötigsten Funktionen werden für die Steuerung des SelectMap Ports realisiert. Nachteilig ist, dass die Konfigurationsinformationen fix im BlockRAM gespeichert sind und nach der FPGA-Initialisierung nicht mehr verändert werden können.

Die in [Li11] vorgestellte Architektur beschreibt ein FPGA Communication Framework (CF), basierend auf dem Ethernet Standard und der Netzwerk Socket Technologie. Diese Lösung nutzt ein eigenes FPGA Communication Protocol (FCP), basierend auf dem Netzwerkprotokoll UDP. Auf dem FPGA wird FCP durch einen Paket-Prozessor realisiert, der die Steuerung des Netzwerkverkehrs übernimmt. ICAP ist hier durch Channel Interfaces angebunden. Diese basieren auf LocalLinks, kombiniert mit Xilinx Ethernet MAC to LocalLink Interfaces und Ethernet MAC Wrapper, die teilweise Lizenzgebühren unterliegen.

3 Die Struktur der FPGA Architektur

Im Vergleich zu den vorgenannten Realisierungen basiert die in diesem Papier vorgeschlagene Lösung auf einem Ethernet MAC Controller IP von OpenCores [Mo11], der unter der LGPL Lizenz verfügbar ist. Unter [Mo09] steht ebenfalls ein CAN-Bus Controller bereit, der anstatt des Ethernet MAC Controllers integriert werden kann. Daneben werden freie Xilinx Core Module wie FIFOs und eine PLL IP Core zur Generierung unterschiedlicher Takte für die Netzwerkkomponente und der ICAP Schnittstelle verwendet. Das Netzwerkprotokoll ist in der präsentierten Arbeit beschränkt auf die IP Ebene, sodass der OSI Netzwerkstack in der Implementierung schlanker ausfällt, da auf die Funktionalität der TCP/UDP Sockets verzichtet wird. Die Steuerung übernimmt eine Finite State Machine (FSM), welche die Netzwerk- und ICAP-Schnittstelle kontrolliert, sodass die Anzahl der Komponenten minimiert werden konnte. Über die Ethernet-Schnittstelle werden die Steuer-, Konfigurations- und Ergebnisdaten übertragen, sodass auch kleine FPGA Implementierungen ermöglicht werden.

Klassische Vorgehensweisen zur Programmierung des FPGA-Konfigurationsspeichers benötigen eine direkte Punkt-zu-Punkt Verbindung zwischen PC und FPGA, z.B. über USB oder PCI-Bus. Die hier gezeigte Lösung erlaubt es, die Rekonfiguration über Netzwerke wie Ethernet oder CAN-Bus aus der Ferne durchzuführen, sodass FPGA und Programmierer räumlich getrennt sein können.

Abbildung 1 zeigt den grundsätzlichen Aufbau der vorgestellten FPGA Architektur: Das Netzwerkinterface XETHERNET empfängt die Daten aus dem Netzwerk und sendet diese, kontrolliert durch XETHERNET_CONTROL, an die Reconfiguration Unit ICAP_SP6. Die Reconfiguration Unit hat eine direkte Verbindung in die Dynamic Reconfiguration Area, in die letztlich alle synthetisierten Hardwarekomponenten programmiert werden. Daten und Verarbeitungsergebnisse werden ebenfalls von der ICAP_SP6 Komponente aus der Dynamic Reconfiguration Area gelesen und durch das XETHERNET Modul an das externe Netzwerk gesendet.

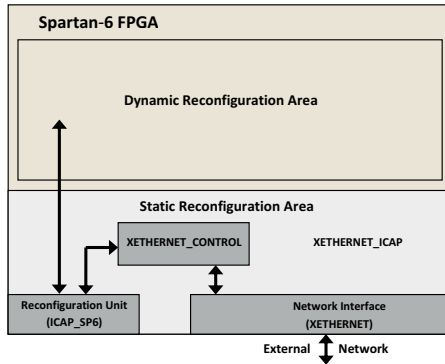


Abb. 1: Grundsätzliche Struktur des dynamisch rekonfigurierbaren FPGA

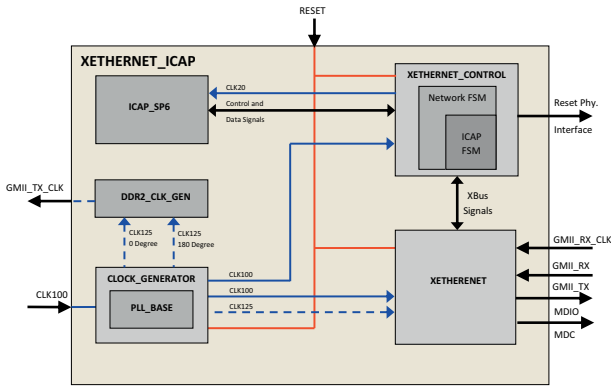


Abb. 2: XETHERNET_ICAP Komponente

Abbildung 2 zeigt den Aufbau der XETHERNET_ICAP Komponente zur Steuerung des FPGA Konfigurationspeichers. Diese besteht aus vier Hauptkomponenten: ICAP_SP6 bildet die Schnittstelle zum Spartan-6 Konfigurationspeicher und XETHERNET die Schnittstelle zum Netzwerk. XETHERNET_CONTROL überwacht beide vorgenannten Komponenten und steuert deren Aktionen durch die Netzwerk FSM und die ICAP FSM. Der CLOCK_GENERATOR generiert die notwendigen Takte für ICAP_SP6, XETHERNET und XETHERNET_CONTROL. Diese Architektur braucht im Gegensatz zu den prozessorgetriebenen Lösungen weniger Ressourcen und beschränkt sich auf die nötigsten Komponenten, um die Transferlast von Konfigurationsdaten zwischen FPGA und den Host-Anwendungen gering zu halten.

Abbildung 3 zeigt die XETHERNET Komponente im Detail. Als Basis für die Gigabit-Ethernet Realisierung dienen der GEMAC Controller [Et09] und Informationen über ethmac aus [Mo11]. Diese Lösung ermöglicht Herstellerunabhängigkeit mit eventuell anfallenden Lizenzkosten. Die Taktfrequenz des XETHERNET Komponente ist 100 MHz und die der Ethernet Schnittstelle liegt bei 125 MHz im GBit Modus. XETHER.Transmitter und XETHER.Receiver dienen in der GEMACFIFO Komponente als Schnittstelle zwi-

schen diesen beiden Frequenzbereichen. Die REGISTER Komponente dient der Steuerung und der Konfiguration der XETHERNET Komponente. Der modulare Aufbau der vorgestellten Architektur ermöglicht den Austausch der Kommunikationsschnittstelle durch Alternativen, wie z.B. durch einen CAN-Bus Controller anstatt des GEMAC Controllers [Mo09].

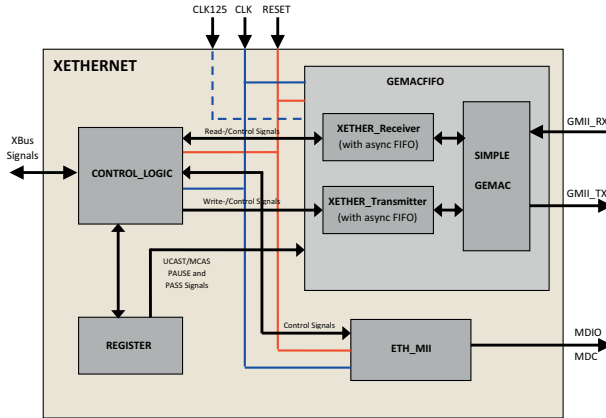


Abb. 3: XETHERNET Komponente

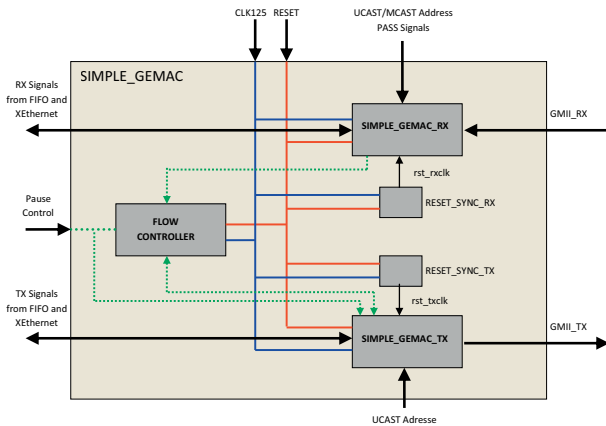


Abb. 4: SIMPLE.GEMAC Komponente

Abbildung 4 zeigt die SIMPLE.GEMAC Komponente. Diese wurde für die XETHERNET Komponente übernommen, da sie in anderen Projekten bereits ihre Funktionalität bewiesen hatte [Et09] [Wi16]. SIMPLE.GEMAC setzt sich im Wesentlichen aus einer Empfangs- und einer Sendeeinheit zusammen, die den Datenfluss steuern, d.h. Pausen-Pakete und Multi- oder UniCast-Signale empfangen und versenden, die Prüfsummen erzeugen und den Vorspann eines Ethernets-Paketes generieren.

4 Das Konzept der netzwerkkontrollierten FPGA-Rekonfiguration

4.1 Die Rekonfiguration des Spartan-6

Das Spartan-6 FPGA lässt sich dynamisch, d.h. im laufenden Betrieb, rekonfigurieren [Se06]. Im Gegensatz zu den Virtex FPGAs, bei denen die Partitionen direkt beschrieben werden können, wird bei den Spartan-6 FPGAs das "Differencing-Based Partial Reconfiguration" [Et07] Verfahren verwendet, d.h. die Rekonfigurationsinformationen werden durch Differenzbildung zweier Bitstreams generiert. Die Befehlssequenz im Bitstream wird dann in die interne ICAP Schnittstelle geladen. Der Vorteil der Differencing-Base Methode ist, dass während der Programmierung der Teil des FPGAs weiter arbeitet, der nicht geändert werden soll. Die Schnittstellen zwischen dem nicht konfigurierten und dem konfigurierten Teil sind nach der Rekonfiguration undefiniert und müssen z.B. durch ein Reset-Signal in einen definierten Zustand gebracht werden [UG15b].

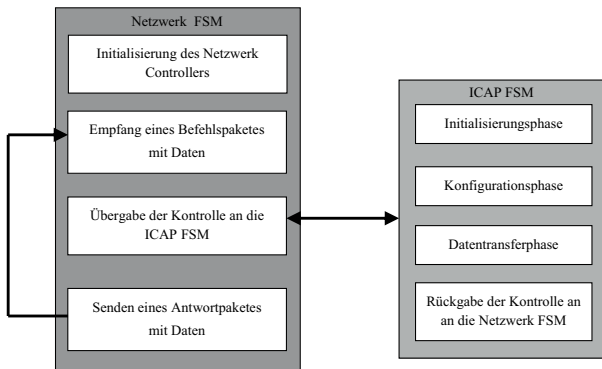


Abb. 5: Innere Struktur der XETHERNET_CONTROL Komponente

4.2 Die Rekonfigurations-Steuereinheit

Die Netzwerk FSM (Abbildung 5) hat vier übergeordnete Zustände: Initialisierung des Netzwerkcontrollers, Empfang von Befehlen/Daten, Übergabe der Kontrolle an die ICAP FSM und Senden von Antwortpaketen/Daten. Diese Zustände können je nach Bedarf unterschiedlich ausgeprägt werden, z.B. benötigt eine Ethernet Schnittstelle eine andere Initialisierung als eine CAN-Bus Schnittstelle. Die ICAP FSM folgt einer festen Abfolge von Schritten, um die Daten in den Rekonfigurationspeicher des FPGAs zu schreiben.

4.3 Steuerung der ICAP FSM

Die Steuerung der ICAP FSM orientiert sich an der fest definierten Struktur des Konfigurations-Bitstreams [UG15a] aus Tabelle 1:

Block	Beschreibung
DUMMY	16 Dummy Worte für BPI Address shift cycle
SYNC	Zwei Synchronisationsworte
HEADER	Initialisierung der Konfigurationsregister
CFG BODY	Start Adresse Schreib/Lese-Befehle FDRI/FDRO Inhalte des FPGA Konfigurationsspeichers AUTO CRC Worte
HEADER2	Initialisierung der Konfigurationsregister für die nächste Operation
DESYNC	Desynchronisierungsbefehl

Tab. 1: Bitstream Struktur des Spartan-6

Die Bitstream-Sequenzen lassen sich in standardisierte Blöcke gruppieren, sodass diese für verschiedene Einsatzzwecke, wie Lese- oder Schreiboperationen, wiederverwendet werden können. In der VHDL-Realisierung korrespondieren diese Blöcke mit den Zuständen der ICAP FSM. Abbildung 6 zeigt die derzeit definierten Abläufe:

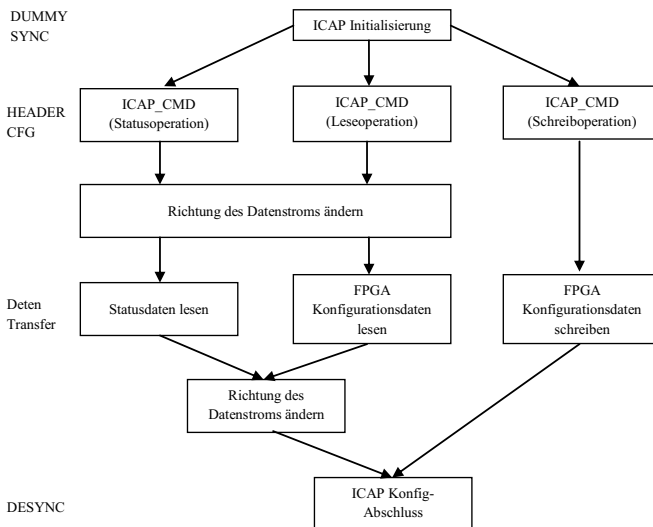


Abb. 6: Ablaufdiagramm der ICAP FSM Steuerung

1. **Konfigurations-Ablauf:** Der Zugriff auf das Konfigurationsregister erlaubt es, den Zustand und die Kennung eines FPGAs zu ermitteln [UG15a]. Die Netzwerk FSM empfängt durch die XETHERNET Komponente den Befehl, die Konfigurationsdaten zu ändern. Dies bewirkt, dass die ICAP FSM auf dem FPGA startet und die Blöcke im Pfad der Statusoperationen bearbeitet.

2. **Lese-Ablauf:** Die Leseoperation geschieht analog zum Konfigurations-Ablauf. Der Unterschied liegt im Detail, da nun der Konfigurationsspeicher anstatt der Statusregister gelesen wird. Der Ablauf ist im Pfad der Leseoperationen verdeutlicht.
3. **Schreib-Ablauf:** Bei der Schreiboperation ist die Anzahl der Blöcke im Pfad kleiner, sodass die ICAP-FSM in diesem Ablauf mit weniger Zuständen auskommt.
4. **Test-Ablauf:** Zurzeit existieren zwei Test-Abläufe. Mit dem ICMP-Ping oder dem ICAP-Ping testet man, ob das FPGA ansprechbar ist und die Netzwerk FSM aktiv ist. Der ICMP-Ping reagiert auf Pings von einem Kommandozeilen Werkzeug und der ICAP-Ping wird von einer Software-API heraus aufgerufen. Diese generiert ein ICAP-Netzwerkpaket, das während der Initialisierungsphase die Funktionsfähigkeit der FSM testet.

4.4 Steuerung der Netzwerk FSM durch ein Netzwerkprotokoll

Das Netzwerkprotokoll zwischen Host und FPGA arbeitet auf der IP Schicht des OSI-Referenzmodells und insbesondere auf der ICMP-Protokoll Ebene. Der Grund für diese Auswahl ist, dass man sich immer noch innerhalb eines weitverbreiteten Netzwerkstandards bewegt, sodass handelsübliche Geräte wie Router, Switches und Hub zur Kommunikation zwischen Host und FPGA oder zwischen FPGA und FPGA verwenden werden können, da die Adressierung über die IP- und MAC-Adresse des FPGAs stattfindet. Die nächsthöhere Protokollschicht im OSI-Referenzmodell [SZ14] wäre die TCP/UDP Schicht, die zwar mehr Funktionalität, z.B. Sockets, bietet, für die vorliegende Anwendung jedoch nicht nötig ist und nur zusätzliche Hardwareressourcen innerhalb des FPGAs zur Verarbeitung des Protokolls bindet.

	Bits 0-3	Bits 4-7	Bits 8-11	Bits 12-15	Bits 16-19	Bits 20-23	Bits 24-27	Bits 28-31
Ethernet Header (14 Bytes)	MAC Ziel						MAC Quelle	
	MAC Quelle				Ethernet Typ		Version	IHL
IP Header (20 Bytes)	Typ des Services		Länge				Identifikator	
	Identifikator		Flags and Offset				TTL	
	Protokoll (ICMP)		Prüfsumme				IP Quelladresse	
	IP Quelladresse						IP Zieladresse	
	IP Zieladresse						Nachrichtentyp	
ICAP Header	Nachrichtencode		Prüfsumme (RFC 1071)				ICAP Identifikation	
	ICAP Identifikation		ICAP Konfigurationsinfo					
	ICAP Konfigurationsinfo							
	ICAP Daten							

Abb. 7: ICAP Netzwerkprotokoll Headerstruktur

Abbildung 7 zeigt die Schichten anhand der Headerstruktur des ICAP Protokolls. Das ICAP Protokoll besteht aus zwei Schichten, die zum Data Link Layer und zum Network Link Layer des OSI-Referenzmodells korrespondieren [SZ14]. Im Network Link Layer benutzt das ICAP Protokoll nicht zugewiesene ICMP Nachrichtentypen und Nachrichtencodes des ICMP Standards [Po81]. Das ICAP und das ICMP Protokoll sind Teil des IP Protokolls und können aus diesem Grund die Paket Routing Funktionalität des IP Headers nutzen.

Die neuen Nachrichtencodes und -typen für Kommandos (Request) und Antworten (Reply) sind in Abbildung 8 und Abbildung 9 gezeigt. Zum Vergleich ist die ICMP-Ping Funktion zusammen mit den neuen ICAP Funktionen dargestellt.

	Typ	Code	Prüfsumme (16 Bit)	Identifikation (16 Bit)	Konfigurationsinfo		Daten
ICMP-Request-Ping	0x08	0x00	Wert	Wert	Sequenz-Nummer	Füllwerte	
ICAP-Request-Ping	0x18	0x00	Wert	Wert	Sequenz-Nummer	Füllwerte	
ICAP-Request-ReadConfig	0x18	0x01	Wert	Wert	RegisterNummer		
ICAP-Request-Status	0x18	0x02	Wert	Wert			
ICAP-Request-IDCODE	0x18	0x03	Wert	Wert			
ICAP-Request-Read	0x18	0x04	Wert	Wert	Major	Minor	
ICAP-Request-Write	0x18	0x05	Wert	Wert	Major	Minor	Data
ICAP-Request-WriteConfig	0x18	0x06	Wert	Wert	RegisterNummer	1. Wert	2-n. Wert

Abb. 8: Struktur des ICAP Header für Request Netzwerkpakete

	Type	Code	Prüfsumme (16 Bit)	Identifikation (16 Bit)	Konfigurationsinfo		Daten
ICMP-Reply-Ping	0x00	0x00	Wert	Wert	Sequenz-Nummer	Füllwerte	
ICAP-Reply-Ping	0x19	0x00	Wert	Wert	Sequenz-Nummer	Füllwerte	
ICAP-Reply-Config	0x19	0x01	Wert	Wert	RegisterNummer	1. Wert	2-n. Wert
ICAP-Reply-Status	0x19	0x02	Wert	Wert	1. Wert	2. Wert	3-n. Wert
ICAP-Reply-IDCODE	0x19	0x03	Wert	Wert	1. Wert	2. Wert	
ICAP-Reply-Read	0x19	0x04	Wert	Wert	Major	Minor	Data
ICAP-Reply-Write	0x19	0x05	Wert	Wert	Major	Minor	Status
ICAP-Reply-WriteConfig	0x19	0x06	Wert	Wert	RegisterNummer	Status	

Abb. 9: Struktur des ICAP Header für Reply Netzwerkpakete

Auf der Hostseite wird die Kommunikation zurzeit mit einer in C geschriebenen Software durchgeführt. Die Netzwerkkommunikation basiert auf Socket und PCAP Libraries von Windows und Linux. Auf der FPGA Seite wird die Steuerung zurzeit durch eine FSM durchgeführt. Die Quelldateien sind in VHDL implementiert.

5 Evaluierung der XETHERNET_ICAP Schnittstelle

Der FPGA Ressourcenverbrauch des vorgestellten Ansatzes ist in Tabelle 2 zusammengefasst. Die logischen Elemente (LUT) verbrauchen etwa 6% der verfügbaren Gesamtsourcen. Die IOs liegen bei 12% und die RAMB16 (18kb Block RAM) bei etwa 3%.

Typ der Logik	Verfügbar	Anzahl	Verbrauch in %
Anzahl der Slice Register	54.576	1.364	2%
Anzahl der Slice LUTs	27.288	1.746	6%
Verbraucher Speicher	6.408	46	1%
Anzahl der IOs	218	27	12%
Anzahl der RAMB16	116	3	3%

Tab. 2: Ressourcenverbrauch auf einem Spartan-6-slx45 FPGA

Abbildung 10 zeigt, wie die statische Logik in einem engen Bereich des FPGAs konzentriert ist. Dies wurde durch entsprechende Vorgaben während der Synthese erreicht. Ziel war es, die Logik in der Nähe der physikalischen Hardwareschnittstellen des Spartan-6 FPGA anzuordnen, um lange Signalwege und das Kreuzen von Signalen aus statischen und dynamischen Komponenten zu minimieren. Die Netzwerk I/O-Ports sind rechts unten von der Steuerlogik und die ICAP Komponente links unten davon untergebracht.

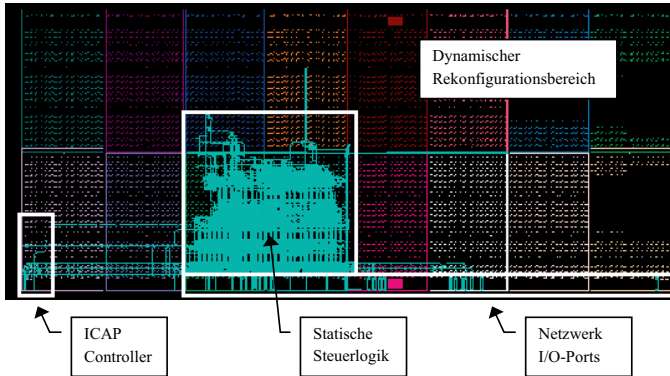


Abb. 10: Logik- und Schnittstellenverteilung auf dem FPGA

Durch die in Abbildung 10 gezeigte Anordnung der statischen Logik, lassen sich große freie Flächen für die dynamische Rekonfiguration auf der oberen FPGA-Hälfte schaffen. Die ICAP-Schnittstelle hat im Spartan-6 eine Arbeitsfrequenz von 20 MHz bei einer Datenbusbreite von 16 Bit. Eine Konfigurationsdatei enthält für diesen FPGA-Typ 1.484.525 Bytes, sodass in ca. 37 ms ein komplettes FPGA rekonfiguriert werden kann. Ein einzelner Frame hat eine Größe von 130 Bytes, sodass hier die Programmierung $3.25\mu\text{s}$ dauert. Die eingesetzte Ethernet Schnittstelle hat einen Arbeitstakt von 125 MHz bei 8 Bit Datenbreite, sodass ein Konfigurationsframe weniger als $1\mu\text{s}$ zur Übertragung in die XETHERNET_ICAP Komponente benötigt.

6 Zusammenfassung und Ausblick

Das in diesem Papier präsentierte Konzept zeigt wie synthetisierte Hardware "remote" über gängige Netzwerkschnittstellen dynamisch und im laufenden Betrieb in ein FPGA konfiguriert werden kann. Die bisher erforderliche Punkt-zu-Punkt Verbindung zwischen FPGA und Host entfällt. Vergleichbar mit der Aktualisierung von Software für klassische MCU/DSP, erlaubt dies insbesondere die kosteneffiziente und nachträgliche Optimierung, Erweiterung und Fehlerbehebung synthetisierter Hardware in komplexen vernetzten Systemen (z.B. ADAS in Fahrzeugen). Diverse Netzwerkschnittstellen können durch die flexible und modulare Struktur integriert werden. Hier sind im ADAS-Umfeld (u.a.) FlexRay, CAN, LIN, und Ethernet im Gebrauch.

6.1 Flexible Struktur

Die XETHERNET_ICAP Komponente ist modular entworfen und könnte neben dem Ethernet Controller oder dem CAN-Bus Controller auch andere Netzwerktechnologien innerhalb des XETHERNET Moduls einbinden. Das XETHERNET Modul benutzt eine einfache Busverbindung zur XETHERNET_CONTROL Komponente, sodass es sich für andere Anwendungen mit wenig Aufwand anpassen lässt.

6.2 Sicherheitsaspekte

Die FPGA Struktur ist aufgeteilt in einen dynamischen und statischen Teil (s. Abbildung 1 und Abbildung 10). Die statische Logik lässt sich bei der Initialisierung des FPGAs aus einem lokalen Flash-Speicher laden, der über das Netzwerk nicht erreichbar ist. Dieser Speicher lässt sich jedoch über eine separate Wartungsschnittstelle aktualisieren. Die Netzwerk FSM und die ICAP FSM lassen sich so konfigurieren, dass nur Änderungen in den von außen dynamisch rekonfigurierbaren Bereichen des FPGAs möglich sind. Damit ist es nicht möglich den statischen Bereich zu überschreiben und somit die Komponenten zur Rekonfiguration zu entfernen, bzw. zu modifizieren.

Neben der operativen Sicherheit gibt es auch Sicherheitsanforderungen entsprechend dem Standard ISO26262. FPGA Hersteller wie XILINX und Altera haben bereits ihre Entwicklungswerkzeuge und Methoden dementsprechend vom TÜV zertifizieren lassen [A113] [PB14].

6.3 Anwendungsszenarien

Die vorliegende Arbeit dient als Grundlage für die Entwicklung komplexer Systeme basierend auf vernetzter, rekonfigurierbarer Hardware. Abbildung 11 und Abbildung 12 zeigen Nutzungsmöglichkeiten im Rahmen verteilter Fahrerassistenzsysteme (ADAS). Anstelle von MCUs/DSPs übernehmen applikationsspezifisch konfigurierbare FPGAs die Datenverarbeitung und steuern sowohl Sensoren als auch Aktuatoren an.

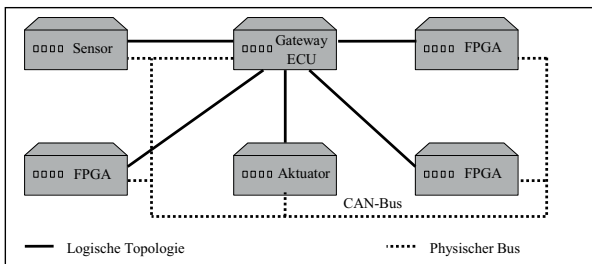


Abb. 11: Zentral kontrolliertes Netzwerk

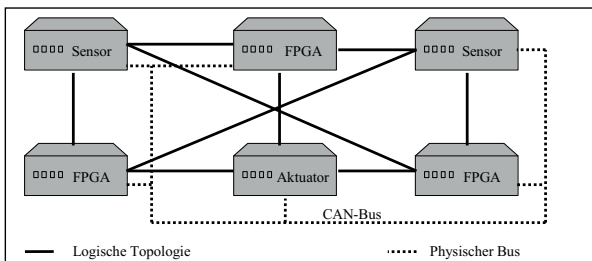


Abb. 12: Verteilt kontrolliertes Netzwerk

Im ersten Szenario (Abbildung 11) kontrolliert ein zentrales Steuergerät (ECU) sowohl die Konfigurationen der FPGAs als auch die Datenverteilung auf die entsprechend konfigurierten Verarbeitungseinheiten. Fällt eine Verarbeitungseinheit aus, wird ein anderes FPGA rekonfiguriert und die Daten flexibel zu diesem neuen Ziel umgeleitet. Das zweite Szenario (Abbildung 12) veranschaulicht die selbständige Zusammenarbeit der FPGAs ohne zentrales Steuergerät. Die Daten werden autonom von den Sensoren empfangen, im Netzwerk aufbereitet und an die Empfänger weitergeleitet. Durch ständige Selbstkontrolle wird überprüft, ob die Verarbeitungseinheiten noch einwandfrei funktionieren. Im Fehlerfall wird die Funktionalität eines defekten FPGAs durch ein anderes FPGA bereitgestellt. Dadurch wird das Gesamtsystem robust und ausfallsicher.

6.4 Ausblick

Die nächsten Schritte sind, weitere Netzwerkschnittstellen an die FPGA-Architektur anzubinden. Beispielsweise sollen mehrere CAN-Bus Module in ein FPGA integriert werden, um ein Netzwerk von selbständig arbeitenden und selbstkontrollierenden FPGAs aufzubauen. Das Kommunikationsprotokoll würde für diesen Zweck um Befehle für die Selbstkontrolle erweitert.

Aktuelle MCUs/DSPs bewegen sich mit ihrer Rechenleistung im Bereich von etwa 1.5 GFlops/Core (ARM Cortex-A15). Hingegen können die FPGAs, wie auch die aktuellen Grafikprozessoren (GPUs) Rechenleistungen im Bereich von TFlops (FP32) bieten. Die Kosten für FPGAs und GPU sind zur Zeit im Bereich um 0.30€/GFlops [GP16]. Durch den Rechenleistungshunger von ADAS Systeme werden die FPGAs wegen ihrer Flexibilität und Leistungsfähigkeit aber immer attraktiver.

Literaturverzeichnis

- [Al13] Alteras Leadership in Functional Safety Expands to ISO 26262, 2013.
www.altera.com/solutions/industry/automotive/applications/safety/overview.tablet.html,
Stand: 17.06.2016.
- [An10] Angermeier, Josef; Bobda, Christophe; Majer, Mateusz; Teich, Jürgen: Erlangen Slot Machine: An FPGA-Based Dynamically Reconfigurable Computing Platform. In (Platzner, Marco; Teich, Jürgen; Wehn, Norbert, Hrsg.): Dynamically Reconfigurable Systems: Architectures, Design Methods and Applications. Springer Netherlands, Dordrecht, S. 51–71, 2010.
- [BY08] Bayar, Salih; Yurdakul, Arda: Self-reconfiguration on Spartan-III FPGAs with compressed partial bitstreams via a parallel configuration access port (cPCAP) core. In: Research in Microelectronics and Electronics, 2008. Ph.D. S. 137–140, June 2008.
- [Et07] Eto, Emi, XAPP290, Difference-Based Partial Reconfiguration, 2007.
www.xilinx.com/support/documentation/application_notes/xapp290.pdf,
Stand: 11.06.2016.
- [Et09] Ettus, Matt, usrp2 (simple_gemac), 2009.
www.github.com/EttusResearch/fpga/tree/master/usrp2, Stand: 11.06.2016.

- [FF12] Fons, Francisco; Fons, Mariano: FPGA-based Automotive ECU Design Addresses AUTOSAR and ISO26262 Standards. *Xcell Journal* 12/01, S. 20–31, 2012.
- [GP16] GPU vs FPGA Performance Comparison, 2016. www.bertendsp.com/pdf/whitepaper/BWP001_GPU_vs_FPGA_Performance_Comparison_v1.0.pdf, Stand: 17.06.2016.
- [Hü10] Hübner, Michael; Göhringer, Diana; Noguera, Juanjo; Becker, Jürgen: Fast dynamic and partial reconfiguration data path with low hardware overhead on Xilinx FPGAs. In: *Parallel Distributed Processing, Workshops and Phd Forum (IPDPSW)*, 2010 IEEE International Symposium on. S. 1–8, April 2010.
- [Li11] Lieber, Peter Andrew: , *FPGA Communication Framework for Communication, Debugging, Testing, and Rapid Prototyping*, 2011. scholarsarchive.byu.edu/etd/3039, Stand: 11.06.2016.
- [Mo09] Mohor, Igor: , *CAN Protocol Controller*, 2009. www.opencores.org/project,can, Stand: 10.05.2016.
- [Mo11] Mohor, Igor: , *Ethernet MAC 10/100 Mbps*, 2011. www.opencores.org/project,ethmac, Stand: 10.05.2016.
- [PB14] PB015, Xilinx All Programmable Functional Safety Design Flow Solution, 2014. www.xilinx.com/publications/prod_mktg/safety-guidelines.pdf, Stand: 17.06.2016.
- [Po81] Postel, Jon: , *Internet Control Message Protocol Specification (RFC792)*, 1981. tools.ietf.org/html/rfc792, Stand: 11.06.2016.
- [Se06] Sedcole, Nicholas Pete: , *Reconfigurable Platform-Based Design in FPGAs for Video Image Processing (Chapter 6)*, PHD Thesis, 2006. cas.ee.ic.ac.uk/people/nps/publications.html, Stand: 14.10.2014.
- [SZ14] Spurgeon, Charles; Zimmerman, Joann: *Ethernet: The Definitive Guide, Second Edition*. O'Reilly Media, Inc, 2014.
- [UG15a] UG380, Spartan-6 FPGA Configuration, Version 2.8, 2015. www.xilinx.com/support/documentation/user_guides/ug380.pdf, Stand: 11.06.2016.
- [UG15b] UG702, Partial Reconfiguration User Guide, Version 14.1, 2015. www.xilinx.com/support/documentation/sw_manuals/xilinx14_1/ug702.pdf, Stand: 11.06.2016.
- [We13] Weber, Marc: *AUTOSAR lernt Ethernet*. In: *HANSER automotive networks*. Hanser, S. 30–33, 2013.
- [Wi16] Williams, Joel: , *Digilent Atlys*, 2016. www.joelw.id.au/FPGA/DigilentAtlysResources, Stand: 11.06.2016.

Systematische und effiziente Erhebung hochgenauer Straßengeodaten für Geoinformationssysteme und Fahrsimulationen

Andreas Richter¹ und Michael Scholz¹

Abstract: Die Entwicklung von Fahrerassistenz- und Automationssystemen benötigt domänenspezifische Datenformate für Fahrsimulation und Automation, die ein Straßennetzwerk logisch und vor allem geometrisch genau beschreiben. Da immer komplexere Situationen wie städtische Umgebungen benötigt werden und die dafür notwendigen Informationen mit der geforderten Genauigkeit nicht immer aus aktuell verfügbaren Geodaten bezogen werden können, sind spezielle Vermessungen notwendig. Um diese Datenerhebungen kosten- und zeiteffizient durchführen zu können, hat das DLR zusammen mit OEMs einen Leitfaden für die Erhebung und Konvertierung von Straßengeodaten in ein leichtgewichtiges Zwischenformat entwickelt, um daraus automatisiert in domänenspezifische Formate konvertieren zu können. Der Leitfaden wird mittels prototypischer Erhebung und anschließenden Datentransformation nach OpenDRIVE einer deutschen Stadtstrecke validiert und steht danach zur freien Verfügung durch Dritte.

Keywords: Fahrsimulation, Straßenbeschreibung, Geodaten, Leitfaden, OpenDRIVE

1 Einleitung

Die aktuell in der Entwicklung befindlichen Fahrerassistenz- und Automationssysteme werden immer komplexer, entsprechend werden auch der Test und die Absicherung immer komplizierter. Als allgemein akzeptiert gilt, dass wesentliche Bestandteile der Entwicklung und Tests durch Simulationen unterstützt werden, um den Zeit- und Kostenrahmen nicht zu sprengen. Entsprechend muss also auch in der Simulation die Komplexität abgebildet werden. Dies betrifft beispielsweise auch die Simulationsumgebung mit komplexen Straßennetzwerken, um urbane Umgebungen abbilden zu können. Der Einsatz von bereits existierenden Geodaten erscheint daher naheliegend, um mittels automatisierter Werkzeugkette benötigte 3D-Umgebungen und logische Straßennetze zu generieren und dadurch den immensen Zeitaufwand bei der manuellen Erstellung zu reduzieren. Das Projekt „Virtuelle Welt“ zeigt beispielhaft, welche Geodaten genutzt werden können und welche Verarbeitungsschritte notwendig sind, um das gewünschte Ergebnis zu erreichen [Ri16]. Es zeigt sich jedoch auch, dass es weiterhin keine ideale Datenquelle gibt, auch wenn immer mehr Geodaten digital verfügbar gemacht werden. Bei öffentlichen, kommerziell verfügbaren sowie

¹ Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Institut für Verkehrssystemtechnik, Technologiefeld Datenerfassung und Informationsgewinnung, Lilienthalplatz 7, 38108 Braunschweig {andreas.richter, michael.scholz}@dlr.de

Katasterdaten müssen hinsichtlich der Genauigkeit oft Abstriche gemacht werden oder es fehlen zum Teil wichtige Detailinformationen, die für Fahrsimulatorumgebungen notwendig sind. Daher werden für Simulatordatenbasen häufig spezielle Vermessungen durchgeführt, um alle wesentlichen Informationen in der benötigten Genauigkeit aus einer Hand zu erhalten. Meist werden diese jedoch nur manuell aufbereitet, in proprietären Datenformaten gespeichert, und wegen des Aufwands nur kleinräumig erhoben. Die speziellen Datenformate lassen sich in der Regel nicht in bestehende amtliche Daten integrieren und eine Vermessung zur Aktualisierung amtlicher Daten erhebt weiterhin nicht für die Simulation relevante Daten.

Ein Leitfaden zur Erhebung von Geodaten soll diese unterschiedlichen Welten zusammenbringen: Großflächig erhobene, amtliche Geoinformationssystem- (GIS) bzw. CAD-Daten, die zusätzliche Detailinformationen beinhalten, um diese Daten auch für Simulationszwecke und als hochgenaue Karte für das automatische Fahren nutzen zu können. Dabei muss die Erhebung auf einem einfachen Prinzip basieren, damit die Kosten für großräumige Vermessung nicht ins Unermessliche steigen, aber dennoch eine Konvertierung in die Zielformate einfach und möglichst automatisiert erfolgen kann.

1.1 verfügbare Geodaten

Im Zeitalter der „Industrie 4.0“ ist die Digitalisierung auch von Geodaten nicht mehr aufzuhalten, zunehmend sind Geodaten frei verfügbar, aber auch das kommerzielle Angebot wächst stetig. Kommunen nutzen digitale Stadtmodelle, um bspw. Solarpotentialanalysen oder Lärmausbreitung berechnen zu können. Stadtgrün, Beschilderung und Lichtsignalanlagen sowie Stadtbeleuchtung werden in Geodatenbanken verwaltet, Stadtpläne mit genauer Straßentopographie digitalisiert. Aus Befliegungen werden nicht nur Luftbilder erstellt, sondern auch digitale Geländemodelle abgeleitet.

Qualität und Umfang der Daten schwanken jedoch von Kommune zu Kommune, je nachdem, wie viel Aufwand in die Erhebung und Speicherung der Daten fließen kann. So werden beispielsweise Verkehrsschilder und Lichtsignalanlagen in Braunschweig in einer Geodatenbank mit den essentiellen Informationen (Schildtyp, Ausrichtung, usw.) georeferenziert abgelegt, in Stuttgart jedoch liegen die Daten in CAD-Kacheln vor und sind teilweise nur als Grafik eingezeichnet. Oft leiten sich diese Daten aus historischen, analog gepflegten Datensätzen ab und somit ist ein maschinelles Auslesen der essentiellen Informationen nur mittels komplexer Algorithmen möglich.

Die Automobilbranche nutzt inzwischen seit Jahrzehnten Geodaten. Ein Fahrzeug ohne Navigationsgerät ist heute außerhalb der Sportwagendomäne kaum noch vorstellbar. Die großen Kartenhersteller wie TomTom und HERE liefern in klassischen Datenbasen großflächige Knoten-Kanten-Modelle mit für die Navigation hilfreichen Zusatzinformationen (Anzahl der Fahrstreifen und deren Typ) sowie „points of interest“. Hinzu kommen so genannte „high definition maps“, in denen weitere Informationen wie ein Streckenhöhenprofil oder Krümmungsradien für eine

Unterstützung der Fahrweise enthalten sind. Hochgenaue geometrische Informationen sind zurzeit nicht enthalten, es sind beispielsweise gerade einmal Anfang und Ende einer Fahrbahnaufweitung als punktuelle Information hinterlegt. Für Navigationsaufgaben ist diese vereinfachte Form der Modellierung ausreichend, da die ungenaue GNSS-Position einfach auf das Knoten-Kanten-Modell gemached werden kann und spurgenaue Informationen rein informativ angezeigt werden. An der Integration geometrischer Informationen wird jedoch gerade intensiv gearbeitet.

Nicht zu unterschätzen sind allerdings auch so genannte „crowd sourced“ Daten wie beispielsweise OpenStreetMap (OSM). Gerade in Deutschland ist die OSM-Community sehr aktiv und digitalisiert beispielweise Innenstädte sehr detailliert. Dabei werden auch die Informationen über die Straßennetze immer genauer. Informationen über einzelne Fahrbahnen, deren Verknüpfung und Typ sind bereits im Datenmodell hinterlegt, allerdings fehlen auch hier noch genaue geometrische Informationen [RFS16]. Kommerzielle Kartenportale, wie von Google oder Apple betrieben, kaufen die Daten bei Dritten ein.

1.2 genutzte Datenformate

Möchte man nun die Geodaten für Fahrsimulationen nutzen, müssen bei der Datentransformation einige Informations- und Abbildungslücken überbrückt werden [FR12]. Für Fahrsimulationen kommen oft spezielle Datenformate zum Einsatz (beispielsweise OpenDRIVE, IPGRoad oder RoadXML), deren Kern eine mathematische Abbildung des Straßenverlaufs ist, an den relativ verortet alle wesentlichen Elemente und Eigenschaften der Straße modelliert werden. Die Elemente werden dann zusätzlich logisch miteinander verknüpft, um für die Simulation einfach auslesbar zu werden. Die mathematische Modellierung kann dabei sehr präzise oder generisch ausgelegt werden.

Kommunale Geodaten sind im Gegensatz dazu oft historisch gewachsen und basieren häufig auf dem Ansatz, die Informationen rein visuell und flächig in Karten darzustellen. Beispielsweise wird ein Straßenbereich in einem einfachen CAD- oder GIS-Format visuell nur durch die begrenzenden Bordsteinkanten beschrieben, jedoch gibt es keine logische Verknüpfung zur tatsächlichen Fläche oder auch zu Daten über Lichtsignalanlagen. Auch können solche Datensätze in Teilen unsauber und mit verschiedenen Methoden erhoben sein, daher Fehler sowie heterogene Genauigkeit aufweisen. Eine Transformation in die Simulatorformate muss zusätzlich noch die logische Verknüpfung hinzufügen. Eine mathematische Modellierung, wie in der Simulation genutzt, wäre für kommunale Anforderungen oft zu aufwendig und schwieriger zu handhaben.

Die Knoten-Kanten-Modelle der Kartenhersteller oder von OSM basieren logischerweise auf Graphen-Abbildungen. Für das automatische Fahren besteht jedoch einhellige Meinung, dass genauere Daten von Nöten sind, daher entwickeln sich auch die Navigationsdatenbasen weiter. Der Navigation Data Standard (NDS) [NA16]

beispielsweise erweitert das Knoten-Kanten-Modell ab Version 2.5 in den unteren Genauigkeitsebenen um geometrisch modellierte Verläufe von Begrenzungen, die an logische Kanten angehängen werden (so genanntes „Lane Product“). Die Daten werden in einem binären Tabellenformat gespeichert. Zurzeit modelliert NDS allerdings nur den Fahrbahnbereich, eine Abbildung von daran anschließenden Räumen und weiteren Objekten findet nicht statt.

Somit haben die Datenformate für jede Domäne ihre Vorteile, die für die anderen Einsatzgebiete eher nachteilig sind, weil die Daten entweder zu komplex modelliert werden oder wesentliche Informationen wie Zusammenhänge fehlen.

2 Datenerhebung

Wie in der Einleitung bereits geschrieben, ist eine Konvertierung von unterschiedlichen Geodaten in eine Fahrsimulatorumgebung möglich. Wie außerdem kurz angerissen, ist eine mobile Vermessung der Straße eine gute Alternative, um alle notwendigen Daten aus einer Hand zu erhalten. Da jedoch die Vermessung historisch mit der CAD-Domäne eng verbunden ist, ist auch hier die Abbildung der Messrohdaten in Spezialformate wie OpenDRIVE für die Dienstleister nicht trivial. Somit wird auch die Vermessung für Automobilentwickler wieder kostenintensiv und zeitaufwändig. Beauftragen Kommunen eine Vermessung und lassen sich die Daten wieder in einem klassischen CAD-Format liefern, kann die Simulatordomäne weiterhin nur mit Aufwand etwas damit anfangen. Die flächendeckende Erhebung von genaueren Navigationsdatenbasen steht noch am Anfang, selbst Google baut sich für seine autonomen Fahrzeuge die Karten von Hand [Go16].

Damit nicht jede Domäne eine eigene Vermessung starten muss und dadurch unnötige Doppelarbeit entsteht, sollen die Welten amtlicher, kommunaler und kommerzieller CAD- und GIS-Daten mit den Fahrsimulator-Daten zusammengebracht werden. Die Automobilbranche ist auf eine kosteneffiziente Methode zur Erhebung von Geoinformationen angewiesen, für Kommunen eröffnen sich neue Einsatzfelder für ihre Geodaten, wenn diese wesentliche Informationen enthalten würden. Das DLR hat in Zusammenarbeit mit den OEMs Audi, BMW, Daimler, Porsche und Volkswagen das Projekt „Road2Simulation“ gestartet, um eine Methode zur Überbrückung der Kluft zwischen den Domänen und Ihren unterschiedlichen Anforderungen an die Daten zu erarbeiten. Das Ergebnis ist ein Leitfaden [RS16] zur systematischen Erhebung, egal ob dies mit einer mobilen Vermessung oder beispielsweise durch Digitalisieren von Plänen oder hochgenauen Luftbildern passiert. Der Leitfaden ist auch für Dritte frei verfügbar.

2.1 Kernelemente des Leitfadens

Die grundlegende Idee des Leitfadens ist eine vereinfachte Vermessung durch einen externen Dienstleister sowie eine automatisierte Datentransformation durch den

Endanwender. Mit Hilfe eines definierten aber einfachen Zwischenformats soll die Datenkonvertierung ermöglicht werden, da in diesem Format alle wesentlichen Detailinformationen enthalten sind. Das Zwischenformat wurde auf das OpenDRIVE-Format [Du15] für die Straßenverlaufsbeschreibung und auf das OpenCRG-Format [Va15] für die Straßenoberflächenbeschreibung ausgerichtet, da es sich bei beiden Formaten um de-facto Standards in der Fahrsimulation handelt. Hinzu kommt, dass diese Formate frei verfügbar sind. Abb. 1 zeigt die Schlüsselemente des Leitfadens für den Straßenverlauf und deren Verknüpfung untereinander.

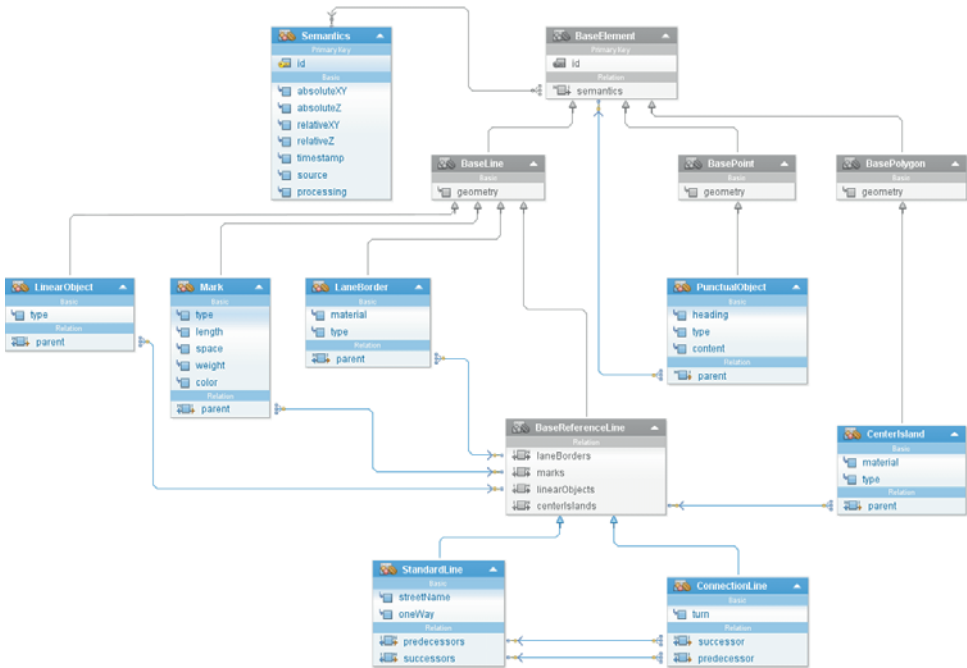


Abb. 1: Auf das wesentliche reduzierte Datenmodell des Leitfadens mit Schlüsselementen (blau) und hierarchischen Elementen (grau).

Basis für den Leitfaden sind so genannte Polylinien mit heterogener Punktdichte der Messpunkte, um den Verlauf der Straße zu modellieren. Für genaue und effiziente Modellierung sollen die Stützstellen eines Linienzugs im Krümmungsbereich möglichst dicht sein, in geraden Abschnitten möglichst groß. Diese unterschiedliche Dichte der Stützstellen der Polylinie ermöglicht sowohl eine hochgenaue Beschreibung als auch eine vereinfachte Darstellung, wenn Stützstellen entsprechend nachträglich interpoliert werden. Auch eine nachträgliche Transformation in eine mathematische Beschreibung kann entweder sehr genau dem Linienzug folgen oder ihn stärker abstrahieren.

Fahrstreifen, Markierungen, Begrenzungen und andere lineare Objekte bilden das Hauptgerüst. Hinzu kommen Punktkoordinaten, die Straßeninfrastrukturelemente wie

Beschilderung, Lichtsignalanlagen, spezielle Straßenmarkierungen und Straßenmöblierung beschreiben. Einige Bestandteile der Straße wie beispielweise Verkehrsinseln sind zurzeit in Simulationsformaten schwer in ihrer Erscheinung zu modellieren. Daher werden sie als flächige Elemente erhoben, um keine Informationen im Zwischenformat zu verlieren. Die Elemente werden mit Attributinformationen versehen. Wichtige Information ist dabei ihre logische Zugehörigkeit zu den anderen Elementen, aber auch Informationen über ihren Typ und ihr Aussehen werden gespeichert. Eine Fahrbahn wird somit durch ihre begrenzende Kante beschrieben, die Informationen zum Typ der Fahrstreifen (beispielsweise Richtungsfahrbahn, Parkstreifen, Grünstreifen oder Gehweg) als auch eine Angabe zum verwendeten Material der Fahrstreifen (beispielsweise Asphalt, Beton oder Kleinpflaster) enthält. Zusätzlich wird auch noch eine Aussage über die Datenqualität verknüpft, die als „Datensemantik“ bezeichnet wird. Die Datensemantik beschreibt Aufnahmezeitpunkt der Information, relativen und absoluten Fehler sowie Methode der Messung (beispielsweise Angabe des Sensors inklusive Firmware-Versionsnummer) und Nachverarbeitung (beispielsweise: roh, gesäubert, verarbeitet oder fusioniert). Somit kann jeder Nutzer im Nachgang selbst entscheiden, ob die Daten seinen Ansprüchen genügen.

2.2 Anwendungsbeispiele und Modellierungshilfen

Ein Leitfaden kann jedoch kein theoretisches Konstrukt bleiben und wird erst dann nutzbar, wenn er durch Beispiele begreifbar wird. Entsprechend wurden für alle wesentlichen Straßensituationen in Deutschland Beispiele zur Umsetzung entworfen. Zu den Situationen gehören, Straßen mit zwei oder mehr Fahrstreifen inklusive Einbahnstraßen, Autobahnen mit baulicher Trennung, komplexe Kreuzungen, Kreisverkehre usw. An diesen Beispielen wird Schritt für Schritt erklärt, wie die Messrohdaten aufbereitet werden sollten und welche Informationen, Details und Besonderheiten erhoben werden müssen.

Die Grundelemente des Leitfadens sind der Straßenverlauf zwischen Kreuzungen, Fahrstreifenverknüpfungen in den Kreuzungen, Fahrstreifenbegrenzungen, Fahrstreifenmarkierungen, lineare und punktuelle Objekte sowie spezielle bauliche Trennungen. Diese wesentlichen Grundelemente müssen teils in bestimmter Art und Weise modelliert werden, um eine automatische Datentransformation unterstützen zu können. Ein Straßenverlauf modelliert immer die Verbindung zwischen zwei Kreuzungen, es werden nie mehrere hintereinander gegangen. Der Verlauf der Straße wird als Linienzug modelliert, der entweder dort, wo die Fahrstreifen in Hin- und Rückrichtung einer Straße aneinander angrenzen, liegt, so lange es keine bauliche Trennung gibt. Gibt es eine Trennung oder handelt es sich um eine Einbahnstraße, muss der Linienzug an die linke Seite der am weitesten links liegenden Fahrstreifen gelegt werden. Abb. 2 zeigt einige Beispiele unterschiedlicher Straßensituationen.



Abb. 2: unterschiedliche Straßensituationen mit modelliertem Straßenverlauf (grün) und Fahrstreifenbegrenzungen (türkis)

Die Fahrstreifenbegrenzung ist das Element, das der Verkehrsfläche seine Erscheinung gibt. In der Regel folgen diese Linienzüge dem Verlauf der Straße, des Bürgersteigs oder auch Fahrstreifenmarkierungen, es besteht aber auch die Möglichkeit, sehr akkurat die Außenkante beispielsweise einer Asphaltfläche zu modellieren, falls dies für bestimmte Anwendungen notwendig ist. Dies kann beispielsweise im Zusammenhang mit Oberflächenvermessungen der Fall sein. In Abb. 2 unten links ist beispielhaft eine Parkplatzausbuchtung zu erkennen.

Kreuzungen mit ihren Abbiegefahrstreifen benötigen besondere Aufmerksamkeit, denn oft sind in der Realität nicht alle für die Simulatorformate notwendigen Informationen Modellierungen vorhanden. Abb. 3 zeigt eine Kreuzung mit allen notwendigen Verbindungen pro Himmelsrichtung. Dort sind beispielsweise Wendemöglichkeiten modelliert, die in der Realität vielleicht nicht zugelassen sind, aber das Aussehen der baulichen Trennung modellieren. Auch müssen für eine saubere Modellierung alle Fahrstreifen möglichst auf einer Höhe enden, durchlaufende Fahrstreifengrenzen müssen dort also unterbrochen werden.

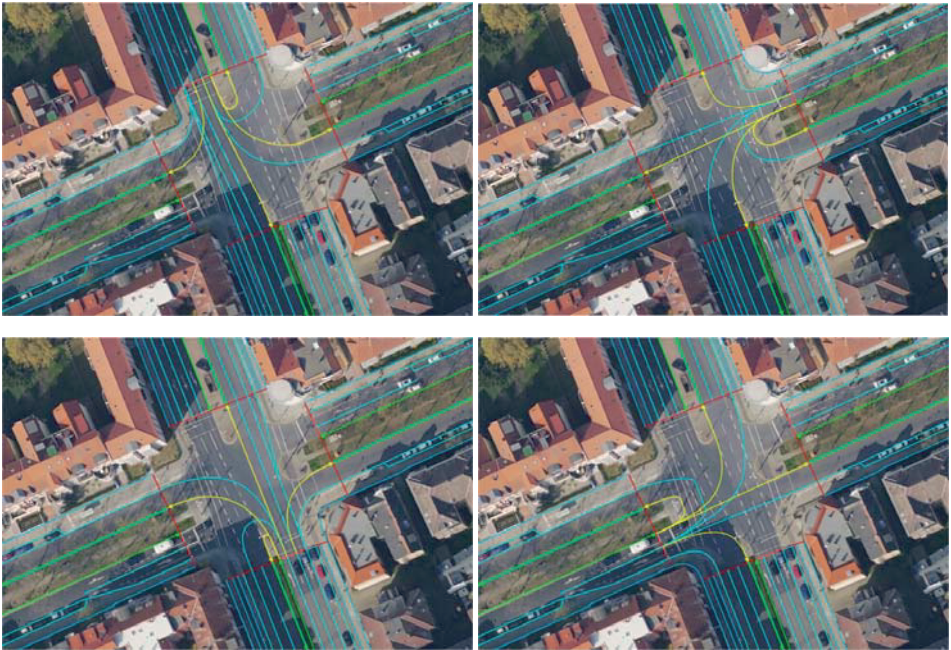


Abb. 3: Kreuzung mit allen Verbindungen (gelb) pro Himmelsrichtung

Komplexe Kreuzungen müssen darüber hinaus möglichst in kleinere Kreuzungen unterteilt werden. Beispielsweise Kreisverkehre sowie Autobahnauf- und -abfahrten (siehe Abb. 2 unten rechts) lassen sich nur auf diese Weise vernünftig modellieren.

Bei den Punktinformationen zur Straßeninfrastruktur zeigt sich schnell, welche komplexe Umgebung abgebildet werden soll. Zur logischen Abbildung sind prinzipiell nur Position und Typ notwendig, für eine vernünftige graphische Modellierung kommen jedoch weitere Informationen über Ausrichtung und Zugehörigkeit hinzu. Abb. 4 zeigt die Fülle an Punktinformationen für nur einen Kreuzungsast.

Um den Aufwand für die Datenerhebung an den Bedarf anpassen zu können, gibt es einen so genannten Standardabbildungsumfang und einen erweiterten Abbildungsumfang. Im Standard werden nur die für die Straßenmodellierung wesentlichen Informationen erhoben, somit ist sozusagen an der Bordsteinkante Schluss, es werden nur Beschilderung und Lichtsignalanlagen berücksichtigt. Nur der erweiterte Abbildungsumfang gibt das vollständige Bild des Verkehrsraums wider und kann je nach Bedarf des Zielsystems modular erweitert oder reduziert werden.



Abb. 4: markierte Punktinformationen (magenta) für Beschilderung, Lichtsignalanlagen und Stadtmöblierung

3 Anwendung und Ausblick

Der Leitfaden wird zudem auch in realer Durchführung getestet. Im Projekt „Road2Simulation“ wurde eine 14 km lange Referenzstrecke in Stuttgart bestehend aus städtischen und Landstraßenabschnitten sowie Tunnel und Autobahn ausgewählt, um die Erhebung nach dem Leitfaden und anschließende Datenkonvertierung zu testen. Die Vermessung startet im zweiten Quartal 2016. Dafür wurde ein Bieterwettbewerb gestartet, in dem sich mögliche Datenlieferanten durch grafische Bearbeitung einer Beispielkreuzung qualifizieren mussten. Dadurch kann sichergestellt werden, dass die Vermessung auch korrekt durchgeführt wird. Die anschließende Konvertierung der entsprechend aufbereiteten Daten wird im dritten Quartal 2016 erwartet. Es ist davon auszugehen, dass die Kosten und Aufwände deutlich geringer sein werden, als in früheren Beauftragungen, in denen die Spezialformate direkt geliefert werden mussten. Zusätzlich werden die erhobenen Daten nicht nur für die Fahrsimulation genutzt werden können, sondern können zugleich leicht beispielsweise in kommunale Geodateninfrastrukturen integrierbar sind.

Durch die Vermessung wird der Leitfaden nicht nur validiert, sondern basierend auf den Erfahrungen bei der Durchführung auch aktualisiert. Auch zukünftige Versionen des Leitfadens bleiben für Interessenten frei Verfügbar und werden sowohl in der Automobilbranche als auch bei den Kommunen beworben, um dort Anwendung zu finden. Denn ein Loslösen von Spezialanbieter, auch wenn es die großen Kartenhersteller sind, ist eine sinnvolle Lösung, um relevante und notwendige Daten flächendeckend zu erheben und aktuell vorhalten zu können. Dies soll wie anfangs

bereits beschrieben eine Win-Win-Situationen für alle Beteiligten schaffen und den Weg freimachen, dass Dritte Interesse an maschinell auswertbaren, amtlichen Geodaten haben und ggf. selbst welche beisteuern.

Der Leitfaden wird in jedem Fall für entstehende Testfelder und entsprechende Projekte eingesetzt werden, in denen hochgenaue Daten in unterschiedlichen Formaten für den Test von automatischen Fahrfunktionen benötigt werden. Daher wird der Leitfaden um Anforderungen weiterer Formate wie beispielsweise NDS erweitert werden, sobald die Definition für die fahstreifengenaue Modellierung fest stehen. Da die unterschiedlichen Simulator- und Navigationsdatenformate auch langfristig ihre Existenzberechtigung behalten werden, ist eine entsprechende Berücksichtigung im Leitfaden nur folgerichtig. Auch die Erweiterung um Besonderheiten weiterer Länder sind geplante Schritte.

4 Zusammenfassung

Die Notwendigkeit der Nutzung von Realdaten zum Test von Assistenz- und Automationssystemen wächst rasant. Da zurzeit keine sofort nutzbare Datenquelle für Fahrsimulatoranwendungen existiert und die spezifische Erhebung kostenintensiv und zeitaufwändig ist, sollte ein Ansatz entwickelt werden, der die Komplexität reduziert und einfache Datenkonvertierung ermöglicht, da es eine entsprechende allgemein zugängliche Methode bis jetzt nicht gab. Das DLR entwickelte in Zusammenarbeit mit OEMs einen Leitfaden, deren Kernelement ein einfaches Geodatenzwischenformat alle wesentlichen Detailinformationen beinhaltet, die für die Datenkonvertierung notwendig sind. Der frei verfügbare Leitfaden enthält neben dem Modell viele Modellierungsbeispiele und Umsetzungsvorschläge zur einfachen Adaption. Er wird mit einer prototypischen Vermessung und Datenkonvertierung getestet und danach gegebenenfalls angepasst. Somit steht nun eine allgemein zugängliche Methode zur effizienten Erhebung von Straßengeodaten für Geoinformationssysteme und Fahrsimulation.

Literaturverzeichnis

- [Du15] Dupuis, M.; et.al.: OpenDRIVE® Format Specification, Rev. 1.4 Issue H, <http://opendrive.org/docs/OpenDRIVEFormatSpecRev1.4H.pdf>, Stand 15.06.2016
- [FR12] Friedl, H.; Richter, A.: Fusion heterogener Geodaten zur Erstellung realer 3D-Welten am Beispiel einer Fahrsimulation. GEOINFORMATIK 2012, 28-30.03.2012, Braunschweig, ISBN 978-3-8440-0888-3. ISSN 1618-1034
- [Go16] Google Self-Driving Car Project. Monthly Report March 2016, <https://static.googleusercontent.com/selfdrivingcar/files/reports/report-0316.pdf>, Stand 15.06.2016

- [NA16] NDS Association: On the way – NDS success stories from 2014, http://www.nds-association.org/wp-content/uploads/NDS_Flyer-Success-stories_2014.pdf, Stand 15.06.2016
- [RFS16] Richter, A.; Friedl, H.; Scholz, M.: Beyond OSM – Alternative Data Sources and Approaches Enhancing Generation of Road Networks for Traffic and Driving Simulations. SUMO2016 – Traffic, Mobility, and Logistics. 23-25.05.2016, Berlin
- [Ri16] Richter, A.; et.al.: Challenges and Experiences in Using Heterogeneous, Geo-Referenced Data for Automatic Creation of Driving Simulator Environments. In: Transactions of The Society for Modeling and Simulation International, Special Issue Driving Simulation, 2016, vorveröffentlicht 08.04.2016, DOI: 10.1177/0037549716641201
- [RS16] Richter, A.; Scholz, M.: Road2Simulation Guidelines – Leitfaden zur Erhebung von Straßendaten für Simulation und Entwicklung, http://www.dlr.de/ts/Portaldata/16/Resources/projekte/road2simulation/Road2Simulation_Guideline.pdf, Stand 15.06.2016, Braunschweig, Germany
- [Va15] various: OpenCRG® User Manual, Issue M, <http://www.vires.com/opencrg/docs/OpenCRGUserManual.pdf>, Stand 15.06.2016

Adding Atmospheric Scattering and Transparency to a Deferred Rendering Pipeline for Camera Based ADAS Tests

Sabrina Heppner,¹ Marius Dransfeld² and Gitta Domik³

Abstract: This paper describes the implementation of a real-time global illumination model (atmospheric scattering and transparency) based on deferred rendering for the realization of virtual test drives used for the evaluation of camera based ADAS. Atmospheric scattering and transparent objects (e.g. windows) contribute to a realistic visualization of outdoor street scenes. However, both are not considered in the basic deferred rendering pipeline. To create an adequate adaption, the influence of the atmospheric scattering on transparent objects has to be considered. This work compares several techniques for the integration of transparency in the deferred rendering global illumination model. The realization of an A-Buffer implemented as Linked List provides the best results concerning the image quality. Weighted Blended Order-Independent Transparency also achieves adequate results. Inferred Lighting is inapplicable for the given use case.

Keywords: ADAS, advanced driver assistance system, global illumination, atmospheric scattering, transparency, deferred rendering, real time

1 Introduction

Camera based Advanced Driver Assistance Systems (ADAS) have gained increasing importance. Testing a new camera based ADAS during its development performing test drives with real cars is impractical, as they are cost- and time-intensive. Also the existence of a car prototype, which is produced late in the development process, is a precondition. An alternative for real-world test drives are tests in a three-dimensional virtual environment embedded in a HIL (Hardware-in-the-Loop) simulation. To ensure that the virtual tests create sufficient results as closely as possible to the results of real test drives, it is important that the rendering is as realistic as possible and real-time capable. A global illumination model for atmospheric scattering is one possibility to create realistic direct and indirect illumination. Global illumination describes a class of algorithms in computer graphics to create more realistic images by taking indirect illumination into account. By default, only local illumination is used and only direct illumination is computed. The interaction of light with particles in the air is not considered. Global illumination additionally takes this interaction into account and thus integrates indirect illumination. If light interacts with particles in the air, it is called atmospheric scattering. Because of the high recursion depth, the calculation of indirect illumination the direct way is very time-consuming.

¹ Paderborn University, CVB, Fürstenalle 11, 33102 Paderborn, sabrina.heppner@uni-paderborn.de

² Paderborn University, CVB, Fürstenalle 11, 33102 Paderborn, mariusdr@mail.uni-paderborn.de

³ Paderborn University, CVB, Fürstenalle 11, 33102 Paderborn, domik@uni-paderborn.de

To establish rendering at real-time frame rates, a deferred rendering pipeline⁴ can be used. Deferred rendering is an image based technique used in games and real-time applications gaining in popularity [La10] [Le11]. In contrast to forward rendering⁵, which is used by default, deferred rendering is a two-phase technique which decouples the rendering of geometry and lights into distinct passes: Geometry phase and illumination phase. During the geometry phase, data (e.g. surface normal, diffuse color, depth) needed for the illumination calculation is stored in the G-Buffer (geometry buffer). During the illumination phase, the data stored in the G-Buffer is read and combined with the information of the light to compute the final surface color. As this computation is done once per pixel, there is no overdraw possible, which is one major performance problem of forward rendering.

As with many other applications with three-dimensional graphics output, the test scenes used in software for the visualization of ADAS tests use transparent objects. Examples are windows in cars, buildings, noise barriers, bus stops or any other objects including parts made of glass. Another type of object uses textures with a dedicated alpha channel. It can be used to display objects like trees or fences. To keep the state of realism high, it is important that transparent objects like these are rendered correctly. Unfortunately, this is not feasible with a default deferred rendering pipeline described before. Therefore, an adequate adaption has to be found. This paper will first take a look at the physical background of atmospheric scattering and transparency. Afterwards, possibilities for the realization inside a rendering pipeline are considered and compared. In the end there will be a consistent concept to render outdoor scenes including atmospheric scattering and transparency based on deferred rendering. For readers who are interested in the physical background, section 2 gives a short overview. Those who like to get to know the concepts in depth are invited to read subsections 3.1 to 3.3 (atmospheric scattering) and subsection 4.1 (transparency). If the interest mainly applies to the results, skip to subsection 3.4 and 4.2.

2 Physical Background

We first consider the physical principles of atmospheric scattering and transparency.

2.1 Atmospheric Scattering

When light passes through the earth's atmosphere, it interacts with particles, namely molecules and aerosols. Three different types of interaction happen: Absorption, outscattering and inscattering. Absorption and outscattering cause a reduction of radiation. Absorption converts light into a different form of energy. During an outscattering event, some light rays are diverting into different directions than the initial traveling direction. In contrast to that, during an inscattering event additional light is redirected into the traveling direction which

⁴ A rendering pipeline is a logical model of computations needed for the process from 3D scene data to a 2D image in a raster-display system [LHL14].

⁵ Each object is rendered separately using material color, lights, etc. to determine the final color of the pixel associated with an object [Va14].

causes an increase of the radiation. Incoming light can either come directly from a light source or it is itself the result of scattering off other particles. The effect that light changes its direction several times due to scattering is called multiple scattering. In particular the scattering effect causes the indirect illumination in outdoor scenes.

The results of light interaction are visible effects in outdoor scenes. The most important effect is the indirect illumination. Other effects are the blue color of the sky, the coloring of the sky depending on the weather and location, the color gradient on the horizon, the coloring of dawn and twilight, the aerial perspective, the aureole and the “blue hour” [Ha96] [Ho07] [Ka91] [KRZ06] [LL01] [Ne14] [Mi12]. For the human perception these are familiar effects. Already Leonardo Da Vinci made use of them to give paintings a more realistic look [Ko12]. Not implementing these effects, e.g. by coloring the sky in an uniform blue, has the implication that a rendered scene receives an unrealistic appearance (best seen in our results shown in Figure 5). Consequently, by implementing these effects, a virtual scene becomes more realistic.

2.2 Transparency

An object can be described as transparent, translucent (partly transparent and partly opaque) or opaque (non-transparent) [Ho13]. These characteristics depend on how much light is absorbed and scattered while passing through the matter. When light emitted by a light source hits an object, some rays will be reflected and the rest will be refracted. While the refracted light is traveling through the object, it might be absorbed and scattered. If there is no scattering, the scene behind the object will be visible through the object. This effect is called transparency. In the case of water and glass very little light of the visible spectrum is absorbed. Therefore, they appear not just transparent, but very clear.

3 State of the Art: Rendering Global Illumination in Outdoor Scenes

In outdoor scenes atmospheric scattering has an important influence on illumination. Because of multiple scattering, the possible depth of recursion is much higher as if just surface reflection is considered. Hence, the standard rendering pipeline and traditional rendering methods, like ray tracing and radiosity, are not suitable to create a physical based rendering of outdoor illumination [Kl87] [Wa07]. In the following, different modern approaches are considered.

3.1 Previous Approaches

Much research has been done to create mathematical models for integrating atmospheric scattering into a virtual scene. The developed approaches can be divided into three groups: Simulation based methods, analytical models, and approximations. Simulation based methods make use of numerical solutions. Analytical models and approximations avoid numerical calculations with the help of analytical or rather approximated calculations. Figure 1

shows an overview of approaches including the chronological order and relations to each others.

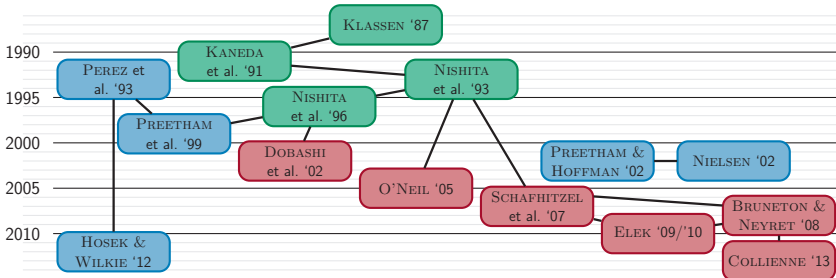


Fig. 1: Rendering of atmospheric scattering (green: Simulation based approaches, blue: Analytical approaches, red: Approximations)

Figure 1 shows that the most modern approaches make use of approximations. This way the calculation in real time can be realized. A comparison of the given approaches shows that the approach of BRUNETON & NEYRET [BN08] is the most innovative one fitting the given use case. It is an enhancement of the approach of SCHAFHITZEL et al. [SFE07] and the only approach implementing multiple scattering and shadow [BN08] [Co13] [Ko12]. Several other researchers use the approach of BRUNETON & NEYRET for their work.

3.2 Precalculated Lookup Tables

In the approach of BRUNETON & NEYRET, multiple scattering is implemented by an incremental calculation of the scattering values performed as precalculation on the GPU [EK10]. Scattering values are written into lookup tables stored as textures. For performing the precalculation, nine shaders, three 2D textures (one for intermediate results and two for final results), and four 3D textures (three for intermediate results and one for final results) are required. Figure 2 shows the complexity of the precalculation process and the relationship between different shaders and textures. In each step, exactly one scattering order is calculated using data of the previous scattering order [EK10].

During run time, the illumination of each pixel has to be calculated. It is composed by the directly incoming light, the reflected light, and the inscattered light. Equation 1 is used to perform this calculation:

$$L(x, \vec{v}, \vec{s}) = L_{direct}(x, \vec{v}, \vec{s}) + L_{reflected}(x, \vec{v}, \vec{s}) + L_{inscattered}(x, \vec{v}, \vec{s}) \quad (1)$$

x is the illuminated point, \vec{s} the direction to the sun and \vec{v} the view direction. How the single factors are composed is visualized in Figure 3. As marked in this figure, some components include integrals. Because of the huge calculation effort arising from these integrals, they are (completely or partially) calculated offline as a preprocess, stored in lookup tables called transmittance, inscatter, and radiance texture and just have to be read during run time. Some additional components are calculated online. For creating an illumination of high quality, the values have to be very precise. Offering precise values while performing

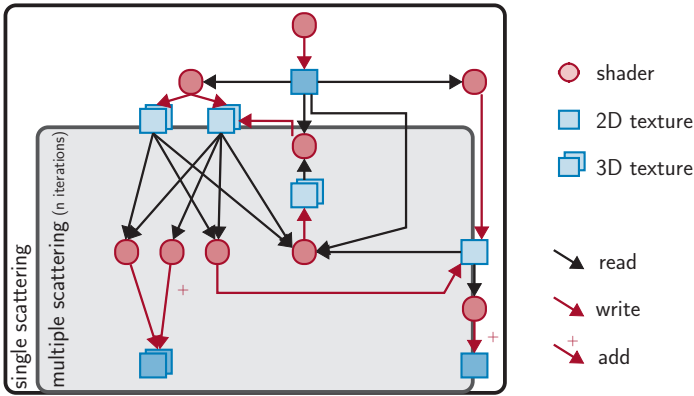


Fig. 2: Precalculation process

the calculation offline and storing the results in textures, would cause a high memory requirement. Therefore, these values are calculated at run time. There is no noticeable slowdown with this implementation. The real-time capability is still maintained.

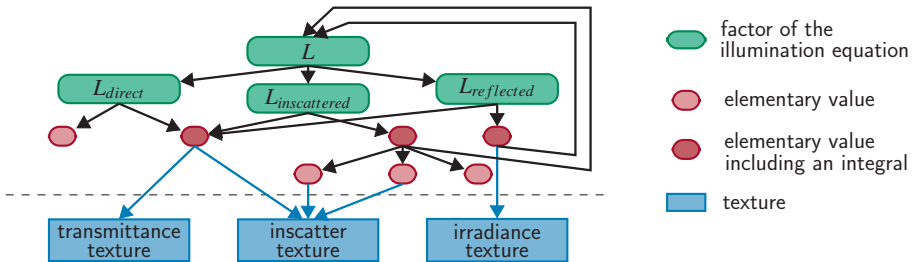


Fig. 3: Factors for the illumination calculation

Because of the consideration of the Mie theory and the Rayleigh theory in the precalculated values, all effects triggered by atmospheric scattering can be reproduced. Every time scattering parameters are changed, a rerun of the precalculation has to be performed [Ko12].

3.3 Deferred Rendering Pipeline

The approach of BRUNETON & NEYRET is adequate to be performed in a deferred rendering pipeline. Figure 4 shows this concept. In an offline phase, the precalculation is performed. As input for the precalculation parameters for the illumination are needed. The results of the calculation are stored in lookup tables to be used in the illumination phase. The execution of the online phase is divided into geometry and illumination phase. During the geometry phase, all data of the 3D scene needed for illumination calculations is stored in the G-Buffer. In the illumination phase the data from the G-Buffer is used in combination with the lookup textures to construct the illumination of the final scene. The illumination is

adapted immediately when the 3D scene is changed. Changes of the scattering parameter cause a rerun of the precalculation.

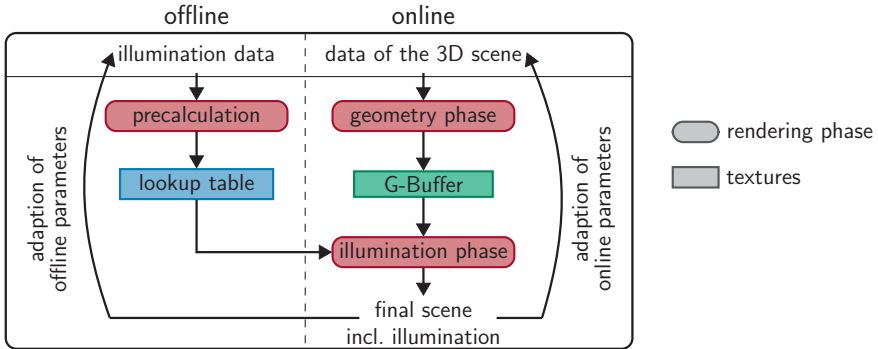


Fig. 4: Rendering of atmospheric scattering in a deferred rendering pipeline

3.4 Our Results of Deferred Rendering by BRUNETON & NEYRET

We implemented the previously described technique by BRUNETON & NEYRET for several traffic scenes and discussed our results (examples in Figure 5) with our domain experts.

3.4.1 Performance

Typically a camera has a frame rate of approximately 30 fps, but faster cameras exist. Therefore, developer of HIL simulators offer a frame rate of 60 fps. Thus, we define 30 fps as strong and 60 fps as weak real-time requirement. To evaluate the performance of the global illumination model for atmospheric scattering, two different types of hardware are used:

- Computer 1: Intel Core i7-3770, 3.40 GHz, nVidia Quadro 600, 8 GB RAM
- Computer 2: Intel Core i7-3770, 3.40 GHz, nVidia GeForce GTX 660 Ti, 8 GB RAM

The rendering time for 10 263 to 219 826 vertices on a standard PC (Computer 1) guarantees frame rates of more than 60 fps. Using a more powerful GPU (Computer 2) decreases rendering time to one fifth of the original rendering time. The rendering time excludes the one-time cost of the precalculation. The time of the precalculation strongly depends on the scattering order (recursion depth of multiple scattering) and on the hardware. For scattering order 1, less than one second⁶ is needed, for multiple scattering time increases⁷. In the case of the powerful GPU, the performance increases slightly by less than a half second. The weaker GPU records a stronger degradation of more than 10 seconds. A higher scattering

⁶ scattering order 1; Computer 1: $\sim 0.95s$, Computer 2: $\sim 0.19s$

⁷ scattering order 4; Computer 1: $\sim 13.30s$, Computer 2: $\sim 1.10s$

order than 4 is redundant because the adjustments are very small and not recognizable by the human eye. As expressed by our domain experts, all measured performance values for the precalculations are acceptable. Even the long measurements can be tolerated because of the infrequent execution of the precalculation. Most important is the run time itself.

3.4.2 Visual Quality

The visual quality is optimized while using a global illumination model for atmospheric scattering. The improvement can be comprehended best by the effects of the atmospheric scattering. Directly implemented by [BN08] are the following effects: Indirect illumination, blue color of the sky, color gradient on the horizon, coloring of dawn and twilight, aureole, aerial perspective. Furthermore, the effect “blue hour” is integrated by adding a blue coloring when the sun’s position is under a certain threshold value. The more realistic look of the scene is assessed as useful and pleasing and will be implemented in the product of our domain experts. Figure 5 contrasts a scene rendered without and including atmospheric scattering. It is significant that the integration of atmospheric scattering creates a much more realistic look.

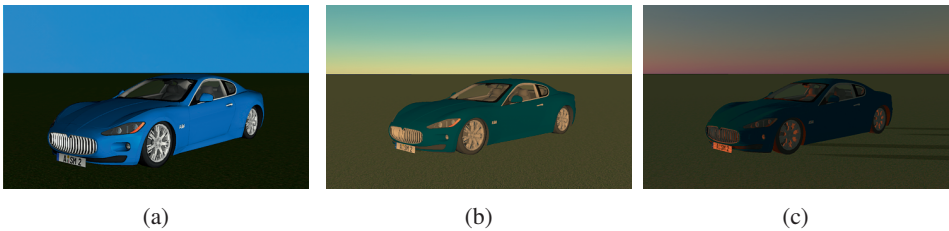


Fig. 5: Comparison: Rendering without (a) and including ((b) morning sun, (c) “blue hour”) atmospheric scattering

4 State of the Art: Rendering Transparency

In the typical situation of rendering transparency an opaque object is partially obscured by a transparent object. The light travels through the transparent object in a straight line (special cases like scattering inside the model are ignored). While rendering such a scene, the shading for both, opaque and transparent objects, are computed separately. The two colors are combined by a process called blending to produce a single color which, if done correctly, gives the visual impression that the transparent object is in front of the opaque object. The blending process can be described formally with the help of Equation 2 where A and B represent the color of a pixel associated with object A and object B and the alpha value of A (A_α) describes the opacity of A .

$$A \text{ OVER } B = A + B \cdot (1 - A_\alpha) \tag{2}$$

This underlying equation defining this process is commutative, meaning the order of the rendered objects is important. As an additional simplification, transparent objects which are

not captured by the camera but placed in front of the light source and thus influencing the color of the light are not considered.

To integrate transparency into a deferred rendering pipeline, the use of different approaches is conceivable. The techniques we will focus on are Forward Transparency, Weighted Blended Ordered-Independent Transparency, Inferred Lighting, and an A-Buffer implemented as Linked List on the GPU.

While integrating transparency into the atmospheric scattering concept, the precomputation remains unchanged. It is only important to define how transparent objects are rendered with the effect of atmospheric scattering. The illumination equation (see Equation 1) and the equation of standard alpha (see Equation 2) blending can be combined trivially. The colors of the pixels associated with object *A* and *B* are calculated by the illumination equation and combined with the help of alpha blending. Thus, the atmospheric scattering model can be extended intuitively to include transparent objects. There are no modifications needed to add neither the atmospheric scattering nor the blending operator.

4.1 Techniques

The straight forward technique is called Forward Transparency. Here the transparent objects are rendered in a separate forward rendering pass called transparent pass after the deferred rendering. During this transparent pass, the transparent objects are rendered as in a forward rendering pipeline. Each object is rendered separately using material color, lights, etc. to determine the final color of the pixel associated with an object [Va14]. Thus, the transparent pass has the same disadvantages as normal forward rendering for opaque objects, e.g. performance decrease with an increasing number of lights. Furthermore, Forward Transparency does not solve the sorting issues arising from ensuring correct blending of multiple transparent objects placed successive.

Forward rendering can be improved by a technique called Weighted Blended Order-Independent Transparency. This technique removes the need to sort transparent objects prior to rendering. The idea of blended Order-Independent Transparency in general is to redefine the blending order to be commutative which allows performing the blending process in arbitrary order. This technique was shaped by the following approaches: [Me07] [BM08] [MB13] [MB14]. In total five rendering passes are needed: Geometry pass, light pass, clear pass, transparent pass and composition pass.

Another approach, Inferred Lighting, was invented by KIRCHER & LAWRENCE [KL09]. It is a variation of deferred rendering to render transparent objects directly into the G-Buffer and separate them from opaque objects in a later pass. Thus, the main advantage of Inferred Lighting over deferred rendering is that transparent and opaque pixels can be lit in the same way. The rendering is separated into three passes: Geometry pass, light pass, and material pass. The lighting calculation can be performed at a lower resolution than the final output. This can improve the rendering performance significantly because the illumination equation has to be computed for fewer pixels and the smaller buffer requires less memory bandwidth.

The A-Buffer (anti-aliased, area-average, accumulation buffer) was first introduced by CARPENTER [Ca84] as a general hidden surface mechanism and an improvement over the Z-Buffer. Among other advantages, it allows rendering of order-independent transparency. During rendering, for each pixel either a single fragment or a pointer to a list of fragments is saved. Hardware accelerated implementation of an A-Buffer was proposed by YANG et al. [Ya10] and LEFEBVRE et al. [LHL14]. They create and maintain linked lists of fragments for each pixel entirely on the GPU. Previous A-Buffer implementations are designed without considering a deferred rendering pipeline. SCHOLLMAYER et al. [SBF15] describe a full deferred rendering pipeline which integrates an A-Buffer implementation using linked lists. A-Buffer implemented as Linked List approach consists of three rendering passes: Geometry pass, light-culling pass and shade-composing pass.

There exist various further techniques which are either too slow to be used in real-time rendering or exhibit other properties which makes them unsuitable for the given domain [Mal1].

4.2 Our Results of Three Transparency Techniques

To compare results of transparency in a deferred rendering pipeline including atmospheric scattering, we implemented three techniques⁸: Weighted Blended Order-Independent Transparency (WBOIT), Inferred Lights (IL), and A-Buffer/Linked Lists (LL). As basis for evaluation we chose performance and visual quality.

4.2.1 Two Different ADAS Scenes

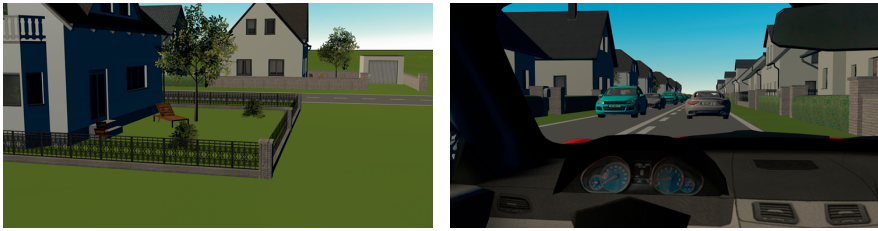
For evaluation, two suitable ADAS scenes of different complexity are created. The setting of the first scene is a small village with several houses, cars, streets, and a forest with 400 randomly placed and oriented trees. There are several surfaces using the alpha channel, e.g. windows, foliage, and fences (see Figure 6a). The second scene is more complex. It is used to evaluate the implementations for a traffic scene. It contains a road with numerous cars surrounded by houses on both sides (see Figure 6b). All tests are conducted with a resolution of 1920 x 1080 pixels and all statements in the following refer to this resolution.

4.2.2 Performance

For the performance test, two different types of hardware are used:

- Computer 1: Intel Core i7 860, nVidia Quadro K2000, 4GM RAM (slow, less memory)
- Computer 2: Intel Core i5 4750, nVidia GTX 660, 16GB RAM (fast, more memory)

⁸ Forward rendering is not considered because the performance is much lower than the performance of the other techniques.



(a) Test scene 1: Small village with forest (b) Test scene 2: Traffic scene

Fig. 6: Test scenes

The performance is recorded from 13 different views of the first scene and 5 different views of the second scene. These views are used to test a variation of different conditions. In the case of scene 1, depending on the view, 25 to 2130 objects, 4436 to 92 833 triangles, and 10 263 to 219 826 vertices are given. In the second scene there are significantly more triangles and vertices to be rendered. On average there are ~600 objects, ~350 000 triangles, and ~420 000 vertices visible. For example, the view in Figure 6b, a close-up view of the windshield from inside of a car, includes a large amount of transparent fragments (~1.5 million).

Because precomputation remains unchanged, its performance is unaffected. During run time, the light distribution function has to be evaluated at every transparent fragment. The most expensive operation is the adding of inscattered light as it involves many lookups in a 3D texture.

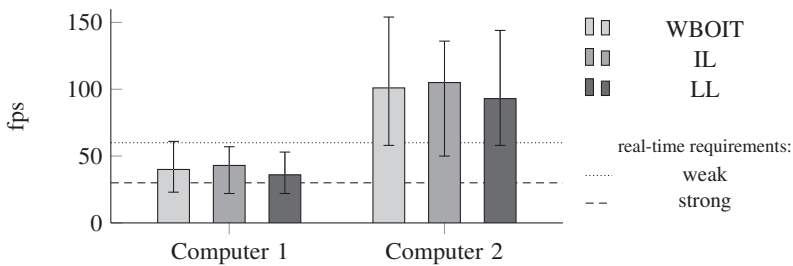


Fig. 7: Average and minimal/maximal performance

Figure 7 shows the results of the performance test. Each bar visualizes the average performance of the implementation of one technique on one computer. Furthermore, minimal and maximal performance occurring during the test are marked. The dashed line and the dotted line visualize the strong (30 fps) and the weak (60 fps) real-time requirements⁹. It is significant that all three techniques offer similar results. The weaker computer (Computer 1) reaches the weak real-time requirement. In the case of the more powerful computer (Computer 2) even the worst measured performance values exceed the number of 60 fps and thus the hard real-time requirement is fulfilled.

⁹ see section 3.4.1

4.2.3 Visual Quality

On visual quality, we again rely on our domain experts for a qualitative assessment. The domain experts observed scenes including transparent objects rendered with different methods. They identified features that differ for the given methods. Table 1 shows screenshots used for comparison. In addition to two ADAS scenes we include a teapot scene demonstrating rendering mistakes arising from a high number of completely transparent overlapping objects not apparent in the ADAS scenes.

WBOIT produces images with varying quality. For surfaces with low alpha values, including objects rendered with alpha mask textures like leaves (leaves (a)) or fences, the results are of desired quality indistinguishable from Linked Lists (leaves (c)). In the case of transparent surfaces with high alpha value or very saturated colors, rendering mistakes occur. Such an effect is recognizable in the very bright specular highlights on the windshield (windshield (a)). Furthermore, the screenshot of the teapots (teapots (a)) reveals that there is a problem concerning the perceived ordering. The colors of teapots in the background dominate too much.

Inferred Lighting shows the worst overall image quality of all three tested techniques. This low quality is caused by the low resolution rendering and subsequent upscaling of transparent surfaces. This is most notably seen at the border of transparent surfaces like the car window (windshield (b)). Lighting fails completely for cases where several surfaces with the same stipple pattern (see [KL09]) are rendered on top of each other (leaves (b)). This also applies for transparent objects overlapping themselves as seen in the case of the teapots (teapots (b)).



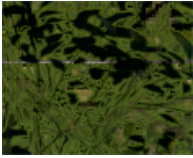

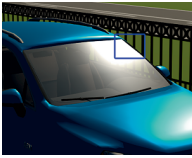



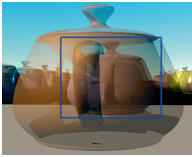



Linked Lists shows desirable image quality while rendering transparent surfaces due to the per-pixel sorting of transparent fragments (leaves, windshield, teapots (c)). The results should be identical to an offline rendering with an A-Buffer implementation. This only applies as long as there is enough video memory available to store all rendered fragments. If memory overflows, fragments have to be discarded, which leads to visual artifacts. Using older GPUs¹⁰, as the ones used for our evaluation, a flickering effect, not visible on still images, occurs. These older GPUs do not support atomic operations needed for the ordering of fragments with similar depth. Thus, the ordering can vary and this leads to a flickering effect.

Due to the multiple rendering mistakes, WBOIT and IL are rejected by our domain experts. In case of suitable graphics cards, LL is acceptable.

5 Summery & Conclusion

Our solution employs a deferred rendering pipeline which allows atmospheric scattering to be added without noticeable slowdown to the real-time requirements while at the same time

¹⁰ E.g. the `atomicMax` operation for 64-bit values is supported by nVidia GPUs starting with the Maxwell generation, e.g. nVidia GeForce GTX 970.

full view on scene	(a) WBOIT	(b) Inferred Lighting	(c) Linked List
leaves 	 ✓	 ✗	 ✓✓
windshield 	 ✗	 ✗	 ✓✓
teapots 	 ✗	 ✗	 ✓✓

Tab. 1: Comparison of visual quality (✓✓: good, ✓: acceptable, ✗: bad)

providing a pleasing natural effect. Regarding the implementation of transparent objects (e.g. windshield) and realistic rendering of fences and foliage (textures with dedicated alpha channel), we have found Weighted Blended Order-Independent, Inferred Lighting and Linked List to be techniques all satisfying the real-time requirements, while only Linked List reaches the visual realism requirements of our domain experts. However, best visual quality is attained at the cost of sufficient video memory and a modern GPU supporting atomic operations.

Acknowledgment

The models placed in the test scenes are provided and copyrighted by dSPACE GmbH.

References

- [BM08] Bavoil, Louis; Myers, Kevin: Order Independent Transparency with Dual Depth Peeling. NVIDIA OpenGL SDK, pp. 1–12, 2008.
- [BN08] Bruneton, Eric; Neyret, Fabrice: Precomputed Atmospheric Scattering. Computer Graphics Forum, 27(4):1079–1086, 2008.

- [Ca84] Carpenter, Loren: The A-buffer, an antialiased hidden surface method. *ACM Siggraph Computer Graphics*, 18(3):103–108, 1984.
- [Co13] Collienne, Peter; Wolff, Robin; Gerdt, Andreas; Kuhlen, Torsten: Physically Based Rendering of the Martian Atmosphere. In (Latoschik, Marc Erich, ed.): *Virtuelle und Erweiterte Realität – 10. Workshop der GI-Fachgruppe VR/AR*. Shaker, Aachen, pp. 97–108, 2013.
- [EK10] Elek, Oskar; Kmoch, Petr: Real-Time Spectral Scattering in Large-Scale Natural Participating Media. In (Hauser, Helwig; Klein, Reinhard, eds): *Proceedings of the 26th Spring Conference on Computer Graphics*. ACM, New York, pp. 77–84, 2010.
- [Ha96] Haltrin, Vladimir I.: Haltrin. A Real-Time Algorithm for Atmospheric Corrections of Airborne Remote Optical Measurements above the Ocean. In: *Proceedings of the Second International Airborne Remote Sensing Conference and Exhibition*. volume 3, pp. 63–72, 1996.
- [Ho07] Hoeppe, Götz: *Why the Sky is Blue – Discovering the Color of Life*. Princeton University Press, Princeton, 2007.
- [Ho13] Hoffman, Naty: *Background: Physics and Math of Shading*. SIGGRAPH, 2013.
- [Ka91] Kaneda, Kazufumi; Okamoto, Takashi; Nakamae, Eihachiro; Nishita, Tomoyuki: Photorealistic Image Synthesis for Outdoor Scenery under Various Atmospheric Conditions. *The Visual Computer*, 7(5-6):247–258, 1991.
- [Kl87] Klassen, R. Victor: Modeling the Effect of the Atmosphere on Light. *ACM Transactions on Graphics*, 6(3):215–237, 1987.
- [KL09] Kircher, Scott; Lawrance, Alan: Inferred Lighting: Fast Dynamic Lighting and Shadows for Opaque and Translucent Objects. In: *Proceedings of the 2009 ACM SIGGRAPH Symposium on Video Games*. Sandbox '09, ACM, New York, NY, USA, pp. 39–45, 2009.
- [Ko12] Kol, Timothy R.: *Analytical Sky Simulation – An Implementation and Analysis of Daytime Skylight Models*. Technical report, University, 2012.
- [KRZ06] Kment, Thomas; Rauter, Michael; Zotti, Georg: *Modelling of Daylight for Computer Graphics*. Technical report, Institut für Computergraphik und Algorithmen, Wien, 2006.
- [La10] Lake, A.: *Game Programming Gems 8*. IT Pro. Course Technology, 2010.
- [Le11] Lengyel, E.: *Game Engine Gems 2*. Taylor & Francis, 2011.
- [LHL14] Lefebvre, Sylvain; Hornus, Samuel; Lasram, Anass: Per-Pixel Lists for Single Pass A-Buffer. *GPU Pro 5: Advanced Rendering Techniques*, 2014.
- [LL01] Lynch, David K.; Livingston, William Charles: *Color and Light in Nature*. Cambridge University Press, Cambridge, 2001.
- [Ma11] Maule, Marilena; Comba, Joo L.D.; Torchelsen, Rafael P.; Bastos, Rui: A Survey of Raster-Based Transparency Techniques. *Computers & Graphics*, 35(6), 2011.
- [MB13] McGuire, Morgan; Bavoil, Louis: Weighted Blended Order-Independent Transparency. *Journal of Computer Graphics Techniques (JCGT)*, 2(2):122–141, 2013.
- [MB14] McGuire, Morgan; Bavoil, Louis: Weighted, Blended Order Independent Transparency. *ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, 2014.
- [Me07] Meshkin, Houman: Sort-independent alpha blending. *GDC Talk*, 2007.

- [Mi12] Minnaert, Marcel: *Light and Color in the Outdoors*. Springer, New York, 2012.
- [Ne14] Nentwig, Mirko: *Untersuchungen zur Anwendung von computergenerierten Kamerabildern für die Entwicklung und den Test von Fahrerassistenzsystemen*. PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2014.
- [SBF15] Schollmeyer, Andre; Babanin, Andrey; Froehlich, Bernd: Order-Independent Transparency for Programmable Deferred Shading Pipelines. In: *Computer Graphics Forum*. volume 34. Wiley Online Library, pp. 67–76, 2015.
- [SFE07] Schafhitzel, Tobias; Falk, Martin; Ertl, Thomas: Real-Time Rendering of Planets with Atmospheres. In (Rossignac, Jarek; Skala, Václav, eds): *Journal of WSCG*. University of West Bohemia, Plzen, pp. 91–98, 2007.
- [Va14] Varcholik, P.: *Real-Time 3D Rendering with DirectX and HLSL: A Practical Guide to Graphics Programming*. Game Design. Pearson Education, 2014.
- [Wa07] Wang, Changbo: Real-Time Rendering of Daylight Sky Scene for Virtual Environment. In (Ma, Lizhuang; Rauterberg, Matthias; Nakatsu, Ryohei, eds): *Entertainment Computing – ICEC 2007*, volume 4740, pp. 294–303. Springer, Berlin & Heidelberg, 2007.
- [Ya10] Yang, Jason C.; Hensley, Justin; Grün, Holger; Thibieroz, Nicolas: Real-time Concurrent Linked List Construction on the GPU. In: *Proceedings of the 21st Eurographics Conference on Rendering*. Eurographics Association, Aire-la-Ville, Switzerland, pp. 1297–1304, 2010.

Additive Fertigung/3D Druck – Technologie, Auswirkungen und Chancen

Additive Fertigung/3D Druck – Technologie, Auswirkungen und Chancen

Felix Baumann¹, Dieter Roller²

Vorwort

Der europäische Maschinen- und Anlagenbau steht im internationalen Wettbewerb (vor allem im Low-Cost-Segment) der Marktübernahme durch chinesische Hersteller entgegen. Für die mittleren und oberen Qualitäts- und Preissegmente bieten komplexe Fertigungsverfahren den hiesigen Produzenten gute Chancen, um ihren Kunden durch zusätzliche Dienstleistungen einen wettbewerbsrelevanten Mehrwert zu bieten.

Vor dem Hintergrund des durch den Kunden implizit aufgebauten Drucks durch steigende Variantenvielfalt, kleinere Losgrößen und erhöhte Produktkomplexitäten wird die Steuerung und -überwachung solcher Fertigungssysteme allerdings zunehmend unübersichtlich und bedarf systematischer Unterstützung durch Computersysteme. Eine Vielzahl von Prozessparametern unterschiedlicher integrierter Produktionsressourcen steht in unmittelbarer Wechselwirkung mit den späteren Eigenschaften und der Beschaffenheit des Fertigungsergebnisses. Das gilt in besonderem Maße für moderne additive Fertigungsverfahren, da hier die mechanischen Funktionen einer Anlage und unterschiedliche Medien und Werkzeuge im Fertigungsprozess zusammengeführt werden.

Durch die weite Verbreitung der Additiven Fertigung erscheint es aktuell möglich, die Produktion, deren Arbeitsplätze in den letzten Jahren vollständig oder teilweise ins Außer-Europäische Ausland ausgegliedert wurde, wieder nach Europa zurück zu tragen. Dies erfolgt durch den zunehmenden Drang der Konsumenten ihre Produkte individualisieren zu können. Es handelt sich hierbei nicht um eine Massenfertigung, sondern um eine Massen-Individualisierung. Die Diskussionsgrundlage hierfür ist gelegt, doch bleibt es offen, ob sich dieses Versprechen auch einhalten lässt.

Weiterhin können durch die Additive Fertigung einige Vorteile in der Produktentwicklung, der Prototypenfertigung und der Kleinserienfertigung („Losgröße 1“) wie beispielsweise kurze Entwicklungszyklen, traditionell nicht oder nur sehr schwer erreichbare Objektgeometrien ("Design for function"), Ressourceneinsparung und flexibler Maschinen- und Werkzeugeinsatz realisiert werden. Diese Technik erlaubt es

¹ Universität Stuttgart, Institut für Rechnergestützte Ingenieursysteme (IRIS), baumann@informatik.uni-stuttgart.de

² Universität Stuttgart, Institut für Rechnergestützte Ingenieursysteme (IRIS), roller@informatik.uni-stuttgart.de

relativ unmittelbar Werkstücke mittels digitaler Modelle in physische Modelle umzuwandeln und erlaubt damit von Menschen entworfene Modelle für Menschen zu produzieren, wobei sich die Produktentwicklung aufgrund physikalischer Eigenschaften des Herstellungsprozesses von der traditionellen Produktentwicklung unterscheidet.

Die räumliche Verflechtung dieser Technologie bringt in ihrer Ursprungsform eine gewisse Lokalität mit sich, die durch sogenannte Cloud Services entzerrt und zu einem global nutzbaren Konstrukt für Menschen und Unternehmen gemacht werden. Dadurch ergibt sich der Bedarf, die Diskussion auch über die Verflechtung verschiedener Fertigungsverfahren und Anbieter sowie deren softwareseitige und architekturelle Unterstützung zu führen.

Viele europäische Unternehmen sind durch Selbstverpflichtungen wie ISO 9000ff oder durch gesetzliche Haftung für ihre Produkte und Prozesse gezwungen, ihre Produktionsprozesse selbst zu kontrollieren, zu dokumentieren und zu beherrschen. Jedoch stellen die Komplexität der neuartigen Fertigungssysteme, der rasche technische Fortschritt sowie die enge Verknüpfung von Hard- und Software in dem Bereich der industriellen Produktion die Mitarbeiter solcher komplexen Fertigungstechnologien vor große Herausforderungen bei deren Handhabung. Neben neuartigen Verfahren, Techniken und Konzepten additiver Fertigungstechnologien, sind daher auch neue Benutzerschnittstellen und Unterstützungswerkzeuge erforderlich, welche die Anwender in die Lage versetzen mit der technologischen Entwicklung Schritt zu halten um effektiv sowie effizient zu arbeiten. Eine besondere Bedeutung kommt angesichts der Komplexität und der sozio-technischen Ausrichtung solcher Systeme Fragestellungen der Mensch-Maschine Interaktion zu. Dieser Workshop ist ein Forum die verschiedenartigen Aspekte dieser Fertigungstechnologie vorzustellen und zu diskutieren. Die Teilnehmer tragen, durch ihre Expertisen in den verschiedenen Fachbereichen, dazu bei, die vielfältigen Aspekte möglichst breit und aus verschiedenen Blickwinkeln zu diskutieren.

Aus den Einsendungen zu diesem Workshop wurden die folgenden, kurz dargestellten, Beiträge ausgewählt: Der Beitrag "Universal API for 3D Printers" behandelt den Aspekt der Ansteuerung und Integration verschiedenartiger Soft- und Hardware im Bereich der Additiven Fertigung in ein System, welches über das Internet im Sinne eine Cloud ansteuerbar ist. In "Bridging Gaps in Cloud Manufacturing with 3D Printing" behandeln die Autoren die Möglichkeit Produktionsfehler oder Lieferverzögerungen in der Produktionskette durch den Einsatz von Additiver Fertigung auszugleichen und bieten ein Konzept der Cloud basierten Fertigung mit Integration bestehender Anlagen und Services. "Supporting the Set-up Processes by Cyber Elements based on the Example of Tube Bending" erläutert anhand des Beispiels der Rohrbiegung die Komplexität des Rüstprozesses und zeigt Möglichkeiten der Unterstützung anhand von Cyber-physikalischen Systemen auf. Die Autoren des Beitrags "The custom-made body – Legal aspects of bioprinted tissue and organs" erläutern die juristischen Voraussetzungen von künstlich hergestellten Organen und Geweben. Im Beitrag "Infrastructuring, globale Innovation & digitale Fabrikation" wird ein Einblick in digitale Fabrikation aus Sicht

von Infrastructuring als holistische Forschungs- und Praxislinse aus der Computerunterstützten Gruppenarbeit gegeben.

Universal API for 3D Printers

Felix Baumann,¹ Oliver Kopp² and Dieter Roller³

Abstract: With this research we propose the implementation of an overlay restful API for 3D printers to expose these machines to the Internet for utilization within cloud services. This is to abstract the underlying communication structure and means for accessing and controlling a 3D printer resource which is performed in one of three ways. The first method of accessing and controlling a 3D printer is via a proprietary protocol or a printer driver in Microsoft Windows. The second method is the control via a USB-serial connection between a controlling computer and the printer resource. This protocol can either be proprietary or based on open standards like GCODE (ISO 6983-1:2009). The third method of control is based on physical storage devices attached to the printer with machining instructions stored on them. This research excludes the communication and control means involving proprietary protocols or drivers due complexity restrictions within the implementation. The approach is designed with extensibility in mind so that future access to proprietary protocols can be added to the control API. Printer resources with only the third control method available are also excluded from this research as they are currently lacking the capability to be remotely controlled. This work describes the design and implementation of an abstraction API layer between varying soft- and hardware components with an extensible architecture for future hard- and software components for within the domain of Additive Manufacturing (AM).

Keywords: 3D Printing, Additive Manufacturing, Cloud Based Service, Cloud Based Manufacturing, CBM, CPS, Abstraction Layer, API

1 Introduction

Additive Manufacturing (AM) denotes a method to fabricate physical objects from digital models by means of layer-wise deposition or curing of a material [KMD00, GL13]. It originates from the technology of Rapid Prototyping (RP) but is no longer restricted to just the creation of prototypes. Its strengths are the creation of objects without specialized tools other than a 3D printer, the possibility to produce objects not restricted by the machining tool geometry, i. e., to allow for design for function, no tool creation beforehand necessary and reduced material waste as the process is additive and not subtractive resulting in fewer material necessary. On the downside the price per object created is much higher than with mass produced items, i. e., it is primarily suitable for low volume series, the quality of additively manufactured objects may vary and be lower than with mass-produced or

¹ University of Stuttgart, Institute of Computer-aided Product Development Systems, Universitätsstr. 38, 70569, Stuttgart, baumann@informatik.uni-stuttgart.de

² University of Stuttgart, Institute for Parallel and Distributed Systems, Universitätsstr. 38, 70569, Stuttgart, kopp@informatik.uni-stuttgart.de

³ University of Stuttgart, Institute of Computer-aided Product Development Systems, Universitätsstr. 38, 70569, Stuttgart, roller@informatik.uni-stuttgart.de

custom machined items, depending on the used technology and the embeddability within the manufacturing process is not sufficiently researched and adopted in comparison to traditional machining methods. AM surmises different technologies such as Fused Deposition Modeling (FDM, or also Fused Filament Fabrication), (Selective) Laser Melting or Sintering (SLM/SLS), Electron Beam Melting (EBM), 3D Printing (3DP), Laminated Object Manufacturing (LOM), or Stereolithography (SLA). Classification schemes are proposed by ASTM F42 working group and ISO/ASTM 52900:2015 but without consensus [GRS15]. The ASTM classification schema distinguishes the technologies by means of material processed. A common technology for consumer grade 3D printers is FDM/FLM where semi-molten thermoplastics are extruded through a nozzle mounted on a gantry system as a print-head movable in two dimensions (X-Y plane). The thermoplastics commonly used are Acrylonitrile Butadiene Styrene (ABC) or Polylactic Acid (PLA). This material is most often supplied as a uni-color filament on spools. The material can also be supplied as granulate but no commercially available systems for consumers exist. The material is pushed into the print-head by metal gears driven by an electro-motor and are heated in the print head to a temperature of approximately 230-245 Degrees Celsius for ABS which is over the materials glass-transition temperature [Bo14]. Pushing additional filament into the print-head extrudes material through the nozzle onto the print-bed or an existing layer of extrudate. After the completion of one layer either the print-bed lowers or the print head moves upwards by a specified distance. The accuracy and precision of the involved electro-motors and gears determine the objects quality and printing duration.

Fig. 1 presents a process for creating an object from a digital model. Following 5 steps are performed:

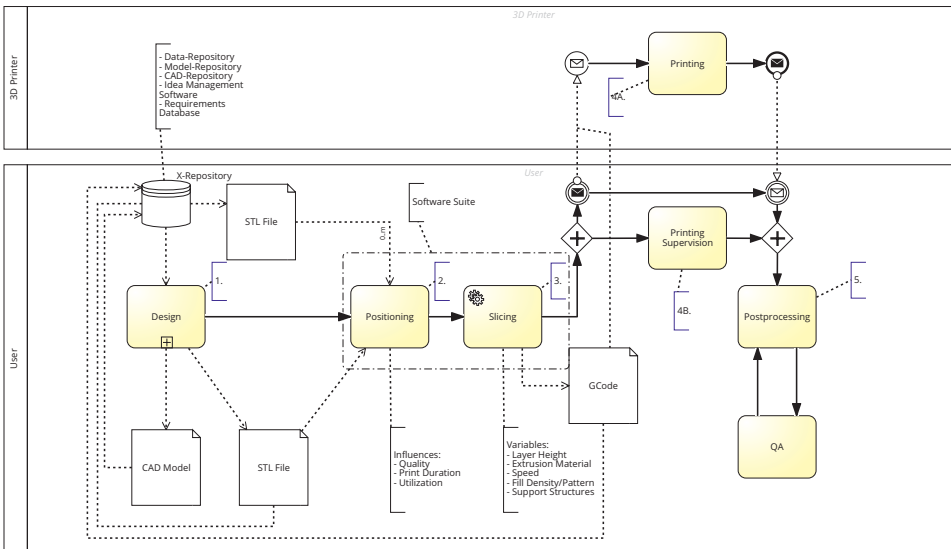


Fig. 1: 3D Printing Process

The process starts with the design of the product [KMD00], which can differ from the design of a subtractive or mass-produced object.

Step two of this process is the positioning of the model in the virtual space that represents the 3D printer and its physical restrictions. Positioning can encompass single objects or multiple objects for increased printer utilization.

After the print object is positioned it is sliced using slicer software. A variety of slicing software exists and they differ in aspects like speed, precision, quality and strategies for printing support structures. The following steps include the upload to the printer if it is a networked device or other means like deployment on memory devices (e. g., SD-Card, USB Stick) and the start of the print which can either require manual interaction or be handled from software. During printing the user is often required to supervise the printing progress as this is error prone especially for consumer grade devices [GS14].

Post-processing and Quality Assurance (QA) follow when the object has been printed and influence each other. These steps involve different hardware (e. g., different printer models) and different software (e. g., proprietary or standardized protocols to communicate with or control of the printer, proprietary or open-source software to transform a digital model from a specific input type to a specific output type or proprietary or open-source software to position and slice the digital model for fabrication). With this research we propose a unifying API that abstracts the controls of the involved soft- and hardware thus enabling a user to communicate with a variety of 3D printers in an uniform way. The proposed API also enables the use of the AM resources as a service. Additive Manufacturing machinery differs in its capabilities regarding material procession and means of communication with and controlling of the machine. A commonly utilized instruction set of 3D printers is GCODE (ISO 6983-1:2009) which is a text and line based machine code with machine instructions identified by a single character code symbol and an opcode. Instrumenting the machine with a pre-computed toolpath in this format differs widely across different printers [ALI12]. The machine code file can either be uploaded to an internal storage device via a protocol over cable (e. g., USB serial communication), manually transferred to a removable storage device, or transferred to the machine ad-hoc through a network or direct-link connection using an open-source or standardized protocol or proprietary means like a device driver. Figure 2 presents an overview over the different methods of transferring information or machine code to a 3D printer. This work is an extension to [BR16] and presents a universal API for 3D printers.

The remainder of this paper is organized as follows: We display current research in this area in Sect. 2 and derive implementation requirements from established approaches. The functionality of the universal API is described in Sect. 3. It is followed by a short discussion of how to add new printers in Sect. 4. An example usage of the API is outlined in Sect. 5. Section 6 presents the implementation and the usage of the API an a real world setting. Then we introduce an example and discuss problems encountered with the implementation. Finally, in Sect. 7, we conclude the paper by discussing our approach, its application, and benefits.

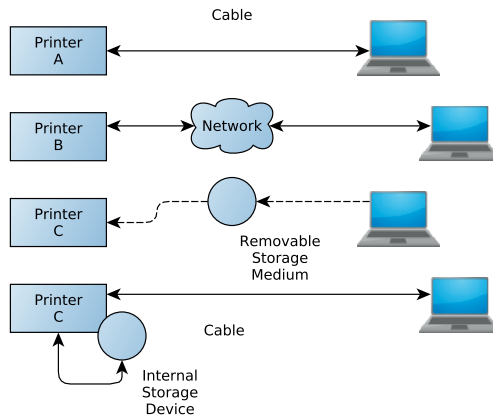


Fig. 2: Communication Variety to Transfer Data to 3D Printers

2 Related Work

Similar systems or services already exist in form of closed source commercial services. Examples are 3D Hubs³ and 3D Printer OS⁴. Being commercial entities their focus is on financial viability. These services allow adding ones own 3D printer and manage it from within the service with a varying degree of granularity. They lack an extension mechanism or plug-in architecture. In contrast to our approach, they are not intended as open services.

The software octoprint⁵ offers remote printing and object management capabilities but does not provide an interface to a business process management systems, user-selectable slicing solutions or support for consolidated information on printing information. Despite these shortcomings we regard this software as a viable possibility for connecting printing hardware to our proposed service.

Further research provides proposals by Wu et al. [Wu15] for CBM systems. Our system differs from those approaches as our focus is the tight integration of business process management (BPM) and 3D printing as well as the sensory upgrade of this technology. From Dong et al. [Do08] we will implement the video supervision approach for the printing process and its remote error detection. Extensions of CBM in the form of Cloud Based Design and Manufacturing [Wu12] provide further insight into the concept of Hardware-as-a-Service (HaaS) and the connection to the broader concept of flexible manufacturing spanning every phase of product development and involvement of different stakeholders. While the availability of affordable consumer grade 3D printers certainly has helped the progression of research in and distribution of 3D printers the scenario where every individual will own a digital fabricator [Mo11] is debatable as the general direction is to offer and consume services [Ba11]. Van Moergestel et al. [Mo15] proved the concept of

³ <http://www.3dhubs.com>

⁴ <http://www.3dprinter.os.com>

⁵ <http://octoprint.org>

Manufacturing-as-a-Service (MaaS) on cheap, distributed and reconfigurable production machines (equiplets) with a focus on interaction in a multi-agent system.

Kubler et al. [Ku15] provide a theoretical background on the relationship between the concepts of Internet of Things (IoT) and Cloud Manufacturing. With this work they connect concepts from cloud service to the manufacturing domain. Vukovic [Vu15] researches on the role and importance of APIs within the IoT paradigm and with that on cloud manufacturing. This work gives justification to our research on the definition of a API for 3D printing service.

Our work is similar to the work by Harrer et al. [Ha14], but they target the management layer of BPEL engines, whereas we target the management layer of 3D printers.

3 The Universal API for 3D Printers

The goal of the API is to abstract a variety of required software and different hardware in a uniform way in order to be consumed by other services (See Fig. 3). An exemplary list of software and input parameters for it is listed below:

- Positioning Software – Software that positions and orients a model representation within the build envelope of the 3D printer. The positioning and orientation can influence the quality of the printed object, the utilization the printer (e. g., multi object print versus single object print) and the material required for a print (e. g., scaffolding material). Exemplary list of parameters required for abstraction:
 - Position of the object within the build envelope
 - Dimensions of the build envelope
 - Information on possible other objects to be printed concurrently
- Slicing Software – Software that transforms a model representation, e. g., a STL or AMF file, into machine code (e. g., GCODE) to be consumed by the printer. The transformation is performed by a slice-wise segmentation of the original model depending upon a list of user selected parameters e. g., slice height or machine speeds. Exemplary list of parameters required for abstraction:
 - Slice height
 - Machine speed
 - Requirements for scaffolding
 - Slicing strategies
- Control Software – Software that transfers pre-programmed machine code to the printer or instructs the printer to start, stop or pause a print. Furthermore, maintenance operations can be executed through the control software.

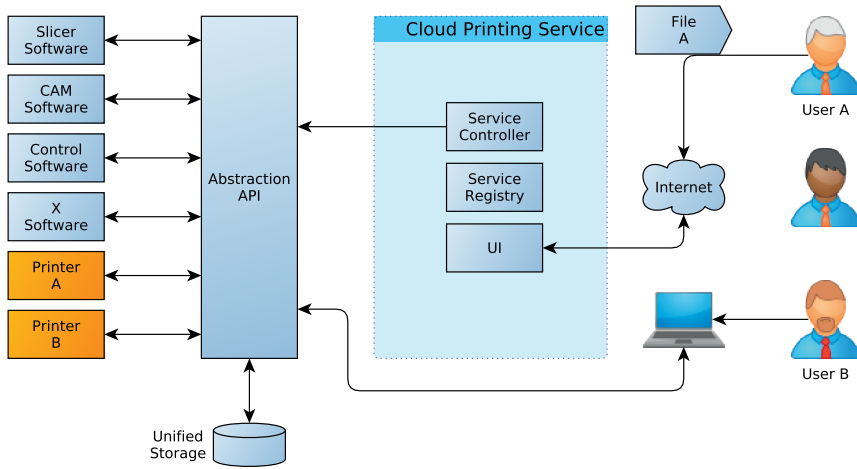


Fig. 3: Overview of the Implemented API

For the API we identified the following six object groups as sufficient for an proof of concept implementation.

- User
- File
- Printer
- Software
- Model
- Print

With the user object operations on users can be executed, which is necessary for accountability. Operations on user objects include:

- Create
- Update
- Delete
- Display information on user
- Associate user with print
- Associate user with printer (e. g., owner or administrator)

With the file object operations on uploaded and stored files can be performed:

- Upload
- Delete
- Update
- Display information on file
- Transform from file format A to file format B
- Associate file with printer (e. g., configuration or description file)
- Associate file with model (e. g., model is described by file)

With the printer object operations on a printer can be performed. Associated operations include:

- Install printer within the system/service
- Mark as available / unavailable
- Remove
- Display information on printer
- Associate printer with print
- Start / Stop / Pause print
- Rate printer quality / reliability / availability or other defined attributes

With the software object operations on software can be performed. The software is installed within the service by an administrator and mappings to the abstracted parameters are defined by the administrator. Usage restrictions for users or file types is defined and persistently stored within the service. In the concept of a 3D printing service the software execution is performed by a scheduler (Fig. 4) that dynamically allocates computing resources necessary for the completion of software tasks:

- Install software within the system / service
- Mark as deprecated
- Remove
- Display information on software
- Execute software

Figure 4 presents an overview of the intended architecture with a business process management system (BPMS) supporting the main service controller.

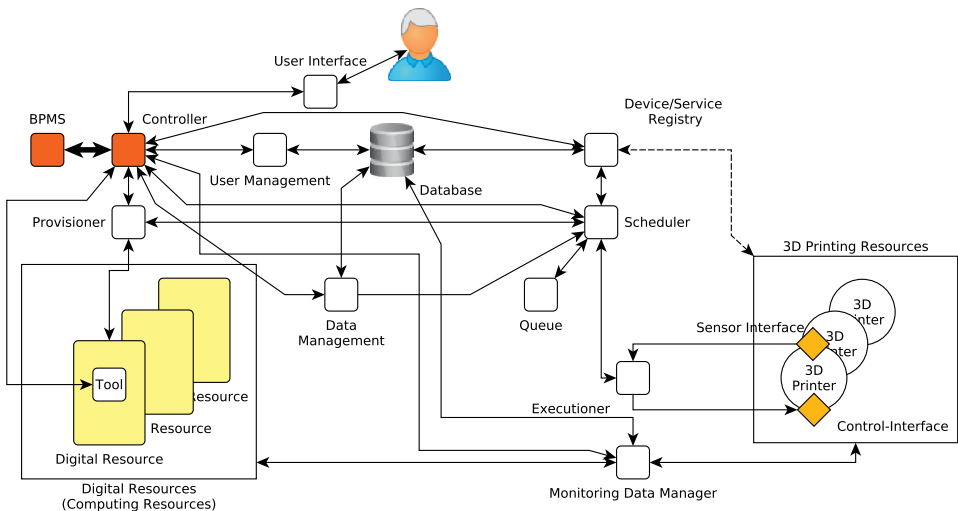


Fig. 4: Architecture for 3D Printing Service

4 Adding Support for new 3D Printers

For integration of various 3D printer types it is necessary to describe their capabilities in an abstract form for re-use. To our knowledge such a description format or language does currently not exist. Resource Description Language (RDL [Sa12]) is a proposition for this issue for the domain of network embedded resources. Capabilities required for interaction with tools includes a) GCode dialect b) Quality settings c) Processing speed and d) Material capabilities. This information is also required for utilization planning and optimization strategies. As a solution for this problem we propose a derivative RDL tailored towards additive manufacturing for subsequent publication. Further problems arise from the firmware of our research printer that limits the transmission speed (ca. 3.5 KiB/s) over the USB connection to the device storage resulting in long transmission times. Solutions include flashing a different firmware and utilizing WiFi enabled SD cards.

5 Example Usage

To clarify the flow of information (Fig. 4) and data within our proposed service we discuss this by an example of a user printing an object. The first process steps of designing and modeling the object with a CAD or modeling tool are not discussed and we assume the user, which already has an account within the service, logs in and has an AutoCAD DXF⁶ file stored on his computer. As a first action the file is uploaded through the web-interface to the controller that instructs the data management service to store the file in the database, then the file is transformed into STL and AMF [AS13] format for future use and stored in the database. For this first step the user uploads the provided CAD in the web based UI that then instructs the operations via the provided API endpoint. The user then selects a printer from a list provided through the UI for printing. This list is manually generated in this version as the adaption of printer models requires manual interaction due to the lack of missing abstract hardware description capabilities. This information is provided by the device/service registry. Future implementations can suggest an appropriate printer (based on availability or matching capability for the object) to the user. After selecting the printer the user is able to select slicing parameters and position the object in the virtual build environment. Future implementations can suggest appropriate parameters based on analysis of the model file and positioning on optimization criteria (e. g., strength, build time or utilization) to the user. The user is able to add other tool steps to the processing of the object file which are orchestrated by the provisioner and associated virtual computing resources. These processing steps are limited by the capability of the underlying software package to be instructed without graphical interaction (i. e., headless system, e. g., as libraries or command line tools). After the model file is sliced the printing job is instantiated with the scheduler that checks if the requested printer resource is available and if so sends it to the executioner. If not a queue is used to store the job until the resource becomes available. The executioner communicates with the control interface of the 3D printer in order to transfer

⁶ Drawing Interchange Format, http://images.autodesk.com/adsk/files/autocad_2012_pdf_dxf-reference_enu.pdf

and start the print. Sensor data is transmitted back to the executioner from the sensor interface. Sensor data is then stored in the database via the scheduler and the controller.

During the print the user is informed on the progress and possible failure of the print via web interface and the underlying API for access through other means. After completion of the print the user is informed through a notification by the API or any registered service. Data acquired during the print is stored in the database for later analysis and accessible through the API.

This data includes slicing logs, parameters for slicing and positioning, information on the printer, possible sensor data acquired during print, original files and derivative intermediate transformed files and video or picture data from the print.

6 Evaluation

The proposed API is implemented as a REST API as a proof of concept. We chose REST over Web Services [PZL08], because we aim for integration over the Web. The prototype builds upon the LoopBack⁷ framework. This framework is based on node.js⁸ and allows for rapid development and deployment of RESTful APIs by utilizing angularJS⁹ on the client side, providing persistent storage through a number of services like MySQL¹⁰, MongoDB¹¹ or other REST services and providing a backend using StrongLoop PM¹². Through these architectural decisions the implementation satisfies the requirements of scalability, reliability and ease of deployment.

We evaluated it on a Makerbot Replicator 2X¹³ and a German RepRap Model X150¹⁴ to assess the capabilities of controlling different 3D printers. For the assessment of the capabilities of multiple software packages versions of Slic3r¹⁵, MiracleGrue¹⁶, FreeCAD¹⁷, GPX¹⁸, and a custom GCODE parse are implemented. The user is able to select a printer from the web based user interface and add processing steps for an uploaded file. The following example list of operations are possible for users to perform:

- Analyze geometry of provided STL file with GPX
- Convert DXF file to STL with FreeCAD
- Slice STL file with Slic3r or MiracleGrue

⁷ <http://loopback.io>

⁸ <http://nodejs.org>

⁹ <http://angularjs.org>

¹⁰ <http://www.mysql.com>

¹¹ <http://www.mongodb.org>

¹² <http://strong-pm.io>

¹³ <http://www.makerbot.com>

¹⁴ <https://www.germanreprap.com/produkte/3d-drucker/x150-3d-drucker>

¹⁵ <http://slic3r.org>

¹⁶ <http://makerbot.github.io/Miracle-Grue>

¹⁷ <http://www.freecadweb.org>

¹⁸ <http://github.com/whpthomas/GPX>

- Position and transpose object in virtual build envelope with user interface
- Upload file from printing service to 3D printers internal memory

The evaluation is performed by a group of 5 students with low to medium experience in the domain of AM. This group is observed performing a set of the following tasks:

- Upload DXF file to print service and convert it to a printable file (GCODE) for a specific printer
- Rotate and rescale an existing file by a set of provided parameters
- Share a model file and a build log with another user of the service
- Analyze and name the failure reason for a build log provided by the service

7 Conclusion and Future Work

To the best of our knowledge no open source 3D printing service or API for it is published yet. There are existing solutions that focus on separate parts and provide solutions to different aspects of the 3D printing process.

Our approach is characterized and differs from other approaches by: a) Focus on 3D printer, b) Focus on communication with manufacturing device, c) Expandable support for future 3D printers and software, d) User customizable selection of processing operations for files associated with the printing process, e) Platform for testing BPMN extension, f) Smartifying 3D printer, g) Platform for testing sensor array, and h) Interchange format for print related information. This 3D printing API and the associated service is designed as an open research platform for academic users to embed experiments and utilize distributed resources. Further projects are aimed at 1. providing means of control of 3D printers from within process models as we are writing an BPMN extension, based on the work by Zor et al. [ZSL11] tailored for 3D printers and 2. utilize sensors for print status observation and as a means for quality research into 3D printing (see ICRM 2016¹⁹). Those projects are to be incorporated in the umbrella project described in this work. As a related project we develop a BPMN extension for 3D printer integration into BPMN where the hardware resources and data flows can be modeled using the extension.

Acknowledgments

This work is partially funded by the BMWi project “SmartOrchestra” (01MD16001F). We would also like to thank Anna Kulischkin, Marc Tuscher, Jonas Koss, and Niklas Kleinhans for their participation in this work.

¹⁹ <http://icrm-aachen.com>

References

- [ALI12] Afsharhanaei, Ali; Leman, Zulkiflle; Ismail, N.: Comparison of STEP-NC with ISO6983 in Terms of Manufacturing Interoperability. *Applied Mechanics and Materials*, 229–231:2556–2561, 2012.
- [AS13] ASTM ISO: . ASTM52915-13, Standard Specification for Additive Manufacturing File Format (AMF) Version 1.1. ASTM_International, West Conshohocken, PA, 2013.
- [Ba11] Banerjee, Prith; Friedrich, Richard; Bash, Cullen; Goldsack, Patrick; Huberman, Bernardo; Manley, John; Patel, Chandrakant; Ranganathan, Parthasarathy; Veitch, Alistair: Everything as a Service: Powering the New Information Economy. *Computer*, 44(3):36–43, 2011.
- [Bo14] Bonten, Christian: *Kunststofftechnik*. Carl Hanser Verlag GmbH & Co. KG, 2014.
- [BR16] Baumann, Felix; Roller, Dieter: 3D Printing Process Pipeline on the Internet. In: *Proceedings of the 8th ZEUS Workshop*. pp. 29–36, 2016.
- [Do08] Dong, Baoli; Qi, Guoning; Gu, Xinjian; Wei, Xiuting: Web service-oriented manufacturing resource applications for networked product development. *Advanced Engineering Informatics*, 22(3):282–295, 2008.
- [GL13] Guo, Nannan; Leu, Ming C.: Additive manufacturing: technology, applications and research needs. *Frontiers of Mechanical Engineering*, 8(3):215–243, 2013.
- [GRS15] Gibson, Ian; Rosen, David; Stucker, Brent: *Additive Manufacturing Technologies*. Springer-Verlag New York, 2 edition, 2015.
- [GS14] Gribbins, Cassandra; Steinhauer, Heidi M.: Experimental Analysis on an Additively Manufactured ABS Living Hinge. In: *Proceedings of the 2016 Annual International Solid Freeform Fabrication Symposium*. 2014.
- [Ha14] Harrer, Simon; Lenhard, Jörg; Wirtz, Guido; van Lessen, Tammo: Towards Uniform BPEL Engine Management in the Cloud. In: *Informatik*. pp. 259–270, 2014.
- [KMD00] Kulkarni, Prashant; Marsan, Anne; Dutta, Debasish: A review of process planning techniques in layered manufacturing. *Rapid Prototyping Journal*, 6(1):18–35, 2000.
- [Ku15] Kubler, Sylvain; Holmström, Jan; Främling, Kary; Turkama, Petra: *Technological Theory of Cloud Manufacturing. Service Orientation in Holonic and Multi Agent Manufacturing and Robotics*, 2015.
- [Mo11] Mota, Catarina: *The Rise of Personal Fabrication*. In: *C&C*. ACM, 2011.
- [Mo15] van Moergestel, Leo; Puiik, Erik; Telgen, Daniël; Meyer, John-Jules: Implementing Manufacturing as a Service: A Pull-Driven Agent-Based Manufacturing Grid. In: *ICTERI*. 2015.
- [PZL08] Pautasso, Cesare; Zimmermann, Olaf; Leymann, Frank: RESTful Web Services vs. Big Web Services: Making the Right Architectural Decision. In: *WWW '08*. ACM, 2008.
- [Sa12] Santos, André ©.; Pedrosa, Luãs D.; Kuipers, Martijn; Rocha, Rui M.: Resource Description Language: A Unified Description Language for Network Embedded Resources. *International Journal of Distributed Sensor Networks*, 2012:11, 2012.
- [Vu15] Vukovic, Maja: Internet Programmable IoT: On the Role of APIs in IoT: The Internet of Things. *Ubiquity*, pp. 3:1–3:10, November 2015.

- [Wu12] Wu, Dazhong; Thames, J. Lane; Rosen, David W.; Schaefer, Dirk: Towards a Cloud-Based Design and Manufacturing Paradigm: Looking Backward, Looking Forward. In: Computers and Information in Engineering Conference. 2012.
- [Wu15] Wu, Dazhong; Rosen, David W.; Wang, Lihui; Schaefer, Dirk: Cloud-based design and manufacturing: A new paradigm in digital manufacturing and design innovation. *Computer-Aided Design*, 59:1–14, 2015.
- [ZSL11] Zor, Sema; Schumm, David; Leymann, Frank: A Proposal of BPMN Extensions for the Manufacturing Domain. In: Proceedings of the 44th CIRP International Conference on Manufacturing Systems. 2011.

Bridging Gaps in Cloud Manufacturing with 3D Printing

Christoph Hochreiner,¹ Philipp Waibel,² Michael Borkowski³

Abstract: In order to keep up with the flexibility and cost-efficiency requirements, the manufacturing industry has begun a transformation process towards a service-oriented production approach. This approach is in line with the cloud manufacturing paradigm that applies established concepts from cloud computing to the manufacturing domain. Although the cloud manufacturing paradigm provides promising concepts for a flexible manufacturing industry, it still exposes several challenges. One of the most important challenges is the recovery from production failures and delays. Due to the physical configuration and transportation overhead in the manufacturing domain, it is often not possible to quickly replace one production facility with another one. To bridge the gap of these production delays, we propose the integration of additive manufacturing technologies tightly into cloud manufacturing processes to ensure the continuity of the production processes.

Keywords: Cloud Manufacturing, Additive Manufacturing, 3D Printing, Industry 4.0

1 Introduction

In the last couple of years, the manufacturing industry has begun a fundamental transformation from static production processes towards flexible network-based ones. Due to the increasing flexibility of customer requests towards quantity, but also customizability of goods, the manufacturing industry needs to continuously adapt the production processes to keep up with these demands [TST12]. Furthermore, the manufacturing industry also needs to stay competitive towards competitors and therefore aims at maximizing the usage of their production facilities.

One promising approach to tackle these challenges is to adopt the successful concepts from the field of cloud computing and transform the manufacturing capabilities from static building blocks into flexible services. Flexible services can then be leased on-demand, in a utility-like manner and allow to compose dynamic production processes that can be adapted to the ever-changing customer demands. These concepts have been introduced as cloud manufacturing [Xu12] and although the theoretical foundation for cloud manufacturing is present [Sc14], there are still several challenges for adopting these concepts in real-world scenarios. To resolve these challenges and to provide a holistic solution approach, the CREMA research project [Sc16] realizes a framework to coordinate the services across numerous manufacturing companies to realize vast manufacturing networks. These manufacturing networks allow a dynamic and efficient realization of production processes by orchestrating individual manufacturing capabilities in a service oriented manner, as described in detail in Section 2.

¹ TU Wien, Distributed Systems Group, Argentinierstrasse 8, 1040 Wien, c.hochreiner@infosys.tuwien.ac.at

² TU Wien, Distributed Systems Group, Argentinierstrasse 8, 1040 Wien, p.waibel@infosys.tuwien.ac.at

³ TU Wien, Distributed Systems Group, Argentinierstrasse 8, 1040 Wien, m.borkowski@infosys.tuwien.ac.at

Although the CREMA framework provides a promising solution for most application scenarios, there are still scenarios that cannot be covered with the current approach. This specifically applies to scenarios, where one specialized service fails, e.g., a manufacturing machine failure or shortage on starting material, and cannot be replaced by other services. In such a situation, the CREMA framework currently needs to wait until the service is mended again or another identical service becomes available within the manufacturing network. In order to eliminate the resulting production delay, the concept of additive manufacturing is predestined to fill the gap [KLN98]. Modern additive manufacturing techniques are capable of producing objects following any arbitrary shape and expose similar quality levels as traditionally produced objects.

Despite the fact that current additive manufacturing machines, e.g., 3D printers, cannot cope with the production makespan of specialized machines, they excel at the almost negligible setting-up time for new product shapes. This makes these machines a perfect fit to support the transition from traditional production processes towards utility based ones [KPH14, Wu15].

In this paper, we are going to provide more detailed information about the CREMA framework and argue how additive manufacturing technologies can become a catalyst for the cloud manufacturing era.

2 Framework Overview

In order to discuss the CREMA framework, we consider an exemplary scenario from the manufacturing domain, as shown in Figure 1. In this process, Company 1 produces a product, based on parts from two suppliers. These parts are then assembled in their plant in Germany before they get their final polish in the plant in the United Kingdom. To realize the production process, the CREMA framework decomposes the overall process into individual steps, which are represented by virtualized services.

At the beginning of the process realization phase, a process designer of Company 1 decomposes the production process into virtualized services. After the decomposition, the process designer queries the Manufacturing Network to find suitable services. While the latter two services are already fixed due to the fact that Company 1 wants to use their own production capabilities, the supplier steps are chosen in an ad-hoc manner, based on the availability, price and quality aspects. After the initialization and configuration phase is finished, the CREMA framework takes care of the instantiation of the production process, i.e., triggers the real world manufacturing activities and takes care of the billings aspects [Sc16].

Whenever a problem, like a production failure or transportation delay, occurs, the CREMA framework needs to update the production process. This is done by finding new services on the Manufacturing Network, which are able to replace the failing ones. For our example scenario, we assume that Supplier 2 is not able to deliver the required parts in time. In this situation, the CREMA framework queries the Manufacturing Network to find a suitable replacement on short notice. While this replacement operation can be solved quickly in a cloud computing scenario, it is more complex for a cloud manufacturing one.

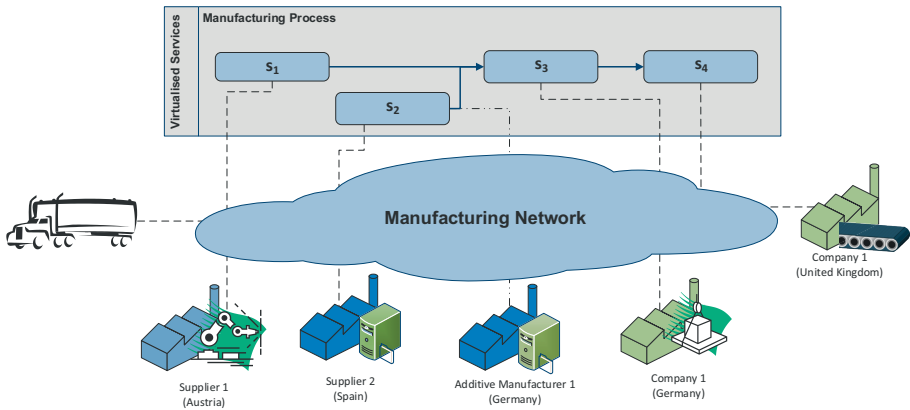


Fig. 1: Exemplary Scenario

In the manufacturing domain, we also have to consider the transport aspects to move the goods from one location to another as well as the long setup time to start the production of a new product shape. Although the Manufacturing Network offers numerous services, which may be suitable as a replacement, it is rather unlikely to find one which is willing to produce a small quantity of goods and is geographically near the plant of Company 1 in Germany that requires the parts, to minimize the transport routes.

In this situation, additive manufacturing services, like the UPS 3D printing service⁴ or 3D Hubs⁵, provide promising solutions for a short term replacement. These service providers offer printing services in a service oriented manner and can be easily integrated into the existing Manufacturing Network. Although the price per unit is typically higher for 3D printing services, compare to traditional manufacturing techniques, these services do not require long setup times and do not care about the quantity of parts they produce. Furthermore, the already existing printing providers offer their services at multiple geographic locations, which makes it easy to find a nearby replacement service, such as the Additive Manufacturer 1 in our exemplary scenario.

3 Outlook

The previous section sketches one of the major challenges for the widespread implementation of cloud manufacturing, namely the slow recovery possibilities whenever a failure or a delay occurs within one specific production step.

To mitigate this, we propose the integration of additive manufacturing techniques into cloud manufacturing paradigm. This allows companies to be more flexible in terms of replacing faulty services to mitigate negative effects on their production processes.

⁴ <https://www.theupsstore.com/print/3d-printing>

⁵ <https://www.3dhubs.com>

Our exemplary scenario shows, that additive manufacturing may be the most promising solution to bridge gaps in cloud manufacturing, these technologies can be also used to foster the individualization of products. Due to the fact that 3D printers require hardly any physical setup time, they are applicable to produce very small quantities of products, which is not feasible with traditional production approaches.

Although additive manufacturing seems to be the perfect solution for cloud manufacturing, there are still several challenges that need to be solved before they can be integrated in the same manner as traditional manufacturing machines. The most prominent challenge is the quality of the printed products. While some 3D printers are capable of printing products in a high quality, there are still numerous printers which cannot provide the required quality. Additionally, each printed object is unique and requires individual quality control. Here it is necessary to develop effective as well as efficient quality assurance techniques, which are suitable for this individual characteristic of manufacturing products.

Acknowledgments

This paper is supported by TU Wien research funds. This work is partially supported by the Commission of the European Union within the CREMA H2020-RIA project (Grant agreement no. 637066).

References

- [KLN98] Kruth, Jean-Pierre; Leu, Ming-Chuan; Nakagawa, Terunaga: Progress in additive manufacturing and rapid prototyping. *CIRP Annals-Manufacturing Technology*, 47(2):525–540, 1998.
- [KPH14] Khajavi, Siavash H.; Partanen, Jouni; Holmström, Jan: Additive manufacturing in the spare parts supply chain. *Computers in industry*, 65(1):50–63, 2014.
- [Sc14] Schulte, Stefan; Hoenisch, Philipp; Hochreiner, Christoph; Dustdar, Schahram; Klusch, Matthias; Schuller, Dieter: Towards process support for cloud manufacturing. In: 2014 IEEE 18th Int. Enterprise Distributed Object Computing Conf. (EDOC). IEEE, pp. 142–149, 2014.
- [Sc16] Schulte, Stefan; Borkowski, Michael; Hochreiner, Christoph; Klusch, Matthias; Murguzur, Aitor; Skarlat, Olena; Waibel, Philipp: Bringing Cloud-based Rapid Elastic Manufacturing to Reality with CREMA. In: Workshop on Intelligent Systems Configuration Services for Flexible Dynamic Global Production Networks (FLEXINET) at the 8th Int. Conf. on Interoperability for Enterprise Systems and Applications (I-ESA 2016). pp. NN–NN, 2016.
- [TST12] Taisch, Marco; Stahl, Bojan; Tavola, Giacomo: ICT in manufacturing: Trends and challenges for 2020 – An European view. In: 10th International Conference on Industrial Informatics (INDIN). IEEE, pp. 941–946, 2012.
- [Wu15] Wu, Dazhong; Rosen, David W.; Wang, Lihui; Schaefer, Dirk: Cloud-based design and manufacturing: A new paradigm in digital manufacturing and design innovation. *Computer-Aided Design*, 59:1–14, 2015.
- [Xu12] Xu, Xun: From cloud computing to cloud manufacturing. *Robotics and computer-integrated manufacturing*, 28(1):75–86, 2012.

Supporting the Set-up Processes by Cyber Elements based on the Example of Tube Bending

Nils Darwin Abele¹, Sven Hoffmann², Christopher Kuhnhen³, Thomas Ludwig⁴, Walter Schäfer⁵, Marcus Schweitzer⁶ and Volker Wulf⁷

Abstract: The increasing demand for individual and customized products constrains industrial companies to produce lower quantities which caused higher set-up times of the machines and therefore higher costs per piece compared to their traditional mass production. Especially within small and medium enterprises, the set-up processes often take place manually. However, the result is a critical factor for production success itself. During the set-up process, the worker is confronted with an information overload caused by the complex production system such as those like a tube bending process. Within this paper we outline a project called ‘Cyberrüsten 4.0’ that aims to provide a cyber element which is able to display needed information and adapted feedback for the set-up process. For this purpose, there are several steps of a practice-based research necessary. These steps were framed by the Design Case Study.

Keywords: Cyber Elements, Design Case Study, Mental Models, Usability, Set-up process, Information Overload, Feedback, Tube bending process, Additive Manufacturing

1 Introduction

Globalization and the resulting larger markets as well as the cheaper production possibilities offer a lot of potential to industrial companies. However, globalization is often accompanied by a number of challenges, too, because global markets imply global competitors. Industrial companies need to adapt their products to market trends at even shorter intervals. In addition, market positions must be expanded with new more advanced products offering higher quality at competitive prices. It is a constant tension

¹ Universität Siegen, Lehrstuhl für Technologiemanagement, Unteres Schloss 3, 57072 Siegen, darwin.abele@uni-siegen.de

² Universität Siegen, Lehrstuhl für Wirtschaftsinformatik und Neue Medien, Kohlbettstraße 15, 57072 Siegen, sven.hoffmann@uni-siegen.de

³ Universität Siegen, Lehrstuhl für Umformtechnik, Paul-Bonatz-Straße 9 – 11, 57076 Siegen, christopher.kuhnhen@uni-siegen.de

⁴ Universität Siegen, Institut für Wirtschaftsinformatik und Neue Medien, Kohlbettstraße 15, 57072 Siegen, thomas.ludwig@uni-siegen.de

⁵ Universität Siegen, Lehrstuhl für Wirtschaftsinformatik und Neue Medien, Kohlbettstraße 15, 57072 Siegen, walter.schäfer@uni-siegen.de

⁶ Universität Siegen, Lehrstuhl für Technologiemanagement, Unteres Schloss 3, 57072 Siegen, marcus.schweitzer@uni-siegen.de

⁷ Universität Siegen, Lehrstuhl für Wirtschaftsinformatik und Neue Medien, Kohlbettstraße 15, 57072 Siegen, volker.wulf@uni-siegen.de

area between customer-driven cost pressure, quality demands and features expected of the products that customers request.

The satisfaction of customer requests and their steadily increasing demand for individual and customized manufacturing products led to an increased number of product variations. An early attempt trying to tackle this challenge was the process-centered concept of 'lean production' that aims at meeting the requirements of the flexibility with which companies must react to global market changes. One characteristic of lean production is that production planning does not primarily optimize machine capacity, but is geared towards customer demands [WJR90]. Lean production and cyber-physical systems offer new options for the physical process design [Bi07]. The changes brought about by lean production were so far-reaching that they signify a paradigm shift – especially in European and Northern American industrial companies. More recently, the concept of lean production has been expanded towards a socio-technical system that integrates the entire company, customers, suppliers, maintenance and process control as well as appropriate feedback loops [SW07]. But following the concept of lean production, the increasing demand for individual and customized products constrains industrial companies to produce lower quantities that caused higher set-up times of the machines and therefore higher costs per piece compared to their traditional mass production. Especially the effort of a machine set-up rises dramatically related to its switch-on time. According to individual products, the workers are not able to develop appropriate routine with reference to set-up processes.

Based on the background of increasing complexity, individualization and industrialization the additive manufacturing plays an important role [LKP15]. As a flexible manufacturing technology it contributes to reduce the production time of high quality products in small batches. In the framework of this technique the operator can be assisted by cyber-physical elements respectively digital models. Due to fast operating and set-up concepts, the additive manufacturing is used economically and efficiently. The social-technical field of application can be transferred into additive manufacturing process properties.

Therefore, we want to outline an approach that will tackle the high set-up times machine workers are challenged with during the manufacturing processes. For example, the rotary draw bending process requires six to eight tools which have to be adjusted within the set-up process. Set-up times are in the range of 60 to 90 minutes under laboratory conditions. In contrast, the set-up of a basic injection molding process only includes one tool and takes place in a time interval of about five to fifteen minutes. The question we try to answer is how to support the worker during the set-up processes based on modern information and communications technology (ICT) tools composed of a software managed physical tool. In section 2 we outline the process of rotary draw bending and the challenges workers are confronted with. Section 3 introduces the methodological framework, whereby section 4 constitutes our empirical-based approach for gathering information about the actual and informal practices of workers during set-up processes. Based on this method, we will take a look at future work.

2 Rotary Draw Bending of Profiles and the Information Overload for the Manufacturing Workers

The rotary draw bending is a manufacturing process with a complex tool setting (see fig. 1) [EM13], [VDI3430]. The complexity is caused by a trial period in which the different independent machine axes are adjusted. The semi-finished product is fixed between the outer and inner clamp die. After that, it is bended around the tool center point of the bend die. The pressure die is a tool for fixing the straight part of the tube in front of the forming area. With the collet's help, the tube can be fed and pushed in the forming direction.

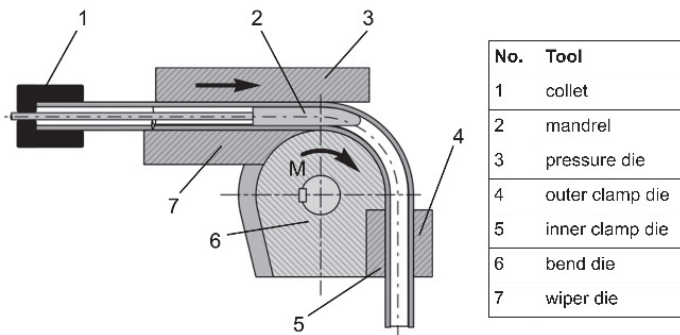


Fig. 1: Tool setting of a rotary draw bending process [VDI3430]

Depending on the geometrical factors – in this context the bending factor B (bending radius to tube diameter) and wall-thickness factor W (tube diameter to wall-thickness of the tube) are worth mentioning – the number of required forming tools increases immediately with the raise of forming complexity (see fig. 2) [En14]. A mandrel, for instance, has to be used to fill up the gap in the middle of the tube. Out of these circumstances, the machine operator has to decide which tool has to be adjusted in which position and which force has to be tuned in. Furthermore, different paths of trying out the production process lead to the goal of producing an accurate tube.

The current state of the art reveals to adjust the process on the staff's knowledge base comparable to the principle of trial and error [He15], [Ko13]. The cause lies in the minimal standardization of bending processes. Therefore, the expert's know-how leads to an island of information. These information will not be regular in other bending operation. Due to this, the result of a good bending operation getting an accurate bent part and even the time to try out the manufacturing process depends on the machine operator. If an employee with an extensive experience adjusts the machine, she/he will reach the goal faster than an employee with less experience in job.

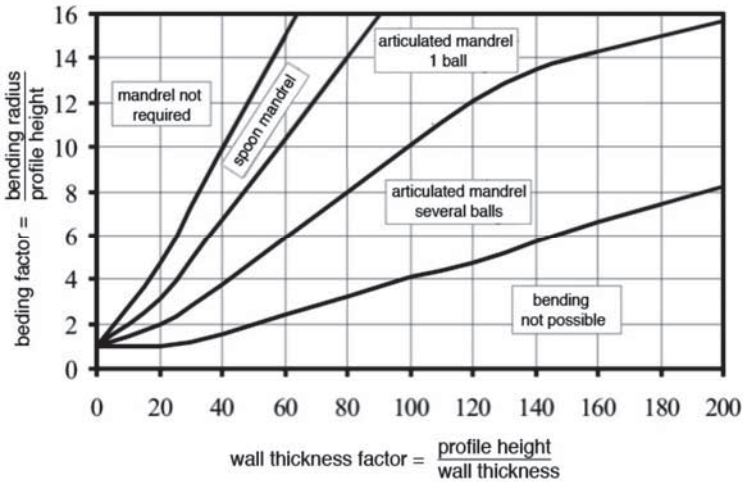


Fig. 2: Bending diagram for mandrel selection [VDI3430]

Next to the problems of saved and badly communicated information about the adjustment steps, different material characteristics of the semi-finished products (straight tubes) lead to more complex situations. For example, a well operating bending process can change into a bad one if the delivery charge of the semi-finished product is changed.

At the end of the bending operation the manufacturer gains a lot of specific knowledge regarding the bent part. These characteristics have to be measured and evaluated to the technical usability of the bent part as shown in [Ku16] to sell the product to the customer. To the operator this induces a situation of a highly complex task, because there are a lot of dependent characteristics of the machinery and the material, which cannot be anticipated in advance and may be influenced in a dynamic procedure [BJ95]. Observing this procedure therefore makes it necessary to consider the different personalities for the operators. That leads to different levels of perceived complexity [Ca88] and different levels of information are processed without an overload [EJ04], [Wi01].

3 Design Case Study as Methodological Framework

As methodological framework, the project focuses on Design Case Studies. The main goal of new developments in an industrial surrounding aims to reinforce the economic efficiency of companies. In addition to that, formal organizational structures will decline and informal aspect as well as the differentiation of work practices will increase. ICT tools help to reach this goal by combining technological, organizational and social perspectives. However, to integrate such a tool, it is necessary to follow up a practical oriented approach of designing this tool. Typically, three main steps are necessary in the

context of a Design Case Study. The first step is the empirical pre study with a focus on social practices. It even consists of micro level descriptions of social practices. In the following step a context-oriented design process of the ICT tool takes place. Next to the social aspects, the tool's description from a product and a process perspective should not be neglected. Finally, the Design Case Study records the introduction, appropriation and potential re-design of the ICT tool [Wu09], [Wu11].

In this case the main task of the ICT tool or other cyber elements will be to provide information and feedback. The evaluated data defines which information is needed and mental models contribute to prepare feedback appropriate to the user.

The whole design work will be attended by the usability idea. This means a participatory work progress between the development sector and the subsequent users [Wu15]. Nevertheless, the ICT tool will be used in an industrial surrounding. As a result of this, the tool must exhibit a robust construction and an ergonomic shape, that is adapted to the work-day.

After certain design and construction phases, there will be different appropriations with the users of the tool. The guiding principle is to go through several iteration loops. In the course of the iteration loops, the ICT tool is continually adjusted by the users' needs. However, we expect to explore some opportunities to improve the tool. Empirical methods will help us to detect the changed user behavior and even to recognize improvements while the set-up process takes place.

4 Empirical Pre-Study

In order to get insights into the worker's specific knowledge and informal practices, we will conduct an empirical study. This empirical pre-study will take place in the context of a Design Case Study and fulfills the task of providing a knowledge base consisting of a large technical and social understanding. Building on the knowledge base the ICT tool will be developed. Furthermore, a common project language for communication with the practitioners will be established. The result of the pre-study is gaining a common understanding of problems and aspects which offer potentials to be improved [Ro16].

4.1 Empirical Access in the Production Environment

The testing field for the investigation of the empirical data will be an initial set-up process. Later on, it is still possible to expand the testing field on further set-up actions.

The common empirical pre-study includes several well-known methods as described in [PD15]. In this case, the challenge consists of the complex production environment in combination with a manually embossed set-up process. To face the given challenge, we need several perspectives to create an empirical data base [Wu15]. In a first step, we will

concentrate on the technological background. Next to the machine-equipment, several other influences have to be taken into consideration, for example the bending devices or the semi-finished products. The analysis of documents (set-up plan, work instructions and user manuals) will be extended by intensive observations and expert discussions.

In a second step, the organizational structures will be determined. The empirical study tries to answer the question of how the set-up process is integrated in the entire manufacturing process. The set-up process is a manual process. As a result, the social perspective plays an important role. The human resource as well as the work flow has therefore to be examined. In this context, ethnographical methods such as interviews and expert discussions will take place. We expect to reveal both formal and informal organizational structures.

4.2 Possibilities for Using Mental Models

As a logical consequence of personalized internal cognitions, mental models represent a reduced image of the reality respectively of an original system [Jo83]. This is also known as tacit knowledge that controls action- and object-related human behavior on a hierarchical basis [FZ94], [Ha86], [Ha96]. Thus, every human being constitutes different homomorphic models to avoid an information overload through the reduction of cognitive effort and complexity [Ha96], [Si76]. To enable an effective transfer of tacit and formal knowledge, studies have shown that the variation of instruction and feedback is context-sensitive [LSZ12]. Concretization and contextualization of linguistic, metaphorical or sensual expressions take place via using analogies by falling back on known issues or experiences [GG83], [La87], [Sc94].

To support the machine operator within the set-up process by means of cyber-physical elements, initially mental models have to be derived. These models help to interpret work plans correctly and to control decision-making efficiently. Building on this, information of prevailing-processes as well as the cyber-physical adjustment instructions should be deduced, linked with the machine sensors and returned via user-friendly visualization. The adjustment implies a complex task. Therefore, it is to be expected that the operators use different heuristics to prevent an information overload. According to this, the development of various mental models is foreseeable, whereby a wide variety of features can influence the model design. At first, personal characteristics, e.g. perceived complexity [Wo87], [Ca88], motivation and attractiveness [LL90], [LL02] as well as risk perception and risk behavior [WBB02] or risk-reducing strategies [CL06] and specific knowledge [Ce08] have to be tested on their effects. Furthermore, context attributes that influence the operator's decision have to be used [JL09], [HNM07]. Therefore, his hierarchical embedding and his strains through the workplace should be considered, too.

Regarding the system to be developed as a feedback system, it leads to the approach of mental and conceptual mental models to get an accurate, consistent and complete system's representation [No83]. For an application the feedback-intervention-theory has

to be considered. Following the theory feedback mediates positive and negative effects on performance both [KD96]. This finding is partly supported by Illies and Judges [IJ05] who show diminishing effects of positive feedback. The adaptation of mental models to further set-up processes and also to further sectors gains a crucial importance to generate synergies, especially with reference to the transfer of tacit knowledge. Based on essential influencing factors, categories or cluster should be constituted to map different learning effects depending on time and quality.

5 Conclusion and Outlook

The project's main target is to support the operator using a cyber-physical element within the set-up based on the example of a tube bending process. The reduction of the set-up process' complexity should be generated by the creation of decision-makings. The connection of setting procedures and 'best practice' provides the basis for IT-supported assistance through visualization tools. Those tools should explicate the tacit knowledge in a simple manner and initiate an optimized feedback and learning process to reduce the recorded cognitive challenges. Essentially, the approach aims at the generation of low-waste production. At a later point of time, the initial limitation on the tube bending process will be abrogated and other set-up processes will be taken into account.

Acknowledgements

This paper belongs to the research project 'Cyberrüsten 4.0: Cyber-physische Unterstützung des Menschen beim Rüstvorgang am Beispiel eines Biegeprozesses zur Kleinserienfertigung auf Basis eines Wissenstransferansatzes' and was funded by a grant of the European Union and EFRE.NRW (No. EFRE-0800263).

We also thank Matthias Betz for his support on this project.



EUROPEAN UNION
Investing in our Future
European Regional
Development Fund



EFRE.NRW

Investitionen in Wachstum
und Beschäftigung

References

- [Bi07] Birkhan, C.: Smart Production Systems – intelligente Konzepte zur Gestaltung von Produktionssystemen. 2007.
- [BJ95] Byström, K.; Järvelin, K.: Task complexity affects information seeking and use. In: *Information Processing & Management* 31, S. 191-213, 1995.
- [Ca88] Campbell, D.J.: Task Complexity: A Review and Analysis. In: *Academy of Management Review* 13, 40-52, 1988.
- [Ce08] Cardinaels, E.: The interplay between cost accounting knowledge and presentation formats in cost-based decision-making. In: *Accounting, Organizations and Society* 33, S. 582-602, 2008.
- [CL06] Cho, J.; Lee, J.: An integrated model of risk and risk-reducing strategies. *Journal of Business Research* 59, S. 112-120, 2006.
- [EJ04] Eppler, M.J.; Mengis, J.: The Concept of Information Overload – A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. In: *The Information Society: An International Journal* 20, S. 1-20, 2004.
- [EM13] Engel, B.; Mathes, C.: Erweiterung der Prozessfähigkeit des Rotationszugbiegens durch ein alternatives Faltenglätterkonzept, EFB-Forschungsbericht, 378, 2013 Hannover, Europäische Forschungsgesellschaft für Blechverarbeitung e.V. (EFB), ISBN 978-3-86776-421-6.
- [En14] Engel, B.; Kuhnhen, C.; Mathes, C.; Groth, S.; Frohn, P.; Heftrich, C.: A new approach for quantitative evaluation of bending components. *Metalforming 2014*, Palermo, Italien, September 2014, S. 1095-1102, 2014.
- [FZ94] Frese, M.; Zapf, D.: Action as the Core of Work Psychology: A German Approach. In (Triandis, H.C.; Dunnette, M.D.; Hough, L.M., eds.): *Handbook of Industrial and Organizational Psychology*, 2nd Ed., Vol. 4, Palo Alto/CA, S. 271-340, 1994.
- [GG83] Gentner, D.; Gentner D.R.: Flowing waters and teeming crowds: Mental models of electricity. In (Gentner, D.; Stevens, A.L., eds.): *Mental models*, Erlbaum, Hillsdale/New York, S. 99-129, 1983.
- [Ha86] Hacker, W.: *Allgemeine Arbeitspsychologie: psychische Regulation von Arbeitstätigkeiten*. 2. Aufl., Huber, Bern, 1986.
- [Ha96] Hacker, W.: Handlungsleitende psychische Abbilder (“Mentale Modelle”). In (Kühl, J.; Heckhausen, H., Hrsg.): *Motivation, Volition und Handeln*, Serie IV, Bd. 4, Hogrefe, Göttingen/Bern/Toronto/Seattle, S. 769-794, 1996.
- [He15] Heller, B.; Chatti, S.; Schikorra, M.; Tekkaya, A.; Kleiner, M.: Blechbiegen. In (Siegert, K., Hrsg.): *Blechumformung*. Springer-Verlag, Berlin, S. 141–216, 2015
- [HNM07] Humphrey, S.E.; Nahrgang, J.D.; Morgeson, F.P.: Integrating Motivational, Social, and Contextual Work Design Features: A Meta-Analytic Summary and Theoretical Extension of the Work Design Literature. In: *Journal of Applied Psychology* 92, S. 1332-1356, 2007.

- [IJ05] Illies, R.; Judge, T.: Goal Regulation Across Time: The Effects of Feedback and Affect. In: *Journal of Applied Psychology* 90, S. 453-467, 2005.
- [JL09] Joo, B.-K.; Lim, T.: The Effects of Organizational Learning Culture, Perceived Job Complexity, and Proactive Personality on Organizational Commitment and Intrinsic Motivation. In: *Journal of Leadership & Organizational Studies* 16, S. 48-60, 2009.
- [Jo83] Johnson-Laird, P.N.: *Mental Models*. Harvard University Press, Cambridge, 1983.
- [KD96] Kluger, A.; DeNisi, A.: The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. In: *Psychological Bulletin* 119, S. 254-284, 1996.
- [Ko13] Korte, J.: Biegen, Messen und Nachbiegen in einem Arbeitsschritt. *MM – Maschinenmarkt*, Nr. 14, Jahrgang 58, S. 27, 2013.
- [Ku16] Kuhnhen, C.; Engel, B.; Zehner, B.-U.; Alles, E.: FE-Simulation zur Ermittlung der Widerstandskräfte gebogener Rohrkonstruktionen in Abhängigkeit der Faltenbildung. In: *35. Verformungskundliches Kolloquium (05.03.-09.03.2016)*, Lehrstuhl für Umformtechnik, Montanuniversität Leoben, S. 147-152, 2016.
- [La87] Lakoff, G.: *Women, Fire, and Dangerous Things. What Categories Reveal About the Mind*. University of Chicago Press, Chicago, 1987.
- [LSZ12] Letmathe, P.; Schweitzer, M.; Zielinski, M.: How to learn on new tasks: Shop floor performance effects of knowledge transfer and performance feedback. In: *J. Oper. Manag.* 30, S. 221-236, 2012.
- [LL90] Locke, E.A.; Latham, G.P.: *A Theory of Goal Setting and Task Performance*. Prentice Hall, Englewood Cliffs, NJ, 1990.
- [LL02] Locke, E.A.; Latham, G.P.: Building a Practically Useful Theory of Goal Setting and Task Motivation. In: *American Psychologist* 57, S. 705-717, 2002.
- [LKP15] Ludwig, T.; Kotthaus, C.; Pipek, V.: Should I try turning it off and on again? Outlining HCI Challenges for Cyber-Physical Production Systems. In: *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)*, 7 (3) pp. 55-68, 2015.
- [No83] Norman, D.A.: Some Observations on Mental Models. In (Gentner, D.; Stevens, A.L., eds.): *Mental Models*, Erlbaum, Hillsdale/New York, S. 7-14, 1983.
- [PD15] Preim, B.; Dachsel, R.: *Interaktive Systeme*. Springer-Verlag, Berlin, 2015.
- [Ro16] Rohde, M.; Brödner, P.; Stevens, G.; Betz, M.; Wulf, V.: Grounded Design – a Praxeological IS Research Perspective. In: *Journal of Information Technology (JIT)*, 2016, accepted for publication.
- [Sc94] Schnotz, W.: *Aufbau von Wissensstrukturen. Untersuchungen zur Kohärenzbildung beim Wissenserwerb mit Texten*. Beltz PsychologieVerlagsUnion, Weinheim, 1994.
- [Si76] Simon, H.: *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*. 3. Aufl., Free Press, New York, 1976.
- [SW07] Shah, R.; Ward, P.T.: Defining and developing measures of lean production. In: *Oper. Manag.* 25, S. 785–805, 2007.

- [WBB02] Weber, E.U., Blais, A.-R.; Betz, N.E.: A Domain-specific Risk-attitude Scale: Measuring Risk Perceptions and Risk Behaviors. In: *Journal of Behavioral Decision Making* 15, S. 263-290, 2002.
- [Wi01] Wilson, T.D.: Information overload: implications for healthcare services. In: *Health Informatics Journal* 7, S. 112-117, 2001.
- [WJR90] Womack, J.P.; Jones, D.T.; Roos, D.: *The Machine that Changed the World: The Story of Lean Production*. 1990.
- [Wo87] Wood, R.E.; Mento, A.J.; Locke, E.A.: Task Complexity as a Moderator of Goal Effects: A Meta-Analysis. In: *Journal of Applied Psychology* 72, S. 416-425, 1987.
- [Wu09] Wulf, V.: Theorien sozialer Praktiken zur Fundierung der Wirtschaftsinformatik: Eine forschungsprogrammatische Perspektive. In (Becker, J.; Krcmar, H.; Niehaves, B., eds.): *Wissenschaftstheorie und gestaltungsorientierte Wirtschaftsinformatik*, Springer/Physika, S. 211–224, 2009.
- [Wu11] Wulf, V.; Rohde, M.; Pipek, V.; Stevens, G.: Engaging with Practices: Design Case Studies as a Research Framework. In: *CSCW Proc. ACM Conf. Comput. Coop. Work.* S. 505–512, 2011.
- [Wu15] Wulf, V.; Müller, C.; Pipek, V.; Randall, D.; Rohde, M.; Stevens, G.: Practice based Computing: Empirically grounded Conceptualizations derived from Design Case Studies. In (Wulf, V.; Schmidt, K.; Rendall, D., eds.): *Designing Socially Embedded Technologies in the Real World*, Springer, London, S. 111–150, 2015.
- [VDI3430] VDI 3430 - Rotary Draw Bending of profiles, Beuth Verlag, 2014.

The custom-made body – Legal aspects of bioprinted tissue and organs

Jessica Mihalyi, Ass.Iur.¹ and RAin Anne-Kathrin Müller²

Abstract: Progress in bioprinting, combining 3D printing technology and the science to grow human tissue in the laboratory, elicits legal analysis as to if and how the law protects a patient's life, health and confidentiality. We argue that the Federal Data Protection Act and the Medicinal Products Act provide ample protection to the recipient of printed tissue and organs. This paper focuses on the legal situation in Germany.

Keywords: bioprinting, tissue engineering, data protection, liability for medicinal products

1 Introduction

While additive manufacturing is widely used for producing implantable medical devices, the thought of 3D printed tissue and organs is still unsettling.

Tailor-made tissue, such as skin, kidney or liver tissue, is critical not only because in Germany alone more than 10.000 people are waiting for a donor organ [Eu16], but also because of the risk of the immune system rejecting a donor organ. A risk which can only be mitigated by heavy medication. In the future, patients providing their own cells for organ growth could potentially minimize or eliminate those risks altogether. In addition, further benefits of printed tissue may include facial reconstruction, reconstruction of limbs as well as the reduction of reliance on animal testing.

Research on tissue engineering and bioprinting is on a fast track. In 2010 researchers at the Wake Forest Institute for Regenerative Medicine (WFIRM) bioprinted skin on burn injuries, which led to the closure of the wound with the printed skin being incorporated with regular skin in the healing process. [Ba16, S.4, Bi10] Although it wasn't fully functional, a kidney was printed in 2011. [K115] In general, the main challenge in developing living tissue with more than 100-200 µm thickness is the lack of nutrition diffusion to the cells. [K115] One solution may be provided by a 3D printing technology, developed by the EU funded research project „ArtiVasc 3D“, which is able to print dendritic and porous blood vessels. [K115] WFIRM also developed an integrated tissue organ printer (ITOP) able not only to print tissue but also micro tunnels, which assures the proper flow of nutrition within the printed tissue. [Wal6a] A human-sized ear printed

¹ Technische Universität Berlin, Fakultät VII, Fachgebiet Wirtschafts-, Unternehmens- und Technikrecht, Straße des 17. Juni 135 (Sekt. H 41), 10623 Berlin, j.mihalyi@tu-berlin.de

² Technische Universität Berlin, Fakultät VII, Fachgebiet Wirtschafts-, Unternehmens- und Technikrecht, Straße des 17. Juni 135 (Sekt. H 41), 10623 Berlin, anne-kathrin.mueller@tu-berlin.de

with the ITOP was successfully implanted on mice. Examination of the mice showed that blood vessels had formed between adjacent tissues. [Wa16a]

With scientific progress come new legal challenges. This paper seeks to identify and assess legal problems of bioprinting. To this end, basic terms (see sect. 1.1) will be explained before describing the process of printing tissue (see sect. 1.2) in order to isolate the relevant facts and legal questions for further investigation (see sect. 1.3.).

1.1 Basic Terms

Tissue Engineering is the growing of tissue and organs for the purpose of replacing damaged or diseased tissue. [Ba16 p.1,4; Fa14 p.97] *Bioprinting* is a manufacturing technique combining 3D printing technology with tissue engineering to bring together biological material, e.g. cells or biomolecules, into a 3D structure aimed at fulfilling biological functions. [Ba16 p.4f.; IG16]

The bioprinting process starts like traditional 3D printing by using a *computer-aided design (CAD)-file*. CAD-files are created either with a 3D scan or a specific CAD software program.

Bioinks, or the printing agents used to print tissue, come in two varieties: hydrogels containing cells or a biodegradable, plastic-like gel used for tissue structure. [Wa16a; Fa14 p.98; Ba16 p.5,7]

1.2 The Process of Printing Tissue

CAD-Software uses patient data from computer tomography (CT) or magnetic resonance imaging (MRI) scans to create a model of the desired tissue or organ to be printed. [Wa16a; Ba16 p.6]

For maximum acceptance of the printed tissue, cells must be isolated from the patient's tissue samples. If a patient's tissue is damaged, stem cells can be substituted because they can be grown into any kind of cell. [Fa14 p.97f.] Bioink, containing the patient's cells, is loaded into toner cartridges. The printer, guided by the computer model, utilizes traditional inkjet technology to print a 3D organ prototype layer by layer. [Wa16b]

After printing, the tissue or organ is surgically implanted by a medical team.

1.3 Legal Challenges

Although experts argue that bioprinting is not yet ready for medical approval, it is only a matter of time until the first organ is used in a medical procedure. The use of patient's data to engineer tailor made CAD-templates for printing tissue and organs requires that data protection laws are in place to maintain the patient's confidentiality (see chap.2).

Bioprinting technology also raises product liability questions, especially because health and life of the organ recipient are at stake (see chap.3).

As bioprinting pertains to individual patient's data used for custom 3D printed tissue and organs, the subject of intellectual property rights will not be examined in this paper.

2 Data Protection Law

The Federal Data Protection Act (Bundesdatenschutzgesetz) and the Data Protection Acts of the sixteen German states are laws based on the right to data protection (Recht auf informationelle Selbstbestimmung) which is part of the general right of privacy (Allgemeines Persönlichkeitsrecht) following article 1 (1), article 2 (1) Basic Law of the Federal Republic of Germany (Grundgesetz der Bundesrepublik Deutschland). [Go15 section 1, recital 6] Consequently, the German data protection laws do not protect all kinds of data but only data with reference to a specific person affecting his or her right of privacy. [Ze15 p.1154] Following section 1 (2) Federal Data Protection Act³, only the collection, processing and use of personal data falls within the material scope of the Federal Data Protection Act.⁴ According to the legal definition provided by section 3 (1) Federal Data Protection Act, personal data means any "Einzelangabe" (engl.: information)⁵ concerning the personal or material circumstances of an identified or identifiable individual.

In order to determine if 3D printed tissue and organs are protected under the Federal Data Protection Act (see sect. 2.2), it is necessary to examine to what degree components used for tissue printing are protected (see sect. 2.1).

2.1 The components: data from MRI or CT scans and the patient's cells

The components used for engineering bioprinted organs fall within the material scope of the Federal Data Protection Act if they qualify as "Einzelangabe" concerning the personal or material circumstances of an identified or identifiable individual. The term "Einzelangabe" includes any kind of information. [Dammann, in: Si14, section 3 recital 5] Data generated by MRI or CT scans contain information about surface structure, shape and volume of the patient's tissue which is about to be replicated. These individual properties, used for constructing a computer model, are information and consequently "Einzelangaben" in the sense of the Federal Data Protection Act.

³ Henceforth only the Federal Data Protection Act is referenced since the term *personal data* is used in both federal and state data protection laws.

⁴ Bioprinting technology does not raise specific issues relative to the personal or territorial scope of the Federal Data Protection Act.

⁵ The Federal Data Protection Act uses the wording "Einzelangabe" whereas European legislation prefers the term "Information". Detailed account [Ha15] p.116 ff.

While the classification of data generated by MRI or CT scans as „Einzelangaben“ doesn't raise any problems, cells of human origin could be classified either as information or as information storage device. Classification of information as “Einzelangabe” is independent of its representation or form. [Dammann, in: Si14, section 3 recital 4] Nevertheless, according to *Dammann* procedures and processes of the real world, for example traces as bloodstains, skid marks, fingerprints or hair can't be considered information per se. [Dammann, in: Si14 section 3 recital 5] Information is by nature intellectual rather than material. [Dammann, in: Si14 section 3 recital 5] Therefore procedures and processes of the real world only become information either by symbolic representation or by intentional use as means of communication. [Dammann, in: Si14 section 3 recital 5] According to this view cells are an information storage medium. But by being used with the intention to replicate genetic information and specific properties of the patient's tissue, cell samples become information as well as “Einzelangaben” according to section 3 (1) Federal Data Protection Act. *Haase* argues that the human body and physical substances as original data storage devices should fall as such within the material scope of the Federal Data Protection Act. [Ha15 p.119f, 351] Collecting data storage devices or collecting personal data pose the same threat to confidentiality and personal freedom. [Ha15 p.120] Plus, information usually depends on data storage devices. [Ha15 p.120] Following *Haase*, tissue samples fall as such within the scope of section 1 (2), section 3 (1) Federal Data Protection Act. Since both legal opinions come to the same conclusion that tissue samples fall within the purview of the Federal Data Protection Act, it's immaterial which view is preferable.

A more in depth study needs to be conducted whether MRI or CT scans and tissue samples could also be considered a special category of personal data according to the legal definition in section 3 (9) Federal Data Protection Act. Considerations may include that tissue samples containing diseased cell material most likely won't be used as basis for bioprinting or that for now genetic information is not per se included in the term personal data on health but will be in the General Data Protection Regulation.

2.2 Printed Tissue and Organs

Having determined that the Federal Data Protection Act applies to MRI and CT data as well as the patient's tissue samples, the assessment whether a bioprinted organ also falls within the Act's material scope can start from two different vantage points. On the one hand, the printed organ could be considered personal data since it is the result from processing personal data. On the other hand, the printed organ could qualify as personal data irrespectively of materials used to manufacture it.

The printed organ could represent “new” personal data according to section 3 (1) Federal Data Protection Act because it represents the outcome of processing personal data. If personal data is processed and altered while reference to a specific person is maintained, new categories of personal data are created. [KW05 p.67] In this line, the probability value determined by a scoring system is considered to be personal data. [Go15 section 3

recital 3a] The printed organ combines the genetic and cellular properties with the exact shape and structure provided by the patient's CT or MRI scans. Thus, it is a new set of personal data which also falls within the legal definition provided in section 3 (1) Federal Data Protection Act.

Regardless of its constitutive parts, printed tissue or organs may meet the requirement "Einzelangabe", section 3 (1) Federal Data Protection Act. As mentioned above, the printed organ does not only contain information about the patient's genetic and cellular properties but also about the shape and structure of the patient's organ. The question, whether the printed organ as data storage unit falls within the scope of the Federal Data Protection Act or an act of intentional use is necessary, does not need to be addressed because both views come to the same conclusion. Either the printed organ itself falls within the scope of the Federal Data Protection Act, or the printed organ becomes information by means of being maintained, stored or any other intentional use of it.

Although two different vantage points exist for assessing if printed organs fall within the material scope of the Federal Data Protection Act, both come to the same affirmative conclusion. Hence, bioprinted tissue and organs fall within the purview of German data protection laws.

3 Liability Law

The ensuing analysis of product liability focuses on specific problems raised by bioprinting technology. Product liability laws allocate responsibility to compensate damages caused by defective devices that were put on the market. In Germany, legal basis for a claim are in general either section 823 German Civil Code or section 1 Product Liability Act. The first determines fault-based liability, whereas the second is an example of absolute liability. With regard to health care, specific legislation exists for different sets of products: the Medical Devices Act (Medizinproduktegesetz) and the Medicinal Products Act (Arzneimittelgesetz).

Liability claims based on the violation of the Medical Devices Act (see sect. 3.1.) revert to general product liability laws, whereas the Medicinal Products Act (see sect. 3.2.) provides its own legal basis for a claim relative to a defective medicinal product. Consequently, the critical question is whether a printed organ qualifies as medical device or as medicinal product (see sect. 3.3.) and which legal consequences the proper qualification entails (see sect. 3.4.).

3.1 Medical Devices Act

The Medical Devices Act was shaped by European legislation especially Council

Directive 90/385/EEC⁶ relating to active implantable medical devices and Council Directive 93/42/EEC⁷ concerning medical devices. The Medical Devices Act pursues the two-fold goal of consumer protection and high technical standards for medical devices. [De12 Introduction, p. 46] The Medical Devices Act's Leitmotiv is patient's safety. [*Deutsch*, in: De12, section 1 recital 4] The Act's purpose according to section 1 Medical Devices Act is to ensure safety, suitability and performance of medical devices. The Medical Devices Act further stipulates requirements for putting medical devices on the market. Within the scope of the Medical Devices Act fall medical devices that were put as such on the market, section 2 (1), as well as devices which are considered by law as medical devices, section 2 (2). The Medical Devices Act does not provide its own legal basis for claims. Liability for defective medical devices is generally determined according to contract law, section 823 (1), section 823 (2) and section 831 German Civil Code or section 1 (1) Product Liability Act.⁸ [*Pannenbecker*, in: Te13, section 14 recital 337]

3.2 Medicinal Products Act

On August 1st, 1961 Germany's first Medicinal Product Act came into force. Subjected to significant alterations by European legislation most importantly by Council Directive 65/65/EEC⁹, today's Medicinal Product Act differs considerably from its predecessor. According to section 1 the Medicinal Products Act's purpose is to ensure in particular the quality, efficacy and safety of medicinal products. Accordingly, the Medicinal Products Act is a classic example of consumer protection and general risk prevention legislation. [*Kügel* in: KMH16 Introduction recital 1] It is also a law on animal protection since it guarantees both human beings and animals a proper supply of safe medicinal products. [*Kügel* in: KMH16 Introduction recital 1] The term *medicinal product* laid down in section 2 simultaneously defines the notion itself and the material scope of the law. [*Kügel* in: KMH16 Introduction recital 64] Absolute liability for

⁶ Council Directive of 20 June 1990 on the approximation of the laws of the Member States relating to active implantable medical devices (90/385/EEC), p. 17; reviewed and amended by Directive 2007/47/EC of the European Parliament and of the council of 5 September 2007 amending Council Directive 90/385/EEC on the approximation of the laws of the Member States relating to active implantable medical devices, Council Directive 93/42/EEC concerning medical devices and Directive 98/8/EC concerning the placing of biocidal products on the market.

⁷ Council Directive 93/42/EEC of 14 June 1993 concerning medical devices, OJ No L 169/1 of 1993-07-12, specified via Commission Directive 2005/50/EC, of 11 August 2005 on the reclassification of hip, knee and shoulder joint replacements in the framework of Council Directive 93/42/EEC concerning medical devices; recently reviewed and amended by Directive 2007/47/EC of the European Parliament and of the council of 5 September 2007 amending Council Directive 90/385/EEC, Council Directive 93/42/EEC and Directive 98/8/EC.

⁸ Accordingly, also OLG Saarbrücken, Urt. v. 3.8.2011 – 1 U 316/10, MPR 2011, 156.

⁹ Council Directive 65/65/EEC of 26 January 1965 on the approximation of provisions laid down by Law, Regulation or Administrative Action relating to proprietary medicinal products replaced by Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use, most recently amended by Directive 2012/26/EU of the European Parliament and of the Council of 25 October 2012.

defective medicinal products follows the provisions in section 84 Medicinal Products Act. Legal basis for fault-based liability is section 823 (1), section 823 (2) and section 831 German Civil Code.

3.3 Are printed organs medicinal products or medical devices?

Following section 2 (3) number 7 Medicinal Products Act the term medicinal product shall not apply to medical devices within the meaning of section 3 of the Medical Devices Act. Pursuant section 3, a medical device is any article intended by the manufacturer to be used for human beings for the purpose of treatment or alleviation of disease and which does not achieve its principal intended action in or on the human body by pharmacological, immunological or metabolic means. Section 3 number 15 Medical Devices Act does not postulate production criteria. That means medical devices are not determined according to a specific production process. Rather the notion of medical device is tied to the producer. Thus, power of definition for what is considered to be a medical device lies with the producer. [*Ratzel*, in: De12, section 3 recital 21] A producer in the legal sense is also a person distributing devices manufactured by third parties. However, to reiterate a sentence coined by *Ratzel*: the *power of definition* (Definitions-macht) lies with the producer but not *sovereignty of definition* (Definitions-hoheit). [*Ratzel*, in: De12, section 3 recital 2]. One could argue that bioprinted tissue could be defined as articles used with the objective purpose of treating bodily injuries or organ failure as well as replacing diseased tissue, thus, as being medical devices. But bioprinted tissue is constructed out of patient's cells. Section 2 (5) number 4 states clearly that the Medical Devices Act shall not apply to transplants or tissues or cells of human origin nor to products, incorporating or derived from tissues or cells of human origin. Hence, printed organs do not fall within the purview of the Medical Devices Act.

Since bioprinted organs are built from of human cells, one could argue that the applicable legislation is not the Medicinal Products Act but rather the Transplantation Act. In this line, section 2 (3) number 8 Medicinal Products Act states that the term medicinal product (and thus the material scope of the Medicinal Products Act) shall not apply to organs within the meaning of Section 1a number 1 of the Transplantation Act if they are intended for transplantation to human beings. The Transplantation Act codifies the legal aspects of transplantation medicine, e.g. time of death, agreement to sample the organ etc. Its main purpose is to increase the willingness as well as the transparency of organ donations, section 1 Transplantation Act. According to the legal definition provided by section 1a Transplantation Act, organs, except the skin, are various tissues of the human body which are a functional unit with regard to structure, vascularization and capability to fulfill physiological functions with exception of advanced therapy medicinal products according to section 4 (9) Medicinal Products Act. On the one hand, the argument could be made that only tissue originally grown in a living organism falls within the definition of section 1a. On the other hand, bioprinted organs are built out of the same biomaterial, have the exact same form and (ideally) fulfill the same

physiological functions as human organs. But if bioprinted tissue fits the criteria of an advanced therapy medicinal product, arguments for or against either view can be suspended.

Pursuing the legal definition in section 4 (9) Medicinal Products Act, advanced therapy medicinal products are (among others) tissue engineered products pursuant to Article 2, paragraph 1, letter a and b of Regulation (EC) No. 1394/2007¹⁰. The aforementioned European regulation defines tissue engineered products as products that contain or consist of engineered cells or tissues and which are presented as having properties for, or are used in or administered to human beings with a view to regenerating, repairing or replacing human tissue. A bioprinted organ consists of engineered cells that have been subjected to substantial manipulation so that biological characteristics of the specific organ and its physiological functions are achieved. Plus, they aim to be a viable alternative to organ donations hence to be administered to humans. The wording “tissue engineered product” provided by Article 2 (1) lit. a and b of Regulation (EC) No. 1394/2007 applies also to bioprinted tissue because bioprinting is tissue engineering by means of a 3D printer. Thus, bioprinted organs fall not within the purview of the Transplantation Act. Instead they fall within the material scope of the Medicinal Products Act because bioprinted tissue qualifies as advanced therapy medicinal products in the sense of section 4 (9).

3.4 Legal Consequences

The pharmaceutical entrepreneur who placed a medicinal product on the market which killed a person, or substantially damaged a person’s body or health faces a two-fold liability: either a fault-based liability or an absolute liability. The fault-based liability following section 823 German Civil Code is generally deemed insufficient with regard to medicinal products [*Brock/Stoll* in: KMH16, section 84 recital 2]. As for absolute liability, regulations of the Medicinal Products Act, especially section 84, take precedence over the Product Liability Act’s regulations, section 15 (1) Product Liability Act. Although legal scholars criticized section 15 (1) Product Liability Act, its compatibility with Regulation 85/374/EEC¹¹ was confirmed by the European Court of Justice in 2014¹².

Section 84 Medicinal Products Act does mention the term *defect* as does the Product Liability Act. Nevertheless, the categories to determine if a product is defective, namely a defect of design, production or instruction, are also used by the courts to determine if a medicinal product is faulty. [*Brock/Stoll* in: KMH16 section 84 recital 64] The defect of

¹⁰ Regulation (EC) No 1394/2007 of the European Parliament and the Council of 13 Nov 2007 on advanced therapy medicinal products and amending Directive 2001/83/EC and Regulation (EC) No 726/2004, L 324/121.

¹¹ Council Directive of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products, No L 210/29.

¹² EuGH, 20 Nov 2014 – C-310/13.

design within the purview of the Medicinal Products Act also comprises development risks. [*Brock/Stoll* in: KMH16 section 84 recital 64] The burden of proof rests with the pharmaceutical entrepreneur, section 84 (3) Medicinal Products Act.

As the pharmaceutical entrepreneur is liable for design defects, it should be noted that CAD-files for bioprinting need to maintain general safety standards with respect to design just as CAD-files do in product design. CAD-files should respect the necessary safety standards and should not pose safety-related threats. [*Foerste*, in: FW12, section 24 recital 71] Nevertheless, more stringent requirements apply for products that pose a potential risk for a person's life. In order to define these safety requirements, the product's intended use and specificities of the intended user group have to be taken into account.¹³ IT security of CAD-files will gain increased importance with regard to safety requirements. In principle, the producer or the pharmaceutical entrepreneur isn't liable for third parties intentionally tampering with his or her products. [BK15b p. 87.] Nevertheless, with regard to digital interconnectedness of the production and increased risk of safety breaches, the producer will have to ensure that no digital tampering occurs. Accordingly, the current changes in production processes call for a concept of combined protection for mechanical and digital product safety [BK15a p.1140.]

4 Conclusion

Medical approval to use bioprinted tissue on humans may well be decades away. Despite new challenges existing laws, such as the Federal Data Protection Law and the Medicinal Products Act, provide ample protection. Bioprinted tissue can be considered (containing) personal data and thus falling within the material scope of the Federal Data Protection Act. As bioprinted tissue is an advanced therapy medicinal product, absolute liability of the pharmaceutical entrepreneur is rooted in section 84 Medicinal Products Act. However, the science of Tissue Engineering and Bioprinting will also call on to national and European legislators to keep pace with technological advancement with regard to quality control and safety standards. As of the time of this writing, the German Bundesrat approved the Government's draft bill 18/8580 implementing Commission Directive (EU) 2015/565 and 2015/566 of 8 April 2015 which regard certain technical requirements for the coding of human tissues and cells as well as procedures for verifying the equivalent standards of quality and safety of imported tissues and cells.

References

- [Ba16] Baier, Daniel, Konzeption und Prototypenbau eines Zweikomponenten-Druckkopfs mit wechselbaren Dual-Extrudern für das 3D-Bioprinting im Bereich des Tissue Engineering, Studienabschlussarbeit, 08. Januar 2016, München.

¹³ Confirming EuGH, Urt. v. 5.3.2015 – C- 503/13.

- [Bi10] Binder, K. W.; Zhao, W., et al.: *In situ bioprinting of the skin for burns*. In: Journal of the American College of Surgeons, 2010, (3), S. 76.
- [BK15a] Bräutigam, Peter; Klindt, Thomas, NJW 2015, 1137ff.
- [BK15b] Bräutigam, Peter; Klindt, Thomas, Digitalisierte Wirtschaft/ Industrie 4.0, Ein Gutachten der Noerr LLP im Auftrag des BDI zur rechtlichen Situation, zum Handlungsbedarf und zu ersten Lösungsansätzen, November 2015.
- [De12] Deutsch/Lippert/Ratzel/Tag, Kommentar zur Medizinproduktegesetz, 2. Aufl. 2012.
- [Eu16] Eurotransplant International Foundation, http://www.eurotransplant.org/cms/index.php?page=pat_germany, Stand: 06.04.2016.
- [Fa14] Fastermann, Petra, 3D-Drucken, Wie die generative Fertigungstechnik funktioniert, 1. Auflage, Springer-Verlag, Berlin, 2014.
- [FW12] Foerste/Graf von Westphalen, Produkthaftungshandbuch, 3. Aufl. 2012
- [Go15] Gola, Peter; Klug, Christop; Körffer, Barbara; Schomerus, Rudolf, BDSG, Bundesdatenschutzgesetz, Kommentar, 12. Auflage 2015.
- [Ha15] Haase, Martin Sebastian, Datenschutzrechtliche Fragen des Personenbezugs, Eine Untersuchung des sachlichen Anwendungsbereiches des deutschen Datenschutzrechts und seiner europarechtlichen Bezüge, 1.Auflage, Mohr Siebeck, Tübingen, 2015.
- [IG16] IGB, Das Fraunhofer-Institut für Grenzflächen- und Bioverfahrenstechnik <http://www.igb.fraunhofer.de/de/kompetenzen/grenzflaechentechnik/biomaterialien/bioprinting.html>, Stand: 07.04.2016.
- [KMH16] Kügel, Wilfried; Müller, Rolf-Georg; Hofmann, Hans-Peter, Arzneimittelgesetz Kommentar, 2. Aufl. 2016.
- [KW05] Kamp, Meike, Weichert, Thilo, Scoringsysteme zur Beurteilung der Kreditwürdigkeit – Chancen und Risiken für Verbraucher, Forschungsprojekt des Unabhängigen Landeszentrum für Datenschutz Schleswig-Holstein, 2005.
- [KI15] Klempert, Oliver, Schilddrüse aus dem 3-D-Drucker, vdi-nachrichten Ausgabe 47, 20. November 2015, <http://www.vdi-nachrichten.com/Technik-Wirtschaft/Schilddruese-3-D-Drucker>, Abgerufen: 07.04.2016.
- [MP15] Müller/Pramann, InTeR 2015, 148ff.
- [Te13] Terbille/Clausen/Schroeder-Printzen, Münchner Anwaltshandbuch medizinrecht, 2.Auflage 2013.
- [Wa16a] Wake Forest Baptist Medical Center, Scientists prove feasibility of 'printing' replacement tissue, 15. Februar 2016, http://www.eurekalert.org/pub_releases/2016-02/wfbm-spf021116.php, Abgerufen: 07.04.2016.
- [Wa16b] Wake Forest Institute For Regenerative Medicine, <http://www.wakehealth.edu/Research/WFIRM/Research/Engineering-A-Kidney.htm>, Abgerufen: 11.04.2016.

- [Ze15] Zech, Herbert, "Industrie 4.0" - Rechtsrahmen für eine Datenwirtschaft im digitalen Binnenmarkt, GRUR 2015, 1151ff.

Infrastructuring & Digital Fabrication

Oliver Stickel¹ and Volkmar Pipek²

Abstract: We try to provide an insight into digital fabrication from the point of view of *infrastructuring* as a holistic research and practice lens from Computer Supported Collaborative Work. Our position paper serves as a discussion contribution for the workshop "Additive Fabrication/3D Printing – Technology, Impact and Chances" at *Informatik 2016*.

Keywords: position paper; digital fabrication; infrastructuring; cscw

1 Introduction and Background

In computer supported collaborative work (CSCW), *infrastructuring* is a school of thought which identifies the division of the domains of *design* and *use* as a critical factor in the development of technological systems. Instead, widening one's view to encompass *socio-technical* systems is proposed; systems in which people, organisations and communities actively design, develop, engineer, use, apply, hack or are otherwise engage in "infrastructuring" activities without clear role boundaries. These activities take place in specific contexts which sometimes overlap. They lead to the formation of localized practices and often comprise both digital and physical artefacts and places as important elements. At certain points the activities converge which makes the action, use development and other activities of the various actors, which were formerly not (consciously) perceived, visible. *Infrastructuring* can also be understood as a methodological framework for development processes of socio-technical systems, whereby key findings in this area include i.a. criticism of heavyweight processes (steered from the top down) and development models; the importance of design and development as *in-situ* activities; as well as the diversity of reciprocal references of socio-technical infrastructures to social practices, requirements and values.

The background of this paper consists of the *infrastructuring* activities of numerous actors, organisations and communities scattered around the world, with whom the authors cooperate and interact within the framework both of ethnographically motivated fieldwork as well as participation in the sense of *research through design* and *action research*, including the incubation and development of Fab Lab Siegen as a local community, fabrication and research infrastructure (including close involvement in the organisational practices of the public sector, the development of new inter-facultative teaching concepts and collaboration with regional medium-sized enterprises, etc.). Our argumentation is also based on data and findings from discourses around the CHI- and (E)CSCW-conferences;

¹ University of Siegen, oliver.stickel@uni-siegen.de

² University of Siegen, volkmar.pipek@uni-siegen.de

the summit on *Rethinking technology innovation: Factories, Fabrication & Design Research*³; and collaboration and research work over several years in the *global innovation gathering*⁴, a global network of community and fabrication hub managers, entrepreneurs from the socio-technical area and hackers, as well as (N)GOs. In the following, we would like to sketch some of the places which we believe digital fabrication informs socio-technical infrastructures. In doing so, there are a number of reasons which currently rather lead us to understand additive fabrication to be a very topical (and legitimately very promising) subarea and the driving force behind digital fabrication itself.

2 Fabrication Infrastructures

2.1 Plurality

Digital fabrication is thought, developed and used in numerous domains which have long exceeded the boundaries of traditional production. In the subset of additive fabrication, the bandwidth ranges from printing of components within interdisciplinary engineering contexts, e.g. in airplane construction or medical applications, through the fabrication of commodities from chocolates to prostheses up to completely open activities in hackspaces, fab labs and similar hubs. In such different areas, related practices are often developed and might benefit from better collaboration. One case example of this is the professional (vocational) training and work with (CNC-) milling machines in comparison to the approach taken and tools employed in nonprofessional communities which engage in additive fabrication (such as the *Maker*-culture). In many aspects, both share aims, values and tools – but are separated by different vocabularies, practices and constraints. The means inefficiency at best, friction and (resource-)conflicts at worse.

2.2 Politics, Economy and Governance

A growing but sometimes diffuse cloud of actors and communities is engaging in innovation and (digital) fabrication domains – which adds to questions about concepts, control(lability), and value of work and resources – and, ultimately, about power structures - to quote the call for papers for this workshop: "*Due to the spread of additive fabrication, it currently seems possible to bring [...] production back to Europe again.*". Such questions can be analysed and discussed from very different perspectives. Some proclaim the next (industrial and/or) social revolution; post-Marxist, some understand them as the appropriation of (digital-physical) means of production. Yet from a capitalistic growth perspective, immense potential, e.g. with a view to mass individualisation, is also attributed to digital, distributed fabrication – especially additive fabrication.

³ See <https://hci.sbg.ac.at/sites/ffdr/> - proceedings currently in production.

⁴ See <http://globalinnovationgathering.com/>

2.3 Globality and Locality

One of the central aspects of digital fabrication is the prospect to get *from bits to atoms* (and back again), anywhere, anytime. This goes hand in hand with concepts of mobile work, the understanding of the term "work" itself, collaborative tools and distributed infrastructures as well as the sharing of knowledge and resources. At the same time, it can be noticed that particularly new, creative opportunities for use, business models and methods often emerge from (hyper-)local places (hubs) such as co-working spaces, labs, think tanks, maker and hacker spaces or fab labs. Global networks which adopt open source and cross-border collaboration on equal terms allow such local hubs to be active at a global level as a matter of course. This position may help to develop and sustain innovative, socio-technically grounded business models with new value chains.

2.4 Education

There is significant market movement and discussion about the role of digital fabrication for education. Especially simple FDM-3d printers and microelectronics are being used more and more in educational settings. They seem to show significant potential, especially when combined with project-based and learner-focussed learning theories such as constructionism - as illustrated by a considerable bandwidth of well published research in this subject area. It should be underlined that exactly these methodical aspects in connection with the tangibility and concreteness of digital-physically (and ideally collaboratively) produced artefacts as well as the ever more differentiated range of tools and machines available also seem to support a more gender-equitable and inclusive approach to socio-technical educational topics. Examples for resulting changes in educational systems include Great Britain, where curricula now include socio-technical aspects and every pupil gets a single-board-computer to hack and make things with or the USA where the Maker culture seems set on bringing back tangible, fabrication-related aspects back into educational practices after the politically ordained end of shop classes in schools. Finally, it is worth looking at the domestic scene: Digital fabrication infrastructures can help people to think and understand appropriate design and fabrication as a global, collaborative process with enough room for local autonomy, participation and individuality.

At this point, critical voices come into play, speaking out against the inclusion of interdisciplinary aspects of fabrication and making in broader educational contexts and warn against e.g. overload, the disintegration of competence profiles, technical determinism, faith in technology and many other valid criticisms. Very serious attention should also be paid to these voices but on the basis of the state of art and as well as our own experiences⁵ over several years utilizing digital fabrication in educational contexts,

⁵ i.a. the development and implementation of experimental educational events on digital fabrication in tertiary education, projects in developmental contexts as well as the development and implementation of introductory workshops in 3D-printing in which we have worked together over the years with hundreds of participants of all ages and in a variety of contexts.

we believe open access to digital fabrication and community innovation during youth and education may not only be a tool for a very broad and interdisciplinary spectrum of concrete educational projects but can also be understood as (creative) media including benefits for self-conception and self-expression, communication, collaboration and quite a few other educational challenges.

2.5 Closing Remarks

We were able to thematize only a few infrastructure aspects of digital fabrication. However, we do hope that these points will stimulate discussion at the workshop and can be expanded upon collaboratively. Infrastructuring as a very broad lens on (global) innovation and fabrication necessitates – not least on the part of researchers – new, interdisciplinary collaboration (particularly *in situ*) as well as taking seriously an unexpected variety of actors and hubs in formerly professional domains.

Der Arbeitsplatz der Zukunft

Perspektiven des Arbeitsplatzes der Zukunft

Jan Haase¹, Gerhard Leitner², Daniel Versick³ und Raphael Zender⁴

Abstract: Der Arbeitsplatz der Zukunft ist ein breites und kontroversiell diskutiertes Thema, dessen sich der Workshop gleichen Titels annimmt. Ein bereits sehr weit gefasster Call for Contributions resultierte in einer Vielfalt von Einreichungen, deren Ausrichtungen und Perspektiven zum Thema unterschiedlicher nicht sein könnten. Einige der als Ergebnis des Begutachtungsprozesses akzeptierten Arbeiten beschäftigen sich mit der Arbeitsumgebung als solches, der Optimierung von Gebäudeinfrastruktur oder der Möglichkeiten des arbeitenden Menschen in die Abläufe des Gebäudes einzugreifen. Den kritischen Aspekten der für eine Automatisierung erforderlichen Sammlung von Aktivitäts- und Nutzungsdaten und den Möglichkeiten sich einer Überwachung zu entziehen widmet sich ein anderer Beitrag. Die Sensibilität des eigenen Zuhauses ist in diesem Zusammenhang von besonderer Bedeutung. Einer der Workshopbeiträge beleuchtet die Möglichkeiten, Laien bei der Konfiguration von Smart Home Systemen zu unterstützen und die Privatsphäre insofern besser zu wahren, als durch die Unterstützung durch Smart Glasses qualifiziertes Fachpersonal nicht mehr vor Ort präsent sein muss, sondern entsprechende Remote-Unterstützung bieten könnte. Smart Glasses sind auch Thema eines weiteren Beitrages, allerdings im industriellen Kontext. Es werden Möglichkeiten aufgezeigt, wie Datenbrillen im Training von Mitarbeitern auf Produktionslinien durchgeführt werden können. Der Einsatz von neuen Technologien wie Smart Glasses ist aber kaum friktionsfrei. Im Vorfeld ihres Einsatzes müssen daher entsprechende Aktivitäten gesetzt werden, sie zu evaluieren und auf die Nutzerbedürfnisse anzupassen. Aufgrund ihrer Neuartigkeit scheinen jedoch konventionelle Vorgehensmodelle und Richtlinien zur Überprüfung der Gebrauchstauglichkeit nicht geeignet. Ein Workshopbeitrag widmet sich daher der Entwicklung eines auf die Besonderheiten dieser neuen Kategorie von Endgeräten angepassten Usability-Frameworks.

Keywords: Digitaler Arbeitsplatz, Gebäudeinfrastruktur, Sensornetzwerke, Privatsphäre, Teleworking, Usability, Wearables.

1 Einleitung

Der digitale Arbeitsplatz der Zukunft soll nicht nur das Arbeiten erleichtern und die Produktivität der Mitarbeiter erhöhen, sondern auch maximale Flexibilität und Kostenreduktion für das Unternehmen und den CIO mit sich bringen.

¹ Institute of Computer Engineering, Universität zu Lübeck, Ratzeburger Allee 160, D-23562 Lübeck, haase@iti.uni-luebeck.de

² Institut für Informatik-Systeme, Alpen-Adria-Universität Klagenfurt, Universitätsstraße 65-67, A-9020 Klagenfurt, gerhard.leitner@aau.at

³ Institut für Informatik, Universität Rostock, Albert-Einstein-Straße 22, D-18059 Rostock, daniel@versick.de

⁴ Institut für Informatik und Computational Science, Universität Potsdam, August-Bebel-Str. 89, D-14482 Potsdam, zender@uni-potsdam.de

Der klassische Arbeitsplatz mit seiner i. d. R. eher fixen Struktur kann aktuelle Trends wie Consumerization und Mobilität nur teilweise bedienen, so dass entscheidende strukturelle Änderungen notwendig sind, um den Arbeitsplatz der Zukunft zu realisieren. Der Trend Consumerization bezeichnet die verstärkte Nutzung privater Endgeräte, wie bspw. Smartphones oder Tablets für Unternehmenszwecke. Diese Entwicklung bringt für den Anwender zahlreiche Vereinfachungen, da aus dem privaten Umfeld bekannte Strukturen nun auch im Unternehmen eingesetzt werden. Der Arbeitgeber kann bei Unterstützung der Nutzung privater Geräte einen echten Nutzungsvorteil und damit für bessere Motivation auf Seiten der Arbeitnehmer sorgen und u. U. sogar die Einarbeitung in neue Technologien einsparen, da der Mitarbeiter diese bereits aus dem privaten Umfeld kennt. In vielen Unternehmensumfeldern ist z. B. Dropbox als Austauschmedium für Daten oder WhatsApp als Nachrichtenkanal der Mitarbeiter nicht mehr wegzudenken, wengleich dies für die Unternehmen zahlreiche technische und rechtliche Probleme mit sich bringt. Teilweise geben Strategien für Bring-Your-Own-Device (BYOD) Antworten auf diese Probleme - diese sind aber in aller Regel keine endgültigen Lösungen. Consumerization ist weiterhin die Ursache der verstärkten Nutzung mobiler Endgeräte, die der Nutzer problemlos zwischen seinem Heim und der Arbeit transportieren kann. Mobiltelefone, Tablets und Laptops als mobile Geräte ermöglichen wahlweise eine Arbeit zu Hause oder im Unternehmen und sind damit eine Antwort auf den Wunsch der Mitarbeiter nach mehr Flexibilität bei Arbeitszeit und Arbeitsplatz.

Mitarbeiter wollen in der Lage sein, in Heimarbeit Ihrer Tätigkeit nachzugehen oder unterwegs während einer Dienstreise zu arbeiten. Der ehemals feste Arbeitsplatz, den das Unternehmen seinen Mitarbeitern zur Verfügung stellt, wandelt sich zu einem Arbeitsplatz bestehend aus zahlreichen mobilen Endgeräten, von denen zumindest ein Teil aus dem Privatbesitz des Mitarbeiters stammt.

Da mobile Arbeitsgeräte in Gewicht und Größe minimiert werden müssen, gibt es eine Reihe von Ressourcenbeschränkungen, die das klassisch feste Arbeitsplatzrechensystem nicht hatte. Die Größenreduktion der Geräte führt zu kleineren Displays mit besonderen Anforderungen an die Darstellung. Außerdem führt sie zu einer reduzierten Batteriekapazität, die ihrerseits Beschränkungen bei der Laufzeit der Systeme und Leistungsfähigkeit verursacht. Das hat wesentlichen Einfluss auf die Architektur des zukünftigen Arbeitsplatzes.

1.1 Themenbereiche

Der Workshop „Arbeitsplatz der Zukunft“ der GI/ITG-Fachgruppe APS beleuchtet diese neuen Herausforderungen des digitalen Arbeitsplatzes der Zukunft und wird daher u. a. Inhalte ausfolgenden Themenschwerpunkten diskutieren:

- Hard- und Software-Architekturen zur effizienten Kollaboration von Endgeräten und Geräteensembles (u. a. Middleware-Architekturen) des digitalen Arbeitsplatzes der Zukunft
- Kommunikationsstrukturen zur effizienten Kommunikation der Arbeitsplatzgeräte untereinander sowie mit Backend-Strukturen (wie bspw. Cloud-Systemen)
- Cloud-Infrastrukturen und Parallelrechnersysteme, die leistungsfähige Backend-Strukturen für den Arbeitsplatz der Zukunft bilden, energieeffiziente Geräte, da die Ressourcenbeschränktheit mobiler Endgeräte eine stetige Steigerung der Energieeffizienz erfordert
- Strategien für Bring-Your-Own-Device (BYOD)
- Usability-Prinzipien am Arbeitsplatz der Zukunft

2 Workshop Komitee

Wir könnten ein hochkarätiges Gremium für den Workshop gewinnen, sowohl aus Forschung, als auch aus der Industrie. Wir bedanken uns an dieser Stelle bei den folgenden Personen, die als Mitglieder des Programmkomitees und Reviewer fungierten.

- Jürgen Brehm (Uni Hannover)
- Hauke Coltzau (Fernuni Hagen)
- Jan Haase (HSU Hamburg)
- Martin Hoffmann (FH Bielefeld)
- Wolfgang Karl (KIT)
- Andreas Kohne (Materna GmbH)
- Gerhard Leitner (AAU Klagenfurt)
- Ulrike Lucke (Uni Potsdam)
- Dominik Meyer (HSU Hamburg)
- Hayk Shoukourian (LRZ München)
- Djamshid Tavangarian (Uni Rostock)
- Daniel Versick (Uni Rostock)
- Raphael Zender (Uni Potsdam)

3 Einreichungen und akzeptierte Beiträge

Die Einreichungen wurden im Zuge eines single-blind Begutachtungsprozesses von mindestens 3 Gutachtern bewertet und kommentiert. Dies führte schließlich zur Auswahl von 6 Arbeiten für den Workshop. Bei den eingereichten Arbeiten handelt es sich bei allen um Langfassungen technisch-wissenschaftlicher Arbeiten, zu unterschiedlichen, dem Call entsprechenden Themenbereichen.

Mit der Arbeitsumgebung als solche beschäftigen sich die Beiträge von Haase und Meyer sowie Schröder et al. In ihrem in englischer Sprache verfassten Beitrag beleuchten Haase und Meyer die technologischen Gegebenheiten der Gebäudeinfrastruktur und die Möglichkeiten einer Optimierung von Gebäudetechnik. Viele der heutigen gewerblich genutzten Gebäude basieren auf 40 Jahre alter Technik. Aufgrund limitierten Budgets haben vor allem öffentliche Körperschaften nicht die Möglichkeit in die Erneuerung der Infrastruktur zu investieren. Dennoch könnten moderne Gebäudetechnologien die Sicherheit, den Komfort, die Energiebilanz und andere Aspekte erheblich verbessern. Diesbezügliche Trends sind z.B. moderne Aufzugskonzepte, die sowohl den Durchsatz an Fahrgästen erhöhen, zur Optimierung der Halte und Verbesserung der Energieeffizienz beitragen und zusätzlich in der Wahrnehmung der Benutzer eine erhöhte Effizienz aufweisen und somit zum Wohlbefinden der Nutzerinnen beitragen könnten. Dasselbe gilt auch für Systeme, die Nutzerverhalten analysieren und darauf aufbauend die Steuerung elektrischer Verbraucher automatisieren können oder adaptive Leitsysteme, die zum Beispiel Besuchern in Gebäuden die Wege weisen können. Die Systeme basieren auf der Verwendung unterschiedlicher Sensoren und sind adaptiv und lernend.

Einer anderen Perspektive auf die Arbeitsumgebung und das Wohlbefinden am Arbeitsplatz nehmen sich Schröder et al an. Die Autorin und Autoren beschäftigen sich mit dem Einfluss der Arbeitsumgebung auf das Wohlbefinden und dessen Einfluss auf die Produktivität. Neben der Berücksichtigung der Forschung des Bereiches „Indoor Environment Quality“ stützen die Autoren ihre Arbeiten auf der Beobachtung der Menschen selbst, ihrer Emotionen und ihrer Aktivitäten. Auf Basis von vielfältiger Sensorik, die es ermöglicht relevante Parameter zu erfassen, kann das Zusammenspiel von Umweltfaktoren und die Interaktion zwischen Mensch und Umwelt erforscht werden. Außerdem von Bedeutung sind in der Arbeit die Einflussmöglichkeiten des Menschen auf seine Arbeitsumgebung, deren Berücksichtigung und deren Einfluss auf die Arbeitszufriedenheit. Für die technische Umsetzung schlagen die Autoren Multiagentensysteme vor, welche die Komplexität auf Agenten verteilen, aber trotzdem eine übergeordnete Instanz besitzen, welche ein menschliches Eingreifen ohne Umwege ermöglicht.

Die in den vorher genannten Arbeiten Möglichkeiten des Nutzer-Trackings und der umfassenden Überwachung durch Sensoren sind bei allen Vorteilen, die sie mit sich bringen, auch kritisch zu beurteilen. Auch im Arbeitskontext ist das Gefühl des ständigen Überwachtseins der Leistung und dem Wohlbefinden nicht zuträglich. Coltau diskutiert in seinem Beitrag Möglichkeiten von Nutzern in ihrer Rolle als Privatpersonen oder Arbeitnehmer, sich der verdeckten Profilbildung, dem Fingerprinting und Tracking zu entziehen bzw. diese Form der Überwachung aktiv einzuschränken. Am Beispiel der Standardtechnologie jQuery werden mögliche Formen der Implementierung diskutiert.

Die Arbeit von Fercher und Leitner bewegt sich im Hinblick auf Privatheit und Datenschutz in einen noch sensibleren Bereich, nämlich das eigene Zuhause und dessen Tauglichkeit als Arbeitsplatz der Zukunft.

Die Autoren zeigen vor allem die Problematik für Laien auf, sich das eigene Zuhause als digitalen Arbeitsplatz zu etablieren und in Stand zu halten. Das Konzept des Smart Homes könnte die technische Basis dafür bilden, birgt jedoch angesichts inhärenter Komplexität sehr viele Barrieren, die durchschnittlich technisch versierte Personen nur schwer eigenständig bewältigen können. In der Arbeit werden Möglichkeiten aufgezeigt, wie unter Zuhilfenahme von Augmented Reality in Form von Smart Glasses sowohl die Eigenständigkeit von Laien in der Konfiguration von intelligenten Umgebungen unterstützt, als auch die Einbindung von qualifiziertem Fachpersonal über die Entfernung hinweg erleichtert bzw. verbessert werden kann.

Smart Glasses sind auch Thema der Arbeit von Brauns et al., jedoch in einem industriellen Kontext. Sie beschäftigt sich mit den Möglichkeiten Datenbrillen zur Unterstützung des Trainings von Mitarbeitern einzusetzen, die direkt in den Produktionsprozess eingebunden sind. Der Beitrag beschreibt ein System für das Training von Montagetätigkeiten. Der Einsatz von Datenbrillen ermöglicht es, den am Training teilnehmenden Mitarbeitern benötigte Informationen direkt in ihr Blickfeld einzublenden. Das System erlaubt die Durchführung des Trainings mit mehreren Trainern und Trainingsteilnehmern und zeichnet sich durch hohe Flexibilität aus. Die Autoren stellen schließlich die Ergebnisse der Evaluierung einer frühen Version des Systems vor, bei der die Tauglichkeit des Trainingskonzepts sowie die Beanspruchung der Probanden getestet wurde.

Mit Aspekten der Evaluierung von Wearables im Allgemeinen und Smart Glasses im Besonderen, beschäftigt sich die Arbeit von Zobel et al. Die zunehmende Verbreitung mobiler Endgeräte am Arbeitsplatz inkludiert auch Smart Glasses und andere Wearables. Diese sind einfach auf unterschiedliche Anwendungsfälle adaptierbar und haben den großen Vorteil gegenüber anderen mobilen Geräten, dass sie die Bewegungs- und Handlungsfreiheit kaum beeinträchtigen. Für eine effiziente und zielorientierte Einführung solcher Technologien ist jedoch deren Gebrauchstauglichkeit zu berücksichtigen. Gängige Usability-Richtlinien sind nicht auf die speziellen Anforderungen von, z. B. Smart Glasses angepasst. Der Beitrag stellt daher ein adaptiertes Framework vor, das für Forschung und Praxis Möglichkeiten zur Untersuchung der Usability von Smart Glasses und ähnlichen Wearables bietet.

Das Spektrum der Arbeiten verspricht sowohl interessante Präsentationen, als auch anschließende Diskussionen mit – wie die Anknüpfungspunkte der einzelnen Arbeiten untereinander verdeutlichen – einem hohem Potenzial für gemeinsame Aktivitäten der Teilnehmerinnen und Teilnehmer in der Zukunft.

4 Workshop Programm

Der Workshop findet am Tag der Wirtschaft der Informatik 2016 statt. Diesem Kontext wird in verschiedener Hinsicht Rechnung getragen.

Der Workshop beginnt mit einer Keynote von Andreas Kohne mit dem Titel „Arbeitsplatz der Zukunft, Vision und Realität“, der zu Beginn des Workshops einen starken Impuls aus der Industrie-Perspektive setzt.

Inhalt der Keynote „Arbeitsplatz der Zukunft, Vision und Realität“

Der Vortrag stellt aus Sicht der Wirtschaft vor, welche Visionen zum Arbeitsplatz der Zukunft existieren und welche Aspekte daraus praktische Relevanz haben. Dabei wird erläutert, welche Technologien, Methoden und Verhaltensweisen sich bereits durchgesetzt haben und welche es noch werden.

Der Keynote-Speaker

Andreas Kohne arbeitet als Assistent der Geschäftsleitung bei Materna und promoviert derzeit an der TU Dortmund. Zuvor arbeitete er drei Jahre als Business Development Manager für Materna. Sein Arbeitsgebiet umfasste die Schwerpunkte Virtualisierung, Cloud-Computing-Beratung und E-Learning. Vom Einstieg bei Materna im Jahr 2008 bis 2012 war Kohne als technischer Consultant im Bereich Infrastruktur Management tätig. Zuvor war er im Forschungs- und Entwicklungslabor von IBM in Böblingen tätig. Andreas Kohne studierte zwischen 2003 und 2008 Kern-Informatik an der TU Dortmund mit dem Nebenfach BWL. Weiterhin ist er Fachbuchautor und Vorstandsmitglied der Alumni der Informatik an der TU Dortmund.

Nach der eigenen Keynote pausiert der Workshop um den Teilnehmern die Möglichkeit zu geben der Keynote von Dr. Sabine Herlitschka Vorstandsvorsitzende & CTO Infineon Technologies Austria AG beizuwohnen und weitere Impulse aus industrieller Perspektive zu erhalten.

Der Workshop startet nach der Mittagspause dann mit der Präsentation und Diskussion der Workshop-Beiträge.

Auf der nächsten Seite befindet sich das Programm in einer Übersicht.

WORKSHOP: Der Arbeitsplatz der Zukunft

09:00-10:30 Keynote Session, Session Chair: Jan Haase

Andreas Kohne, Materna

Arbeitsplatz der Zukunft, Vision oder Realität

11:00-12:00 Keynote Tag der Wirtschaft

Sabine Herlitschka, Infineon

Industrie 4.0 und zukünftige Arbeitswelten: Praxis und Vision von Infineon Austria

12:00 – 14:00 Mittagspause

14:00-15:30 Session 1, Session Chair: Raphael Zender

Anton Fercher, Gerhard Leitner, Alpen-Adria-Universität Klagenfurt

Das eigene Zuhause als Arbeitsumgebung - Augmented Reality zur Konfiguration, Adaptierung und Steuerung im Smart Home

Jan Haase, Dominik Meyer, Universität Lübeck, Helmut-Schmidt-Universität der Bundeswehr Hamburg

New trends in building automation for offices of the future

Hauke Coltzau, Fernuniversität Hagen

Integration von User Privacy Mechanismen in clientseitige Standardbibliotheken

16:00-17:30 Session 2, Session Chair: Gerhard Leitner

Lilian Schröder, Aljoscha Pörtner, Matthias König, Martin Hoffmann,
Fachhochschule Bielefeld

Überlegungen zu Multi-Sensor-Aktor-Systemen und Kontrolle in intelligenten Arbeitsumgebungen

Sarah Brauns, Tobias Käfer, Dirk Koriath, Andreas Harth Volkswagen AG, Karlsruher Institut für Technologie

Individualisiertes Gruppentraining mit Datenbrillen für die Produktion

Benedikt Zobel, Lisa Berkemeier, Sebastian Werning, Oliver Thomas,
Universität und Hochschule Osnabrück

Augmented Reality am Arbeitsplatz der Zukunft: Ein Usability-Framework für Smart Glasses

5 Danksagung

Dieser Workshop wird von der Fachgruppe der GI und ITG zum Thema "Appliances, Personal and Smart End Devices (APS)" durchgeführt. Diese Fachgruppe nannte sich früher "Arbeitsplatzrechner und Personal-Computer (APS+PC)" und wurde Anfang 2016 umbenannt, um der neuen Ausrichtung zu entsprechen. Interessenten können sich unter <http://fg-aps.gi.de/> über die Arbeit der Fachgruppe informieren.

Das eigene Zuhause als Arbeitsumgebung - Augmented Reality zur Konfiguration und Steuerung im Smart Home

Anton Josef Fercher¹ und Gerhard Leitner¹

Abstract: Durch die Omnipräsenz digitaler Endgeräte, speziell mobiler Geräte wie Smartphones verschmilzt berufliches und privates Leben zusehends, denn sowohl auf softwaretechnischer Ebene (Profile, Nutzerdaten) als auch betreffend die Hardware (Laptop, Tablet) werden multiple Identitäten (eine Person ist, zum Beispiel, Mitarbeiter und gleichzeitig Familienvater) durch singuläre, eventuell mit Kontextinformation versehene Identitäten abgelöst (Skype mit KollegInnen an anderen Firmenstandorten, Skype mit Familienmitgliedern, im gleichen Profil). Die vorliegende Arbeit setzt sich mit jenen Aspekten auseinander, die das eigene Zuhause in seiner Rolle als Arbeitsplatz (Homeoffice) und dessen Auswirkungen sowohl auf Beruf als auch auf Privatleben betreffen. Ein Schwerpunkt wird dabei auf die Herausforderungen gelegt, die eine arbeitstechnische Integration eines Arbeitsplatzes der Zukunft in ein Haus der Zukunft (Smart Home) mit sich bringen können. Im Gegensatz zum Arbeitsplatz im Unternehmen weist der Arbeitsplatz zu Hause zwar größere Freiheitsgrade in seiner Gestaltbarkeit auf, birgt jedoch auch höhere Herausforderungen da, typischer Weise, keine ServicetechnikerInnen, SystemadministratorInnen oder NetzwerkspezialistInnen vor Ort sind, um die erforderliche infrastrukturelle Basis für das Arbeiten „von Zuhause“ zu etablieren und am Laufen zu halten. Aktuell obliegt diese Art der Vorbereitung der technisch versierten bzw. einschlägig ausgebildeten Person selbst. Welche Möglichkeiten aber haben technisch durchschnittlich begabte bzw. interessierte Personen aus Berufsgruppen, für welche die Arbeit von Zuhause ebenfalls von Interesse sein könnte? Die Arbeit zeigt am Beispiel der Augmented Reality Wege auf, es auch diesen Personen zu ermöglichen sich zu einem gewissen Grad smarte Arbeitsumgebungen selbst zu etablieren.

Keywords: Telearbeit, Arbeit von Zuhause, Smart Home, Infrastruktur, Intelligente Umgebungen und Smart Homes, Augmented Reality.

1 Einleitung

Die stetig zunehmende Anzahl von in unserem Arbeits- und Alltagsumfeld eingebetteten computergestützten Komponenten formt durch ihre systematische Vernetzung so genannte intelligente Umgebungen. Diese versprechen erhöhten Komfort, bessere Unterstützung bei, zum Beispiel, täglichen Aufgaben, Sicherheit und Energieeffizienz. Bei allen möglichen Vorteilen erhöht Vernetzung jedoch auch die Komplexität von teilweise schon individuell komplizierter Technik.

Intelligente Umgebungen machen auch vor dem privaten Wohnumfeld nicht halt. Die

¹ Alpen-Adria-Universität Klagenfurt, Institut für Informatik-Systeme, Universitätsstraße 65-67, A-9020, Klagenfurt, {antonjosef.fercher, gerhard.leitner}@aau.at

Nutzung des eigenen Heims als Arbeitsplatz birgt in diesem Zusammenhang spezielle Anforderungen, derer sich diese Arbeit fokussiert widmet. Mit sich ändernden Lebensweisen ist ein zunehmender Variantenreichtum von Endgeräten und Anwendungen zu beobachten, sichtbar zum Beispiel, an der Vielfalt von Laptops, Tablets und Smartphones, welche über Lebensbereiche (Arbeit, Familie, Freizeit) hinweg genutzt werden. Auch damit schwappt die Notwendigkeit des Vorhandenseins entsprechender Infrastruktur unter anderem in Gestalt von Umgebungszintelligenz, von anderen Lebensbereichen wie dem Arbeitsplatz, auch ins private Umfeld über.

Dieses stellt jedoch mehrere besondere Herausforderungen dar, denn, wie [Hi99] aufzeigt, sind die Basistechnologien für diese Systeme in anderen Kontexten (wie industriell genutzten oder öffentlichen Gebäuden) entstanden und müssen sich auf die besonderen Umstände des Zuhauses erst anpassen.

Eine der Herausforderungen ist jene, die je nach Wohnsituation sehr unterschiedlichen strukturelle Gegebenheiten, Bedürfnisse und Gewohnheiten der Personen entsprechend zu berücksichtigen. Diesbezüglich besteht auch das Problem, dass im Gegensatz zum betrieblichen oder öffentlichen Sektor Standards bzw. verallgemeinerbare Abläufe fehlen. Auch wenn gewisse tägliche Routinen über viele Haushalte hinweg auf den ersten Blick ähnlich erscheinen, so werden diese von den BewohnerInnen teilweise dennoch in individueller Weise umgesetzt [e.g. Da08]. Zusätzlich stellt die Vielzahl der Möglichkeiten bestehende und neu hinzukommende Geräte zu vernetzen NutzerInnen vor eine neue Aufgabe: Aus diesem Überangebot die individuell richtigen Lösungen zu finden [MVH14].

Standardsysteme zu entwickeln, wie dies im industriellen und öffentlichen Gebäudesektor funktioniert, die im Vorhinein die Heterogenität des (häuslichen) Alltags abdecken und in richtiger Art und Weise unterstützend funktionieren können, ist daher schwierig. Eine initiale Installation und Konfiguration reicht in der Regel nicht aus, es muß mit laufenden Anpassungen gerechnet werden bis Mensch und Technik zumindest in akzeptabler Weise zusammenarbeiten. Hinzu kommt, dass sich Lebensumstände bzw. Anforderungen und Bedarfe mit der Zeit ändern können, wodurch sich die Notwendigkeit einer laufenden Adaption erhöht [e.g. Da06].

Die Basis für das zentrale Konzept der Umgebungszintelligenz können zukünftig Smart Home Systeme bilden. Diese sind per Definition [Ch08] dadurch gekennzeichnet, dass sie Verbindungen zu unterschiedlichsten, im Idealfall allen, in der Wohnumgebung vorhandenen Subsystemen haben. Dies beginnt bei der elektrotechnischen Basis wie Verteilung, Verkabelung und Absicherung. Auf der Geräteebene bestehen Verbindungen zu den grob zu unterscheidenden Kategorien Weißware (klassische Haushaltsgeräte) und Braunware (Unterhaltungszintelligenz). Eine besondere Bedeutung nimmt naturgemäß die IT-Infrastruktur ein, die auf unterschiedlichen Endgeräten wie Computer, Drucker, Tablets sowie kabelgebundenen oder drahtlosen Verbindungen und den dafür erforderlichen Geräten (Modems, Router, Switches) basiert. Die einzelnen Gewerke können auch Verbindungen zu anderen aufweisen, zum Beispiel verfügen aktuelle

Fernsehgeräte standardmäßig über Netzwerkschnittstellen, nahezu alle Geräte haben Verbindung zur Basisinfrastruktur, da sie mit Strom versorgt werden müssen.

Ausgehend von der Anforderung das eigene Zuhause als Arbeitsplatz zu nutzen, sollten im Idealfall alle, zumindest aber die wichtigsten Gewerke miteinander kooperieren und auf die besonderen Umstände anpassbar sein. Das erscheint einfacher als es – nach aktuellem technischen Stand – tatsächlich ist. Typische am Markt erhältliche als „smart“ bezeichnete Systeme sind nicht zu anderen Systemen kompatibel und verfügen über eine starre Programmlogik mit festen vorzudefinierenden Regeln.

Ein wirklich „intelligentes“ bzw. adaptives Verhalten, wie es der Name verspricht, ist derzeit von einem Smart Home jedenfalls noch nicht zu erwarten. Weiter entwickelte Systeme, die Verfahren der künstlichen Intelligenz (Inferenz, Lernalgorithmen) einsetzen, zeigen Tendenzen in diese Richtung, denn sie können über entsprechende Sensorik alltägliche Abläufe erfassen und von sich aus bestehende Regeln der Programmlogik anpassen bzw. diese um neue erweitern. Diese Systeme müssen aber über einen entsprechenden Unterbau verfügen (umfassende Sensorik, Serverinfrastruktur zur Datenspeicherung und Analyse) und sind daher noch eher in experimentellen Settings anzutreffen.

Aber auch solche „intelligenteren“ Systeme werden nicht auf wundersame Weise immer richtig auf die Intentionen von AnwenderInnen reagieren können, sondern es wird durch verschiedene Unsicherheitsfaktoren (ungenauere Sensordaten, neue Situationen, etc.) auch zu Fehlentscheidungen kommen. AnwenderInnen müssen daher die Möglichkeit haben, auf diese zu reagieren, zumindest sollte der Grund für das fehlerhafte Verhalten sichtbar sein [EG01].

1.1 Der Arbeitsplatz im eigenen Zuhause

Im Wesentlichen bezieht sich diese Arbeit auf die Tätigkeiten so genannter WissensarbeiterInnen, die mehr geistige als körperliche Tätigkeiten vollbringen, die nicht ortsgebunden sind und somit auch Zuhause durchgeführt werden können. Wesentliches Merkmal der Tätigkeiten ist, dass sie ohne IT kaum denkbar oder zumindest wesentlich schwieriger wären. Das IT Know-how der in diesem Sinne Zuhause arbeitenden Personen kann dabei jedoch sehr unterschiedlich sein. Auf der einen Seite des Kontinuums zum Beispiel der Schriftsteller, der von den im Vergleich zur in der Vergangenheit für die Arbeit genutzten Schreibmaschine vorhandenen Vorzügen des Computers hinsichtlich der Speicherungs-, Sicherungs- und Editiermöglichkeiten profitiert. In der Mitte des Spektrums eine breite Vielfalt von Berufsgruppen wie BuchhalterInnen, VertreterInnen oder Grafik-DesignerInnen, die einen Teil oder die Gesamtheit ihrer Arbeit zu Hause erledigen können, bis hin zu stark IT-lastigen Tätigkeiten welche von ProgrammiererInnen oder SystemadministratorInnen theoretisch von Zuhause durchgeführt werden könnten.

Allen beispielhaft genannten Berufsgruppen ist gemeinsam, dass ein Bedarf nach entsprechender IT-Infrastruktur besteht, der vom einfachen Stand-Alone Computer, über die Standard-Internetanbindung (über LAN oder WLAN), bis hin zu lokalen bzw. remote Speicher- und Sicherungssystemen, Anbindung an Cloud-Systeme oder VPN-Verbindungen zu anderen Infrastrukturen gehen kann.

Die Möglichkeit hochvernetzte intelligente Umgebungen in Kombination mit Smart Home Technologie auch zuhause nutzen zu können bringt weitere Vorteile. Professionelle Gebäudetechnik im Industrie- und Dienstleistungszweckbau sorgt beispielsweise für ein ideales Raumklima unter gleichzeitiger Optimierung des Energieaufwands, oder verfügt über Sicherheitstechnik, die bei einem unberechtigten Zugriff alarmiert – Funktionen die auch der Arbeitsplatz zuhause (zukünftig) bieten könnte. Um die genannten Geräte und Funktionen nutzen zu können, ist jedoch im Vorfeld ein mehr oder weniger hoher Installations- und Konfigurationsaufwand sowie Adaptierungen erforderlich. Das Vorhandensein der entsprechenden Hardwareinfrastruktur setzt diese Arbeit im Haus der Zukunft gedanklich voraus und beschäftigt sich fokussiert mit der Konfiguration und Anpassung.

Während bei einigen AnwenderInnen ein tiefgehendes Verständnis für IT-Belange und die Fähigkeit die Infrastruktur selbst konfigurieren und warten zu können angenommen werden kann, ist dieses bei anderen Gruppen sehr unterschiedlich und kann sehr gering sein. Im Gegensatz zu Arbeitskontexten oder öffentlichen Bereichen sind im privaten Umfeld explizit für das reibungslose Funktionieren der Systeme abgestellte Personen (HaustechnikerInnen, AdministratorenInnen) nicht bzw. nicht permanent vor Ort verfügbar. Um sowohl potenzielle Probleme, Herausforderungen und Vorteile dieser Technologien in den Griff bekommen zu können, werden - ähnlich wie in der Vergangenheit bei Unterhaltungselektronik, Weißware oder konventionellen PC- und Netzwerksystemen - die BewohnerInnen zukünftiger intelligenter Wohnumgebungen die Rolle von SystemadministratorInnen zumindest zu einem gewissen Grad selbst übernehmen müssen.

Lösungen, die AnwenderInnen bei der Anpassung ihrer individuellen intelligenten Umgebungen unterstützen spielen angesichts der skizzierten Herausforderungen eine wichtige Rolle. Wie auch [No11] erkennt, sind gewisse technische Systeme einfach von Grund auf komplex - durch gutes Benutzerschnittstellen- und Interaktionsdesign können sie trotzdem beherrschbar gemacht werden.

Damit diese durchschaubar und auch für Laien nutzbar werden, müssen mehrere Kriterien erfüllt werden. Es muss ein richtiges mentales Modell über deren Funktionsweise und die daraus resultierenden Abläufe aufgebaut werden. Dieses dient als Grundlage um die Möglichkeiten des Smart Home Systems in Bezug auf die eigenen Bedürfnisse zu verstehen und im Falle einer Fehlfunktion bzw. eines nicht gewünschten Verhaltens entsprechende Änderungen vorzunehmen zu können. Das zweite Kriterium ist die Anforderung, dass die Technik sich bei normaler Funktion im Hintergrund hält und nur bei Bedarf bzw. im Falle einer Notwendigkeit sicht- und manipulierbar ist.

Dies erscheint besonders für die eigenen vier Wände erstrebenswert. Bei Bedarf, wie eben zur Konfiguration sollten die entsprechenden Funktionen dennoch leicht und unmittelbar zugreifbar sein. Dies stellt eine der Grundideen des Calm Computing [We99] dar.

Den in dieser Arbeit vorgestellten, auf Augmented Reality basierenden Ansatz, kann man sich als Pendant zur sprichwörtlichen rosarote Brille vorstellen. Diese basiert ja auf einer durch ein entsprechendes mentales Modell eingegengten Perspektive, welche nur angenehme Dinge sichtbar macht. Im Fall der hier vorgestellten Beispiele ermöglicht beispielsweise eine reale Datenbrille angereichert durch entsprechende Informationen eine Veränderung des mentalen Modells von BenutzerInnen und somit eine erweiterte Sicht auf die für die technische Funktion relevanten Komponenten im Smart Home.

2 Augmented Reality als mögliche Realisierung

Auf Augmented Reality [e.g. MRS11] basierende Applikationen erweitern den Blick auf die reale Welt bei Bedarf durch eine zusätzliche Informations- und Interaktionsschicht. Im aktuellen Stand der Technik bieten sich vor allem Tablets und Smartphones an, da sie über entsprechende Leistungsfähigkeit und Ausstattung verfügen, und nicht zu Letzt wegen ihrer weiten Verbreitung.

Daneben drängen neue Generationen von smarten Devices auf den Markt, die für die genannten Zielsetzungen sehr gut geeignet erscheinen. Es sind dies meist ebenfalls auf Smartphone-Technologien basierende AR-Brillen (Smartglasses). Diese können zumindest über kleinere Zeiträume hinweg (typischerweise ~ 2 Stunden) ohne drahtgebundene Stromquelle betrieben werden, was deren Einsatz für Anwendungsszenarien in der Praxis zunehmend interessant macht.

Obwohl diese neue Generation von Smartglasses auch bezüglich ihrer „Tragbarkeit“ in der Praxis (Größe, Gewicht, Aussehen und Design) gegenüber früheren Generationen von AR-Brillen verbessert wurde, scheint eine breitere Akzeptanz für Anwendungen im öffentlichen Leben noch nicht gegeben zu sein. Zu möglichen Hindernissen können fehlende soziale Akzeptanz (Aussehen) aber auch datenschutzrechtliche Bedenken (laufende Videoaufnahme) [Du14] als auch fehlender persönlicher Nutzen [WAE16] gezählt werden.

Im Bereich der Arbeitswelt könnten bestimmte Akzeptanzprobleme weniger problematisch sein, da sich Technologien in diesem Kontext auch in der Vergangenheit als notwendiges Werkzeug etabliert haben - entsprechend Nutzen bringende Anwendungsszenarien natürlich vorausgesetzt (Niemand würde mit einer Schweißbrille durch die Stadt laufen, in der entsprechenden Arbeitssituation ist jedoch ein deutlicher Nutzen vorhanden). Der Umstand, dass im Gegensatz zu anderen Plattformen und Geräten die Hände für die eigentliche Aufgabe weitgehend frei bleiben könnten, erweitert deren Einsatzmöglichkeiten nicht nur im Arbeitsumfeld.

Im privaten Wohnumfeld könnten Nachteile wie Akzeptanzprobleme weniger stark ausgeprägt sein, da in den eigenen vier Wänden Verunstaltungen durch das Tragen von Geräten oder datenschutzrechtliche Bedenken weniger ins Gewicht fallen.

Die weitere Verbreitung von AR blieb jedoch bisher noch aus, was vermutlich mehrere Gründe hat, die Parallelen zur Entwicklung von Smart Home Technologie aufweisen. Die bisherige Entwicklung ist oft technologiegetrieben und mögliche Bedürfnisse der Benutzer bzw. Anforderungen an benötigte Funktionalitäten werden nicht entsprechend berücksichtigt [Le15].

3 Installation und Wartung von Smart Homes

Konkretere Aufgaben, welche die Installation und Wartung von Smart Home Systemen betreffen, sollen hier aus Sicht von zwei Gruppen von Stakeholdern diskutiert werden.

Die erste Nutzergruppe sind die BewohnerInnen selbst, die Installations- und Konfigurationsaufgaben übernehmen (müssen), insbesondere in Hinblick auf die in der Einleitung beschriebene Notwendigkeit der laufenden Anpassung auf geänderte Bedürfnisse und technische Rahmenbedingungen innerhalb des Haushalts. Speziell die nicht technisch versierten, zu Hause arbeitenden Personen werden die Aufgaben nicht allein bewältigen können. Für sie tun sich jedoch mit Augmented Reality neue Möglichkeiten auf, andere Gruppen von Stakeholdern besser einzubinden. Dabei handelt es sich um TechnikerInnen, zum Beispiel SystemadministratorInnen aus der eigenen Firma, oder Smart Home SpezialistInnen.

Stellen wir uns ein Szenario vor, von dem angenommen werden kann, dass es zu Hause ein häufig vorkommendes ist, nämlich das unspezifische „Internet geht nicht“ Problem. Verallgemeinerbar sind in diesem Szenario folgende Punkte. 1) Alle beteiligten Komponenten müssen mit Strom versorgt sein. 2) Die Internetverbindung nach außen muss vorhanden sein, 3) das Netzwerk innerhalb des Zuhauses muss funktionieren, bzw. es dem angeschlossenen Gerät erlauben sich nach außen zu verbinden. Dieses einfache Szenario hat gut und gerne 10 mögliche Fehlerquellen, die einzeln oder in Kombination den Zugriff ins Internet verhindern können.

Der in dieser Arbeit vorgestellte Ansatz wäre zwar auch in Kombination mit konventionellen Technologien denkbar, Smart Home Systeme werden aber als zukünftiger Standard gesehen und sind auch aufgrund der beschriebenen Vernetzung der im Haushalt vorhandenen Gewerke funktional besser dazu geeignet, Ansätze wie den vorgestellten zu unterstützen.

In dem einfachen Szenario muss zunächst geprüft werden, ob die beteiligten Komponenten entsprechend mit Strom versorgt sind. Beim Computer oder Laptop ist das relativ einfach möglich. Zumindest eine weitere Komponente wie Modem oder Router muss jedoch noch überprüft werden, die sich nicht unmittelbar in der Nähe befinden

muss. Sowohl eine Sichtprüfung am Gerät als auch eine Prüfung ob der entsprechende Stromkreis aktiv ist (Kontrolle der Sicherungsautomaten im Zählerschrank) ist auch dem Laien zumutbar, sofern er oder sie entsprechende Informationen über Standort und Position hat bzw. erhält. In einem Smart Home wäre es darüber hinaus aber auch denkbar, dass die Stromzufuhr über ein vom System gesteuertes Relais deaktiviert wurde, etwa weil das Internet zu dieser Tageszeit in der Regel ohnehin nicht verwendet wird und Energie gespart oder Elektrosmog reduziert werden soll. Letztere sind wünschenswerte Funktionen, die Komplexität kann sich aber im Fehlerfall schnell erhöhen.

Die nächste Prüfung beinhaltet die logische Verbindung zum Internet, die unter Anderem beinhaltet ob eine Verbindung nach außen (Verbindung zwischen Modem/Router und Vermittlungsstelle/Knoten), oder nach innen (Gerät ist per LAN Kabel am Router angeschlossen bzw. hat eine WLAN Verbindung, entsprechende Berechtigungen existieren, Ports sind offen etc.) besteht. Alles in allem bei einem auch kleinen Beispiel viele Punkte, die schuld an einer Fehlfunktion sein können und den Laien überfordern.

Derartige Probleme könnten mit Augmented Reality entweder allein oder in Kooperation mit versiertem Fachpersonal gelöst werden. Dabei sind jedoch im Vergleich zu industriellen oder öffentlichen Gebäuden mehr oder weniger gravierende Unterschiede zu erwarten. Gebäude der genannten Sektoren basieren im allgemeinen auf einer überschaubaren Anzahl von Standards wie KNX, desweiteren müssen entsprechende Installationen schon im Stadium der Ausschreibung, der Planung und der Installation detaillierter dokumentiert werden.

Im intelligenten Zuhause ist man mit einem breiteren Spektrum an möglichen Systemen und Einzelkomponenten konfrontiert. Neben bereits vereinzelt anzutreffenden Systemen aus der professionellen Gebäudetechnik gibt es auch vielfältige smarte Lösungen speziell für den privaten Heimbereich (z.B. funk-, powerline basierte Systeme) sowie weniger integrierte Insellösungen (Smart-Objects), die in Kombination mit konventioneller Installationstechnik zum Einsatz kommen können. Weiters muss bei älteren Gebäuden damit gerechnet werden, dass vorhandene Elektroinstallationen nicht dem heutigen Standard entsprechen, ausreichend dimensionierte Leerverrohrung für eine Nachrüstung bzw. Adaptierung fehlen, sowie Dokumentation von unsystematisch gewachsenen Systemen wenig bis gar nicht vorhanden ist. Auf der Ebene der smarten Adaptierung bieten funk- und powerline basierte Systeme zwar vor allem in der Nachrüstung von smarten Funktionen Vorteile in Bezug auf den Installationsaufwand, verwenden jedoch in der Regel individuelle, nicht herstellerübergreifende Verbindungs- und Übertragungstechniken. Entsprechend groß sind auch die Unterschiede bei der Vernetzung und Programmierung der integrierten Komponenten, was beispielsweise für den Laien als auch für die Fachperson, die in einen Haushalt mit einem beliebigen System zur Störungsbeseitigung gerufen wird, den Aufwand erheblich erhöhen kann, da gegebenenfalls erst entsprechendes Know-How über das verbaute System eingeholt werden muss.

Mit Middlewareplattformen wird versucht dem System- und Protokollwirrwarr im Heimbereich Herr zu werden, was aus Anwendersicht für den alltäglichen Betrieb Vorteile bringt (einheitliche Benutzerschnittstellen und Mechanismen der Bedienung, Adaptierung und Konfiguration, etc.). Im Fehlerfall beispielsweise wird jedoch oft auch wieder ein Eingreifen im jeweiligen Subsystem notwendig sein bzw. wird die Komplexität des Gesamtsystems durch die zusätzlichen Komponenten für die Middleware noch erhöht.

Möglichkeiten der Fernwartung von Smart Home Systemen sind sowohl für den Laien im Hinblick auf Kurzfristigkeit als auch für Anbieter in Hinblick auf die Aufwands- und Kostenreduktion von Interesse. Besonders gilt das, zum Beispiel, für ländliche Gebiete wo für angefordertes Fachpersonal lange Anfahrtswege anfallen können, gleichzeitig aber für Smart Home Technologien beispielsweise zur Unterstützung von Ambient Assisted Living Lösungen große Potentiale vorhanden sind [e.g. Le15]. Auch im Zusammenhang mit dem Home-Office und, zum Beispiel, der Notwendigkeit der Anpassung eines Heimnetzwerkes auf firmenspezifische Sicherheitsrichtlinien können die (ev. einzigen entsprechend ausgebildeten und autorisierten) ServicetechnikerInnen des Arbeitsgebers nur mit entsprechend hohem Aufwand Hilfe vor Ort zur Verfügung stellen.

Konventionelle Supportlösungen via Telefon können bei komplexeren technischen Systemen schnell an ihre Grenzen stoßen, insbesondere wenn die erwähnten Unterschiede und individuellen Gegebenheiten vor Ort und die im Durchschnitt geringe technische Versiertheit der BenutzerInnen berücksichtigt werden. Damit eine Fachperson aus der Entfernung einen Laien vor Ort möglichst erfolgreich unterstützen kann, müssen beide ein gemeinsames Grundverständnis über die Fragestellung und das konkrete Problem entwickeln. Fehlt der unterstützenden Person die direkte Sicht auf das individuelle Gesamtsystem, so können durch Missverständnisse und andere ungewollte Nebeneffekte von Lösungsschritten Probleme noch verschlimmert bzw. neu generiert werden [PEJ09].

Selbst wenn eine Einschränkung auf einen Hersteller möglich ist und auf den Kunden-Support des jeweiligen Anbieters zurückgegriffen werden könnte, so ergeben sich dennoch je nach Haushalt individuell verschiedene Ausbaustufen und mögliche Kombinationen von Einzelgeräten. Typischer Weise müssen verschiedene Systeme von verschiedenen Herstellern und deren Vernetzung untereinander identifiziert werden. Desweiteren hat die bauliche Infrastruktur (Grundriss, Elektroinstallation, etc.) Einfluss auf den Aufbau und die Vernetzung des Gesamtsystems und die Platzierung von Komponenten. Nicht zuletzt bestimmen auch die individuellen Bedürfnisse und alltäglichen Routinen der BenutzerInnen den Einsatz und die Kombination von Komponenten. Ob dieser vielschichtig komplexen Problemstellung ist das Erreichen des zuvor angesprochenen Grundverständnisses schwierig.

Remote Zugriff auf Konfigurationsschnittstellen von Smart Home Systemen (falls vorhanden) könnten zumindest einige der angesprochenen Hürden erleichtern.

Das würde zwar über die Standardverbindung im oben genannten „Internet geht nicht“ Szenario auch nicht funktionieren, da ja die Internetverbindung unterbrochen ist. In diesem Fall könnte aber theoretisch eine Alternativverbindung via Smartphone (mobiles 4G/3G Netzwerk) Abhilfe schaffen. Bei nicht ohnehin cloudbasierten Lösungen ist dafür in der Regel jedenfalls eine entsprechende, nicht immer triviale Konfiguration des Heimnetzwerkes (Port-Forwarding, etc.) notwendig, was vielfach ohne zusätzlichen Support von außen wiederum nicht möglich sein wird.

Leichter bzw. zweckmäßiger wäre ein Zugriff auf cloudbasierte Lösungen möglich, die zumindest Teile der Programmlogik auf leichter zugängliche Server auslagern. Allerdings sind hier Bedenken bezüglich des Datenschutzes zu berücksichtigen, da speziell bei Smart Home Systemen auch sensible private (Sensor-) Daten erfasst und verarbeitet werden. Durch die reine Zugriffsmöglichkeit auf Konfigurationsoberflächen können jedenfalls nicht alle potentiellen Probleme eines auf verschiedene Hardwarekomponenten basierenden Systems gelöst werden, da oft auch an der jeweiligen im Gebäude verteilten Hardware selbst Hand angelegt werden muss (Kabelverbindungen, Batteriewechsel, etc.).

4 Augmented Reality im Smart Home

Augmented Reality basierte Lösungen könnten für die in Abschnitt 3 beschriebenen Aufgaben in unterschiedlicher Weise von den diversen Stakeholdern eingesetzt werden, was in diesem Abschnitt anhand von verschiedenen Anwendungsszenarien diskutiert werden soll.

4.1 Szenario 1: Hilfe und Information vor Ort

In einem simplen Fall könnte das Einblenden von Bedienungsanleitungen für einzelne Komponenten des Smart Homes realisiert werden, die beispielweise nach dem Einlesen eines Barcodes am möglicherweise fehlerhaften Gerät in der AR Sicht einer Datenbrille eingeblendet werden. Dies könnte einerseits schon dem Laien helfen, speziell aber für Fachpersonal von Nutzen sein, das mit neuen Komponenten konfrontiert unverrichteter Dinge wieder abreisen müsste. Entsprechende Webservices für den Zugriff auf die jeweilige Dokumentation könnten vorhanden sein und werden idealerweise über Herstellergrenzen hinweg zentral abgelegt und gewartet (der Zugriff könnte analog zum Szenario „Internet geht nicht“ notfalls wieder alternativ über ein mobiles Netzwerk erfolgen). Zu den Aufgaben einer derartigen zentralen Stelle könnten folglich auch der einheitliche Aufbau der Inhalte und die Anpassung der Darstellung für AR Zwecke gehören.

Weitergehendere Zugriffsmöglichkeiten einer AR Applikation in Systeme vor Ort könnten auf der Bereitstellung von kontextspezifischeren Informationen basieren. So könnten beispielsweise Daten über Geräteverknüpfungen und die dahinter liegende

Programmlogik abgefragt und visualisiert werden. Abbildung 1 skizziert ein Beispiel, das Informationen über die Verknüpfung und physische Verteilung für mehrere Geräte in einer AR-Sicht darstellt. Abbildung 2 zeigt eine mögliche Sicht durch eine AR-Brille für einen Steckdosen-Aktor, der auf Signale eines Bewegungsmelders reagiert, mit Möglichkeiten zur Konfiguration der Verbindung. Derartige Möglichkeiten Geräteverknüpfen zu verwalten werden beispielsweise auch von [Kr10] oder [HKM13] auf Basis von Smartphone bzw. Tablet basierten Applikationen realisiert.

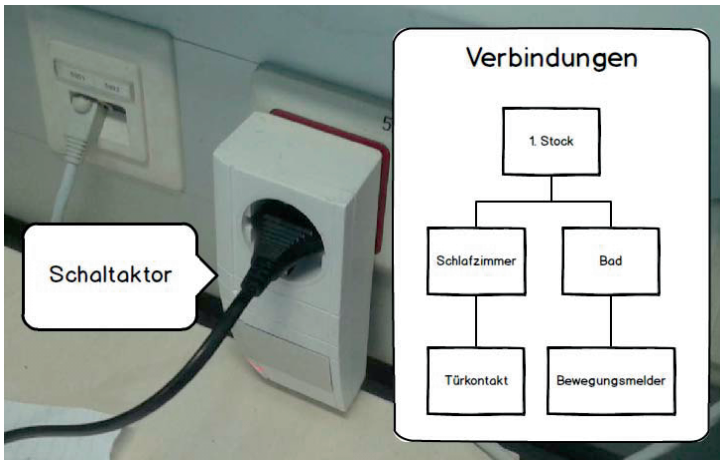


Abb. 1: AR basierte Visualisierung von Geräteverbindungen

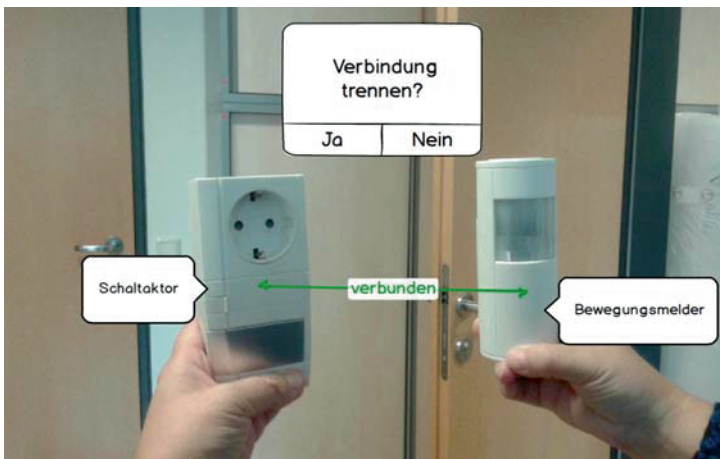


Abb. 2: Verwaltung von Geräteverbindungen via AR Brille

4.2 Szenario 2: Datensicherheit

Einen wesentlichen Faktor beim Betreiben von intelligenten Umgebungen stellt das Thema Sicherheit dar. Auch nicht cloudbasierte Plattformen bieten oft die Möglichkeit einer Integration in das heimische LAN bzw. WLAN um beispielsweise Bedienoberflächen über Smartphones oder Tablets zugänglich zu machen. Bei entsprechendem Internetzugang und geeigneter Routerkonfiguration kann dies auch von außerhalb des heimischen Netzwerks erfolgen. Der Zugriff von außen erfordert aber ein entsprechend abgesichertes Heimnetzwerk um unberechtigten Zugriff auf persönliche oder arbeitsrelevante Daten zu verhindern. Aber nicht nur von außen droht Gefahr. Es muss auch sichergestellt sein, dass z.B. Familienmitglieder nicht auf sensible Firmendaten zugreifen können. Auch wenn keine böse Absicht unterstellt werden muss, könnten Daten irrtümlich überschrieben oder gelöscht werden. Schließlich muss gewährleistet sein, dass durch nicht kontrollierten Datentransfer (z.B. herstellerspezifische Updates oder Bugfixes) keine Sicherheitslücken entstehen. Auch hier könnten AR Lösungen eingesetzt werden um den ein- und ausgehenden Datenfluss eines Geräts live vor Ort in einem entsprechenden Overlay zu visualisieren, wie es beispielsweise von [MHS14] vorgeschlagen wird. Selbst wenn der jeweilige Datenverkehr verschlüsselt sein sollte so kann es dennoch sinnvoll sein zu wissen, dass von oder zu einem Gerät ein nicht erklärbarer Datenfluss stattfindet.

4.3 Szenario 3: Remote Unterstützung

Für AR verwendbare Geräte (insbesondere auch AR Brillen wie Smartglasses) verfügen üblicherweise über eine für Videoaufnahmen geeignete Kamera, die für die Erkennung und Verfolgung von Markern oder anderwertig identifizierbaren Objekten (Stichwort: Natural Feature Tracking) verwendet wird. Der beispielsweise über eine AR Brille aufgenommene Videostrom kann aber auch über ein Netzwerk weitergeleitet werden um sich mit einer anderen Person sein derzeitiges Blickfeld zu teilen (Ausschnitt und Qualität hängen von den Parametern und der Positionierung der Kamera am AR Gerät ab). Genau solche Anwendungsmöglichkeiten können eingesetzt werden um zwischen einer Person vor Ort und einer aus der Ferne unterstützenden Fachperson das in Abschnitt 3 angesprochene gemeinsame Problemverständnis zu gewinnen. Die individuelle Smart Home Installation vor Ort wäre damit für die Fachperson über eine vom Laien getragene AR Brille live sichtbar und Einzelheiten, die einen Überblick über das gesamte System ermöglichen, müsste nicht erst von einer nicht technisch versierten Person erfragt werden.

Je klarer das Bild der Problemstellung für die unterstützende Technikerin oder den unterstützenden Techniker desto eher kann mit einer erfolgreichen Problemlösung aus der Ferne gerechnet werden. Analog kann sich auch eine Fachperson vor Ort bei Bedarf weitere Unterstützung von anderen ExpertInnen für bestimmte Teilbereiche einholen. Ebenfalls kann die Entscheidung bezüglich der Notwendigkeit eine Fachperson an den Standort des jeweiligen Haushalts zu schicken unterstützt werden bzw. kann diese sich

schon vorab ein Bild über die Gegebenheiten vor Ort machen und sich entsprechend besser auf den Einsatz vorbereiten. Abbildung 3 skizziert ein Beispiel für den Batteriewechsel bei einem Funkschalter. Eine Person vor Ort teilt ihre Sicht der Problemstellung über eine AR Brille mit einer remote unterstützenden Fachperson. Über eine Voice over IP Verbindung können Informationen via Sprache ausgetauscht werden („Hier öffnen“, „Batteriefach“), weiters könnte die Fachperson auch die Möglichkeit haben remote weitere Hinweise in grafischer Form auf der AR Ebene darzustellen (grüne Pfeile im Beispiel).

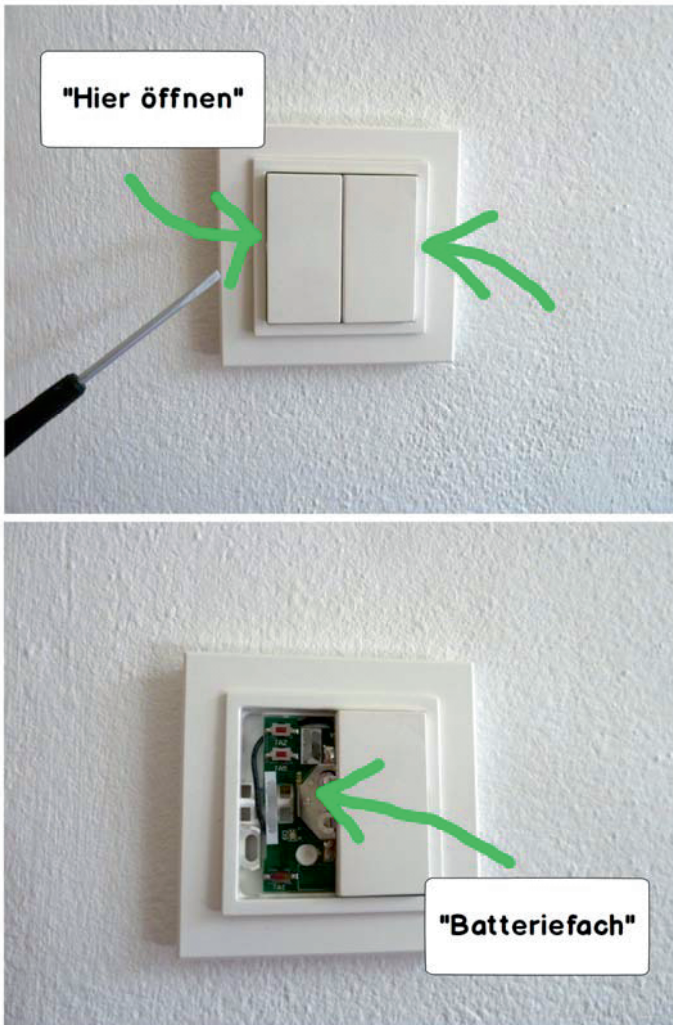


Abb. 3: Remote Unterstützung via AR Brille

5 Ausblick

AR Technologien versprechen interessante Anwendungsszenarien für die Installation und Wartung von zukünftigen Wohn- und Arbeitsumgebungen. Für den Einsatz in der Praxis sind diesbezüglich jedoch noch einige Fragestellungen offen. Eine grundlegende Frage könnte die Wahl der Hardwareplattform sein. Smartphones und Tablets sind weit verbreitet, mehr oder weniger praxistaugliche AR Brillen in Form von Smartglasses sind relativ neu am Markt. Letztere hätten den Vorteil, dass AnwenderInnen für die eigentliche Aufgabe grundsätzlich beide Hände frei hätten. Voraussetzung dafür ist jedoch der Einsatz von innovativeren Interaktionsformen wie Sprach- oder (3D-) Gestensteuerung während auf den herkömmlichen mobilen Plattformen etabliertere Interaktionsformen auf Basis von Touchinterfaces zum Einsatz kommen. Generell waren die Entwicklungen in den Bereichen Smart Home sowie auch AR in der Vergangenheit eher industrie- bzw. technologiegetrieben und die Sicht der eigentlichen AnwenderInnen wurde oft vernachlässigt. Eine weiter gehende Orientierung dieser Technologien hin zum Menschen mit seinen konkreten Anforderungen und Bedürfnissen wird daher Gegenstand laufender und zukünftiger Forschung im Bereich der Mensch-Maschine Interaktion sein müssen.

Literaturverzeichnis

- [Ch08] Chan, Marie; Estève, Daniel; Escriba, Christophe; Campo, Eric: A review of smart homes - Present state and future challenges. *Comput. Methods Prog. Biomed.* 91, 1, 55-81, 2008.
- [Da06] Davidoff, Scott; Lee, Min Kyung; Yiu, Charles; Zimmerman, John; Dey, Anind K.: Principles of smart home control. In: *Proc. of UbiComp'06*, Springer, Berlin, Heidelberg, 19-34, 2006.
- [Da08] Dahl, Y.: Redefining smartness: The smart home as an interactional problem. In: *Proc. of 2008 IET 4th International Conference on Intelligent Environments*, Seattle, WA, 1-8, 2008.
- [Du14] Due, Brian L.: An essay about challenges and possibilities with smart glasses, Working papers on interaction and communication, 1(2), Centre for Interaction Research and Communication Design, University of Copenhagen, 1-21, 2014.
- [EG01] Edwards, W. Keith; Grinter, Rebecca E.: At Home with Ubiquitous Computing: Seven Challenges. In: *Proc. of UbiComp '01*, Springer-Verlag, London, UK, 256-272, 2001.
- [Hi99] Hindus, Deborah: The importance of homes in technology research. In: *Cooperative Buildings. Integrating Information, Organizations, and Architecture*, Springer Berlin Heidelberg, 199-207, 1999.
- [HKM13] Heun, Valentin; Kasahara, Shunichi; Maes, Pattie: Smarter Objects: Using AR technology to Program Physical Objects and their Interactions. In: *Extended abstracts of CHI 2013*, ACM, New York, 961-966, 2013.

- [Kr10] Kriesten, Bastian; Mertes, Christian; Tünnermann, René; Hermann Thomas: Unobtrusively Controlling and Linking Information and Services in Smart Environments. In: Proc. of NordiCHI 2010, 276-285, 2010.
- [Le15] Leitner, Gerhard: The Future Home is Wise, Not Smart. Springer Verlag London, 2015.
- [MHS14] Mayer, Simon; Hassan, Yassin N.; Sörös, Gábor: A magic lens for revealing device interactions in smart environments. In: Proc. of SIGGRAPH Asia 2014 Mobile Graphics and Interactive Applications. Shenzhen, China, 1–6, 2014.
- [MRS11] Mehler-Bicher, Anett; Reiß, Michael; Steiger, Lothar: Augmented Reality: Theorie und Praxis, Oldenbourg, München, 2011.
- [MVH14] Mennicken, Sarah; Vermeulen, Jo; Huang, Elaine M.: From today's augmented houses to tomorrow's smart homes: new directions for home automation research. In: Proc. of UbiComp '14, ACM, New York, USA, 105-115, 2014.
- [No11] Norman, Donald A.: Living with Complexity, The MIT Press, Cambridge, Mass. [u.a.], 2011.
- [PEJ09] Poole, Erika Shehan; Edwards, W. Keith; Jarvis, Lawrence: The Home Network as a Socio-Technical System: Understanding the Challenges of Remote Home Network Problem Diagnosis. Comput. Supported Coop. Work 18, 2-3, 277-299, 2009.
- [WAE 16] Weiz, Daniel; Anand, Gagat; Ernst, Claus-Peter H.: The Influence of Subjective Norm on the Usage of Smartglasses. In: The Drivers of Wearable Device Usage: Practice and Perspectives, Springer, Cham, Switzerland: 1-11, 2016.
- [We99] Weiser, Mark: The computer for the 21st century. SIGMOBILE Mob. Comput. Commun. Rev. 3, 3, 3-11, 1999.

New Trends in Building Automation for Offices of the Future

Jan Haase¹, Dominik Meyer²

Abstract: Many office buildings today have the serviceableness and comfort of 40 years old standards. Due to limited budget especially the public authorities are not able to change the edificial status very often. However, there are developments of features or devices in modern building automation that can drastically increase the level of safety, security, comfort, energy saving, and other areas.

This includes trends such as modern elevator concepts moving more than one cabin per shaft and thus generating a much better throughput. Such a system is able to optimize the order of stops and thereby save energy while enhancing the individual's sensation of efficiency. Other examples are electronic awareness-systems which can track people in the building and automatically turn on and off lighting and heat, and adaptive door signs. This can be achieved by deploying sensors of different kinds (such as photo sensors, proximity sensors, and cameras) and interpreting the sensor data received. Further examples can be emergency lights embedded in the floor which show the way to the nearest exit. These can also be used to lead guests to specific offices.

Apart from that, modern buildings can learn, i.e., they can collect data and thereby extrapolate or predict human behavior. This is helpful in order to recognize situations out of the ordinary (e.g., intruders or even people in trouble) as well as prepare for predicted excessive demands like the morning rush of arriving employees to elevators. This paper gives an overview on current trends in building automation for office buildings

Keywords: Building automation, office, energy consumption, occupant comfort, elevator, emergency solutions

1 Introduction

Many office buildings existing today were built several decades ago, obeying the construction rules and laws of these times. There are only few cases where existing buildings have to be refitted or upgraded to new laws (such as the elimination of asbestos, which was banned in several states in the 90's and in the European Union in 2005). In these cases normally transitional provisions are granted, postponing the mandatory retrofitting for several years. So even today many buildings still have asbestos-containing pipes, roof tiles, thermal insulation, plasterwork, etc.

¹ Institute of Computer Engineering, Universität zu Lübeck, Ratzeburger Allee 160, D-23562 Lübeck, haase@iti.uni-luebeck.de

² Faculty of Electrical Engineering, Helmut Schmidt University of the Federal Armed Forces Hamburg, Holstenhofweg 85, D-22043 Hamburg, dmeyer@hsu-hh.de

Apart from avoiding health threats, many newer trends find their way into refurbished or newly built buildings. These more and more include comfort features for the inhabitants or – in the case of public buildings – employees and visitors.

Section 2 gives an overview on currently existing requirements (by law or by trend) for modern buildings. Section 3 highlights several use cases. In Section 4 open questions and future topics are touched, before Section 5 concludes this paper.

2 Requirements for Modern Buildings

The current requirements for newly built or refurbished office buildings can be categorized into the following classes:

- Safety
- Security
- Comfort
- Energy saving

2.1 Safety

Safety aspects encompass all parts of office buildings where accidents might happen. This includes disasters like fires, earthquakes, etc. Thus, buildings have to be designed to not trap fleeing people, with a special attention on the needs of the elderly or disabled.

2.2 Security

In this category all parts of surveillance and counteractions against theft or breaking and entering are located.

2.3 Comfort

People working in public buildings are nowadays used to certain comfort features such as escalators, elevators (possibly with a higher capacity due to special control schemes), smooth and quiet floors, doors automatically opening using motion detectors, etc.

2.4 Energy saving

This category has become more and more important in the last years due to efforts to lower the overall energy consumption of office buildings. This covers thermal insulation

as well as distributed low-power sensors. These sensors nowadays use ultra-low power communication protocols like ZigBee or Wireless Hart [GHG11] in order to save energy and are in many cases equipped with harvesting features like small solar panels.

Various adaptable systems and devices including photovoltaics, batteries and electric vehicles have been proposed for the purpose of supporting highly energy sustainable buildings as well as enhancing overall comfort for the people residing or working in it [HZA15].

3 Selected Use Cases

In this Section a number of selected use cases is highlighted which shows the current state of the art regarding the categories described in Section 2.

3.1 Elevators

The first elevators were used for centuries for goods only, but in the second half of the 19th century the triumph of passenger elevators started with the introduction of the safety elevator by Elisha Graves Otis in 1852. This elevator featured a mechanic brake in case the rope the elevator was hanging from broke.

In the next decades the availability of elevators enabled master-builders to design houses much higher than four or five floors which was common until then. This was an important pre-condition for the construction of skyscrapers in the 20th and 21st century.

Nowadays, essentially all office buildings have elevators, but an important issue is the footprint, i.e., how much space it will consume. Furthermore, common elevators cannot be used as an emergency route in case of a fire in the building due to smoke and electrical shutdown; thus, saving elderly or disabled people from higher floors the fire truck's ladder cannot reach is a serious problem. Furthermore, the throughput of elevator systems in terms of passengers/time unit is being maximized, since especially in peak times often long queues of waiting passengers form in front of the elevators. These problems are being addressed by more recent technological developments.

As elevators often present to be a bottleneck for passengers, several approaches were developed to address this problem. Typically, many passengers want to travel from the ground floor to one of the upper floors in the morning, and back down again in the evening. An easy adaption of common elevators is the feature that the cabins wait with closed doors and thus can quickly be sent to any floor they are requested to. Better algorithms [BR15] also send them back automatically to the ground floor in the mornings, even if they are not currently being requested. The traffic patterns can be categorized into peak traffic, down peak traffic, lunch time (two way) traffic and inter-floor traffic [SC10].

A technologically new step was the double deck elevator, which was first installed in the Eiffel Tower in Paris 1889 and then in the Empire State Building 1931 by Otis but is still not very common today. The idea is to have one cabin stop at even floors and the other at odd floors and thus transport twice as many passengers at the same time. Depending on their destination, passengers can take an escalator to a landing of the alternate cabin (see Fig. 1).



Fig 1: Double elevator at Midland Square Nagoya, Japan (taken from [WC16])

Modern elevators may feature a regenerative drive (see Fig. 2), i.e., they feed energy generated when the cabins slow down back to the building's power grid. The difference between the lightly or heavily loaded car and the counterweight is used to harvest energy – a concept being used for modern trains, too, which harvest energy when slowing down before stops at train stations. The overall energy consumption of the elevator can thereby decrease by 30% [HZZ14,LSC15].

REGENERATIVE DRIVE - HOW DOES IT WORK?

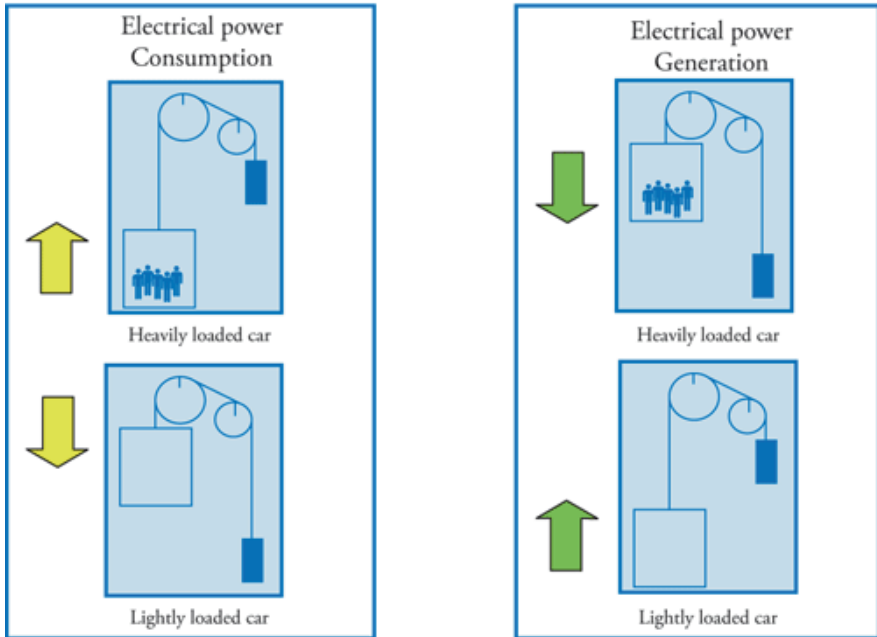


Fig 2: Regenerative drive (taken from [Otis16])

An important new concept invented by ThyssenKrupp in 2003 was the Twin [Thy03]. It runs two cabins in the same elevator shaft, one above the other. The selection of target floors is not done in the cabin, but outside, and therefore the scheduling algorithm can send the better suited cabin to pick up the passenger. The scheduling control also avoids collisions. ThyssenKrupp indicates time saving of about 65%.

A very new and unique approach is ThyssenKrupp's Multi [Thy16a]. It runs cabins without ropes or cables. The engines are not located in an elevator engine room, but in the cabins itself. Thus the cabins can run independently from each other and not only move vertically, but also horizontally (if horizontal shafts exist). This enables the cabins to reach different parts of buildings as well as pass each other (see Fig. 3). A test tower presenting the first implementation of this technology is currently being built in Rottweil, Germany, and expected to be finished by end of 2016 [Thy16b].

The MULTI system and possible competitor products will then open up new ways of elevators, since in case of higher demand simply some additional cabins can be put into operation, and put aside when demand decreases.

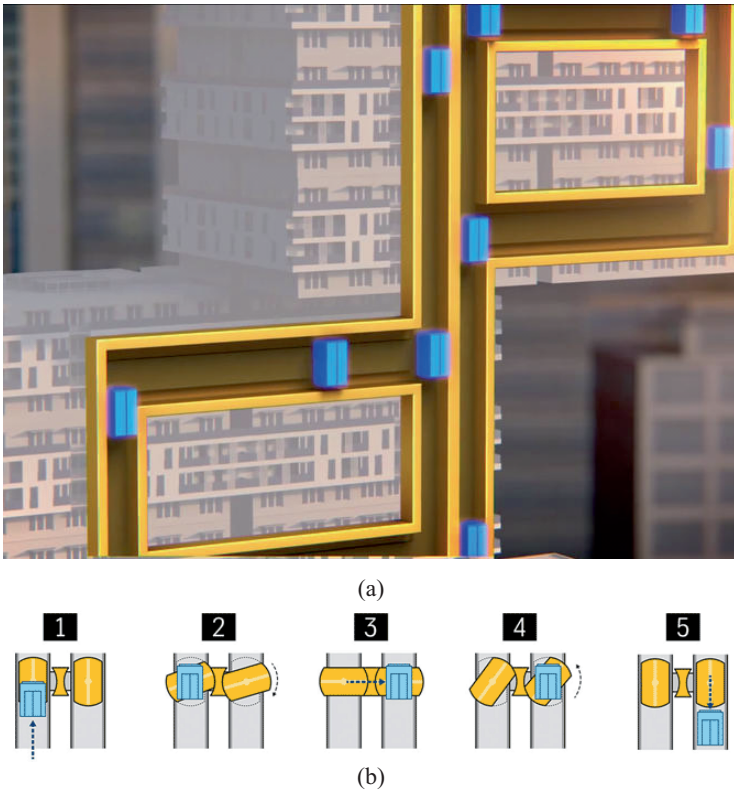


Fig 3: The MULTI system: (a) cabins running independently and even horizontally, (b) an elevator changing from one vertical shaft to another (both taken from [Thy16a])

Other studies try to conserve energy and enhance comfort by estimating the number of passengers in a common elevator by means of cameras [MIY15]. The idea is to reduce unnecessary stops for full elevators.

One of the solutions for the safety problem of rescuing people from higher floors in case of disasters like fire is the evacuation elevator [JD07, KOK14]. It offers continued operation using auxiliary energy and is built in a way that on all floors a certain area in front of the elevator can be sealed off by fire doors, forming a safe location for disabled people. The evacuation elevator can then be boarded and will take all passengers to the ground floor.

3.2 Lighting and Heat

If the building control system has information about where people are currently present, a lot of energy can be saved by only activating heat in non-empty rooms. Unused offices

thus don't consume energy above the minimum [MHR14, ANC15, ADH16].

This also may enhance the comfort of occupants. However, the estimation of comfort parameters is a complex problem [RFR15, SCP15].

In the last years, the topic of automatic ventilation and shading (by sun tracking or shadow outline tracking) was well covered by industry. The idea is to keep out heat in warm regions or summer times while letting in light, avoiding artificial lighting in the day for power consumption reasons. Special kinds of glass shades can be attached to the building surface and automatically rotated depending on the solar factor [Colt15].

Furthermore, long corridors can be (automatically) lighted only where people are, further decreasing overall energy consumption.

Buildings – and particularly public buildings – are viewed as possible objects for storing energy. The idea is to slightly overheat in times of cheaper energy and thus having to heat less later, when the price of energy has risen over the day. Furthermore, solar energy is to be stored, too. Similarly, cooling of buildings in hot regions or summer times can be made more energy-efficient this way. Current projects try to forecast the energy availability as well as the energy consumption [PZJ11, YCZ12, SUT13, KZ15].

3.3 Tracking People

In office buildings it might be convenient to track people. There are several possible reasons. In order to save energy, lighting and heat might only be switched on (possibly automatically) when people are present, as described in the previous Section. Typical technologies for tracking are motion sensors, evaluation of employees' access cards, cameras recognizing and tracking faces [LLZ15], or wearable devices [HRT14].

Apart from that an information about the whereabouts of people in a building is very helpful in a case of emergency like a fire or earthquake. The rescuing team (e.g., firemen) can then focus on rooms with (possibly incapacitated or unconscious) persons and ignore empty offices.

However, the gathering of data to get this information is legally problematic in many countries – it violates the security rights of the people being monitored. There are still open issues in terms of finding a viable path between emergency preparation and data privacy protection [GH89].

3.4 Security Features

A very common security application in office buildings is access control.

Currently, the usual access control systems (see Fig. 4) are more and more being replaced by smarter systems which have intelligent readers. Other studies focus on fully

automated face recognition systems [IZ11] instead of relying on tracking data (see Section 3.3).

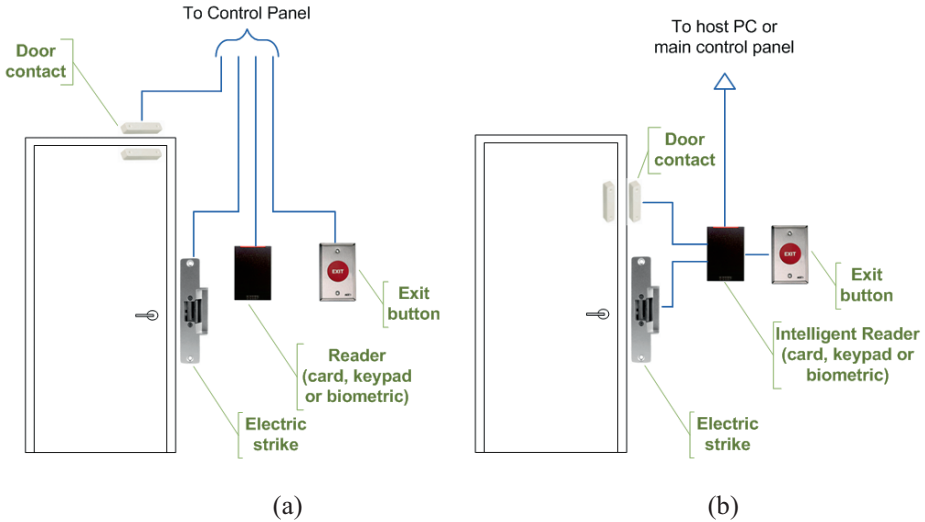


Fig. 4: Typical door access control. (a) basic readers send read or detected data to a control panel which then in turns decides whether the door is to be opened, (b) intelligent readers can directly control the door opening

3.5 Emergency Solutions

Modern office buildings feature several approaches to lessen the impact of disasters like fire or earthquakes. Typical examples for this are non-flammable floors and wallpapers, emergency lighting in case of a power failure, or windows being opened automatically in case of fire which act as smoke outlets.

Other approaches focus on the rescue team (e.g., the firemen), who have to enter the burning building and quickly save people as well as extinct the fire. Current office buildings do have a floor plan available for firemen which indicates all relevant locations like hose couplings or dangerous fuel tanks.

This might in the future be enhanced to include more or less current data about the people present in the building (see Section 3.3). Furthermore, currently blocked areas can be marked in the floor plan. This requires an adaptable or even interactive presentation, e.g., a floor plan on a fireman's tablet pc instead of an ordinary printout.

Taking a step further, this floor plan with additional information could be loaded into standardized (and thereby widely available) display systems of fireman's masks, giving them an augmented reality view, similar to Google's glass (see Fig. 5) or other wearable

computers [Bar16].

The office building control system needs to unlock doors for rescue teams in order to offer the most direct way to trapped victims. This is a feature not yet widely implemented.



Fig. 5: Augmented reality for firemen using Google Glass (taken from [FG14])

4 Future Topics

Future topics for elevator technologies will be the ever increasing number of passengers due to higher and bigger buildings. Particularly office buildings tend to be located in city centers, where land prices are steep – thus they will almost certainly be built higher and higher. If the passenger queues cannot be processed in time, though, the value of upper-floor offices (and apartments) will decrease. Thus this is still an open problem. The new approach of independently running cabins also gives the opportunity to create “premium” services for special guest, i.e., cabins reserved for important passengers which will then be given priority on their way to the destination.

Another future topic might be to use lighting in the floor or walls as a guidance system for guests or visitors, leading them to the exact office they need to be. If the employee the visitor has an appointment with currently stays in another room (e.g., a conference room), the guidance system automatically adapts and shows the quickest route.

Future security related topics could be focusing on the electronic desktop. Currently, it is already possible in many offices to log in at any computer and access personal data from there, as it is stored on a central server. In the future, all office equipment might be connected this way and thus also incoming phone calls might automatically be rerouted to the room an employee currently is in, door plates adapt according to the current occupants of a room (which might be even more useful for conference rooms), and even personal comfort data like preferred room temperatures is migrated with the employee. All these features can also extended to home office via secure VPNs.

Furthermore, new algorithms using face recognition or similar methods could make use of tracking data and check the plausibility of results, e.g., if an employee was detected and identified in the lobby seconds ago and now wants to enter a restricted room. If these algorithms are adequately reliable, future offices might get on without access cards or entry codes at all.

The area of emergency management is a growing topic for public buildings. An important subject is the automated communication with rescuing teams entering the building. For this, new protocols might be needed. Other communication, suddenly of much lower priority, be choked off – an example might be a massively used WiFi network which will automatically disconnect all clients in case of an emergency in order to minimize traffic and keep enough bandwidth for emergency communication. Furthermore, the guidance system for visitors mentioned before could be used for rescuing teams, too, showing the quickest way to a given destination.

5 Conclusion

This paper gave an overview on current developments in the area of building automation with regard to office buildings. Due to the longevity of public buildings the spectrum of buildings with more or less modern features is large; usually only in case of newly built or refurbished buildings new technology is introduced. However, there are several interesting approaches that might become trendy or even compulsory in the years to come for existing buildings.

6 References

- [ADH16] I. B. A. Ang, F. Dilys Salim and M. Hamilton, "Human occupancy recognition with multivariate ambient sensors," 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops), Sydney, NSW, 2016, pp. 1-6.
- [ANC15] A. Akbar, M. Nati, F. Carrez and K. Moessner, "Contextual occupancy detection for smart office by pattern recognition of electricity consumption data," 2015 IEEE International Conference on Communications (ICC), London, 2015, pp. 561-566.

- [Bar16] Woodrow Barfield, Ed., "Fundamentals of Wearable Computers and Augmented Reality", 2nd Ed., CRC Press, 2016, ISBN 978-1-4822-4351-2
- [BR15] D. T. Bamunuarachchi and D. N. Ranasinghe, "Elevator group optimization in a smart building," 2015 IEEE 10th International Conference on Industrial and Information Systems (ICIIS), Peradeniya, 2015, pp. 71-76.
- [Colt15] Colt International Ltd., "Shadoglass and Shadovoltaic Solar Shading Systems", 2015, http://www.colt.es/descargas.html?file=tl_files/pdf/Solar%20Shading/Shadoglass%20and%20Shadovoltaic%20solar%20shading%20systems.pdf
- [FC14] Fast Company, "This Firefighter Built His Own Google Glass App And It's Saving Lives", 2014, <http://www.fastcompany.com/3026559/this-firefighter-built-his-own-google-glass-app-and-its-saving-lives>
- [GH89] Grant, R. and Higgins, C. (1989), Monitoring service workers via computer: The effect on employees, productivity, and service. *Natl. Prod. Rev.*, 8: 101–112. doi: 10.1002/npr.4040080203
- [GHG11] K. Gravogl, J. Haase, C. Grimm, "Choosing the best wireless protocol for typical applications", *Proceedings of 24th International Conference of Architecture in Computing Systems (ARCS)*, 2011
- [HRT14] M. Hardegger, D. Roggen, G. Tröster, "3D ActionSLAM: wearable person tracking in multi-floor environments", *Journal of Personal and Ubiquitous Computing*, Vol. 19, No. 1, pp. 123-144, Sept 2014, ISSN 1617-4917
- [HZA15] J. Haase, G. Zucker, F. AlJuheshi, M. khair Allah and M. Alahmad, "A survey of adaptive systems supporting green energy in the built environment," *Industrial Electronics Society, IECON 2015 - 41st Annual Conference of the IEEE, Yokohama*, 2015, pp. 004009-004014.
- [HZZ14] J. Han, Z. Zheng, J. Zheng and J. Li, "A survey of braking energy recovery and management technology," *Transportation Electrification Asia-Pacific (ITEC Asia-Pacific)*, 2014 IEEE Conference and Expo, Beijing, 2014, pp. 1-4.
- [IZ11] R. Ibrahim and Z. M. Zin, "Study of automated face recognition system for office door access control application," *Communication Software and Networks (ICCSN)*, 2011 IEEE 3rd International Conference on, Xi'an, 2011, pp. 132-136.
- [JD07] Theodore L. Jarboe, John O'Donoghue, "Elevator and Escalator Rescue, a Comprehensive Guide", PennWell, 2007
- [KOK14] Max T. Kinateder, Hidemi Otori, Erica D. Kuligowski, "The Use of Elevators for Evacuation in Fire Emergencies in International Buildings", *National Institute of Standards and Technology, NIST Technical Note 1825*, July 2014, <http://dx.doi.org/10.6028/NIST.TN.1825>
- [KZ15] K. J. Kircher and K. M. Zhang, "Model predictive control of thermal storage for demand response," 2015 American Control Conference (ACC), Chicago, IL, 2015, pp. 956-961.
- [LLZ15] Jun Liu, Ye Liu, Guyue Zhang, Peiru Zhu, Yan Qiu Chen, "Detecting and tracking people in real time with RGB-D camera", *Pattern Recognition Letters*, Volume 53, 1 February 2015, Pages 16-23, ISSN 0167-8655,

- [LSC15] S. Lin, W. Song, Y. Chen, L. Luo and Z. Feng, "Study on the model of elevator regeneration energy and its energy storage control method," 2015 International Conference on Smart Grid and Clean Energy Technologies (ICSGCE), Offenburg, 2015, pp. 125-128.
- [MHG10] J. M. Molina, J. Haase and C. Grimm, "Energy Consumption Estimation and Profiling in Wireless Sensor Networks," Architecture of Computing Systems (ARCS), 2010 23rd International Conference on, Hannover, Germany, 2010, pp. 1-6.
- [MHR14] N. Masoudifar, A. Hammad and M. Rezaee, "Monitoring occupancy and office equipment energy consumption using real-time location system and wireless energy meters," Proceedings of the Winter Simulation Conference 2014, Savannah, GA, 2014, pp. 1108-1119.
- [MIY15] F. Mohamudally, C. S. Inn, L. S. Yeong and C. W. Chong, "Estimating free space in an elevator's confined environment," TENCON 2015 - 2015 IEEE Region 10 Conference, Macao, 2015, pp. 1-6.
- [Otis16] Otis Elevator Company, Website, <http://www.otis.com>, 2016.
- [PZJ11] P. Palensky et al., "Demand response with functional buildings using simplified process models," IECON 2011 - 37th Annual Conference on IEEE Industrial Electronics Society, Melbourne, VIC, 2011, pp. 3230-3235.
- [RFR15] M. L. Ruz, S. Fragosó, D. Rodríguez and F. Vázquez, "Real-time estimation of thermal comfort indices in an office building with a solar powered HVAC system," Control and Automation (MED), 2015 23th Mediterranean Conference on, Torremolinos, 2015, pp. 803-808.
- [SC10] George R. Strakosch, Robert S. Caporale, Eds., "The Vertical Transportation Handbook", 4th Ed., Wiley, 2010, ISBN: 978-0-470-40413-3
- [SCP15] S. S. Shetty, H. D. Chinh and S. K. Panda, "Strategies for thermal comfort improvement and energy savings in existing office buildings using occupant feedback," 2015 IEEE International Conference on Building Efficiency and Sustainable Technologies, Singapore, 2015,
- [SUT13] L. Spitalny, D. Unger, J. Teuwsen, V. Liebenau, J. M. A. Myrzik and B. Van Reeth, "Effectiveness of a building energy management system for the integration of net zero energy buildings into the grid and for providing tertiary control reserve," PowerTech (POWERTECH), 2013 IEEE Grenoble, Grenoble, 2013, pp. 1-6.
- [Thy03] ThyssenKrupp, TWIN elevator system, <https://twin.thyssenkrupp-elevator.com/>, 2003
- [Thy16a] ThyssenKrupp, "MULTI, a new era of elevators to revolutionize high-rise and mid-rise constructions", <http://www.thyssenkrupp-aufzuege.de/en/new-installation/elevators/passenger-elevators/multi/>, 2016
- [Thy16b] ThyssenKrupp, "The Test Tower, starting point of ground-breaking projects", <http://testturm.thyssenkrupp-elevator.com/en/>, 2016
- [WC16] Wikimedia Commons, https://en.wikipedia.org/wiki/File:Double_elevator_at_Midland_Square_Nagoya.JPG, 2016

- [YCZ12] Xiaohui Yan, Haisheng Chen, Xuehui Zhang and Chunqing Tan, "Energy storage sizing for office buildings based on short-term load forecasting," 2012 IEEE 6th International Conference on Information and Automation for Sustainability, Beijing, 2012, pp. 290-295.

Integration von User Privacy Mechanismen in clientseitige Standardbibliotheken

Hauke Coltzau¹

Abstract: In diesem Beitrag wird überblicksartig diskutiert, wie clientseitige Standardbibliotheken wie jQuery [JQ16] erweitert werden können, um Möglichkeiten zur verdeckten Profilbildung, sowie zum Fingerprinting und Tracking von Nutzern über Webseitengrenzen hinweg einzuschränken. Er dient als Diskussionsgrundlage für nachfolgende Implementierungstätigkeiten.

Keywords: User Privacy, Datenhoheit.

1 Verdeckte Profilerzeugung und User Tracking

Den vielen Vorteilen, die sich aus der Nutzung smarterer, mobiler Endgeräte im Alltag ergeben, steht unter anderem die direkte Bedrohung der Privatsphäre und Anonymität ihrer Nutzer vor allem gegenüber Dienst- und Inhaltsanbietern gegenüber. Dies geschieht zum einen durch die als freiwillig anzunehmende Datenherausgabe der Kunden an soziale Onlinenetzwerke, zum anderen aber durch verdeckte Sammlung und Analyse von Nutzerdaten über Betreibergrenzen hinweg und unter Einsatz von Big-Data Methoden.

Canvas Fingerprinting (u.a. [AEE14]) und Nutzerverhaltensanalyse ([ARA11]) zeigen exemplarisch, wie es auf Anbieterseite gelingen kann, ohne Verwendung üblicher Authentifizierungsverfahren über Systemgrenzen hinweg Nutzer mit hoher Sicherheit zu identifizieren. In Kombination mit Realweltdaten, die durch die große Anzahl von Sensoren in modernen Smart Devices problemlos erfasst werden können, entstehen umfangreiche Profile der Nutzer, über deren Erstellung, Auswertung und Weitergabe ebendiese keinerlei Kontrolle ausüben können. Die mittlerweile als permanent anzunehmende Vernetzung der Geräte in Kombination mit deren eigener Leistungsfähigkeit ermöglicht die Erstellung und Übertragung von Profildaten in erheblichem Umfang und mit feiner zeitlicher Granularität. Der Impact ist besonders groß, wenn weitere, nicht kontextzugehörige Daten beispielsweise aus sozialen Netzwerken zur Profilerstellung herangezogen werden.

Datenanalyse im Backend und Verknüpfung von auf den ersten Blick unzusammenhängenden Daten führen zu Erkenntnissen, auf deren Basis ohne tatsächliche Kenntnis der Person unternehmerischen Entscheidungen getroffen werden. Diese wiederum können sich erheblich auf den Nutzer und seinen Alltag auswirken, wie

¹ Fernuniversität Hagen, hauke.coltzau@fernuni-hagen.de

es beispielsweise bei wohnortbezogener Bonitätsbewertung bereits gängige Praxis ist. Im höchsten Maße bedenklich ist hierbei, dass Nutzer weder Kenntnis darüber besitzen, welche Daten genutzt werden, noch Einfluss nehmen können, die Auswertung, Interpretation und Weitergabe derselben zu steuern.

Es liegt also im Interesse der Nutzer, der ungesteuerten und für sie unkontrollierbaren Verwertung eigener Daten sowie der Profilerzeugung und -nachverfolgung entgegenzuwirken. Aufgrund vielfältiger Szenarien und Kontexte, in denen Profildaten entstehen und ausgewertet werden können, scheint ein einheitlicher und umfassender Ansatz, der alle diese Szenarien abdeckt, in der Praxis nur schwer umsetzbar. Stattdessen erscheint es zielführender, den Nutzern eine Menge unabhängiger Werkzeuge und Hilfsmittel zur Verfügung zu stellen, die sie an ihre jeweiligen Bedürfnisse und Anwendungskontexte anpassen können.

2 User Privacy

So naheliegend und fundiert die Kritik an freiwilliger Herausgabe eigener Daten beispielsweise bei Registrierung in und Nutzung von sozialen Onlinenetzwerken (Facebook et. al.) auch ist, so stellt dieses Verhalten aus Sicht des Autors und im Kontext dieses Beitrags keine Verletzung der User-Privacy dar. Schließlich stimmt der User dieser Weitergabe und der daraus resultierenden weitläufigen Nutzung aktiv zu und kann sich in Anbetracht breiter gesellschaftlicher Diskussion auch nicht auf Unwissenheit berufen. Eine technische Lösung ist hier weder zielführend noch ist sie geeignet, an Stelle des notwendigen gesellschaftlichen und politischen Diskurses zu treten.

Im Fokus der diskutierten Fragestellung stehen stattdessen diejenigen Nutzerdaten, für die mindestens drei der folgenden Kriterien zutreffen, wobei die Grenzen zwischen zulässiger und unzulässiger oder fragwürdiger Datenverwertung fließend sind, sodass diese Kriterien selbst nur Richtlinien sein können:

- sie werden ohne Zustimmung der Nutzer erhoben
- ihre Erhebung erfolgt individualisiert
- der Umfang ihrer Erhebung ist unklar
- ihre Erhebung dient nicht oder nicht hauptsächlich dem für den Nutzer erkennbaren Zweck der besuchten Seite
- sie werden ohne explizite Zustimmung über Seiten-/Betreibergrenzen hinweg geteilt
- sie werden in einer Art und Weise verwendet, die geeignet ist, für den Nutzer nicht transparente, an sich vermeidbare Nachteile zu generieren.

Während über die Verwendung der Daten im Backend und deren weiterer Auswertung naturgemäß nur wenig Informationen vorliegen, ist über Techniken des Profilings, des Fingerprintings und des Trackings bekannt, dass diese weitflächig eingesetzt werden.

Beim Profiling (s. [M+14][PHV+13]) werden vom Nutzer explizit oder implizit zur Verfügung gestellte Informationen zur Erstellung eines individuellen (ggf. auszugsweisen) Benutzerprofils herangezogen. Hierbei lässt sich grob an Hand der verwendeten Quellen unterscheiden: Öffentliche Quellen stehen dem Grunde nach ohne nennenswerte Beschränkungen für jedermann zum Profiling zur Verfügung. Hierzu zählen im Wesentlichen soziale Onlinenetze aber auch Onlineangebote von Lokalzeitungen und privaten Betreibern (Vereine, etc.). Diese Form des Profilings kann mit Hilfe von Big-Data Methoden auf eine große Anzahl von Nutzern angewendet werden (s.a. [HRN13]).

Geschlossene Quellen entstehen dort, wo Nutzer in einem als privat angenommenen Verhältnis zu einem Dienstanbieter bei der Inanspruchnahme der Dienste Daten erzeugen, die zum Profiling durch den Dienstanbieter selbst geeignet sind. Diese Daten können hochsensibler und sehr persönlicher Natur sein, wie es beispielsweise bei E-Mails oder Kalenderdaten der Fall ist. Entsprechend „wertvoll“ sind diese Daten für den Dienstanbieter.

Gegen Profiling aus offenen Quellen können Nutzer vor allem durch Nicht-Veröffentlichung entgegenwirken, technische Lösungen stehen hier nicht im Vordergrund. Die Abwehr von Profiling aus geschlossenen Quellen gestaltet sich bei der Nutzung von Onlinediensten ungleich schwieriger, da hierzu in den meisten Fällen ein Mitwirken des Dienstanbieters vorausgesetzt ist, der jedoch das Profiling im eigenen Interesse durchführt. Auch hier sind nicht in erster Linie technische Lösungen gefragt.

Die Nachverfolgung besuchter Seiten und ggf. deren Navigation durch diese Seiten wird als (User-)Tracking bezeichnet. Hierzu müssen Nutzer eindeutig identifizierbar sein (nicht notwendigerweise jedoch *persönlich* identifizierbar). Soll diese Identifizierbarkeit auch ohne Wissen und Zutun des Nutzers erfolgen, spricht man von *Fingerprinting*. Klassische Verfahren wie die Verwendung von Cookies sind aufgrund ihrer guten Abwehrbarkeit auf dem Rückzug. Neuere Verfahren (s. u.a. [AJN+13], [NKJ+13]) versuchen, den User über Eigenschaften seines Browsers (Canvas Fingerprinting), seines Systems oder über das Verhalten des Nutzers selbst (bspw. durch Mausbewegungen) zu identifizieren. Anonymes Browsing, wie es beispielsweise von *The Onion Router* (TOR) zur Verfügung gestellt wird, verschleiert nur den Absender, schützt jedoch keinesfalls sicher vor Tracking und Profiling.

Nikiforakis et. al. beschreiben in [NJL15] Varianten zur Verschleierung von für das Fingerprinting geeigneten Daten durch gezielt eingefügte Zufallsinformationen und zeigen dessen Wirksamkeit gegen einige in der Praxis verbreitete Tools im Rahmen einer eigenen Browserimplementierung. Ihre Arbeit kann als Grundlage für diejenigen im nachfolgenden Kapitel diskutierten Erweiterungen genutzt werden, in denen Randomisierung von Daten verwendet werden soll.

Allen Verfahren zum Fingerprinting und Tracking ist gemein, dass sie nicht ohne aktiven Code auf Seite des Clients durchgeführt werden können. Der sicherste Ansatz zur Vermeidung von Fingerprinting ist daher zweifellos, die Ausführung von clientseitigen Skripten zu unterbinden. Dies führt jedoch dazu, dass eine Vielzahl von Webanwendungen und Webseiten nicht oder nur sehr eingeschränkt nutzbar ist. Skriptblocker wie *NoScript* verhindern die Ausführung einzelner Skripte, nichtversierte Nutzer können jedoch kaum fundierte Entscheidungen treffen, welche Skripte zulässig sind und welche ihre Datenhoheit untergraben.

Automatisierte Lösungen zur Vermeidung von Tracking und (in Grenzen) Profiling wie *Privacy Badger*, *Disconnect Private Browsing* und *Ghostery* unterbinden die Ausführung von bekannten Clientbibliotheken mit Trackingfunktionalität. Sie sind jedoch browserabhängig und setzen stetige Pflege im Sinne der Aufnahme neuer Blockadeandidaten voraus.

3 Möglichkeiten zur Datenkontrolle an der Schnittstelle zum User

Eine Alternative zum Blockieren von Skripten kann sein, Standardbibliotheken an der Schnittstelle zum User derart zu härten, dass dem Nutzer erfasste und übertragene Daten transparent gemacht werden und er deren Übermittlung nach seinen Bedürfnissen unterbinden oder steuern kann.

Zur Implementierung solcher Schutzansätze bietet sich eine Erweiterung der jQuery-Bibliothek ([JQ16]) an, weil diese Bibliothek als leistungsfähiges Standardwerkzeug nahezu alle Aspekte der Implementierung von Clientanwendungen abdeckt und einen besonders hohen Verbreitungsgrad hat. Sie bietet Schnittstellen zur Erfassung von Nutzerereignissen, zur Manipulation des DOM und zur asynchronen Kommunikation im Hintergrund an und ist damit eigentlich als Werkzeug zur Aushöhlung der Datenhoheit des Nutzers prädestiniert.

Nachfolgend wird übersichtsartig gezeigt, welche Eingriffsvarianten sich direkt in jQuery implementieren lassen und auf welche bekannten Verfahren zur Vermeidung von Fingerprinting, Tracking und ungewollter Datenerhebung im Allgemeinen dabei zurückgegriffen werden kann.

3.1 Kommunikationsbeschränkung

Eine der Kernfunktionen von jQuery ist die Durchführung synchroner und asynchroner http-Requests sowie die Behandlung der entsprechenden Antworten. Asynchrone Requests eignen sich naturgemäß besonders für Datenübertragung im Hintergrund, aber auch über eingeschleuste Parameter synchroner Requests lassen sich am Nutzer vorbei Daten übertragen.

Ähnlich wie in bestehenden Browserplugins kann die Möglichkeit zum Absetzen von http-Requests gezielt eingeschränkt werden. Sowohl Vermeidung von Cross-Site Kommunikation als auch Techniken wie Black- und Whitelisting kommen als naheliegende Varianten in Frage. Durch die weite Verbreitung von jQuery ist eine communitygetriebene Pflege der jeweiligen Listen ein realistischer Ansatz.

Eingeschleuste Parameter lassen sich auf diesem Wege jedoch nur begrenzt erkennen. Dies gilt insbesondere dann, wenn durch Verschlüsselung oder ähnliche Verfahren eine Analyse derselben nicht möglich ist. Als einzige Maßnahme bleibt in diesem Fall die Information des Netzers über potentielle Verletzungen seiner Datenhoheit.

3.2 Content Sandboxing

Bereits bevor Daten versendet werden, kann daher eine Beschränkung des Zugriffs auf diese sinnvoll sein. Inhalte von Bedienelementen wie Suchfeldern sind im Regelfall nur für einen kleineren Teil der Anwendung von Bedeutung und können vor dem Zugriff durch andere Anwendungszweige geschützt werden, ohne dass Funktionsverluste entstehen. Detektionspunkte entstehen im Wesentlichen dort, wo auf DOM-Elemente aus mehreren Skripten heraus zugegriffen wird. Die Abwägung, ob ein Zugriff generell erlaubt wird und welche Informationen je nach aufrufendem Skript zur Verfügung gestellt werden, ist sicher nicht trivial. Dennoch kann das Auftreten von skriptübergreifendem Zugriff zumindest als Indiz für ungewollte Datenerhebung gewertet werden und im einfachsten Fall auch hier zu einer Information des Nutzers führen, selbst wenn keine weiteren Aktionen daraus abgeleitet werden.

3.3 Ereignisnormalisierung und -verschleierung

Eine zentrale Anlaufstelle zur Erfassung von Verhaltensdaten ist die Ereignisverwaltung auf dem Client. Jedes Ereignis, dessen Auslösecharakteristik entweder durch den Nutzer oder durch das verwendete System beeinflusst wird, ist ein potentieller Anknüpfungspunkt für Tracking und Profiling. Sobald jedoch sichergestellt wird, dass sämtliche Ereignisse durch eine geeignete Normalisierungs- oder Verschleierungskomponente gefiltert werden, bevor sie zur Bearbeitung an registrierte Listener weitergeleitet werden, sind die Möglichkeiten hierzu wesentlich stärker beschränkt. Dieses Verfahren kann vollständig transparent durchgeführt werden.

Verschleierung und Normalisierung sind unterschiedliche Ansätze, den Informationsgehalt der Ereignisdaten zu reduzieren. Bei der Ereignisverschleierung wird in dem Ereignis in geeignetem Maße Information verändert, was dazu führt, dass mehrere gleichartige Ereignisse weniger voneinander unterscheidbar sind und somit die Ereignischarakteristik nicht mehr oder zumindest nur in reduziertem Maße zum Fingerprinting herangezogen werden kann. Dasselbe Ziel verfolgt die Normalisierungskomponente, die mitgelieferte Ereignisdaten auf ihre wesentliche Aussage reduziert. So ist beispielsweise bei der Behandlung von Mausclick-Ereignissen

im Regelfall nur relevant, welche DOM-Elemente davon betroffen sind, nicht aber die exakten Koordinaten des Klicks. Dennoch müssen Normalisierung und Verschleierung dynamisch an die Bedürfnisse des Nutzers und der jeweiligen Anwendung angepasst werden können, um Funktionalitätseinbußen zu vermeiden.

Selbstverständlich können Skripte, die Daten für Fingerprinting erfassen sollen, eigene Ereignisbehandlungsfunktionen implementieren und diese an jQuery vorbei direkt über die Javascript Kernfunktionen registrieren. Hier lassen sich jedoch die Sprachmöglichkeiten zum sehr einfachen Überladen von Funktionen in Javascript direkt nutzen, um sicherzustellen, dass die diskutierte Erweiterung von jedem dieser Registrierungsversuche Kenntnis erlangt und diese über die eigenen Behandlungsroutinen transparent umlenkt.

3.4 Ereigniskettenanonymisierung

Während Ereignisnormalisierung und -verschleierung nur das einzelne Ereignis betrachtet, richtet sich Ereigniskettenanonymisierung gegen die Ermittlung von Erkenntnissen, die sich aus der Kombination mehrerer in Zusammenhang stehender Ereignissen ableiten lassen, also beispielsweise einer Analyse des Zeitverhaltens bei der Tastatureingabe oder bei Mausbewegungen oder der Verwendung des Scrollrades. Normalisierung und Verschleierung sind bei solchen Ereignisserien leicht durchführbar, beispielsweise durch Einfügen von interpolierten Blindereignissen, durch Auslassung oder Verzögerung von Ereignissen, wenn Folgeereignisse existieren oder durch Glättung von Bewegungskurven.

3.5 Anonyme geteilte Profilnutzung

Noch einen Schritt weiter führt der Gedanke, zusammenhängende Ereignisketten (also Nutzerverhalten) anonymisiert über mehrere Systeme hinweg als Pseudoverhaltensprofile zu teilen und anstelle der vom Nutzer angestoßenen Ereignisketten zu verwenden. Über die Zeit oder mit Wechsel auf andere Webseiten werden auch die verwendeten Profile gewechselt. Fingerprinting ist dann zwar möglich, verfehlt aber seinen Zweck, denn weder wird der eigentliche Nutzer sicher identifiziert noch ist Tracking über mehrere Webseiten hinweg möglich.

3.6 Systemanonymisierung

Fingerprinting auf Basis von Systemeigenschaften (Canvas Fingerprinting und ähnliche Verfahren) macht sich zu Nutze, dass sich Systeme in ihrem Verhalten und in ihrer Konfiguration in vielen Details unterscheiden. Die Analyse einzelner Eigenschaften (wie beispielsweise die Anzahl installierter Schriften) reicht zur Identifikation typischerweise nicht aus, ermöglicht aber durch Kombination eine oft ausreichend hohe Identifikationsgenauigkeit.

Einige der herangezogenen Eigenschaften wie beispielsweise Variationen bei der Darstellung von Texten betreffen noch Grenzbereiche der Zuständigkeit von jQuery, weil die dazu notwendigen Operationen das DOM manipulieren und Informationen aus diesem abfragen. Viele Systemeigenschaften wie Details zur Browserumgebung liegen jedoch klar außerhalb der Kompetenz von jQuery. Es wäre demnach eine eigenständige Erweiterung zu implementieren.

Die Abwehrmöglichkeiten sind dem Verfahren nach ähnlich zu den bereits diskutierten Ansätzen. Systemeigenschaften lassen sich ausblenden oder manipulieren (beispielsweise die Anzahl geladener Plugins), randomisieren oder mit anderen Nutzern über Systemgrenzen hinweg teilen.

Fingerprinting auf Basis der Leistungsfähigkeit des Systems (CPU Rechenzeit für Testaufgaben, verfügbarer Arbeitsspeicher) lässt sich mit diesen Mitteln jedoch nicht oder nur schwer vermeiden. Hier sind systemnähere Ansätze gefordert.

4 Zusammenfassung und Ausblick

In diesem explizit als Ideenpapier zu verstehenden Beitrag werden Ansätze diskutiert, clientseitige Standardbibliotheken wie jQuery derart zu erweitern, dass ungewollte Datenerfassung und –übertragung eingeschränkt werden kann. Die Vorschläge richten sich insbesondere gegen Fingerprintingtools und Tracker und erheben dabei keinen Anspruch auf Vollständigkeit.

In zukünftigen Arbeiten sollen die hier vorgeschlagenen Ansätze im Rahmen eines quelloffenen Projektes implementiert und ihre Wirksamkeit gegenüber bekannten Tracking- und Fingerprintingwerkzeugen überprüft werden.

Literaturverzeichnis

- [AEE14] Acar, G. et al.: The web never forgets: Persistent tracking mechanisms in the wild. In: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2014. S. 674-689.
- [ARA11] Angeletou, S.; Rowe, M.; Alani, H.: Modelling and Analysis of User Behaviour in Online Communities. In: (Aroyo, L. et.al. Hrsg.): Proc. 10th International Semantic Web Conference, Bonn 2011. Springer, Berlin/Heidelberg S. 35-50, 2011.
- [JQ16] jQuery Projektseite, jquery.com, Stand: 27.06.2016.
- [NJL15] Nikiforakis, N.; Joosen, W.; Livshits, B.: PriVaricator: Deceiving Fingerprinters with Little White Lies. Proc. 24th International Conference on World Wide Web. 2015. ACM, New York, S. 820-830.
- [M+14] Mitrou, Lilian, et al.: Social media profiling: A Panopticon or Omnipticon tool? In: Proc. of the 6th Conference of the Surveillance Studies Network. 2014.

- [PHV+13] Peñas, P.; del Hoyo, R.; Vea-Murguía, J.; González, C.; Mayo, S.: "Collective Knowledge Ontology User Profiling for Twitter -- Automatic User Profiling," Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2013, Atlanta, 2013, S. 439-444.
- [HRN13] Hoppe, A. ; Roxin, A.; Nicolle, C.: Dynamic, Behavior-Based User Profiling Using Semantic Web Technologies in a Big Data Context", In Proc. On the Move to Meaningful Internet Systems: OTM 2013 Workshops, Graz, 2013, Springer, Berlin/Heidelberg, S. 363-372.
- [AJN+13] Acar, G.; Juarez, M.; Nikiforakis, N.; Diaz, C.; Gürses, S.; Piessens, F.; Preneel, B.: FPDetective: dusting the web for fingerprinters. In Proc. ACM SIGSAC conference on Computer & communications security (CCS '13), 2013, ACM, New York, 1129-1140.
- [NKJ+13] Nikiforakis, N.; Kapravelos, A.; Joosen, W.; Kruegel, C.; Piessens, F.; Vigna, G.: Cookieless Monster: Exploring the Ecosystem of Web-Based Device Fingerprinting. In Proc. IEEE Symposium on Security and Privacy (SP), 2013, Berkeley, S. 541-555

Überlegungen zu Multi-Sensor-Aktor-Systemen und Kontrolle in intelligenten Arbeitsumgebungen

Lilian Schröder¹, Aljoscha Pörtner¹, Matthias König¹ und Martin Hoffmann¹

Abstract: Zu einem modernen Arbeitsplatz gehört eine angenehme Arbeitsumgebung. Zahlreiche Forschungsergebnisse aus dem Bereich „Indoor Environment Quality“ liefern Informationen dazu, welche Parameter der Umgebung einen Einfluss auf Produktivität sowie Wohlbefinden der Personen haben. Erweitert werden diese Faktoren durch das Beobachten der Menschen selbst, ihrer Emotionen und ihrer Aktivitäten. Um eine intelligente Arbeitsumgebung zu schaffen, wird ein System gebraucht, das diese Parameter erfasst und darauf intelligent reagiert. Dafür muss zunächst vielfältige Sensorik miteinander verbunden werden. Nur dann kann das Zusammenspiel der Umweltfaktoren und die Interaktion zwischen Mensch und Umwelt berücksichtigt werden. Nicht zuletzt ist die Kontrolle eines Menschen über seine Arbeitsumgebung entscheidend für die Arbeitszufriedenheit. Die Wünsche und Einflussmöglichkeiten des Menschen müssen daher verstärkt berücksichtigt werden. Für die technische Umsetzung eines solchen Systems bietet sich ein Multiagentensystem an, welches die Komplexität auf Agenten verteilt, aber trotzdem eine übergeordnete Instanz besitzt, welche ein menschliches Eingreifen ohne Umwege ermöglicht. Es werden Vorschläge gemacht, wie eine solche Architektur gestaltet werden könnte.

Keywords: Indoor Environment Quality, Multiagentensystem, Intelligente Gebäude, Organic Computing, Autonomic Computing

1 Einleitung

Was ist nötig, um eine angenehme, produktive Arbeitsumgebung zu schaffen? Zahlreiche wissenschaftliche Arbeiten haben zu dieser Frage Ergebnisse aus unterschiedlichen Bereichen geliefert (s.u.). Verschiedene Umwelteinflüsse, persönliche Faktoren und das Zusammenspiel aus Aufgabe und Umgebung beeinflussen nachweislich sowohl die Produktivität als auch die Zufriedenheit der Menschen an ihrem Arbeitsplatz. Moderne Technologien ermöglichen es, die Arbeitsumgebung auf vielen Ebenen flexibel an persönliche Bedürfnisse anzupassen. Im ersten Teil dieser Arbeit wird aufgezeigt, wie verschiedene Sensoren und Aktoren dafür genutzt werden können, alle interessanten Parameter zu erfassen, miteinander zu verknüpfen und so die Umgebung für den Menschen zu optimieren. Der zweite Teil dieser Arbeit befasst sich mit der technischen Umsetzung eines solchen Systems. Es wird ein Überblick über Multiagentensysteme als Basis für intelligente vernetzte Systeme gegeben und die Relevanz der Kontrolle über diese Systeme in den Fokus gerückt. Darauf folgend werden eine mögliche Architektur und die daraus resultierenden Forschungsfragen für das Gebäude der Zukunft diskutiert.

¹ Fachhochschule Bielefeld, Campus Minden, Artilleriestr. 9, 32427 Minden, [Vorname].[Nachname]@fh-bielefeld.de

Der erste Schritt in Richtung eines solchen Systems ist es, die relevanten Parameter zu bestimmen. Studien zu „indoor environment quality (IEQ)“ beschäftigen sich damit, welche Einflüsse der Umwelt Produktivität und Wohlbefinden sowie Zufriedenheit beeinflussen. Im Folgenden sollen die wichtigsten Erkenntnisse der bisherigen Forschung dargestellt werden, die die Grundlage für die Entwicklung eines Sensorsystems darstellen. Für viele einzelne Faktoren wurde bereits gezeigt, dass sie sich negativ auswirken und sogar die Gesundheit beeinträchtigen können. Die meisten Untersuchungen beschäftigen sich mit Arbeitsumgebungen und versuchen, diese zu bewerten. Um eine Bewertung zu erreichen, wird einerseits die Zufriedenheit der Mitarbeiter gemessen, andererseits werden die Arbeitsleistung und die Produktivität in Abhängigkeit von der Umgebung gemessen, z.B. über Performanz-Messungen [Ma04][BB98]. Die Zufriedenheit und das Wohlergehen können z.B. über Fragebogen ermittelt werden (Beispiele: Depression Anxiety and Stress Scale, Perceived Stress Questionnaire [MAT13]). Die Konzepte Job-Zufriedenheit und Produktivität müssen dabei für eine genaue Betrachtung durchaus getrennt gesehen werden [Ju01], es wird aber häufig nicht klar dargestellt, wie genau Zufriedenheit und Produktivität zusammenhängen. Insgesamt sind für die Forschung im Bereich intelligente Gebäude sowie IEQ beide Konzepte relevant. Weitere Faktoren, die in der Literatur als Einfluss auf die Produktivität genannt werden, sind Motivation [EJ00], sowie Stimmung, Emotion und Persönlichkeit (s.u. und [OPS09]). Kreativität hingegen kann als Variable gemessen werden, um die Qualität der Arbeit über die Produktivität hinaus bestimmen zu können. Dafür gibt es auch einige Methoden, z.B. einen Remote Associates Test oder eine Aufgabe, bei der Ideen generiert werden sollen [MZC12].

1.1 Indoor Environment Quality

Schon seit den achtziger Jahren gibt es Interesse daran, die Auswirkungen der Arbeitsbedingungen im Büro auf den Menschen zu erforschen (z.B. [CN89]). Klassische Themen sind Luftqualität und Temperatur, und seit dem Aufkommen von Großraumbüros besonders auch das Geräuschlevel in der direkten Arbeitsumgebung. Einen Überblick über die Thematik Indoor Environment Quality geben z.B. Kim und de Dear [KdD12]. Häufiger Gegenstand der Untersuchungen ist das Messen eines Zusammenhangs zwischen Umweltfaktoren und der Leistung und/oder dem Wohlbefinden. Für zahlreiche Faktoren der Umgebung gibt es Ergebnisse, die zeigen, wie eine Arbeitsumgebung gestaltet sein sollte:

- Temperatur: Sie ist ideal zwischen 21 und 22° [SFL06]; beeinflusst viele Parameter wie Konzentration und Arbeitsleistung aber auch gesundheitliche Aspekte, und ist abhängig von persönlichen Vorlieben, Kleidung, Außentemperatur und Luftbewegungen; eine Komfortzone kann aus mehreren Variablen für jede Situation berechnet werden [SSJ06];
- Geräusche: zu laute Umgebung lenkt ab, macht müde und reizbar und senkt die Arbeitsleistung; sie beeinflusst die Faktoren Konzentration, Performanz/work rate, Müdigkeit (z.B. [WWC02]) und Reizbarkeit [Ma04]; 70% der Mitarbeiter in einer Studie sagen, ihre Produktivität würde verbessert werden, wenn sie weniger

durch Geräusche gestört würden [Ma04]; durch hohe Lautstärke werden ergonomische Möglichkeiten weniger genutzt [EJ00]; der Einfluss der Lautstärke wird beeinflusst durch: u.a. speech intelligibility (STI, Sprachverständlichkeit) [Ma04] [Ve02] [KYE07] und Bedeutung des Geräuschs [Su94], Vorhersagbarkeit und Kontrolle über die Quelle, Variabilität des Geräuschs [Ma04], Alter und Art der kognitiven Aufgabe [Be05].

- **Licht:** Die Lichtverhältnisse am Arbeitsplatz beeinflussen Produktivität, Unfallrate, Performanz, Schlaf, Arbeitsmoral, Krankheitsrate, Aufmerksamkeit, Vitalität, Stimmung [PL00] [ET02] [VBVDB04] und Energiekosten. Am besten für den Menschen sind Tageslicht und Fenster, sie sind wichtig für biologische Funktionen (Hormone, Tagesrhythmus, Vitamin D-Produktion) und beeinflussen u.a. Arbeitsmoral, Laune und Ermüdung der Augen. Je ähnlicher das Licht dem Tageslicht ist, umso positiver ist der Einfluss des Lichts (bezogen auf Helligkeit und Spektrum, vgl. [ET02]). Die Auswirkungen des Lichts werden beeinflusst durch persönliche Präferenz [VBVDB04]. Diese Erkenntnisse finden bereits durch die „Human Centric Lighting“-Initiative den Weg in die Praxis und durchgeführte Marktstudien des Verbands untermauern dabei die Relevanz der Thematik [Ge13].
- **Luftqualität:** Die Luftqualität wird bestimmt durch die Konzentration bestimmter Stoffe in der Luft, besonders durch die „volatile organic compounds“ (VOCs), u.a. CO₂ (siehe z.B. [FDA98]). Die Luftqualität hat starke Auswirkungen auf die Produktivität und das Wohlbefinden [Wy04].
- **Weitere Faktoren:** Auch andere Faktoren haben einen Einfluss auf die Menschen, unter anderem Sauberkeit, verbaute Materialien, „visual comfort“, „sound privacy“, „ease of interaction“, „comfort of furnishing“ und Farben (siehe z.B. [KdD12][SJ09]). Diese sind i. A. nicht manipulierbar und daher hier nicht relevant.
- **Gesundheitsbezogene Faktoren:** Der Aufenthalt in Gebäuden kann durch vielerlei Faktoren gesundheitsschädigende Wirkung haben, beispielsweise zu Irritationen an Auge, Nase, Haut, sowie Asthma, Atemwegserkrankungen im Allgemeinen, Allergien und die Verbreitung ansteckender Krankheiten führen ([Fi00][Ja98][Ch03], vgl. Forschung „Sick Building Syndrome (SBS)“). Diese beeinträchtigen auch die Zufriedenheit. Die meisten Faktoren, die die Gesundheit betreffen, sind nicht durch Aktoren beeinflussbar, trotzdem sollten Sensoren vorhanden sein, die Handlungsbedarf erkennen, z.B. für die Erkennung von Pilzsporen.
Ein weiterer Aspekt, der hier berücksichtigt werden sollte, ist, dass auch die Arbeit an sich krank machen kann, z.B. durch zu starken dauerhaften Stress (z.B. [Wi13]).

Zusätzlich zu den direkten Auswirkungen einzelner Variablen wurden auch Zusammenhänge und Interaktionseffekte untersucht. Beispielsweise moderiert die Lautstärke den Einfluss, den die Temperatur auf die Fehlerrate bei einem Leistungstest hat [WWC02]. Es gibt weiterhin einen Zusammenhang zwischen Temperatur, Gerüchen und Luftqualität: je wärmer es ist, umso stärker werden Gerüche wahrgenommen, und je wärmer, umso schlechter ist die wahrgenommene Luftqualität [WWC02][Wy04]. Auch wurde festge-

stellt, dass die Art der Aufgabe die Bedürfnisse verändert, z.B. in Bezug auf Lautstärke oder Musik [Ma04].

2 Erweiterung der Analyse

Die bisher geschilderten Untersuchungen konzentrieren sich auf die Umgebungsanalyse, unabhängig von den Personen, die sich in der Umgebung aufhalten und ihren Aufgaben, Aktivitäten, Stimmungen und persönlichen Präferenzen. Ein modernes System für die Gestaltung einer Arbeitsumgebung sollte diese Aspekte allerdings berücksichtigen. Es könnte beispielsweise für die Temperaturregelung relevant sein, wie viel die Person sich bewegt, oder für die Gestaltung des Umfelds insgesamt, welche Art von Aufgabe sie gerade erledigt. Die Stimmung und Emotionen einer Person sind ebenfalls relevante Parameter. Untersuchungen haben ergeben, dass Stimmung die kognitiven Prozesse („cognitive style“) und auch messbar die Arbeitsleistung beeinflusst. Leicht positive Stimmung sorgt für effektiveres und flexibleres Denken [SJ09]. Lisetti und Nasoz [LN04] schreiben, dass gesteigerte Emotionsintensität auch die Performanz steigere, aber nur bis zu einem bestimmten Punkt. Emotionen sind insgesamt stark involviert in Aufmerksamkeit, Wahrnehmung, Lernen, Gedächtnisprozesse, Problemlösen, Entscheidungen treffen, kreatives Denken und menschliche Interaktion (vgl. [AS05]). Eine Umgebung, die sich günstig auf die Stimmung auswirkt, ist also sowohl für Zufriedenheit als auch für die Produktivität relevant. Ein besonders wichtiger Aspekt des inneren Zustands einer Person am Arbeitsplatz ist das Stresslevel, das sie empfindet. Stress ist das zweithäufigste arbeitsbezogene Gesundheitsproblem in der EU [MAT13] und kann ernste Gesundheitsprobleme verursachen [EJ00][Wi13]. Durch die enge Verbindung physiologischer Reaktionen mit dem erlebten Stress ist es möglich, über Sensoren das Stresslevel mit sehr hoher Genauigkeit festzustellen [Sa11]. Um diese „Biosignale“ zu analysieren, werden spezielle Sensoren und Analysemethoden gebraucht. Der Trend der letzten Jahre hin zu immer kleineren, immer weiter verbreiteten „wearables“ ermöglicht eine immer komfortablere und unauffälligere Ermittlung physiologischer Parameter.

2.1 Kontrolle

Zusätzlich zu den genannten Parametern darf ein weiterer Aspekt nicht vernachlässigt werden. Mitarbeiter haben das Bedürfnis, Kontrolle über ihre Arbeitssituation zu haben. Dazu gehört auch die Kontrolle über das direkte physikalische Umfeld. Die Zunahme automatisierter Funktionen intelligenter Gebäude birgt die Gefahr, dass die Kontrolle des Einzelnen abnimmt und sich die Mitarbeiter möglicherweise bevormundet fühlen. Nach weit verbreiteter Ansicht beeinflusst der Umfang der Kontrolle, die eine Person über den Job hat, die Arbeitsleistung, den empfundenen Stress und die Gesundheit [HRC04]. Es wurde sogar ein signifikanter Zusammenhang zwischen einem niedrigen Level an Kontrolle und der Wahrscheinlichkeit, an kardiovaskulären Erkrankungen zu sterben, gefunden [Jo96]. Ein Aspekt der „job control“ ist die Kontrolle über die physikalische Umgebung und darin enthaltene externe Stressoren wie Geräusche. Dies wird bei Huang et al. [HRC04] als

„environmental control“ bezeichnet. Sie schreiben, Kontrolle über die Arbeitsumgebung könne positive Effekte auf die Gesundheit und Effektivität der Arbeiter haben. Umgebungskontrolle trage laut Untersuchungen direkt zu Arbeitsplatzzufriedenheit und Leistung bei. Indirekt beeinflusse sie diese Parameter, indem sie die Wahrnehmung verändere, z.B. von Ablenkungen, Privatsphäre, Stress und Kommunikation. Bei dem Entwurf eines umfassenden Systems mit vielen automatischen Funktionen sollte also der Aspekt der persönlichen Kontrolle berücksichtigt werden. Nicht nur bei Fehlfunktionen des Systems ist dies wichtig. Reijula et al. [Re11] betonen, dass wenn die Umgebung zu viel autonome Aktivität zeige, dies den Stress der Person vergrößern könne statt zu verringern. Es müsse daher eine übergeordnete Funktion („override“) geben, sodass der Benutzer die intelligente Umgebung abschalten könne, falls er von der Interaktion mit der Umgebung in irgendeiner Form irritiert sei. Laut Leaman und Bordass [LB99] wird die wahrgenommene Kontrolle über das eigene Umfeld beeinflusst durch: tatsächliche Kontrollmöglichkeiten, mögliches Feintuning durch den Benutzer sowie die Reaktionszeit des Systems.

3 Vernetzung und Aktorik

Basis für die Nutzung der wissenschaftlichen Erkenntnisse zur Steigerung des Wohlbefindens und der Produktivität innerhalb intelligenter Umgebungen ist ein System aus zahlreichen Sensoren und Aktoren. Reijula et al. [Re11] geben eine Definition für den Begriff „intelligent environment“ und nennen wichtige Kriterien, die eine intelligente Arbeitsumgebung erfüllen sollte: die Interaktion zwischen Umgebung und Benutzern solle erfasst, verarbeitet und im Kontext interpretiert werden; im Sinne der Benutzerbedürfnisse solle darauf reagiert werden; sie solle aktiv, autonom und allgegenwärtig sein und den Arbeitsfluss des Benutzers unterstützen, sowie die Wahrnehmung des Benutzers seines körperlichen und psychologischen Wohlbefindens verbessern [Re11, S. 234]. Auch bezieht sich die Definition nicht nur auf die physikalischen Aspekte, sondern betont, dass eine Wahrnehmung und Reaktion auf psychologische Aspekte wichtig ist. Ein weiteres gutes Beispiel für die Notwendigkeit eines komplexeren Bilds der Situation beschreiben Ramos et al. [RHD14]. Stresserkennung basiert darauf, dass das Stresslevel physiologische Vorgänge beeinflusst, wie z.B. eine Steigerung des Pulses. Es muss aber unbedingt bedacht werden, dass auch andere Dinge die gleichen Parameter beeinflussen, besonders die körperliche Aktivität. Wenn eine Person schnell geht, verändert sich der Puls, aber deswegen liegt noch kein Stress vor. Ramos et al. versuchen daher, Aktivitätserkennung mit Stressmonitoring zu verbinden und die Einflüsse so zu kontrollieren. Ein Ziel der weiteren Forschung sollte also sein, die Fülle an Informationen sinnvoll zu vernetzen.

Mithilfe der IEQ-Forschungsergebnisse kann ein konkreter Sollzustand für jede Variable formuliert werden und es gibt Manipulationsmöglichkeiten, um diesen Zustand zu erreichen: z.B. können die Lichtverhältnisse über spezielle Lampen gesteuert werden oder die Geräuschkulisse durch Maskierungsgeräusche oder im Vorfeld durch die Architektur des Arbeitsplatzes verbessert werden. Auch der innere Zustand des Menschen ist durch Sensorik erfassbar, hier gibt es aber kaum konkrete Handlungsanweisungen, wie auf diese Informationen reagiert werden soll. Ein Ansatz ist, die Musik in einer Stresssituation leiser zu machen oder unwichtige Nachrichten abzuschalten [Ha04, HP05]. Hier entstehen also neue Forschungsfragen, deren Betrachtung durch umfangreiche Sensorsysteme

möglich wird. Für die zahlreichen Parameter, die den inneren Zustand oder die Aktivität des Menschen betreffen, müssen erst Möglichkeiten der Einflussnahme erforscht und definiert werden. Sie hängen natürlich stark von der individuellen Zielsetzung ab.

Im Rahmen des Affective Computing wird vorgeschlagen, die Emotionen aktiv durch Maßnahmen zu beeinflussen, um positive Emotionen zu fördern. Ein System, das die Emotionen erkennen kann, könnte den Einfluss von Emotionen nutzen, indem es Stimuli präsentiert, die erwünschte Stimmungen aufrechterhalten oder alternativ unerwünschten entgegenwirken. Frustrierte Benutzern könnten beispielsweise zu einer anderen Aufgabe oder einem anderen Aspekt der Arbeit geleitet werden, oder es könnte ihnen zu einer Pause geraten werden [Zi03, S. 2].

4 Kenetic Software Design in intelligenten Umgebungen

Innerhalb der beschriebenen intelligenten Umgebungen stellt die technische Kontrolle dieser komplexen Sensor-Aktor-Netze eine nicht minder schwere Aufgabe dar. Diese müssen sich auf der einen Seite an die Bedingungen der Umgebung anpassen (vgl. Absatz 1.1) und auf der anderen Seite die implizit oder explizit geäußerten Präferenzen des Nutzers beachten (vgl. Absatz 2.1) und so gut wie möglich erfüllen. Hierbei bietet sich die Modellierung als Multiagentensystem (MAS) und die Umsetzung mittels des durch Ferber in [Fe99] beschriebenen Kenetic Software Design an. Ferber bezieht hierbei auf Systeme, welche durch Interaktion, Adaption und Reproduktion evolvieren [Fe99]. Die Umgebung kann dabei beschrieben werden als eine endliche Menge E aus diskreten, momentanen Zuständen:

$$E = \{e, e', \dots\}.$$

Die einzelnen Komponenten besitzen die Möglichkeit, die Umgebung auf Basis der ermittelten Daten und übereinstimmend mit vorgegebenen Zielen zu transformieren. Hierzu besitzt jeder Aktor eine endliche Menge an Aktionen. Sei

$$Ac = \{\alpha, \alpha', \dots\}.$$

die endliche Menge an Aktionen, welche durch den einzelnen Aktor durchgeführt werden können. Dann ist die typische Sequenz einer Interaktion mit der Umgebung innerhalb eines Sensor-Aktor-Netztes eine alternierende Folge von verzahnten Umgebungszuständen und darauf basierenden Aktionen:

$$s : e_0 \xrightarrow{\alpha_0} e_1 \xrightarrow{\alpha_1} e_2 \xrightarrow{\alpha_2} e_3 \xrightarrow{\alpha_3} \dots \xrightarrow{\alpha_{u-1}} e_u.$$

wobei R die Menge aller möglichen endlichen Sequenzen über E und Ac ist, R^{Ac} die Teilmenge der Sequenzen, welche mit einer Aktion $\alpha \in Ac$ enden und R^E die Teilmenge der Sequenzen, welche in einem Zustand $e \in E$ enden. Der Effekt, welcher sich auf die Umgebung durch die Ausführung von Aktionen ergibt, wird dargestellt durch eine *state transformer function* [Fa95]:

$$\tau : R^{Ac} \rightarrow \rho(E),$$

wobei $\rho(\cdot)$ ebendiese Funktion darstellt. Formal lässt sich eine Umgebung Env nun als ein Tripel $Env = \langle E, e_0, \tau \rangle$ beschreiben, wobei E die Menge der Umgebungsstatus ist, $e_0 \in E$ der Initialzustand ist und τ die *state transformer function* ist. Somit lässt sich eine Kombination aus Sensoren und Aktoren als *Agent* innerhalb des Systems auffassen. Dieser

Agent lässt sich als Funktion, welche Sequenzen ausführt, modellieren [Wo09][RS95]:

$$Ag : R^E \rightarrow Ac.$$

Der Agent wählt also eine entsprechende Aktorkonfiguration auf Basis der Umgebungszustandshistorie. Hierbei gilt die Annahme, dass, obwohl Umgebungen nicht-deterministisch sind, die entsprechenden Agenten deterministische Entscheidungen treffen [Wo09]. Betrachtet man die heutigen vernetzten Umgebungen als eine Menge von Sensoren und Aktoren, wobei vertikale Kombinationen wiederum einen Agenten bilden [Fo83] (siehe Abbildung 1), ist AG die Menge aller Agenten innerhalb einer intelligenten Umgebung. Die derzeit existierenden Umgebungen bestehen auf Basis dieser Annahme häufig nur aus *reaktiven*, manchmal auch als *tropistisch* (von griechisch $\tau\rho\epsilon\pi\omicron$, "wende") [GN87] bezeichneten, Agenten, welche Perzeptionen auf direktem Wege in Aktionen umwandeln:

$$Ag : E \rightarrow Ac.$$

Als Beispiel hierfür lässt sich die Lichtsteuerung eines Raumes anführen. Hier werden typischerweise zwei Sensoren verwendet - Lichtsensor und Lichtschalter. Ein reaktiver Agent besteht dabei aus diesen beiden Sensoren und einem Aktor - der Lichtsteuereinheit -, welche folgendermaßen definiert werden kann:

$$Ag(e) = \begin{cases} \text{Licht an} & \text{falls (Helligkeit} = \neg OK) \vee (\text{Lichtschalter} = \text{ON}) \\ \text{Licht aus} & \text{sonst.} \end{cases}$$

Modelliert werden kann dieser Agent mittels einer Funktion $percept : E \rightarrow Per$, welche eine Zuweisung von einer Menge an Umgebungsstatus E zu einer nicht-leeren Menge an Perzeptionen Per , typischerweise mittels der Sensoren, vornimmt und einer Funktion $action : Per^* \rightarrow Ac$, welche einer Sequenz an Perzeptionen eine Menge an Aktionen zuordnet. Der Nutzer interagiert in diesem Fall nur indirekt mit dem System und gibt keine oder nur wenig direkte Anweisungen. Ferber [Fe99] nutzt hierfür den Vergleich des Schäfers und seiner Schafherde, welcher Aktionen der Individuen (Agenten) nur durch sein, mittels der Sensoren der Schafe aufgenommenes, Verhalten auslöst. Er nennt dieses Beispiel unter Bezug auf *Kenetic Software Design* und konkretisiert damit das Konzept der Nutzung von Agenten und deren Interaktion zur Lösung von Problemen in komplexen Systemen. Der Einsatz von MAS in drahtlosen bzw. drahtgebundenen Sensor-Aktor-Netzen ist durchaus kein neuer Lösungsansatz. So nutzen Sandhu et al. [SAA04] verteiltes Lernen in MAS zur intelligenten und adaptiven Steuerung der Beleuchtung in Gebäuden. Boman et al. [BDY99] realisieren ein MAS mit komplexen Agenten zur Wärmeregulierung in Gebäuden auf Basis von Entscheidungsbäumen und *Influence Diagrams* (IDs), wobei das System an sich nicht evolviert, sondern auf Basis von definierten Plänen mit *a-priori*-Informationen agiert. Hagrais et al. [Ha03] studieren das Lernen und Adaptieren in Agenten-orientierten autonomen Architekturen am Beispiel eines integrierten Gebäudesystems, wobei sie Autonomie als die Fähigkeit beschreiben sich zu adaptieren und eigene Regeln zu erzeugen [Ha03]. Wang et al. [WYW10] adressieren in ihrem Paper ausdrücklich das Konfliktpotenzial zwischen Energieeffizienz des Gebäudes und dem gewünschten Komfortlevel des Nutzers. Sie nutzen für die Lösung des Problems ein System aus miteinander verhandelnden zentralen und lokalen Agenten. Die zentralen Agenten fokussieren dabei die Energieaufnahme und die lokalen Agenten die jeweiligen

Bedürfnisse des Nutzers. Um die zugrundeliegenden nicht-linearen Systeme bestmöglich an die Nutzerwünsche zu adaptieren, ist der Controller mit einer auf Fuzzylogik basierenden Kontrolleinheit ausgestattet. In [WYW11] wird dieses System zur Nutzung in Smart-Grids (SGs) erweitert. Zur Optimierung des Systems wird eine Abwandlung des Partikelschwarmoptimierungsalgorithmus (PSO) von Kennedy und Eberhart [KE95] angewandt, welches sich in der Arbeit durch seine hohe Konvergenzgeschwindigkeit bezüglich globaler Extrema, bei gleichzeitiger Vermeidung lokaler Extrema, auszeichnet.

4.1 Kontrolle in nutzerzentrierten intelligenten Umgebungen

Problematisch an dieser Form der Selbst-Organisation bzw. Selbst-Optimierung ist die fehlende Möglichkeit des Nutzers auf die Entscheidungen und Regeln des Systems expliziten Einfluss zu nehmen. Daher wird im Bereich des Organic Computing (OC)

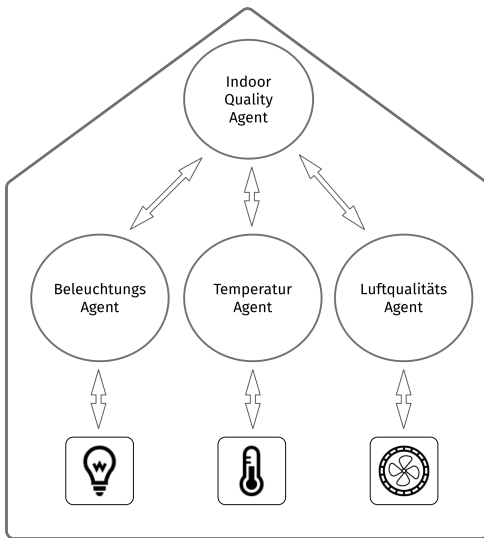


Abb. 1: MAS in intelligenten Gebäuden

der Begriff der kontrollierten Selbst-Organisation eingeführt [Sc10]. Hierbei handelt es sich um die Möglichkeit, ungewolltes emergentes Verhalten innerhalb des organischen Systems zu beheben und somit letztendlich dem Nutzer die volle Kontrolle über das System zu gewährleisten [Sc10]. Eine typischerweise hierfür verwendete Architektur ist die *Observer/Controller-Architektur* [SMS05][MS04][Ri06]. Hierbei wird das häufig als zentralistisch angenommene System von Sensoren und Aktoren dezentralisiert und somit in diverse Subsysteme aufgeteilt. Diese Subsysteme oder auch *Systems under Observation/Control* (SuOC) werden innerhalb der Architektur einer darüberliegenden Monitoring und Kontrollschicht untergeordnet. Der Obser-

ver überwacht und aggregiert die Informationen (*Situations Parameter*) und trifft Vorhersagen über das zukünftige Verhalten des Systems auf Basis eines definierten Observationsmodells. Sollten die Ergebnisse nicht mit dem gewünschten Verhalten des Systems übereinstimmen, greift der Controller in das Verhalten der einzelnen SuOCs ein und versucht dieses zu optimieren [Ri06]. Diese Form der Kontrolle in technischen Systemen wurde bereits erfolgreich erprobt, bspw. in der Verkehrsflusskontrolle [Pr11] oder der Robotik [Mö06].

5 Status Quo und Vision eines intelligenten Gebäudes

Beide beschriebenen Beispiele sowie das Beispiel der intelligenten Umgebung haben zwei Aspekte gemein: (1) Dezentralisierung eines zentralistischen Systems und (2) Hetero-

genität bezüglich der technischen Komponenten. Und obwohl moderne *Building Automation Systems* (BAS) bereits durch eine verteilte Architektur gekennzeichnet sind, bilden sie in der Praxis häufig ein unflexibles monolithisches System. Komponenten werden installiert und aufwendig konfiguriert [Di10]. Ein geeignetes Applikationsmodell für erweiterte Logik bzw. Intelligenz des Gebäudes fehlt [PGK09]. Praus et al. [PGK09] schlagen für die Lösung des Problems einen erweiterten Applikationslayer vor, welcher Entwicklern die Möglichkeit gibt, vordefiniertes Verhalten einfacher in das System einzubringen. Diese Form der Kontrolle hat zwei Nachteile: Zum einen wird die Komplexität des Systems nicht abstrahiert, sondern transformiert, d.h. der Entwickler muss sich auf einer höheren Ebene mit dieser Problematik beschäftigen, und zum anderen evolviert dieses System nicht auf Basis der Wünsche und Präferenzen des Nutzers, sondern maximal auf Basis der Regeln des Entwicklers. Wir schlagen daher eine Architektur auf Basis der in Kapitel 4 diskutierten MAS-Perspektive vor (siehe Abbildung 1). Subsysteme bestehend aus Sensor-Aktor-Konfigurationen werden als modulare regelbasierte Agenten realisiert. Ziele und Regeln werden auf höchster Ebene symbolisch definiert. Hierzu werden die Erkenntnisse aus Kapitel 4.1 miteingebunden (siehe Abbildung 2). Innerhalb der verschachtelten SuOCs kann sowohl mit symbolischen als auch mit subsymbolischem Wissen gearbeitet werden, um beispielsweise die Helligkeit innerhalb eines Raumes zu adaptieren. Hieraus ergibt sich, dass der spätere Nutzer des Systems und seine Umgebung die Aktionen des Systems explizit und implizit beeinflussen kann und somit die geforderte Kontrolle behält. Die vorgeschlagene Architektur wirft dabei einige zukünftige Forschungsfragen auf, welche teilweise mit den in der Forschung zu intelligenten technischen Systemen im Allgemeinen und organischen technischen Systemen im Speziellen genannten Forschungsfragen übereinstimmen. Neben den Fragen auf Ebene der Spezifikation müssen unter anderem die Fragen nach der Kooperation zwischen den Agenten und der Interaktion zwischen dem MAS und dem Nutzer adressiert werden. Weiterhin müssen Möglichkeiten der Selbst-Konfiguration und Optimierung untersucht werden, sowie Maßnahmen und Techniken zur Stabilisierung des Systems. Es muss also die Frage verfolgt werden, inwieweit das integrierte Gebäudesystem ausgefallene Komponenten kompensieren kann. Im Hinblick auf Datenschutz und Sicherheit des Systems sollten auch Maßnahmen zur autonomen Abwehr von Angriffen und Fehlern erforscht und entwickelt werden.

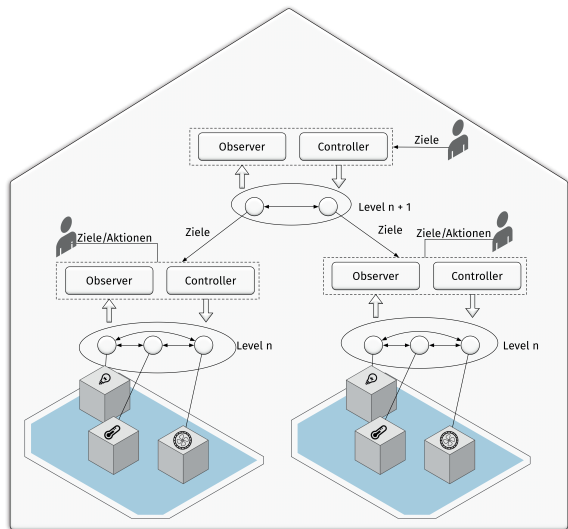


Abb. 2: Vision des Gebäudes der Zukunft

6 Zusammenfassung und Ausblick

Vernetzte Umgebungen werden immer mehr Einzug in unseren Alltag erhalten. Die Gebäude, in denen wir leben und arbeiten, sollten in Zukunft nicht nur energetisch bestmöglich agieren, sondern auch im Hinblick auf die Bedürfnisse der darin lebenden Menschen optimiert werden. Die vorliegende Arbeit gibt für die dafür nötige Sensorik eine zusammenfassende Darstellung der wissenschaftlich relevanten Parameter sowie Faktoren und diskutiert die Auswirkung auf die Arbeitsleistung und das Wohlbefinden der darin lebenden und arbeitenden Personen. Des Weiteren wird die Relevanz der Kontrolle über diese Systeme und die Autonomie der Menschen innerhalb dieser Umgebungen in den Fokus gerückt und erörtert. Hierbei wird festgehalten, dass obwohl die Steuerung und Verwaltung vernetzter Umgebungen und Gebäude bereits einen gewissen Grad der Reife erreicht haben, die Konzepte zur Kontrolle intelligenter Umgebungen noch unausgeprägt sind. Methoden und Techniken aus dem Bereich des *Autonomic* bzw. *Organic Computing* bieten dabei Möglichkeiten und Vorschläge, um Intelligenz und die damit zusammenhängende Selbst-Organisation dieser Systeme beherrschbar und durch den Anwender kontrollierbar zu gestalten. Als Basis hierfür wird die Nutzung von *Multiagentensystemen* als konzeptioneller Rahmen vorgeschlagen und an Beispielen erläutert. Für den Ausblick lässt sich festhalten, dass neben den bereits andiskutierten konkreten technologischen Fragestellungen zur Nutzung von Methoden der künstlichen Intelligenz in technisch-physikalischen Systemen, auch die Erforschung und Beobachtung der Nutzerakzeptanz eine gravierende Rolle für die Etablierung solcher Systeme im Alltag spielen wird. Hierfür müssen die adressierten Personen ein subjektives Relevanzempfinden entwickeln und die Systeme sowie die damit zusammenhängende Sensorik und Aktorik als sinnvoll und unterstützend wahrnehmen. Unabdingbar hierfür ist die interdisziplinäre Zusammenarbeit zwischen den Fachgebieten der Architektur, des Bauingenieurwesens, der Psychologie sowie der Informatik und Elektrotechnik, um diese Fragen auf Basis eines wissenschaftlichen Syntheseprozesses zu beantworten. Teile dieser Fragen werden im Rahmen zukünftiger Forschungsprojekte und auf Basis der hier dargestellten Ergebnisse adressiert. Eine Umsetzung und Evaluation der vorgeschlagenen Architektur ist geplant.

Literaturverzeichnis

- [AS05] Anttonen, Jenni; Surakka, Veikko: Emotions and heart rate while sitting on a chair. In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, S. 491–499, 2005.
- [BB98] Banbury, Simon; Berry, Dianne C.: Disruption of office-related tasks by speech and office noise. *British Journal of Psychology*, 89(3):499–517, 1998.
- [BDY99] Boman, Magnus; Davidsson, Paul; Younes, Håkan L.: Artificial Decision Making Under Uncertainty in Intelligent Buildings. In: Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence. UAI'99, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, S. 65–70, 1999.
- [Be05] Beaman, C. Philip: Auditory distraction from low-intensity noise: a review of the consequences for learning and workplace environments. *Applied Cognitive Psychology*, 19(8):1041–1064, 2005.

- [Ch03] Chao, H Jasmine; Schwartz, Joel; Milton, Donald K; Burge, Harriet A: The work environment and workers' health in four large office buildings. *Environmental Health Perspectives*, 111(9):1242, 2003.
- [CN89] Crouch, Andrew; Nimran, Umar: Perceived facilitators and inhibitors of work performance in an office environment. *Environment and Behavior*, 21(2):206–226, 1989.
- [Di10] Dietrich, D.; Bruckner, D.; Zucker, G.; Palensky, P.: Communication and Computation in Buildings: A Short Introduction and Overview. *IEEE Transactions on Industrial Electronics*, 57(11):3577–3584, Nov 2010.
- [EJ00] Evans, Gary W.; Johnson, Dana: Stress and open-office noise. *Journal of Applied Psychology*, 85(5):779–783, 2000.
- [ET02] Edwards, L; Torcellini, Paul A: A literature review of the effects of natural light on building occupants. National Renewable Energy Laboratory Golden, CO, 2002.
- [Fa95] Fagin, R.; Vardi, M.; Moses, Y.; Halpern, J. Y.: Reasoning About Knowledge. The MIT Press, Cambridge, MA, 1995.
- [FDA98] Fisk, William J; De Almeida, Anibal T: Sensor-based demand-controlled ventilation: a review. *Energy and buildings*, 29(1):35–45, 1998.
- [Fe99] Ferber, Jacques: Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999.
- [Fi00] Fisk, William J: Review of health and productivity gains from better IEQ. Lawrence Berkeley National Laboratory, 2000.
- [Fo83] Fodor, J.A.: The Modularity of Mind: An Essay on Faculty Psychology. The MIT Press, Cambridge, MA, 1983.
- [Ge13] German Electrical and Electronic Manufacturers Association (ZVEI): Human Centric Lighting: Going Beyond Energy Efficiency. http://www.lightingeurope.org/uploads/files/Market_Study-Human_Centric_Lighting_Final_July_2013.pdf, Juli 2013. [Online; Zugriff 01.05.2016].
- [GN87] Genesereth, Michael R.; Nilsson, Nils J.: Logical Foundations of Artificial Intelligence. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1987.
- [Ha03] Hagra, Hani; Callaghan, Victor; Colley, Martin; Clarke, Graham: A Hierarchical Fuzzy-genetic Multi-agent Architecture for Intelligent Buildings Online Learning, Adaptation and Control. *Information Sciences - Informatics and Computer Science: An International Journal - Special issue on recent advances in soft computing*, 150(1-2):33–57, Marz 2003.
- [Ha04] Haag, Andreas; Goronzy, Silke; Schaich, Peter; Williams, Jason: Emotion recognition using bio-sensors: First steps towards an automatic system. In: ADS. Springer, S. 36–48, 2004.
- [HP05] Healey, Jennifer A; Picard, Rosalind W: Detecting stress during real-world driving tasks using physiological sensors. *Intelligent Transportation Systems, IEEE Transactions on*, 6(2):156–166, 2005.
- [HRC04] Huang, Yueng-Hsiang; Robertson, Michelle M; Chang, Kuo-I: The role of environmental control on environmental satisfaction, communication, and psychological stress effects of office ergonomics training. *Environment and Behavior*, 36(5):617–637, 2004.

- [Ja98] Jaakkola, Jouni J. K.: The Office Environment Model: A Conceptual Analysis of the Sick Building Syndrome. *Indoor Air*, 8(S4):7–16, 1998.
- [Jo96] Johnson, Jeffrey V; Stewart, Walter; Hall, Ellen M; Fredlund, Peeter; Theorell, Tor: Long-term psychosocial work environment and cardiovascular mortality among Swedish men. *American Journal of Public Health*, 86(3):324–331, 1996.
- [Ju01] Judge, Timothy A; Thoresen, Carl J; Bono, Joyce E; Patton, Gregory K: The job satisfaction–job performance relationship: A qualitative and quantitative review. *Psychological bulletin*, 127(3):376, 2001.
- [KdD12] Kim, Jungsoo; de Dear, Richard: Nonlinear relationships between individual IEQ factors and overall workspace satisfaction. *Building and Environment*, 49(3):3e40, 2012.
- [KE95] Kennedy, J.; Eberhart, R.: Particle swarm optimization. In: *Proceedings of the IEEE International Conference on Neural Networks*. Jgg. 4, S. 1942–1948, Nov 1995.
- [KYE07] Kitapci, Kivanc; Yilmazer, Semiha; Erkip, Feyzan: Effect of speech intelligibility on visual short-term memory performance. In: *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*. Jgg. 2007. Institute of Noise Control Engineering, S. 4370–4380, 2007.
- [LB99] Leaman, Adrian; Bordass, Bill: Productivity in buildings: the „killer“ variables. *Building Research & Information*, 27(1):4–19, 1999.
- [LN04] Lisetti, Christine Lætitia; Nasoz, Fatma: Using noninvasive wearable computers to recognize human emotions from physiological signals. *EURASIP Journal on Advances in Signal Processing*, 2004(11):1–16, 2004.
- [Ma04] Mardex, Justin: Auditory, visual, and physical distractions in the workplace. Cornell University, Department of Design and Environmental Analysis, [online] <http://www.justinmardex.com/media/jrmavpdpdf.pdf>, 2004.
- [MAT13] Muaremi, Amir; Arnrich, Bert; Tröster, Gerhard: Towards measuring stress with smartphones and wearable devices during workday and sleep. *BioNanoScience*, 3(2):172–183, 2013.
- [Mö06] Mösch, Florian; Litza, Marek; El Sayed Auf, Adam; Maehle, Erik; Großpietsch, Karl E.; Brockmann, Werner: Self-Organizing Systems: First International Workshop, IWSOS 2006, and Third International Workshop on New Trends in Network Architectures and Services, EuroNGI 2006, Passau, Germany, September 18-20, 2006 Proceedings. Springer Berlin Heidelberg, Berlin, Heidelberg, Kapitel ORCA – Towards an Organic Robotic Control Architecture, S. 251–253, 2006.
- [MS04] Müller-Schloer, C.: Organic computing - on the feasibility of controlled emergence. In: *Proceedings of the International Conference on Hardware/Software Codesign and System Synthesis*. S. 2–5, Sept 2004.
- [MZC12] Mehta, Ravi; Zhu, Rui Juliet; Cheema, Amar: Is noise always bad? Exploring the effects of ambient noise on creative cognition. *Journal of Consumer Research*, 39(4):784–799, 2012.
- [OPS09] Oswald, Andrew J; Proto, Eugenio; Sgroi, Daniel: Happiness and productivity. Bericht, IZA discussion papers, 2009.
- [PGK09] Praus, F.; Granzer, W.; Kastner, W.: Enhanced control application development in Building Automation. In: *Proceedings of the 7th IEEE International Conference on Industrial Informatics*. S. 390–395, June 2009.

- [PL00] Partonen, Timo; Lönnqvist, Jouko: Bright light improves vitality and alleviates distress in healthy people. *Journal of Affective disorders*, 57(1):55–61, 2000.
- [Pr11] Prothmann, Holger; Tomforde, Sven; Branke, Jürgen; Hähner, Jörg; Müller-Schloer, Christian; Schmeck, Hartmut: *Organic Computing — A Paradigm Shift for Complex Systems*. Springer Basel, Basel, Kapitel Organic Traffic Control, S. 431–446, 2011.
- [Re11] Reijula, Jori; Gröhn, Matti; Müller, Kiti; Reijula, Kari: Human well-being and flowing work in an intelligent work environment. *Intelligent Buildings International*, 3(4):223–237, 2011.
- [RHD14] Ramos, Julian; Hong, Jin-Hyuk; Dey, Anind K: Stress Recognition-A Step Outside the Lab. In: *PhyCS*. S. 107–118, 2014.
- [Ri06] Richter, Urban; Mnif, Moez; Branke, Jürgen; Müller-Schloer, Christian; Schmeck, Hartmut: Towards a generic observer/controller architecture for Organic Computing. In (Hochberger, C.; Liskowsky, R., Hrsg.): *Proceedings of Informatik für Menschen!* (INFORMATIK 06). Jgg. P-93, S. 112–119, 2006.
- [RS95] Russell, Stuart J.; Subramanian, Devika: Provably Bounded-optimal Agents. *Journal of Artificial Intelligence Research*, 2(1):575–609, Januar 1995.
- [Sa11] de Santos Sierra, Alberto; Ávila, Carmen Sánchez; Casanova, Javier Guerra; Pozo, Gonzalo Bailador Del: A stress-detection system based on physiological signals and fuzzy logic. *Industrial Electronics, IEEE Transactions on*, 58(10):4857–4865, 2011.
- [SAA04] Sandhu, Jaspal S.; Agogino, Alice M.; Agogino, Adrian K.: Wireless Sensor Networks for Commercial Lighting Control: Decision Making with Multi-agent Systems. In: *Proceedings of the Association for the Advancements of Artificial Intelligence (AAAI) Workshop on Sensor Networks*. S. 131–140, 2004.
- [Sc10] Schmeck, Hartmut; Müller-Schloer, Christian; Çakar, Emre; Mnif, Moez; Richter, Urban: Adaptivity and Self-organization in Organic Computing Systems. *ACM Transactions on Autonomous and Adaptive Systems*, 5(3):10:1–10:32, September 2010.
- [SFL06] Seppänen, Olli; Fisk, William J; Lei, QH: Effect of temperature on task performance in office environment. Lawrence Berkeley National Laboratory, 2006.
- [SJ09] Sears, Andrew; Jacko, Julie A: *Human-Computer Interaction Fundamentals*. CRC Press, 2009.
- [SMS05] Schöler, Thorsten; Müller-Schloer, Christian: In (Beigl, Michael; Lukowicz, Paul, Hrsg.): *Proceedings of the 18th International Conference on Architecture of Computing Systems*. Innsbruck, Austria, S. 139–153, 2005.
- [SSJ06] Saberi, Ommid; Saneei, Parisa; Javanbakht, Amir: Thermal comfort in architecture. *WINDSOR 2006: Comfort and Energy Use in Buildings-Getting them right*, 2006.
- [Su94] Sundstrom, E.; Town, J. P.; Rice, R. W.; Osborn, D. P.; Brill, M.: Office Noise, Satisfaction, and Performance. *Environment and Behavior*, 26(2):195–222, 1994.
- [VBVDB04] Van Bommel, WJM; Van Den Beld, GJ: Lighting for work: a review of visual and biological effects. *Lighting Research & Technology*, 36(4), 2004.
- [Ve02] Veitch, Jennifer A; Bradley, John S; Legault, Louise M; Norcross, Scott; Svec, Jana M: Masking speech in open-plan offices with simulated ventilation noise: noise level and spectral composition effects on acoustic satisfaction. Institute for Research in Construction, Internal Report IRC-IR-846, 2002.

- [Wi13] Wijsman, Jacqueline; Grundlehner, Bernard; Liu, Hao; Penders, Julien; Hermens, Hermie: Wearable physiological sensors reflect mental stress state in office-like situations. In: *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*. IEEE, S. 600–605, 2013.
- [Wo09] Wooldridge, Michael: *An Introduction to MultiAgent Systems*. Wiley Publishing, 2nd. Auflage, 2009.
- [WWC02] Witterseh, Thomas; Wyon, David; Clausen, Geo: The effects of moderate heat stress and open-plan office noise distraction on office work. In: *Proceedings of Indoor Air 2002*. 2002.
- [Wy04] Wyon, DP: The effects of indoor air quality on performance and productivity. *Indoor air*, 14(s7):92–101, 2004.
- [WYW10] Wang, Zhu; Yang, Rui; Wang, Lingfeng: Multi-agent intelligent controller design for smart and sustainable buildings. In: *Proceedings of the 4th IEEE Systems Conference*. S. 277–282, April 2010.
- [WYW11] Wang, Z.; Yang, R.; Wang, L.: Intelligent multi-agent control for integrated building and micro-grid systems. In: *Proceedings of the IEEE Conference on Innovative Smart Grid Technologies (ISGT)*. S. 1–7, Jan 2011.
- [Zi03] Zimmermann, Philippe; Guttormsen, Sissel; Danuser, Brigitta; Gomez, Patrick: Affective computing - a rationale for measuring mood with mouse and keyboard. *International journal of occupational safety and ergonomics*, 9(4):539–551, 2003.

Individualisiertes Gruppentraining mit Datenbrillen für die Produktion

Sarah Brauns¹ Tobias Käfer² Dirk Koriath³ Andreas Harth⁴

Abstract: In großen Unternehmen sowie im Mittelstand wird heutzutage versucht, komplexe interne Prozesse durch Unterstützung von Informationstechnologien zu vereinfachen. Ein Beispiel dafür ist das Training der Mitarbeiter, die direkt in den Produktionsprozess eingebunden sind, da die Vermittlung von Arbeitsinhalten an diese Mitarbeiter aufgrund vielfältiger Produktvarianten immer schwieriger wird. Im vorliegenden Beitrag beschreiben wir ein System für das Training von Montageumfängen mithilfe von virtuellen Techniken, genauer Datenbrillen. Der Einsatz von Datenbrillen ermöglicht es, den am Training teilnehmenden Mitarbeitern benötigte Informationen direkt in ihr Blickfeld einzublenden. Im Gegensatz zu bestehenden Arbeiten erlaubt unser System die Durchführung des Trainings mit mehreren Trainern und Trainingsteilnehmern. Trainer können mittels Tablet den Fortschritt der einzelnen Trainingsteilnehmer verfolgen und flexibel auf die Trainierenden reagieren, z.B. indem Trainer Inhalte unmittelbar anpassen. Des Weiteren sind die Trainingsinhalte so gestaltet, dass die Teilnehmer zwischen verschiedenen Schwierigkeitsgraden wählen können. Wir beschreiben die verteilte Systemarchitektur des aktuellen Demonstrators, wobei auf Komponenten mittels Web-Protokollen zugegriffen wird und Daten mittels Semantic-Web-Technologien ausgezeichnet werden. Außerdem stellen wir die Ergebnisse der Evaluierung einer frühen Version des Demonstrators in Grundzügen vor, bei der die Tauglichkeit des Trainingskonzepts sowie die Beanspruchung der Probanden getestet wurde.

1 Einleitung

Durch die zunehmende Einsicht in die Vorteile der Digitalisierung versuchen Unternehmen, neue Medien in bestehende Prozesse zu integrieren oder mithilfe moderner Technologien neue, effiziente Prozessformen zu schaffen [HK14]. In diesem Bereich stellt gerade der Trainingsprozess eine Herausforderung in produzierenden Unternehmen dar. Es ist schwierig, auf die individuellen Bedürfnisse abgestimmte Gruppentrainings von Prozessen durchzuführen, wenn Mitarbeiter mit verschiedensten Kenntnisständen in einem starren Trainingskonzept die gleichen Prozessschritte erlernen sollen. Zugleich ist gewünscht, die Trainings möglichst realitätsnah und ansprechend für die Teilnehmer zu gestalten, um den Transfer in die tägliche Arbeit zu erleichtern und die Teilnehmermotivation zu erhöhen. In traditionellen Trainingskonzepten ist die Realitätsnähe eine Herausforderung, da Mitarbeiter ständig zwischen der Arbeitsaufgabe und der Lektüre der Trainingsinhalte wechseln müssen. Im vorliegenden Beitrag untersuchen wir, ob sich mittels virtueller Techniken diese Schwachstellen – fehlende Individualisierung und Realitätsnähe – in der Mitarbeiterschulung verringern lassen.

¹ Volkswagen AG, Konzernforschung, Virtuelle Techniken, 38436 Wolfsburg, sarah.brauns@volkswagen.de

² Karlsruher Institut für Technologie, Institut AIFB, 76128 Karlsruhe, tobias.kaefer@kit.edu

³ Volkswagen AG, Konzernforschung, Virtuelle Techniken, 38436 Wolfsburg, dirk.koriath@volkswagen.de

⁴ Karlsruher Institut für Technologie, Institut AIFB, 76128 Karlsruhe, harth@kit.edu

Für in Großserie produzierende Unternehmen ist der Trainingsprozess von wesentlicher Bedeutung. Bei der Volkswagen AG erarbeitet sich jeder Produktionsmitarbeiter im Bereich Montage und Kommissionierung seine Fertigkeiten im Grundlagentraining und vertieft sie später im sogenannten Profiraumtraining, in dem alle Prozesse direkt im montagenahen Umfeld trainiert werden. Diese Form des dynamischen Trainings soll die Mitarbeiter im Umgang mit Vorgaben zu Ergonomie, Gesundheit oder Sicherheit schulen. Die Lerninhalte decken dabei auch potentielle Qualitätsmängel ab, die sich auf Folgeprozesse auswirken. Inhalte im Hinblick auf Steigerung von Produktivität und Werksleistung werden ebenfalls vermittelt.

In der Vergangenheit gab es bereits Ansätze zur Integration von virtuellen Technologien in den Trainingsprozess (siehe Abschnitt 2). Diese Anwendungen setzten virtuelle Techniken und augmentierte Inhalte ein, z.B. in Form digitaler Reparaturleitfäden eines Fahrzeugs oder Schritt-für-Schritt-Ausbauanleitung für Servicemitarbeiter. Allerdings wurden die Ansätze entweder außerhalb der Produktion oder nur in Umgebungen mit hohem Immersionsgrad eingesetzt. Im Gegensatz dazu fokussiert sich unser Beitrag auf den Einsatz virtueller Techniken für das Training im Produktionsumfeld mit augmentierten Informationen direkt im Blickfeld des Montagearbeiters mittels Datenbrillen (Head-Mounted Displays – HMDs). Dadurch werden Schulungen unter Produktionsbedingungen ermöglicht. Die Mitarbeiter müssen nicht in einer komplett virtuellen Umgebung trainieren, sondern können im gewohnten Umfeld arbeiten, haben beide Hände für die Durchführung ihrer Aufgabe zur Verfügung, und können zusätzliche, augmentierte Informationen zur Unterstützung des Prozesses abrufen, während sie ihre Montageaufgaben durchführen. Somit können die Mitarbeiter alle notwendigen Prozesse vor Ort erlernen und in der realen Umgebung mit virtuellen Techniken trainieren.

Die Möglichkeit, virtuelle Techniken flächendeckend im Training einzusetzen wird aufgrund der weiterentwickelten Hardware und Software immer greifbarer. Produkte aus dem Bereich der Virtuellen Realität werden aktuell für den Massenmarkt reif gemacht. Im Vergleich zu frühen Versionen von Datenbrillen hat sich diese Technologie stark weiterentwickelt, z.B. hinsichtlich Auflösung, Bildwiederholrate und Mobilität. Die fallenden Preise für entsprechende Geräte erlauben ebenfalls einen breiteren Einsatz. Allerdings gab es bisher bei den Versuchen, das Training in der Produktion mit virtuellen Techniken zu unterstützen, oft technische Probleme, die bei dem Einsatz von Hardware zum Anzeigen virtueller Inhalte sowie Software zum Erstellen dieser Inhalte auftraten. Beispielsweise nutzen Datenbrillen eine Optical-See-Through-Kalibrierung zur korrekten Darstellung kontextsensitiver Inhalte. Nur mit einer plausiblen Kalibrierung können die augmentierten Geometrien an der richtigen Position relativ zum Betrachter dargestellt werden. Darüberhinaus soll die Darstellung nach Möglichkeit auch an die Augenposition des Nutzers angepasst werden, um optische Verzerrungen zu vermeiden [Hu10]. Um diese Kalibrierungen zu realisieren, werden aufwändige Prozesse benötigt, welche die Trainierenden zusätzlich durchführen müssen, um auf sie zugeschnittene und an ihr Blickfeld angepasste augmentierte Informationen angezeigt zu bekommen. Zudem werden in komplexeren Szenarien Systeme von verschiedenen Herstellern benötigt, die selbst aufwändig kalibriert werden müssen und zueinander meist inkompatibel sind, was die Integration in eine bestehende Softwarelandschaft erschwert.

Die Kernpunkte dieser Forschungsarbeit sind:

- Wir präsentieren ein Szenario, in dem Mitarbeiter mittels Datenbrillen hinsichtlich eines Produktionsablaufes sowie der Vermeidung bestehender Qualitätsfehler geschult werden, um Nacharbeit zu verhindern. Die Mitarbeiter können neue Inhalte auf audio-visuelle Art erlernen, indem sie zugleich die benötigten Informationen direkt im Blickfeld haben, während sie die Montagetätigkeit durchführen.
- Wir stellen eine im Rahmen des Verbundprojekts ARVIDA¹ entwickelte Architektur vor, welche die flexible Integration und Austauschbarkeit von Komponenten gewährleisten soll. Da in dem Trainingsszenario verschiedenste Hardware- und Softwarekomponenten flexibel kombiniert werden müssen, ist die Modellierung der zu übertragenden Daten sowie die Spezifikation der Schnittstellen notwendig. Die Architektur basiert auf Semantic-Web-Technologien, in der Daten mittels RDF (Resource Description Framework) und RDF Schema modelliert, und einzelne Komponenten über Web-Schnittstellen angesprochen werden.
- Wir evaluieren, welche Auswirkungen die Arbeit mit digitalen Medien und erweiterter bzw. virtueller Realität auf die Mitarbeiter hat, insbesondere im Hinblick auf Trainingszeit und Motivation.

Im weiteren Verlauf des Textes stellen wir verwandte Arbeiten in Abschnitt 2 vor, beschreiben das Szenario in Abschnitt 3 und die Architektur des Gesamtsystems in Abschnitt 4. Die Grundzüge der Evaluierung eines Demonstrators werden in Abschnitt 5 behandelt. Wir schließen mit Zusammenfassung und Ausblick in Abschnitt 6.

2 Verwandte Arbeiten

Im Folgenden stellen wir zunächst den Konzepten aus RAMI 4.0 (Referenzarchitektur für Industrie 4.0) die in der ARA (ARVIDA Referenzarchitektur) entwickelten Techniken und Methoden gegenüber. Danach stellen wir Ansätze vor, die virtuelle Technologien in den Trainingsprozess integrierten.

In RAMI 4.0² werden Ansätze für Datenrepräsentation und Architekturen entwickelt, die Unternehmen die Kommunikation und Anbindung von Systemen in Industrie-4.0-Umgebungen und den übergreifenden Informationsaustausch ermöglichen sollen. Dazu werden verschiedene Ebenen eines Unternehmens herangezogen, sodass Hierarchien, Lebenszyklen, Wertschöpfungsprozesse und Layer der Unternehmenssoftware abgebildet und mit Richtlinien für Datenmodellierung und Kommunikationsprotokolle belegt werden können. Dies geschieht vorrangig auf der Ebene von Industrie-4.0-Komponenten, deren Kommunikationsfähigkeit durch IT-Systeme oder Verwaltungsschalen (Wrappern) erreicht wird. Die dafür zu verwendenden Technologie-Standards werden aktuell noch diskutiert.

¹ Angewandte Referenzarchitektur für virtuelle Dienste und Anwendungen, <http://www.arvida.de/>

² http://www.zvei.org/Downloads/Automation/Industrie%204.0_Komponente_Download.pdf

Dahingegen sind die Technologie-Standards in ARVIDA bereits spezifiziert. Zwar beschränkt sich dieses Referenzarchitektur-Modell vorrangig auf die Umgebung virtueller Dienste und Anwendungen (z.B. Tracking, generisches Menschmodell), konkretisiert jedoch die Vorstellung einer allgemeinen, einheitlichen Schnittstelle, mithilfe deren Komponenten beliebig ausgetauscht werden und trotzdem problemlos in ein Gesamtsystem eingebunden werden können. Die entwickelte ARVIDA Referenzarchitektur soll schlussendlich grundlegend kompatibel zu den Konzepten einer Industrie-4.0-Architektur sein, sodass sich auch weitere Unternehmensbereiche damit erschließen und vernetzen lassen. Dabei basiert ARVIDA auf Diensten, die Systemkomponenten darstellen und mithilfe von REST-Schnittstellen und Vokabularen (modelliert in RDF, RDFS und OWL, der Web Ontology Language) in Architekturkonzepte von Industrie-4.0-Umgebungen eingebunden werden können. Somit ist die Architektur in ARVIDA zwar spezifischer, aber auch detaillierter und anschaulicher beschrieben, weshalb die ARA einfacher und unkomplizierter in den Unternehmen umgesetzt werden kann. Dieses wird bereits durch die generischen Anwendungsszenarien des Projektes gezeigt.

Dahingegen zielt RAMI 4.0 aktuell auf eine zwar weniger detailgetreue, aber allgemein gültige und abstrahierte Struktur, in die sich sämtliche Unternehmen und Komponenten eingliedern lassen. Im Rahmen von RAMI 4.0 wird angestrebt, die europäische Standardisierung und Normung für Unternehmensstrukturen von Industrie-4.0-Umgebungen voran zu treiben. Im Gegensatz dazu basiert die ARA auf etablierten globalen Standards des World Wide Web Konsortiums (W3C), die offen und kostenlos zugänglich sind.

Alles in allem können die beiden Referenzarchitekturmodelle nebeneinander existieren und gegebenenfalls ergänzend verwendet werden, da RAMI 4.0 eine allgemeine Struktur und Normung für Industrie-4.0-Umgebungen anstrebt, wohingegen die ARA spezifische Schnittstellen und konkrete Vokabulare für den Bereich virtueller Techniken entwickelt hat. Somit können Unternehmen allgemeine Konzepte aus RAMI 4.0 nutzen und sich gleichzeitig für virtuelle Dienste und Anwendungen der ARA bedienen. Die ARA ist jedoch potentiell auch für generelle Industrie-4.0-Umgebungen einsetzbar.

Die verwandten Arbeiten im Bereich Training mit virtuellen Techniken gliedern wir in zwei Teile: zum einen Arbeiten, welche Datenbrillen in der Schulung von Servicemitarbeitern einsetzen; zum anderen Arbeiten, die vollimmersive Systeme für die Produktion nutzen.

Im Verbundprojekt ARTESAS wurde ein Szenario für den Kühler austausch am Fahrzeug in einer Service-Werkstatt³ entwickelt. Hier hatten die Projektpartner BMW und DFKI mittels einer halbtransparenten Datenbrille sowie einem mobilen PC augmentierte Bauteile, Hilfsmittel sowie Reparaturschritte im Sichtfeld eines Servicetechnikers eingeblendet. In einem weiteren Anwendungsbeispiel haben die Partner Bosch und Re'flect zusammen mit Range Rover eine mobile Trainings-Applikation entwickelt, mithilfe derer Komponenten wie Kabelbäume und Sensoren hinter dem Armaturenbrett sichtbar gemacht werden konnten, um unnötigen Ausbau von Bauteilen zu vermeiden und so Kosten in den

³ http://www.dfki.de/cebit2006/mti/pdf/8.2_ARTESAS_AR_BMW.pdf

Servicewerkstätten zu verringern⁴. Mithilfe eines mobilen Computers konnte der Servicetechniker so einzelne Bauteile selektieren und diese von allen Seiten betrachten, um sich einen Überblick über deren Umgebung zu verschaffen. Mithilfe dieser Lösung konnte ein Training am realen Fahrzeug stattfinden.

Auch bei der Volkswagen AG wurde ein solcher Reparatur-Fall bereits betrachtet. Da das Modell XL1 nur in geringer Auflage gebaut wurde und zur seiner Reparatur mobile Servicemitarbeiter mit spezifischem Wissen, „Flying Doctors“, eingesetzt werden, entstand die App MARTA (Mobile Augmented Reality Technical Assistance)⁵. Diese App ist ebenfalls auf einem mobilen PC nutzbar und zeigt Reparaturfälle des XL1 an, sodass ein mit dieser Technologie ausgestatteter Mitarbeiter vor Ort das Fahrzeug mit Inhalten der erweiterten Realität überlagern kann und so angezeigt bekommt, welche Schritte er für den jeweiligen Reparaturfall durchzuführen hat und welche Werkzeuge dazu benötigt werden. Dabei erfolgt die Initialisierung des Trackings mithilfe der Fahrzeugsilhouette.

Für das direkte Montagetraining gibt es ebenfalls erste Versuche, diesen Prozess mit virtuellen Techniken zu unterstützen. Das Interaktionskonzept von Circulis und Brigmanis [CB11] ist ein Ansatz, mit dem Teilnehmer die Möglichkeit bekommen sollen, individualisiert mit einer Umgebung aus 3D-Objekten zu interagieren. Mittels Datenschuttschuh und Kopftracking-Sensor kann die Position des Teilnehmers bestimmt werden. Für die Kommunikation zwischen Ein- und Ausgabegerät haben die Autoren versucht, ein Modell darzustellen, das Hintergrundprozesse beschreibt und den Datenaustausch zwischen den Systemen skizziert. Daten wurden für die Applikation aufbereitet und nach vorgegebenen Prinzipien und Regeln analysiert, wodurch aufgrund von Einstellungen und Ergebnissen definierte Operationen in der virtuellen Welt ausgelöst werden.

Zur Unterstützung der Produktionsplanung wurden durch Schenk u.a. [SSK05] voll interaktive, immersive 3D-Visualisierungen von Montagelinien und Fabriken entwickelt, die die Arbeit des Fabrikplaners und des Montaguearbeiters unterstützen sollen. Letzterer hat die Möglichkeit, in einer virtuellen Umgebung die Prozesse zu trainieren und sich somit vorab für seine Aufgabe zu qualifizieren. In der Simulation ist es möglich, intuitiv mit der Umgebung zu interagieren und diese zu beeinflussen. Auch ein digitales Fabrikmodell steht zum Training zur Verfügung, in dem Trainingsteilnehmer die Aufgaben jedes Arbeiters übernehmen können. Dieses Simulationskonzept ist vollständig virtuell gestaltet.

Auch Haase und Termath [HT15] haben ein On-the-Job-Trainingskonzept entwickelt, mit dessen Hilfe ein Servicemitarbeiter Aufgaben im Rahmen einer 3D-Applikation mit didaktischem Konzept bearbeiten kann. Die Trainingsinhalte sind hier mit realen Arbeitsprozessen verknüpft, sodass der Prozess möglichst real durchgeführt werden kann. Der Trainingsteilnehmer darf unabhängig handeln und hat am Ende des Prozesses die Gelegenheit, in einem integrierten Selbstkontroll-Test sein Ergebnis zu evaluieren. Das Training findet dabei in einer vollimmersiven, virtuellen Umgebung statt, die gleichzeitig die Motivation des Trainingsteilnehmers durch den Einsatz neuer Medien erhöhen soll. Die Funktion des Trainers besteht darin, Aufgaben zu kreieren und zu validieren. Dieses System kann fle-

⁴ <https://www.re-flekt.com/archive/de/automotive-ar/165-range-rover-dashboard-de>

⁵ http://www.volkswagenag.com/content/vwcorp/content/de/innovation/Virtual_technologies/MARTA___innovative_service_support_tool_for_the_Volkswagen_XL1.html

xibel in verschiedenen Arbeitsumgebungen eingesetzt werden und dient somit ebenfalls zum Erlernen von Montage-, Kommissionierungs- sowie Instandhaltungsprozessen.

Im Gegensatz zu den bestehenden Arbeiten beschreiben wir ein verteiltes System zur Schulung von Mitarbeitern mit augmentierten Inhalten in der Montage, in dem verschiedene Nutzer den Lernvorgang miteinander gestalten und durchführen können.

3 Szenario

Im Trainingsszenario unterscheiden wir zwei Rollen: zum einen Trainer, die Lerninhalte formulieren, zu Lektionen zusammenstellen und mit Lernmaterialien versehen sowie die Lerninhalte in Trainingssitzungen vermitteln; zum anderen die Trainingsteilnehmer (Mitarbeiter), die an in Sitzungen unterteilten Schulungen teilnehmen, um bestehende Kenntnisse zu vertiefen bzw. weitere Fertigkeiten zu erlernen.

Ein Trainer bekommt in einer Sitzung die Lerninhalte auf einem Tablet angezeigt. Mithilfe des Tablets kann er das Training steuern und ggf. Inhalte hinzufügen, um die Trainingsteilnehmer in der aktuellen Sitzung gemäß ihres aktuellen Wissensstandes bedarfsgerecht zu unterstützen. Die Trainingsteilnehmer tragen eine Datenbrille, die über eine Drahtlosverbindung an den Rest des Systems angebunden ist. So können den Trainigsteilnehmern dieselben Inhalte dargestellt werden, die dem Trainer auf dem Tablet vorliegen. Im Szenario führen Trainingsteilnehmer Arbeitsschritte aus, während der Trainer die Sitzung begleitet und gegebenenfalls um zusätzliche Informationen ergänzt. Sowohl Trainer als auch Trainingsteilnehmer sind berechtigt, in den Inhalten einer Sitzung vor- und zurückzuspringen.

Im Rahmen von ARVIDA wird das virtuelle Training in mehreren Stufen entwickelt: in Stufe 1 besteht die Darstellung in der Datenbrille aus Bildern und Text; in Stufe 2 werden statt der Bilder 3D-Geometrien dargestellt; in Stufe 3 werden zusätzlich 3D-Geometrien durch Aufprojektion dargestellt. Das System wird stufenweise erweitert, da das Erstellen von Trainingsinhalten aufwändig ist [PPS13] und diese erst in Form von Bildern, Video oder 3D-Modellen integriert und bearbeitet werden müssen. Eine Trainingslektion liegt dabei in drei verschiedenen Schwierigkeitsgraden (leicht, mittelschwer, schwer) vor mit unterschiedlich detaillierten Inhalten und Prozessinformationen. Während einer Trainingssitzung kann zwischen den Schwierigkeitsgraden umgeschaltet werden, um bei Bedarf mehr oder weniger Informationen zur Verfügung zu haben. Dies soll den Mitarbeitern ein stufenweises Erlernen der Trainingsinhalte trotz unterschiedlicher Kenntnisse und Erfahrungen ermöglichen.

Die Stufe 1 des Trainings mit Datenbrillen besteht aus Bildern und Text, die dem Trainingsteilnehmer auf der Datenbrille angezeigt werden und somit die einzelnen Arbeitsschritte visualisieren (Abb. 1). Die Darstellung der Arbeitsschritte erfolgt mittels App auf einer EPSON Moverio BT-200 Datenbrille. Der Trainingsteilnehmer hat dabei die Transferteistung zwischen statischem Bild und realem Objekt durchzuführen.

Die Trainingsapplikation für Datenbrillen im Trainingseinsatz wurde über das Verbundprojekt ARVIDA hinaus hinreichend produktisiert, sodass Stufe 1 bereits im Profiraum-



Abb. 1: Training mit einem Trainer (Tablet) und einem Teilnehmer (Datenbrille) mit Bildern und Text zur Beschreibung der Arbeitsschritte (Stufe 1).

Training verwendet wird. Im Zuge dessen wurde für die Trainingserstellung die App um ein „Train-the-Trainer“-Modul erweitert, die dem Trainer Inhalte im Bereich Methodik und Ziele sowie technische Tipps vermittelt. Mithilfe dieser App ist der Trainer in der Lage, sein Training selbst aufzubauen und die Daten in einer Datenbank zu hinterlegen, von der aus auch andere Trainer auf die Inhalte zugreifen können. Dabei verdeutlicht die Train-the-Trainer-Anwendung dem Trainer, wie er für sein Training eine Story schaffen und damit auch die Teilnehmer anspornen kann, um den Ehrgeiz während des Trainings zu wecken. Diese Form des interaktiven Lernens motiviert die Mitarbeiter sowie die Trainer und erhöht den Spaßfaktor bei der Trainingsdurchführung [PDS03].

Die Vorteile dieser Trainingsform bestehen in der Flexibilität und Individualität, die die verschiedenen Schwierigkeitsgrade und Anpassungen während der Trainingssitzung sowohl dem Trainer als auch dem Trainingsteilnehmer bieten. Während einer Sitzung kann der Trainer auf seinem Tablet Notizen anfertigen, die für die Reflektion zur Verfügung stehen sollen. Im Anschluss an die Sitzung ist die Möglichkeit zur Reflektion gegeben und die Trainer bekommen auf ihr Tablet die Dauer der einzelnen Arbeitsschritte sowie die Gesamtzeit des Trainings angezeigt. Dadurch können die Sitzungsteilnehmer gemeinsam Verbesserungspotential erkennen und ggf. gezielt Trainingsinhalte festlegen. Die Daten werden dabei lediglich für die Rückmeldung direkt nach der Sitzung temporär gespeichert.

In Stufe 2 wird das Training um augmentierte Inhalte in Form von Hilfsgeometrien ergänzt. Diese Geometrien lassen sich in die Datenbrillen-App laden und können im Training durch ein Trackingverfahren positionsgenau am Objekt angezeigt werden. Dies sind z.B. virtuelle Pfeile, die im Display an das Trainingsobjekt angefügt werden und Verschraubungsorte anzeigen (Abb. 2). Dadurch verringert sich die Transferleistung des Mitarbeiters erheblich.

Bereits in dieser Ausbaustufe des Trainings kann ein Trainer mittels der entwickelten Applikation mehrere Mitarbeiter betreuen. Diese können sich auch in unterschiedlichen Arbeitsschritten befinden, wenn sie das Training über die Datenbrille eigenständig steuern.



Abb. 2: Training mit einem Trainer (Tablet) und einem Teilnehmer (Datenbrille) mit zusatzlichen 3D Geometrien zur Beschreibung der Arbeitsschritte (Stufe 2).

Zeitgleich konnen sich weitere Trainer in der Sitzung befinden, die sich mit ihren jeweiligen Teilnehmern ber die mobilen Endgerate koppeln. In Stufe 2 sind ebenfalls drei verschiedene Schwierigkeitsgrade in der Trainings-Datenbank hinterlegt, mithilfe derer das Training individualisiert gestaltet werden kann.



Abb. 3: Training mit einem Trainer (Tablet) und einem Teilnehmer (Datenbrille) mit AR-Inhalten und zusatzlichen 3D-Geometrien zur Beschreibung der Arbeitsschritte plus Aufprojektion (Stufe 3).

In Stufe 3, die sich derzeit in Entwicklung befindet, sollen zusatzlich Geometrien mithilfe eines Projektors angezeigt werden, die das Training mit Datenbrillen erganzen (Abb. 3). Mit dieser Erweiterung soll in zukunftigen Arbeiten geklart werden, inwieweit eine Aufprojektion den bestehenden Trainingsprozess weiter unterstutzt.

Die Evaluation im vorliegenden Beitrag behandelt Stufe 1. Die Anbindung der Systeme untereinander in Stufe 2 und 3 erfolgt hierbei mithilfe der im Folgenden beschriebenen Architektur.

4 Architektur

Da an einer Trainingssitzung mehrere Mitarbeiter sowie mehrere Trainer teilnehmen können (n Trainer, m Teilnehmer), wird eine Architektur benötigt, mit der die verschiedenen Applikationen flexibel Daten austauschen können. Des Weiteren wird in Stufe 3 ein Projektor angebunden, der ebenfalls in das Gesamtsystem integriert werden muss. In diesem Abschnitt beschreiben wir die Systemarchitektur für Stufe 2 und 3 als Grundlage für die Integration der einzelnen Komponenten zu einem Gesamtsystem.

Es gibt folgende Anwendungen für Endnutzer:

- Autorenwerkzeug zur Definition der Lerninhalte
- Trainer-Anwendung (mit Train-the-Trainer-Modul) zur eigenständigen Erstellung und Durchführung von Trainingssitzungen
- Trainingsteilnehmer-Anwendung zur individuellen Darstellung der Lerninhalte

Die Anwendungen sind als Apps implementiert, die im Falle der Trainer auf Tablet-Geräten (Android oder Windows) und im Falle der Trainingsteilnehmer auf Datenbrillen ausgeführt werden.

Alle Anwendungen kommunizieren über einen Server („Workflow-Server“), der Trainingssitzungen und Lerninhalte verwaltet und den Anwendungen bereitstellt. Die Kommunikation zwischen den Anwendungen und dem Server findet mittels HTTP⁶ statt. HTTP folgt dem Architekturstil REST (Representational State Transfer) für verteilte Anwendungen. REST propagiert eine lose Kopplung der Komponenten der verteilten Anwendung, indem es zustandslose Kommunikation mit einer beschränkten Menge von Operationen empfiehlt. In HTTP beinhalten diese Operationen GET, PUT, POST und DELETE, welche den grundlegenden Operationen Create, Read, Update und Delete (CRUD) entsprechen. Die Operationen in HTTP sind für Web-Ressourcen definiert, welche beliebige Dinge im System repräsentieren können.

Wir halten daher die Daten auf dem Server in der Form von Web-Ressourcen vor: Für die verschiedenen Anwendungen sind die Daten einerseits Lerninhalte sowie andererseits Beschreibungen von für das Lernszenario relevante Entitäten (z.B. Trainer, Trainingsteilnehmer, Trainingssitzungen, Lektionen). Die Lerninhalte (Bilder, Videos, 3D-Geometrien) sind dabei in binären Formaten gespeichert, die Beschreibungen werden in RDF⁷ ausgezeichnet. Die in RDF zu beschreibenden Web-Ressourcen sind daher Linked-Data⁸-Ressourcen. Um sich aktuell zu halten, rufen die Anwendungen in der vorliegenden Architektur stetig den aktuellen Zustand der relevanten Ressourcen ab (sogenanntes Polling). Um die Menge der übertragenen Daten zu verringern, verwenden wir die HTTP-Header Etag und If-Not-Match.

⁶ <http://tools.ietf.org/rfc/rfc7231>

⁷ <http://www.w3.org/TR/rdf11-concepts/>

⁸ <http://www.w3.org/DesignIssues/LinkedData.html>

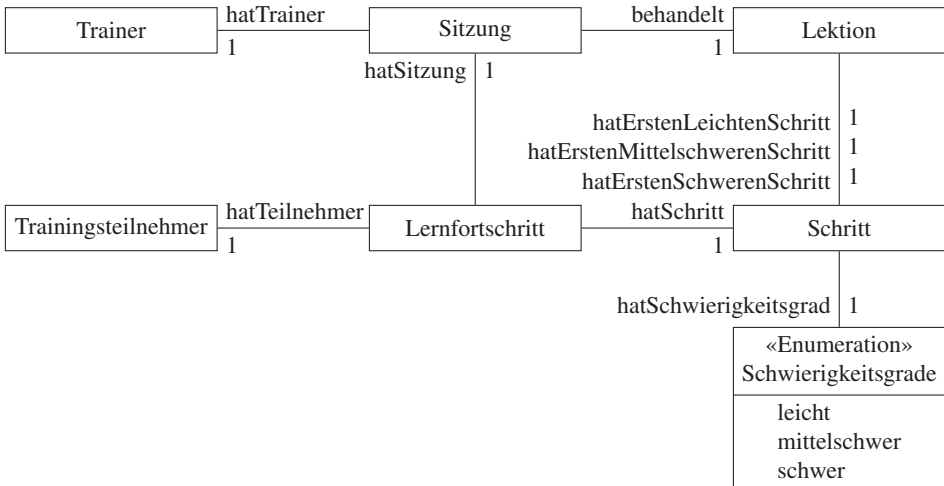


Abb. 4: UML-Klassendiagramm für die zwischen den Anwendungen ausgetauschten Daten.

Im Folgenden beschreiben wir zuerst die Modellierung der Daten für unser Szenario und danach die Interaktion zwischen den Apps und dem Server. Wir modellieren die Beziehungen mithilfe von RDFS⁹ (siehe UML-Klassendiagramm in Abb. 4). Wenn Trainer und Trainingsteilnehmer zu einer Sitzung zusammenkommen wird eine Lektion gelernt. In den Lektionen werden Arbeitsabläufe gelehrt, daher unterteilen wir eine Lektion in mehrere Schritte. In jeder Lektion wird der behandelte Arbeitsablauf auf drei Schwierigkeitsgraden beschrieben. Dabei werden z.B. in höheren Schwierigkeitsgraden weniger detaillierte Beschreibungen gegeben. Die Schritte in den Arbeitsabläufen auf den drei Schwierigkeitsgraden einer Lektion sind untereinander vernetzt, um während einer Trainingssitzung den Grad des Informationsgehaltes anpassen zu können. Um den Trainingsteilnehmer eine individuelle Lerngeschwindigkeit zu ermöglichen, repräsentieren wir den Lernfortschritt jedes Trainingsteilnehmers in einer Web-Ressource, die einer Sitzung zugeordnet ist.

Daten gemäß diesem Klassendiagramm werden zwischen den Anwendungen und dem Server ausgetauscht. In weiteren Abbildungen beschreiben wir die Kommunikation der Anwendungen am Beispiel einer Lektion, die erstellt und anschließend gelehrt wird.

- Der Autor definiert eine Lektion. Er legt die binären Lerninhalte auf den Workflow-Server (erster Request in Abb. 5). Er legt einen Schritt an und fügt dem Schritt einer Lektion zu (folgende Requests in Abb. 5 nutzen RDF zur Beschreibung der Daten).
- Der Trainer erstellt eine Ressource für eine Sitzung auf dem Workflow-Server und fügt der Sitzung Links zur Ressourcen für den individuellen Lernfortschritt von Trainingsteilnehmer 4 hinzu (Abb. 6).

⁹ <http://www.w3.org/TR/rdf-schema/>

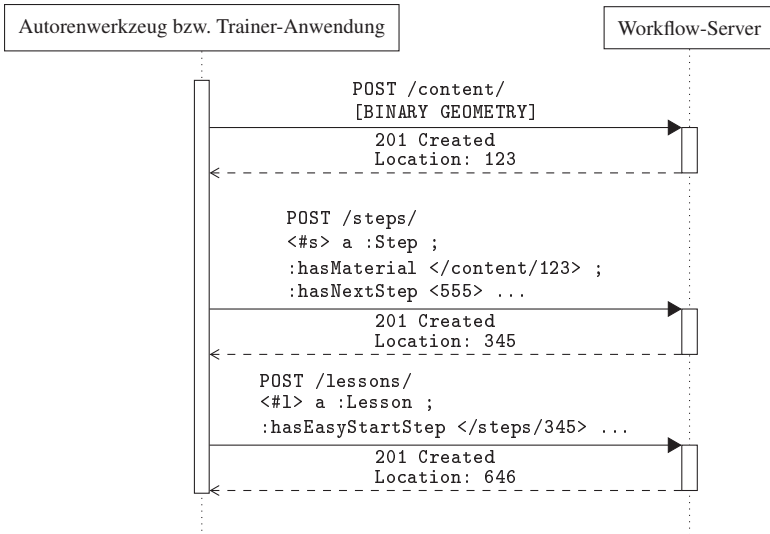


Abb. 5: Erstellung einer Lektion durch den Trainer.

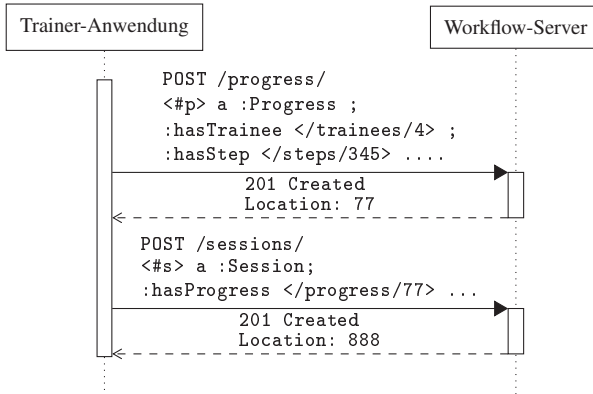


Abb. 6: Erstellung einer Sitzung (/sessions/888) und Hinzufügen des Trainingsteilnehmers 4.

- Die Trainingsteilnehmer stellen ihre Anwendungen auf die ihnen zugeordnete Lernfortschritts-Ressource ein. Die Anwendung fordert vom Workflow-Server die Lerninhalte des aktuellen Schritts an (Abb. 7).
- Der Trainer schaltet den Lernfortschritt für den Trainingsteilnehmer weiter und die Trainingsteilnehmer-App registriert die Änderung (Abb. 7).

5 Evaluierung eines Demonstrators

Die Stufe 1 des Trainings mit statischen Bildern wurde bereits im Volkswagen-Konzern evaluiert, um die Tauglichkeit des Trainingskonzeptes sowie die Beanspruchung der Pro-

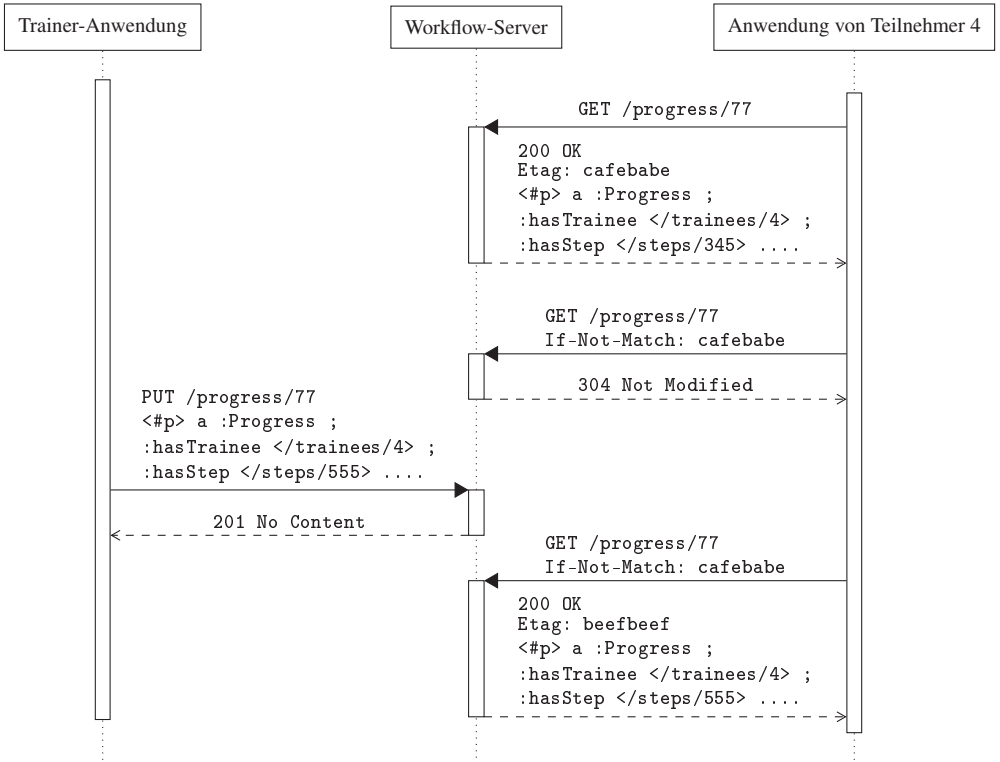


Abb. 7: Die Anwendung von Trainingsteilnehmer 4 pollt den Zustand des Lernfortschritts und die Inhalte. Der Trainer schaltet den Lernfortschritt des Teilnehmers weiter.

banden zu testen. Dabei lag der Fokus auf den Bereichen Müdigkeit und Wachheit, um Aussagen über die medizinische Beanspruchung treffen zu können. Alle internen Genehmigungen wurden erteilt, sodass das Szenario „Verbau des Kabelstranges am Fahrzeug-Leiterrahmen“ mit mehreren Probanden getestet und bewertet werden konnte.

Die Evaluierung wurde über zwei Monate mit 76 Fertigungsmitarbeitern im Zweischichtbetrieb mit einer jeweiligen Trainingsdauer von zwei Stunden betrieben. Es wurden die Art der Instruktion (mit oder ohne Datenbrille), die Anzahl der Hilfestellungen durch den Trainer und die Technikaffinität bewertet sowie Daten zu Befinden, Beanspruchung und Reaktionszeiten erhoben. Die Durchführung erfolgte ohne die Nutzung verschiedener Schwierigkeitsgrade. Dabei wurde ersichtlich, dass sich jüngere Mitarbeiter generell kompetenter im Umgang mit Informationstechnik fühlen als ältere Probanden. Gleichsam konnte bei allen eine Trainingszeit mit Datenbrillen nachgewiesen werden, die nicht länger dauerte als ein Training ohne Datenbrille. Das Feedback der Probanden war ebenfalls zustimmend, da die Darstellung der Inhalte auf der Datenbrille durchgehend als positiv empfunden wurde. Insbesondere der Vorteil der Informationen im Blickfeld sowie die Eigenschaft, dass beide Hände für die Durchführung der Aufgabe zur Verfügung stehen, wurden als positiv wahrgenommen. Zudem wurde auch der Erfolg des Autorenwerkzeugs nachgewiesen, da

die Trainer nach kurzer Einarbeitung selbst in der Lage waren, Trainingsinhalte zu erstellen. Der Aufwand ist im Vergleich zum bisher üblichen Prozess geringer, da die Erstellung interaktiv und dynamisch möglich ist.

Insgesamt wird das Training mit Unterstützung virtueller Techniken als hoher Motivationsfaktor für die Mitarbeiter sowie die Trainer und somit als sinnvolle, lohnende Bereicherung des Prozesses wahrgenommen, in dem durch die Weiterentwicklung von Hardware und Software noch weiteres Potenzial ausgeschöpft werden kann. Daher soll eine weitere Evaluierung der zweiten Ausbaustufe erfolgen, in der auch die zugrundeliegende Architektur bewertet werden soll.

6 Zusammenfassung und Fazit

Der vorliegende Beitrag stellt ein System zu Schulung von Mitarbeitern mittels virtueller Techniken vor. Mehrere Trainingsteilnehmer können interaktiv und mit verschiedenen Schwierigkeitsgraden geschult werden. Der vorgestellte Demonstrator ermöglicht zudem die Durchführung der Fortbildung durch mehrere Trainer. Wir haben die Systemarchitektur beschrieben sowie die Ergebnisse einer initialen Evaluierung des Trainings mittels Datenbrillen vorgestellt.

Durch die individuelle Trainingsgestaltung mit verschiedenen Schwierigkeitsgraden werden den Teilnehmern gleichzeitig nur diejenigen Inhalte dargestellt, welche für ihren Wissensstand passend sind. Somit kann das Training bedarfsgerecht durchgeführt werden, wodurch sich auch die Motivation und der Lernerfolg erhöhen sollen. Mit der Möglichkeit verschiedener Schwierigkeitsgraden in der Trainingsapplikation kann der Trainingsprozess individuell an den Lernstand des Teilnehmers angepasst und Informationen noch während des Trainings ergänzt werden, um optimale Lernbedingungen zu schaffen.

Für den Produktiveinsatz gibt es Hindernisse, sodass derzeit noch kein Serieneinsatz für ein Training mit Datenbrillen direkt in der Fertigung möglich ist. Zum einen ist noch nicht geklärt, inwieweit sich Tablets, Kameras oder Datenbrillen für den Einsatz in der Produktion eignen. Darüber hinaus fehlt die durchgängige Netzwerkanbindung mit WLAN. Zudem stellen die im produktiven Einsatz in der Montage verwendeten Betriebssysteme ein Hindernis dar und können nicht ohne weiteres eingebunden werden. Daher wurde das Trainingsszenario zunächst in einem Trainingszentrum aufgebaut, das dem Produktionsort nachempfunden war und nicht diesen technischen Beschränkungen unterliegt.

Nach Möglichkeit soll das Training an bestehende Produktionssysteme angebunden werden, um auf Daten und Inhalte einfach zugreifen zu können. Mittels der ARA könnten diese Daten ebenfalls in RDF ausgezeichnet sowie über HTTP zugänglich gemacht werden und so einfach integrierbar sein, vorausgesetzt die nicht-technischen, organisatorischen Herausforderungen hinsichtlich der innerbetrieblichen Sicherheit werden adressiert. Die Verfügbarkeit von Daten aus den Produktionssystemen kann den hohen Aufwand für die Erstellung virtueller Inhalte weiter reduzieren. Außerdem ist die derzeitige Performanz der Datenbrillen nicht ausreichend. Allerdings ist eine graduelle Anpassung und Eignung an

produzierende Systeme zu erwarten, da die Hardware- und Software-Hersteller den industriellen Bereich als potenziellen Kunden zunehmend in den Blick nehmen.

Danksagung

Die Arbeiten wurden vom BMBF im Rahmen des Projekts ARVIDA (Angewandte Referenzarchitektur für virtuelle Dienste und Anwendungen) gefördert (FKZ 01IM13001G).

Literatur

- [CB11] Cirulis, Arnis; Brigmanis, Kristaps; Brigis: Software Modules Development for Input Devices in VR/AR Learning Systems. In: Proc. Int. Conf. on Virtual and Augmented Reality in Education. S. 35 – 40, 2011.
- [HK14] Hirsch-Kreinsen, Hartmut: Wandel von Produktionsarbeit – „Industrie 4.0“. Soziologische Arbeitspapiere, 38, 2014. Fakultät 11, TU Dortmund.
- [HT15] Haase, Tina; Termath, Wilhelm: A Virtual Interactive Training Application for Supporting Service Technicians in the Field of High Voltage Equipment. *Procedia Computer Science*, 77:207 – 214, 2015.
- [Hu10] Huckauf, Anke; Urbina, Mario H.; Grubert, Jens; Böckelmann, Irina; Doil, Fabian; Schega, Lutz; Tümler, Johannes; Mecke, Rüdiger: Perceptual Issues in Optical-See-Through Displays. In: Proc. 7th Symp. on Applied Perception in Graphics and Visualization. S. 41–48, 2010.
- [PDS03] Pivec, M.; Dziabenko, O.; Schinnerl, I.: Aspects of game-based learning. In: Proc. I-KNOW. S. 216–225, 2003.
- [PPS13] Petersen, Nils; Pagani, Alain; Stricker, Didier: Real-time Modeling and Tracking Manual Workflows from First-Person Vision. In: Proc. Int. Symp. on Mixed and Augmented Reality. S. 117–124, 2013.
- [SSK05] Schenk, Manfred; Straßburger, Steffen; Kißner, Heike: Combining Virtual Reality and Assembly Simulation for Production Planning and Worker Qualification. In: Proc. 1st Int. Conf. on Changeable, Agile, Reconfigurable and Virtual Production. S. 411–414, 2005.

Augmented Reality am Arbeitsplatz der Zukunft: Ein Usability-Framework für Smart Glasses

Benedikt Zobel¹, Lisa Berkemeier¹, Sebastian Werning² und Oliver Thomas¹

Abstract: Mobile Endgeräte werden immer präsenter am Arbeitsplatz. Insbesondere für nicht-stationäre Anwendungsgebiete und dynamische Prozesslandschaften, wie in der Logistik, sind Mobilität und Flexibilität entscheidend. Smart Glasses und weitere Wearables vereinen diese Eigenschaften, sind einfach auf unterschiedliche Anwendungsfälle adaptierbar und beeinträchtigen dabei nicht die Bewegungsfreiheit wie herkömmliche handgeführte Bediengeräte. Für eine effiziente und zielorientierte Auswahl und Einführung solcher Technologien ist die Gebrauchstauglichkeit zu berücksichtigen. Gängige Usability-Richtlinien sind nicht auf die speziellen Anforderungen von Smart Glasses angepasst. In diesem Beitrag wird deshalb ein, durch eine systematische Literaturrecherche erarbeitetes, Usability-Framework vorgestellt, das konkret auf diese Anforderungen eingeht und auf konventionellen Richtlinien aufbaut. Dieses Framework eröffnet für Forschung und Praxis Möglichkeiten zur Untersuchung der Usability von Smart Glasses und ähnlichen Wearables und bietet durch die Literaturrecherche einen Überblick des aktuellen Wissenstandes als Ausgangspunkt für weitere Forschung im Bereich Usability mobiler Endgeräte.

Keywords: Usability, Smart Glasses, Wearables, Logistik, DIN EN ISO 9241-11

1 Einleitung

In aktuellen Trends wie Consumerization und Bring-Your-Own-Device (BYOD) wird die verstärkte Integration mobiler Endgeräte am Arbeitsplatz deutlich [DK13, HIJ12, WL12]. Diese mobilen Technologien werden zum zentralen Treiber einer effizienten Prozessunterstützung und bieten somit neue Möglichkeiten der Informationsbereitstellung für wissensintensive Tätigkeiten [NTF14], deren Beschaffenheit u.a. freie Hände bei der Ausführung erfordert. Dies sind Anforderungen, die verstärkt in der Logistikbranche auftreten. Richtig eingesetzt liegt der Innovationsgehalt des Einsatzes von Wearables und speziell Smart Glasses als Primäranwendung von Augmented Reality in der verbesserten Integration von Mitarbeitern in Unternehmensprozesse. Die Einsetzbarkeit dieser Technologie in der Logistik wurde bereits im Rahmen von Pilotprojekten und Studien erprobt [BS16, P115]. So unterstützen Datenbrillen den Anwender durch kontextadaptive Informationsbereitstellung sowie freihändige Bedienung bei informationsintensiven und bimanuellen Aktivitäten [P115]. In diesem

¹ Universität Osnabrück, Informationsmanagement und Wirtschaftsinformatik, Katharinenstr. 3, 49074 Osnabrück, vorname.nachname@uni-osnabrueck.de

² Hochschule Osnabrück, Betriebswirtschaftslehre, insbesondere Logistikmanagement, Kaiserstr. 10c, 49809 Lingen, s.werning@hs-osnabrueck.de

Zusammenhang hat die Gebrauchstauglichkeit (engl. Usability) der eingesetzten Geräte einen signifikanten Einfluss auf die Arbeitslast der Mitarbeiter [AGR11].

Usability wird in der Industrie häufig mit Frameworks wie der Norm DIN EN ISO 9241 beurteilt [Ma14]. Durch den Einsatz von Smart Glasses ergeben sich neue Anforderungen an Aspekte, wie die Ergonomie und Interaktion, die von konventionellen Rahmenwerken nicht abgedeckt werden [BB12]. Darüber hinaus basiert die ISO-Norm auf dem Technologie- und Wissensstand von 1998, ihre Aktualität und Übertragbarkeit auf neue Technologien, wie Smart Glasses, ist somit kritisch zu betrachten.

In diesem Beitrag wird daher eine Erweiterung dieser Rahmenwerke durch spezifische Usability-Anforderungen an Smart Glasses vorgestellt. Hierzu wird in Kapitel 2 ein Überblick der betrachteten Technologie sowie aktuellen Ausprägungen des Logistikarbeitsplatzes geschaffen. Erweitert durch potenzielle Entwicklungen, wird ein Ausblick auf einen Arbeitsplatz der Zukunft geschaffen. Darauf folgend werden in der Wissenschaft verbreitete Usability-Standards analysiert sowie ein geeigneter Standard als Ausgangspunkt selektiert. Durch eine systematische Literaturrecherche, dessen methodisches Vorgehen in Kapitel 3 erklärt wird, werden entsprechende Bewertungsaspekte und Anforderungen für Smart Glasses identifiziert. Diese werden in Kapitel 4 zu Usability-Kriterien aggregiert und anschließend in das selektierte Grundlagen-Rahmenwerk integriert. Das entstehende Usability-Framework für Smart Glasses bildet mit den Erkenntnissen aus der Literaturrecherche eine Grundlage für weitere Forschung im Bereich Gebrauchstauglichkeit mobiler Endgeräte. Des Weiteren wird die derzeit bestehende Wissensbasis der Evaluation soziotechnischer Systeme am Arbeitsplatz erweitert. Abschließend erfolgen eine kritische Würdigung der Ergebnisse sowie weiterer Forschungsbedarf in Kapitel 5.

2 Grundlagen

2.1 Smart Glasses-basierte Arbeitsplatzgestaltung in der Logistik

Seit Erreichen der Marktreife mobiler Endgeräte wurde die Digitalisierung stark vorangetrieben. Neue Technologien, wie z. B. Smartphones, Tablets und schnellere mobile Internetverbindungen, lösten eine Innovationswelle aus [Et15]. Nach der Expansion des Smartphone Marktes wird eine ähnliche Entwicklung auch für innovative tragbare Endgeräte wie Wearables erwartet [Lu13, WEA14]. Serienreife Produkte wie die Vuzix M100 Smart Glasses bestätigen, dass dieser Technologie-Trend bereits begonnen hat [Vu16]. Für das Jahr 2019 wird so eine Absatzmenge von 156 Mio. Wearables prognostiziert [Id14, Id15].

Die Logistik ist Deutschlands drittgrößte Branche und ein vom technologischen Wandel besonders gezeichneter Dienstleistungszweig [CG13]. Sie ist geprägt von komplexen Arbeitsschritten mit vielfältigen Prozessvarianten. In Zukunft wird sich ein produktiver

Arbeitsplatz durch ein hohes Maß an Mobilität und Flexibilität definieren [MN89]. Speziell für den Einsatz in der Logistik können heute bereits Kommissionierszenarien durch die Anwendung von Smart Glasses in Form von Pick-by-Vision-Systemen umgesetzt werden [BS16, We16]. Zukünftig werden blickgeführte Funktionen der Smart Glasses, wie zum Beispiel das Scannen von Barcodes, MDE-Geräte ersetzen und eine integrierte Bestandserfassung ermöglichen. Mitarbeitern der Transportlogistik werden in Echtzeit mit Informationen zu Lieferaufträgen versorgt ohne den eigentlichen Arbeitsprozess zu unterbrechen und Kunden werden bei abstimmungsintensiven Prozess direkt in die Arbeitsschritte integriert [BS16, Er16]. Des Weiteren bieten Wearables Lösungen für die Anforderungen globaler Wertschöpfungs-systeme. Zukünftig ist vorstellbar, dass im Rahmen des Qualitätsprüfungsprozesses einzelne Prüfprotokolle mit Hilfe der Smart Glasses visualisiert und durch Verwendung der Smart Watch interaktionsbasiert abgearbeitet werden. So entsteht durch den Einsatz von Wearables die Möglichkeit der Erkennung von stark individuellen Waren. Für die Wahl adäquater Systeme in einem speziellen Anwendungsfall sind Bewertungskriterien der Usability notwendig, welche im Rahmen dieser Publikation gezielt für Smart Glasses erarbeitet werden.

2.2 Konventionelle Usability-Frameworks

Usability bewertet alle Aspekte der Interaktion von Anwender und System hinsichtlich ihrer Qualität der Nutzung von systemseitigen Funktionen [Ni93]. Die Messung von Usability erfolgt durch Frameworks, diese können detaillierte Aspekte umfassen oder kontextbezogene Kategorien vorgeben. Detaillierte Rahmenwerke geben häufig einen Fragenkatalog zur Messung der Usability vor [Ma14]. Ein anerkanntes Framework ist das NASA-TLX, das die Einflussfaktoren der subjektiven Arbeitslast eines Anwenders von soziotechnischen Systemen im betrieblichen Kontext erfasst [AV16, BB12, Bru15, Ol13, Ra14, Th16, TLR14, WD06]. Neben dem NASA-TLX wird auch die System-Usability-Scale als Rahmenwerk verwendet, das durch einen detaillierten Fragebogen gekennzeichnet ist [Bra15, KA12, TLR14]. Des Weiteren ist eine heuristische Evaluation von Usability nach Nielsen [NM90] verbreitet [BGK06, MAA13, Ni92]. Die Anforderungen an die Usability für Bildschirmgeräte der Normenreihe DIN EN ISO 9241 („Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten“) resultieren aus entsprechenden Richtlinien [Ma14]. Der zugehörige Teil 11 („Anforderung an die Gebrauchstauglichkeit – Leitsätze“) definiert weiterführend abstrakte Maße für die Gebrauchstauglichkeit. Obwohl es sich um ein Rahmenwerk für Bürotätigkeiten handelt, ist es durch seine generische Formulierung flexibel auf einen anderen Kontext übertragbar [Ce98]. Diese Flexibilität ist auch ein Anspruch an die Ableitung und Strukturierung Smart Glasses-spezifischer Usability-Aspekte. Die DIN EN ISO 9241-11 ist darüber hinaus nicht nur ein in der Literatur verbreitetes Framework, sondern gleichzeitig ein internationaler Standard für Usability. Als initiales Rahmenwerk für die Konstruktion eines Usability-Frameworks für Smart-Glasses wird daher die DIN EN ISO 9241-11 untersucht. Die Norm nutzt normierte Kriterien für eine vereinheitlichte Messung von Usability. Die Maße dieser Gebrauchstauglichkeit sind

Effektivität, Effizienz und Zufriedenstellung; sie werden in Relation zum Anwendungskontext bewertet. Der Anwendungskontext wird beeinflusst durch Gebrauchsfaktoren, wie die Nutzer, die technische Infrastruktur, soziale Einflüsse oder auch physische Einwirkungen auf den Anwender. Effektivität misst die Aufgabenerfüllung hinsichtlich der Übereinstimmung der Ergebnisse mit den Zielvorgaben im spezifischen Anwendungsfall. Die Effizienz setzt diese Aufgabenerfüllung in Relation zu dem erbrachten Aufwand. Das dritte Maß der Gebrauchstauglichkeit, die Zufriedenstellung, erfasst schließlich die Interdependenzen zwischen Mensch, Endgerät und Aufgabe durch die subjektive Einstellung des Anwenders zum System. Diese kann durch die Intensität der Nutzung oder geäußerte Reaktion des Anwenders auf die Interaktion im System erfasst werden. Die Gestaltung des Endgerätes und die Interaktion im soziotechnischen System sind dabei starke Einflussfaktoren der Zufriedenstellung [Ce98]. Der Einsatz der ISO 9241-11 wird limitiert durch die abstrakten Messkriterien, deren kontextbezogene Spezifizierung in messbare Einzelfaktoren, nicht statistisch validiert ist und weiterhin keine Angaben zur Interpretation der Ergebnisse vorliegen [Se06]. Die Usability-Maße der Norm DIN EN ISO 9241-11 werden im Folgenden für Smart Glasses spezifiziert.

3 Methodisches Vorgehen

Zur Erforschung des aktuellen Kenntnisstandes in der Wissenschaft wurde eine systematische Literaturrecherche ausgewählt. Die Recherche wurde nach den Richtlinien von Webster & Watson [WW02] durchgeführt. Zur Strukturierung der qualitativen Analyse wurden die von Dyba & Dingsoyr [DD08] definierten Phasen durchlaufen: (1) Datenbanken zur Identifizierung relevanter Forschung durchsuchen, (2) Beiträge aufgrund von Titeln ausschließen, (3) Beiträge aufgrund von Abstracts ausschließen, sowie (4) Primärbeiträge identifizieren und Forschung kritisch würdigen. Die von [WW02] empfohlene Vorwärts- und Rückwärtssuche wurde als letzter Schritt angewandt.

Aktuelle Forschung im Bereich Smart Glasses, Wearables und Usability ist in verschiedenen Forschungsfeldern zu finden, wie z.B. in der Informatik, der Wirtschaftsinformatik, und weiteren. Um diese Bereiche in die Suche miteinzubeziehen, wurden insgesamt 9 unterschiedliche Forschungsdatenbanken durchsucht. Diese sind *Elsevier ScienceDirect*, *ISI Web of Knowledge*, *EBSCO Host*, *Springer Link*, *Emerald*, *Wiley Online Library*, *AIS Electronic Library*, *IEEE Xplore* sowie *WISO*. Um möglichst breitenwirksam den aktuellen Stand der Forschung im Bereich Usability für Smart Glasses identifizieren zu können, wurde eine entsprechende Schlüsselwortmatrix entworfen. In der ersten Dimension wurden alle gängigen Bezeichnungen von Smart Glasses sowie eventuell vorhandene Abkürzungen aufgeführt, also *Smart Glasses*, *Data Glasses*, *HMD*, *Head Mounted Display*, *Wearables*, *HUD*, *Head-Up Display* sowie *Head Worn Display*, die zweite Dimension beinhaltet das Wort *Usability*. Da die Thematik von Datenbrillen erst mit der Ankündigung der ersten alltagstauglichen

Datenbrille, der *Google Glass*, Aufmerksamkeit erhielt, wurde die Anfangssuche auf Beiträge mit Veröffentlichungsdatum ab 2012 beschränkt. Ferner wurden nur Beiträge in den Sprachen Englisch und Deutsch betrachtet.

Von den in Phase 1 identifizierten 3232 Beiträgen konnten durch das Herausfiltern von nicht verwertbaren Einträgen und Duplikaten 69 Elemente entfernt werden. In Phase 2 konnten weitere 2537 Beiträge aufgrund ihrer Titel eliminiert werden. So wurden Beiträge entfernt, die offensichtlich keinen thematischen Bezug zu Smart Glasses oder Usability haben. Durch das Filtern von Abstracts der verbliebenen 626 Beiträge konnten 513 weitere Quellen entfernt werden. So wurden Betrachtungen von reinen Smartphone-Apps als „Augmented Reality“-Anwendung eliminiert. In Phase 4 wurden die verbliebenen 113 Beiträge durch Analyse des Volltextes weiter eingeteilt. Dabei wurden die Identifikation von Usability-Prinzipien, Untersuchungen oder Kriterien die zur Herleitung dieser dienen können, und ein konkreter Bezug zu Smart Glasses oder Wearables als Relevanzkriterien festgelegt. Auf diesem Weg wurden 49 Primärquellen identifiziert. Unter Einbezug relevanter Verweise in diese Quellen wurde diese Auswahl durch eine Vorwärts- und Rückwärtssuche auf 53 Beiträge erweitert. Der Auswahlprozess ist in Abb. 1 dargestellt.

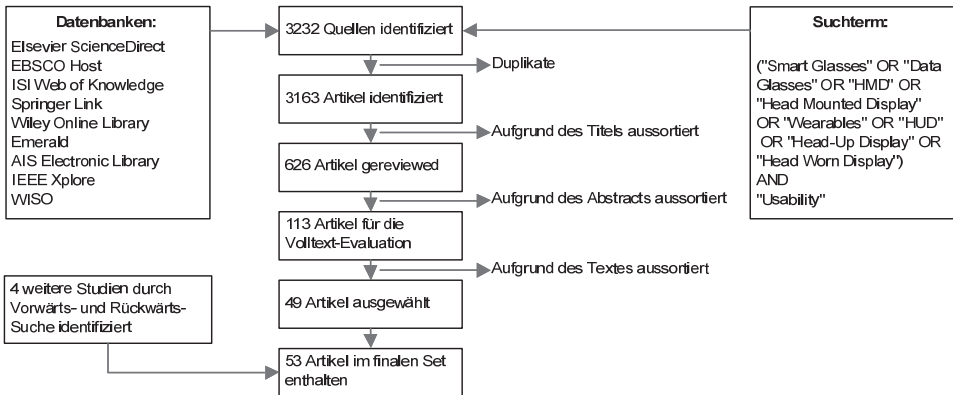


Abb. 1: Prozessdarstellung der systematischen Literaturrecherche

4 Usability-Framework zur Gestaltung von Smart Glasses

In der Literaturanalyse sind 40 Usability-Aspekte identifiziert worden, die mindestens von zwei verschiedenen Quellen als relevant postuliert werden. Diese Aspekte werden in Tab. 1 beschrieben und zu zwölf Usability-Kriterien aggregiert.

<i>Usability Kriterium (UK)</i>	<i>Usability Aspekt</i>	<i>Beschreibung</i>	<i>Quelle</i>
Performanz (UK1)	Leistungsfähigkeit	Durch das System induzierte Leistungsfähigkeit und damit verbundene Kosteneinsparungen.	[AI14, AV16, Ba14, BB12, BLB13, WD13]
	Effizienz	Nutzen-Aufwand-Verhältnis des mittels Smart Glasses erzielten Ergebnisses.	[AV16, Bra15, Gi13, Gr13]
	Arbeitslast	Subjektive Arbeitslast insbesondere im Verhältnis zur Nichtnutzung des Systems.	[AV16, KS15, Ne12]
Perzeption (UK2)	Wahrnehmung	Die bereitgestellten Informationen können einfach erfasst werden (Farbe, Helligkeit, Kontrast, Positionierung virtueller Objekte).	[AV16, Ba14, BB12, Bra15, DM15, MAB14, Ya13]
	Informationsverarbeitung	Der Nutzer erfasst die Informationen unmittelbar.	[BGK06, DM15, MAA13, Ra14, St16, TSK15, Ya13]
	Immersion	Vertiefung in die virtuelle bzw. erweiterte Realität.	[Ba14, BGK06, DM15, MAA13, Ne12, Ra14, RMP15, WD13]
Präferenz (UK3)	Zufriedenheit	Zufriedenheit des Nutzers mit dem System.	[Bra15, CCL15, RMP15, WD06, Gi13, Gr13, Ha14, HLH15, IR15, JL13, Lv15, O113]
	Einstellung	Die Einstellung gegenüber der Nutzung des Systems.	[AI14, CG12, KA12, KS15, RMP15]
	Motivation	Das System motiviert zur Nutzung und Aufgabenerfüllung.	[O113, Ra14]
User Interface (UK4)	Informationsdarstellung	Die Informationen werden für eine qualitativ hochwertige Darstellung visuell aufbereitet.	[AV16, BGK06, Ha14, IR15, LMN04, MAA13, Ne12, St16, WD13]
	Menüführung	Strukturierte Navigation durch das Menü, z.B. durch eine hierarchische Menüstruktur.	[JL13, TSK15]
	Individualisierung / Adaption	Das User Interface kann an die Bedürfnisse und soziokulturellen Anforderungen des Nutzers angepasst werden.	[BGK06, Bo13, LS12, TSK15]
	Konsistenz	Konsistente Gestaltung und Funktionsweise des User Interface.	[BGK06, BL12, KA12, LMN04, MAA13]
	Schlichtheit	Ein minimalistisches Design steigert die Benutzerfreundlichkeit des Interface.	[BGK06, KA12, LMN04, MAA13]

<i>Usability Kriterium (UK)</i>	<i>Usability Aspekt</i>	<i>Beschreibung</i>	<i>Quelle</i>
Interaktion (UK5)	Feedback	Rückmeldung des Wearables über Systemstatus an den Nutzer durch ein haptisches oder auch visuelles Feedback.	[BGK06, BL12, JL13, LMN04, Ne12]
	Natürlichkeit	Die Interaktion erfolgt intuitiv und durch für den Nutzer natürliche Bewegungen.	[Ba14, KA12, Ne12, TSK15, WD13]
	Intuitivität (Steuerung)	Die Bedienung erfolgt intuitiv, z.B. mittels Gesten eines simplen Gestensets.	[BL12, Lv15, Or15, Pi13]
	Bedienkomfort	Der Aufwand der für die Interaktion mit dem Gerät und der Nutzung der Funktionen erforderlich ist. Z.B. ist eine automatisierte Erkennung der Situation und Anzeige der benötigten Informationen weniger aufwändig als bewusstes Abrufen von Informationen durch den Nutzer. Ein hoher Bedienkomfort spiegelt sich z.B. in einer leichten Erlernbarkeit der Systemsteuerung wieder.	[AI14, BGK06, Ne12, WD13, Bra15, CCL15, IR15, KA12, KP14, KS15, MAA13, MAB14]
	Integration des Nutzers	Der Nutzer wird in das System integriert, z.B. durch Feedback.	[Ce12, CG12, Ne12, St16]
Kognitiver Aufwand (UK6)	Mentaler Anspruch	Mentaler Aufwand der zur Nutzung des Systems erforderlich ist.	[AV16, Ba14, BB12, Bra15, Ha14, Ne12, WD06]
	Komplexität	Den Umfang des Interaktions- und Funktionsrahmens angemessen gestalten.	[KA12, MAA13, MAB14, WD06]
Tauglichkeit (UK7)	Nützlichkeit	Das System bildet einen Mehrwert bei der Aufgabenerfüllung.	[AI14, AV16, Bo13, Bra15, KS15, MAA13, Pu12, RMP15, St16, Ta15]
	Aufgabengemessenheit	Das System ist geeignet für die Aufgabenerfüllung ohne überdimensioniert zu sein.	[AV16, LB13, Ma14, RMP15, TSK15, WD13]
	Effektivität	Das mit dem System erzielte Ergebnis entspricht der Zielvorstellung.	[Do14, Gi13, Gr13, Ne12]
Systemstabilität (UK8)	Robustheit	Fehlertoleranz und Laufzeitstabilität des Systems.	[AI14, AV16, LMN04, MAA13, Ne12]
	Fehlermanagement	Fehlerprävention, -behandlung und Userunterstützung.	[Ba14, BB12, BGK06, HLH15, Ma14, TSK15]
Anwendungskontext (UK9)	Kontextsensitivität	Besondere Anforderungen aus dem Kontext der Nutzung, wie z.B. eine Helmpflicht auf Baustellen.	[BL12, LMN04, Ta15]
	Einsatzort	Umwelteinflüsse in Form von Lichtverhältnissen, Umgebungsgeräuschen, etc.	[AV16, Ne12, Ta15]

<i>Usability Kriterium (UK)</i>	<i>Usability Aspekt</i>	<i>Beschreibung</i>	<i>Quelle</i>
Ergonomie (UK10)	Tragekomfort	Physischer Komfort bei der Nutzung des Systems.	[AV16, BGK06, CCL15, Na12, Ne12, St16, TSK15, WD06, WD13]
	Physische Beeinträchtigung	Gesundheitliche Folgen einer Nutzung des Systems.	[Bo13, Bra15, Ne12, Od14, Or15]
soziale Erfahrung (UK11)	Konventionen	Das System entspricht den soziokulturellen Normen.	[BGK06, KS15]
	Datenschutz	Datenschutz wird durch das System gewährleistet.	[BGK06, MAA13, OI13]
	Kollaboration	Eine Interaktion mit anderen Personen wird gefördert.	[BB12, OI13],
Technologie (UK12)	Sensorik	Die Sensorik Unterstützung ermöglicht einen adäquaten Funktionsumfang.	[Bru15, JL13, MAB14, Ne12, OI13]
	Display	Das Display ist geeignet für die Informationsbereitstellung hinsichtlich Auflösung, Transparenz und ggf. 3D-Darstellung.	[Do14, MAB14, Ne12, Ta15]
	Mobilität	Das System ist mobil einsetzbar.	[Bru15, KS15]
	Systemintegration	Das System kann in die vorhandene Systemlandschaft integriert werden.	[Bru15, Ce12, DM15, MAB14, Ne12]
	Latenz	Die Wartezeit des Systems ist akzeptabel.	[Ba14, HLH15, IR15, Ne12]
	Eingabemedien	Es sind verschiedene Eingabemedien als Basis für die Interaktion vorhanden.	[Bru15, Ce12]
	Standards	Industriestandards und Erwartungen an die Hardware werden erfüllt.	[Bo13, Do14]

Tab. 1: Usability-Kriterien und -Aspekte

Die identifizierten Usability-Kriterien werden in Abb. 2 durch eine Zuordnung zu den Maßen der Gebrauchstauglichkeit in die DIN EN ISO 9241-11 integriert. Eine eindeutige Abgrenzung der Kriterien ist jedoch nicht möglich, da Interdependenzen zwischen den Aspekten bestehen. Die übergeordneten Usability-Faktoren der DIN EN ISO 9241-11 sind durch ihre Abstraktheit für verschiedene Anwendungsfälle geeignet. Das vorliegende Framework spezifiziert diese Variablen zunächst hinsichtlich der Technologie Smart Glasses und bleibt dadurch kontextadaptiv für verschiedene Einsatzszenarien, und somit auch für Smart Glasses im Anwendungsfall Logistik.

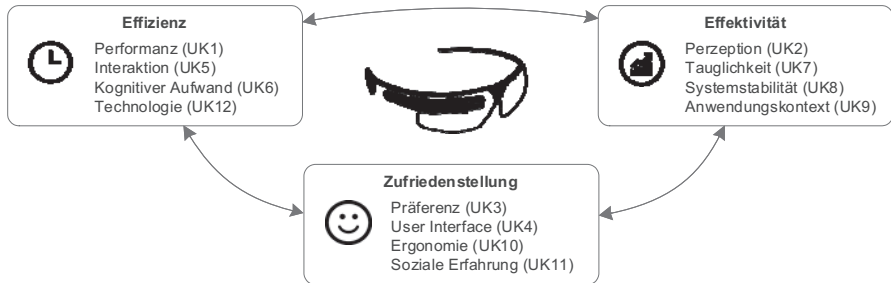


Abb. 2: Integration von Usability-Kriterien und ISO

5 Diskussion und Fazit

In diesem Beitrag wird der aktuelle Wissensstand zum Thema Usability für Smart Glasses in der Wissenschaft erhoben. Die durchgeführte Recherche reagiert auf ein Defizit von konkreten Gestaltungsrichtlinien für diese Technologiegruppe. Durch Kombination von aus der Literatur entnommenen Usability-Kriterien und dem verbreiteten Standard DIN EN ISO 9241-11 ist ein Usability-Framework für Smart Glasses definiert worden. Bei der Literaturanalyse hat sich allerdings ein abweichendes Verständnis von Usability gezeigt. Daraus folgt die Möglichkeit einer nicht trennscharfen Abgrenzung der Usability-Kriterien zu angrenzenden Forschungsgebieten, wie der Akzeptanzforschung. Darüber hinaus sind die identifizierten Experimente mit verschiedenen Arten von Smart Glasses durchgeführt worden, teilweise auch mit nicht mehr aktuellen Modellen. Eine Weiterentwicklung der Usability der Geräte wird somit nicht konsequent berücksichtigt. Ebenso ist das Framework durch die Durchführung einer Literaturrecherche auf bereits benannte Kriterien aus wissenschaftlicher Forschung limitiert. Das vorliegende Framework ist u.a. aus diesen Gründen in Experimenten und weiterer Forschung zu ergänzen, validieren und schließlich in dem fokussierten Anwendungskontext, Smart Glasses in der Logistik, zu evaluieren. Für die Anwendbarkeit des Frameworks auf Smart Glasses Systeme ist die Erstellung eines korrespondierenden Fragen- und Messkatalogs erforderlich. Dazu sind die folgenden Schritte geplant: Erstellung des Usability Katalogs aus der vorliegenden Literatur, Ergänzung durch Experteninterviews und weitere Literaturrecherche sowie eine abschließende Konsolidierung der Ergebnisse in einem Fragen- und Messkatalog. Weiterhin ist eine Evaluation des Katalogs durch eine Anwendung im Experiment geplant, um das Modell um konkrete Vorgehenshilfen anzureichern.

Das Usability-Framework ist Basis für die Bewertung von Smart Glasses-Systemen und wird somit zu einem effizienten Werkzeug der Entscheidungshilfe bei der Auswahl und Einführungen von Smart Glasses-Technologien für Unternehmen in der Praxis. Gleichzeitig bietet es eine wissenschaftliche Grundlage zur Evaluation und Gestaltung der Usability mobiler Endgeräte der entsprechenden soziotechnischen Systeme und stellt

dadurch eine Erweiterung der Wissensbasis dar. Für die Logistikbranche ist das hier aufgezeigte Framework zur Gebrauchstauglichkeit von Smart Glasses notwendig, um heutige Arbeitsplätze, welche von statischen Strukturen geprägt sind, mit adäquaten Smart Glasses-Technologien auszurüsten. Hierdurch entsteht ein mobiler, flexibler und somit zukunftsfähiger Arbeitsplatz. Das vorliegende Framework kann dahingehend auf weitere einzelne Wearables wie Smart Watches oder auch eine holistische Perspektive, die die Kombination verschiedener Wearables berücksichtigt, erweitert werden. Technologische Innovationen am Arbeitsplatz, wie Wearables, können nachhaltig umgesetzt werden, wenn deren Usability berücksichtigt wird.

Literaturverzeichnis

- [AGR11] Ayyagari, R.; Grover, V.; Russell, P.: Technostress: Technological Antecedents and Implications. In: *MIS Quarterly* Bd. 35 (2011), Nr. 4, S. 831–858
- [AI14] Ahlan, A. R.; Isma, B.: User Acceptance of Health Information Technology (HIT) in Developing Countries: A Conceptual Model. In: *Procedia Technology* Bd. 16, Elsevier B.V. (2014), S. 1287–1296
- [AV16] Aromaa, S.; Väänänen, K.: Suitability of virtual prototypes to support human factors/ergonomics evaluation during the design. In: *Applied Ergonomics* Bd. 56 (2016), S. 11–18
- [Ba14] Barbieri, L. et al.: Effects of device obstruction and tool-hand misalignment on user performance and stiffness perception in visuo-haptic mixed reality. In: *International Journal of Human-Computer Studies* Bd. 72 (2014), Nr. 12, S. 846–859
- [BB12] Bai, Z.; Blackwell, A. F.: Analytic review of usability evaluation in ISMAR. In: *Interacting with Computers* Bd. 24 (2012), Nr. 6, S. 450–460
- [BGK06] Bertini, E.; Gabrielli, S.; Kimani, S.: Appropriating and assessing heuristics for mobile computing. In: *AVI '06: Proceedings of the working conference on Advanced visual interfaces (2006)*, S. 119–126
- [BL12] Billinghamurst, M.; Lee, M.: Multimodal Interfaces for Augmented Reality. In: Dill, J. et al. (Hrsg.): *Expanding the Frontiers of Visual Analytics and Visualization* : Springer London, 2012, S. 449–465
- [BLB13] Budhiraja, R.; Lee, G. A.; Billinghamurst, M.: Using a HHD with a HMD for mobile AR interaction. In: *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* : IEEE, 2013, S. 1–6
- [Bo13] Borisov, N. et al.: Integrating Production Workers into User Interface Design for Diagnosis Devices in Automotive Production Environments: Field Experiences and Lessons Learned. In: Stephanidis, C. (Hrsg.): *HCI International 2013 - Posters' Extended Abstracts*. Bd. 374 : Springer Berlin Heidelberg, 2013, S. 469–473
- [Bra15] Brancati, N. et al.: Usability Evaluation of a Wearable Augmented Reality System for the Enjoyment of the Cultural Heritage. In: *2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)* : IEEE, 2015, S. 768–774

- [Bru15] Brusie, T. et al.: Usability evaluation of two smart glass systems. In: Systems and Information Engineering Design Symposium (SIEDS) : IEEE, 2015, S. 336–341
- [BS16] Bechtle; SAP: Smarte Brillen optimieren Lagerlogistik Bd. 03/2016 (2016), S. 27–28
- [CCL15] Cha, J.; Choi, E.; Lim, J.: Virtual Sliding QWERTY : A new text entry method for smartwatches using Tap-N-Drag. In: Applied Ergonomics Bd. 51, Elsevier Ltd (2015), S. 263–272
- [Ce12] Cernea, D. et al.: Tangible and Wearable User Interfaces for Supporting Collaboration among Emergency Workers. In: Hutchison, D. et al. (Hrsg.): Collaboration and Technology. Bd. 7493 : Springer Berlin Heidelberg, 2012, S. 192–199
- [Ce98] CEN: ISO 0241: Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten; Teil 11: Anforderungen an die Gebrauchstauglichkeit - Leitsätze (1998)
- [CG12] Chong, M. K.; Gellersen, H.: Usability classification for spontaneous device association. In: Personal and Ubiquitous Computing Bd. 16 (2012), Nr. 1, S. 77–89
- [CG13] Clausen, U.; Geiger, C. (Hrsg.): Verkehrs- und Transportlogistik : Springer Berlin Heidelberg, 2013
- [DD08] Dybå, T.; Dingsøy, T.: Empirical studies of agile software development: A systematic review. In: Information and Software Technology Bd. 50 (2008), Nr. 9-10, S. 833–859
- [DK13] Disterer, G.; Kleiner, C.: BYOD - Bring Your Own Device. In: HMD Praxis der Wirtschaftsinformatik Bd. 50 (2013), Nr. 2, S. 92–100
- [DM15] Dini, G.; Mura, M. D.: Application of Augmented Reality Techniques in Through-life Engineering Services. In: Procedia CIRP Bd. 38 (2015), S. 14–23
- [Do14] Donovan, A. et al.: Understanding How Network Performance Affects User Experience of Remote Guidance. In: Baloian, N. et al. (Hrsg.): Collaboration and Technology. Bd. 8658. Cham : Springer International Publishing, 2014, S. 1–12
- [Er16] Ernst, C.-P.: The Drivers of Wearable Device Usage : Springer International Publishing, 2016
- [Et15] Etezadzadeh, C.: Smart City - Stadt der Zukunft? Die Smart City 2.0 als lebenswerte Stadt und Zukunftsmarkt : Springer Fachmedien Wiesbaden, 2015
- [Gi13] Di Gironimo, G. et al.: A virtual reality approach for usability assessment: case study on a wheelchair-mounted robot manipulator. In: Engineering with Computers Bd. 29 (2013), Nr. 3, S. 359–373
- [Gr13] Gross, T.: Mensch-Computer-Interaktion in Wissenschaft und Praxis. In: HMD Praxis der Wirtschaftsinformatik Bd. 50 (2013), Nr. 6, S. 6–15
- [Ha14] Ham, J. et al.: Smart Wristband: Touch-and-Motion–Tracking Wearable 3D Input Device for Smart Glasses. In: Streit, N.; Markopoulos, P. (Hrsg.): Distributed, Ambient, and Pervasive Interactions, Lecture Notes in Computer Science. Bd. 8530. Cham : Springer International Publishing, 2014, S. 109–118
- [HIJ12] Harris, J.; Ives, B.; Junglas, I.: IT Consumerization: When Gadgets Turn Into Enterprise IT Tools. In: MIS Quarterly Executive Bd. 11 (2012), Nr. 3, S. 99–112

- [HLH15] Hall, N.; Lowe, C.; Hirsch, R.: Human Factors Considerations for the Application of Augmented Reality in an Operational Railway Environment. In: *Procedia Manufacturing* Bd. 3 (2015), S. 799–806
- [Id14] IDC: Absatz von Wearables weltweit in den Jahren 2014 und 2015 (in Millionen Stück). URL <http://de.statista.com/statistik/daten/studie/515723/umfrage/absatz-von-wearables-weltweit/>. - abgerufen am 2016-05-11. — Statista - Das Statistik-Portal
- [Id15] IDC: Unit shipments forecast basic/smart wearables worldwide from 2014 to 2019 (in millions). URL <http://www.statista.com/statistics/296565/wearables-worldwide-shipments/>. - abgerufen am 2016-05-11. — Statista - Das Statistik-Portal
- [IR15] Irshad, S.; Rambli, D. R. A.: User Experience Satisfaction of Mobile-Based AR Advertising Applications. In: Badioze Zaman, H. et al. (Hrsg.): *Advances in Visual Informatics*. Bd. 9429. Cham : Springer International Publishing, 2015, S. 432–442
- [JL13] Jeon, M.; Lee, J.-H.: The Ecological AUI (Auditory User Interface) Design and Evaluation of User Acceptance for Various Tasks on Smartphones. In: Hutchison, D. et al. (Hrsg.): *Human-Computer Interaction. Interaction Modalities and Techniques*. Bd. 8007 : Springer Berlin Heidelberg, 2013, S. 49–58
- [KA12] Kluge, M.; Asche, H.: Validating a Smartphone-Based Pedestrian Navigation System Prototype. In: Hutchison, D. et al. (Hrsg.): *Computational Science and Its Applications – ICCSA 2012*. Bd. 7334 : Springer Berlin Heidelberg, 2012, S. 386–396
- [KP14] Kim, Y.; Park, J.: Study on interaction-induced symptoms with respect to virtual grasping and manipulation. In: *International Journal of Human-Computer Studies* Bd. 72 (2014), Nr. 2, S. 141–153
- [KS15] Kim, K. J.; Shin, D.-H.: An acceptance model for smart watches: Implications for the adoption of future wearable technology. In: *Internet Research* Bd. 25 (2015), Nr. 4, S. 527–541
- [LB13] Lee, M.; Billinghamurst, M.: A usability study of multimodal input in an augmented reality environment. In: *Virtual Reality* Bd. 17 (2013), Nr. 4, S. 293–305
- [LMN04] Longoria, R. G.; McGee, M.; Nash, E.: Heuristics for Designing Mobile Applications. In: Sammes, A. J.; Longoria, R. (Hrsg.): *Designing Software for the Mobile Context* : Springer London, 2004, S. 109–134
- [LS12] Law, E. L.-C.; Sun, X.: Evaluating user experience of adaptive digital educational games with Activity Theory. In: *International Journal of Human-Computer Studies* Bd. 70 (2012), Nr. 7, S. 478–497
- [Lu13] Luftman, J. et al.: Key information technology and management issues 2012–2013: an international study. In: *Journal of Information Technology* Bd. 28, Nature Publishing Group (2013), Nr. 4, S. 354–366
- [Lv15] Lv, Z. et al.: Extending touch-less interaction on vision based wearable device. In: *2015 IEEE Virtual Reality (VR) : IEEE*, 2015, S. 231–232
- [Ma14] Maguire, M.: Socio-technical systems and interaction design: 21st century relevance. In: *JERG* Bd. 45, Elsevier (2014), Nr. 2, S. 162–170
- [MAA13] Marcus, A.; Abromowitz, S.; Abulkhair, M. F.: Heuristic Evaluation of

- iCalamityGuide Application. In: Hutchison, D. et al. (Hrsg.): Design, User Experience, and Usability. User Experience in Novel Technological Environments. Bd. 8014 : Springer Berlin Heidelberg, 2013, S. 130–139
- [MAB14] Mantzios, V.-M.; Apostolopoulos, T.; Beltramello, O.: Augmented Reality Applications Assisting Maintenance Activities in Extreme Environments: HCI Issues. In: Junqueira Barbosa, S. D. et al. (Hrsg.): HCI International 2014 - Posters' Extended Abstracts. Bd. 434. Cham : Springer International Publishing, 2014, S. 613–618
- [MN89] Macrae, N.: Workplace flexibility: Past, present, and future. In: Journal of Labor Research Bd. 10 (1989), Nr. 1, S. 51–55
- [Na12] Navab, N. et al.: First Deployments of Augmented Reality in Operating Rooms. In: Computer Bd. 45 (2012), Nr. 7, S. 48–55
- [Ne12] Nee, A. Y. C. et al.: Augmented reality applications in design and manufacturing. In: CIRP Annals - Manufacturing Technology Bd. 61 (2012), Nr. 2, S. 657–679
- [Ni92] Nielsen, J.: Finding usability problems through heuristic evaluation. In: Proceedings of the SIGCHI conference on Human, 1992, S. 373–380
- [Ni93] Nielsen, J.: Usability Engineering. Cambridge : Academic Press, Inc., 1993
- [NM90] Nielsen, J.; Molich, R.: Heuristic Evaluation of user interfaces. In: CHI '90 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1990, S. 249–256
- [NTF14] Nüttgens, M.; Thomas, O.; Fellmann, M. (Hrsg.): Dienstleistungsproduktivität: Mit mobilen Assistenzsystemen zum Unternehmenserfolg : Springer Fachmedien Wiesbaden, 2014
- [Od14] Odenthal, B. et al.: A Comparative Study of Head-Mounted and Table-Mounted Augmented Vision Systems for Assembly Error Detection: Augmented Vision Systems for Assembly Error Detection. In: Human Factors and Ergonomics in Manufacturing & Service Industries Bd. 24 (2014), Nr. 1, S. 105–123
- [OI13] Olsson, T.: Concepts and Subjective Measures for Evaluating User Experience of Mobile Augmented Reality Services. In: Huang, W.; Alem, L.; Livingston, M. A. (Hrsg.): Human Factors in Augmented Reality Environments. New York : Springer New York, 2013, S. 203–232
- [Or15] Orlosky, J. et al.: ModulAR: Eye-Controlled Vision Augmentations for Head Mounted Displays. In: IEEE Transactions on Visualization and Computer Graphics Bd. 21 (2015), Nr. 11, S. 1259–1268
- [Pi13] Piumsomboon, T. et al.: User-Defined Gestures for Augmented Reality. In: Hutchison, D. et al. (Hrsg.): Human-Computer Interaction – INTERACT 2013. Bd. 8118 : Springer Berlin Heidelberg, 2013, S. 282–299
- [Pl15] Plutz, M.: Gastkommentar / Datenbrillen : Neuer Blick. In: Logistik Heute Bd. 6 (2015), Nr. c, S. 61
- [Pu12] Puig, J. et al.: Towards an efficient methodology for evaluation of quality of experience in Augmented Reality. In: 2012 Fourth International Workshop on Quality of Multimedia Experience (QoMEX) : IEEE, 2012, S. 188–193

- [Ra14] Radu, I.: Augmented reality in education : a meta-review and cross-media analysis. In: Personal and Ubiquitous Computing Bd. 18 (2014), Nr. 6, S. 1533–1543
- [RMP15] Rashid, Z.; Melià-Seguí, J.; Pous, R.: Bridging Offline and Online World Through Augmentable Smart Glass Interfaces. In: De Paolis, L. T.; Mongelli, A. (Hrsg.): Augmented and Virtual Reality. Bd. 9254. Cham : Springer International Publishing, 2015, S. 420–431
- [Se06] Seffah, A. et al.: Usability measurement and metrics: A consolidated model. In: Software Quality Journal Bd. 14 (2006), Nr. 2, S. 159–178
- [St16] Stocker, A. et al.: Datenbrillengestützte Checklisten in der Fahrzeugmontage, 2016
- [Ta15] Talip, B. A.: Dimension of structural usability for Augmented Reality system. In: 2015 IEEE Confernece on Open Systems (ICOS) : IEEE, 2015, S. 34–39
- [Th16] Theis, S. et al.: Head-Mounted Displays – Bedingungen des sicheren und beanspruchungsoptimalen Einsatzes, 2016
- [TLR14] Tiefenbacher, P.; Lehment, N. H.; Rigoll, G.: Augmented Reality Evaluation : A Concept Utilizing Virtual Reality. In: Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments (2014), S. 226–236
- [TSK15] Tomberg, V.; Schulz, T.; Kelle, S.: Applying Universal Design Principles to Themes for Wearables. In: Antona, M.; Stephanidis, C. (Hrsg.): Universal Access in Human-Computer Interaction. Access to Interaction. Bd. 9176. Cham : Springer International Publishing, 2015, S. 550–560
- [Vu16] Vuzix: M100 Smart Glasses. URL <https://www.vuzix.com/Products/m100-smart-glasses>. - abgerufen am 2016-05-11
- [WD06] Wang, X.; Dunston, P. S.: Compatibility issues in Augmented Reality systems for AEC: An experimental prototype study. In: Automation in Construction Bd. 15 (2006), Nr. 3, S. 314–326
- [WD13] Wang, X.; Dunston, P. S.: Tangible mixed reality for remote design review: a study understanding user perception and acceptance. In: Visualization in Engineering Bd. 1 (2013), Nr. 1, S. 8
- [We16] Wesp, R.: Picken per Datenbrille wird salonfähig. In: Lebensmittel Zeitung Bd. 15 (2016), S. 41–42
- [WEA14] White, J.; Edmondson, J.; Anderson, W.: Next Generation Mobile Computing. In: Next Generation Mobile Computing Bd. 31 (2014), Nr. 2, S. 44–47
- [WL12] Weiß, F.; Leimeister, J. M.: Consumerization. In: WIRTSCHAFTSINFORMATIK Bd. 54 (2012), Nr. 6, S. 351–354
- [WW02] Webster, J.; Watson, R. T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. In: MIS Quarterly Bd. 26 (2002), Nr. 2, S. xiii – xxiii
- [Ya13] Yang-Mao, S.-F. et al.: Exploring Psychophysical Factors Influencing Visibility of Virtual Image Display. In: Hutchison, D. et al. (Hrsg.): Universal Access in Human-Computer Interaction. User and Context Diversity. Bd. 8010 : Springer Berlin Heidelberg, 2013, S. 327–335

IT-Unterstützung in Emergency Management & Response (IT-Rettung)

IT-Unterstützung in Emergency Management & Response (IT-Rettung)

Jens Pottebaum¹, Christian Erfurth²

1 Vorwort zum Workshop

Notfallsituationen sind kritische Situationen, in denen eine Gefahr für Menschen, Infrastrukturen (z. B. Verkehr, Energie, Informationstechnik) und die Umwelt besteht. IT-Unterstützung für das Notfallmanagement kann in der Vermeidung sowie der Verbesserung der Vorbereitung, der Abwehr und der Bewältigung von Schadenslagen helfen. Anwendungsbeispiele zeigen, dass sie z.B. die Koordination zwischen Einsatzkräften verbessern und die Lagefeststellung beschleunigen kann. IT soll in diesen Situationen helfen, komplexe und kritische Situationen zu beherrschen. Allerdings zeigen Erfahrungen, dass der Einsatz von IT als operatives Einsatzmittel oder Entscheidungsunterstützungswerkzeug auch für eine Zunahme der gefühlten Komplexität einer Einsatzlage bei vielen Beteiligten sorgen kann.

Im Fokus des Workshops stehen die besonderen Herausforderungen und technischen Konsequenzen, die sich für die IT in diesem Umfeld ergeben, sowie ein Austausch über aktuelle Erkenntnisse aus der Entwicklung und der Erforschung von Lösungen. Uns interessieren dabei sowohl Ansätze, deren Fokus auf der Unterstützung bei der Rettung und Versorgung einer einzelnen Person liegt, als auch Ansätze, die sich mit einer breiten IT-Unterstützung im Notfall- und Krisenmanagement beschäftigen. Beide Aspekte können z.B. in Großschadenslagen gleichzeitig eine hohe Relevanz haben, wenn beispielsweise ein Rettungseinsatz durch die Beteiligung unterschiedlicher Organisationen und Behörden schnell zu einer komplexen Situation für einzelne Akteure (z.B. Notärzte) und Führungsstellen (z.B. einen Führungsstab) wird. Ein schneller Lageüberblick und ein effizienter, bedarfsgerechter Informationsaustausch zwischen den Beteiligten sind Ziele, die durch geeignete technische Lösungen und (IT-)Equipment vor, während und nach Notfallsituationen erreicht werden sollen.

Georg Neubauer, Projektkoordinator des EU-Projekts EPISECC für das Austrian Institute of Technology (AIT), motiviert die Diskussion im Workshop durch einen Impulsvortrag zur Interoperabilität von Organisationen und IT-Systemen im Krisenmanagement. Er stellt die Ergebnisse einer breiten Analyse der Verwendung von Daten und Informationssystemen in der europäischen Gefahrenabwehr vor und leitet

¹ Universität Paderborn, C.I.K. / HNI-PE / DMRC, Warburger Straße 100, 33098 Paderborn,
pottebaum@cik.upb.de

² Ernst-Abbe-Hochschule Jena, FB Wirtschaftsingenieurwesen, Carl-Zeiss-Promenade 2, 07745 Jena,
christian.erfurth@eah-jena.de

Herausforderungen für die länderübergreifende Zusammenarbeit ab. Er stellt das Konzept eines ‚Common Information Space‘ vor, das Handlungsempfehlungen zur Gewährleistung von Interoperabilität ganzheitlich zusammenfasst und z.B. Standards für den Datenaustausch empfiehlt.

Das Workshop-Thema greifen fünf Beiträge aus unterschiedlichen Perspektiven auf.

Klafft et al. „Challenges in Designing and Distributing a Not For Profit First Aid App Worldwide“ beschreiben anhand einer Fallstudie die Anforderungen und Erkenntnisse bezüglich der Entwicklung von Anwendungen, die in Notfallsituationen die Ersthilfe verbessern sollen. Reuter et al. berichten zur “Evaluation of Alerts and Notifications for Emergency Services based on Cross-Platform Social Media Data” aus dem EU-Projekt EmerGent, wobei sie insbesondere die Feuerwehr-Organisationen betrachten.

Zwei inhaltlich verwandte Beiträge stellen den Einfluss kommunikationstechnischer Komponenten als Bestandteile ganzheitlicher System-Architekturen dar. Barth et al. präsentieren “An Open Service Platform for Multi-Hazard in Action - the PHAROS Pilot Demonstration”, während Raffelsberger et al. unter der Überschrift „Flexible Kommunikations- und Informationslösungen für eine optimierte Einsatzführung von Interventionskräften“ den Mehrwert im Anwendungsbereich „Einsatzunterstützung“ herausarbeiten.

Komplementär zu diesen Themen adressieren Reuter et al. die Abhängigkeiten zwischen „Sicherheit vs. Privatsphäre“. Sie liefern durch die Auswertung ihrer explorativen Studie „Zur Akzeptanz von Überwachung in sozialen Medien im Kontext von Terrorkrisen“ einen Diskussionsbeitrag, der das Workshop-Thema auch in Richtung von Prävention und Terrorabwehr öffnet.

Alle Artikel wurden in einem Begutachtungsprozess durch das Programmkomitee bewertet und erhielten ausführliche Rückmeldungen zu Inhalt und Qualität.

2 Programmkomitee

Christian Erfurth	EAH Jena
Jens Pottebaum	Universität Paderborn
Wolf Engelbach	Fraunhofer IAO
Simon Nestler	Hochschule Hamm-Lippstadt
Georg Neubauer	Austrian Institute of Technology (AIT)
Christian Reuter	Universität Siegen
Thomas Rose	Fraunhofer FIT

Challenges in Designing and Distributing a Not For Profit First Aid App Worldwide

Michael Klafft¹, Maurice Said, Susan Anson, Hayley Watson², Amanda Hughes³, Eridy Lukau⁴

Abstract: First aid apps can assist people in preparing for emergencies, and they can help citizens to bridge the gap between the emergence of a disaster and the time when professional rescuers arrive on the scene. In this paper, we discuss key requirements and factors influencing the uptake, acceptance, and use of first aid apps from an end-user's as well as a distributor's perspective. Based upon theoretical considerations as well as practical experiences from a multi-national first aid app program, best practices for the adaptation of a first aid app to the local user context are identified, recommendations for successful market introduction are formulated, and the capabilities and limitations of first aid apps are discussed. Results presented are based on semi-structured interviews with 29 experts from eight countries, as well as a secondary analysis of Google Analytics data.

Keywords: First aid app, requirements, critical success factors, user behaviour.

1 Introduction

Smartphones and the mobile internet offer new opportunities to provide health information to members of society in developed and developing countries alike. In line with the United Nations' Millenium Development Goals, Hagar and Kartzintel [HK14] identify first-aid as a promising area for mobile health information initiatives. However, they also state that as of 2014, there is still a gap between the need for information and the availability of suitable mobile applications to deliver this information. This paper will address the issue of providing first aid information via mobile applications worldwide at limited cost by discussing the Universal App Program currently managed by the Global Disaster Preparedness Center hosted by the American Red Cross [GD13, GD15]. Section 2 discusses the state of the art in relation to first aid apps. Section 3 introduces the Universal First Aid App program and its key concepts. In section 4 we discuss the success factors that influence the uptake of not for-profit first aid apps that are disseminated globally but adapted and marketed locally by regional organizations in

¹ Jade Hochschule, Management Information Technology Department, Friedrich Paffrath Str. 101, 26389 Wilhelmshaven, Germany. Michael.Klafft@jade-hs.de

² Trilateral Research Ltd., Crown House 72 Hammersmith Road, London W14 8TH, United Kingdom. {Maurice.Said | Susan.Ansan | Hayley.Watson}@trilateralresearch.com

³ Utah State University, Computer Science Department, 4205 Old Main Hill, Logan, UT, United States. Amanda.Hughes@usu.edu

⁴ Fraunhofer Institute for Open Communication Systems, Electronic Safety and Security Systems for the Public and Industries, Kaiserin Augusta Allee 31, 10589 Berlin, Germany, Eridy.Lukau@fokus.fraunhofer.de

different target countries. We present the methodology used for primary and secondary research in our study (section 5), followed by a discussion of key requirements for First Aid Apps. In the final section we discuss user behavior and user feedback on the apps. The article concludes by providing an outlook on future research activities.

2 State of the art

The potential impact of first aid apps is continuously improved by declining smartphone prices and an increased availability of feature phones worldwide. Royston et al. [Ro15] point out that even in low to middle income countries, feature phones that are able to display video and other media, are now becoming commonplace. Using such phones to provide high-quality healthcare information to people who do not have access to trained health workers could protect them from being exposed to ineffective or even harmful treatment [Ro15]. It could also enable people at the site of an emergency to help themselves and each other in cases where professional assistance will take time to arrive, or is unaffordable to affected people due to a lack of health insurance. A challenge exists, however, in providing first aid apps to those many feature phones in developing countries which do not support any mobile internet or Wi-Fi (or where the user cannot afford an internet connection). To counter this problem, Royston et al. [Ro15] propose asking manufacturers to store health-related information in their products by default at the factory, or to disseminate health content and health apps on micro-SD cards. However, the authors also identify a shortage of appropriate content and a lack of investment in such content. Quality issues with first aid app content were also identified by Thygeson et al. [Th12] who analyzed the content of 65 English-language first aid apps and calculated adherence scores of the apps with official guidelines. In many first aid categories, the apps scored on average less than 40% of the maximum points. This includes very important categories like: “assessing each victim for responsiveness, airway patency, circulation, breathing, and medical alert tags”, or “performing logical head to toe checks for injuries”. The study also found that free apps fared particularly poorly in a number of categories, e.g., “appropriate handling of amputated body parts”, where they only received 7.7% of the maximum score (compared to 35.3% of paid apps). Among the top six first aid apps in the study, none were provided for free. The highly varying quality of healthcare apps has raised some regulatory concerns, and some state-run healthcare providers like the British National Health Service (NHS) took the issue into their own hands by creating an Online Health Apps Library (for an extract, see [Na16]), where it “lists and recommends some carefully selected apps”, and where “developers can submit their apps for review and possible listing” [Ka14]. Such signalling helps citizens to choose a reliable and high-quality app among the alternatives on offer. However, to our knowledge, the NHS is the only authority so far which provides a vetting program for health related apps. As a result, in most countries people have to choose a suitable health app on their own. Consequentially, they may select a first aid app that does not meet global and local quality standards or guidelines. In the following sections, we will analyze an app program that places a particular focus on up-

to-date and quality-assured content and is distributed by trustworthy organizations (Red Cross / Red Crescent societies), thus addressing the abovementioned problems related to the quality of content that have been identified in the literature.

3 The Universal App Program

In 2012 the American Red Cross released a free First Aid App for iOS and Android devices (based upon app content licensed from the British Red Cross) that offers easy-to-follow advice for a variety of first aid scenarios. These scenarios are based on the International Federation of Red Cross and Red Crescent Societies' Guidelines for first aid [IF11]. The advice aims to help users to learn and understand the fundamentals of first aid, and to reinforce first aid concepts with those users who already have basic first aid training. The app provides users with information through animations, videos, quizzes and links to complementary information held on the respective national Red Cross society's websites. Figure 1 provides an overview of key topics presented within the app.⁵ While some training courses are related to specific types of medical conditions (e.g. "diabetic emergency", "anaphylaxis"), other categories are relevant for large scale disasters as well as personal injuries, such as "bleeding", or courses explaining how to deal with unconscious people.

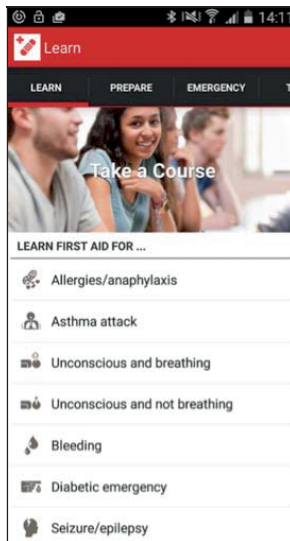


Fig. 1: Key topics of training courses in the First Aid App

Figure 2 shows an example screen for the topic "unconscious and not breathing", with

⁵ All screenshots are taken from the first aid app variant which is distributed by the Swiss Red Cross / Samaritans.

key instructions on how to treat people in such a state and an embedded explanatory video which visually illustrates the instructions.

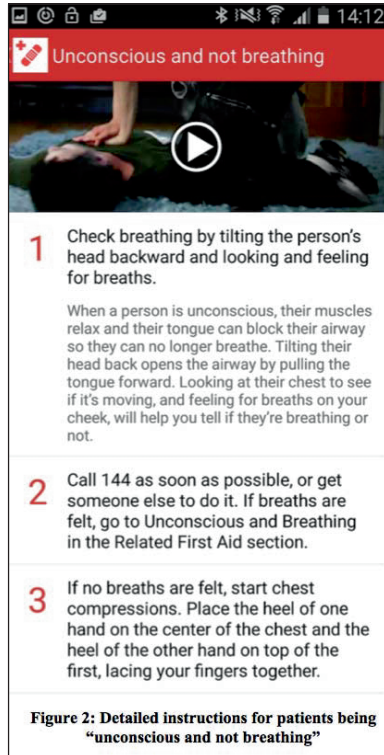


Fig. 2: Detailed instructions for patients being “unconscious and not breathing”

Overall, app content may differ between countries but is structured into five areas.

Learn: Users are guided by simple step-by-step and easy to understand advice on multiple first aid scenarios in order to learn first aid concepts. Such concepts range from allergies/anaphylaxis to strains and sprains, missing persons, burns, shocks, and broken bones, among others. For each emergency case, simple and easy-to-understand explanations are presented, which are supplemented by a video tutorial. The app also provides answers to frequently asked questions for each of the cases. In some cases like allergies/anaphylaxis the app offers a panic button that directly leads to a national allergy center (if available in the country) in order to provide more information and advice. Step-by-step instructions are also equipped with short animations, teaching users how to treat injuries or how to apply a bandage in an emergency.

Prepare: The app offers information to users in order to prepare them for multiple types of emergency situations such as fire, chemical emergencies, earthquakes, landslides, heat

waves, traffic accidents and more. Users benefit from checklists and advice with specific instructions before, during, and after an incident. The information also varies with whether the emergency occurred indoors or outdoors. In any emergency the prepare segment of the app offers a deeplink to related first aid topics of the learn section. In case of an earthquake, users are equipped with checklists such as "Brace overhead light fixtures", hints such as "Stay away from windows to avoid being injured by shattered glass" as well as a deeplinks to "Broken Bone", "Head Injuries" and "Missing Person" topics in the Learn-section of the app.

Emergency: The emergency section differs from the "Learn" section by focusing more directly on instructions for an emergency. Additionally, each of the emergency cases includes a so called "panic button" which directly makes an emergency call if activated.

Tests: By providing a test section a user is able to test how much they have learnt from the app by taking part in offline interactive quizzes. The user can choose between six categories: "Unconscious and breathing, Bleeding, Choking, Heart attack, Prepare and Burns". Each category consists of up to three questions with several answers (multiple choice) or picture puzzles where they have to choose the correct picture. By solving the quizzes in one of the named categories correctly, the user gets rewarded with a badge belonging to that category. These badges can be shared on social media platforms such as Facebook or Twitter. With this so called gamification method, a user is able to obtain achievements in all of the six categories in order to validate the status of their knowledge of first aid skills.

Information/More: This section offers general information about national Red Cross first aid courses, blood donations, first aid shops or volunteer programs. Users can also donate to the national Red Cross or share their stories on using the app on social media channels from the app.

In May 2013 the Global Disaster Preparedness Center (GDPC)⁶ launched the Universal App Program in order to introduce the application into new markets within the Red Cross/Red Crescent network across the world. The idea behind this approach was to exploit synergies and economies of scale, making the introduction of the app possible in countries independent of resources. Nevertheless, the internationalization of the app requires a strong localization of the application. In order to facilitate the development process as well as the translations, and video editing, the GDPC offers access to a content management system and different tools to participating national Red Cross societies. This concept enables an easy adaptation to each country with regard to language, content, emergency numbers as well as local additional information on training courses. As of June 2016, 79 national Red Cross societies decided to join the program and have launched a local version of the First Aid App.

⁶ The Global Disaster Preparedness Center is a joint initiative between the American Red Cross and the International Federation of Red Cross/Red Crescent societies that strives to enhance disaster preparedness capacities at community, local, and national levels throughout the world.

4 Critical success factors for app uptake

As each Red Cross national society (in the following labeled as „host organization“) participating in the Universal App Program has full ownership of their app once it is launched, app uptake depends on (a) the assessment of the first aid app program by the host organization, and (b) the evaluation and assessment of the app by ordinary citizens, and necessary facilitating conditions for them to actually use the app. Success factors related to these two groups of actors will be discussed in the following subsections.

4.1 Success factors related to the citizens

As discussed in the introduction, the availability, perceived quality of content, and costs are key factors for app uptake from the perspective of citizens. These factors influence technology acceptance, which is the crucial factor influencing the uptake of software systems. Previous research by Davis [Da85] indicates that design features of an information system induce a cognitive response with potential users, based upon the system’s perceived usefulness and perceived ease of use. In a second step, this cognitive response is translated into an affective response, that is, the intention to use the system. The prevalence of facilitating conditions (in our case, for example, the availability of smartphones with Android 4.1. and above or iOS8 and above) enables the transition of intending to use an app to actual use. This early model by Davis was later modified and extended to be more suitable for a consumer and healthcare context. Venkatesh et al. [VTX12] focus on the consumer’s perspective. Similarly to Davis, they identify “performance expectancy” and “effort expectancy” as key success factors. However, in the consumer context, additional factors need to be taken into account, including: hedonic motivation (i.e., user enjoyment), price value characteristics of the application, and habit (that is the ability to perform actions in the application automatically without learning, due to prior knowledge and experience). Another key factor is social influence, which is the influence of others on whether or how to use or not to use an application (e.g., because it is perceived as cool by others). While the models by Davis and Venkatesh et al. are not specific to a certain type of application, Cho et al. [Ch14] explicitly studied determinants of adoption of smartphone health apps. They identify three additional factors that influence app uptake: health consciousness (i.e., the “extent to which individuals have interest in and are aware of their own health conditions and well-being”), health information orientation (that is, “the extent to which a person actively seeks health information through diverse sources”), and internet health information use efficacy (i.e., the individual’s “cognitive ability to seek health information through the internet”). Although none of the above-mentioned studies looked into first aid apps in particular, findings should have some relevance for first aid apps as they are a specific sub-category of health apps and information systems in general. However, one caveat needs to be ventured: the most specific of the studies cited [Ch14] was performed with college students only, thus, whether the findings can also be applied to substantially different user groups, e.g., individuals with low literacy levels,

still needs to be verified. The overview of success factors indicates that app uptake will be affected by the performance of the app itself (by being easy to use and by providing useful information to the citizen), but also by the marketing and dissemination activities of the host organization (as these may affect factors like social influence), and by addressing hedonic motivation (e.g., enhanced gamification features of the app). Other factors that indirectly influence app uptake are the availability of competing, for-free apps, and the presence of local threats in a country. Those who already have a first aid app, are unlikely to install a second app (due to limited additional usefulness), and the likelihood of needing first aid (due to an increased risk of, e.g., natural disasters or terrorism) will affect the perceived usefulness and value of the first aid app.

4.2 Success factors related to the host organizations

Host organizations will only adopt and market the app if these activities are in line with their own organizational goals and if they can manage the effort for translation, localization of content, and adaptation to the national context. This will be influenced by (a) the flexibility of the app structure, (b) the quality and usability of the underlying content management system, and (c) the local need for a new first aid app (mainly influenced by the eventual availability and quality of competing not-for-profit first aid apps in the national market). Other factors that influence app uptake by the host organization are legal issues (e.g., product liability), and their resources (both in terms of financial resources and skilled personnel, in particular personnel with some knowledge of and affinity to information technology, and personnel with medical and language expertise to provide precise translations of app content). The perceived effort needed to introduce the app will be determined by the expected quality and quantity of support provided by the GDPC. Figure 3 provides an illustration of key factors that influence the uptake of a first aid app that is globally provided yet locally adopted and marketed by not for-profit intermediaries, in this case the host organizations.

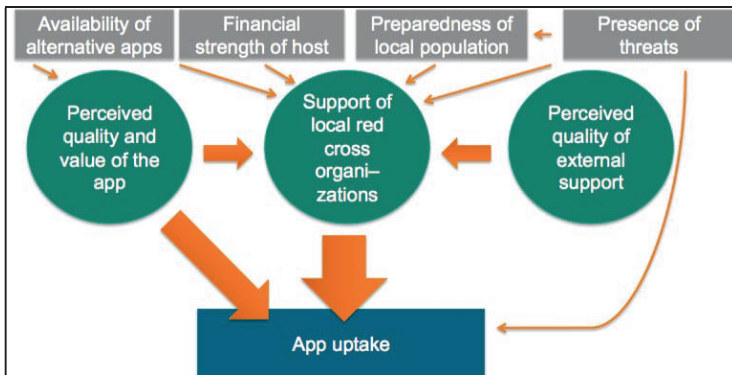


Fig. 3: Key factors influencing app uptake at the host organization and by citizens

5 Method

In order to identify requirements for a first aid app framework that is promoted globally but adapted to the local context, semi-structured interviews were conducted in 2015 with 29 interviewees from 8 socially, economically and culturally diverse countries. The interviews addressed issues including expectations, the goals and purposes of introducing the app, and experiences during customization, testing, and roll-out. Interviewees were members of the national Red Cross / Red Crescent organizations in the countries involved in the study and, had been involved in the actual adaptation or launch of the First Aid App in their respective country. Wherever possible, individual interviews were conducted. However, in a few instances, group interviewees were conducted (e.g. in cases where one of the interviewees had only limited knowledge of English). In most cases, interviews were conducted in person, whilst others were conducted virtually via Skype. The following table provides an overview of the countries analyzed, the number of interviewees in each country, and the interview format used.

Country	Number of interviewees	Interview format
Czech Republic	3	1 individual face-to-face interview 1 face-to-face group interview
Hong Kong	5	5 individual face-to-face interviews
Ireland	3	3 individual virtual interviews
Israel	2	2 individual virtual interviews
Malta	5	5 individual face-to-face interviews
Myanmar	4	1 face-to-face group interview
Iceland	5	5 individual face-to-face interviews
Switzerland	2	1 face-to-face group interview

Tab. 1: Setting of the interviews

Expert interviews were conducted to analyze the host organizations' views. Experts were also asked about the feedback that they received from the citizens who used the app. The interviews were recorded and transcribed and transcripts were used to identify system requirements with open coding.

In order to gain a deeper understanding of the citizens' views, user comments on Google Play and the Apple App Store were analyzed. User comments were coded and categorized into "positive", "negative", or "neutral". A comment was considered positive when it only contained positive adjectives or attributes, or a recommendation to

other users to use the app. A comment was considered negative when it only contained negative adjectives and attributes, or when it reported problems with the app or voiced criticism of the app. All other comments, including those with mixed attributes, were classified as neutral. In addition to the analysis of user comments, data on user behavior were extracted from Google Analytics to analyze patterns determining app download and usage activities.

6 System requirements from the host organizations' perspective

The analysis of expert interviews revealed five recurring items consistently mentioned by several host organizations, which will be discussed in this section.

User friendliness of the content management system. Since first aid information material is often prepared by medical experts, the content management system for the app must be easy to use even for non-IT professionals without the need for time-consuming training. Most host organizations do not have the resources to hire external IT professionals for system adaptation and implementation.

Localization of content. It should be possible to adapt the basic content of the app, and to add additional topics of local relevance. Examples of locally specific added content included information on snake bites in Switzerland (with an emphasis on specific poisonous snake species that are prevalent in the country), information on stabbing attacks (terrorism) or accidents related to swimming pools in Israel, information for beachgoers in Malta, and a special information feature on Ebola in the Czech Republic (there weren't any Ebola cases, but the additional feature addressed a general public interest on the Ebola disease during the recent epidemic). A second aspect is the adaptation of first aid content to local guidelines and the local context. The initial content of the app is based on IFRC first aid guidelines [IF11], but these always have to be reviewed against the local context. For example, recommendations in countries where ambulances are available within minutes and hospitals are reachable in a short time will substantially differ from recommendations for remote areas without any medical infrastructure. Furthermore, some first aid practices are internationally contested, so that adaptation to what is locally perceived as the best practice must be feasible. One example for a contested practice is whether to administer aspirin orally to a person who appears to be suffering from a cardiac infarction. Doing so is recommended in some countries (typically those with a poor medical infrastructure), but it is considered counterproductive in many countries or areas where ambulances arrive within minutes and medical experts will be able to administer more powerful drugs intravenously so that these take effect long before the aspirin has been digested.

Language adaptation. An app that is distributed and marketed worldwide has to be able to cope with multiple languages without any restrictions. This refers to the number of languages in which the content has to be implemented (Switzerland, for example, required 4 national languages plus English), and the handling of different character sets

and text orientations within the same app. The latter is, for example, relevant for Israel, where content is offered in English (Latin alphabet, text from left to right), and in Hebrew (Hebrew alphabet, text written from right to left). Similar issues were reported in Hong Kong and Myanmar. Language adaptation should not only consider characters and text orientation, but also alphabetic ordering of content, which has to be easy to implement for any language offered in the app. In order to enhance usability and for political reasons (equal representation of all population groups), it must be possible to select and store any of the available languages as the default language in the app's settings.

Local storage of the app. In order to assure instant availability of the app in all emergency situations, all app content including video material must be stored locally on the smartphone (in order to guarantee that the app works without delay even in case of breakdowns or loss of the mobile phone network). In the Czech Republic, Red Cross members also required the possibility to store and launch the app from an SD card (a feature requested by citizens who complained that the initial version of the app consumed too much memory on the smartphone itself). This requirement is also in line with recommendations from Royston et al. [Ro15].

Sharing of the app between smartphones. In some countries where Internet access is either very expensive or unreliable or unavailable, it will be necessary to provide the possibility to share the app between devices. In addition to copying / exchanging the app using SD cards (as already mentioned above), users in Myanmar reported that they use Zapyta for app distribution, which is an app that establishes a Wi-Fi-connection between smartphones and then shares content between these devices.

The interviewees pointed out that the First Aid App's main purpose is (a) to generate or renew interest in the topic of first aid (with the goal to attract more participants to first aid courses), and (b) to help people in refreshing basic first aid knowledge, as practical experience shows that within three months, people already forget about 85% of the information that they received during a first aid course. By keeping users better trained, the app can enhance emergency preparedness and strengthen the resilience of laypeople in case of disasters, thus bridging the gap between the onset of disasters and the time when professional help arrives. In areas where first aid training is not available at all or too expensive, the app can provide citizens with some basic information on how to deal with emergencies once they occur.

7 User behaviour and user feedback

In order to identify when and how often people downloaded and used the app, Google Analytics data from the eight countries listed in Table 1 plus data from Mexico was analyzed. Central to this analysis was how successful the app had been in different countries in terms of downloads (as a percentage of the population). Table 2 shows the prevalence of the app in the countries under analysis. Some additional data, which were

analyzed, was the average duration of a session and the number of screens viewed in a session.

Country	Percentage of population who downloaded the app	Average session duration (minutes)	Average number of screens per session
Czech Republic	0.43%	2:29	7.42
Hong Kong	0.32%	2:54	9.92
Ireland	0.23%	2:16	7.21
Israel	0.72%	2:26	7.70
Malta	0.70%	2:25	8.20
Myanmar	0.01%	2:41	13.53
Iceland	3.97%	2:44	7.02
Switzerland	0.31%	1:58	7.27
Mexico	0.08%	2:50	8.66

Tab. 2: Downloads and usage statistics (as of Nov. 2015)

Data indicates that the success of the app varies between countries. In Iceland, the app has been very successful with downloads at 3.97% of the population. Typically, crisis communication apps reach up to 3% of the population directly, depending on the type of app, the threat levels in the region, and marketing activities (for examples see [BK16, Kr14]). The success in Iceland may be partly due to extremely high internet affinity (98.2% of the population, [WB15]). Additionally, the success in Iceland seems to have been affected by three factors in the app's rollout: 1) it coincided with the Icelandic Red Cross' 90th anniversary celebrations which invited a lot of publicity, 2) the Icelandic Red Cross benefited from wide and frequent media exposure and also created a cartoon character to promote first aid and, 3) unlike in other countries most of the Icelandic population live in the country's capital where the bulk of the app's marketing was focused making it easier for a larger percentage of the population to learn about the app. At the other end of the scale, in Myanmar only 0.01% of the population downloaded the app. This was partly due to the rather low available income and therefore limited availability of smartphones, but also due to the fact that sharing apps peer-to-peer is common because of high internet costs (Google analytics does not reliably detect such shared apps). App usage duration (between 2 and 3 minutes) and average number of screens viewed (typically 7-9) are similar in all countries under analysis, with Myanmar being again an exception with an average of 13.53 screens viewed per session.

A second point of inquiry was if the download rates were influenced by disasters or other major events. In order to analyze this a number of significant events in each country were selected, and download statistics were examined for a peak during this event. Although no general pattern could be identified, some incidents did in fact appear to result in a substantial increase in downloads. Figure 4 shows how a series of terrorism-related stabbing incidents (and the timely release of first aid information related to such incidents) influenced the download rate and app usage in Israel. The first stabbing incident was reported on October 3rd, 2015. The intensity of the attacks increased from October 7th onwards, with app usage starting to increase rapidly around October 8th.

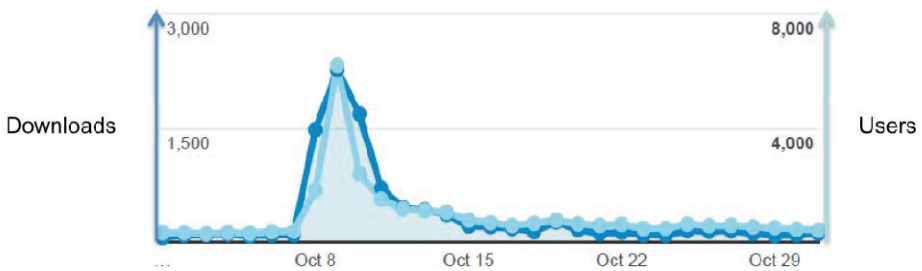


Fig. 4: Effect of stabbing attacks in Israel on downloads (left scale) and user statistics (right scale)

However, it is not always terrorism and disasters that drive app usage. In the Czech Republic, for example, usage and download intensity increased following a large scale sports event (the Fintail world championship).

Coding of the comments left by users on Google Play and the Apple App Store revealed that the overwhelming majority (>85%) of users gave positive feedback and / or found the app useful / helpful. In Switzerland, Israel, and the Czech Republic, for example more than 85% of comments were positive. Negative comments referred mostly to some bugs and technical errors. However, these results cannot be regarded to be representative because they only include the views of those who actively decided to leave a comment in the Google Play / App Store, which is only a tiny fraction (<<1%) of overall users and most likely contains a self-selection bias.

8 Summary and Outlook

Within the study, we were able to derive key requirements that have to be met if a not-for-profit first aid app shall successfully be distributed worldwide (section 6). The study then analyzed an example of such an app (the First Aid App provided by the GDPC) and showed that this app does by and large meet the requirements of the national

distributors (here: the Red Cross / Red Crescent societies who have to decide whether to launch the app in a country), as well as requirements of the citizens using it (section 7). Some additional development focus still has to be placed on expanding the multi-language capabilities of the app, such as displaying content alphabetically ordered in any selected language. Figures on app uptake indicate that, under optimal circumstances (like in Iceland), approximately 4% of the population can be reached by a first aid app if it is distributed via app stores on a voluntary basis. This number may, however, be further increased through cooperations with mobile phone manufacturers (pre-installation of the app), or if governments decide to make installation of a first aid app a legal requirement for all smartphones to be sold in a country.⁷ Expert interviews indicate that first aid apps can potentially increase preparedness levels for an emergency

- by motivating people to look into the topic of first aid and to register for a first aid course,
- by refreshing existing knowledge on first aid (which is often initially obtained when acquiring a driving license or doing military / civil service, but rarely updated ever since), and
- by giving access to first aid information to people who otherwise would not have any access to first aid training.

However, they also indicate that first aid apps are a complement rather than a replacement for regular in-person training courses, because the latter are deemed to be much more effective in preparing citizens for emergencies. From the citizens' perspective, app uptake is driven by (a) media coverage of the app, and (b) sometimes by the prevalence of disasters, threats, or large scale events. However, in our study, only a limited number of incidents could be looked into, which did not reveal any clear pattern on which types of incidents and events actually drive app uptake and why some other incidents did not. Analyzing the relation between external threats / incidents and app uptake, complemented by a multifactorial statistical analysis, was beyond the scope of this study and remains an issue for future research.

Acknowledgements

We would like to thank the Global Disaster Preparedness Center (GDPC) for funding this study. In particular, we would like to thank Karin Metz (GDPC) for her comments on this paper. We would also like to thank all interviewees for the valuable input provided to our study. Furthermore, we would like to thank all project partners (Trilateral Research, Utah State University, Asian Disaster Preparedness Center, Fraunhofer FOKUS) for their valuable assistance.

⁷ A similar approach was chosen by the Dutch government when introducing its NL-Alert disaster warning service. Mobile phone manufacturers were required to pre-configure all new cellular phones so that they could automatically receive cell broadcasts issued by this service.

References

- [BK16] BKS Portal (Ed.). KATWARN mit 100.000 Nutzern; Land sowie acht Landkreise und drei Kreisfreie Städte setzen bereits KATWARN ein. <https://www.bks-portal.rlp.de/organisation/isim/aktuelles/katastrophenwarnsystem-katwarn-mit-100000-nutzern>, as of June 20th, 2016
- [Ch14] Cho, J., Quinan, M. M., Park, D., and Noh, G.-Y. 2014. Determinants of Adoption of Smartphone Health Apps among College Students. *American Journal of Health Behaviour* 38 (2014) 6, pp. 860-870.
- [Da85] Davis, F. D. A Technology Acceptance Model for Empirically Testing new End-User Information Systems: Theory and Results. Massachusetts Institute of Technology, Cambridge (USA), 1985.
- [GD13] Global Disaster Preparedness Center. Universal App Program: First Aid - Frequently Asked Questions. http://preparecenter.org/sites/default/files/universal_app_program_faqs.pdf, as of December 12th, 2015.
- [GD15] Global Disaster Preparedness Center.. A Universal App for First Aid. <http://preparecenter.org/content/universal-app-first-aid>, as of December 12th, 2015.
- [HK14] Hagar, C.; Kartzinel, H. 2014. Healthcare Information for All By 2015: Preliminary findings and future directions. *Information Development* 9/14, pp. 1-8.
- [IF11] International Federation of Red Cross and Red Crescent Societies. International first aid and resuscitation guidelines 2011 - For National Society First Aid Programme Managers, Scientific Advisory Groups, First Aid Instructors and First Responders. IFRC, Geneva, 2011.
- [Ka14] Kamel Boulos, M. N.; Brewer, A. C.; Karimkhani, C.; Buller, D. B.; Delavalle, R. P. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. *Online Journal of Public Health Informatics* 5 (2014) 3, pp. 229 ff.
- [Kr14] Kraudzun, H. Eine App für Katastrophen. *Märkische Oderzeitung* 20.11.2014. <http://www.moz.de/artikel-ansicht/dg/0/1/1347999> as of June 20th, 2016.
- [Na16] National Health Service. Tools. <http://www.nhs.uk/tools/pages/toolslibrary.aspx>, as of April 15th, 2016.
- [Ro15] Royston, G.; Hagar, C.; Long, L.-A.; McMahon, D.; Pakenham-Walsh, N. Mobile health care information for all: a global challenge. *The Lancet Global Health* 3 (2015) 7, pp. e356-e357
- [Th12] Thygerson, S. M.; West, J. H.; Rassbach, A. R.; Thygerson, A. L. iPhone Apps for First Aid: A Content Analysis. *Journal of Consumer Health on the Internet* 16 (2012) 2, pp. 213-225.
- [VTX12] Vankatesh., V.; Thong, J. Y. L.; Xu, X. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly* 36 (2012) 1, pp. 151-178.
- [WB15] The World Bank. Internet user per 100 people. <http://data.worldbank.org/indicator/IT.NET.USER.P2>, as of May 2nd, 2016

Sicherheit vs. Privatsphäre: Zur Akzeptanz von Überwachung in sozialen Medien im Kontext von Terrorkrisen

Christian Reuter¹, Gordian Geilen² und Robin Gellert³

Abstract: Nach den terroristischen Anschlägen in Paris 2015 und Brüssel 2016 wurde das Bedürfnis nach mehr Sicherheit und Überwachung im Internet laut. Als Folge der Enthüllungen der Überwachungs- und Spionagetechniken der National Security Agency (NSA) durch Edward Snowden 2013 konnte in der Bevölkerung aber auch ein Aufschrei nach erhöhtem Schutz der Privatsphäre im Internet wahrgenommen werden. Die geschilderten Ereignisse verdeutlichen die gegensätzlichen Wünsche nach Sicherheit und Überwachung im Internet sowie Schutz der Privatsphäre. Im ersten Teil dieses Beitrags stellen wir den Stand der Forschung im Bereich Terror, Sicherheit und Privatsphäre in sozialen Medien dar. Im zweiten Teil führen wir eine explorative Studie durch, um zu beleuchten, ob Bürgerinnen und Bürger in Krisenzeiten bereit wären, ihre Privatsphäre im Internet, vor allem in sozialen Netzwerken, zugunsten von Sicherheit zu reduzieren. Basierend auf qualitativen Daten zeigt diese Arbeit Meinungscluster und Tendenzen in Bezug auf das Nullsummenspiel „Sicherheit und Privatsphäre“.

Keywords: Soziale Medien, Sicherheit, Privatsphäre, Notfallmanagement

1 Einleitung

Sicherheit oder Privatsphäre? Lassen sich diese beiden Maxime in Einklang miteinander bringen? Oder befinden sie sich in einem nicht kompromissfähigen Nullsummenspiel? Edward Snowden löste mit seinen Spionageaufdeckungen einen enormen Anstieg an dem Bedarf nach mehr Privatsphäre im Internet aus und stieß damit auf positiven Widerhall. Konträr dazu steht die Omnipräsenz des Terrorismus, sodass mittlerweile sogar von „Terrorkrisen“ [We02] gesprochen wird. Der schwerwiegendste Angriff auf Europa war der Anschlag auf die Pariser Innenstadt im Dezember 2015, bei dem über 130 Menschen getötet wurden. Daraufhin wurden rapide neue Sicherheitskonzepte entworfen, um Anschläge wie diesen in Zukunft zu verhindern. Dabei war vor allem ein Detail von essentieller Bedeutung: Die Attentäter hatten sich vorab im Internet über ihren bevorstehenden Anschlag ausgetauscht.

Die beiden Leitgedanken „Sicherheit“ und „Privatsphäre“ haben beide großen Stellenwert. Doch wenn sie sich in einem augenscheinlichen Widerspruch befinden, ist

¹ Universität Siegen, Institut für Wirtschaftsinformatik, Kohlbettstr. 15, 57072 Siegen (alle Autoren), christian.reuter@uni-siegen.de

² gordian.geilen@student.uni-siegen.de

³ robin.gellert@student.uni-siegen.de

dann deren Koexistenz überhaupt möglich? Im Zuge dieser Arbeit soll daher zunächst der Wunsch nach Privatsphäre der Bürger analysiert werden, um anschließend über das Potential zur Identifizierung und Überwachung von Terroristen im Internet zu diskutieren. Nach der Darlegung des Stands der Forschung werden der Fokus und die Ziele dieser Arbeit erläutert (Kap. 2). Anschließend werden die Methodik der empirischen Untersuchung sowie die erzielten Ergebnisse dargestellt (Kap. 3). Fortan folgt eine Diskussion über die Resultate (Kap. 4), um abschließend ein Fazit geben zu können (Kap. 5).

2 Stand der Forschung: Terror, Sicherheit und Privatsphäre in sozialen Medien

Terror, Sicherheit, Risiko, Vertrauen und Privatsphäre stehen in der Fachliteratur in einem nicht immer eindeutigen Verhältnis. Dies hängt teilweise mit der vagen und nicht eindeutigen Begriffsverwendung sogenannter „schillernder Begriffe“ [We07] zusammen, deren „wahre Bedeutung“ sind je nach Kontext und Disziplin unterscheidet. Selbst innerhalb des Systems von Behörden und Organisationen mit Sicherheitsaufgaben (BOS) können terminologische Unterschiede attestiert werden [Re11].

[Wo12] zeigen, dass die norwegische Bevölkerung nach den Terroranschlägen im Juli 2011 ein erhöhtes interpersonelles und institutionelles Vertrauensgefühl gewann. Da Vertrauen und Angst als Gegensatzpaar verstanden werden kann, kann dies als Kontrast zu den Ergebnissen von [Gr09] stehen, wonach in der Bevölkerung die Angst vor terroristischen Anschlägen (gerade nach den Anschlägen auf das World Trade Center) relativ hoch ist, was sie auch auf die hohe mediale Präsenz des Themas zurückführen. Auch [Ja07] betont die Relevanz von Medien, vor allem die Fernsehberichterstattung, wenn es um die Angst der Bevölkerung nach Terroranschlägen geht. Doch nicht nur das Fernsehen, sondern gerade das Internet wird von terroristischen Gruppen zunehmend genutzt, um eine größere Zielgruppe zu erreichen [SeJa11], was beispielsweise durch Sharing-Plattformen wie YouTube oder soziale Netzwerke wie Facebook begünstigt wird.

Die Frage ist nun, ob die Bekämpfung dieser Angst und somit die Schaffung eines Gefühls der Sicherheit der Bevölkerung nur durch gravierende Einschnitte in die Privatsphäre erreichbar wäre. [DS04] untersuchten nach den Anschlägen am 11. September 2001 die Bereitschaft der amerikanischen Bevölkerung, Bürgerrechte gegen höhere persönliche Sicherheit einzutauschen und fanden heraus, dass sich diese Bereitschaft mit zunehmendem Gefühl nach Bedrohung verstärkt. Eine Studie, die diesen Zusammenhang in Europa untersuchte, konnte hingegen keine signifikante Korrelation erkennen [FR15]. [De02] diskutiert den Zusammenhang von Freiheit (und Privatsphäre gehört dazu) und Sicherheit im Kontext des Terrorismusbekämpfungsgesetz, schlussfolgert jedoch, dass dieses eher der Dialektik von Schutz und Angst folgt. Zur Abwägung von Sicherheit und Freiheit im Kontext von

Terror stellt [Le04] fest, dass Sicherheit auch keine abwägungsfähige Position darstellt, da sie sich nur negativ als Abwehr von Gefahren definieren lässt.

[BR15, S.10] machen deutlich, dass seit der NSA-Affäre 2013 und den Enthüllungen durch Edward Snowden, die aufkommende Prämisse die Verschlüsselung der eigenen Daten im Internet und gerade in sozialen Netzwerken war. Jedoch gab es in der Bevölkerung auch vor Snowden schon den Wunsch nach mehr Privatsphäre [MD03]. [BR15, S.10] sprechen hier von einer erhöhten öffentlichen Besorgnis über persönliche Daten und Privatsphäre, wodurch sich ebenso der Einsatz von Verschlüsselungs-Tools, bspw. Programme wie „Tor“, bei Privatpersonen erhöhte. Doch auch die Betreiber sozialer Netzwerke selbst verwenden nun erhöhte Sicherheitsvorkehrungen, wenn es den Schutz der Privatsphäre betrifft. Eine weitere wichtige Erkenntnis als direkte Folge der Snowden-Enthüllungen ist die Etablierung sogenannter „Anti-Facebook“-Netzwerke (ebd.), welche vermehrt auf Privatsphäre der Nutzer achten und Software verwenden, die eben jene Privatsphäre schützen. Aus den bisherigen Ausführungen lässt sich konstatieren, dass zumindest ab Mitte 2013 Privatsphäre einen immensen Stellenwert innerhalb der Gemeinschaft der Internetnutzer und Mitglieder sozialer Netzwerke einnahm. Eine solche Entwicklung in der Etablierung von Domänen „unterhalb des Radars“ kann und wurde, wie die Pariser Anschläge verdeutlichen, jedoch auch von Terroristen verwendet, wie diverse Internetpräsenzen von Al-Qaida, der Taliban, FARC oder nicht zuletzt dem IS darlegen.

[Ch15] betont, dass erst durch die von Terroristen verfolgte Kommunikationsstrategie im Web 2.0 das Verbrechen zu einem terroristischen Akt wird. [Tu15] beschäftigt sich mit den generellen Maßnahmen und Strategien der Terrorbekämpfung, unter anderem auch mithilfe von Kommunikation. Mit der Ermittlung möglicher Methoden zur Identifizierung Terrorverdächtiger in sozialen Netzwerken haben sich Jeberson und Sharma [JS15] befasst. Diese sind jedoch nicht dazu geeignet, die terroristischen Zellen an sich ausfindig zu machen bzw. ihre Aktivitäten zu identifizieren, sondern beziehen sich auf jene Gruppen, die radikal handeln, aber die Grenze zum Terrorismus noch nicht überschritten haben, wie in Deutschland beispielsweise die Salafisten. Darüber hinaus besteht auch die Möglichkeit der Einbeziehung von öffentlich zugänglichen Nutzer-Postings zu Zwecken der Informationsgewinnung durch die Behörden in Krisensituationen. Sutton et al. [Su08] sprechen in diesem Kontext von sogenannten *Backchannels*, einer besonderen Form des Data Minings. Hierzu zählen Nutzerbeiträge, Videos, Geotags, Bewegungsprofile und dergleichen. Hinsichtlich der Früherkennung bzw. Überwachung terroristischer Aktivitäten im Netz können diese -Informationen in Kombination mit anderen Methoden als weiteres Überwachungsinstrument zum Einsatz kommen. [Re13] beschäftigt sich mit der Förderung der Entstehung von Communities in sozialen Medien, was [Lu16] mit Methoden der Nachvollziehbarkeit des Übergangs von Publics zu Communities aufgreift.

Eine generelle Grundlage zur Anwendung von Methoden zur Identifizierung und Überwachung potentieller terroristischer Aktivitäten ist die schon seit langem kontrovers diskutierte Vorratsdatenspeicherung, bei der es sich um ein Konzept zur kurzzeitigen

Aufbewahrung aller benutzergenerierten Daten handelt. Dies erlaubt ein Abbild der Kommunikation im (deutschsprachigen) Web, auf welche die Behörden durch die Kooperation mit IT-Dienstleistern Zugriff erhielten. Auf diese Weise ist zwar keine Überwachung in Echtzeit möglich, jedoch wird sich davon eine erhöhte Aufklärungsrate nach begangenen Straftaten versprochen. Dies stößt in der deutschen Bevölkerung bislang auf erheblichen Widerstand aufgrund der damit einhergehenden immensen Einschränkung der Privatsphäre der Internetnutzer [Ko16]. Beispielsweise haben die Politiker Burkhard Hirsch, Gerhart Baum und Sabine Leutheusser-Schnarrenberger in diesem Sinne eine Erklärung zur Vorratsdatenspeicherung abgegeben, um unbegründete Eingriffe in die Privatsphäre der Internetnutzer zu verhindern [Ku16]. Inwiefern dieses Instrument somit tatsächlich zur Anwendung kommen könnte, bleibt vorerst fraglich.

Zusammenfassend zeigt die Literaturrecherche, dass das Internet im Allgemeinen und die sozialen Netzwerke im Speziellen von allen beteiligten Akteuren (Nutzer, Behörden als auch terroristische Gruppierungen) aktiv genutzt werden. Zum anderen steht ebenso fest, dass es Methoden gibt, um Antiterrormaßnahmen im Internet zu etablieren, diese sich jedoch alle im bisherigen legalen Rahmen bewegen sollten und durch die Sensibilität in Folge der Spionageaffäre um Edward Snowden begrenzt zu sein scheinen. Der entscheidende Aspekt hierbei ist die Grenze zwischen der gewünschten Privatsphäre der Bevölkerung und deren Sicherheitsbedürfnis. Da in Mitteleuropa ein Anschlag stattfand und dieser über ein Netzwerk geplant worden war, stellt sich die Frage, ob eine Bereitschaft besteht, die bisherigen Methoden, wie von Jeberson und Sharma [JS15] erläutert, weiter auszubauen und die Privatsphäre der Nutzer sozialer Medien einzuschränken, um den Behörden einen größeren Handlungsspielraum innerhalb des Internets zu ermöglichen. Anhand der vorliegenden Literatur kann folgende Hypothese formuliert werden: Im Kontext der Terrorismuswahrnehmung sind trotz des Wunsches nach Privatsphäre die Nutzer sozialer Medien bereit, Teile ihrer Privatsphäre zugunsten erhöhter Sicherheit zumindest partiell und temporär aufzugeben. Es soll eruiert werden, ob für erweiterte Sicherheitsmaßnahmen (weltliche und digitale) Akzeptanz in der Bevölkerung besteht – gerade wenn ein gewisser Freiheitsverlust damit einherginge.

Die Forschungsfrage lautet: *Welche Auswirkungen haben terroristische Aktivitäten auf das Sicherheitsempfinden und inwieweit ist die Bevölkerung bereit, auf Freiheiten zu Gunsten von erhöhter Sicherheit zu verzichten?*

3 Empirische Studie

Um die dargestellten Fragestellungen und Ziele dieser Arbeit zu erreichen, wurde eine mehrheitlich quantitative, jedoch dennoch explorative, Umfrage durchgeführt, welche sowohl das Sicherheitsempfinden der Bevölkerung als auch deren Wünsche erfassen soll. Der Fokus dieser Umfrage lag auf der Erschließung der Frage, ob die Probanden für das Versprechen nach mehr Sicherheit auf Teile ihrer Privatsphäre verzichten würden. Die quantitativen Fragen enthielten sowohl nominales als auch ordinales Skalenniveau.

Ersteres bezog sich vor allem auf Fragen der Wünsche oder bisheriger Erfahrungen der Probanden, letzteres auf von uns gestellte Aussagen, welche die Probanden bewerten sollten. Dadurch konnte nicht nur ein Cluster an individuellen Erfahrungen und Meinungen erstellt, sondern auch, im Falle der persönlichen Meinungen der Probanden zu utopischen Aussagen, ein Stimmungs- bzw. Prioritätencluster angelegt werden.

1. Nutzen Sie soziale Netzwerke?
2. Welche sozialen Medien nutzen Sie am häufigsten?
3. Weshalb nutzen Sie soziale Medien?
4. Ich habe auch schon einmal in sozialen Netzwerken beobachtet, dass radikale Gruppen oder Einzelpersonen dort Präsenz zeigen.
5. In welcher Form haben Sie bereits Extremismus bzw. Radikalismus in sozialen Medien erlebt?
6. Um welchen Inhalt ging es dabei?
7. Wie reagieren Sie, wenn Sie Kommentare und Postings solcher Gruppen oder Personen lesen?
8. Nach den Anschlägen von Paris: Wie sicher fühlen Sie sich in Deutschland?
9. Wie empfinden Sie die vorhandenen Sicherheitsmaßnahmen gegen Terrorismus (z. B. Polizei und Grenzschutz) in Deutschland?
10. Welche der folgenden Sicherheitsmaßnahmen würden Ihr Sicherheitsgefühl verstärken?
11. Wenn es der Aufspürung von Terroristen diene, würde ich auf Bereiche meiner Privatsphäre in sozialen Netzwerken verzichten (z. B. durch generelles Erfassen privater Nachrichten).
12. In welchen Bereichen sind Sie dazu bereit, auf Teile Ihrer Privatsphäre in sozialen Netzwerken zu verzichten?
13. Geschlecht
14. Wie alt sind Sie?
15. Welchen Schulabschluss haben Sie erlangt?
16. Welcher Tätigkeit gehen Sie nach?

Tabelle 1: Fragen

Auf die mithilfe von Google Forms erstellte Online-Umfrage, welche von uns über soziale Medien, E-Mail-Verteiler und persönliche Ansprache an die Teilnehmer der Studie verteilt wurde, erhielten wir Antworten von 61 Probanden aus Deutschland ab 18 Jahren. Der größte Anteil der Probanden lag in der Altersgruppe zwischen 18 und 30 Jahren (67%), mit einem ausgewogenen Verhältnis beim Geschlecht der Teilnehmer (49% weiblich, 51% männlich). Mit etwa der Hälfte der Probanden machten Studierende (49%) den Großteil der Teilnehmer aus (Abb. 1).

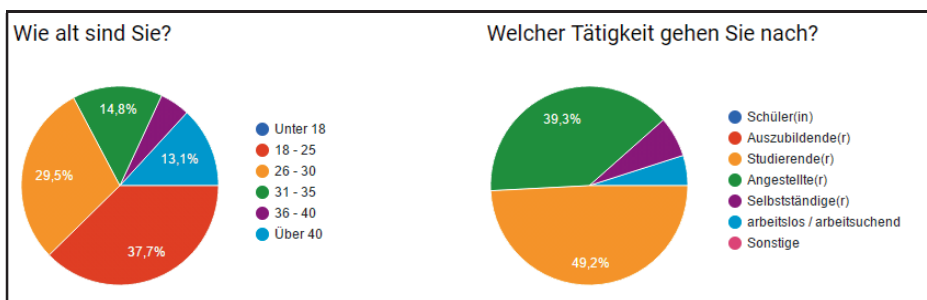


Abb. 1: Angaben der Probanden auf die Fragen nach ihrem Alter und der Tätigkeit (n = 61).

Die Teilnehmer können selbstverständlich weder von der Anzahl noch von der Verteilung als repräsentativ angesehen werden, was den explorativen Charakter unserer Studie erklärt. Es geht uns hierbei weniger darum, vorher definierte Hypothesen

repräsentativ zu bestätigen oder zu widerlegen, sondern eine erste empirische Basis für diese Fragestellung in diesem speziellen Kontext zu schaffen. Aus diesem Grund wurden neben geschlossenen auch offene Fragen gestellt. Der Rücklauf bei der qualitativen (offenen) Antwortmöglichkeit betrug knapp 56% (n = 34). Das ist hinsichtlich der überschaubaren Anzahl an Teilnehmern als relativ hohe Rücklaufquote zu bewerten (mehr als die Hälfte) und gewährleistet ein präziseres Clustering der Probanden.

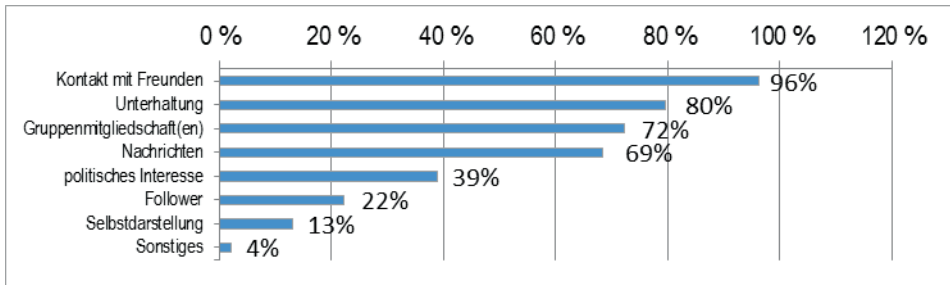


Abb. 2: Gründe für die Nutzung sozialer Medien. Mehrfachnennung möglich (n = 54).

Im Folgenden werden wir zuerst die wichtigsten Erkenntnisse der Auswertung der erhobenen Daten zusammenfassen. Anschließend folgt ein Clustering der Ergebnisse, um kausale Wirkungszusammenhänge zwischen den gegebenen Antworten zu eruieren.

3.1 Ergebnisse I: Nutzerverhalten und Umgang mit Missbrauch

Die Mehrheit der Probanden (89%) hat bereits Missbrauch in sozialen Medien (durch radikale Gruppen oder Einzelpersonen) persönlich erfahren, ob im direkten Umfeld oder über geteilte Inhalte. Auch wenn dies nicht immer im Kontext des Terrorismus erfolgt ist, erschließt sich hieraus, dass es sich bei radikalem oder extremistischem Missbrauch sozialer Medien kaum um eine Randerscheinung handeln kann. Dies ist bereits eine wichtige Erkenntnis im Hinblick auf eine potentiell positive Einstellung der Probanden zu verschärften Überwachungsmaßnahmen in sozialen Medien, denn persönliches Erleben und eigene Betroffenheit fördern im Allgemeinen die Bereitschaft zur Änderung des Status quo. Von den Probanden, die angegeben haben, schon einmal mit Extremismus oder Radikalismus in sozialen Netzwerken in Berührung gekommen zu sein (Abb. 3), geben rund 94% an, dies in Form von Kommentaren, Likes oder Shares erfahren zu haben, gefolgt von eindeutigen Bild- oder Videomaterial (69%), privaten oder öffentlichen Profildseiten (60%) sowie innerhalb von Gruppen in sozialen Medien (42%).

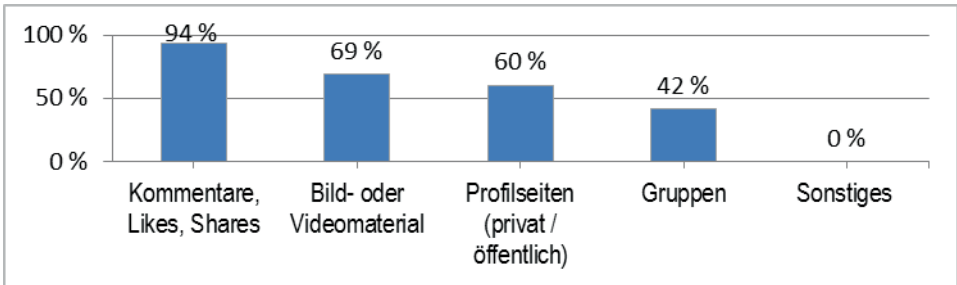


Abb. 3: Bereits erlebte Formen von Extremismus bzw. Radikalismus in sozialen Netzwerken. Mehrfachnennung möglich (n = 48).

Als interessant erweisen sich dabei auch die Antworten auf die Folgefrage nach den erlebten extremistischen bzw. radikalen Inhalten (Abb. 4). Die beiden Topantworten mit jeweils rund 94% bzw. 90% der Probanden sind erlebte Ausländerfeindlichkeit und Rassismus, gefolgt von Hetze gegen ethnische Minderheiten (56%) und gegen Homosexuelle (29%). Eine weitere wichtige Erkenntnis an dieser Stelle ist der erlebte Aufruf zur Radikalisierung in sozialen Medien durch Gruppierungen wie die Salafisten oder gar den IS von immerhin einem Viertel der Befragten (25%), die zuvor angegeben hatten, Hetze oder Extremismus erlebt zu haben. Dies ist als ein Indiz für eine aktive Präsenz solch radikaler Gruppierungen in sozialen Medien anzusehen.

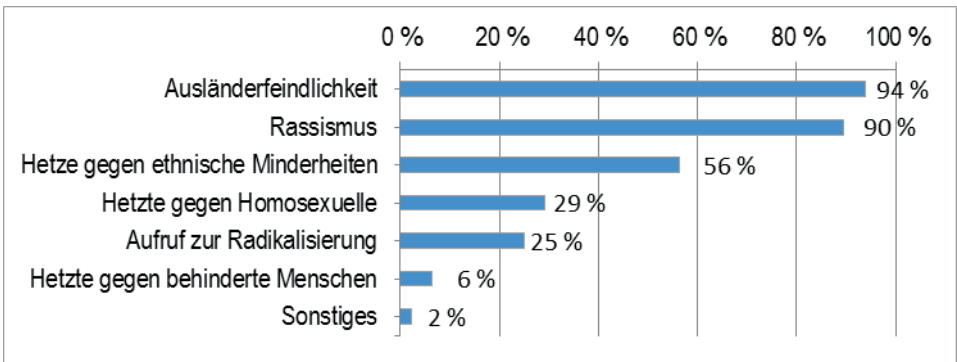


Abb. 4: Antworten auf die Frage nach bereits erlebten Inhalten von Extremismus bzw. Radikalismus in sozialen Netzwerken. Mehrfachnennung möglich (n = 48).

3.2 Ergebnisse II: Sicherheitsempfinden und Maßnahmen

Der nächste Themenblock befasste sich damit, wie die Probanden die aktuelle Sicherheitslage in Deutschland bewerteten. Wir stellten fest, dass sich trotz der Anschläge von Paris knapp 49% der Probanden in Deutschland derzeit sicher fühlen, wohingegen sich 28% gegenteilig äußern, sicherlich auch von der Berichterstattung beeinflusst.

Damit einhergehend sollten die Probanden weiter die aktuellen Sicherheitsmaßnahmen in Deutschland bewerten. Auffällig ist hier, dass keiner der Probanden mit der Ausprägung „5“ (optimal) geantwortet hat und es somit keinen einzigen Probanden gibt, der die Sicherheitsmaßnahmen als optimal bezeichnen würde. Rund 33% der Probanden beurteilten diese jedoch mit einer „4“ (gut), während weitere 30% diese als ausreichend ansehen würden. Fassen wir diese beiden Gruppen zusammen, erhalten wir knapp 62% der beteiligten Probanden, welche die Sicherheitsmaßnahmen als mindestens ausreichend betrachten würden. Etwa 32% und damit die zweithäufigste gewählte Antwort auf diese Frage stellt die Ausprägung „2“ (schlecht) dar, während weitere 7% die Sicherheitsmaßnahmen als sehr mangelhaft bezeichnen würden. Das Ergebnis dieser Frage ist demnach sehr ambivalent. Die beiden Topantworten stellen dabei zwei gegensätzliche Pole der Einschätzung von Sicherheitsmaßnahmen dar. Erst mit Zunahme jener Probanden, die die Sicherheitsmaßnahmen als ausreichend bezeichnen, kann festgestellt werden, dass der Großteil der Probanden (62%) die aktuellen Sicherheitsmaßnahmen als mindestens ausreichend bezeichnen würde.

Wir konnten ein Cluster darüber erstellen, welche Wünsche die Probanden bezüglich Sicherheitsmaßnahmen haben (Abb. 5). Dabei antworteten die Meisten (60%), sie wünschten sich verstärkte Polizeipräsenz, gefolgt von einer naheliegenden Schlussfolgerung, nämlich einer besseren Ausrüstung der Polizei (58%). An dieser Stelle folgt der Wunsch nach stärkerer Überwachung der Online-Kommunikation (48%).

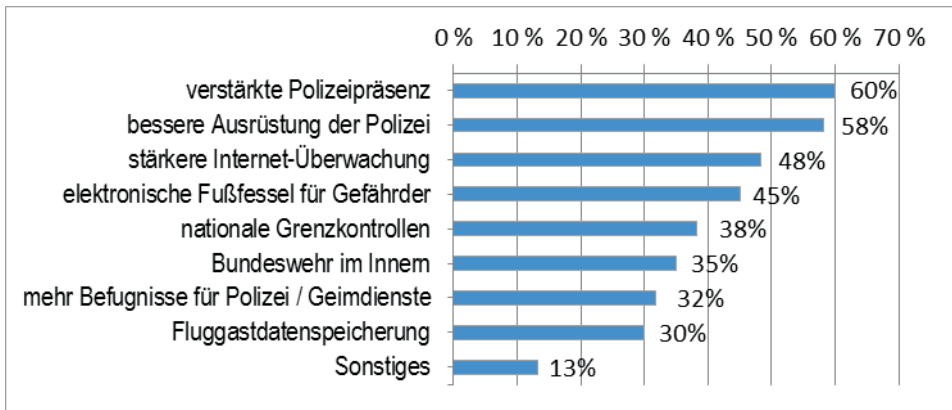


Abb. 5: Antworten der Probanden auf die Frage nach gewünschten Sicherheitsmaßnahmen gegen terroristische Aktivitäten und Bedrohungen. Mehrfachnennung möglich (n = 60).

Alleine die Betrachtung der am meisten genannten Antworten lässt darauf schließen, dass die Probanden weniger Maßnahmen im Netz als im weltlichen Raum fordern, was natürlich auch durch die Fragestellung nach radikalem Verhalten beeinflusst sein kann. Der Wunsch nach körperlicher Unversehrtheit durch erhöhte Polizeipräsenz ist dabei der Primus der genannten Antworten. Dies zeigt, dass sinnlich und körperlich erfahrbare Maßnahmen, in diesem Fall die verstärkte physische Präsenz der Exekutive, von der Bevölkerung eher begrüßt würde als beispielsweise die Überwachung von sozialen

Netzwerken, wobei es möglich ist, dass sich die Probanden digitale Maßnahmen schlechter vorstellen können, wie weltliche. Warum genau die Probanden sich durch Polizeipräsenz sicherer fühlen, kann durch die Umfrage nicht beantwortet werden. Hypothetisch könnte darauf geschlossen werden, dass die Polizei eine direktere und repressivere Maßnahme gegen Terrorismus darstellt als digitale Aufklärungsmaßnahmen, also das Sammeln und Auswerten von Daten. Dies würde sich auch mit dem Wunsch nach einer besseren Ausrüstung für die Einsatzkräfte decken.

Um jedoch weitere Meinungen und Wünsche bezüglich möglicher Sicherheitsmaßnahmen zu erfassen, die nicht zur Auswahl standen, wurde eine Option für freie Antwortmöglichkeiten gegeben. Generell können zwei Meinungscluster erstellt werden: (1) *Repressive Maßnahmen*: Der Großteil der Probanden wünscht sich repressive Maßnahmen wie verstärkte Polizeipräsenz oder deren verbesserte Ausrüstung. (2) *Präventive Maßnahmen*: Die Probanden wünschten sowohl stärkere Überwachung im Internet als auch elektronische Fußfesseln für sogenannte potentielle Gefährder. Bei diesen beiden Clustern sollte jedoch beachtet werden, dass die Grenzen zwischen repressiven und präventiven Maßnahmen fließend sind. So kann Polizeipräsenz in dem Sinne einen präventiven Charakter besitzen, indem sie mögliche Straftäter vor Gewalttaten abschreckt.

3.3 Ergebnisse III: Sicherheit vs. Privatsphäre

Auf die Frage, ob die Probanden bereit seien, auf Teile ihrer Privatsphäre in sozialen Netzwerken zu verzichten und im Gegenzug gewährleistet wäre, dass dieser Verzicht die Aufspürung von terroristischen Organisationen und ihren Anhängern begünstigt, antworteten die Probanden sehr unterschiedlich, jedoch mit leichter Tendenz. Rund 44% der Befragten wären bereit, zumindest Teile der Privatsphäre aufzugeben, während rund 36% dies verneint. Bisher muss also festgehalten werden, dass die Probanden geteilter Meinung bezüglich des Verlusts ihrer Privatsphäre im Netz zu Gunsten verbesserter Terrorbekämpfung sind. Erst mit der Ergänzung der folgenden, offenen Frage kann ein präziseres Meinungs-Cluster erstellt werden. Vor allem die Meinungen jener Probanden, die sich bei dieser Frage unschlüssig waren, also die eine „3“ auf der Skala angegeben haben, können damit näher untersucht werden (Abb. 6).

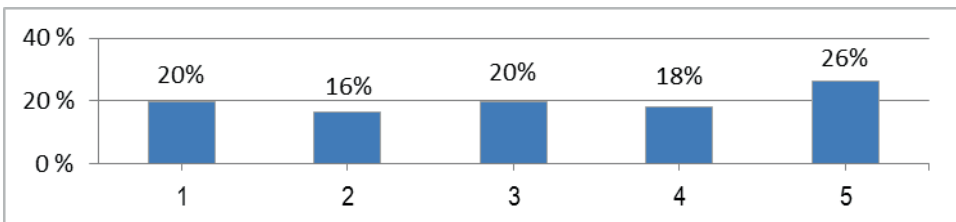


Abb. 6: Antworten auf die Frage nach der Bereitschaft zum Verzicht auf Privatsphäre zu Gunsten erhöhter Sicherheit (n = 61). Skala: 1 (stimme gar nicht zu) bis 5 (stimme vollkommen zu).

Nach der Analyse beider Fragen konnten in Bezug auf den *Verzicht auf Privatsphäre zugunsten von Sicherheit* drei Meinungscluster erstellt werden, welche im Folgenden erläutert werden sollen:

Starke Zuneigung (44%): Diese Gruppe, welche bei der Frage nach der Bereitschaft zum Verzicht auf Privatsphäre zu Gunsten erhöhter Sicherheit mit Ausprägung „4“ oder „5“ geantwortet haben, betrachten soziale Medien als öffentlichen Raum und argumentieren, dass jeder selbst entscheiden müsse, welche Informationen er in diesem öffentlichen Raum zulässt („[...] wer soziale Netzwerke nutzt, sollte sich im Klaren sein, was er dort postet“, P41). Sollte es der Aufspürung von Terroristen dienen, wären jene Probanden bereit, ihre Privatsphäre aufzugeben („Komplette Kommunikation über soziale Netzwerke sollte für Behörden einsehbar sein“; „Facebook und WhatsApp sollten vollkommen überwacht werden“), da sie diese in den sozialen Medien entweder als nicht gegeben betrachten oder bereit sind, einzelne Teilaspekte komplett offenzulegen („Gerne können FÜR SOLICHE ZWECKE sogar auch meine Nachrichten gelesen werden“). Dazu zählen Gruppenzugehörigkeit oder private Nachrichten.

Starke Abneigung (36%): Ein etwas geringerer Anteil steht dem konträr gegenüber und ist strikt gegen eine Überwachung sozialer Medien und nicht dazu bereit, Teile seiner Privatsphäre aufzugeben („Auf keinen Fall. Ich habe nichts mit Terrorismus zu tun“; „In keinem Bereich“). Sehr interessant ist dabei die Antwort eines Probanden, welche sich auch mit denen der restlichen „Gegner“ des Verzichts auf Privatsphäre deckt: „Privatsphäre gegen Sicherheit einzutauschen, erscheint mir generell nicht fair: Sicherheit kann nicht zu 100% gewährleistet werden. Darum wäre ich nicht bereit, auf Privatsphäre zu verzichten“ (P14). Dieses Stimmungsbild durchzieht dieses Cluster in seiner Gesamtheit und zeigt, dass jene Probanden nicht an eine vermehrte Sicherheit durch die Überwachung sozialer Medien glauben, gar deren Missbrauch fürchtet („Generell dient die Überwachung zum größten Teil der Marktforschung oder der Kontrolle der Bürger“), den Behörden nicht vertraut („Weg vom Überwachungsstaat!!!!“) und somit auch nicht bereit sind, Teile seiner oder ihrer Privatsphäre aufzugeben.

Kompromissbereitschaft (20%): Zu denjenigen, die auf der Skala mit einer „3“ geantwortet haben, kann durch die Hinzunahme der offenen Antwort ein Stimmungsbarometer hinzugefügt werden. Diese Personen würden ohne triftigen Grund seitens der Behörden nicht auf ihre Privatsphäre verzichten, würden aber im Gegenzug zur Verhinderung von Straftaten und dergleichen einen Kompromiss eingehen und „Hilfe anbieten“ („Sollte es dazu dienen, eine Auflösung/Aufklärung von beispielsweise Straftaten oder Verdachte jeglicher Art zu unterstützen, ließe ich mich allerdings auf Kompromisse ein und würde meine Unterstützung anbieten“; „Wenn dadurch ein Terroranschlag verhindert wird, in allen Bereichen“).

Nicht zu unterschätzen sind nach dieser Cluster-Bildung zwei Sachverhalte. Zunächst zeigt sich, dass das Cluster „Starke Zuneigung“ die größte Gruppe darstellt. Ebenso erkennt man anhand der offenen Antworten, dass der Verzicht der Privatsphäre in

sozialen Räumen kein Novum für diese Gruppe darstellt („[...] soziale Netzwerke sind ja quasi ein öffentlicher Raum“). Der zweite Sachverhalt betrifft die Synergie zwischen den Clustern „Starke Zuneigung“ und „Kompromissbereitschaft“. Letzteres Cluster ist ebenso bereit, Teile ihrer Privatsphäre zumindest partiell oder temporär aufzugeben („Öffentliche Beiträge/Profile, da diese ein breites Publikum betreffen und beeinflussen können. KEIN Mitlesen privater Nachrichten“). Betrachten wir beide Gruppen als Einheit, erkennen wir den Trend, dass der Verzicht der Privatsphäre tatsächlich eine Möglichkeit für den Großteil der Befragten, nämlich knapp 64%, darstellt.

4 Diskussion: Sicherheit und/oder Privatsphäre

Die anfänglichen Überlegungen hinsichtlich der Auswirkungen der NSA-Spähoffäre um Edward Snowden auf das Bedürfnis nach Privatsphäre in der Bevölkerung scheinen im starken Kontrast zu stehen zum gleichzeitigen Bedürfnis nach Sicherheit – gerade nach den terroristischen Anschlägen von Paris und Brüssel. Bei Betrachtung der Erkenntnisse aus unserer Studie in Bezug auf repressive und präventive Maßnahmen sowie die Meinungscluster starke Zuneigung, starke Abneigung und Kompromissbereitschaft gegenüber dem Verzicht auf Privatsphäre zu Gunsten erhöhter Sicherheit lässt sich zunächst festhalten, dass es scheinbar keine allgemeingültigen Lösungsansätze geben kann, die alle zufriedenstellen. Dies deckt sich auch mit den nicht konsistenten Forschungsergebnissen zu jenem Thema [DS04, FR15].

Greifbar an dieser Stelle ist jedoch eine Tendenz zur Zuneigung respektive Kompromissbereitschaft der Probanden hinsichtlich des (zumindest teilweisen) Verzichts auf Privatsphäre für ein verstärktes Sicherheitsempfinden. Immerhin ist das Gros der Teilnehmer (64%) dieser Auffassung. Somit ergeben sich hier zumindest Chancen für den Staat bzw. Behörden und Organisationen mit Sicherheitsaufgaben (BOS), neue respektive, bislang verwehrte Antiterrormaßnahmen zu etablieren.

An dieser Stelle möchten wir auf eine Studie der R+V Versicherung [RV15] zurückgreifen, die sich als Langzeitstudie seit Mitte der 1990er mit den Ängsten der Deutschen befasst. Die Erkenntnisse dieser Studie korrelieren mit unseren Umfrageergebnissen. Insbesondere seit den Anschlägen vom 11. September 2001 in den USA ist ein signifikanter Anstieg der Furcht vor terroristischen Anschlägen zu verzeichnen. Seither hat sich die Sorge davor in der deutschen Bevölkerung verdoppelt. Zwar sind Schwankungen bei der Angabe von Terrorangst in den letzten 15 Jahren zu beobachten, doch mit durchschnittlich 46% in den Jahren ab 2001 liegt das „Angstniveau“ der Deutschen fast doppelt so hoch wie noch in den 1990er-Jahren. Hier wäre darüber hinaus die Frage zu stellen, welche Einflussfaktoren (neben den Attentaten selbst) dazu geführt haben. Zu den Variablen zählen unter anderem die Medienberichterstattung durch die Presse (insbesondere die Art und Weise der Darstellung) und die damit verbundene Wahrnehmung durch die Rezipienten, die zeitlichen Abstände von Anschlägen (Aktualität) oder die direkte bzw. indirekte

Betroffenheit. Diese Studie führt die Angst der Deutschen vor Terrorismus auf dem zweiten Platz (52%), gefolgt von der Sorge vor Naturkatastrophen (53%).

Wie bereits Gethmann [Ge96] festgestellt hat, ist die Angst jedoch kein guter Ratgeber. Dies ist insbesondere für Behörden und Politik von großer Bedeutung: Kurzfristige, übereilte Reaktionen auf terroristische Bedrohungen sind daher gewissermaßen als ambivalent zu betrachten. Einerseits können schnelle, zeitnahe Antiterrormaßnahmen der Bevölkerung ein Gefühl von Sicherheit im Sinne von Handlungsfähigkeit und Entschlossenheit der Regierung vermitteln. Andererseits – und dies ist ein nicht zu unterschätzendes Risiko – können übereilte Maßnahmen als Kurzschlussreaktion auf Bedrohungen oder Attentate (un-)absehbare, negative Auswirkungen auf die Freiheit und Privatsphäre der Bevölkerung haben. Eine extreme Ausprägung dieser Art wäre beispielsweise die Charakteristik eines Polizeistaats, zu der die totale Überwachung zählt und somit der Freiheit und Privatsphäre entgegenwirkt. Auch der nicht selten geäußerte Wunsch unserer Probanden nach dem Einsatz der Bundeswehr im Innern (Abb. 5) kann als intuitiver „Schnellschuss“ interpretiert werden.

Aus diesem Grund empfiehlt sich stets eine mittel- bis langfristige Evaluierung einer Bedrohung und entsprechender Gegenmaßnahmen sowie deren Konsequenzen für die Bevölkerung, bevor endgültige Entscheidungen gefällt werden. In Anlehnung an das *Time-lag*-Prinzip aus den Wirtschaftswissenschaften oder die *Cultural Lag*-Theorie nach Ogburn [Og57] ist es im übertragenen Sinne empfehlenswert, stets einen zeitlichen Verzögerungseffekt bei Entscheidungsfindungen zu beachten. Denn das Abwarten der Effekte terroristischer Anschläge kann von Vorteil sein, bevor politische Entscheidungen im Hinblick auf den Einsatz der genannten Methoden gefällt werden. Hierdurch kann gewährleistet werden, dass keine voreiligen Entscheidungen zu ungewollten Konsequenzen führen wie z. B. ein nicht-intendierter Freiheitsverlust seitens der Bevölkerung.

5 Fazit: Keine klare Tendenz, aber Kompromissbereitschaft

Privatsphäre und Sicherheit stehen in einem Spannungsfeld. Der 2013 ausgelöste NSA-Skandal durch Edward Snowden und der daraus resultierende „Privatsphäre-Boom“ scheinen noch aktuell zu sein. Viele der Probanden unserer Studie, welche dem zweiten Cluster „Starke Abneigung“ zugeordnet werden konnten, argumentieren damit, dass eine Überwachung aufgrund nie erreichter absoluter Sicherheit nur der Spionage der eigenen Bevölkerung aufgrund von Marktforschung bzw. Kontrolle diene. Auf der anderen Seite, dem ersten Cluster „Starke Zuneigung“, wünschen sich gerade diese Probanden eine Überwachung des Internets und betrachten dieses als öffentlichen Raum. Diese beiden Cluster, da ungefähr gleich groß in ihrer Anzahl an Probanden, zeigen die zwei Extreme, welches dieses Thema hervorruft. Interessant ist dabei das dritte Cluster derjenigen, die eine „Kompromissbereitschaft“ zeigen. In diesem Fall zeigt sich, dass jene Probanden zwar auf ihre Privatsphäre achten, in dringenden Verdachtsfällen aber zumindest auf

Teile ihrer Privatsphäre verzichten würden.

Ein Trend (Privatsphäre oder Sicherheit) kann nach unserer Umfrage nicht ausgemacht werden, wenn nur die einzelnen Cluster betrachtet werden. Zwar besteht die Kompromissbereitschaft zur Aufgabe der Privatsphäre, jedoch nur im Gegenzug des Versprechens einer Prävention oder Aufspürung der Verursacher terroristischer Attacken. In diesem Fall bleibt festzuhalten, dass der Wunsch nach Privatsphäre immer noch gegeben ist, in unserer Studie jedoch auffiel, dass sich das Bewusstsein gegenüber sozialen Netzwerken als öffentlicher Raum geändert hat und eine generelle Kompromissbereitschaft besteht, Teile der Privatsphäre in sozialen Netzwerken aufzugeben. Erst wenn wir aufgrund ihrer inhaltlichen Übereinstimmung die Cluster „Starke Zuneigung“ und „Kompromissbereitschaft“ als eine Einheit betrachten, kann jener Trend, soziale Netzwerke als öffentliche Räume mit eingeschränkter Privatsphäre zu begreifen, die bei Bedarf wie in der Realität, durch exekutive Organe interveniert werden können, aufgezeigt werden.

Die Kompromissbereitschaft bzw. die starke Zuneigung zu einer Überwachung von Teilen der sozialen Netzwerke kann aus einer anderen Feststellung erfolgen, nämlich der, dass etwa 89% aller Probanden bereits Auftritte radikaler Vereinigungen bzw. Zeuge von radikalen und extremistischen Äußerungen in sozialen Netzwerken wurden. Zieht man noch hinzu, dass in den Nachrichten durch den Vormarsch des IS, das Erstarken rechtspopulistischer Parteien und der Flüchtlingsdebatte, die Diskussion über den Umgang mit diesen Themen im Vordergrund steht, so kann man davon ausgehen, dass einige der Probanden, die zu den Clustern „Starke Zuneigung“ und „Kompromissbereitschaft“ gehören, mit einer anderen Sichtweise auf diese Parteien, Gruppierungen und extremistisches Gedankengut reagieren bzw. dieses öfter oder in einer kritischeren Form in den sozialen Medien erlebt haben. Um diese Querverweise jedoch fundiert zu eruieren, benötigt es einer weiteren Studie, die sich genauer mit der Erfahrung extremistischen Gedankenguts in sozialen Medien beschäftigt. Darüber hinaus ist eine Untersuchung notwendig, welche soziokulturellen und -ökonomischen Merkmale die Probanden der einzelnen Cluster aufweisen.

Als Fazit kann und muss nach dieser Studie jedoch weiterhin davon ausgegangen werden, dass Sicherheitsmaßnahmen auf Kosten von Privatsphäre im öffentlichen Raum zwiespältig betrachtet werden. Die Enthüllungen durch Snowden haben möglicherweise immer noch Einfluss auf die Sensibilisierung der Privatsphäre in sozialen Netzwerken, die Anschläge des IS hingegen zeigen, dass Kompromissbereitschaft existiert, ihre Privatsphäre in diesen Medien einzuschränken. Eine generelle Tendenz hin zur bedingungslosen Bereitschaft, die Privatsphäre zugunsten von Sicherheit aufzugeben, konnte nicht gezeigt werden.

Danksagung: Die Forschungsarbeiten wurden im Rahmen des BMBF-Projekts „KOKOS“ (Fö.-Kz. 13N13559) sowie im Rahmen des EU-FP7-Projekts „EmerGent“ (Fö.-Kz. 608352) gefördert. Die Fragestellung und kompakte Eindrücke basieren auf [Re16].

Literaturverzeichnis

- [BR15] Bartlett, Jamie; Reynolds, Louis: State of the art 2015: a literature review of social media intelligence capabilities for counter-terrorism, Demos (2015)
- [Ch15] Christoph, Stefan: Funktionslogik terroristischer Propaganda im bewegten Bild. In: Journal for Deradicalization Bd. Fall/15 (2015), Nr. 4, S. 145–205
- [DS04] Davis, Darren W.; Silver, Brian D.: Civil Liberties vs. Security: Public Opinion in the Context of the Terrorist Attacks on America. In: American Journal of Political Science Bd. 48 (2004), Nr. 1, S. 28–46
- [De02] Denninger, E: Freiheit durch Sicherheit? Anmerkungen zum Terrorismusbekämpfungsgesetz. In: Aus Politik und Zeitgeschichte Bd. 10/11 (2002), S. 22–30
- [Fr15] Friedewald, Michael; Lieshout, Marc van; Rung, Sven; Ooms, Merel; Ypma, Jelmer: Privacy and Security Perceptions of European Citizens: A Test of the Trade-off Model. In: IFIP Advances in Information and Communication Technology Bd. 457 (2015), S. 54–70
- [Ge96] Gethmann, Carl Friedrich: Ist die Angst ein schlechter Ratgeber? In: 5. Essener Forum für psychosoziale Versorgung, 1996
- [Gr09] Grimm, A.; Hulse, L.; Schmidt, S.: Risikowahrnehmung und psychologische Reaktionen in öffentlichen Krisensituationen am Beispiel von Terrorattentaten. In: Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz Bd. 52 (2009), Nr. 12, S. 1129–1140
- [Ja07] Jakob, Nikolaus: Die Diffusion von Terrormeldungen, die Wirkung von Anschlägen auf die öffentliche Meinung und die Folgen für das Vertrauen in der Demokratie. In: Glaab, S. (Hrsg.): Medien und Terrorismus - auf den Spuren einer symbiotischen Beziehung.. Band 3. Aufl. Berlin : Berliner Wissenschafts-Verlag, 2007 — ISBN 978-3-8305-1435-0, S. 155–174
- [Je15] Jeberson, W; Sharma, Lucky: Survey on counter Web Terrorism. In: COMPUSOFT, An international journal of advanced computer technology Bd. 4 (2015), Nr. 5, S. 1744–1747
- [Ko16] Koshan, Mansoor: Vorratsdatenspeicherung – verfassungsrechtliche Rahmenbedingungen und rechtspolitische Verortung. In: Datenschutz und Datensicherheit - DuD Bd. 40 (2016), Nr. 3, S. 167–171
- [Ku16] Kurz, Constanze: Erklärung zur Vorratsdatenspeicherung: „Eingriff in die Privatsphäre von Millionen Menschen“. . — netzpolitik.org
- [Le04] Lepsius, Oliver: Freiheit, Sicherheit und Terror: Die Rechtslage in Deutschland. In: Leviathan Bd. 32 (2004), Nr. 1
- [Lu16] Ludwig, Thomas; Reuter, Christian; Pipek, Volkmar: From Publics to Communities: Tracing the Path of Shared Issues through ICT. In: Computer Supported Cooperative Work: The Journal of Collaborative Computing (JCSCW) Bd. 25 (2016), Nr. 2-3, S. 193–225

- [MD03] Metzger, Miriam J.; Docter, Sharon: Public Opinion and Policy Initiatives for Online Privacy Protection. In: *Journal of Broadcasting & Electronic Media* Bd. 47 (2003), Nr. 3, S. 350–374
- [Og57] Ogburn, William F.: Cultural lag as theory. In: *Sociology and Social Research* Bd. 41 (1957), Nr. 3, S. 167–174
- [Re16] Reuter, Christian; Gellert, Robin; Geilen, Gordian: Reception of Terror in Germany – Security, Privacy and Social Media. In: *Adjunct Proceedings of 30th EnviroInfo Conference*. Berlin, 2016
- [Re13] Reuter, Christian; Heger, Oliver; Pipek, Volkmar: Combining Real and Virtual Volunteers through Social Media. In: Comes, T.; Fiedrich, F.; Fortier, S.; Geldermann, J.; Müller, T. (Hrsg.): *Proceedings of the Information Systems for Crisis Response and Management (ISCRAM)*. Baden-Baden, Germany, 2013, S. 780–790
- [Re11] Reuter, Christian; Pohl, Patrik; Pipek, Volkmar: Umgang mit Terminologien in interorganisationaler Krisenkooperation - eine explorative Empirie. In: Eibl, M. (Hrsg.): *Mensch & Computer 2011: Übermedien Übermorgen*. München, Germany: Oldenbourg-Verlag, 2011.
- [Rv15] R+V Versicherung: R+V Studie – Die Ängste der Deutschen. URL <https://www.ruv.de/presse/aengste-der-deutschen>
- [SJ11] Seib, Philip; Janbek, Dana M.: *Global Terrorism and New Media: The post-Al Qaeda Generation*. New York : Routledge, 2011
- [Su08] Sutton, Jeannette; Palen, Leysia; Shklovski, Irina: Backchannels on the Front Lines: Emergent Uses of Social Media in the 2007 Southern California Wildfires. In: Friedrich, F.; Van de Walle, B. (Hrsg.): *Proceedings of the Information Systems for Crisis Response and Management (ISCRAM)*. Washington D.C., USA, 2008.
- [Tu15] Turk, Austin T.: Terrorism and Counterterrorism. In: Goode, E. (Hrsg.): *The Handbook of Deviance*. Hoboken, NJ : John Wiley & Sons, Inc, 2015, S. 537–548
- [We02] Werthes, Sascha; Conrad, C; Kim, R: Die Terrorkrise als Medienereignis? Internationale Krisenkommunikation - Eine Herausforderung im 21. Jahrhundert. In: Schicha, C.; Brosda, C. (Hrsg.): *Medien und Terrorismus*. Münster : Lit-Verlag, 2002 — ISBN 3-8258-5923-1
- [We07] Weichhart, Peter: Risiko - Vorschläge zum umgang mit einem schillernden Begriff. In: *Berichte zur Deutschen Landeskunde* Bd. 81 (2007), Nr. 3, S. 201–214
- [Wo12] Wollebæk, Dag; Enjolras, Bernard; Steen-Johnsen, Kari; Ødegård, Guro: After Utøya: How a High-Trust Society Reacts to Terror—Trust and Civic Engagement in the Aftermath of July 22. In: *PS: Political Science & Politics* Bd. 45 (2012), Nr. 01, S. 32–37

An Open Service Platform for Multi-Hazard in Action - the PHAROS Pilot Demonstration

B. Barth, M.A. Marchitti, J. Mulero Chaves¹, U. Raape, C. Strobl², M. Borràs, O. Vilalta, H. Ballart, J. Vendrell, N. Prat³, M. Mendes⁴, T. Ladoire⁵, G. Gardikis, S. Pantazis, S. Costicoglou⁶, K. Jäckel⁷, W. van Setten⁸

Abstract: Natural and man-made disaster, whose impact has been increasing on the last years, present a series of common dynamics which can be exploited for a more efficient emergency management. Current disaster management systems focus on single types of emergencies and cannot be flexibly adapted to the respective situation. The presented approach developed during the PHAROS project describes an integrated service platform which offers tools for multi-hazard emergency management using a modular and flexible structure. The assets provided by PHAROS include: (i) data collecting and processing, (ii) communication and alerting means, (iii) risk and emergency detection and situation assessment, (iv) decision support. In addition to the system scope and description, the paper presents the PHAROS pilot demonstration used to validate the system. For this evaluation the system was developed for a forest fire scenario and tested in Solsona, Spain in March 2016 during a real operational prescribed burn.

Keywords: PHAROS, public alerting, forest fire, IT supported crisis management, risk assessment, earth observation, decision support

1 Introduction

The Earth trembles, rivers burst their banks, tsunamis destroy coastal regions. There is a higher concentration and impact of natural and man-made disasters, like floods, forest fires, nuclear accidents and earthquakes. Emergencies and disasters generally present some similarities with regard to their basic dynamics but at the same time pose particularities which must be tackled differently. Current disaster and emergency management tools are generally hazard-specific and cannot be easily adapted neither to support management

¹ DLR, Institut für Kommunikation und Navigation, Postfach 1116, 82230 Weßling, {Benjamin.Barth ; Maria-Antonietta.Marchitti ; Javier.MuleroChaves}@dlr.de

² DLR, Deutsche Fernerkundungsdatenzentrum, Postfach 1116, 82230 Weßling, {Ulrich.Raape@dlr.de ; Christian.Strobl}@dlr.de

³ Pau Costa Foundation, R&D Department, c/ Mossèn Cinto Verdaguer 42, baixos, 2A, 08552 Taradell , {mariona ; oriol ; helena ; jvendrell ; nprat}@paucofoundation.org

⁴ Tecnosylva, Engineering Department, Parque Tecnológico de León , 24009 León, Mendes, mmendes@tecnosylva.com

⁵ Avanti Communications PLC, Consulting Department, 20 Black Friars Lane, EC4V6EB London, Thomas.Ladoire@avantiplc.com

⁶ Space Hellas, R&D Department, Messogion ave., 15341 Athens, {ggar ; span ; scostic}@space.gr

⁷ IQ wireless GmbH, R&D Department, Carl-Scheele-Str. 14, D-12489 Berlin, klaus.jaekel@iq-wireless.com

⁸ Stichting Platform Mobile Messaging (SPMM) research, Crown Business Center, Planetenweg 5, 2132 HN Hoofddorp, wim.van.setten@spmm.org

of other types of emergencies nor to efficiently handle cascade events and inter-related crisis. Therefore, possible existing synergies between available tools are difficult to be exploited. Furthermore, the potential of space based technologies, such as satellite-based Earth observation (EO), satellite communications and satellite navigation, which are robust against disaster events, is not fully exploited for emergency management purposes. The PHAROS approach is to provide an integrated platform which offers a wide range of IT-based tools for multi-hazard emergency management during the complete emergency management cycle [Mu14a]. The system has a modular structure which is flexible and scalable, i.e. it allows easy tailoring to address different hazards (e.g. forest fires or floods) and needs. PHAROS assets include: (i) tools for data collection, processing and distribution, (ii) communication technologies with emphasis on public alerting, (iii) risk and emergency detection and forecast of the evolution and of the disaster itself, finally, (iv) decision support services.

Regarding the development of information and knowledge management systems in the context of disaster risk reduction, a list of issues which underline the need for a systematic approach have been identified, according to [Th16]. Among these issues, the following can be highlighted:

- Information is scattered among various agencies and institutions with limited coherence, coordination and sharing;
- Information about hazard events, exposure, vulnerability, and the impacts of disasters is often not systematically collected;
- Risk information is not systematically used for policy and decision making;
- There is little integration of knowledge systems at national, regional and community levels.

Therefore, there seems to be a need for tools that allow the gathering and integration of different information sources in a systematic way, making it possible to use the collected information for decision support purposes. Additionally, these tools shall enable information sharing among different jurisdictional and organisational levels, identifying the relevant information items and providing data formats supporting efficient data sharing. The modular PHAROS system approach is able to provide flexibility and scalability. On one hand, with respect to the functionalities that it is able to provide, thus allowing to add and remove system elements according to the hazard and scope to be addressed. On the other hand, scalability allows adapting the system deployment to the different organisational structures, at local, regional, national and international level.

Although keeping in mind a multi-hazard system approach, the design and implementation phases took into account forest fires as exemplary scenario. The system was validated and demonstrated in Solsona, Spain in March 2016 during a real operational situation in the context of a prescribed burn. The demonstration was done in cooperation with the Catalan Fire Brigades and the DLR VABENE++ project. Aerial real time image acquisition (optical and thermal) by means of a helicopter was included in the demonstration and the gathered data was presented to users for situation assessment.

This paper is organized as follows: section 2 describes the approach and applications of the system. Section 3 shows the structure of the system and the individual modules. Section 4 describes the pilot demonstration using a real forest fire and in section 5 the user evaluation is presented. Finally, section 6 concludes the paper.

2 The PHAROS Approach

The present document introduces the final version of the PHAROS system architecture with focus on the forest fire hazard. The description includes the identification of the different modules (sub-systems) that build the system, explaining their interaction with other modules and describing their functionalities. The developed system is in pre-operational state which integrates a number of modules, namely data systems (such as space-based and aerial EO data and terrestrial in-situ sensor data) together with tools for processing this data (such as simulators, alerting and decision support tools) as illustrated in figure 1. Each of these data systems and tools can be considered as an independent module to be integrated within the overall system, providing different features. The modules can be interconnected to the overall system by means of a service platform (SP), which acts as a mediator between them. It must be highlighted, for the sake of clarity, that even if the overall PHAROS system is described as a platform, the term service platform will be used to refer to the system mediating core module.

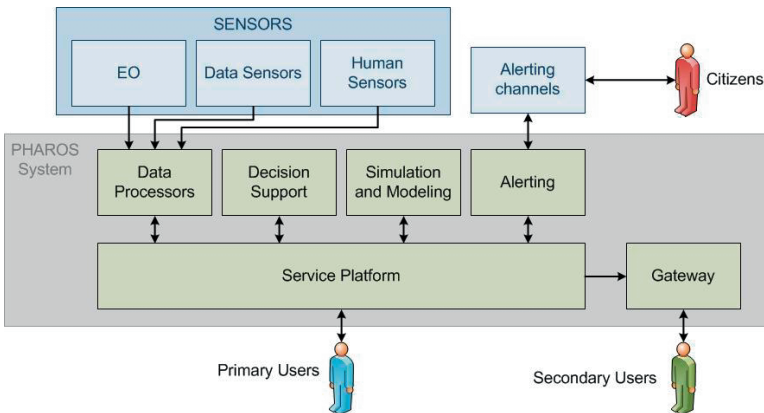


Fig. 1: PHAROS Functional Architecture

According to their purpose when using the system, users have been classified in three different categories: (i) primary users which are any responsible authority which uses the system to directly confront a particular situation and/or incident, (ii) secondary users, i.e. any third-party entity, for instance, research institutions or companies that exploit a service of the system and (iii) recipients like citizens that are provided with information that is sent/made available through the PHAROS system. Tackling these users, the PHAROS system combines the four applications described in the following.

Risk and threat assessment which is considered a key feature of the PHAROS platform. Primary users are interested in obtaining information that allows them to assess the ex-

isting risk or threat for different areas and hazards, taking into account the characteristics of the area, the current situation and the existence of critical infrastructures. Critical infrastructures must be understood as the assets, systems and networks, whether physical or virtual, so vital that their incapacitation or destruction would have a debilitating effect on security, national economic security, national public health or safety, or any combination thereof [Of16]. In risk assessment scenarios, not only the real but also the potential damages to the existing infrastructures play an important role together with the possibility that an event is initiated by the infrastructure itself. Together with the mentioned cases, users may interact with the system in order to get predictions of the evolution of the crisis, either during real crisis or for training, mitigation or preparedness purposes.

Furthermore, the system can be used for risk and disaster detection and monitoring scenarios. Automatic detection is important not only to timely detect hazards, which is a topic where already existing tools can offer satisfying solutions, but also to timely detect a high risk of any hazard likely to happen. This involves the use of different data inputs, such as EO data and in-situ sensor data, to be monitored together with the establishment of rules that will define the behaviour of the system for the different cases.

Another application area covered by the system is disaster management. This usually defines the interaction between the system and the different actors which need proposals on the suitable strategies to be carried out (for instance, sending an alert message, or fire fighting strategies), based on the input coming from the sensors and the data processing carried out by the system. Apart from primary users in the respective control centres, disaster management takes into account the involvement of first responders on the field, since many decisions are actually taken by them.

PHAROS is a composite system bringing together multiple and diverse actors, thus the efficient support of communications among the actors is also considered as a crucial aspect. The focus of communication is given to dissemination of alerting messages to alert recipients using a wide variety of communication technologies (alerting channels). Alerts are sent, for instance, to dedicated mobile applications for direct presentation in the users mobile terminals, data being transmitted over cell broadcast. Nevertheless, communication is not restricted to alerting; it also includes interconnection of the different elements of the system, as well as allowing sharing of data gathered and processed by PHAROS (e.g. simulation results, EO data) to third-party platforms in order to allow interoperability between different authorities in different areas or with different jurisdictional responsibilities.

3 System Description

The implemented PHAROS system architecture is depicted in figure 2. In the following, a detailed description of the implementation of each module focusing on the forest fire case is given. The arrows in the picture indicate the direction of the communication. Arrows in both directions indicate a bidirectional connection, for instance the simulation services and the service platform communicate with each other in a bidirectional way: the simulation services provide simulation results to the service platform and the last provides data that is used by the simulation services for the simulation.

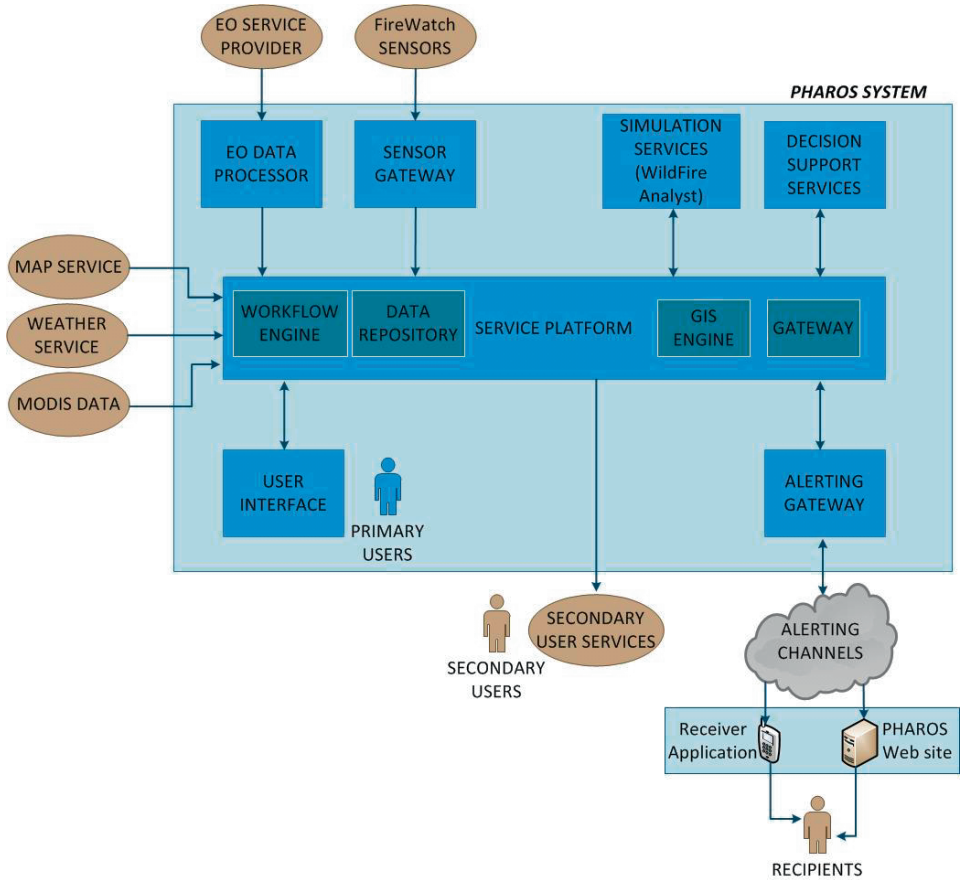


Fig. 2: PHAROS System Architecture

3.1 Service Platform

The Service Platform (SP) is the central component of the system. Its purpose is to interconnect the different modules and orchestrate the communication workflows between them. Its primary role is to act as a data mediator (information hub) between the different modules, so that they can exchange information in a transparent way, using common open APIs and interfaces. Furthermore, the SP is in charge of orchestrating the PHAROS system by implementing the rules established in the work-flow engine and invoking the appropriate modules according to pre-defined sequences. The SP exposes the services of the system via a graphical user interface (GUI) tailored for the primary users as well as via a gateway devised for secondary users to access the PHAROS services.

Most of the interfaces exposed by the SP are based on Web technologies and open standards, so that all modules can publish and acquire information in a uniform and coherent manner. The SP offers a data repository service for sensor and raster/vector geo-referenced

data, where raw sensor information as well as processed data and simulation/assessment/decision support data can be stored and acquired on-demand. A common information bus based on an enterprise service bus (ESB) architecture enables modules to asynchronously push and pull information also supporting publish/subscribe-based communication and allows the interconnection of a virtually unlimited number of heterogeneous sensors and processing modules, thus promoting the scalability and expandability of the system.

Finally, the SP provides connectivity with third-party services. It implements dedicated interfaces to external service providers as plug-ins, conforming to the providers API, in order to acquire information from sources external to the PHAROS platform. So far the SP interfaces two particular external service providers: a map and a weather service provider. Regarding the provision of services, it exposes an interface so that PHAROS-generated information can be consumed by third-party services, operated by secondary users.

3.2 User Interface

The user interface gives primary users access to the various functionalities of the system. It integrates together all the other sub-systems and offers multiple services to the users. It is also used by administrators to configure the system. It offers a login interface to identify users and their access rights. The interface intuitively makes sure that the user can find functionalities easily and quickly.

3.3 Earth Observation Systems and Processors

The EO data processor gives the possibility to use EO data (satellite and airborne based) and its derived products in the SP, making it available to other PHAROS components such as the simulation services and the decision support system). The SP and the EO Data Processor communicate using INSPIRE/OGC compliant services [Mu14b].

The primary use in the context of forest fire is the provision of an operational fire hot spot service, which is capable of detecting thermal anomalies in near-real time. For this purpose two independent EO platforms are used, the MODIS sensor from NASA [MO16] and the MSG Seviri Sensor [Me16] from EUMETSAT. Both of them are acting as a triggering input for the simulation services and the decision support services. The MODIS and MSG Seviri hotspot data does not need to be processed in the PHAROS system. Therefore these data can be considered as an additional data input, which is directly provided to the corresponding service through the SP.

3.4 Sensor System and Gateway

Sensor systems provide sensor data to be used for risk and emergency monitoring and management. In the particular forest fire case, the FireWatch system [IW16] is used as

sensor network. The FireWatch in-situ sensor system performs a permanent and automatic surveillance of a predefined area regarding smoke as a first indicator of an incipient wild-fire. In case that a smoke cloud is recognised, a detection message is generated, including further detection details (e.g. fire location, images) and transferred to the sensor gateway. The sensor gateway combines various sensor networks. Here, all messages are further processed and verified by human personnel. The sensor gateway controls the sensor network elements and provides the interface to the SP.

3.5 Simulation Services

The aim of the simulation services is twofold: (i) on the one hand, it enables the evaluation of the potential development of a detected crisis situation and on the other hand (ii), it enables to assess the risk of that situation related to the existence of critical areas or infrastructures. Concerning the first case, the simulation of the behaviour of the detected hazard can be manually triggered by the user or automatically by the operation of the decision support service. At the same time, the outcome of the simulation services can be used by the decision support service as described in the following paragraph. The simulation module is based on the WildFire Analyst capabilities [Te16]. This module provides, among others, the assessment of the forest fire potential based on critical infrastructures, the visualisation of fire evolution in an area and an estimation of the necessary evacuation time.

3.6 Decision Support Services

The Decision Support Services aggregate all available information about a situation (e.g. EO data, weather and mapping data provided by the external suppliers, data coming from the FireWatch system, the simulator outputs and any data manually introduced by the users) in order to provide tools for improving situational awareness by using a common operational picture (COP). A decision support system (DSS) has been developed which provides the user with notifications about important situation updates and alerting proposals based on the knowledge stored in the system by means of pre-defined alerting plans. The DSS can be extended to apply additional domain knowledge (work-flows, rules) provided by end users and domain experts to also cover additional proposal types and hazard types.

3.7 Alerting Services

Following a multi alerting channel approach for alerting services an alerting gateway (AG) has been developed. With the multi-channel approach the number of possible receivers is increased and with this the number of people reached, respectively. The objectives of the AG are: (i) to compose alert/information messages, either by combining options selected by the user through the GUI or by using the results of the operation of the DSS and (ii)

to transmit the corresponding messages through selected available alerting channels. For transmitting the messages, the AG can use either the common alerting protocol (CAP) to ensure compatibility with already existing alerting systems as well as a new protocol for efficient transmission over narrow-band channels. An RSS-feed presented on a website and cell broadcast have been selected as alerting channels to evaluate the system. A receiver application for Android mobile phones has been implemented for the reception and presentation of the messages.

4 System Demonstration - Forest Fire in Solsona

The PHAROS system was validated and demonstrated in Solsona, Spain, March 2-4, 2016 in cooperation with the Catalan fire-brigades. Its main goal was twofold: (i) evaluating performance and viability of the integrated set of tools, services and products developed during the project, and (ii) testing, validating and gathering feedback from key stakeholder groups and end users (primary users) of the PHAROS platform in its pre-operational status. To achieve those objectives, the pilot demonstration was based on the forest fire scenario. A prescribed burn, a technique commonly used in forest management in order to minimise risks of fire in a region, was carried out during the demonstration. It was suitable to simulate a low-medium intensity forest fire [Co11],[RST13]. Some exercises to be carried out by the end users were prepared. In order to collect feedback a questionnaire was distributed among the end users and evaluated. The feedback gathered during the pilot demonstration revealed how the end users held a very positive opinion of the platform and the different components that they were able to test during the demo. For each of the components, recommendations and suggestions to improve the platform in possible future developments were collected.

The following different types of exercises were conducted: (i) Exercises using real-time sensor data and (ii) non-real-time exercises. When it was not possible to gather data from the field, manual input was used. Apart from real time data coming from the different sensors, the prepared setup allowed to manually configure the system parameters to see their effects during emergency management (e.g. higher wind speed in order to achieve larger fire perimeters).

4.1 Burning area and location

In figure 3 the selected area of the prescribed burn is shown. It includes virtual objects for the DSS which are described later on in the paper. The selection of the location depended essentially on three factors: weather conditions, type of vegetation and extension of the plot. However, there were additional objectives of the prescribed burn besides the PHAROS pilot which affected the area selection. These basically were forest fire prevention and forestry management [RST13],[ALM13],[Fe10]. The fire was used to reduce the fuel load, in particular the undergrowth vegetation was burned in order to reduce its amount for the fire season. The plot had a regular structure of *Pinus nigra* with undergrowth

of *Buxus sempervivens* and *Juniperus communis*. Dominant *Pinus nigra* height is between 7 to 10 meters. The tree density was between 800 and 2700 trees/ha with complementary species such as *Quercus ilex* 200 to 500 tree/ha and regenerated of holm oak and pine.

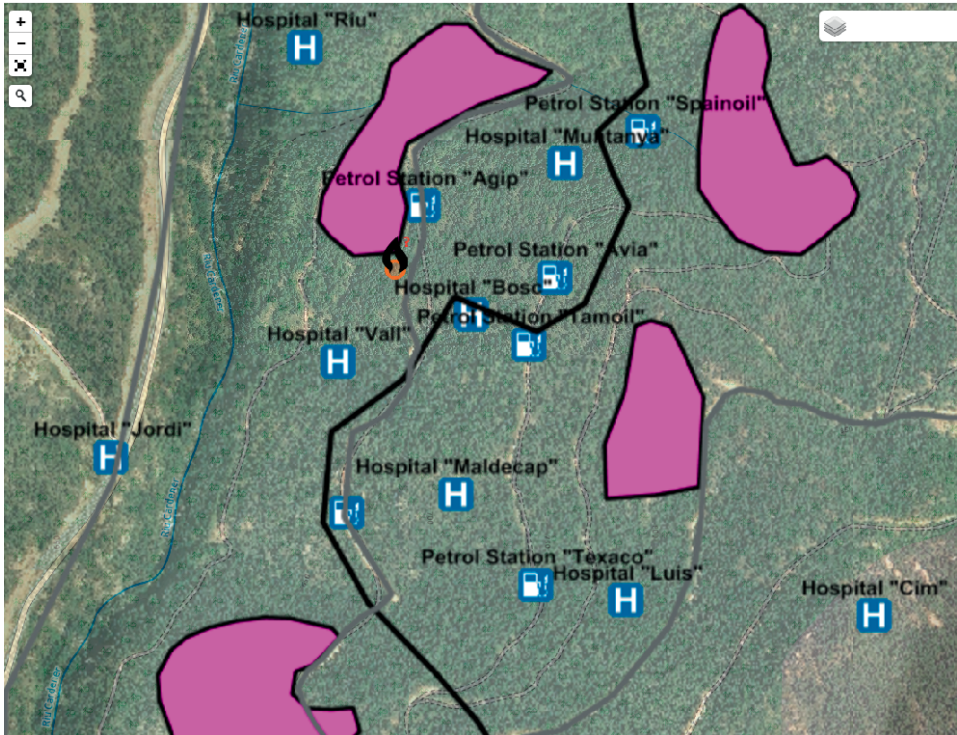


Fig. 3: Area of the prescribed burn with some exemplary virtual objects: hospitals (H), petrol stations, settlements (pink), power lines and streets

4.2 Description of the Prescribed Burn

Thanks to the support provided by the Forest Science Centre of Catalonia (CTFC), whose headquarters are located in Solsona, a control room was established in the CTFC premises. The fire could be monitored from the control room and a set of pre-defined exercises were carried out to test the PHAROS platform.

The execution of a prescribed burn (PB) was based on a prescribed burn plan, in which all detailed information of the PB is compiled. Before starting the PB, a briefing was carried out between all actors involved in the execution of the PB to clarify aspects related to the development of the PB and safety procedures for the staff:

- Review of PB objectives, as defined in a prescribed burn plan;
- Roles assignment, e.g. the head of burning, the head of ignition, the head of safety, the torches, the observers or watchers;

- Check of the radio communication;
- Definition of the smoke management;
- Recognition and tour of the plot, on foot or by car.

After this briefing, a test fire was carried out to evaluate if the fire behaviour was appropriate to achieve the PB objectives. Otherwise, the head of burning operations would reserve the right of postponing the PB by some hours or even cancelling it. If the tests were successful the exercises started.

Around 100 invited stakeholders from different organizations attended the PHAROS presentation and 25 of them could use and validate the platform. The demonstration involved the fire fighters experts that are commonly taking decisions during wildfires in the test region. Six experienced supporters, familiar with the PHAROS system, were distributed across four workstations in the control room. The supporters guided the end users through a forest fire exercise in order to explain them how the PHAROS platform worked; they simulated an emergency and recorded the participants feedback. Each workstation had a PHAROS platform running during the pilot demonstration. Groups of 2-4 end users gathered around each workstation while the main supporter explained the platform and performed the intervention on a simulated emergency.

The fire fighters performed the prescribed burnings during three days each with a duration of four to five hours. Thereby, the open fire line was around 80-100 meters. 12-20 fire fighters formed the team that carried out the burn. To give an idea of the fire some pictures of the fire can be seen in figure 4.



Fig. 4: Forest fire of the pilot demonstration

4.3 Fire Detection and Crisis Mapping

Besides the satellite data provided by the MODIS system a helicopter was used in cooperation with the project VABENE++ to provide the PHAROS system with real time data of the fire site. The BO 105 helicopter was equipped with two sensor systems: (i) AIR-

Sig sensors (thermal camera), (ii) a 4K camera. Different maps made from the helicopter during the burn are available at [Ze16].

For the fire detection the already mentioned FireWatch system was used. The FireWatch equipment was installed in a strategic location for fire surveillance. From there, the FireWatch cameras delivered smoke notifications to the PHAROS system at the start of every days exercises (figure 5, right side zoomed). In order to interconnect the FireWatch system with the SP a satellite link using F-SIM was selected. Satellites offer a robust connection in case of disasters and hazards. A 72 cm satellite dish for the F-SIM terminal was installed with line of sight to the geostationary satellite Ka-Sat at 9 degree East at same location hosting the FireWatch gateway.



Fig. 5: Forest fire detected by the FireWatch system

4.4 Decision Support with Virtual Objects

Considering that the chosen pilot area was in a remote area with few infrastructure and nearly no high-risk objects, in order to be able to test and show the features and capabilities of the PHAROS system during the pilot demonstration, the area and its surroundings needed to be filled with virtual objects. These virtual objects were fed to the PHAROS system (e.g. added to the geographic databases and the DSS configuration) in order to create a scenario of a populated (urban/rural) area with settlements, hospitals, petrol stations, streets and power lines. The placement of the virtual objects are shown in figure 3.

In order to assess the potential impact of wildfires assuming different environmental and weather conditions, a set of simulation runs were conducted and documented. An example where the prediction of the evolution of a fire is shown can be seen in figure 6 . The contours in the results show the evolution of the fire area on an hour basis, considering the wildfire to spread without control of the fire fighters. The results of the simulations were used by the DSS to assess the potential impact of the fire. The DSS successfully used the gathered data and simulations to support the user with notifications and alerting proposals. The DSS warned if critical infrastructure was affected so that the users could prioritize and plan next steps in fighting the fire and send warning messages to the public.

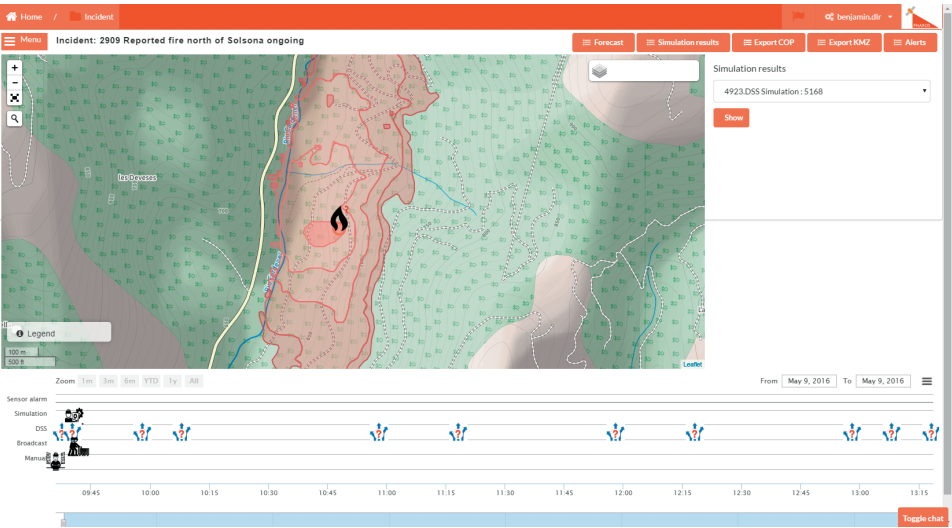


Fig. 6: PHAROS GUI showing simulation results

4.5 Communication with First Responders and Public Alerting

During the demonstration the users could use the communication tools at any time to communicate with the fire fighters on the field and warn the public. First responders were equipped with smart phones with a derivative of the mobile application that also provides a chat feature and the possibility to send the location of the device to the PHAROS SP. In addition, they also receive the message intended for the public warning.

In order not to disturb the public peace alerts were not transmitted via real channels to the public. They were transmitted and received in a lab environment. Smart phones with receiver applications connected to the alerting services via a VPN were present in the control room for this.

A second channel demonstrated for alerting purposes was cell broadcast. The cell broadcast alerting services consists of two building blocks: (i) a cell broadcast broker (CBB) which basically receives the message from the AG and forwards it to the users (described in more detail in [Mu14c], [Ba15]), (ii) the PHAROS cell broadcast network, which is implemented in a lab environment with OpenBTS, to implement the Mobile Operator Domain. For lab type research purposes, we were using a very small transmitter of only 100 mWatt, being allowed for research purposes, strong enough to test a couple of mobile receivers or mobile handsets, roamed on the used research network and placed within about a meter of the software radio set up.

5 End User Evaluation Results

The PHAROS system received a generalised positive feedback. The end users described the system as a spine of the emergency system. The novel component of this platform is that it integrates an emergency prevention and prediction system, as well as simulation and evaluation tools. Altogether, the different components of the platform will allow a better decision making before and during the early stages of the emergency, allowing then to prevent natural and humanitarian disasters. In general, minor improvements were suggested to increase the effectiveness of the platform to manage disaster emergencies. The highly valuable feedback given by the end users has been translated into recommendations to constitute a basis to advance in the development and implementation of the PHAROS platform. These recommendations and detailed feedback on the single components can be found at [Pr16]. Besides suggesting improvements on the existing features, new additional features were advised, e.g. the integration of a variety of existing/current systems such as TETRA, traffic management (TMC), and 112 services. Integration of additional systems and capabilities with PHAROS modular and scalable design will be relatively straightforward.

The end users reported that the mobile application was very useful and they particularly liked how it was integrated into the wider PHAROS system. A recommendation from the users in relation to the mobile application was that they would have liked to have even more textual and visual information displayed on the smart phone.

Besides the wide spectrum of improvements and augmentations that could be added to the PHAROS system, there also exists a major opportunity to provide a standardized emergency response alert system, in collaboration with mobile network operators, manufacturers and governmental agencies.

6 Conclusion and Outlook

In this paper we show that the PHAROS concept of a open integrated platform is able to be adapted to the forest fire case and successfully supports the crisis management. We describe the recent implementation of the PHAROS system and the execution of the pilot demonstration. For the PHAROS pilot three prescribed burns were organized and executed and the different parts of the PHAROS system were tested. Around 100 invited stakeholders from different organizations attended the PHAROS presentation and 25 of them could interact with the platform. This wide range of participants had a positive impact in order to provide an accurate feedback. The demonstration involved the fire fighters experts that are commonly taking decisions during wildfires in the test region. They had the possibility to explore the platform during a real forest fire event and carry out the simulation exercises that were prepared for them in the context of this pilot demonstration.

7 Acknowledgement

The research leading to these results has received funding from the European Community's Seventh Framework Programme ([FP7/2007-2013]) under grant agreement n [606982].

References

- [ALM13] Ascoli, D.; Lonati, M.; Marzano, R.: Prescribed burning and browsing to control three encroachment in southern European heathlands. 2013.
- [Ba15] Barth, B., ed. PHAROS Deliverable D4.4: Alert Service. 2015.
- [Co11] Costa, P.; Castellnou, M.; Larraaga, A.; Miralles, M.; Kraus, D.: Prevention of large wild-fires using the fire types concept. FireParadox & European Forest Institute. 2011.
- [Fe10] Fernandes, P.M.: Fire-smart management of forest landscapes in the Mediterranean basin under global change. 2010.
- [IW16] IQ-Wireless: , FireWatch - Waldbrand-Früherkennung für grosse Flächen, available at <http://www.iq-wireless.com/de/firewatch>, last accessed on May 2016.
- [Me16] Meteosat Second Generation (MSG) Spacecraft, June 2014, available at <https://directory.eoportal.org/web/eoportal/satellite-missions/m/meteosat-second-generation>, last accessed on March 2016.
- [MO16] MODIS: , available at <http://modis.gsfc.nasa.gov/about/>, last accessed on May 2016.
- [Mu14a] Mulero Chaves, J., ed. PHAROS Deliverable D2.1: PHAROS Service Concept Specification. 2014.
- [Mu14b] Mulero Chaves, J., ed. PHAROS Deliverable D2.9: PHAROS System Architecture-Final. 2014.
- [Mu14c] Mulero Chaves, J., ed. PHAROS Deliverable D4.3: Alert System Specification, volume First Issue. 2014.
- [Of16] Official Website of the Department of Homeland Security: , Definition of Critical Infrastructure, June 2014, available at <http://www.dhs.gov/what-critical-infrastructure>, last accessed on May 2016.
- [Pr16] Prat, N., ed. PHAROS Deliverable D7.3: End-users Evaluation on PHAROS Products and Services. 2016.
- [RST13] Russell-Smith, J.; Thornton, R., eds. Prescribed burning in fire-prone landscapes. *Frontiers in Ecology and the Environment*, volume 11. 2013.
- [Te16] Tecnosylva: , Wildfire Analyst, available at <http://www.tecnosylva.com/index.php/productos/desarrollos/wildfire-analyst>, last accessed on May 2016.
- [Th16] The United Nations Office for Disaster Risk Reduction (UNISDR): , Information and Knowledge Management for Disaster Risk Reduction (IKM4DRR) Framework and Scorecard, May 2013, available at <http://www.preventionweb.net>, last accessed on May 2016.

- [Ze16] Zentrum für Satellitengestützte Kriseninformation (ZKI): , Forest fire exercise in Catalonia (Pilot demonstration PHAROS-System), March 2016, available at <https://www.zki.dlr.de/article/2773>, last accessed on May 2016.

Flexible Kommunikations- und Informationslösungen für eine optimierte Einsatzführung von Interventionskräften

Christian Raffelsberger¹, Martina Umlauf¹, Andreas Kercek¹, Alexander Almer²,
Thomas Schnabel², Patrick Luley²

Abstract: Im Rahmen des Projektes „4C4FirstResponder“ wird die Entwicklung einer Kommunikations- und Informationslösung zur optimierten Einsatzführung von Interventionskräften angestrebt. Dazu werden multisensorale Daten (Text, Audio, Bilder, Video, etc.) generiert, zu einem Lagebild zusammengeführt sowie rollenbasiert verteilt (an mobile Teams, Einsatzleitung, etc.). Kern der Kommunikationslösung ist ein Kommunikationsmanager, der TETRA und andere Netzwerktechnologien integriert, die Vorteile aller Technologien nutzen kann und Daten entsprechend von QoS-Anforderungen über diese verschiedenen Technologien verschickt. Eine weitere wichtige Komponente der Lösung ist eine Szenarien-orientierte Darstellung der Daten und Bedienung auf (mobilen) Endgeräten. Das Projekt verfolgt einen agilen Entwicklungsansatz und startete mit einer Bedarfserhebung bei den Endbenutzern (Einsatzkräfte). Damit konnte eine erste Identifikation von Einsatzszenarien und Anwendungsfällen erfolgen und ein Architekturentwurf des Systems und des Kommunikationsmanagers erstellt werden, welche in dieser Arbeit vorgestellt werden.

Keywords: First Responder, Breitbandkommunikation, TETRA, mobile Assistenzsysteme, echtzeitnahes Lagebild, mobile Einsatzzentrale, Kommunikationsmanager.

1 Einleitung und Motivation

Komplexe Anforderungen an Einsatzkräfte durch sehr unterschiedliche und sich oft dynamisch verändernde Sicherheitslagen, bei gleichzeitiger Anforderung eines effizienten Personaleinsatzes, erfordern einen gezielten Einsatz von neuen Informations- und Kommunikationslösungen [LH08], [PP08]. Die echtzeitnahe Erstellung eines aktuellen Lagebildes, mobile multisensorale Assistenzlösungen sowie eine stabile Sprach- und Datenkommunikation unterstützen das koordinierte Zusammenwirken unterschiedlicher Einheiten bzw. Einsatzkräfte und sind somit wesentliche Faktoren für eine optimierte Einsatzführung und damit für einen gezielten und effizienten Einsatz [Ge10]. Daher sind umfangreiche nationale und internationale Innovationsbestrebungen und Forschungsaktivitäten im Gange, um Einsatzkräfte mit umfangreicher Sensorik und neuen Kommunikationsmöglichkeiten auszustatten (siehe Abschnitt 2).

¹ Lakeside Labs, Lakeside B04, 9020 Klagenfurt, {raffelsberger | umlauf | andreas.kercek}@lakeside-labs.com

² JOANNEUM RESEARCH, Steyrergasse 17, 8010 Graz, {firstname.lastname}@joanneum.at

TETRA ist der derzeit eingesetzte digitale, sichere Funkstandard für Einsatzkräfte in Österreich und in vielen europäischen Staaten. TETRA eignet sich jedoch lediglich für die Übertragung von Sprache und Textnachrichten, bietet jedoch nicht genügend Bandbreite für komplexe multimediale Dienste. Es gibt zwar internationale Aktivitäten die Bandbreite von TETRA zu erhöhen [St11], aber wie in [Fe13] beschrieben, existiert eine Reihe von technischen, organisatorischen sowie wirtschaftlichen Faktoren, die eine rasche Umsetzung erschweren.

Als Kern des hier beschriebenen Projekts 4C4FirstResponder wird daher eine flexible, multimediale Kommunikationslösung angestrebt, die neben TETRA auch Breitband Technologien wie UMTS/LTE und WiFi einbindet. Dies ermöglicht eine Szenarienorientierte und zeitoptimierte Koordination von First Respondern und Einsatzstäben (Polizei, Feuerwehr, Rettung, Hilfsorganisationen, Sicherheitspersonal, etc.). In einem im Rahmen des Projekts durchgeführten User Workshop (siehe Abschnitt 3) wurden Anforderungen der Einsatzkräfte erhoben, wobei die Verwendung von TETRA als zentrales Kriterium zur Sicherstellung der Kommunikationssicherheit identifiziert wurde. Daher wird TETRA als sichere Übertragungstechnik im Projekt berücksichtigt (siehe Abschnitt 4.2).

Mobile Sicherheitskräfte werden als „Human Sensor“ eingebunden, um eine effizientere Interaktion zwischen der Einsatzzentrale und den mobilen Teams zu ermöglichen. Die Ausstattung mit körpergetragenen Sensoren und Videokameras sowie mobilen Displays („handheld“ oder „helmet-mounted“) sind dabei grundlegende Voraussetzungen. Mobile Einsatzteams sind somit Sensoren und Akteure zugleich und ermöglichen so optimierte Datengewinnung und vernetzte Einsatzführung (siehe Abschnitt 4.1). Ein wesentlicher Faktor ist die situationsbezogene, fokussierte Unterstützung der Einsatzkräfte durch mobile Assistenzsysteme und die Vermeidung von „Ablenkung“ durch digitale Systeme.

Wesentliches Verbesserungspotential in Bezug auf existierende Lösungsansätze besteht in folgenden Bereichen:

- Sichere und robuste Übertragung von Multimediatdaten auch unter rauen Umgebungsbedingungen und im Fall, dass manche Übertragungstechniken nicht bzw. nicht permanent zur Verfügung stehen. Sowie eine flexible Anpassung der Kommunikationstechnologie an diese wechselnden Rahmenbedingungen (dynamische Einbindung von Hierarchien, wechselnde QoS-Anforderungen, dynamisch auftauchende / verschwindende Netze/Verbindungen).
- Szenarien-fokussierte, einsatztaugliche Bedienung und User Interfaces sowie Prozesseinbindung von mobilen Unterstützungssystemen. Dabei ist eine optimale gruppen- und rollenorientierte Vernetzung und Datenverteilung erforderlich.
- Die Anforderungen und die Akzeptanz der Kommunikationslösungen und mobilen Unterstützungssysteme sowie die bedarfsträgerechte und rollenorientierte Einbindung der Einsatzorganisationen und der involvierten Gruppen sollen in 4C4FirstResponder umfangreich untersucht werden. Diese empirisch gesicherten

individuellen Wahrnehmungsprofile sollen dann in eine sogenannte „konvergente Lagebildarstellung“ Eingang finden.

2 Verwandte Arbeiten

Forschung im Bereich neuer Kommunikationstechnologien für Notfalleinsätze ist ein breites Feld. Die meisten Ansätze gehen von einem (Total-) Ausfall der bestehenden kommerziellen Kommunikationsinfrastruktur aus und präsentieren Alternativen, vor allem aus den Bereichen mobiles Ad-Hoc Networking (MANET), opportunistische bzw. verzögerungstolerante Netze (DTN) sowie hybride Ansätze [Le11], [MC13], [RH13], [Re15].

Visionen für pan-europäische Systeme, um die Zusammenarbeit verschiedener Einsatzorganisationen in Großschadensszenarien zu verbessern, werden in [Do15], [Gr14], [He14] und [Ku15] beschrieben. Ähnlich zu 4C4FirstResponder wird in [Gr14] eine Integration von mehreren Kommunikationsnetzen vorgeschlagen. Der Fokus dieser Arbeiten liegt allerdings auf Großschadenslagen, während 4C4FirstResponder auf kleinere, öfter wiederkehrende Einsätze abzielt.

ServalMesh [GS13] ist eine DTN-basierte Plattform, die auf Android Smartphones läuft und sich sowohl an Betroffene wie auch an Einsatzkräfte richtet. [Ji11] stellt eine DTN-basierte Middleware Architektur vor, mit dem Ziel existierende Systeme wie TETRA zu ergänzen, um die Einbindung von Freiwilligen mit Android Mobilgeräten zu ermöglichen. [LCL16] beschreibt ein System mit Android Mobilgeräten, das Mobilfunknetze mit Ad-hoc- und opportunistischen Netzen vereint. Im Gegensatz zu 4C4FirstResponder liegt der Fokus bei diesen Ansätzen auf der Kommunikation zwischen Einsatzkräften und Betroffenen (z.B. verschüttete Personen nach Erdbeben) bzw. Betroffenen untereinander.

Ein Kommunikationssystem für Feuerwehren auf Basis von mobilen Geräten und drahtlosen Routern und Erfahrungen aus Übungseinsätzen werden in [Ng15] präsentiert. Das System besteht neben den mobilen Endgeräten aus einem Ad-hoc Netzwerk zur Vernetzung von Fahrzeugen am Einsatzort und ermöglicht die Einbindung einer Einsatzzentrale per UMTS/LTE. Im Gegensatz zu 4C4Responder bleibt jedoch TETRA ungenutzt.

Im EU FP7 Projekt E-SPONDER [Ca03], [Wal14] wurden Körpersensoren für Einsatzkräfte entwickelt und getestet. Die weitere Netzwerkstruktur ist an die typischen Hierarchieebenen angepasst: die Ersthelfer sind untereinander und mit einer mobilen Einsatzzentrale vernetzt und diese wiederum mit der zentralen Einsatzleitung. Im Gegensatz zu 4C4FirstResponder wird im E-SPONDER System sämtliche Sprachkommunikation zwischen den Einsatzkräften über ein Voice-Over-IP System abgewickelt.

Um den Benutzer- und Umgebungskontext im Einsatzfall einzubeziehen und eine optimale Anpassung des Systems zu ermöglichen, wurden verschiedene elektronische Assistenzsysteme vorgestellt [BL12], [Ca12]. Die Einbeziehung von sogenannten Head Mounted Displays [Wi05] oder Smart Glasses als Unterstützung für First Responder in speziellen Szenarien gewinnt stark an Aufmerksamkeit. Beispielsweise wird ein Smart Glasses Prototyp im Projekt “Augmented Reality Tools for Improved Training of First Responders”³ beschrieben, wobei Einsatzkräfte mit Smart Glasses ausgestattet werden, um Videos des Einsatzortes zu übertragen sowie Echtzeitinformationen zu erhalten. Mobile, zum Teil körpergetragene Sensoren sind ein weiteres Mittel um Einsatzkräfte zu unterstützen [Go12], [Ta12]. Die in 4C4FirstResponder zu entwickelnden Assistenzsysteme werden auf diese Erfahrungen aufbauen.

Wie sich gezeigt hat, ist die Zusammenarbeit mit Endanwendern beim Entwurf eines solchen Systems sehr wichtig [Ch12], [Ch13], [Ge10]. Aus dieser Erfahrung heraus ist auch für 4C4FirstResponder eine frühe Einbindung von Einsatzkräften vorgesehen (siehe Abschnitt 3).

Zusammenfassend kann gesagt werden, dass bei Kommunikations- und Assistenzlösungen in diesem Bereich typischerweise bereits bekannte Protokolle (z.B. TCP/IP, HTTP) und kommerzielle Endgeräte (Laptops, PDAs, Smartphones) zum Einsatz kommen. Weiters werden oftmals bestehende Netze mit Ad-hoc Netzwerken bzw. DTN oder hybriden DTN/MANET Netzwerken erweitert.

3 Der 4C4FirstResponder Ansatz

Aus der Erfahrung anderer Projekte (siehe Abschnitt 2) hat sich gezeigt, dass eine frühe Einbindung der Endanwender (Feuerwehr, Polizei, Rettungskräfte, etc.) wichtig ist, daher wird ein agiler Entwicklungsansatz verfolgt. Es werden den Endanwendern über den gesamten Entwicklungsverlauf Demonstratoren zur Verfügung gestellt. Mit Hilfe des Nutzerfeedbacks werden sowohl Anforderungen als auch Entwicklungsergebnisse iterativ angepasst. So wird sichergestellt, dass am Ende des Projektes Verfahren und Lösungen entstehen, die für die Zielgruppen einen wesentlichen Mehrwert darstellen.

Um den agilen Prozess in Gang zu bringen, fand zu Beginn des Projektes ein halbtägiger Workshop statt, bei dem die Endanwender die aus ihrer Sicht wichtigsten Anwendungsfälle sowie ihre Anforderungen darlegen konnten. An dem Workshop nahmen sechs Vertreter der Polizei sowie ein Feuerwehrkommandant teil.

Am Beginn des Workshops wurden die Nutzer zu ihren derzeitigen Arbeitsabläufen sowie eingesetzten Technologien befragt. Dabei hat sich herausgestellt, dass TETRA für die Kommunikation sensibler Daten zwingend gefordert wird. Beispielsweise sollen Einsatzbefehle nur über TETRA übertragen werden. Neben der bestehenden

³ <http://srl.mcgill.ca/projects/responder/>

Sprachkommunikation sehen die Einsatzkräfte in vielen Szenarien den Bedarf an der Einbindung weiterer Kommunikationsmodalitäten, wie Multimedia oder Sensordaten. In unklaren Situationen können diese Daten hilfreich sein, da Sprachkommunikation alleine nicht immer ausreicht. Als Beispiel wurde genannt, dass die Einsatzzentrale oftmals über Luftaufnahmen verfügt, diese aber nicht den Einsatzkräften vor Ort zur Verfügung stehen. Diese Bilder würden jedoch helfen die Einsatzziele besser zu beschreiben, wodurch Missverständnisse der reinen Sprachkommunikation ausgeräumt werden könnten. Eine weitere Anforderung in diesem Zusammenhang ist die Geo-Referenzierung dieser Daten, um die Darstellung in einer Karte zu ermöglichen.

Basierend auf diesen Anforderungen, wurde gemeinsam mit den Einsatzkräften über Szenarien diskutiert, bei denen diese Anforderungen bisher oft nur unzureichend erfüllt werden. Dabei wurden mehrere Szenarien genannt: spontane Demonstrationen, „Flashmobs“ oder z.B. Fußballspiele, bei denen sich große Fangruppen zwischen Bahnhof, Stadion und Innenstadt bewegen. Diese Szenarien sind in der räumlichen Ausdehnung schwer planbar, wodurch oftmals ein hohes Informationsdefizit in den unterschiedlichen Einsatzsituationen herrscht. Daten und Informationen für eine gezielte und effiziente Einsatzführung sind meist nicht zeitnah und in der erforderlichen Qualität vorhanden bzw. können nicht zeitgerecht an Personen und Teams verteilt werden, was die Koordination geeigneter Maßnahmen im Falle von kritischen Situationen erschwert. Hier würde eine kamerabasierte Lagebilderstellung hilfreich sein.

Ein weiteres, oft genanntes Problem bei Einsätzen ist die Überlastung mit Information, die zu schlechterer Effizienz führen kann. Dieses Problem taucht bei allen Szenarien und Anwendungsfällen auf und ist auch ein wesentlicher Aspekt, der bei den Entwicklungen im Rahmen des Projektes berücksichtigt werden muss.

4 Systemkonzept

Wie zuvor beschrieben, liegt der Schwerpunkt des Projekts einerseits in der Vernetzung und Demonstration der Einsatzfähigkeit von Kommunikationsmöglichkeiten unter Einbindung von Breitbandkommunikationslösungen sowie dem digitalen BOS⁴ Funkstandard TETRA. Andererseits in der Vernetzung von mobilen Sicherheitskräften, ausgestattet mit körpergetragenen Sensoren, Videokameras und mobilen Displays. Die für Einsatzszenarien notwendige Kommunikationssicherheit wird dabei durch einen Kommunikationsmanager (KomMan) gewährleistet, der in kritischen Bereichen TETRA als sichere Übertragungstechnik nutzt. Abb. 1 gibt einen Überblick der geplanten Module bzw. Systemkomponenten für die „Mobile Einsatzzentrale“ und die „Mobilen Unterstützungssysteme“. Wesentlich dabei ist, dass die „Mobile Einsatzzentrale“, je nach Konfiguration, sowohl in einer mobilen Kommandozentrale (z.B. Führungscontainer) als auch bei Abschnittskommandanten (z.B. Kommandofahrzeug) eingesetzt werden kann.

⁴ BOS - Behörden und Organisationen mit Sicherheitsaufgaben

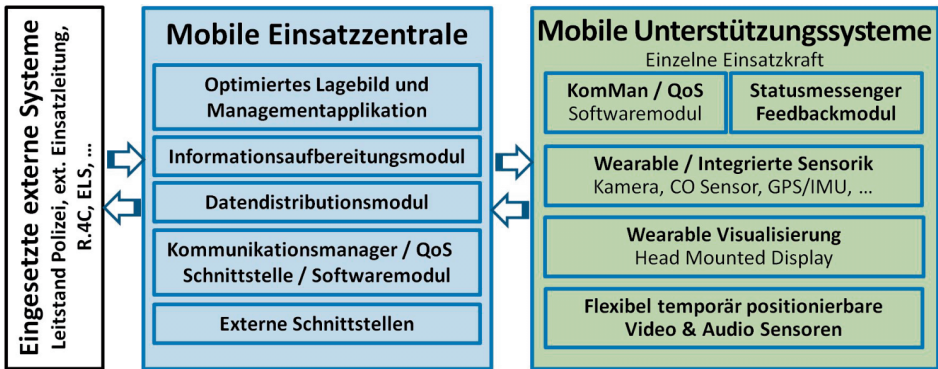


Abb. 1: Schematische Darstellung der 4C4FirstResponder Systemkomponenten.

Nachfolgend werden einige der Kernkomponenten des 4C4FirstResponder-Systems im Detail beschrieben.

4.1 Mobile Assistenzsysteme und Lagebilddarstellung

Um die Kommunikation und Handlungsabläufe für die Einsatzkräfte zu optimieren, ist ein kontextsensitives AR-Assistenz- und Feedbacksystem geplant, welches sich dynamisch an die individuelle Einsatzsituation der Nutzer anpasst und bedarfsorientiert relevante Informationen dem Sehfeld überlagert.

Mobile Sicherheitskräfte werden hierbei als „Human Sensor“ eingebunden und haben auch die Möglichkeit, auf Sensorinformationen zuzugreifen. Mobile Einsatzteams sind somit Sensoren und Akteure zugleich und ermöglichen eine optimierte Datengewinnung und vernetzte Einsatzführung. Die Evaluierung von geeigneter Hardware auf Basis der definierten Nutzeranforderungen sowie Usability-Tests von „Head mounted“ Display Lösungen in spezifischen Einsatzszenarien sind dabei wichtige Projektziele.

Das geplante Feedbacksystem umfasst, neben der Erfassung von Vital- und Umgebungsdaten der Einsatzkräfte, auch die Echtzeitübertragung von videogestützten Lageberichten an die übergeordnete Kommunikationseinheit oder direkt zu weiteren Einsatzkräften. Ergänzt wird das Feedbacksystem durch ein kontextsensitives Assistenzsystem mit minimaler Nutzerinteraktion, welches relevante Informationen direkt im Sehfeld der Einsatzkräfte einblendet. Speziell in stressreichen Situationen, in denen die Aufnahme- und Handlungsfähigkeit beschränkt sind [CN08], kann eine solche Assistenz ein effektives und intuitives Feedback- und Informationssystem darstellen. Um die dynamische Darstellung der richtigen Information im Sehfeld zu ermöglichen, ist das automatische Erkennen des Kontexts der jeweiligen Einsatzkraft und ihrer situationsbezogenen Bedürfnisse eine große Herausforderung, welche im Projekt behandelt wird. Basis der Kontexterkenkung stellen die Daten der am Körper getragenen Sensorik dar. Hierbei wird eine Reihe von Bio- und Kontextsignalen (Atemfrequenz, Puls, Hautwiderstand, Position und Geschwindigkeit) mittels Messgurt und Armband in

Echtzeit erfasst und verarbeitet. Beispiele für Hardware, die diese Aufgaben durchführen kann, sind etwa der Brustgurt Bioharness-3 von Zephyr⁵ und das Armband Empatica E4⁶. Herausfordernd dabei sind die Echtzeitverarbeitung einer Vielzahl komplexer Sensorinputs und deren Wechselwirkung untereinander. Diese Aspekte werden im Zuge des Projekts mit Modellen und Algorithmen, die mit einer dynamischen Umgebung umgehen können, angesprochen.

Multisensorale Daten (Bilder, Videos, Text, akustische Informationen) sowie situationsbezogene Informationen der mobilen Einsatzkräfte (Position, Status, Situation, etc.) ergeben auf Basis von kartographischen Referenzdaten ein umfassendes Lagebild. Die Funktionalität des Kommunikationsmanagers (siehe Abschnitt 4.2) gewährleistet die echtzeitnahe Zusammenführung von unterschiedlichen Informationsquellen und damit eine neue Qualität eines aktuellen Lagebildes zur Unterstützung von zeitkritischen Entscheidungsprozessen. Die Entwicklung einer mobilen, intelligenten Einsatzzentrale, welche z.B. in einem Führungscontainer oder Kommandofahrzeug zum Einsatz kommt, ermöglicht den Nutzern einen rollenorientierten Zugang zur aktuellen Lageinformation. Die mobile Einsatzzentrale inkludiert ein Lagebild- und Managementmodul, dient der Erfassung, aktuellen Visualisierung und Verteilung der Lageinformation und Situationsdaten. In Abstimmung mit dem nachfolgend beschriebenen Kommunikationsmanager werden zur Verfügung stehende Bandbreiten und Kommunikationstechnologien berücksichtigt und Daten entsprechend aufbereitet, um einen optimierten Informationsaustausch zu realisieren.

4.2 Kommunikationsmanager

Eine wichtige Komponente der 4C4FirstResponder Systemarchitektur ist der Kommunikationsmanager. Dieser wählt anhand der aktuellen Eigenschaften der verfügbaren Kommunikationskanäle, QoS-Anforderungen der Applikationen und Charakteristika der zu übertragenden Daten, die passenden Kommunikationsnetze aus. Der Kommunikationsmanager agiert dabei als Vermittler zwischen den Anforderungen der Applikationen einerseits und den verfügbaren Kommunikationsnetzen andererseits. Bei der Auswahl von geeigneten Kommunikationskanälen arbeiten verschiedene Instanzen des Kommunikationsmanagers, die auf den verschiedenen Hierarchieebenen eingesetzt werden, zusammen und treffen Entscheidungen in einer dezentralen und selbstorganisierenden Weise. Abb. 2 zeigt die Einbindung des Kommunikationsmanagers in die Hierarchieebenen des 4C4FirstResponder Gesamtsystems.

⁵ <https://www.zephyranywhere.com/products/bioharness-3>

⁶ <https://www.empatica.com/e4-wristband>

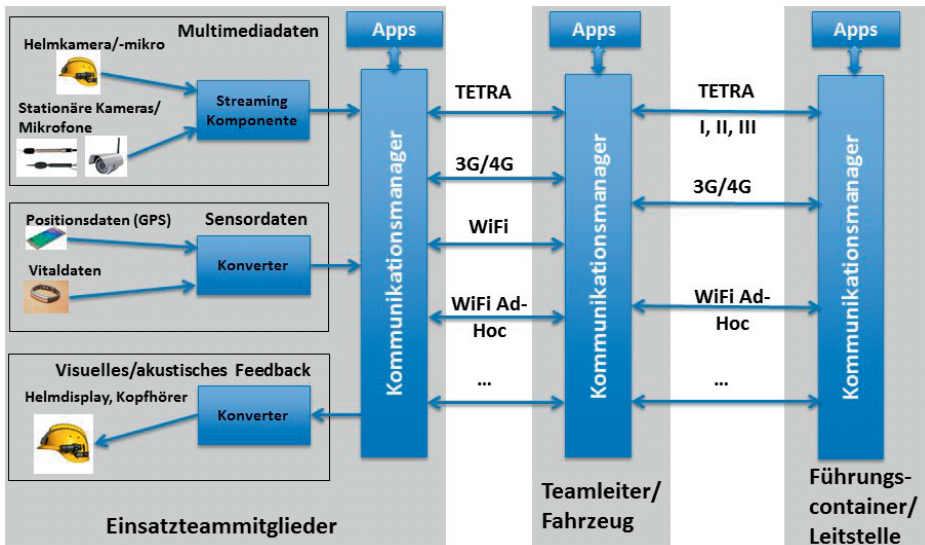


Abb. 2: Einbettung des Kommunikationsmanagers in die verschiedenen Hierarchieebenen.

Der Kommunikationsmanager wird als Software-Lösung realisiert, die auf verfügbarer Standard-Hardware eingesetzt werden kann (z.B. auf Android Smartphones, oder Linux-basierten Routern) und dabei soweit wie möglich auf bestehende Netzwerkprotokolle zurückgreift.

Die Architektur des Kommunikationsmanagers ist in Abb. 3 dargestellt. Der „Netzwerk Monitor“ sammelt laufend Informationen über die Netzwerkschnittstellen. Diese Informationen werden einerseits vom „Scheduler-Modul“ des lokalen Kommunikationsmanagers bei der Auswahl der Schnittstellen verwendet. Andererseits werden diese auch im Netzwerk verteilt (Discovery-Modul), damit eine Koordination der Kommunikationsaufgaben ermöglicht wird. Außerdem werden die Statusinformationen den Applikationen bereitgestellt, damit diese darstellen können, welche Art der Datenübertragung verfügbar ist (z.B. nur Text, falls nur eine TETRA Verbindung verfügbar ist). Der Kommunikationsmanager unterstützt auch eine Priorisierung von Kommunikationsaufgaben. Dies ist insbesondere in Situationen notwendig, in denen die Netzwerkressourcen nicht ausreichen, um alle Kommunikationsanforderungen zu erfüllen. Die Priorität der Daten wird dabei anhand von Applikations- bzw. Benutzervorgaben festgelegt.

Wie zuvor beschrieben, wird der Kommunikationsmanager auf mehreren Hierarchieebenen eingesetzt. Um aktuell verfügbare Kommunikationspartner zu finden, besitzt der Kommunikationsmanager ein Discovery-Modul, welches andere Instanzen des Kommunikationsmanagers in den verfügbaren Kommunikationsnetzen findet. Darüber hinaus hat das Discovery-Modul die Aufgabe, Informationen wie z.B. Netzwerkadressen der Kommunikationspartner zu sammeln und im Netzwerk zu verteilen.

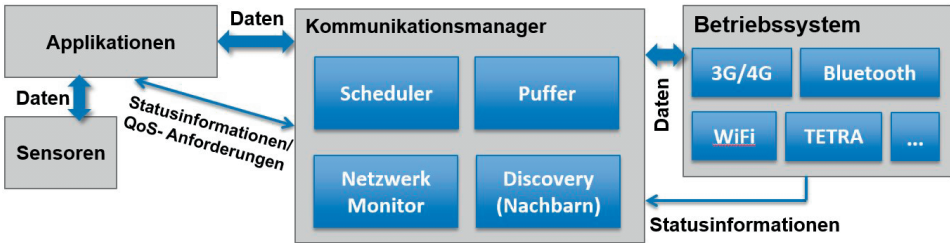


Abb. 3: Architektur des Kommunikationsmanagers.

Aufgrund der lokal gesammelten Statusinformationen und empfangenen Informationen von anderen Kommunikationsmanagern, werden die von den Applikationen übergebenen Daten versendet. Da die vorhandenen Netzwerkressourcen nicht immer ausreichen um Daten sofort zu senden, werden Daten in einem Puffer zwischengespeichert. Die Auswahl welche Daten wann gesendet werden, wird vom Scheduler-Modul getroffen. Dabei werden die QoS-Anforderungen und vorgegebenen Prioritäten berücksichtigt.

Die Unterstützung von mehreren Kommunikationsschnittstellen ist eine wesentliche Aufgabe des Kommunikationsmanagers. Dadurch kann die Bandbreite mehrerer Schnittstellen aggregiert und der Ausfall einzelner Schnittstellen kompensiert werden, oder eine optimierte Wahl der Schnittstelle erfolgen (z.B. „WiFi Offloading“ um Daten nicht über evt. teure Mobilfunknetze zu übertragen). Trotz dieser Möglichkeiten und der Verbreitung von mobilen Geräten, die über mehrere Schnittstellen verfügen, wird die Multischnittstellen-Kommunikation (auch Multipfad-Kommunikation) von den meisten Standard Netzwerk- und Transportprotokollen und Applikationen nicht unterstützt. Daher wurden in den letzten Jahren verschiedene Ansätze vorgestellt, die eine bessere Nutzung mehrerer vorhandener Schnittstellen ermöglichen [Ha15]. Die vorgeschlagenen Lösungen arbeiten auf verschiedenen Schichten des OSI-Schichtenmodells, wobei die meisten Ansätze in der Transport- oder Applikationsschicht implementiert sind. Bei der Entwicklung des Kommunikationsmanagers kann daher auf bestehende Arbeiten auf dem Gebiet der Multi-Schnittstellenkommunikation und Datencharakterisierung zurückgegriffen werden.

Ein prominenter Ansatz ist Multipath TCP (MP-TCP) [Ra12], bei dem eine TCP-Verbindung auf mehrere Pfade (bzw. mehrere Schnittstellen) aufgeteilt werden kann. Der Vorteil von MP-TCP ist, dass keine Änderungen auf Applikationsebene notwendig sind. MP-TCP setzt allerdings eine Änderung des Betriebssystems auf Sender- und Empfängerseite voraus. Ein weiterer Nachteil ist die Middlebox-Problematik [Ra12], bei der Knoten entlang eines Pfades zwischen Sender und Empfänger MP-TCP Segmente verändern oder blockieren können. Andere Ansätze [Ki14], [Ta15] basieren auf HTTP und sind somit in der Applikationsschicht implementiert. HTTP bietet durch die Verwendung des „Range“-Headers die Möglichkeit, bestimmte Teile einer Ressource auszuwählen. Durch Verteilen der HTTP Anfragen auf unterschiedliche Schnittstellen, kann die Bandbreite dieser Schnittstellen aggregiert oder zwischen verschiedenen

Schnittstellen gewechselt werden, um z.B. Ausfälle zu kompensieren. Der Vorteil von HTTP liegt vor allem in der weiten Verbreitung und Unterstützung des Protokolls. So verwenden bereits viele Applikationen HTTP zur Datenübertragung und können somit einfach um Multipfad-Kommunikation erweitert werden [Ki14]. Im Bereich von mobilen Netzen, haben Deng et al. [De14] ein Transportmodul vorgestellt, das versucht eine optimale Kommunikationsstrategie zu finden. Die Strategie wird dabei auf Basis einer Vorhersage von Netzwerkcharakteristika (z.B. Signalstärke, durchschnittlicher Datendurchsatz) und Zielvorgaben der Applikation, wie „minimiere Verzögerung“ oder „maximiere Bandbreite“, gewählt. Bei einem ähnlichen Ansatz [Hi10] können Applikationen die Art der Datennutzung festlegen (z.B. Hintergrundverkehr, der höhere Latenzzeiten erlaubt, oder Interaktive Nutzung), damit die für diese Nutzungsart beste Kombination von Netzwerkanälen gewählt wird.

Bei Gesprächen mit Vertretern verschiedener Einsatzorganisationen (siehe Abschnitt 3) hat sich gezeigt, dass die Anforderung besteht, gewisse Vorgaben zur Datenübertragung (z.B. Priorität der Daten, Sicherheitsanforderungen) machen zu können. Aus diesem Grund sind Multipfad-Ansätze, die aus Sicht der Applikation transparent arbeiten nicht geeignet. Eine wichtige Aufgabe besteht daher in einer Untersuchung inwieweit sich bestehende Multipfad-Ansätze auf Einsatzszenarien umlegen lassen.

Darüber hinaus hat sich gezeigt, dass seitens der Einsatzkräfte die Akzeptanz für die Verwendung von kommerziellen Netzen zur Übermittlung von missionskritischen Informationen fehlt. Außerdem ist die Nutzung von TETRA Sprechfunk ein wesentlicher Bestandteil heutiger Einsatzprozedere. Im Gegensatz zu anderen Architekturen für moderne Einsatzmanagementsysteme (siehe Abschnitt 2), wird TETRA daher bei der 4C4FirstResponder Architektur berücksichtigt. Die Integration von TETRA und kommerziellen Breitbandnetzen kann auf mehrere Arten erfolgen: *Integrated TETRA Gateway*, *On-Site TETRA Gateway* und *Off-Site TETRA Gateway*.

Beim integrierten TETRA Gateway werden Geräte eingesetzt, die über alle notwendigen Kommunikationsschnittstellen verfügen. Der Nachteil dieser Lösung ist, dass bestehende Geräte ausgetauscht werden müssen, was wenig praktikabel ist. Alternativ könnte das Peripheral Equipment Interface (PEI) verwendet werden, das es anderen Geräten ermöglicht, ein TETRA Gerät als Gateway zu verwenden. Der Nachteil hierbei wäre allerdings, dass die Einsatzkraft ein zusätzliches Gerät tragen müsste, was von den Nutzern im Workshop kritisch gesehen wurde. Das TETRA Gateway könnte auch vor Ort, z.B. in einem Fahrzeug, eingesetzt werden. Die Kommunikation zwischen Einsatzkraft und TETRA Gateway erfolgt dabei über ein lokales Netzwerk, wie z.B. WiFi. Alternativ könnte das TETRA Gateway auch an einem externen Standort (z.B. der Einsatzzentrale) eingesetzt werden. Hierbei ist man allerdings von einer bestehenden Kommunikationsinfrastruktur wie z.B. LTE abhängig, damit eine Einsatzkraft das TETRA Gateway erreichen kann. Da durch diese Abhängigkeit die Robustheit der Kommunikation negativ beeinflusst wird, liegt der Fokus im Projekt auf den ersten beiden Lösungsansätzen (Integrated bzw. On-Site TETRA Gateway).

5 Zusammenfassung und Ausblick

In dieser Arbeit wurde das 4C4FirstResponder Projekt vorgestellt. Ziel des Projekts ist die Entwicklung einer flexiblen Kommunikationslösung, die eine innovative Lagebilderstellung und optimierte Einsatzführung ermöglicht. Dabei werden Nutzer auf allen Ebenen, vom Einsatzteammitglied bis zur Einsatzführung, in einer Einsatzleitzentrale eingebunden. Mobile Sicherheitskräfte werden mit körpergetragenen Sensoren und Videokameras, sowie mobilen Displays ausgestattet und somit als „Human Sensor“ eingebunden. Auf Basis der Ergebnisse des Workshops mit verschiedenen Bedarfsträgern wurde ein Architekturentwurf vorgestellt. Basierend auf dieser Architektur wird gerade ein erster Prototyp erstellt, der in realen Tests mit Einsatzkräften getestet werden soll. Der Fokus liegt auf der Demonstration der Einbindung mehrerer Netzwerke zur Kommunikation zwischen der Einsatzzentrale und einem Abschnittskommandanten, der mit einem einsatztauglichen Tablet ausgestattet ist. Dabei soll die Übermittlung von Multimediainformationen im Rahmen eines typischen Einsatzprozederes getestet werden, um Feedback zu Nutzerschnittstellen zu erhalten und weitere Anforderungen zu erheben.

In weiterer Folge wird ein kamerabasiertes System zur Verbesserung des lokalen Lagebildes entwickelt. Das System besteht aus einem mobilen Rechner, der an einer Teleskopstange montiert wird. Dadurch ermöglicht das System die Aufnahme von Bildern und Videos aus einer erhöhten Position. Zusätzlich enthält das System Geo- und Lagesensoren, damit sowohl der Standort als auch die Blickrichtung der Aufnahme ermittelt werden können. Dadurch ist es möglich die Aufnahmen innerhalb einer Karte zu verorten. Die Aufnahmen werden dann vom Kommunikationsmodul (siehe Abschnitt 4.2) weiter versendet.

Um Informationsüberflutung zu vermeiden, sollen die gesammelten Daten rollenbasiert gefiltert werden. Nutzer haben hierbei auch die Möglichkeit in den Filterprozess einzugreifen. Eine Herausforderung, die in diesem Zusammenhang gelöst werden muss, ist Interaktionsmöglichkeiten zu finden, die einerseits die Einsatzkräfte nicht überfordern, es ihnen aber andererseits ermöglichen, bestimmte Informationen als besonders wichtig hervorzuheben.

Das Projekt wird voraussichtlich bis Ende 2017 laufen.

Danksagung: Dieses Projekt wird finanziert im Sicherheitsforschungs-Förderprogramm KIRAS vom Bundesministerium für Verkehr, Innovation und Technologie (bmvit). Die Autoren danken den Projektpartnern Eurofunk Kappacher, Dräger Austria, IFR, FFW Gumpoldskirchen, Karl-Franzens-Universität Graz (Institut für Soziologie) und dem Bundesministerium für Inneres (Sektion I).

Literaturverzeichnis

- [BL12] Bauer, G.; Lukowicz, P.: Can smartphones detect stress-related changes in the behaviour of individuals? In: Proc PERCOM Workshops, 2012.
- [Ca03] Casoni M. et al.: An Ad-hoc Emergency Network for Crisis Events. In: Proc. Future Network and Mobile Summit, 2013.
- [Ca12] Campbell, A.: From Smart to Cognitive Phones. In: IEEE Pervasive Computing, Vol. 11(3), 2012.
- [Ch12] Chipara O. et al.: WIISARD: A Measurement Study of Network Properties and Protocol Reliability during an Emergency Response. In Proc. MobiSys, 2012.
- [Ch13] Chenji, H., et al: DistressNet: A Disaster Response System Providing Constant Availability Cloud-like Services. In: Elsevier AdHoc Networks, Vol. 11(8), 2013.
- [CN08] Chittaro, L.; Nadalutti, D.: Presenting Evacuation Instructions on Mobile Devices by Means of Location-Aware 3D Virtual Environments. In: Proc. Mobile HCI, 2008.
- [De14] Deng, S.; et al: All Your Network Are Belong to Us: A Transport Framework for Mobile Network Selection. In: Proc. ACM HotMobile, 2014.
- [Do15] Dorfinger P. et al.: A Flexible Self-Aligning Communication Solution for Multinational Large Scale Disaster Operations. In Proc. IARIA ICN, 2015.
- [Fe13] Ferrus R. et al.: Public Safety Mobile Broadband: A Techno-Economic Perspective. In: IEEE Vehicular Technology Magazine, Vol. 8(2), 2013.
- [Ge10] George S.M. et al.: DistressNet: A Wireless Ad Hoc and Sensor Network Architecture for Situation Management in Disaster Response. In: IEEE Communications Magazine, Vol. 48(3), 2010.
- [Go12] Gomes, P. et al.: Vital Analysis: Annotating sensed physiological signals with the stress levels of first responders. In: Proc. IEEE EMBC, 2012.
- [Gr14] Grazia, C.A. et al.: Integration between Terrestrial and Satellite Networks: the PPDR-TC Vision. In Proc. IEEE WiMob Workshop EN4PPDR, 2014.
- [GS13] Gardner-Stephen, P. et al.: The serval mesh: A platform for resilient communications in disaster & crisis. In: Proc. IEEE GHTC, 2013.
- [Ha15] Habak, K. et al.: Bandwidth Aggregation Techniques in Heterogeneous Multi-Homed Devices: A Survey. In: arXiv:1309.0542v3, 2015.
- [He14] Hellwagner H.: The BRIDGE Project - Bridging Resources and Agencies in Large-Scale Emergency Management. In: IEEE STCSN E-Letter, Vol. 2(1), 2014.
- [Hi10] Higgins, B.D. et al.: Intentional Networking: Opportunistic Exploitation of Mobile Network Diversity. In: Proc. MobiCom, 2010.
- [Ji11] Jiang, P. et al.: Publish/Subscribe Delay-Tolerant Message-Oriented Middleware for Resilient Communication. In: IEEE Communications Magazine, Vol. 49(9), 2011.
- [Ki14] Kim, J. et al.: Multi-source Multipath HTTP (mHTTP): A Proposal. In Proc. Sigmetrics, 2014.

- [Ku15] Kuhnert M. et al.: Next Generation, Secure Cloud-based Pan-European Information System for Enhanced Disaster Awareness. In: Proc. ISCRAM, 2015.
- [LH08] Lachner J.; Hellwagner H.: Information and Communication Systems for Mobile Emergency Response. In: Springer LNBIP, Vol 5, 2008.
- [LCL16] Lu, Z.; Cao, G.; La Porta, T.: Networking Smartphones for Disaster Recovery. In: Proc. PerCom, 2016.
- [Le11] Legendre, F. et al: 30 years of wireless Ad Hoc networking research: What about humanitarian and disaster relief solutions? What are we still missing? In: Proc. ACWR, 2011.
- [MC13] Martín-Campillo, A. et al.: Evaluating opportunistic networks in disaster scenarios. In: Journal of Network and Computer Applications, Vol. 36(2), 2013.
- [Ng15] Nguyen T.A.B. et al.: Hybrid Communication Architecture for Emergency Response - An Implementation in Firefighter's Use Case. In: Proc. PerCom Workshops, 2015.
- [PP08] Portmann M.; Pirezada A.A.: Wireless Mesh Networks for Public Safety and Crisis Management Applications. In: IEEE Internet Computing, Vol. 12(1), 2008.
- [Ra12] Raiciu, C. et al.: How hard can it be? Designing and implementing a deployable multipath TCP. In: Proc. USENIX NSDI, 2012.
- [Re15] Reina, D.G., et al: A Survey on Multihop Ad Hoc Networks for Disaster Response Scenarios. Intl. Journal of Distributed Sensor Networks, Vol. 2015, 2015.
- [RH13] Raffelsberger, C.; Hellwagner, H.: Overview of Hybrid MANET-DTN Networking and its Potential for Emergency Response Operations. In: ECEASST, Vol. 56, 2013.
- [St11] Stepler, M., et al.: Evolution of TETRA. White Paper, P3 communications GmbH, 2011. URL: <http://www.tetra-applications.com/26973/whitepapers/Evolution%20of%20TETRA.pdf>, letzter Zugriff: 22.06.2016
- [Ta12] Talavera, G. et al.: Protecting Firefighters with Wearable Devices. In: Proc. UCAM, 2012.
- [Ta15] Tang, Z., et al.: An Application Layer Protocol for Energy-Efficient Bandwidth Aggregation with Guaranteed Quality-of-Experience. In: IEEE Trans. On Parallel and Distributed Systems, Vol. 26(6), 2015.
- [Wi05] Wilson, J. et al.: Design of Monocular Head-Mounted Displays for Increased Indoor Firefighting Safety and Efficiency. In: Proc. SPIE, Vol 5800, 2005.

Semi-Automatic Alerts and Notifications for Emergency Services based on Cross-Platform Social Media Data – Evaluation of a Prototype

Christian Reuter, Christoph Amelunxen, Matthias Moi¹

Abstract: The convergence of social networking and mobile media technology is shifting the way how people communicate and gain or share information. People are using social media to a greater extent, also in emergency situations. During disasters throughout the world, such as the 2010 Haiti earthquake, the 2013 European floods, or the terror attacks 2015 in Paris and 2016 in Brussels, this has been illustrated again. Often information about disasters even finds its way faster to social media than it reaches regular news companies and emergency services. However, approaches for processing and analysing the vast quantities of data produced have even more potential. Yet many emergency services still have not found a way to put this potential to an effective use. Within our project EmerGent we are developing a system to process and analyse information from social media particularly tailored for the specific needs of emergency services. The aim is to transform the high volume of noisy data into a low volume of rich content that is useful to emergency personnel. In the first part of this paper we present our approach from a user interface perspective. The second part deals with the evaluation of the approach and the derivation of future potentials of the approach.

Keywords: social media, emergency management, evaluation

1 Introduction

Social media data might be of value for emergency services. However, many studies have identified barriers regarding the high amount of data emergency services need to deal with before, during and after emergencies. A study with 24 district officers in Virginia has demonstrated that the amount of data and the severity to detect or recognise relevant information is a main barrier [Ka11]. A survey with 241 emergency services in the US has shown that information overload is often the case [P115]. Another study in Thailand emphasized ensuing risks of wrong decisions and correction of mistakes [Ka12]. As a conclusion from this, there is a request for proper tools to deal with the high amount of social media data [Re15a], as identified in an interview study on current and potential use of social media in 7 European countries. Based on a survey with 761 emergency service staff across 32 European countries the “majority of emergency

¹ EU-Project EmerGent (www.fp7-emergent.eu)

Matthias Moi, M.Sc. (University of Paderborn; moi@cik.uni-paderborn.de)

Dipl.-Wirt.Ing. Christoph Amelunxen (University of Paderborn; amelunxen@cik.uni-paderborn.de)

Dr. Christian Reuter (University of Siegen; christian.reuter@uni-siegen.de)

services has positive attitudes towards social media”, however it is “more used to share information (44%) than to receive messages (19%)”, while a future increase is expected [Re16, S.3].

In our project the strategy is to process and analyse social media data streams before, during and after emergencies including the steps of (A) information gathering (incl. data preparation), (B) data enrichment, (C) semantic information modelling with ontologies and information mining, (E) information quality and (F) alert detection [Mo15]. The alert detection aims to observe on-going conversations and trying to identify arising alerts from the overall noise. It operates on mined and quality-rated information to detect emergency-related events such as “two injured persons detected” [Mo15]. Emergency services might want to know more upon request. E.g. if they receive an alert, they need the option to investigate the origin or the reason for it. Because of that it has to be possible to visualise the dependencies between alerts and single posts, which initiated an event (temporal representations, graph-based visualisations).

In this paper we present the evaluation of our concept, which has been implemented as a prototype for emergency services, including identified potentials.

2 Related Work

For the realisation of the need for information for decision-makers, the use of social media analytics has to meet specific requirements. As a basis for the comprehensive analysis of social media a sufficient number of platforms has to be connected. Especially in a crisis context, it is not sufficient only to look at microblogging services such as Twitter, emergency services should base their decisions on information from various media [Re15a, RV14]. Social media is also highly heterogeneous, so that the ways of distribution of different types of information varies greatly. An additional factor, which terminated the size of the available data base, is the type of connection. The different platforms offer only limited possibilities to detect an amount of data in a given period [RS14].

The number of different methods of analysis as well as the functional areas, which are covered by this, constitute a separate criterion in selecting appropriate practices for social media analysis. Thus, applications differ for market analysis from those of crisis management on things like focusing on reputation management, marketing and the customer relationship management [St14]. Other methods of analysis point to a general validity for the analysis of social media, such as sentiment analysis [Sh10]. Additionally, there are methods which are especially suitable for the context of crisis management so that an individual assessment of the provided functionality is required.

In addition to automated information review and selection, the possibility of individual weighting and validation by experts is an important point in the use of social media data [Ke10, Re15b]. Basically such systems should be able to adapt to changes in the setting

and to support emergent collaboration, such as ad hoc needed work in unforeseen events [Re14]. Especially in relation to the processing of unforeseen patterns it is necessary to activate this expertise and use. Some approaches try to combine virtual and real activities of volunteers [Re13]. Based on this approaches like XHelp focus on the citizens themselves and the support of their volunteer activities [Re15c], other approaches suggest to allow monitoring their activities [Lu15].

Market surveys reveal that the market for social media intelligence software is very heterogeneous and fast moving, which is also supported by related research [Po13, Tr15]. The reason is that a large number of designed systems often provide identical functionality in analysing social media. Existing approaches focus on the analysis of social media (e.g. Brand Watch, Public Sonar aka Twitcident, Signal Socialmention), provide functionality for organisational interaction and communication in social media (e.g. Coosto, Hootsuite, SproutSocial, TweetDeck) or provide specific functionalities or other approaches like crowdsourcing (e.g. AIDR: Artificial Intelligence for Disaster Response, SensePlace2, Tweedr, TwitInfo, Ushahidi). The transitions between the groups seem fluid and no clear-cut delineation is possible. Systems trying to generate alerts for emergency services have not been detected.

3 Alerts and Notifications: The EmerGent Semi-Automatic Concept for Emergency Services

Following the strategy [Mo15] to process social media data before, during and after emergencies, the visualisation of relevant information was defined as an important part for emergency services in order to handle the vast amount of social media data. In Reuter et al. [Re12] the data generating way of communication is defined as C2A (citizen to authorities), whereas citizens report information related to an incident or emergency. Considering social media this can be done directly or indirectly. For direct communication social apps or direct messaging can be used. For indirect C2A communication citizens do not address messages to concrete recipients (for instance the public timeline on Twitter). To analyse this information, API of different social media have to be used [RS14]. We also plan for A2C communication using a feedback channel allowing answering on posts from social media, e.g. with a warning.

In the requirement analysis, based on several empirical studies with emergency services [Re15a, Re16], we identified that it is not feasible for emergency services to deal with every single message that might be of interested or emergency related [Ak16]. During an emergency thousands of potentially relevant messages may come up and this would lead to an information overload [HP13] again. Information has to be transformed into meaningful and manageable parts.

To target this issue, we introduced the concept of *alerts* and *notifications* [Mo15] with the aim of “transferring high volume, but unclear information content into low volume and rich content suitable for emergency services” (Moi et al 2015., p. 47). An alert is a

set of classified messages sharing a similar context, which is of particular interest for emergency services. The context is defined by but not limited to attributes like date, time, location, full text, identified event types or language. The main aim of the interface is not to present hundreds of messages to the user, but to select relevant messages and to cluster them into alerts. Each notification consists of several messages from Twitter, Facebook, Google+, Instagram or YouTube, as long as they belong to the same setting. To do so, within the information mining process relevant data is filtered, classified and automatically categorised into these alerts and notifications. Furthermore, messages are evaluated for the information quality process to estimate different criteria's like timeliness, understandability, believability or completeness in order to support ES in the evaluation of information [Mo15]. Off course the user also has the possibility to adapt the process with specific key words. Therefore, we call it a semi-automatic progress.

The EmerGent ES-Interface is part of an IT-System (see Figure 1).

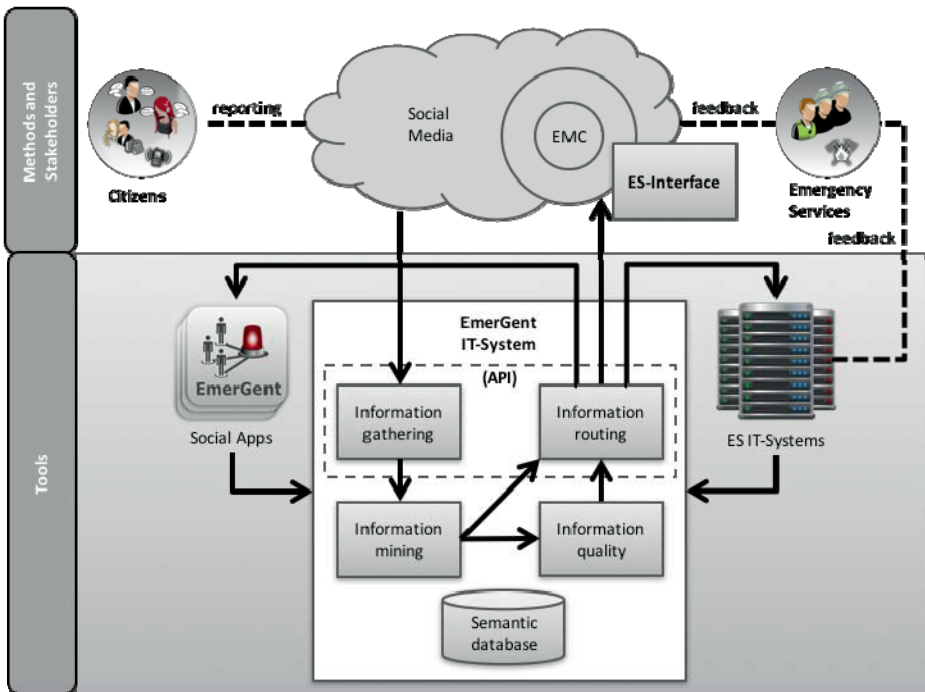
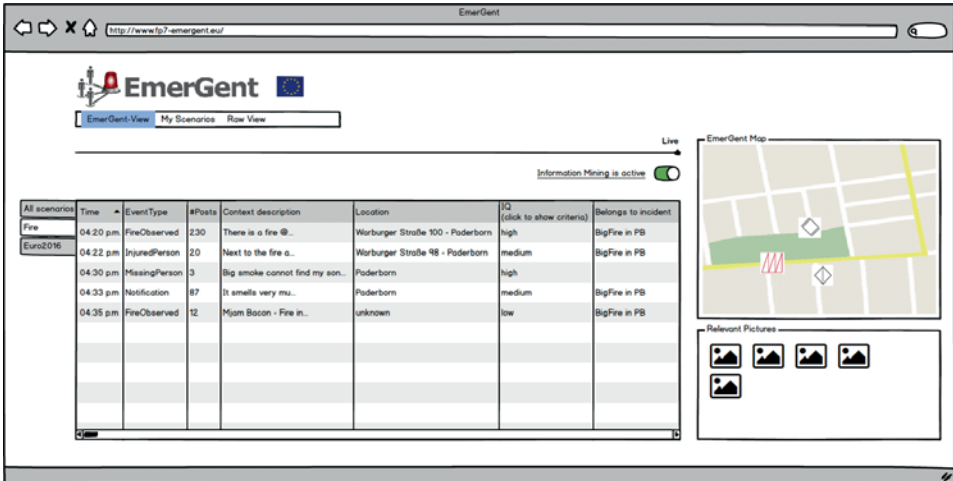


Figure 1 Overview of the EmerGent IT-System

The basic concept of the user interface (UI) is shown in Figure 2 where processed messages are aggregated into alerts and notifications.



Time	Event Type	#Posts	Context description	Location	IQ (click to show criteria)	Belongs to incident
04:20 p.m.	FireObserved	230	There is a fire @...	Warburger Straße 100 - Paderborn	high	BigFire in PB
04:22 p.m.	InjuredPerson	20	Next to the fire a...	Warburger Straße 98 - Paderborn	medium	BigFire in PB
04:30 p.m.	MissingPerson	3	Big smoke cannot find my son...	Paderborn	high	
04:33 p.m.	Notification	87	It smells very mu...	Paderborn	medium	BigFire in PB
04:35 p.m.	FireObserved	12	Mjam Bacon - Fire in...	unknown	low	BigFire in PB

Figure 2 UI concept for visualising alerts & notifications (incl. zoom)

Followed by this concept a first prototype was developed to let emergency services interact with social media in a comfortable way. The relevant parts are summarised in Figure 3. The first functionality is *information gathering*, e.g. to edit your own search criteria based on location and keywords (Examples: you are a commander and need general information of all events (e.g. flood); or: you are a special unit (dike expert) and need special information (e.g. #broken, destroyed, damaged, #dikes, dam, hill, #sandbag)). Based on this, different alerts are displayed in a list (Figure 3, top) as well as on a map (Figure 3, bottom left). Using *filter criteria*, it is possible to filter the notifications/alerts to your specific requirements. The *information quality* is represented by different criteria like timeliness, understandability, believability or completeness (Figure 3, bottom right). A *post popup* includes detailed information and functionality such as sharing, replying and reporting.

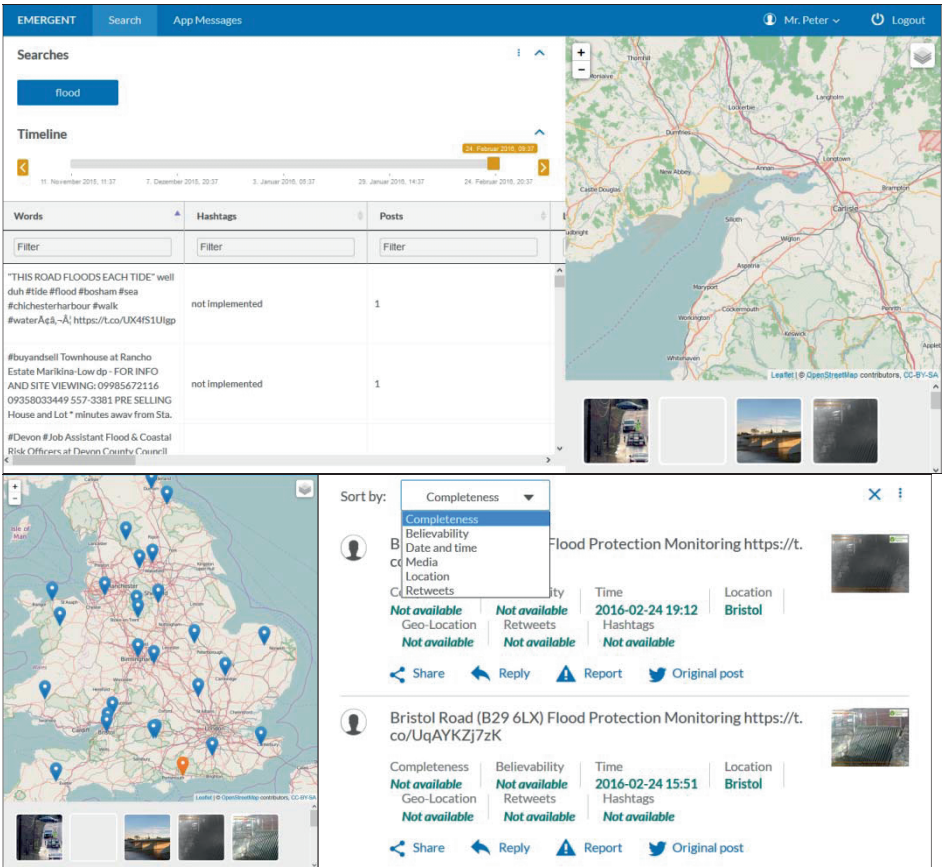


Figure 3: The EmerGent Emergency Service Interface

4 Qualitative Evaluation with Emergency Services

In order to evaluate the concept, we conducted an end user oriented evaluation with the aim to receive feedback from practitioners/emergency services. It followed the structure of a situated evaluation [Tw94] and was carried out in different countries. It is important that we do not focus on feedback such as the change of colour or the size of the system. We mainly wanted to receive feedback on the functionality of the intended system and how it would be of value for their work. The questions were split in two parts: personal details and 10 questions about the system.

4.1 Methodology

Scenarios are recognised as an important strategy towards understanding the interface between the environment and the system as well as a means to elicit and specify software behaviour [Le97]. Broadly, a scenario is an evolving description of situations in the environment. This scenario approach is a part of the evaluation strategy of EmerGent in two different ways. The first one happens in the form of a story about a flood in Germany. The story shows under which real life circumstances social media might provide added value to emergency services and to citizens in distress. The story connects a series of events resulting from the flood regarding the possible uses of social media [Ak16]. The story is a tool to support the interviewer. It helps to introduce the user to the overall situation where the interviewee could use this kind of system. Depending on the individual interviewee, the interviewer decides whether to use the story or not. The second one happens in form of precise examples for every ES-Interface function. It demonstrates a concrete situation in which way this function could be useful. This method builds the system context, which is defined by the context and system boundary [Po09].

The stakeholders are a very important knowledge source to evaluate the system. If important stakeholders are unconsidered, it could have negative effects on the results [Po09]. The above mentioned evaluation was end user oriented, which provides strong basic conditions for the identification of relevant stakeholders. The following stakeholder identification approach is based on an emergency services process analysis. To cover the different ES processes of the countries it is necessary to abstract them. In this case they were related to time- and local-aspects and the rank of the end user. Local-related aspects regard the place and environment where stakeholders work, existing infrastructure and working conditions. This lead to the identification of three working areas: emergency services in the field (e.g. incident commander), an incident support back office (e.g. public-safety answering point (PSAP), incident commander staff) and a general back office or other authorities (e.g. press and public relations). Regarding alerts and notifications two timestamps are relevant: before and after an official emergency call. Before an emergency call one cannot search for a certain event or place. After an emergency call a concrete event in a determined place or area is existent. The rank reflects, in combination with the work years, the experience, education and the perspective to situations. Regarding all these conditions, five relevant roles are identified (Table 1).

These five roles serve as target roles for the selection of the interviewees. In the first evaluation twelve interviews have been conducted within a period of four weeks. All target roles were represented, what was the highest priority in order to get a brought feedback. As the availability of gold level officers is much more limited than silver and bronze, we decided to have them represented in the final evaluation. An overview about the distribution of participants is given in Table2. In total, twelve emergency service staffs (E1-E12) from three countries (Germany, Poland and Slovenia) participated in the

evaluation.

Table 1: identified relevant end-user roles

Role	Command-level			Time-related (emergency-call)		Location-related		
	Gold (strategical)	Silver (tactical)	Bronze (Operational)	Before	After	In the field	Incident support back office	General back office
Emergency service social media manager/team				x	x		x	x
Dispatcher	x	x	x	x	x	x		
Incident commander	x	x	x		x	x		
Incident command staff (special media and press)	x	x	x		x	x	x	
Press and public relations				x	x		x	x

Table 1: Participants

ID	Main role	Command level	Work experience in years	Age
E1	Incident Commander	Silver	Under 5	30-39
E2	Section Leader, Incident Commander	Silver	15+	50-59
E3	Incident Commander, Communication Officer	Silver	15+	40-49
E4	Member of the crew	Silver, Bronze	5-9	20-29
E5	Member of the crew	Silver, Bronze	10-14	40-49
E6	Head/supervisor of organisation	Bronze	15+	50-59
E7	Incident commander	Bronze	5-9	20-29
E8	Head/supervisor of organisation	Bronze	15+	40-49
E9	Member of the crew	Bronze	5-9	30-39
E10	Section leader	Bronze	15+	40-49
E11	Section leader	Bronze	15+	40-49
E12	Member of the crew, PSAP Operator	Silver, Bronze	5-9	30-39

4.2 Results I: Potentials of Alerts and Notifications

First Impression (Q1): Nine of the 12 interviewees liked the presented interface or thought it would be helpful, however three could not see a possible use for it or were not interested. E1 and E3 saw its potential especially for the control centre or the high command, E3 also mentioned the possibility to forward information as very useful. E1

and E8 said that additional information would always be positive. However, the interface still seems to need some improvements as mentioned by E4 and confirmed by E5 with his addition “*when it’s finished*”. This is based on the limitations of the current development. E12 was already satisfied with the current state and thought it was “*simple to understand*”. E1 mentioned “*each additional source of information is helpful, but it has to be considered as a tool for a larger scale operation*”.

Potentials (Q2): Almost all participants were able to identify useful potential uses of the ES-interface. This included forwarding information to the head of operations or the officer-in-charge as well as informing the citizens which was mentioned multiple times by the interviewees. Most of them saw its potential mostly for big scale incidents and in getting information directly from the scene. E1, E10 and E12 also saw potential for identifying risks or emergencies beforehand. “*It could be used to define the development of some event via timeline and geo-location. Therefore, it could be useful to predict the future development of some event – so we can plan ahead with the RSue units’ deployment, logistics*” (E12). The filtering function and location map were also perceived as helpful by E1 and E12. Just one participant could not see any potential – this was the same user (E9) who thought that the interface would not be useful for them overall.

Most Useful Functionality (Q5): Forwarding information was once again mentioned as one of the favourite functionalities by the interviewees (E1, E2, E3, E7). E8 and E10 found the connection of location and incidents also very useful. E6 and E12 mentioned the option to filter the information through own criteria and therefore extracting useful “*information out of thousands of posts*” (E12).

4.3 Results II: Challenges of Social Media in Emergencies

Weaknesses and Risks (Q3): The main concern expressed in nearly every answer was false information and its impact on operations or the people causing panic. E1 and E12 said that the system cannot be accurate a 100%, either missing important information due to language barriers or overreaction due to the used wording. E3 thought the main barrier of using the application would include the needed skills in social media and the pure manpower needed to handle the amount of information. Also E1, E2 and E10 mentioned concerns in terms of privacy and data protection: “*Authorities keep citizens and their social media platforms under surveillance. Citizens could feel monitored*” (E1).

Additional Functionality Needed (Q7): Only four participants provided an answer to this question. E2 said that a better connection is needed to other involved actors such as the police or other emergency services, while E3 suggested that the interface needs to distinguish between private and official accounts. Official ones should be treated as more trustworthy and the interface should also allow the user to identify particular accounts as more trustworthy than others.

Challenges of Social Media in Emergencies (Q8): The interviewees saw false information and possible negative consequences on acting on such false information as the main challenges. This is followed by a need to be able to filter out unimportant and unhelpful information, as well as lacking skills in the handling of social media.

4.4 Results III: Impact on the Work

Work Improvement (Q4): The majority of the participants thought that the application could help them, especially in the use case of information and communication. As the additional text answers also showed, it was particularly seen as potentially helpful for big and mass events. Generally, the opinion seems to be positive towards the interface and its capability of supporting the interviewees' jobs. However, the answers differ in which use case it would help but, with the exception of only one, they all got at least six approvals. **(Q9)** The majority of the participants would use the application for the reasons already mentioned in the previous questions; the main reason for not using it was lack of time or not using social media at all. A suggestion mentions: *"keep it simple on the UI and complex in the backend"* **(Q12)**.

5 Discussion and Conclusion

Social media plays a role in emergency management, however the amount of information sometimes exceeds the ability to work with them. Some tools exist to deal with this amount, although fewer systems were tailored for emergency services who do not have time to work with this system in full time (section 2). Therefore – based on related work – our approach comprises alerts, notifications and grouping of messages from social media with a possible relevance for emergency services. These are generated in a semi-automatic way, which means that the generation progress is adjustable to the individual needs of the officer in charge (section 3). As a contribution of this article, a system, including this functionality, has been evaluated (section 4), detecting the following findings (table 1):

Table 2: Summary of the findings

Aspect	Summary
Potentials	forwarding information to the head of operations, identifying risks, filtering using own criteria – precise information (notifications) are needed (Q2, Q5), especially in mass events (Q4), semi-automatic approaches can help
Weaknesses	false information (Q3) and negative consequences (Q8)
Final remarks	"keep it simple on the UI and complex in the backend" (Q10)

The majority of the participants expressed a positive attitude towards the ES-interface and they see potential for supporting them in their work. Only a few did not think it would be helpful, mostly because they do not have the time to use it or are not familiar with social media at all. Interviewees thought that being able to forward social media information to others is one of the most useful features and they wanted the interface to allow them to communicate with other services and organisations involved in crisis situations. Another positively perceived feature is filtering: it helps ES to deal with the massive amount of posts and therefore allows focusing on the helpful information only. On the other hand, participants thought that this might also be problematic, as an automatic system cannot replace the human expertise and might therefore lead to missing important information. In this context, E1 and E2 recommend to check the EmerGent information in the back office (e.g. public-safety answering point) by using humans and to share relevant information with the incident commander. However, EmerGent notifications are meant to do this in a semi-automatic way. Unfortunately, most participants were not able to answer the question about the EmerGent notification as it was impossible to demonstrate this in the current version of the interface.

The main challenge and problem of social media use in general, in emergency services and in crisis situations at all, is the possibility of false information and how this could lead to wrong decisions during a crisis. Interviewees were worried that false information or rumours shared via social media could cause panic – this was one of the main reasons why many were reluctant to rely on social media during emergencies. In future work we aim to improve our system in order to detect alerts even more appropriate and to address the weaknesses.

Acknowledgements

The research project EmerGent was funded by a grant of the European Union (FP7 No. 608352). This article is based on Deliverable 3.7 “Potentials of Social Media Usage by EMS and citizens’ involvement in the EMC enabled by EmerGent”.

References

- [Ak16] Akerkar, Rajendra; Friberg, Therese; Amelunxen, Christoph: EmerGent Deliverable 3.5: User Requirements, Version 2. Paderborn, 2016
- [HP13] Hiltz, SR; Plotnick, Linda: Dealing with Information Overload When Using Social Media for Emergency Management: Emerging Solutions. In: Proceedings of the Information Systems for Crisis Response and Management (ISCRAM). Baden-Baden, Germany, 2013, S. 823–827
- [Ka12] Kaewkitipong, Laddawan; Chen, Charlie; Ractham, Peter: Lessons Learned from the Use of Social Media in Combating a Crisis: A Case Study of 2011 Thailand Flooding Disaster. In: Proceedings of the International Conference on Information Systems (ICIS).

Orlando, USA, 2012, S. 1–17

- [Ka11] Kavanaugh, Andrea; Fox, Edward A; Sheetz, Steven; Yang, Seungwon; Li, Lin Tzy; Whalen, Travis F.; Shoemaker, Donald; Natsev, Paul; u. a.: Social Media Use by Government: From the Routine to the Critical. In: Proceedings of the International Digital Government Research Conference. Maryland, USA, 2011.
- [Ke10] Keim, Daniel; Mansmann, Florian; Thomas, Jim; Keim, Daniel: Visual Analytics : How Much Visualization and How Much Analytics? In: ACM SIGKDD Explorations Newsletter Bd. 11 (2010), Nr. 2, S. 5–8 — ISBN 1931-0145
- [Le97] Leite, J C S P; Rossi, G; Balaguer, F; Maiorana, V; Kaplan, G; Hadad, G; Oliveros, A: Enhancing a requirements baseline with scenarios. In: Requirements Engineering, 1997., Proceedings of the Third IEEE International Symposium on, 1997, S. 44–53
- [Lu15] Ludwig, Thomas; Reuter, Christian; Siebigteroth, Tim; Pipek, Volkmar: CrowdMonitor: Mobile Crowd Sensing for Assessing Physical and Digital Activities of Citizens during Emergencies. In: CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. Seoul, Korea : ACM Press, 2015, S. 4083–4092
- [Mo15] Moi, Matthias; Friberg, Therese; Marterer, Robin; Reuter, Christian; Ludwig, Thomas; Markham, Deborah; Hewlett, Mike; Muddiman, Andrew: Strategy for Processing and Analyzing Social Media Data Streams in Emergencies. In: Proceedings of the International Conference on Information and Communication Technologies for Disaster Management (ICT-DM). Brest, France, 2015
- [Pl15] Plotnick, Linda; Hiltz, Starr Roxanne; Kushma, Jane a; Tapia, Andrea: Red Tape : Attitudes and Issues Related to Use of Social Media by U . S . County- Level Emergency Managers. In: Proceedings of the Information Systems for Crisis Response and Management (ISCRAM). Kristiansand, Norway, 2015
- [Po09] Po, Klaus: Basiswissen requirements engineering. Aus- und Weiterbildung zum Certified Professional for Requirements Engineering : dpunkt, 2009
- [Po13] Po, Daniela: Social Media Analysis for Crisis Management: A Brief Survey. URL <http://stcsn.ieee.net/e-letter/vol-2-no-1/social-media-analysis-for-crisis-management-a-brief-survey>.
- [Re12] Reuter, Christian; Marx, Alexandra; Pipek, Volkmar: Crisis Management 2.0: Towards a Systematization of Social Software Use in Crisis Situations. In: International Journal of Information Systems for Crisis Response and Management (IJISCRAM) Bd. 4 (2012), Nr. 1, S. 1–16
- [RS14] Reuter, Christian; Scholl, Simon: Technical Limitations for Designing Applications for Social Media. In: Koch, M.; Butz, A.; Schlichter, J. (Hrsg.): Mensch & Computer 2014: Workshopband. München, Germany : Oldenbourg-Verlag, 2014, S. 131–140
- [Re13] Reuter, Christian; Heger, Oliver; Pipek, Volkmar: Combining Real and Virtual Volunteers through Social Media. In: Comes, T.; Fiedrich, F.; Fortier, S.; Geldermann, J.; Müller, T. (Hrsg.): Proceedings of the Information Systems for Crisis Response and Management (ISCRAM). Baden-Baden, Germany, 2013, S. 1–10
- [Re14] Reuter, Christian: Emergent Collaboration Infrastructures: Technology Design for Inter-Organizational Crisis Management (Ph.D. Thesis). Siegen, Germany : Springer Gabler,

2014 — ISBN 978-3-658-08585-8

- [Re15a] Reuter, Christian; Ludwig, Thomas; Friberg, Therese; Pratzler-Wanczura, Sylvia; Gizikis, Alexis: Social Media and Emergency Services? Interview Study on Current and Potential Use in 7 European Countries. In: International Journal of Information Systems for Crisis Response and Management (IJISCRAM) Bd. 7 (2015), Nr. 2
- [Re15b] Reuter, Christian; Ludwig, Thomas; Ritzkatis, Michael; Pipek, Volkmar: Social-QAS: Tailorable Quality Assessment Service for Social Media Content. In: Proceedings of the International Symposium on End-User Development (IS-EUD). Lecture Notes in Computer Science, 2015
- [Re15c] Reuter, Christian; Ludwig, Thomas; Kaufhold, Marc-André; Pipek, Volkmar: XHELP: Design of a Cross-Platform Social-Media Application to Support Volunteer Moderators in Disasters. In: Proceedings of the Conference on Human Factors in Computing Systems (CHI). Seoul, Korea : ACM Press, 2015
- [Re16] Reuter, Christian; Ludwig, Thomas; Kaufhold, Marc-André; Spielhofer, Thomas: Emergency Services Attitudes towards Social Media: A Quantitative and Qualitative Survey across Europe. In: International Journal on Human-Computer Studies (IJHCS) (2016)
- [RV14] Ruggiero, Aino; Vos, Marita: Social Media Monitoring for Crisis Communication: Process, Methods and Trends in the Scientific Literature. In: Online Journal of Communication and Media Technologies Bd. 4 (2014), Nr. 1, S. 105–130
- [St14] Stieglitz, Stefan; Dang-Xuan, Linh; Bruns, Axel; Neuberger, Christoph: Socialmedia analytics. In: Business and Information Systems Engineering Bd. 6 (2014), Nr. 2, S. 89–96
- [Sh10] Sheth, Amit; Purohit, H; Jadhav, A; Kapanipathi, P; Chen, L: Understanding events through analysis of social media. In: The Ohio Center of Excellence in Knowledge-Enabled Computing (Kno.e.sis) (2010), S. 1–4
- [Tr15] Trateral Research & Consulting: Comparative Review of Social Media Analysis Tools for Preparedness. URL http://Trateralresearch.com/wp-content/uploads/2015/08/GDPC_SMAT_Short-Report-for-GDPC_Final.pdf
- [Tw94] Twidale, Michael; Randall, David; Bentley, Richard: Situated evaluation for cooperative systems Situated evaluation for cooperative systems. Lancaster, UK, 1994

Appendix: Evaluation Questions

Part I: Personal Details

1. What type of organisation do you work for?
2. What is your main role (the role you spend most time on) in this organisation?
3. What is your command level?
4. How many years have you been working for Emergency Services?
5. What is your age?

6. What is your gender?
7. What country do you live in?

Part II: 10 Questions on the Application

1. What is your first impression?
2. What potentials does the application involve?
3. What weaknesses and risks does the application have? How could they be improved?
4. Would a final version of the application help you doing your job better? In which use cases (not)?
5. What functionality of the application do you find most useful?
6. What is your evaluation of the “EmerGent Notifications” (summary of social media posts) and of the “Information Quality”?
7. Is there any additional functionality you would like the application to have?
8. What are the greatest challenges of using social media in emergencies?
9. If the application was operationally available, would you use it? Why (not)?
10. Do you have final remarks?

Softwarearchitekturen und –entwicklung für Smartwatch und Wearable Apps

Vorwort: Softwarearchitekturen und –entwicklung für Smartwatch und Wearable Apps

Klemens Waldhör¹, Rainer Lutze²

1 Workshop Inhalte und Ziele

Smartwatches und Wearables gewinnen, als voraussichtliche Endgerätegeneration »nach« den Smartphones, zunehmend an Aufmerksamkeit. Selbst wenn die Hardware der heute verfügbaren Geräte noch wahrnehmbare Funktions-, Leistungs- und Handhabungsdefizite aufweist, die einer breiten Verwendung entgegenstehen, ist jetzt schon erkennbar, dass diese Gerätegeneration spezifische, neuartige Anforderungen an die Software mit sich bringt. Aufgrund der größtenbedingten sehr begrenzten Batterieleistung dieser Geräte wird ein hocheffizientes, Software gesteuertes Energiemanagement dauerhaft eine zentrale Bedeutung in der App Entwicklung einnehmen. Dies erfordert die geeignete Verteilung der Funktionalitäten einer Anwendung zwischen Wearable und einer leistungsstarken Hintergrundplattform (Server, Smartphone, ...). Unterschiedliche Betriebssysteme und Marktplätze der Marktführer Google mit Android Wear, Apple IOS mit Swift, aber auch Samsung mit Tizen und LG mit Web OS erfordern jeweils unterschiedliche Herangehensweisen und erschweren heute eine plattformübergreifende, systematische Entwicklung, Distribution und Qualitätssicherung solcher Apps. Durch den Einsatz der spezifischen Smartwatch Sensorik, wie sie im klassischen PC Bereich nicht vorhanden sind, ergeben sich einerseits neue Anwendungsmöglichkeiten, aber auch neue softwaretechnische Herausforderungen, wie diese Sensoren geeignet in Apps integriert und die hierauf aufbauenden Softwarefunktionen umfassend und systematisch getestet werden können (z.B. Kombination aus Bluetooth, Wifi, GPS, Mobilfunk). Eine zusätzliche Komplexitätsdimension gewinnt die Softwareentwicklung durch die Sachlage, dass das Anforderungsprofil an B2B Software Apps eine hohe Volatilität aufweist und sich nahezu täglich weiterentwickelt und verändert. Dies wirft die Frage nach einem hierzu einem passenden Anforderungsmanagement und Verfahren für eine wirtschaftliche Wartung und Pflege der Software auf. Natürlich beinhalten die neuen Möglichkeiten der Wearables zusätzlich eine Vielzahl ethischer und juristischer Aspekte, wie die aktuelle Diskussion um den Schutz der Privatsphäre bei kamerabestückten Datenbrillen und die Nutzung von Wearable Informationen im Bereich der Gesundheitswirtschaft zeigt – diese Aspekte sind aber nicht Gegenstand des Workshops.

Ziel des Workshops ist es, folgende Fragen zu diskutieren und zu vertiefen:

¹ **FOM** University of Applied Sciences, D45127 Essen / Nuremberg, Germany
klemens.waldhoer@fom.de

² **Dr.-Ing. Rainer Lutze Consulting**, Wachtlerhof, Wilhermsdorfer Str. 14, D90579 Langenzenn, Germany
rainer.lutze@lustcon.eu

- Softwarearchitekturen, -verteilung: welche Softwarearchitekturen und Verteilungskonzepte passen für Smartwatches am besten? Wie kann eine geeignete Softwareverteilung zwischen Wearable/Smartwatch und leistungsstarker Hintergrundplattform oder Cloud die spezifischen Einschränkungen der Endgeräte wie Batterielaufzeit, beschränkte Rechenkapazität und Verarbeitungsparallelität berücksichtigen? Wie können die Daten zwischen Endgerät und Hintergrundplattform, Cloud effizient und verfälschungs- bzw. abhörsicher übertragen werden?
- Software-Engineering und Qualität, Projektmanagement: was sind die besten Prozesse bzw. Vorgehensmodelle bei Requirements Engineering und der Entwicklung von (plattformübergreifenden) Smartwatch Apps? Welche Projektmanagementmethoden und -prozesse (agile Methoden) eignen sich besonders für die Entwicklung von Smartwatch Apps? Wie werden App Distributionsprozesse eingebunden und gesteuert? Wie kann trotz hoher Änderungsraten und Updates die Qualität in Hinblick auf Stabilität und Sicherheit gewährleistet werden?
- Softwareentwicklung: welche Ansätze gibt es im Crossplattform-Bereich zur Entwicklung von Smartwatch Apps? Welches der vorhandenen Programmiermodelle ist am flexibelsten, am schnellsten erlernbar? Wie können Benutzer geeignet den Entwicklungsprozess eingebunden werden?
- GUI – Entwicklung und Design: was sind die spezifischen Herausforderungen bei der Gestaltung von Benutzeroberflächen für Smartwatches unter Berücksichtigung der minimalen Displaygrößen? Wie können zusätzliche Interaktionsmöglichkeiten wie Touchscreens, Knöpfe, Kronen/Lünetten, Gesten, Sprachein- und -ausgabe für den Benutzer kognitiv effizient eingesetzt werden? Wie werden die unterschiedliche Interaktionsmöglichkeiten – visuell, taktil, haptisch, aktustisch am besten und in welchen Situationen kombiniert? Gibt es barrierefreie Smartwach Apps und durch welche Merkmale zeichnen sie sich aus?
- Spezifische Anwendungsbereiche: welche Anwendungsfelder bieten sich besonders für Smartwatches an, insbesondere im Gesundheits-, Fitness-, Kommunikationsbereich? Was sind die spezifischen Anforderungen auf diesen Gebieten?

2 Beiträge

Im Beitrag „An Extendible Communication as a Service Platform for Wearables and Future-Oriented Devices“ von Englert und Muschiol et. Al wird eine Communication-as-a-Service (CaaS) Plattform für Wearables vorgestellt und evaluiert. Die CaaS-Plattform dient Wearables für die Kommunikation und Contentverwaltung. Die Architektur und die Implementierung der Plattform werden beschrieben. Für die Evaluierung werden 30 reale Probanden genommen, die Kommunikationsaufgaben durchführen und deren Durchführungszeiten gemessen werden. Die Zeitmessungen werden verglichen mit Messungen ohne die Unterstützung der CaaS-Plattform. Dieser Vergleich wird mit einem statistischen Divergenzmaß, der Kullback-Leibler-Divergenz, durchgeführt. Die

Evaluierung zeigt, daß die Nutzung der CaaS-Plattform für Kommunikationsaufgaben mit Wearables eine Effizienzsteigerung von 2,58 ermöglicht.

Der Beitrag „xHealth: Eine MQTT und REST basierte Architektur zum Zugriff auf Sensordaten smarterer Objekte“ (Guttenberger, Waldhör) stellt ein Konzept sowie eine prototypische Implementierung vor, in welcher Sensordaten smarterer Objekte - wie einer Smartwatch - an einen zentralen Service (Broker) übermittelt und mittels Data Mining analysiert werden können. Die Übertragung großer Mengen von Sensordaten an zentrale Services ist insbesondere für Anwendungen im Bereich Smartwatch interessiert, wenn Auswertungen auf der Uhr nicht effizient und performant ausgeführt werden können. Über ein effizientes Protokoll kann diese Aufgabe über einen Broker an andere Systeme delegiert, ausgewertet und das Ergebnis an den Client zurückgemeldet werden. Der Beitrag erläutert dies an Hand zweier Szenarien, einer beispielhaften Implementierung zur Übertragung von Daten einer Wetterstation und bei der Verwendung zur Erkennung von Aktivitäten des täglichen Lebens, z.B. um zu erkennen, ob ältere Personen genug Flüssigkeit zu sich nehmen oder gestürzt sind.

Im Beitrag „The Application Architecture of Smartwatch Apps – Analysis, Principles of Design and Organization“ (Lutze, Waldhör) beschreiben die beiden Autoren, wie Smartwatch-Anwendungen systematisch vom Entwurf bis zur Implementierung entwickelt werden sollen. Die Entwicklung von Smartwatch-Anwendungen weist inhärente Besonderheiten auf, u.a. durch die Größe der Uhr, die eine besondere Herausforderung für das GUI-Design erfordern, aber auch das Energie-Management erfordert besondere Aufmerksamkeit. Dazu schlagen die Autoren fünf Entwurfsprinzipien vor: die Notwendigkeit 1) eines effizienten Dialogmanagements, 2) der Einsatz einer verteilten Softwarearchitektur, 3) der Aufbau der App basierend auf einer Drei Schichten Architektur, 4) Berücksichtigung der Software-Engineering Implikationen zur ökonomischen Wartung der Apps, sowie 5) der Berücksichtigung der technischen Einbindung der Smartwatch in ihrer Umgebung. Die Autoren diskutieren diesen Ansatz an einem praktischen Beispiel der Entwicklung einer Smartwatch-App im Gesundheitsbereich.

Die situationsbewusste Adaption interaktiver Anwendungen an den mentalen und emotionalen Benutzer-Zustand dürfte zu einem wichtigen Softwareentwicklungs-Paradigma werden. Der Beitrag von Christian Martin, Christian Herdin und Sanim Rashid von der Forschungsgruppe Automation in Usability Engineering (AUE) der Hochschule Augsburg stellt eine integrierte Test- und Evaluationsumgebung zur Situationsanalyse vor, die es ermöglicht, Benutzer durch Eye- und Gaze-Tracking sowie emotionserkennende FaceReader-Software während ihrer Aktivitäten zu beobachten. Eine Schnittstelle zu Wearables wie dem Empatica-E4-Wristband erlaubt die Echtzeitauswertung biologischer, für Emotionen relevanter Signalfolgen und ergänzt die Entscheidungsbasis zur Auswahl von Software-Artefakten aus einem Pattern-Repository zur dynamischen Anpassung der Zielsoftware.

Bienhaus beschreibt in seinem Beitrag, wie durch die aktuelle Fitnessbewegung Smartwatches, Lifetracker und Wearables eine rasante Entwicklung erfahren haben. Diese neuen Geräte zur Selbstvermessung bieten Potenziale bei der Digitalisierung im

Healthcare-Bereich. Wearables und Smartphones werden bereits bei mobilen E-Health-Lösungen (M-Health) eingesetzt. Dieser Beitrag geht auf Chancen und Herausforderungen für solche Einsatzszenarien ein. Architekturen und Protokolle werden vorgestellt. Die praktische Umsetzung wird am Beispiel einer Notfall-App auf einer Smartwatch und angeschlossenem Hausnotruf demonstriert.

3 Programmkomitee und Organisation

Das Programmkomitee setzte sich wie folgt zusammen:

- MSc. Lukas Alperowitz, Technische Universität München
- Prof. Dr. Wilhelm Hasselbring, Kiel University
- Prof. Dr. Peter Hoffmann, FOM Hochschule für Oekonomie & Management
- Dr. Rainer Lutze, Dr.-Ing. Rainer Lutze Consulting
- Prof. Dr. Christian Märtin, Hochschule Augsburg
- Prof. Dr. Rene Mayrhofer, Johannes-Kepler-Universität Linz
- Prof. Dr. Klemens Waldhör, FOM Hochschule für Oekonomie & Management (Vorsitzender des Programmkomitees)

Der Workshop wurde organisiert von:

Prof. Dr. Klemens Waldhör
FOM Hochschule für Oekonomie &
Management gemeinnützige GmbH
Hochschulstudienzentrum Nürnberg
klemens.waldhoer@fom.de

Dr. Rainer Lutze
Dr.-Ing. Rainer Lutze Consulting
- Strategie- und Technologiebera-
tung, Langenzenn
info@lustcon.de

Smartwatch und Wearables im Gesundheitsbereich: Grundlagen und Anwendungen

Diethelm Bienhaus¹

Abstract: Durch die aktuelle Fitnessbewegung haben Smartwatches, Lifetracker und Wearables eine rasante Entwicklung erfahren. Eine Vielzahl von Vitaldaten lassen sich mit diesen Geräten erfassen und Apps oder Browser-basierte Anwendungen geben dem Nutzer vielfältige Auswertungen. Diese neuen Geräte zur Selbstvermessung bieten Potenziale bei der Digitalisierung des Gesundheitsbereichs. Wearables und Smartphones werden bereits bei mobilen E-Health-Lösungen (M-Health) eingesetzt. Dieser Beitrag geht auf Chancen und Herausforderungen für solche Einsatzszenarien ein. Architekturen und Protokolle werden vorgestellt. Die praktische Umsetzung auf Basis einer offenen Plattform und Architektur wird am Beispiel einer Notfall-App auf einer Smartwatch demonstriert.

Keywords: Wearables, E-Health, M-Health, Ambient Assisted Living, Internet of Things, MQTT, REST, CoAP

1 Motivation

„In the postmobile world the focus shifts to the mobile user who is surrounded by a mesh of devices extending well beyond traditional mobile devices.“ so David W. Cearley, Vice President Gartner Research. In dieser „postmobilen“ Welt ist der Mensch von einem Netz von Smartwatches, Lifetracker und Wearables umgeben, die eine Vielzahl von Daten erfassen und den Nutzer auf unterschiedlichste Weise unterstützen sollen.

In der Gartner-Studie ([CWB15]) wird aufgrund des technischen Fortschritts in der Sensorik ein großes Potenzial für Gesundheitsanwendungen prognostiziert.

Die Digitalisierung wird im Gesundheitswesen bereits länger diskutiert. Durch die rasante Entwicklung von Wearables im Fitness Bereich - Stichwort Quantified Self - ergeben sich Chancen und Impulse für elektronische Anwendungen im Gesundheitsbereich. In Ländern wie den USA werden Wearables bereits für das Remote Monitoring von Patienten eingesetzt.

Die Digitalisierung des Gesundheitsbereichs umfasst im Wesentlichen „eHealth (also die Anwendungen elektronischer Geräte zur medizinischen Versorgung und Wahrnehmung anderer Aufgaben im Gesundheitswesen), mHealth (mobile eHealth-Lösungen) und Telemedizin (den professionellen Mediziner vorbehalten)“([GSMW16])

¹ Prof. Dr.-Ing. Diethelm Bienhaus, Technische Hochschule Mittelhessen, FB Mathematik, Naturwissenschaften und Informatik, Wiesenstraße 14, 35390 Gießen, diethelm.bienhaus@mni.thm.de

Laut einer im Oktober 2015 von TNS Emnid durchgeführten Umfrage sind Dreiviertel der Menschen in Deutschland offen für Wearables und nutzen diese bereits oder befürworten die entsprechenden Technologien.

Wichtig für die Akzeptanz solcher Techniken ist ein spürbarer Verbrauchernutzen. Anstelle des schriftlichen Dokumentierens von Blutzuckerwerten oder anderen Blutwerten in Tagebüchern könnten die Daten über vernetzte mobile Geräte direkt an den Arzt übertragen werden. Solche Apps sind bereits seit einigen Jahren verfügbar. Beispielsweise sei hier die Telemetrieplattform „smart medication“ genannt. Diese Lösung unterstützt Patienten mit angeborener Hämophilie A und B, die weit entfernt vom betreuenden Hämophiliezentrum wohnen, bei der ärztlich kontrollierten Heimselbstbehandlung. Plattform und App wurden vom Institut für Wirtschaftsinformatik der Universität Marburg gemeinsam mit der IT-Firma Rosch & Associates, Frankfurt entwickelt (vgl. [Me13]).

„Durch neue Techniken der Selbstvermessung (Apps, Wearables) besteht das Potenzial für ein kontinuierliches, hochaufgelöstes Bild des Individuums, bei dem Überschreitungen von individuellen Grenzwerten frühzeitig, unabhängig von einem Arztbesuch, erkannt werden und präventiv Verhalten geädert werden kann.“ ([GSMW16])

Bei chronischen Patienten können mittels Wearables und Smartphone erhobene Daten zur Verbesserung von Therapie und Krankheitsverlauf beitragen ([AH16]). Diabetikern und Patienten mit Bluthochdruck können so digitale Logbücher der relevanten Werte auf bequeme Weise erstellen lassen. Auf Basis der Messwert-Historie ist es dann für den Arzt leichter, eine Therapie abzuleiten.

Ein anderes Modell verfolgen Krankenkassen, die teilweise Wearables kostenlos den Mitgliedern zur Verfügung stellen: Mit Rückzahlungen oder Boni sollen Kundenanreize für eine gesunde Lebensführung geschaffen werden. Hier stellen sich unmittelbar Fragen nach Datenschutz und Datensicherheit.

Ziel deutscher Datenschützer ist es, für den Markt der Gesundheits-Apps und Wearables einen besseren Datenschutz zu erreichen. Hierzu sind gesetzliche Rahmenbedingungen notwendig. In dem am 4. Dezember 2015 im Bundestag beschlossenen „Gesetz für sichere digitale Kommunikation und Anwendungen im Gesundheitswesen“ (E-Health-Gesetz) wird auf den Einsatz von Wearables eingegangen. Im sog. Patientenfach, einem Teil der Online-Anwendung der elektronischen Gesundheitskarte, sollen eigene Daten z.B. ein Patiententagebuch über Blutzuckermessungen oder Daten von Wearables und Fitnessarmbändern, abgelegt werden können. Bis Ende 2018 sollen die notwendigen Voraussetzungen dazu geschaffen werden.

1.1 Lokation und Zugriff auf die gesammelten Daten

Die Marktführer für Fitnesstracker setzen eine zwangsweise Anmeldung des Nutzers an die Cloud-Applikationen des jeweiligen Herstellers für eine Synchronisation und Auswertung der erfassten Daten voraus. Eine solche Synchronisation der Fitnessdaten aus dem Tracker ist praktisch. Vergleiche von Ergebnissen innerhalb einer Nutzergruppe mit dem Hintergrund einer durch die Gruppe verstärkter Motivation sind umgesetzt. Einen anderen Weg an die Daten zu gelangen als über die Cloud des Herstellers ist von diesen nicht gewünscht. Ohne Cloud können die Daten mittels der passenden App, die beim Hersteller mit einem persönlichen Profil zu registrieren ist, abgerufen werden.

Ein Hersteller ohne Datensammelwut ist Beurer, der eine App für seinen Fitnesstracker zur Verfügung stellt, die auch ohne Zwangsregistrierung genutzt werden kann. Um Daten aus Apps an eigene Applikationen weiterzuleiten ist eine aufwändige Eigenentwicklung notwendig. Die Geschäftsmodelle des Fitnessbereichs sind für den Einsatz solcher Geräte im M-Health Sektor hinderlich. Um Wearables hier einsetzen zu können sind offene Schnittstellen und Lösungen für die im folgenden genannten Herausforderungen notwendig.

1.2 Herausforderungen des Einsatzes von Wearables im E- und M-Health

Für den Einsatz von Wearables im Gesundheitsbereich ergeben sich Herausforderungen, wie sie im Fitnesskontext so nicht gegeben sind bzw. nicht als Herausforderung unmittelbar wahrgenommen werden.

Datenschutz und Datensicherheit

Was geschieht mit den erfassten Daten? Wo werden diese gespeichert? Wer hat darauf Zugriff? Diese Fragestellungen gehören zu den großen Problemen beim der Nutzung von Wearables insbesondere für M-Health Anwendungen. Viele Hersteller nutzen die erhobenen Daten auch für Werbezwecke. Versicherungen locken beispielsweise mit reduzierten Beiträgen bei einem "gesunden Lebenswandel", der durch die Wearables erfasst werden soll. Datenschutz und Datensicherheit sind bei Wearables ein blinder Fleck. Der Sicherheitsanbieter Kaspersky Lab konnte mit einer speziellen nicht autorisierten App eine Verbindung mit Fitness-Geräten verschiedener Anbieter herstellen und so an persönliche Daten gelangen.

Validität der Daten

Beim Erfassen von Vitaldaten mittels Wearables oder Smartphones zeigen sich relativ große Schwankungen. Es wird teilweise von einer Schwankungsbreite von bis zu 20% ausgegangen. Damit sind so erfasste Werte kaum für M-Health Aufgaben geeignet.

Interoperabilität und Standardisierung

Fitnessbänder benötigen ein weiteres Gerät, um die Daten an Applikationen, die mit wenigen Ausnahmen, in der Cloud des Herstellers ablaufen, weiterzuleiten. Als Gateway dient dabei meist ein Smartphone mit einer passenden App. Alternativ bieten Hersteller auch Anwendungen für PCs/Notebooks an, die die Daten des Bandes auslesen und an die Cloud-Anwendung übertragen. Problematisch ist dabei aus Datenschutzsicht, dass sich die Daten und Analysetools auf Cloud Servern der Hersteller befinden.

Diese bieten meist nur Web-Schnittstellen für die Daten in der Cloud an. Eine Offenlegung der API der Funkschnittstelle ist nicht gewünscht, wie der Autor aus eigenen Projekten erfahren musste ([K116]).

Neben der unzureichenden Interoperabilität ist die fehlende Standardisierung der Formate, in denen die erfassten Daten ausgegeben und dargestellt werden, eine wesentliche Hürde für den Einsatz im Gesundheitsbereich.

Standardisierte und offene Schnittstellen sind eine notwendige Eigenschaft von Wearables und Sensorgeräten für den Einsatz in M-Health Anwendungen. Dies steht allerdings im Widerspruch zu den Interessen der meisten Hersteller solcher Geräte. Eine Nutzung von solchen Geräten im M-Health Bereich ist daher aktuell noch als sehr problematisch einzustufen.

Vollständigkeit der Daten

Wearables wie Fitnessbänder oder Brustgurte erfassen vor allem Werte wie Puls- bzw. Herzfrequenz, Atemfrequenz, Beschleunigung, EKG-Daten, Sauerstoffsättigung oder Hauttemperatur. Daraus abgeleitet werden Werte wie die zurückgelegten Schritte, Schlafphasen und -dauer oder verbrannte Kalorien.

Diese Daten sind für die Bewertung von Fitness konzipiert. Für M-Health Anwendungen sind weitergehende Werte anzustreben wie z.B. die Blutzuckerkonzentration. Mittels solcher Daten könnte einer großen Anzahl von Betroffenen geholfen werden.

2 Architekturen und Protokolle für M-Health Anwendungen

2.1 RESTful Architekturen

Der Durchbruch des Internets kam mit der Entwicklung des Hypertext Transfer Protocols HTTP. Roy Fielding formalisierte die Prinzipien dieses Protokolls und stellte in seiner Dissertation [Fi00] den *Representational State Transfer* (REST) als Protokoll-Stil vor. RESTful Systeme sind leichtgewichtig und Ressourcen orientiert.

Als Gegensatz dazu können Service orientierte Architekturen angesehen werden. Bei diesen nutzen Clients Services von Dienste anbietenden Servern. Mechanismen zum

Auffinden und Beschreiben der Services gehören ebenso wie spezielle Protokolle (SOAP) zu diesem Ansatz. Die Kommunikation erfolgt in Sessions und Server müssen den Session State verwalten.

Für M-Health Anwendungen ist der schwergewichtige SOA-Ansatz weniger geeignet. RESTful Architekturen bieten hingegen folgende Vorteile:

- **Skalierbarkeit und Einfachheit:** Der Zugriff auf Daten folgt der Hierarchie eines Ressourcenbaums.
- **Leichtgewichtige Elemente und menschenlesbare Ergebnisse:** Im Gegensatz zu entfernten Prozeduraufrufen mit verbundenem Kommunikationsoverhead für Aufruf und Rückgabe sind lediglich wenige Methoden (GET, PUT, ...) für die Abfrage einer Ressource notwendig. Das Ergebnis kann dann im XML-Format zurückgegeben werden, so dass es menschenlesbar und maschinell bearbeitbar ist.
- **Lose Kopplung und Resilienz:** Client und Server sind nur lose gekoppelt. Client-Code ist nicht von einer spezifischen Aufrufsyntax und dem Aufrechterhalten der Verbindung abhängig. Da kein gemeinsamer Zustand vorhanden ist, kann nach einem Ausfall und Neustart sowohl von Client als auch von Server weiter kommuniziert werden.

2.2 IoT-Protokolle für M-Health Anwendungen

Für das Internet der Dinge (Internet of Things, IoT) sind einige auf die Anforderungen von kleinen Prozessoren mit geringerer Rechenleistung und verfügbarem Speicher angepasste Protokolle verfügbar. Weite Verbreitung finden im IoT die Protokolle *Message Queuing Telemetry Transport* (MQTT, vgl. u.a.: [Ed14]) und *Constrained Application Protocol* (CoAP, vgl. u.a.: [Wa15]). Sie sind aufgrund kleiner Nachrichtengrößen, einfacher Nachrichtenverwaltung und geringem Protokollballast besonders für die Kommunikation von Internet-basierten, ressourcenreichen Geräten mit ressourcenbeschränkten Geräten geeignet.

Bei beiden Protokollen ist die Anwendungsschicht reduziert. Es kommen einfache Zuverlässigkeitsstrategien zum Einsatz. Ziels ist es, einfache Rohdaten wie Messwerte an ressourcenreichere Geräte zu übertragen, die dann die Nachbearbeitung und Auswertung übernehmen.

MQTT folgt einem REST-Ansatz und nutzt TCP/IP als Transportschicht. Die Kommunikation folgt dem Publisher-Subscriber-Muster (vgl.: [Ga95]).

CoAP basiert ebenfalls auf einer REST-Architektur. Als Transportschicht kommt UDP zum Einsatz. IP Multicast kann für eine Gruppenkommunikation eingesetzt werden. Die Interaktion selbst folgt dem Client/Server Muster. Designziel bei CoAP war es, den Nachrichten-Overhead im Vergleich zu HTTP zu minimieren.

Beide Protokolle bieten sich für M-Health Applikationen an, bei denen relativ ressourcenschwache Geräte zur Erfassung von Vitaldaten und Umweltzuständen zum Einsatz kommen.

Abbildung 1 illustriert die unterschiedlichen Kommunikationsformen bei CoAP und MQTT.

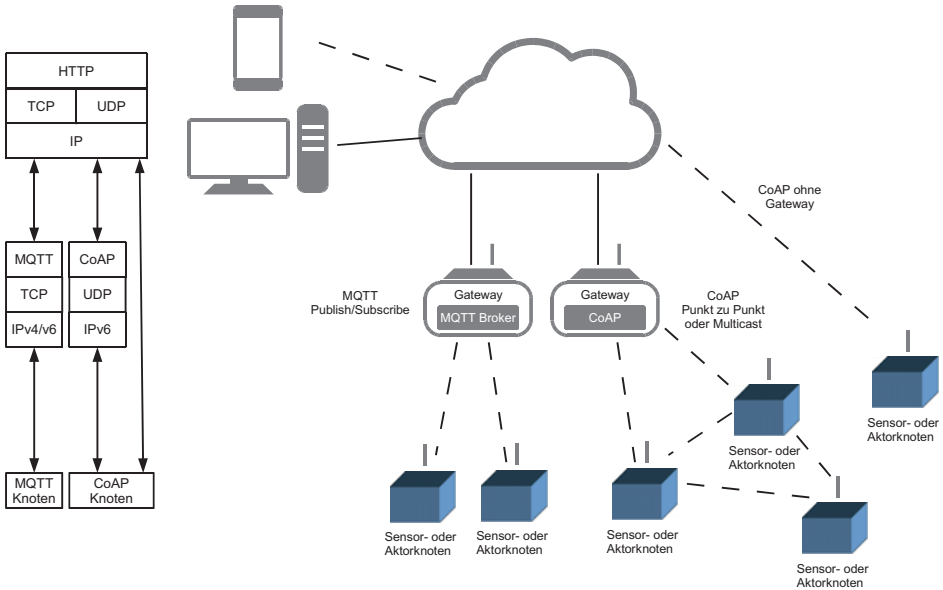


Abb. 1: Kommunikationsformen bei MQTT und CoAP

2.3 Funkstandards

Wesentliche Standards für die drahtlose Verbindung von Sensorknoten und eingebetteten Systemen sind 6LoWPAN (802.15.4) und Bluetooth insbesondere Bluetooth LE (802.15.1). Die Arbeitsgruppe IEEE 802.15 spezifiziert Wireless Personal Area Networks (WPAN). Wi-Fi selbst ist im Standard 802.11 festgelegt. Folgende Tabelle gibt eine Übersicht dieser Funkstandards im Vergleich zu 802.2 Ethernet. Viele weitere Standards wie ANT+, Zigbee oder Z-Wave existieren, werden hier jedoch aus Platzgründen nicht weiter betrachtet.

	802.15.4 6LoWPAN	802.15.1 Bluetooth	802.11 Wi-Fi	802.3 Ethernet
Klasse	WPAN	WPAN	WLAN	LAN
Batterie- lebensdauer	Jahre	ca. 1 Woche	< 5 Tage	—
Netzwerkgröße (Hops)	65k	7	30	1024
Bandbreite	250Kbps	>1Mbps	>11-100 Mbps	>100 Mbps
Reichweite	1 – 75 m	10 m	100 m	—
Ziele	geringer Energie- verbrauch, Skalierbarkeit, geringe Kosten	kabellose Anbindung von portablen Geräten	Durchsatz, Ersatz für drahtgebundene Netzwerke	Durchsatz

Tab. 1: Vergleich Funkstandards

Tabelle 1 listet wichtige Kenngrößen der Funkstandards auf - Ethernet ist hierbei als Vergleichsgröße mit aufgeführt. Aus Gründen des geringen Energiebedarfs verwenden Wearables meist WPAN Funkstandards. Für M-Health Anwendungen kommt darüber hinaus bei Indoor-Szenarien auch Wi-Fi in Frage. Bei fehlendem Netzwerk können leistungsfähigere Geräte über GSM kommunizieren.

2.4 Kommunikationsmodelle für M-Health

Aus den dargelegten technischen Grundlagen lassen sich prinzipielle Kommunikationsmodelle für M-Health-Szenarien unter Verwendung von Fitnesstracker, Smartwatches, Smartphones und selbst entwickelten Sensorknoten ableiten. Abbildung 2 zeigt die diversen Kommunikationswege von der Erfassung bis zur Auswertung auf lokalen Servern oder in der Cloud.

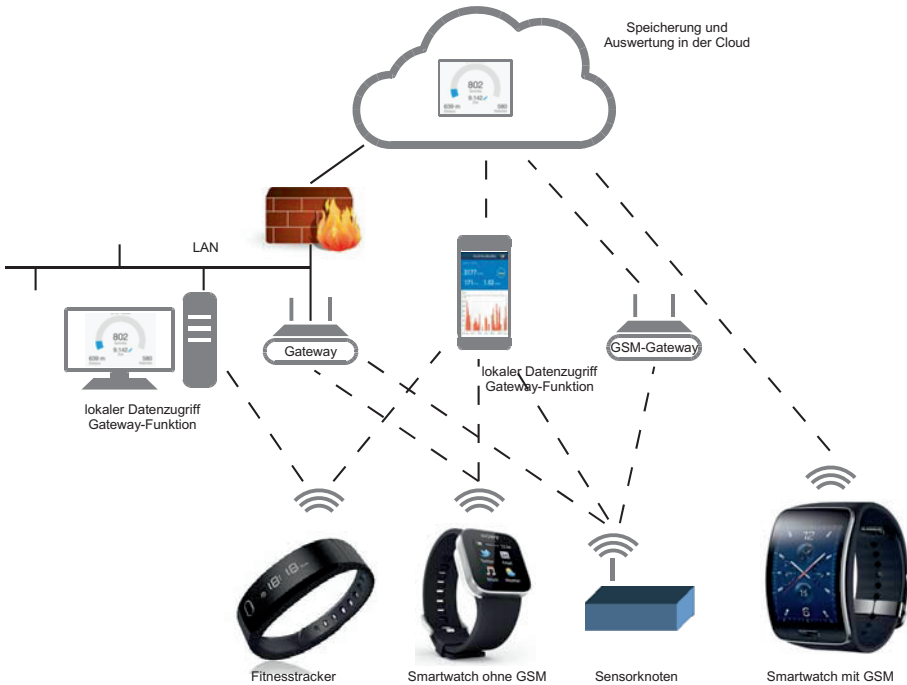


Abb. 2: Kommunikationsmodelle für M-Health Anwendungen

3 Anwendungsbeispiel

In den Studiengängen Informatik und Ingenieur-Informatik am Fachbereich Mathematik, Informatik und Naturwissenschaften der THM werden die Module *Kommunikation in eingebetteten Systemen* (Vorlesung und Praktikum) und *Intelligente Umgebungen und Assistenzsysteme* angeboten.

Um Studierende hier an Einsatzmöglichkeiten von Wearables im M-Health Kontext heranzuführen, wurde eine offene Beispielimplementierung entwickelt ([Eb16]). Ziel war es, eine leichtgewichtige und leicht erlernbare Basis für AAL-Anwendungsfälle zu schaffen, die eine einfache Integration verschiedener Hardware wie Sensorknoten oder Smartwatches ermöglicht. Um Daten solcher Geräte zu übertragen, wurde MQTT als Protokoll ausgewählt. Für komplexere Systemfunktionen stehen REST-Schnittstellen zur Verfügung.

Ausgangspunkt ist ein Szenario, bei dem benutzerbezogene Daten mittels Smartwatch und umgebungsbezogene Daten aus der Wohnung über ein Hausautomationssystem erfasst werden.

Für die Auswertung der Daten und Ereignisse sollen Applikationen bereitgestellt werden können, die beispielsweise bei potenziellen Gefahrensituationen wie einem Sturz oder extremen Vitalwerten eine Alarmkette bis hin zum Hausnotruf auslösen können.

Die Systemarchitektur ist modular aufgebaut und unterscheidet system- und nutzerbezogene Funktionsbereiche. Es wurden folgende vier Modulfelder gewählt:

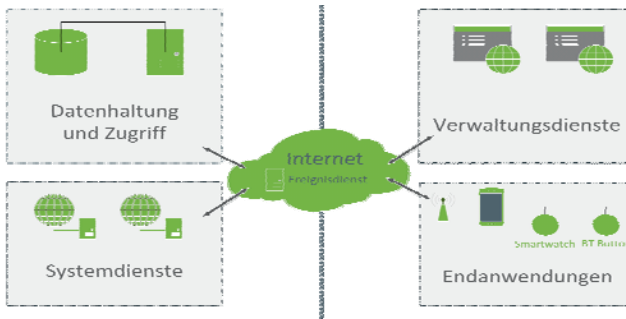


Abb. 3: Modulfelder der generischen Plattform

Zum Bereich *Datenhaltung und Zugriff* gehört beispielsweise das Management von Datenbanken, Objekt-Relationales Mapping und REST-Schnittstellen. *Systemdienste* stellen Basisfunktionen wie z.B. eine Ereignisverarbeitung bereit. Zu den *Verwaltungsdiensten* gehören vor allem die Benutzerverwaltung und die Konfiguration der Ereignisverarbeitung. Verwaltungsdienste greifen auf die systemnahen Dienste zurück.

Endanwendungen sind die Schnittstelle des Nutzers mit dem System. Vielfältige Geräte können hier über MQTT leicht eingebunden werden. Exemplarisch wird dies am Beispiel einer Smartwatch – konkret der Urban von LG - erläutert. Für die Smartwatch wurde eine App entwickelt, die Erinnerungs- und Alarmfunktionen implementiert. Die App ist als Uhren-Face immer im Vordergrund. Die Funktionalität und Bedienoberfläche wurde bewusst einfach gehalten, sodass auch ältere oder eingeschränkte Nutzer diese bedienen können.

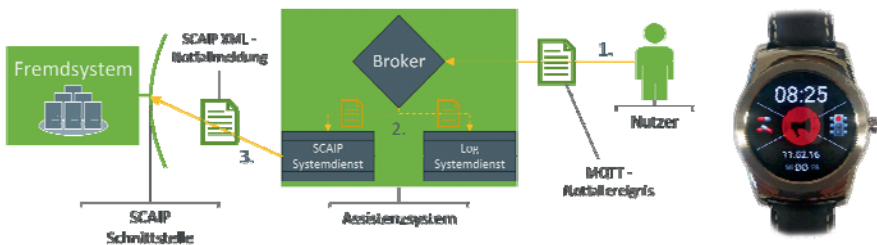


Abb. 4: Kommunikationsstruktur der Notfallmeldung (links) und Smartwatch (rechts)

Geplant war, einen von der Smartwatch ausgelösten Alarm über das System und hinterlegte Regeln an einen Hausnotruf weiterzuleiten. Für die Anbindung an den Hausnotruf existiert eine Reihe von herstellerspezifischen proprietären Protokollen. Ein Ansatz hin zu einer Standardisierung im Bereich des Hausnotrufs wurde vom Swedish Standards Institute SIS unterbreitet. Das SIS hat mit dem "Social Care Alarm Internet Protocol" SCAIP [SIS14] einen offenen Standard verabschiedet, der eine XML-Struktur definiert. SCAIP ist IP-basiert und so ausgelegt, dass "social care service chains" aus verschiedenen Services vom Pflegebringer bis zum Notruf realisiert werden können.

In die entwickelte Plattform lässt sich SCAIP als ein Modul leicht integrieren. Eine praktische Erprobung scheiterte leider an organisatorischen Randbedingungen eines konkreten Hausnotrufanbieters. Realisiert wurde alternativ die Anbindung an ein Gebäudeautomationssystem, über das dann ein Hausnotrufendgerät geschaltet wurde. Das verwendete Gebäudeautomationssystem verbindet sich über das EnOcean-Funkprotokoll mit diversen energieautarken Sensoren oder Aktoren. Parallel zum Schalten des Notrufs erfolgt eine Anzeige des Ortes und der Alarmmeldung in einer Browser-Anwendung auf einer digitalen Landkarte (Abb. 5).

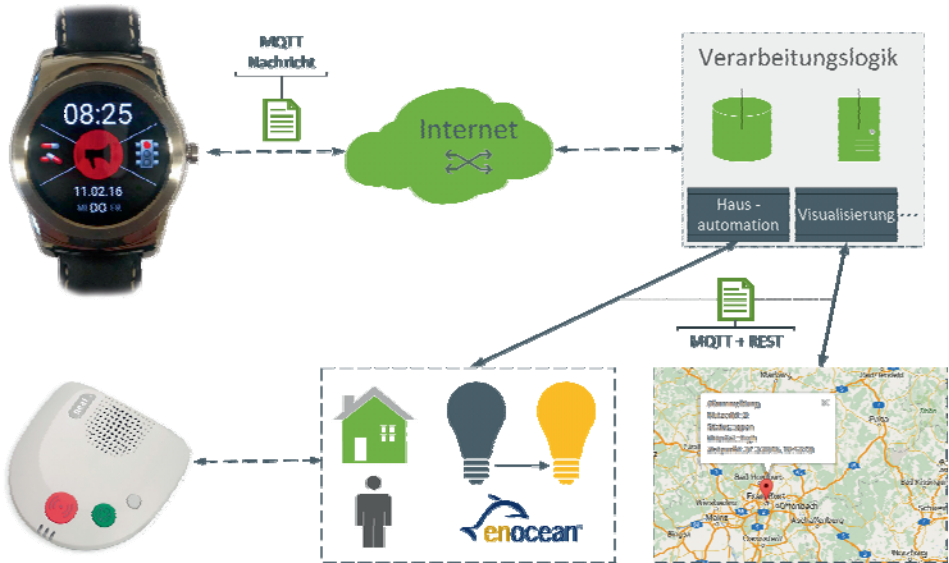


Abb. 5: Von der App zum Hausnotruf und zur Visualisierung

4 Handlungsfelder

Aus dem ersten Prototypen auf Basis der Umgebung und aus den Ergebnissen des Projektes „Gesund, sicher und mobil mit Technik und Serviceerbringung. Prävention - Telecare - Digital Health (GSMTS)“ (vgl.: [K116]) lassen sich einige Handlungsfelder ableiten.

Für den Einsatz von Wearables im Gesundheits- und Assistenzumfeld sind einige technische Voraussetzungen insbesondere offene Geräteschnittstellen notwendig.

Bedienkonzepte und Benutzungsoberflächen von Geräten für den Fitnessbereich lassen sich nicht ohne weiteres für Anwendergruppen mit körperlichen oder kognitiven Einschränkungen oder weniger technikaffine Nutzer übertragen. Hier sind Nutzerzentrierte Designmethoden notwendig. In [NSG13] werden Vorgehensweisen und Methoden aus User Experience (UX) Designansätzen hinsichtlich des Einsatzes speziell für den AAL Bereich untersucht.

Mittels UX Designmethoden wie *Storyboards*, *Paper Prototyping*, *Value Design Method* und *Personas* (vgl. u.a. [Ma16], [Wi08]) könnten vielversprechende Designlösungen für Anwendungen im Gesundheits- und Assistenzumfeld entstehen. Personas beschreiben archetypische Nutzer oder Stakeholder durch spezifische Charakteristika und persönliche Rahmenbedingungen wie Lifestyle, Bildungsstand, Einkommen oder Technikaffinität.

5 Ausblick

Das Angebot an Wearables ist in den letzten Jahren extrem gewachsen. Um solche Geräte für M-Health Anwendungen einsetzen zu können, sind wie hier übersichtsartig dargelegt noch einige Rahmenbedingungen zu erfüllen.

Doch was kommt nach den Wearables? Entwicklungen gehen dahin nach den auf der Haut getragenen 'Geräten als nächstes diese physische Grenze zu überwinden und „unter die Haut“ zu gehen. Implantables könnten der nächste Schritt der digitalen Revolution im Healthcare-Sektor werden.

Chronische Erkrankungen könnten automatisch überwacht werden und Fitnessdaten lassen sich *in vivo* wesentlich präziser erfassen. Über das reine Messen hinaus könnten Implantables einzelne Organ- und Regulationsfunktionen übernehmen. Implantables kommunizieren mit dem Träger, seiner Umwelt und dem Arzt - als Mensch oder KI-Programm.

Eine "Schöne neue Welt" zeigt sich am Horizont - ein verantwortungsbewusster Umgang mit dem technisch Möglichen ist umso mehr gefordert.

Literaturverzeichnis

- [AH16] Andelfinger, Volker P.; Hanisch, Till, Hrsg. eHealth - Wie Smartphones, Apps und Wearables die Gesundheitsversorgung verändern werden. Springer Gabler, 2016.
- [CWB15] Cearley, David W.; Walker, Mike J.; Burke, Brian: Top 10 Strategic Technology Trends for 2016: At a Glance. Bericht, October 2015. URL: https://www.it-systems.de/news-media/gartner-top-10-strategic-technology-trends-for-2016-at-a-glance/14061841/blobBinary/gartner_top10.pdf (Abruf Juni 2016).
- [Eb16] Ebner, A.: Entwurf und Entwicklung eines modularen Assistenzsystems im AAL Kontext, Thesis, Technische Hochschule Mittelhessen, FB MNI, 2016
- [Ed14] Edler, K.; Götz, C.; Pirchner, F.; Obermaier, D.: MQTT im IoT: Einstieg in die M2M-Kommunikation. shortcuts. entwickler.Press, 2014.
- [Fi00] Fielding, Roy Thomas: Architectural Styles and the Design of Network-based Software Architectures. Dissertation, 2000. AAI9980887.
- [Ga95] Gamma, Erich; Helm, Richard; Johnson, Ralph; Vlissides, John: Design Patterns: Elements of Reusable Object-oriented Software. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1995.
- [GSMW16] Gigerenzer, Gerd; Schlegel-Matthies, Kirsten; Wagner, Gert G.: Digitale Welt und Gesundheit. eHealth und mHealth - Chancen und Risiken der Digitalisierung im Gesundheitsbereich. Sachverständigenrat für Verbraucherfragen beim Bundesministerium der Justiz und für Verbraucherschutz, 2016.
- [K116] Klein, B.; Reutzel, S.; Roßberg, H.; Bienhaus, D.; Lutze, R.; Hofmann, J.; Dallwitz, D.; Sütö, S.; Heinrich, H.; Donath, M.: Anbindung von Hausautomation und Wearables: Akzeptanz und Gebrauchstauglichkeit. In: Zukunft Lebensraume Kongress 2016. 2016.
- [Ma16] Markopoulos, Panos; Martens, Jean-Bernard; Malins, Julian; Coninx, Karin; Liapis, Aggelos, Hrsg. Collaboration in Creative Design - Methods and Tools. Springer International Publishing Switzerland, 2016.
- [Me13] Meier: Plattform für Hämophilie-Patienten. Deutsches Ärzteblatt, 110(47), 2013.
- [NSG13] Nedopil, Christoph; Schauber, Cornelia; Glende, Sebastian: The art and Joy of user integration in AAL projects. Ambient Assisted Living Association, 3, Rue de Luxembourg 1000 Brussels, Dec. 2013.
- [SIS14] SIS Swedish Standards Institute. Digital Social Alarm - Social Care Alarm Internet Protocol (SCAIP) - SS91100:2014, 2014.
- [Wa15] Waher, P.: Learning Internet of Things. Community experience distilled. Packt Publishing, 2015.
- [Wi08] Witt, Hendrik: User Interfaces for Wearable Computers - Development and Evaluation. Vieweg+Teubner, 2008.

An Extendible Communication as a Service Platform for Wearables and Future-Oriented Devices

Roman Englert¹, Jörg Muschiol²

Abstract: Wearables are suitable devices for communication, navigation, and orientation, since they are easy to wear, and comfortable companions. A demerit is that most of them need to be connected to a master device or a server that provides the content. In this paper we propose and evaluate a prototype of a communication as a service (CaaS) platform that enables any wearable to be connected with. The architecture, and the implementation of the communication platform are described. Real probands were acquired for the evaluation of the CaaS system. The evaluation is based on communication tasks that were executed by the probands and are compared to the situation without the assistance of wearables and the communication platform in the background. A well-known usability measure, namely the Kullback-Leibler divergence is used to demonstrate the effectiveness of the CaaS platform that increases the usage of wearables for communication tasks by a factor of 2.58.

Keywords: Wearables, Communication as a Service (CaaS) Platform, Architecture, Usability, Evaluation

1 Introduction

Today communication is a ubiquitous appliance. This behavior emerged due to the invention of wireless devices like cellular phones (so-called mobiles) and cellular networks with fast data transmissions. The next generation of mobiles are wearables that are bracelets, smartwatches, glasses, etc. or communication devices that are integrated into clothes [Mu14]. Wearables have limitations, since they are small-sized, lightweight devices with small accumulators. Thus, they need a short-distance connection to a powerful smartphone. In the case they are offline, their internal storage can be used to collect and store data like geo-position, health condition, etc. Those data can later be transferred and processed in the cloud or a cloud-based system can offer multimodal communication with smart services as proposed in this paper.

Cloud-based services that offer (multimodal) communication are known as communication as a service (CaaS) [Sp11]. They may contain voice over IP communication, instant messaging, and video calls / conferencing. Multimodal communication enables different communication modalities like speech, tapping, gestures in parallel [EG06]. The proposed CaaS consists of three different domains for devices like

¹ University of Siegen, Faculty III, Institute of New Media, and FOM (University of Applied Sciences for Economy and Management), Siegen/Essen, roman.englert@t-online.de

² FOM (University of Applied Sciences for Economy and Management), Essen, muschiol@fom.de

wearables, data transmission, and users [Mu13]. All domains are processed in the cloud and thus fit perfectly to wearables. The separation of the domains makes the CaaS system expandable to new requirements by technological developments of wearables.

The evaluation is done with a data set of 30 probands [Mu14]. The underlying scenario is a grandma going to an appointment with her doctor. She needs assistance to finding the way and to arrive in time due to physical, and mental weakens. Assistance is brought to her by wearables in form of communication and navigation services. The probands had to fulfill six communication tasks with smart devices and six without. The execution times were measured and compared for the tasks with, and without wearables. These measurements were also compared with the Gaussian distribution, assuming that the skills for using these smart devices are normal distributed. This comparison is done with the deviation measure Kullback-Leibler divergence (KLD) that is known from statistics [Ha69]. The KLD is insensitive to outliers, which is important for real data. The evaluation shows that the probands were able to reach their appointment with the assistance of the wearable-based services, and the execution times of the tasks were reduced by a factor of up to six. Additionally, the comparison with the KLD showed that the tasks execution improved with the usage of the wearables by a factor of 2.58.

The structure of the paper is as follows: wearables as communication devices are described in Sect. 2. Then, the proposed and implemented CaaS system and its architecture are described in Sect. 3. The evaluation with the real probands is done in Sect. 4, including a description of the scenario and how the measurements were taken. A brief introduction into the KLD is provided and the data evaluation is included. Sect. 5 contains the discussion of the results. The paper concludes with a summary and outlook for further research in Sect. 6.

2 Wearables as Communication Devices

Wearable Computing extends the usage of Information and Communication Technology (ICT) with small, body-worn, hands-free to use devices [De99]. Additionally, it enables new possibilities of mobile, context-sensitive technology usage with the extension of the human senses that can be enhanced by sensor integration in the devices [Ma13]. The devices can interact context aware autonomously or in addition to other devices, e.g. smartphones or embedded systems in cars. Mann [Ma98] describes three modes how wearable devices interact: 1) The constant mode provides a permanent and stable interaction with the user. 2) The augmentation mode offers interchange with the user, the device, and the environment without distraction of the real world. 3) Capsulation between human and the device is supported by the mediation mode which filters information in the solitude mode and provides security of the individual in privacy mode.

Many devices are named “wearables” (e.g., watches, wrists, bracelets, glasses, textiles and accessories) and the Human Computer Interaction (HCI) differs from device to device [Ra15], sometimes with poor approaches of user engagement. Wearables are used in

scenarios where they support human communication activities in domains as health care, activity recognition, fitness, elder care, and entertainment.

Challenges of using wearables efficiently are often small displays, poor computing power, and small accumulator capacity. As a consequence, the user has to recharge them often (e.g. Samsung Gear S average charging period is about 2 days), with complicate and unmanageable procedures and s/he has to remember to wear the device. Many devices need additional steps or more equipment to be enabled and that results in low user involvement [Pa15]. In 2014 Motti and Caine [Mo14] evaluated success attributes for the user acceptance of wearables, e.g.:

- UX/UI & Usability (intuitive, simple and easy to use devices)
- Quality (audio, video and image resolution with high contrast, details and sharpness),
- Contextual awareness (external factors as light, brightness and noise),
- Battery lifetime (time to charge, compatible chargers),
- Price and overall costs (overall costs and benefits),

In a survey with wearable devices users, 75% described themselves as “early adopters of technology” and 48% were younger than 35 years, with an income of more than US \$ 100,000 annually [Ni14]. Smartwatches are one of the first device categories of wearables which move from the specialist market to the mainstream. Recent advances in miniaturization and diminishing costs in processor, sensors, and communication technologies enable multisensor devices with capabilities to measure the heart-rate, galvanic skin response, activity level, and also supporting gesture or speech recognition [Ra15]. Most of them act as an additional display/device (e.g., Pebble, Apple Watch, Motorola Moto 360), and a minority also enable fully independent communication, and interaction with an autarkic voice and data connection (e.g. Samsung Gear S [Sa16]) with an own (e)sim card. The operating systems of, e.g., Samsung Tizen, Android Wear or Apple Watch differ in proprietary features, and usability concepts.

One demerit of smartwatches is the tiny display. The device needs to be small enough to be worn on a wrist and as a result the user interface is restricted by the size (especially for fashion- oriented watches). Supporting multimodality voice in- and output can help to solve this problem. As a wearable acts contextual, the application appears smart for the user interaction, and the combination of GSM, Wi-Fi, and GPS is mandatory. As a result of the combination of all sensor data, the CPU and sensor activity minimize the energy supply. Context awareness is still one of the most important research topics in wearable computing, challenged by the location awareness for indoor and outdoor scenarios [Ro14].

Wearable computing devices as “Google Glass” [Go16], “Oculus Rift” [Oc16], and “Microsoft HoloLens” [Mi16] belong to the category of Eyewear Computing with Head-Mounted Devices (HMD). They extend the real world with contextual digital information layered on the real environment (Google Glass) or they enable virtual reality platforms

where the user walks with an avatar through a virtual environment, enhanced with real time gestures and objects. (Oculus Rift, Microsoft HoloLens). In 2014 Roggen [Ro14] stated that wearables as “Google Glass” should not replace the reality but they seamlessly augment it with micro interactions (only two to four seconds) and the technology has to stay out of the way when it is not needed [Ro14]. Kimura developed an eyeglass based videoconferencing system which fuses the images from four to six fish-eye cameras on the front glass frame to reconstruct the users face and send this to the video conference partner (see Fig. 1, a). Thus, media rich User Experience (UX) video conferencing becomes possible with just one sole wearable-only solution [Ro14].

Research topics as “Augmented Reality in a Contact Lens” [Pa09] demonstrate the future potential of Eyewear computing with the Augmented Reality (AR) technology. Googles project “Contact Lens” aims to assist people with diabetes by constantly measuring the glucose levels in their tears, and supports wearable healthcare scenarios [Fa15]. Hiroki Watanabe supports an approach of a miniaturized, integrated dietary monitoring with acceleration sensors integrated in a tooth (see Fig. 1, b) [Ro14].

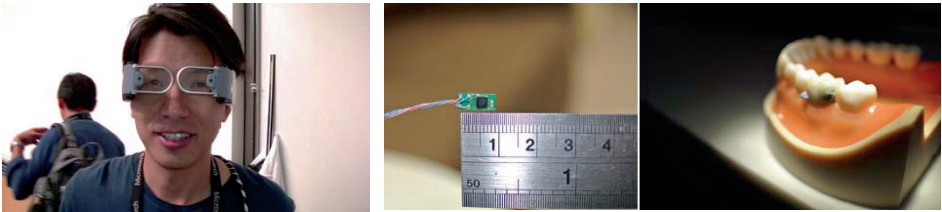


Fig. 1: (a) Eyeglass-based video conferencing system / (b) Dietary monitoring integrated in a tooth

Actually, a lot of research projects and also end-user products of pervasive wearable computing are upcoming. Giving some final examples:

The “dash” (see Fig. 2, a) earables are a new category of autarkic, seamless, wireless earphones, combined with high-level sensors. They support measuring health parameters combined with convenience wearing comfort, an aesthetic design and UX. The integrated microphones enable this device as a smart communication wearable [Br16]. Another innovative approach of a (text) communication device with a high level fashion design is the eyecatcher bracelet (see Fig. 2, b). This e-ink supported wearable is adaptive to the context, enables the communication via text and solves the problem of user acceptance by being a piece of jewelry. It is also nearly energy autarkic through the low energy consumption of the e-ink display [Lo16].



Fig. 2: (a) „Earable“ seamless, wireless smart earphone / (b) Eyecatcher fashion e-ink bracelet

Yokota et al. demonstrate an ultra-flexible, and conformable three-color, highly efficient polymer light-emitting diodes (PLEDs), and organic photodetectors (OPDs) to realize optoelectronic skins (oe-skins). That introduce multiple electronic functionalities such as sensing, and displays on the surface of human skin (see Fig. 3). The total thickness of the devices, including the substrate, and encapsulation layer, is only 3 μm, which is one order of magnitude thinner than the epidermal layer of human skin [Yo16]. This example depicts the next level of body worn technology.

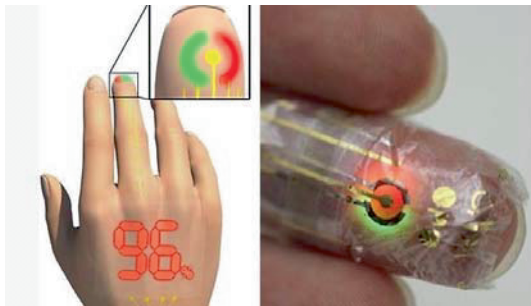


Fig. 3: Ultra-flexible, and conformable three-color display

Although wearable devices have the potential to facilitate communication, and information demands on a variety of application domains, this benefit might not be driven by one devices alone. The successful use, and potential benefits to an interaction of a mass of devices, and their combined interaction of communication.

In the following, an architecture platform including future technologies as expected for future communication will be described and discussed.

3 CaaS and the Architecture

The proposed CaaS system is based on an architecture with three domains [Mu13] (Fig. 4): The domain “anybody (left and right part) contains the user. S/he may communicate in a multimodal manner with speech, gestures, mimic and haptic according to individual

preferences. Multimodal communication must be supported by the applied devices that are part of the domain “any device” (interior left and right side).

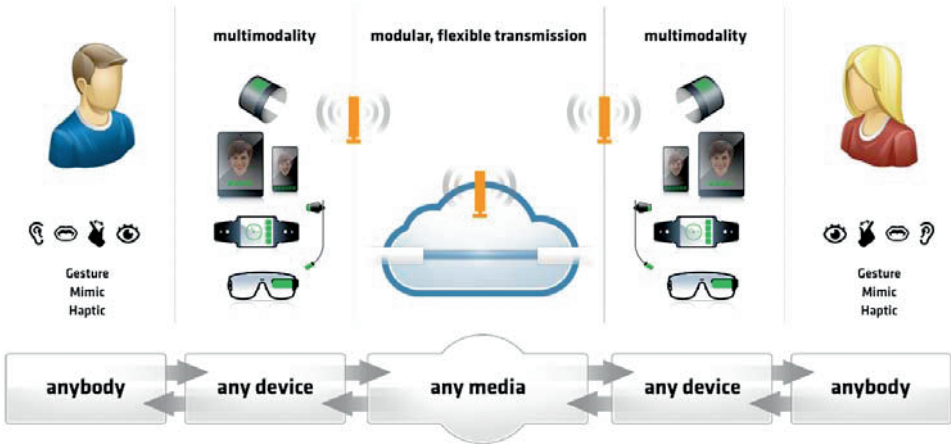


Fig. 4: Universal communication model (UCM) for wearable communication devices

If wearables are used, then it is recommended to connect them to a cellular phone (mobile) that contains a powerful accumulator. Finally, the domain “any media” consists of networks like LTE for the data transmission. These networks enable users to be nomadic and to transfer data with high quality and low latency.

The implementation of this architecture is done by several modules (Fig. 5) that are part of the universal communication platform (UCP): the left column contains the processing modules for the users’ input/output devices. These devices support media richness and multimodality. Here, media richness combines HD video and audio input/output, and provides the interaction application for data aggregation of external information sources like location or weather data. Touch devices enable sensor input that is processed by the application server (center column). Additional devices (bottom left part) are further displays or interaction devices like wearables, smartwatches and loudspeakers. They interact directly via a wireless connection with, e.g., a phablet to support multimodal interaction. The center column of the UCP contains a server running applications for data and client aggregation. Finally, the external services (right column) provide further information sources like NFC or video streams and from value-added content providers for e.g. weather information. Additional functionalities from external services like text-to-speech (TTS) for multimodality and WebRTC (We16) for integrated audio and video communication can also be processed.

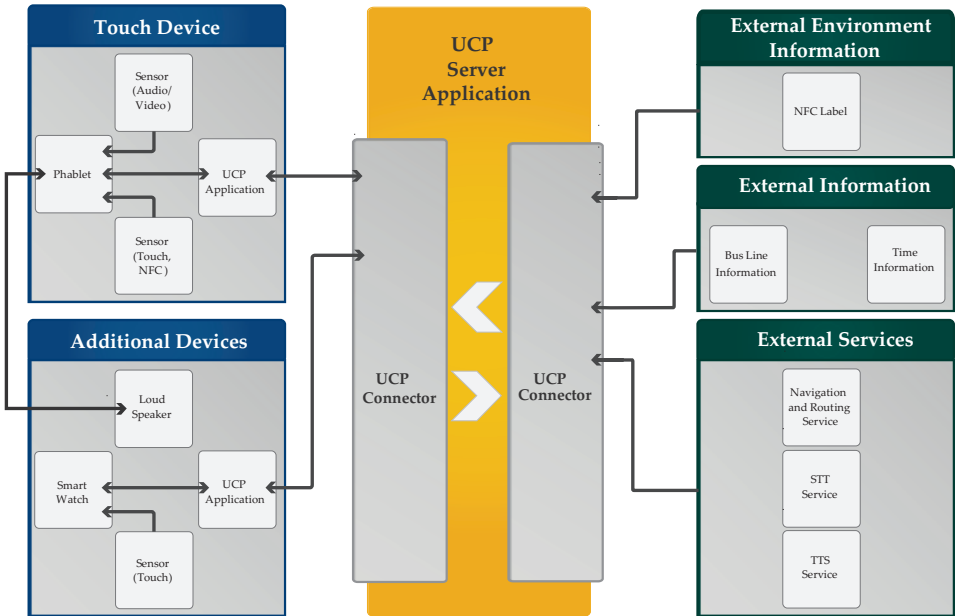


Fig. 5: Architecture of the UCP for wearables

CaaS offers multimodal communication for light devices like wearables, since computing intensive applications are executed in the cloud. The necessary high quality connection is provided by wireless networks like LTE and the wearables can be connected to LTE via a powerful smartphone. The implemented platform offers additional services for any user and any future device, especially wearables.

4 Evaluation: Tasks and Measurements

The evaluation is based on the scenario that a grandma with physical and mental capability loss has an appointment with her medical doctor and she needs to use the public transportation system and to walk the last mile in order to visit the clinic (Fig. 6). Grandma has a wearable equipment consisting of smartglasses, a bracelet with vibration, a phablet and a smartwatch. The smartglasses, the bracelet and the smartwatch are connected to the phablet that holds a LTE connection to the CaaS system which supports grandma on her way to the doctor.

Before grandma starts at home, the appointment is announced on her smartglasses and the bracelet is vibrating. She receives a TTS information for the navigation and starts to walk to the bus stop. After entering the bus, the phablet buys a ticket via NFC and the CaaS system acknowledges the appointment. The information to leave the bus is displayed on

her smartglasses including the direction to walk. Secure street crossing is assisted by bracelet vibrations (warnings), when low contrast of traffic lights makes the separation of the colors red / green difficult. Her vital condition can be checked before she may decide to use the stairs in the clinic.

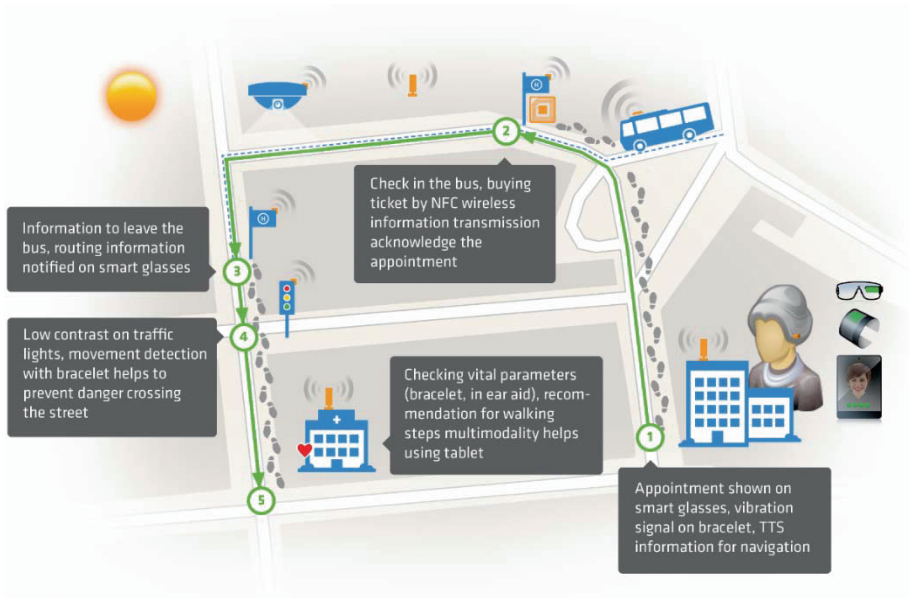


Fig. 6: Grandma's way to the doctor and navigation support due to her physical limitations

This scenario is simulated in a laboratory with six stations, where the probands have to execute different tasks (Fig. 7) [Mu14]:

1. Plan the trip.
2. Select the proper bus line.
3. Determine the current position.
4. Navigate on a street crossing.
5. Check the time schedule during the trip.
6. Confirm the appointment.

The CaaS supports the probands by providing the above in the scenario described information and the user just has to confirm an action. In the case, where the CaaS is not available the probands have to use paper maps and bus plans, and a smartphone. The challenge for the probands is on one side to use the different wearables and the CaaS, and

on the other side to interpret maps and bus plans, and to use the smartphone.



Fig. 7: A proband checks the time schedule during a trip and confirms appointment (tasks 5 and 6)

The 30 probands have been selected according to their ICT usage [Mu14]: more than half of them are using communication devices, digital photography, and a tom-tom. Only one third uses online banking, online shopping, electronic calculators and calendars. All probands are 60+ and many of them had a mental and/or physical capability loss which makes the usage of wearables a challenge, since they are unexperienced with them.

P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P21	P23	P24	P25	P26	P27	P28	P29	P30
114	49	89	153	77	82	86	69	138	71	63	80	89	75	94	45	79	31	29	94	94	46	70	126	90	86	106	93	57	121
91	99	106	65	27	28	70	39	69	25	36	40	28	26	82	47	44	26	31	86	85	20	34	51	90	48	95	44	39	95
32	41	39	5	27	4	14	6	61	22	21	11	28	15	30	10	29	21	14	8	10	5	11	6	17	6	11	20	15	38
12	34	21	14	2	8	17	24	6	85	2	12	7	12	15	17	13	5	15	2	5	3	9	7	8	8	8	10	4	6
65	70	41	32	10	44	33	27	21	15	38	27	63	8	76	19	41	10	14	11	52	24	51	38	27	16	11	28	21	58
40	32	30	30	40	24	27	19	33	26	29	31	18	18	34	24	31	21	16	18	33	22	21	39	25	20	55	27	29	44
62	23	46	29	21	25	11	11	19	29	21	21	25	5	9	7	10	12	9	25	40	22	9	11	18	12	51	9	20	19
37	19	32	13	57	5	10	28	52	19	17	16	21	18	16	16	32	5	5	5	69	29	16	22	44	24	16	13	48	24
19	29	34	27	15	20	14	31	28	34	14	13	20	31	24	16	18	14	15	17	48	14	21	14	21	18	42	22	17	54
12	17	19	13	14	9	11	9	8	9	10	10	9	14	8	13	11	13	7	13	9	8	10	7	7	8	7	7	6	9
10	15	15	11	8	8	8	8	13	10	16	8	10	14	14	9	12	18	6	6	8	10	7	8	7	1	5	6	6	10
4	3	10	12	4	3	5	2	8	4	3	3	3	5	5	3	3	2	1	2	4	3	1	2	6	2	3	1	1	3

Tab. 1: Six task execution measurements for the 30 probands without CaaS support (top) and six measurements with CaaS support (bottom)

Each proband had to perform six tasks without wearables and CaaS support and six tasks with wearables and CaaS support. The execution times were measured resulting in $30 \times 6 \times 2 = 360$ measurements (Tab. 1). In the subsequent section these measurements are evaluated and discussed.

5 Results and Discussion

The evaluation of the measurements is done with two methods: First, the time measurements without and with support of wearables and the CaaS are compared with each other (Sect. 5.1). Second, the measurements are converted into statistical distributions to compute the deviation of the Gaussian (normal) distribution. The deviation provides a measure for the capability of the probands to use wearables, where the normal distribution is the bias (Sect. 5.2).

5.1 Comparison of the Time Measurements

The measurements for the execution of the six tasks without the usage of wearables and CaaS are the following (Tab. 1): Task 1 between 29 and 153 seconds (1st row). Task 2 between 20 and 106 seconds (2nd row). Task 3 between 4 and 61 (3rd row). Task 4 between 2 and 85 seconds (5th row). Task 5 between 8 and 76 seconds (5th row), and task 6 between 18 and 55 seconds (6th row).

The measurements for the six tasks with the usage of wearables and CaaS are the following (Tab. 1): Task 7 between 5 and 62 seconds (7th row). Task 8 between 5 and 69 seconds (8th row). Task 9 between 14 and 54 (9th row). Task 10 between 6 and 19 seconds (10th row). Task 11 between 1 and 15 seconds (11th row), and task 12 between 2 and 12 seconds (12th row).

The time intervals are compared by computing the improvement of the task executions in percentage for the minimum and the maximum of the intervals without and with wearables: Task 1 ($(|5-29|)/29 \cdot 100\% = 83\%$, 60%), task 2 (75%, 35%), task 3 (71%, 13%), task 4 (67%, 347%), task 5 (700%, 407%), and task 6 (800%, 358%). Hence, the minimum execution time improvement is up to 800% and the maximum execution time improvement is up to 358%. The results show that especially the latter two tasks, namely 5 resp. 11 and 6 resp. 12, both are strongly supported by the wearables and CaaS. A limitation of this comparison is the outlier sensitivity. Therefore, an entropy-based measure is applied in the following section.

5.2 Comparison of the Time Distribution and the Gaussian Distribution

The Kullback-Leibler divergence (KLD) is an entropy-based measure in order to compute the divergence of two statistical distributions [CT91]:

$$D(P||Q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}, \quad (1)$$

whereas $P(X)$ is the sample (measurement) and $Q(X)$ the theoretical sample distribution, with $p(x)/q(x) > 0$. In the field of parametric tests sample distributions are compared with a theoretical distribution [Ha69]:

1. The unique distribution, where each data point occurs with the same probability:

$$U(x) = \left(\frac{1}{n}, \dots, \frac{1}{n}\right), \quad (2)$$

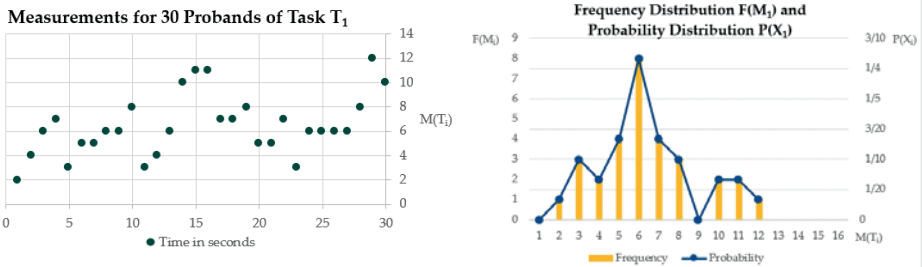
or a curved distribution like

2. The Gaussian distribution, where the assumption is, that the capabilities of the probands are normally distributed. However, the mean $\mu = \frac{1}{n} \sum X_i$ and the variance $\sigma^2 = \frac{1}{n} \sum (X_i - \mu)^2$ of X need to be computed:

$$G(x) = \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}}, \dots, \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}} \right). \quad (3)$$

Summarizing, in the beforehand evaluation compares the KLD (Formula (1)) the distribution of the measurement with the Gaussian distribution (Formula (3)). A larger entropy means, that there is more uncertainty for a distribution. The distribution of the measurement is achieved by computing the frequency distribution from the measurement, and subsequently converting the frequency distribution into a probability distribution [Ha69]. Fig. 8 shows an example for the measurements (left part) and the frequency and probability distributions (right part).

Fig. 8: Exemplary time measurements for 30 probands (left part) and the corresponding frequency (yellow) and probability (blue) distributions (right part)



In the case without the support of wearables and CaaS are the KLDs for the six tasks (Formula (1)): 3.154, 2.691, 1.611, 2.157, 2.157, and 1.932. Since the KLD is an entropy-based calculation, indicate smaller numbers a better fitting of the measured probability distribution with the Gaussian distribution. Thus, the 3rd task measurement has the best fitting of both distributions and the 1st task measurement a less good one with the Gaussian distribution. This means, that the execution times of the probands for the 3rd task are nearly normal-distributed, and in contrast to this have the execution times for the 1st task some outliers according to the Gaussian (normal) distribution.

The other case, where the support of wearables and CaaS is used for the task execution, shows the following KLD computations: 1.962, 2.174, 1.861, 1.036, 0.835, and 0.845. Smaller KLDs indicate lesser outliers for nearly all tasks compared without the usage of wearables, except task 3 “determine your current position” was easier without wearable support. Especially tasks 3-5 were much easier to solve with wearables and CaaS support.

In total the execution of the tasks with wearable and CaaS support has a better fitting of the measurements with the Gaussian distributions than the executions without wearable and CaaS support. This result underlines the usefulness of wearables combined with a CaaS system to solve communication tasks.

6 Summary and Outlook

In this paper we have shown that wearables are useful devices to fulfill communication and navigations tasks. This result has been proven for elderly people with physical or mental loss. Additionally, the chosen probands were digital immigrants, that were not used to wearables. It has been shown that the maximum execution time for the tasks could be improved up to 358%. For five of six tasks the elderly probands were nearly normal distributed with less outliers according to their capabilities to handle wearables. Furthermore, navigation support systems (“when to leave the bus”, “next crossing right”) are more reliable than the human memory.

A central key factor for using wearables are an easy connectivity and an easy interaction interface. Unfortunately, the manufacturers differ in their design and interaction paradigms. Additionally, they differ in their functionality, especially for multimodal communication. Therefore, we plan to develop a common user interface with a common functionality set for wearables that is also suited for outdoor application.

Acknowledgements

The authors are thankful to Christoph Römer for designing the layout and masks of the prototype, and additionally, to Sebastian Möller (Technical University of Berlin) for fruitful discussions.

References

- [CT91] Cover, T.; Thomas, J.: Elements of Information Theory. New York: Wiley & Sons, 1991.
- [Br16] The dash, April 2016. Available at <http://www.bragi.com/>.
- [De99] Dey, A. K., Abowdm, G. D., and Futakawa, M.: The Conference Assistant: Combining

- Context-Awareness with Wearable Computing. In Proceedings of the Third International Symposium on Wearable Computers: Digest of Papers. Los Alamitos, California: IEEE Computer Society. Washington, 1999, pp. 21-28.
- [EG06] Englert, R.; Glass, G.: An Architecture for Multimodal Mobile Applications: Proc. of the Twentieth Symposium on Human Factors in Telecommunication (HFT2006). Sophia Antipolis, France: ETSI, 2006.
- [Fa15] Farandos, N. M. et al.: Contact Lens Sensors in Ocular Diagnostics. In *Advanced Healthcare Materials*, 2015, Vol.4, No. 6, pp. 792-810.
- [Go16] Google Glass, April 2016. Available at <https://developers.google.com/glass/>.
- [Ha69] Hayes, L.: *Statistics*. Rinehart and Winston Publishers, New York, 1969.
- [Lo16] Eyecatcher, Lookseelabs, April 2016. Available at <http://www.lookseelabs.com/#infinitepossibilities>.
- [Ma98] Mann, S.: Keynote: Wearable Computing as Means for Personal Empowerment. In Proceedings of the First International Conference on Wearable Computing. Fairfax, Virginia: IEEE Computer Society Press. Washington DC, 1998.
- [Ma13] Mann, S.: Wearable Computing. In *The Encyclopedia of Human-Computer Interaction*, edited by The Interaction Design Foundation. Denmark, Aarhus, 2013.
- [Mi16] Microsoft Holo Lense, April 2016. Available at <https://www.microsoft.com/microsoft-hololens/en-us>.
- [Mo14] Motti, G.; Caine, V.; Caine, K.: Understanding the Wearability of Head-Mounted Devices from a Human-Centered Perspective. In Proceedings of the 2014 ACM International Symposium on Wearable Computers, 2014, pp. 83-86.
- [Mu13] Muschiol, J. "Universal Communication Model for the Future Society." In *Lecture Notes in Informatics: Proceedings of the Conference Informatik 2013: SUBICO Workshop*. Koblenz: Springer. Berlin, Heidelberg, 2013, pp. 1690–1703.
- [Mu14] Muschiol, J.: *Universal Communication Model and Evaluation with the Elderly Society*. PhD thesis, Universidad Católica de Murcia, Spain, 2014.
- [Ni14] Nielsen: Are Consumers Really Interested in Wearing Tech on Their Sleeves? Nielsen website. <http://www.nielsen.com/us/en/insights/news/2014/tech-styles-are-consumers-really-interested-in-wearing-tech-on-their-sleeves.html>. March 20, 2014. Accessed April 30, 2016.
- [Oc16] Oculus Rift, April 2016. Available at <https://www.oculus.com/de/rift/>.
- [Pa09] Parviz, B. A.: *Augmented Reality in a Contact Lens*, 2009. <http://spectrum.ieee.org/biomedical/bionics/augmented-reality-in-a-contact-lens>, accessed April 2016.
- [Pa15] Patel, M. S.; Asch, D. A.; and Volpp, K. G.: Wearable devices as facilitators, not drivers, of health behavior change. In *Jama*, 2015, Vol. 313, No.5, pp. 459-460.
- [Ra15] Rawassizadeh, R.; Price, B.; Petre M.: Wearables: Has the Age of Smartwatches Finally Arrived?. In *Communications of the ACM*, 2015, Vol. 58, No. 1; pp. 45-47.

- [Ro14] Roggen, D. et al.: ISWC 2013 – Wearables Are Here to Stay. In IEEE Pervasive Computing, 2014, No. 1; pp. 14-18.
- [Sa16] Samsung Gear S. April 2016. Available at <http://www.samsung.com/de/consumer/mobile-devices/wearables/>.
- [Sp11] Spath, D.; Weisbecker, A.; Kokemüller, J.; Stuttgart Fraunhofer IAO (Hrsg.): Cloud Computing und Software-as-a-Service.: Tagungsband des Stuttgarter Softwaretechnik Forums 2011 am 28. September 2011, Fraunhofer Verlag.
- [We16] WebRTC. April 2016. Available at <http://www.webrtc.org>
- [Yo16] Yokota, T. et al.: Ultraflexible Organic Photonic Skin. In Science Advances, 2016, Vol. 2, No.4, e1501856.

xHealth: Eine MQTT und REST basierte Architektur zum Zugriff auf Sensordaten smarterer Objekte

Manuel Guttenberger¹, Klemens Waldhör²

Abstract: Smarte Objekte wie Smartwatches oder Wearables verfügen über Gyro- und Beschleunigungssensoren zur Erfassung der Umwelt. Zur Erkennung von Mustern in Bewegungsdaten mittels Data Mining sind für viele Anwendungsfälle die Sensordaten zunächst an einen zentralen Service zu übermitteln, um komplexe Berechnungen wie Klassifikation durchzuführen. Im Rahmen dieser Arbeit wird ein Konzept, Anwendungsfälle sowie ein Prototyp für das System xHealth vorgestellt. xHealth empfängt Sensordaten über das MQTT-Protokoll sowie einer RESTful API basierend auf HTTP, speichert diese und stellt diese für weitere Analysen aus dem Bereich Big Data zur Verfügung.

Keywords: Healthcare, Digital, Transformation, Internet of Things, Web of Things, M2M, Sensordaten, Cloud, Architektur, REST, HTTP, MQTT, API, ADL/EDL Erkennung

1 Einleitung

Gartner prognostiziert für das Jahr 2020 mehr als 25 Mrd. Geräte im „Internet of Things“ (IoT) [Pe16]. Insbesondere Wearables und Smartwatches werden durch diese Entwicklung an Bedeutung gewinnen. Diese smarten Objekte schaffen die technische Basis für den Service von xHealth. Zur Analyse von Sensordaten smarterer Objekte über einen zentralen Service mittels Data Mining werden geeignete Protokolle benötigt. Die Vielfalt an verfügbaren Protokollen im IoT ist ebenso groß wie die Bandbreite unterschiedlicher Geräte, welche das IoT mangels Interoperabilität vom „Internet of Things“ zu einem „Intranet of Things“ mutieren lässt. Es fehlt ein universelles Protokoll, wie es etwa im Web mit dem HTTP-Protokoll existiert. Hier setzt die Idee des Web of Things (WoT) an, in welchem etablierte Webtechnologien wie HTTP und RESTful APIs für den Datenaustausch zwischen smarten Objekten verwendet werden [GT15]. Ein populäres Protokoll im IoT stellt MQTT (Message Queue Telemetry Transport) dar [W316]. MQTT ist ein effizientes und schlankes Protokoll, welches zur Kommunikation zwischen smarten Objekten entwickelt wurde. Je nach Sensor werden mehrere hundert Messungen pro Sekunde generiert, meist in mehreren Achsen (x,y,z) sowie einem dazu passenden Zeitstempel, ggf. weitere Zusatzinformationen. Werden nur Gyrometer- und Beschleunigungsdaten mit jeweils drei Achsen verwendet, wird pro Sekunde ein Datendurchsatz von mindestens 5000 – 6000 Byte (600 – 1000 Sensordaten) erzeugt. Wichtig

¹ FOM Hochschule für Oekonomie und Management, Hochschulstudienzentrum Nürnberg
City Park Center, Zeltnerstr. 19, 90443 Nürnberg, manuel.guttenberger@mgu.io

² FOM Hochschule für Oekonomie und Management, Hochschulstudienzentrum Nürnberg
City Park Center, Zeltnerstr. 19, 90443 Nürnberg, klemens.waldhoer@fom.de

ist zu berücksichtigen, dass diese Sensordaten asynchron anfallen, daher bei den meisten Geräten mittels Callback-Funktionen des jeweiligen Sensors realisiert sind und übertragen werden sollen. Für Geräte wie Wearables und Smartwatches stellt dies eine hohe Hardwareauslastung dar, insbesondere da diese Geräte meist noch weitere Aktionen parallel durchführen, über eigenes Energiemanagement zur Akkulaufzeitoptimierung verfügen, das die Geräte wann immer möglich in einen Schlafzustand versetzt, und gleichzeitig nicht über die Hardwareausstattung eines PC oder auch eines Mobiltelefons verfügen. Eine effiziente und schlanke Übertragung der Daten stellt damit einen entscheidenden Faktor bei der Implementierung dar.

2 Forschungsfragen und Methodik

Eine erste und naheliegende Lösung wäre, Sensordaten mittels HTTP (neben Alternativen wie Bluetooth in der Kommunikation zwischen Smartwatches und Mobiltelefonen) an einen entsprechenden Webserver zu übertragen. HTTP Requests erzeugen aber einen entsprechenden Overhead an übermittelter Information, was zu einer entsprechenden Belastung der Verbindung und beteiligten Anwendungen führt.

Die zentrale Forschungsarbeit dieser Arbeit ist damit:

- Können mittels MQTT Sensordaten effizienter übermittelt werden als mit HTTP?
- Oder anders formuliert: Ist MQTT das Protokoll der Wahl bei der Übertragung von Sensordaten?

Im Rahmen dieser Arbeit wird ein Konzept für das System xHealth vorgestellt und mittels Prototyping zwei Szenarien exemplarisch implementiert. Im ersten Szenario wird eine xHealth basierte Wetterstation implementiert. Im zweiten Anwendungsfall wird die ADL/EDL-Erkennung auf Smartwatches auf ihre Tauglichkeit in Bezug auf MQTT untersucht. Zur Beantwortung der Forschungsfrage wird der Datendurchsatz von HTTP und MQTT mit einem Arduino Uno Mikrocontroller untersucht.

3 Das MQTT Protokoll

3.1 Entwicklung von MQTT

MQTT wurde im Jahr 1999 von Stanford-Clark (IBM) und Nipper (Eurotech) zur Überwachung von Pipelines in der Wüste entwickelt. Zur Datenübertragung wurden teure Satellitenverbindungen verwendet, weshalb die Effizienz des Protokolls das vordergründige Ziel bei der Entwicklung darstellte [HI16].

Beflügelt wurde das Interesse an MQTT durch das 2011 gestartete Projekt Eclipse Paho, welches u.a. Implementierungen für C, Java, JavaScript, .Net und Android Services bereitstellt [EC16b]. Seit 2014 ist MQTT in der Version 3.1.1 von der Organisation OASIS standardisiert und befindet sich aktuell in der Entwicklung zur Norm ISO/IEC 20922 [BG14], [IS16].

Neben der M2M-Kommunikation und dem IoT wird MQTT auch für Messenger Dienste wie den Facebook Messenger eingesetzt. Durch den Einsatz von MQTT und einer persistenten Verbindung zwischen mobilen Endgeräten und Server konnte die Latenz beim Nachrichtenversand um ein Vielfaches gesenkt werden [Zh16].

3.2 Nachrichtenversand mittels MQTT

MQTT basiert auf einer Client-Server-Architektur und implementiert das Publish/Subscribe Pattern [BG14]. Im TCP/IP-Referenzmodell ist MQTT in der Anwendungsschicht angesiedelt. Zum Versand von Nachrichten muss ein Client mittels eindeutiger ID eine Verbindung zum Broker aufbauen. Ein Broker ist ein Server und kümmert sich um den Nachrichtenversand. Nach dem Verbindungsaufbau kann ein Client (sog. Publisher) eine Nachricht auf einem Topic veröffentlichen und Topics für den Empfang von Nachrichten abonnieren (sog. Subscriber).

Im Beispiel von Abb. 1 wird vom Client „DataMining“ das Topic „gyro“ abonniert. Der Client „Notfallzentrale“ ist Subscriber des Topics „alarm“. Sobald der Client „Smartwatch“ die Daten eines Gyrosensors (x,y,z) auf dem Topic „gyro“ veröffentlicht, werden diese vom Broker an „DataMining“ gesendet. „DataMining“ analysiert die Sensordaten. Wird bspw. ein Ereignis wie Sturz erkannt, wird eine Nachricht auf dem Topic „alarm“ veröffentlicht und vom Broker an den Client „Notfallzentrale“ weitergeleitet. Zur Sicherstellung des Nachrichtenversands stellt MQTT folgende QoS-Level bereit [BG14]:

- QoS-Level 0: Versand von max. einer Nachricht („Fire and Forget“)
- QoS-Level 1: Versand von min. einer Nachricht, Duplikate möglich
- QoS-Level 2: Versand von genau einer Nachricht, keine Duplikate

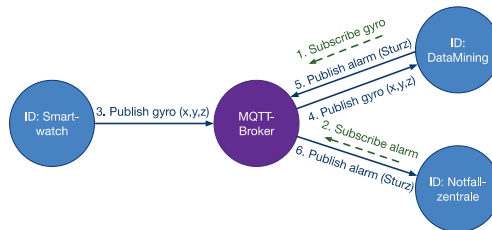


Abb. 1: Grundlegende Funktionsweise des Nachrichtenversands mit MQTT

3.3 Topics und Topic-Filter

MQTT verwendet Topics zum Filtern von Nachrichten. Ein Topic ist eine UTF8-Zeichenkette. Topics müssen mindestens ein Zeichen lang sein und können Leerzeichen enthalten. Die Topic-Namen sind frei wählbar und können zum Abbilden komplexer Strukturen hierarchisch aufgebaut werden (sog. Topic Levels). Ein Beispiel für ein Topic ist: `accelerometer/smartwatch`.

Zum Filtern von Topics stellt MQTT zwei Wildcard-Zeichen bereit. Wie in Tab. 1 ersichtlich können mittels Pluszeichen (+) eine Ebene und mit dem Rautezeichen (#) mehrere Ebenen eines Topics gefiltert werden [BG14]. Zusätzlich wird noch das \$ Zeichen für Broker interne Verwendung verwendet.

Filter	Exemplarische MQTT-Topics
<code>+/smartwatch</code>	<code>accelerometer/smartwatch</code> oder <code>gyrosensor/smartwatch</code>
<code>gyrosensor/#</code>	<code>gyrosensor/smartwatch/001</code> oder <code>gyrosensor/smartwatch/002</code>

Tab. 1: Einsatz von Wildcard-Zeichen zum Filter von MQTT-Topics

3.4 Aufbau eines MQTT-Paketes

MQTT zeichnet sich durch einen geringen Protokoll-Overhead von nur 2 Byte³ im fixen Header aus. Das erste Byte im fixen Header unterteilt sich wie in Abb. 2 ersichtlich in zwei Nibbles. Das erste Nibble dient als Indikator eines von insgesamt 14 Pakettypen. Über das zweite Nibble können zugehörige Flags wie das QoS-Level bei einem Publish-Paket übermittelt werden. Im zweiten Byte wird die verbleibende Paketlänge übermittelt. Optional können gemäß Spezifikation drei weitere Bytes für die Längenangabe verwendet werden. Abhängig vom Pakettyp sind der variable Header sowie die Payload zum Teil optional. So wird bspw. bei QoS-Level 1 und 2 der variable Header zur Identifizierung der Pakete verwendet. Bei einem Publish-Paket wird die Nachricht über die Payload übermittelt [BG14].

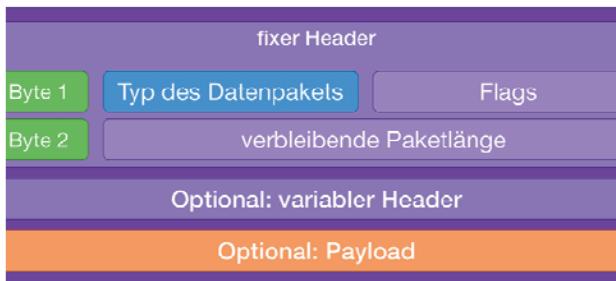


Abb. 2: Aufbau eines MQTT-Datenpakets

³ Mit 2 Byte können max. 127 Byte Payload, bei 5 Byte max. 256 MB übermittelt werden [BG14].

Kritisiert wird in diesem Zusammenhang die fehlende Möglichkeit, Metadaten zur übertragenen Payload zu übermitteln. Im Vergleich zu MQTT stellt bspw. HTTP einen Content-Type-Header für die Spezifizierung des Entity-Bodies bereit [Fi14]. Mitentwickler Stanford-Clark begründet diese Entscheidung mit „MQTT is intended as a transport, not a protocol. Its job is to get the data from one machine to another, not to define what that data is“ [Cr16]. MQTT ist damit als die Basis für ein eigenes Protokoll zu sehen, in welchem das Datenaustauschformat sowie die Semantik der Daten spezifiziert werden muss.

3.5 Mögliche Datenaustauschformate für MQTT

Das Protokoll MQTT gibt kein Datenaustauschformat vor. Dies wirft die Frage auf, welche Austauschformate sich für MQTT anbieten. Stanford-Clark nennt JSON und „Protocol buffers“ von Google (protobuf) als interessante Austauschformate für MQTT [Cr16]. Mit JSON werden Daten textuell übertragen und sind somit menschenlesbar. Mit einem binären Datenformat wie protobuf können Daten bspw. durch die Verwendung von Datentypen platzsparender übermittelt werden [PR16].

4 Architektorentwurf xHealth

4.1 Analyse

In einem möglichen Anwendungsfall empfängt xHealth die Sensordaten eines smarten Objekts über einen MQTT Broker zur weiteren Analyse mittels Data Mining, um bspw. Ereignisse wie einen Sturz zu erkennen. Bei Sturzerkennung wird ein externes System wie eine Notfallzentrale informiert. Zur Analyse sind Daten eines Gyro- und Beschleunigungssensor 245-mal pro Sekunde⁴ an xHealth übermittelt.

Mittels RESTful Services basiert auf HTTP sollen nach der Vision des WoT Sensordaten übermittelt bzw. abgerufen werden können. Hierbei ist insbesondere die Interoperabilität zum Protokoll MQTT von Bedeutung. In einem Anwendungsfall werden Temperatur- und Luftdrucksensordaten mittels MQTT an xHealth übermittelt und von einer Smartwatch über die RESTful API abgerufen. Beide Anwendungsfälle werden in Kapitel 5 exemplarisch implementiert.

4.2 Softwarearchitektur

Ziel der Softwarearchitektur ist eine möglichst unabhängige Entwicklung der Komponenten [Ru12]. Nachfolgend soll xHealth als Whitebox betrachtet und in Komponenten gegliedert werden. Für die Definition der Komponenten wird ein Use-Case-orientierter

⁴ gängige Frequenz des im Prototyping verwendeten Gyrosensors „L3GD20H“ [ST16]

Ansatz gewählt. Dieser macht davon Gebrauch, dass jeder Use-Case aus der Softwareanalyse eine entsprechende Applikationslogik benötigt [Ru12]. Hierzu wird das Layer-Pattern eingesetzt, welches die zukünftige Wartung vereinfacht und eine inkrementelle Entwicklung ermöglicht [Su12].

Auf dem obersten Layer befinden sich die der MQTT-Broker und RESTful API sowie die Benutzeroberfläche. Auf der Applikationsschicht sind sechs Module angesiedelt. Die Komponente „**Data Mining**“ erhält Daten der Schnittstellen zur Erkennung von Mustern in Bewegungsdaten. „**Workflow Management**“ dient zur Verwaltung von Workflows wie bspw. dem Nachrichtenversand bei Mustererkennung. Mittels „**Reporting**“ können bspw. erkannte Muster mittels User Interface (GUI) visualisiert werden. Ebenfalls über das GUI bedienbar sind die administrativen Komponenten. Die Verwaltung smarter Objekte erfolgt über die Komponente „**Device Management**“. Das „**Access Management**“ reglementiert die Zugriffsrechte auf die APIs. Rollen und Benutzerkonten werden im „**User Management**“ verwaltet. Die Datenhaltung erfolgt im untersten Layer. Die Speicherung von Entitäten wie Sensordaten soll in einer **Datenbank** erfolgen. Mittels Dateisystem können Daten wie Konfigurationen gespeichert werden.

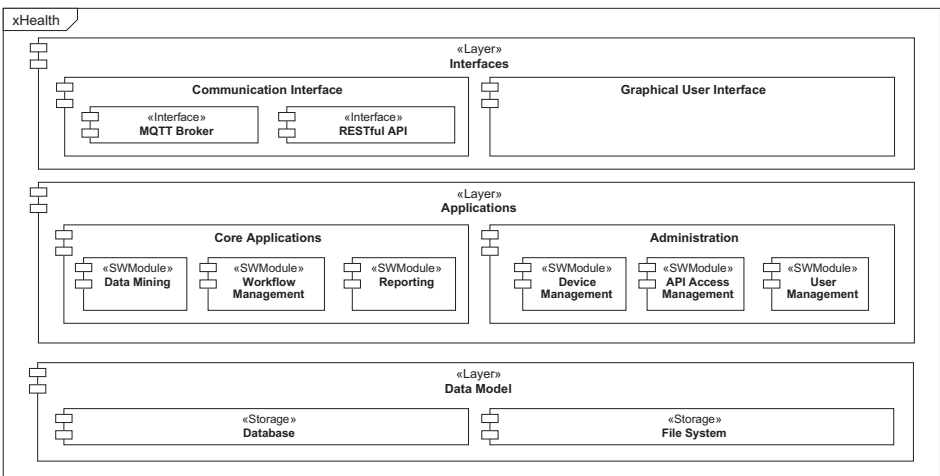


Abb. 3: Komponenten von xHealth sowie die Konnektivität zu weiteren Systemen

Aus technologischer Sicht eignet sich für die Realisierung bspw. Open Source Software basierend auf einer MEAN-Architektur. Der MEAN-Stack besteht aus der NoSQL-Datenbank MongoDB, Express als Webserver-Framework, AngularJS Frontend-Framework und Node.js als Serverplattform [Ha14]. Vorteile einer MEAN-Architektur liegen vor allem in der Skalierbarkeit und dem nichtblockierenden Zugriff auf I/O-Ressourcen durch asynchrone JavaScript-Programmierung [Sn16]. Weiter hat sich MongoDB als Speicher für Sensordaten und Big Data Anwendungen bewährt. So setzt bspw. Bosch für die IoT Cloud auf eine MongoDB basierte Architektur [BO14].

MQTT-API

Zum Nachrichtenversand mittels MQTT wird ein Broker benötigt. Mosquitto ist ein Open Source Broker basierend auf der Programmiersprache C, welcher im Rahmen des Projekts „Eclipse IoT“ entwickelt wird. Der Broker zeichnet sich durch einen besonders schonenden Umgang mit Ressourcen aus. So beanspruchen bspw. 1000 verbundene Clients lediglich 3 MB RAM [EC16a].

Sensordaten werden von smarten Objekten über folgendes Topic-Pattern veröffentlicht: `<api-version>/<device-id>/<sensortyp>`

In der ersten Ebene wird eine API-Version übermittelt, um nicht abwärtskompatible Änderungen an der API vornehmen zu können. In der zweiten Ebene wird die Device-ID übermittelt. Das dritte Level dient der Zuordnung unterschiedlicher Sensortypen. Weiter existiert ein Rückkanal, um Daten von xHealth an das jeweilige smarte Objekt zu versenden. Hierzu wird die Ebene der Device-ID verwendet.

Topic	Beschreibung
v1/123/gyr	Gyrosensor-Daten von Device 123
v1/123/acc	Beschleunigungssensor-Daten von Device 123
v1/123/tem	Temperatur-Daten von Device 123
v1/123/pre	Luftdruck-Daten von Device 123
v1/123	Rückkanal für Device 123

Tab. 2: Topic-Struktur von xHealth am Beispiel von Device 123

Zur Spezifizierung der Semantik der zu übermittelten Daten soll das binäre Format `protobuf` verwendet werden. Tab. 3 zeigt einen Versuch, in welchem ein Datensatz eines Gyrosensors bestehend aus x-, y- und z-Werten in der Einheit rad/s mit verschiedenen Datenformaten übermittelt und die Datenmenge gemessen wurde. Mit `protobuf` können Daten platzsparender als mit JSON oder CSV übermittelt werden.

Datenformat	Typ	Payload	Datenmenge
CSV	textuell	-0.02,-0.09,-0.01	17 Byte
JSON	textuell	{"x":-0.02,"y":0.09,"z":-0.01}	30 Byte
<code>protobuf</code>	binär	0d0ad7a3bc15ec51b8bd1d0ad723bc	15 Byte

Tab. 3: Vergleich der Datenmenge am Beispiel eines Gyrosensor-Datensatzes

Die Validierung der Daten erfolgt durch die Definition von Schemata, welches von jedem Kommunikationspartner zu implementieren ist. Zugriffsklassen für eine spezifische Programmiersprache können mittels Compiler erzeugt werden.

Zugriffsklassen für die Verwendung eines Schemas mit unterschiedlichen Programmiersprachen wie C++ oder Java können mittels Compiler erzeugt werden [PR16]. Mit JavaScript kann mittels dem Modul `protobuf.js` ein `protobuf` Schema ohne Einsatz eines Compilers verwendet werden [Wi16].

5 Anwendungsfälle für xHealth

5.1 Prototypische xHealth basierte Wetterstation

Die prototypische Implementierung von xHealth erfolgt auf einer Amazon EC2-Instanz vom Typ „t2.micro“ mit Ubuntu 14.04 LTS. Zur Realisierung der MQTT- sowie RESTful API wird das IoT-Framework Node-RED basierend auf Node.js und der MQTT Broker Mosquitto eingesetzt.

Abb. 4 zeigt eine App auf einer Smartwatch, welche Daten eines Temperatur- und Luftdrucksensors mittels HTTP-Methode GET über die RESTful API von xHealth abrufen und auf dem Display darstellt. Diese Daten wurden zuvor von einem Arduino Uno ausgelesen und mit dem MQTT-Client „PubSubClient“ an xHealth übermittelt [On16]. Der Nachrichtenversand mittels MQTT erfolgt über die Topics in Tab. 4. Für die Entwicklung der App wird das Garmin Connect IQ SDK sowie die Eclipse IDE eingesetzt [GA16]. Die Programmierung erfolgt in „Monkey C“. Diese Sprache wurde für Wearables entwickelt und weist Ähnlichkeiten zu Java, Ruby und Python auf [GA16]. Als Basis der eigenen Entwicklung dient die mit dem SDK gelieferte App „JsonRequest“.

Topic	Beschreibung
v1/001/tem	Temperatur-Daten von Gerät 001 (Arduino)
v1/001/pre	Luftdruck-Daten von Gerät 001 (Arduino)

Tab. 4: MQTT-Topics des Arduino Uno

Die Übermittlung der Daten erfolgt mit einem protobuf-Schema.

```
message Weather{
  required float temperature = 1;
  required float pressure = 2;}
```

Nachfolgendes Beispiel zeigt einen JSON-Response mittels dem Tool „curl“.

```
curl -X GET http://api.xHealth.io/v1/001/wetter
{"temp":21.60,"pressure":967.309998}
```

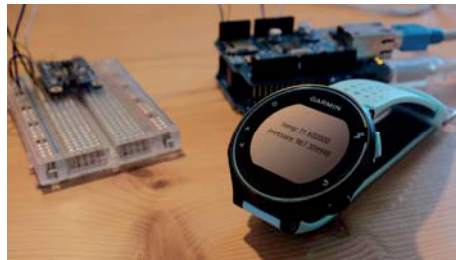


Abb. 4: Abruf von Meteodaten über xHealth auf einer Garmin Forerunner 235 Smartwatch

5.2 Smartwatch basierte ADL/EDL Erkennung

Die Erkennung von Aktivitäten des täglichen Lebens (ADLs) wie Zähneputzen, Rasieren, Trinken, Essen, Kämmen, Schreiben etc. sowie Ereignisse des täglichen Lebens (EDLs) wie Aufstehen, Zubettgehen, aber auch Stürze können helfen, frühzeitig Veränderungen im Verhalten von Personen (z.B. in der Agilität etwa bei altersbedingter Demenz) bzw. kritische, gesundheitsgefährdende Situationen wie Dehydrierung, Sturz, Desorientierung zu erkennen und ggf. notwendige Gegenmaßnahmen zu ergreifen [LW15a], [LW15b], [WA15]. Smartwatches sind ideale Werkzeuge, um die dafür notwendigen Daten erfassen und diese Aktivitäten erkennen und analysieren zu können [BA15], ohne dass deren Träger durch den Einsatz solcher Technologien stigmatisiert werden, wie es bei den derzeit am Markt erhältlichen Produkten der Fall ist [LW15b]. [Ba15] und [LW15] verwenden verschiedene Methoden des Data Mining um aus Sensordaten ADLs in Realzeit erkennen lassen können. Dazu werden verschiedene Modelle entwickelt und trainiert (logistische Regression, neuronale Netze, NMM und Entscheidungsbäume). Mit diesen Modellen werden zur Laufzeit die Sensordaten überwacht und die entsprechenden ADLs/EDLs ermittelt. Ziele von [LW16] ist es eine Notfall-App zu entwickeln, die den Träger im Alltag und in Notfallsituationen unterstützt. [15c] beschreibt einen Architekturvorschlag, wie Anwendungen auf Smartwatches modular gestaltet werden können, sodass sie den Anforderungen. Die App läuft auf einer Samsung Gear S und ist in Tizen mit JavaScript realisiert.

Die Erkennung der ADLs/EDLs erfolgt in mehreren Schritten (Abb. 5):

- 1) Die Sensordaten werden in 1-Sekunden oder 5-Sekunden-Abständen zu 10-Sekunden-Blöcken zusammengefasst,
- 2) aus diesen 10-Sekunden-Blöcken werden statistische Parameter wie Mittelwert, Standardabweichung, Interquartilsabstand, Nulldurchgänge etc. berechnet und
- 3) die Parameter werden auf das entsprechende trainierte Modell angewandt und die passende ADL/EDL ermittelt.

Bei der Erkennung von ADLs unmittelbar auf der Smartwatch ist insbesondere die Ermittlung der statistischen Parameter sehr rechenaufwändig. Es müssen für das beschriebene 10-Sekunden Auswertintervall etwa 4000 Sensorwerte kontinuierlich alle 1 oder 5 Sekunden je nach Implementierung verrechnet werden. Wie verschiedene Experimente, u.a. auf einer Samsung Gear Live Android Wear Implementierung in Java zeigt, überfordert dies die Hardware der Smartwatch trotz verschiedener Optimierung der statistischen Routinen. Ähnliches gilt für die Samsung Gear S mit Tizen und JavaScript als Implementierungssprache. Nach etwa 30 – 40 Sekunden werden die Berechnungen nur mehr stark zeitverzögert durchgeführt, es beginnen Intervalle zu fehlen. Um trotzdem eine Realzeiterkennung zu ermöglichen, wurde in einem ersten Versuch die Auswertung auf einen Server verlagert, auf dem insbesondere das Data Mining Tool RapidMiner zur Anwendung kommt. Wie in Abb. 6b) dargestellt, wurde dazu ein entsprechender HTTP/REST basierter Service mit JSON als Austauschformat aufgesetzt. Erste Versuche

zeigen, dass damit Realzeit-Analysen möglich sind, aber durch die Verwendung von HTTP einen entsprechenden großen Overhead erzeugt wird. Dieser fällt bei einer Verwendung im häuslichen WLAN mit entsprechenden Übertragungsgeschwindigkeit kaum ins Gewicht, bei schlechten Verbindungsqualitäten kann es aber zu merklichen Verzögerungen kommen. Gerade außer Haus hier ist aber eine rasche Reaktion notwendig.

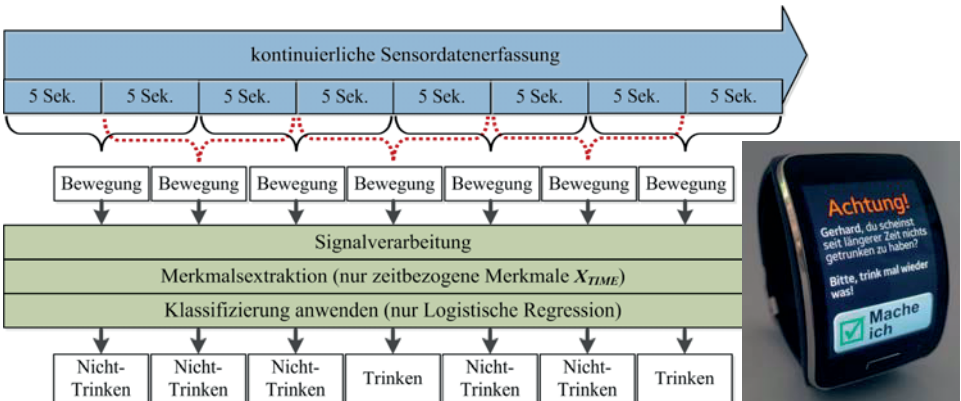


Abb. 5: Datenerfassung und Realzeiterkennung von ADLs/EDLs [WB15], [WL16]

MQTT kann hier Verbesserungen bringen: a) durch den geringeren Protokolloverhead, b) durch eine entsprechende Komprimierung der übertragenen Daten, etwa wie beschrieben mittels protobuf und c) durch die Möglichkeit einen entsprechenden QoS Level zu wählen. So bietet sich hier QoS-Level 0 an, da bei der großen Anzahl übertragener Daten einige fehlende Sensordaten die ADL/EDL-Erkennungsraten kaum beeinflussen. Analog wäre auch der QoS-Level 2 denkbar, um die multiple Übertragung von Daten zu verhindern.

Ein mögliches Einsatzszenario der Architektur wird in Abb. 6 beschrieben. Am zentralen xHealth Broker können sich mehrere Smartwatches und Anwendungen anmelden und unterschiedliche Topics subscribieren bzw. publizieren (sensor, adl, Tab. 5). [Wn]-Smartwatches werden von betreuungsbedürftigen Personen, deren ADLs erfasst werden sollen, getragen. Sie publizieren Sensordaten, die vom ADL BDA Server subscribiert werden. Dieser führt die ADL/EDL Erkennung durch und publiziert seinerseits die erkannten ADLs. Die $W_1 \dots W_n$ Smartwatches werden über ihre ADL Subskriptionen informiert, etwa um im Falle eines Sturzes noch entscheiden zu können, ob eine Benachrichtigung der Notrufzentrale erfolgen soll. Parallel dazu kann der Träger der Smartwatch C1, z.B. eine Betreuungsperson, über den Sturz über die ADL Subskription informiert werden, ebenso wie die Notfallzentrale.

Im Rahmen eines ersten Proof of Concepts (Abb. 7) konnte gezeigt werden, dass mit diesem Ansatz problemlos die erzeugten Mengen an Sensordaten übertragen, am Server verarbeitet und erkannte ADLs an die Smartwatches übertragen werden können.

6 Untersuchung der Effizienz

Zur Ermittlung des Datendurchsatzes werden zwischen dem Arduino Uno und xHealth jeweils 10.000 Pakete in der Reihenfolge 10, 50 und 100 Byte Payload übermittelt. Der Nachrichtenversand mittels MQTT erfolgt unter Verwendung von QoS-Level 0 und dem MQTT-Client „PubSubClient“ [On16]. Die Übermittlung der Daten an die RESTful API erfolgt mittels HTTP-Methode „POST“.

Abb. 8 zeigt mittels Wireshark IO-Graph die zuvor mit tshark serverseitig aufgezeichneten Netzwerkdaten der Protokolle MQTT und HTTP.

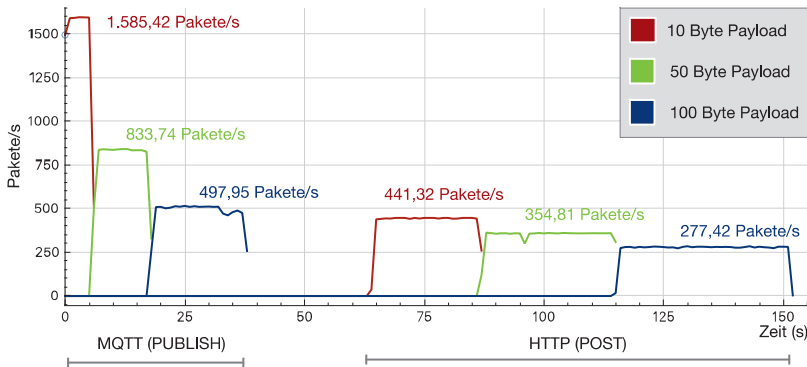


Abb. 8: Analyse der von xHealth empfangenen Pakete mit Wireshark

Der Arduino Uno arbeitet mit einer Taktfrequenz von 16 MHz basiert auf einer 8 Bit Architektur und verfügt über 2 KB RAM, was die Implementierung von TCP/IP sowie die maximale Übertragungseinheit beim Datenaustausch limitieren. Im Vergleich hierzu arbeiten aktuelle Smartwatches mit bis zu vier Prozessorkernen basiert auf einer 32 Bit Architektur mit einer Taktfrequenz von mehr als einem GHz.

7 Fazit

Basiert auf der prototypischen Implementierung von xHealth sowie der Untersuchung der Effizienz von MQTT wird folgendes Fazit gezogen:

- MQTT und xHealth bieten eine interessante Möglichkeit die Sensordaten auf einem Server auswerten zu lassen und den Tradeoff zwischen beschränkter Smartwatch Hardware und Erkennungsgeschwindigkeit zu optimieren.
- Das MQTT Publish/Subscriber Modell von xHealth bieten für die ADL/EDL Erkennung und Einbindung in ein Gesamtsystem zur Notfallunterstützung eine solide Grundlage für eine einfach erweiterbare Kommunikationsarchitektur.

- Die Interoperabilität der übermittelten Daten zwischen MQTT/protobuf und HTTP/JSON konnte am Beispiel des Anwendungsfalls „Wetterstation“ aufgezeigt werden.
- Mit MQTT kann zwischen einem Arduino Uno und xHealth basiert auf einer Payload von 10 Byte mehr als der 3-fache Datendurchsatz erreicht werden.
- Beim Datenaustausch über MQTT kann durch den Einsatz von protobuf die Datenmenge der zu übermittelnden Payload im Vergleich zu JSON am Beispiel der Daten eines Gyrosensors um etwa 50 % reduziert werden.

Literaturverzeichnis

- [Ba15] R. Baldauf, "Mobile sensorbasierte Erkennung von Trinkbewegungen," Thesis, FOM, Nürnberg, 2015.
- [BO14] Bosch Si Whitepaper: IoT and Big Data, www.bosch-si.com/lp/iot-big-data.html, Stand: 15.05.16.
- [BG14] Banks, A.; Gupta, R.: MQTT Version 3.1.1. OASIS, 2014.
- [Cr16] Craggs, I.: Why doesn't MQTT have a payload format, www.modelbasedtesting.co.uk/?p=243, Stand: 02.04.2016.
- [EC16a] Eclipse Mosquitto, projects.eclipse.org/projects/technology.mosquitto, Stand: 15.05.2016.
- [EC16b] Eclipse Paho - Open Source messaging for M2M, www.eclipse.org/paho, Stand: 20.04.2016.
- [Fi14] Fielding, R. et.al.: Hypertext Transfer Protocol – HTTP/1.1 (RFC: 7230). Internet Engineering Task Force, 2014.
- [GA16] Garmin Developer – connect IQ SDK, developer.garmin.com/connect-iq/overview/, Stand: 15.05.2016. [GT15] Guinard, D.; Trifa, V.: Building the Web of Things - ME-AP. Manning Publications, New York, S. 23, 2015.
- [Ha14] Haviv, A.: MEAN Web Development. Packt Publishing Ltd, Birmingham, S. 7, 2014.
- [HI16] HiveMQ – MQTT 101, www.hivemq.com/blog/how-to-get-started-with-mqtt, Stand: 02.05.2016.
- [IS16] ISO/IEC 20922 - Message Queuing Telemetry Transport (MQTT) v3.1.1, www.iso.org/iso/catalogue_detail.htm?csnumber=69466, Stand: 21.04.2016.
- [KK16] Klose, B.; Klose, H.: Meteorologie: eine interdisziplinäre Einführung in die Physik der Atmosphäre. Springer-Verlag, Berlin, S. 3, 2016.
- [LW15a] Lutze, R., Waldhör, K., Baldauf, R.: Dehydration Prevention and Effective Support of Elderly by the Use of Smartwatches, in Proceedings of the IEEE Healthcom 2015, 2015.

- [LW15b] Lutze R., Waldhör K.: “SmartWatches als Hausnotrufsysteme der nächsten Generation,” in 8. AAL Kongress - Zukunft Lebensräume 2015, Berlin: VDE Verlag, 2015.
- [LW15c] Lutze R., Waldhör K.: A Smartwatch Software Architecture for Health Hazard Handling for Elderly People, in ICHI 2015, 2015, pp. 356–361, 2015.
- [LW16] Lutze R., Waldhör K.: Integration of Stationary and Wearable Support Services for an Actively Assisted Life of Elderly People: Capabilities, Achievements, Limitations, Prospects - A Case Study, In: Wahlster W (Hrsg.) Zukunft Lebensräume 2016, Springer, Frankfurt, 2016.
- [MO16] Mosca – MQTT broker as a module, www.mosca.io, Stand: 07.05.2016.
- [On16] O’Leary, N.: PubSubClient - Arduino Client for MQTT, pubsubclient.knolleary.net, Stand: 26.04.2016.
- [Pe16] Pettey, C.: Smarter With Gartner: The Internet of Things and the Enterprise, www.gartner.com/smarterwithgartner/the-internet-of-things-and-the-enterprise, Stand: 08.05.2016.
- [PM16] PM2 - Advanced Node.js process manager, www.pm2.keymetrics.io, Stand: 08.05.2016.
- [PR16] Protocol Buffers, developers.google.com/protocol-buffers, Stand: 15.05.2016.
- [Sn16] InfoWorld - Why Node.js beats Java and .Net for Web, mobile, and IoT apps, www.infoworld.com/article/2975233/javascript/why-node-js-beats-java-net-for-web-mobile-iot-apps.html, Stand: 15.05.2016.
- [ST16] STMicroelectronics - L3GD20H Datasheet, www.st.com/st-web-ui/static/active/en/resource/technical/document/datasheet/DM00060659.pdf, Stand 04.04.2016.
- [Su12] Sommerville, I.: Software Engineering. Pearson Deutschland GmbH, München, S. 194, 2012.
- [Ru12] Rupp, C. et.al.: UML2 glasklar. Carl Hanser Verlag, München, S. 93, 2012.
- [W316] W3C - Bindings To Common Protocols, www.w3.org/WoT/IG/wiki/Bindings_To_Common_Protocols, Stand: 10.05.2016.
- [WB15] Waldhör, K., Baldauf, R.: Recognizing Drinking ADLs in Real Time using Smartwatches and Data Mining. In S. Fischer, I. Mierswa, & G. Schäfer (Eds.), Proceedings of the RapidMiner Wisdom Europe (2015) (pp. 1–18). Aachen: Shaker., 2015
- [Wi16] Wirtz, D. et.al.: GitHub - protobuffs, github.com/dcodeIO/protobuf.js, Stand: 15.05.2016.
- [WL16] Waldhör, K., Lutze, R., Baldauf, R.: ADL Erkennung mit Smartwatches: Data Mining zur Realzeiterkennung von Aktivitäten des täglichen Lebens (ADL). In K. Waldhör & R. Lutze (Eds.), wearables /at /work. Konferenz: wearables \at \work 2016. 17.02.2016, München, Nürnberg: FOM, 2016.
- [Zh16] Zhang, L.: Building Facebook Messenger, www.facebook.com/notes/facebook-engineering/building-facebook-messenger/10150259350998920, Stand: 22.03.2016.

The Application Architecture of Smartwatch Apps – Analysis, Principles of Design and Organization

Rainer Lutze¹, Klemens Waldhör²

Abstract: We present an innovative approach for the systematic construction of smartwatch apps starting from an analysis of the requested functionality. Based on the resulting functional requirements, we recommend to derive a suitable smartwatch app architecture based on the following *five principles*: 1) the need for dialogue management, 2) the need for a distributed architecture, 3) the advantages of a layered app architecture, 4) software engineering implications for the maintainability of the app with economic costs, and 5) the technical boundary conditions on the causal network of the app. The approach taken results from - and is backed by - our two year experiences on building smartwatch apps in the healthcare area.

Keywords: software architecture for smartwatch apps, principles of software construction for smartwatch apps, software engineering of smartwatch apps

1 Known Architectural Approaches

Apple watch apps in principal follow the IOS™ MVC architectural pattern, where the “view” part is dislocated on the watch and the “model” and “controller” portion of the app are running on a coupled iPhone™. The MVC (model view controller) is an approved app design pattern for separating the graphical user interface (GUI) of the app, *view*, from the domain and task specific logic of data manipulation, *controller*, and the persistent storage of data, the *model*. For alleviating the GUI implementation for their distributed apps for Apple watch, watchOS™ features three design patterns for presenting information on the Apple Watch screen ([Ap16]): (*readonly*) *glances*, *actionable notifications* and *complications*. For now, Apple does not support smartwatch apps running independently on the smartwatch.

In contrast, Samsung’s Tizen OS™ and Gear™ smartwatches feature standalone apps. For Apple’s watchOS™ the communication between smartwatch and smartphone works fully automatically (via Bluetooth). For Tizen, the Samsung Accessory Pack (SAP) (cf. [Sa16]) supports a software distribution between smartphone and smartwatch and the corresponding communicating between both devices (via Bluetooth), but it has to be explicitly programmed.

¹ **Dr.-Ing. Rainer Lutze Consulting**, Wachtlerhof, Wilhermsdorfer Str. 14, D90579 Langenzenn, Germany
rainer.lutze@lustcon.eu

² **FOM** University of Applied Sciences, D45127 Essen / Nuremberg, Germany
klemens.waldhoer@fom.de

Android Wear™ OS (cf. [An16a]) supports similar functionality. Programmable notifications based on the Bluetooth protocol are used to exchange data between the smartwatch and the mobile phone. This includes the option to distinguish between actions which are identical both on the mobile phone and the smartwatch and actions where the functionality may differ. In addition, Android Wear supports starting actions of other smartwatch and mobile phone apps using Intents. From a GUI point Android distinguishes three types, so called cards: 1) Cards for displaying information from notifications, typically sent from the mobile phone, 2) single action controls (play, pause) and 3) expandable stacks of cards, combing the information of several notifications. Cards themselves may be composed of several pages. The design of the cards and the app itself should follow the Android Wear design principles (cf. [An16b]). For handling background services Android uses the Java thread concept. This in turn is mostly realized through some listener functionality. Thus Android provides the developer for the construction of an app basically with the same functionality – with some minor limitations – as for an app on a mobile phone. This is one of the advantages of Android compared to other approaches, in addition its well proven and stable wearable OS.

2 Engineering of Smartwatch Apps

Following a well-established workflow in the engineering process (cf. [FG13]), in the first step the **function** of a smartwatch app has to be determined. From this, secondly, the architectural design, **form** of the product, can be derived. Finally, the relevant **technical standards and regulations** for optimizing the app will be considered and applied.

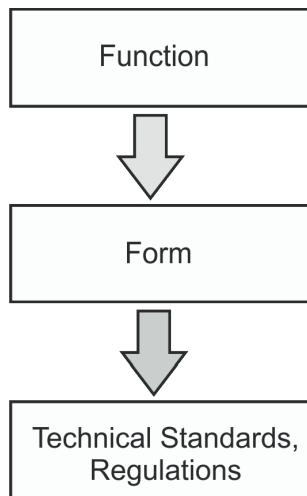


Fig. 1 Product Development Process (PDP) proposed by Pahl/Beitz

2.1 Functional Aspects of Smartwatch Apps

The typical smartwatch app will include one – or a combination – of the following *four basic functional aspects*:

- **Presenting** – the function of the app is the presentation of foreign content (not created on the smartwatch) on the body-worn device. The presentation may be controlled by manual input, typically affirmations or selections, done by the wearer on the smartwatch.
- **Controlling** – the function of the app is to manipulate resp. command foreign activities via control values originating from the smartwatch software itself or issued by the wearer of the watch.
- **Monitoring** – the function of the app is to detect activity patterns or recognize information about the health conditions of the smartwatch wearer by fusing the measured sensor data of the smartwatch. Monitoring does include a local display of such derived information on the watch and may include remote alerting in cases of detected emergency situations by deviant activity patterns.
- **Communicating** – the function of the app is to provide a real-time (chat, speech) communication between the smartwatch wearer and a distant person (or machine).

Many apps include a combination of those basic functional aspects. A popular combination is the “controlling-presenting” bundle, e.g. in a *»news«* app, where the smartwatch user selects his news categories of interest first, and then repeatedly gets latest news for the selected categories, which need to be presented in very compressed style. The “monitoring-presenting” bundle is present in *»fitness tracker«* apps, where the actual heart rate values, done steps and actual geographic position of the smartwatch wearer will be presented together in a map delivered from a foreign, central map service, e.g. Google, Here or Microsoft. Our assistance app for elderly people (cf. [LW16]) features a bundle of the “monitoring-communicating-controlling” aspects. A basic monitoring of the smartwatch wearer takes place and in case of detected emergency situations, health hazards, a speech connection to home emergency call center (HECC) will be established. Additionally, the assistance app controls a home automation (HA) system, in that the departure and arrival of the wearer from resp. at his/her home will be reported to the HA system for switching off/on critical electric appliances during his/her absence and effecting the lighting of the home.

2.2 Architectural Requirements from App Functionality - The Need for Dialogue Management

Even apps with pure presenting functionality will already run into a problem, if more than one piece of information shall be displayed simultaneously on the smartwatch screen. The same is true for monitoring apps, if multiple and different information derived from the smartwatch sensor analysis shall be presented to the smartwatch user at

the same time. A **dialogue management** will be necessary, which prioritizes the information to be displayed. The dialogue manager then exclusively assigns the smartwatch display and smartwatch input/output devices (touchscreen, mic/speaker) for a defined time slice to that smartwatch app thread with highest actual priority. For example, in [LW15] a blackboard is used as a central synchronization mechanism between all different threads which desire to gain the user's attention. A central scheduling algorithm determines the blackboard element with highest priority and passes a temporal exclusive privilege to this element for utilizing the smartwatch screen and I/O devices.

If the smartwatch app includes multiple aspects of presenting, monitoring and communicating, the dialogue manager will be inevitable for coordinating the smartwatch display as a scarce resource. Typically, outgoing emergency calls placed by the smartwatch user or incoming calls from other parties will have the highest priority for display and I/O utilization.

2.3 Architectural Requirements from App Functionality – The Need for a Distributed Software

For the current Apple Watch app architecture with its distributed MVC pattern, a distributed app implementation approach is mandatory. If for control aspects apps, monitoring or communication aspect apps complex configuration information will be necessary for the function of the app, for the input of the configuration information the smartwatch display is typically not well suited. A companion device (e.g. a coupled smartphone) will be needed, on which the configuration details can be visually presented in sufficient space and descriptiveness. Also, for the machine learning and data mining process of the sensor data captured and condensed on the smartwatch, a more powerful additional device will be needed to perform this analysis. For our assistance app for the elderly, data will be passed from the smartwatch to a powerful workstation or a cloud based service (e.g. RapidMiner), in order to perform these learning, data mining computations (cf. [Lu15], [WB15]). The model learned from the sensor data then will be retransmitted to the smartwatch app and applied there.

In addition, a highly available server may act on behalf of the smartwatch and continue and complete its operations, if, for example, the wearable accidentally should run out of power and cannot act by itself anymore. We utilize such a “long-term analysis server” (LTAS) in our assistance app for durably storing and processing data delivered from the smartwatch, especially for state-of-affairs covering a period of more than a day. With respect to the limited computational power, it might be also a reasonable decision to dislocate as much demanding computations as possible from the smartwatch to a more powerful companion device like a smartphone, server or cloud service. For example, also the needed dialogue manager may run on an accompanying smartphone, if the smartwatch is reliable coupled to the smartphone via Bluetooth in a sustainable way.

In conclusion, a software distribution of parts of the smartwatch app software to a companion device with larger screen, more powerful CPU and storage.... will be favorable for all but very simple mono-functional apps.

2.4 Architectural Requirements from App Functionality – The Advantages of a Layered App Software

We have already described the necessity of having an explicit dialogue management module for presentation control as part of multifunctional smartwatch apps. In addition to this GUI module a further modularization especially of monitoring apps is meaningful.

The monitoring task typically consist of:

- 1) the *recognition* of activity patterns and/or health conditions by sensor fusion (accelerometer, gyroscopes, magnetometer, barometer, GPS, heart rate monitor, ...)
- 2) observing a specific *sequencing or clustering* of those patterns / conditions over time.

For subtask 1), sub symbolic reasoning techniques like *neuronal networks* or *logistic regression* have proven their validity (cf. [Lu15]). For subtask 2), the course of operations of the smartwatch wearer's behavior over time can best be modeled and followed *via finite automata* (we have been using UML state machine, see [LW15]) or some variant of *business process modeling (BPM)*. With respect to the strongly different implementation techniques for subtask 1) and 2), and for minimizing dependencies and interfaces, a layered architecture for subtasks 1) and 2) is to be recommended, where the results of subtask 1), the *events and activities of daily living* (cf. [LW16]), will be used as input, atomic elements, for processing the subtask 2). The resulting "uses" hierarchy (cf. [Pa72]) of smartwatch app software layers is depicted in fig. 2.

Another advantage of a layered, modular architecture will be the possibility to execute certain layers (in our case only the bottom layer) always in the background in a guaranteed way, whereas other layers will be executed only on demand, when the smartwatch wearer interacts with his/her watch. Otherwise the execution of those layers usually will be hibernated or may be even ceased (see section 2.4) for energy saving reasons by many smartwatch operating systems.

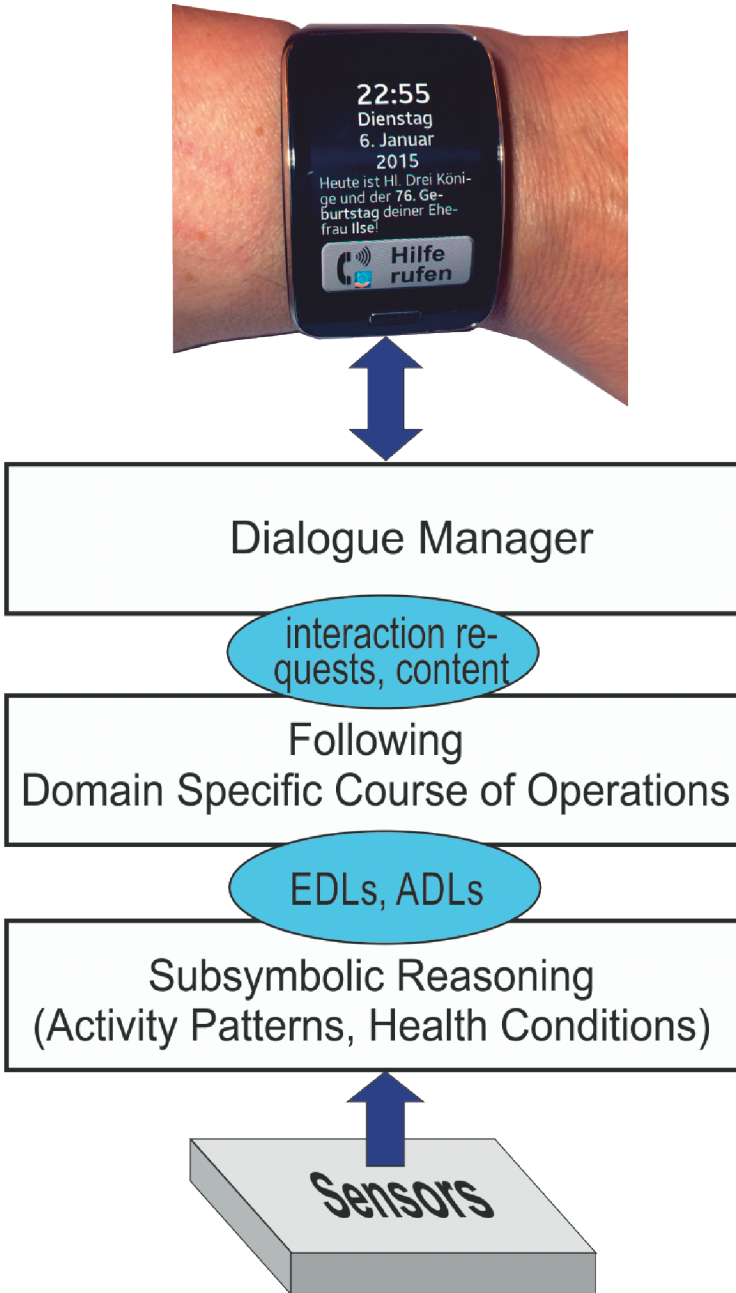


Fig. 2 : Layered Smartwatch App Software Architecture for Monitoring Functional Aspects

2.5 Software Engineering Implication for the App Architecture

Programmable smartwatches worn on the wrist are a relatively new gadget, for which use cases and application scenarios are still under development. This is especially true for business applications of smartwatches in areas as telemedicine or healthcare. As a consequence, the application logic implemented in today's smartwatch apps primarily is best practice, empirical knowledge, which grows and changes from day to day. The situation is complicated by the fact that also the smartwatch hardware typically is of the first or at most second generation and substantial changes, hopefully improvements of the hardware and/or system software have to be considered. It is therefore a real challenge, how this smartwatch application software can be *maintained with economic effort*. Another aspect concerns the *decomposition of the empirical knowledge* in independent chunks which can be considered and supported separately and in isolation from each other.

Based on the long-term experiences in other software areas, **declarative model and description techniques** are well suited for addressing this maintenance problem. For example, in our assistance app for the elderly, each individual health hazard potentially affecting an elderly person will be described by a separate UML state machine (USM). The multitude of all finite state machines will be executed simultaneously on the smartwatch and synchronized by the dialogue management in form of a central blackboard (cf. [LW15]). Fig. 5 depicts the finite state machines for monitoring the absence from home with respect to the hazards of running away from an (agreed) vicinity of the home or staying away from home for an (agreed) excess period of time. In our application domain, the discussion of the app's functionality with the care-giving domain experts takes place using the graphic state machines diagrams. From these depictions the state machine transition tables will be compiled automatically and transferred into the smartwatch. The simultaneous execution of the state machines will be done by a (static) interpreter for these (frequently changing) state transition tables.

2.6 Technical Boundary Conditions for the Causal Network of the App

A decisive factor for successful smartwatch programming is **energy conservation** of the limited capacity of the smartwatch battery. In business applications, the typical requirement is that the smartwatch battery has to last at least a full, extreme working day (let us say: 18 hours) and may have to be recharged each night. The implication from such a harsh requirement is that energy consuming sensors, communication acts and the display illumination of the smartwatch have been activated and used only as little as possible and in those limitations specified by the concise functionality the app solution has to achieve.

It is common practice for most smartwatches to turn on the display and activate the input devices only when the user turns up and rotates his wrist so that he/she can actually view the smartwatch display. Awaking the smartwatch app from a hibernation mode and

executing the smartwatch app code only in those “viewed” situations is also a standard procedure for many app operation systems. This may be well sufficient for many presentation and control functional aspects.

But, for monitoring and communication aspects an always ongoing, **sustainable background operation** of the smartwatch apps is essential and indispensable. Up to now, this is typically only guaranteed for incoming calls, incoming notifications with “breaking news” urgency and the synthetic “step counter” most wearables will provide as a synthetic sensor. Although, the demand for such reliable and steady background operations for parts of the smartwatch apps meanwhile has been acknowledged by many operation system designers. For example, the Tizen OS™ (rel. 2.3 and up) offers the concept of a “service application”, which will be immediately started at boot time of the smartwatch and will run continuously in the background.

3 Experiences with Applying the Principles

We have built our Smartwatch assistance app for elderly persons, targeting at a safe and self-determining living at home until the very high age (cf. [LW16], [LW15] for a detailed description) for the Samsung Gear™ S and S2 smartwatches following the engineering principles described above. The app focuses on the monitoring functional aspect, but also includes control and communication aspects (see fig. 3). The task to be solved by the app requires a standalone app without any dependency on a smartphone carried along (which might have been forgotten by the members of the intended user group).



Fig. 3: The implemented assistance app for Samsung Gear™ S smartwatch

3.1 The App Architecture in Detail

Distributed Architecture. Due to the fact that the Gear S features a large 2 inch, 360x480 pixel AMOLED display it was possible to integrate also the *configuration*

editor for the individual data of the wearer and his/her family data into the smartwatch app and to renounce a distributed architecture. For the follow-up model S2 with its only 1.2 inch 360x360 pixel circular display, an app external configuration editor is mandatory, realized either by a companion app running on a coupled Android smartphone or a web browser based solution. This is currently being implemented.

Layered Architecture. We implemented the proposed 3-layer structure:

- The **bottom layer** contains in fact 2 sub-layers with a neuronal net (NN) structure for ADL, EDL recognition by sensor fusion (see fig. 4). The NN output is used by the upper sub-layer for:
 - *aggregating ADLs* which are defined by a starting and terminating EDL as endpoints (cf. /LW16/ for details),
 - calculating the *wellbeing function* for the smartwatch wearer, consisting of the three wellbeing aspects of *inactivity* between recognized ADLs, the *excess duration* of ADLs and the *agility* of the smartwatch wearer (cf. /LW16/ for a definition of this wellbeing function),
 - updating the *nominal values* (by time series analysis) which are necessary parameters of the wellbeing calculation (cf. /LW16/ for details),

Additionally, the communication module for handling 3G speech connections with the HECC and http communication with the LTAS (for reporting occurred ADLs) is located on this layer.

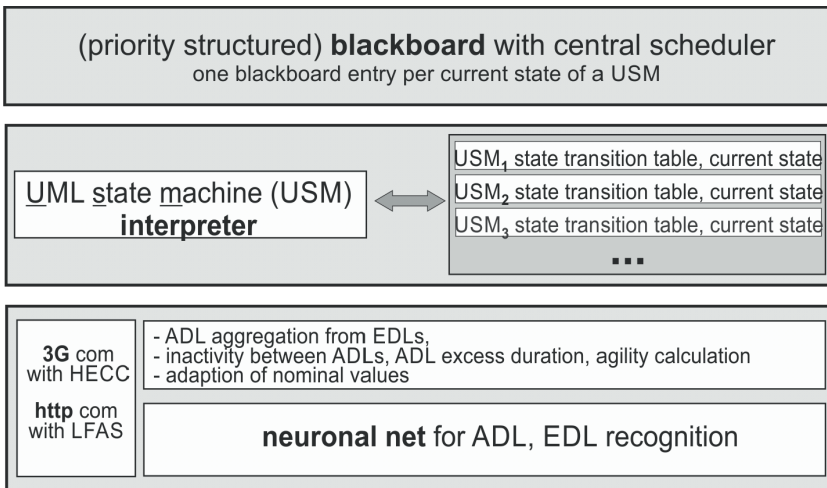


Fig. 4 The implemented 3-layer app architecture for the assistance app

- The **medium layer** contains a) the interpreter for the USMs and b) the state transition tables and current state of the corresponding USMs. Each USM models and describes the handling of a specific health hazards over the course of time (e.g. fig. 5). With each state of an USM typically an output to be displayed on the smartwatch screen is associated [as “do” action as long as the current persists], and 3G, http communication acts with the HECC and/or LTAS to be performed [as “entry”, “exit” actions of the state transitions]. USM states also have a priority attached (cf. fig 5 and /LW15/).
- The **upper layer** contains the dialogue manager implemented as a priority structured blackboard (BB) ([En89]). Each USM pushes its current state on the specific BB level associated with the state’s priority. The BB scheduling algorithm then selects a suitable USM state from those with highest priority on the BB, for which the associated output will actually show up on the smartwatch screen in a specific situation (cf. /LW15 for details).

3.2 Software Engineering Aspects

The implemented layered architecture proved its advantages especially in effectively organizing the implementation work. Following the layer and module structure, the work was assigned to and distributed between different persons / teams with clearly defined interfaces and team responsibilities.

The declarative approach of describing the course of operations for health hazard handling by USMs demonstrated its value in that the graphical depictions of the state machines (“*state machine diagrams*”) could be utilized as a common basis for discussion with the healthcare domain experts about the required behavior of the app.

3.3 Technical Limitations and Aspects

Several limitations in the implementation of the required app features have been evident. The most urgently needed stable background operations are not available at all on the Gear S model, and are only available for *native* (i.e. “C” programmed) apps on the S2, not the *wearable* (i.e. Java Script programmed) app variant we have been utilizing so far. This is a severe drawback, in that the app temporarily ceases its operation due to hibernation in a way which cannot be influenced by the app. Only a manual input operation of the smartwatch wearer will reliably terminate a hibernation phase and will bring back the app into operation. But again, the layered architecture delivers a blueprint to convert exactly the bottom layer of the app into a TIZEN *native service application*, in order to get the EDL, ADL recognition dependably and uninterruptedly working on the S2 smartwatch.

The energy consumption control of special smartwatch sensors needs also to be improved. Especially for the high power consuming GPS sensor, this sensor can only be

influenced indirectly on the Gear devices by not using it, not by directly switching it off/on, which would be strongly preferred due to a much better efficiency.

As far as technical standards and regulations are concerned, for future speech communication a SIP protocol stack (and support for related IP protocols like RTSP) would be desirable besides HTTP, in order to realize standard VOIP communication and to support – in our domain of interest - the SCAIP protocol for social care alerts via IP (cf. [SI14]).

4 Conclusions / Future Work

The proposed engineering principles have proven their validity for our application case. But, in general, the systematic construction and the software engineering of smartwatch apps still is in *statu nascendi*. The current focus is on the principal realization of required functionalities and economic maintenance costs. For the future, as soon as these issues will have been sufficiently solved, the focus will shift and include the formal validation and verification of the smartwatch apps.

With their plenitude of current and future sensors (e.g. for blood pressure measurement, c.f. [LW16]), smartwatch apps will be suited for medical diagnosis and thus may turn into a regulated medicine product in the long-term. Even if they will run on a repeatedly, yearly updated smartwatch consumer hardware will have to cope also with non-reliable / imprecise casual sensor readings. This will put much more emphasis on the structure, principles and methods by which the app has been built, in in order to verify its function, as requested for this product category.

Also privacy concerns and the potential use of smartwatch data in the scope of benefit plans for insurances issues in the long-term do require well established engineering principles and standards for constructing smartwatch apps. Only in this way, a auditable transparency of the smartwatch app software will be achieved, which allows to verify that such most sensitive personal data will be handled in a ethic and legally adequate manner.

5 References

- [An16a] Android Wear, <http://developer.android.com/guide/index.html>, retrieved May 15,2016
- [An16b] Android Wear Design Principles: <http://developer.android.com/design/wear/index.html>, retrieved May 15,2016
- [Ap 16] Apple - App Programming Guide for watchOS - The Watch App Architecture, <https://developer.apple.com/library/watchos/documentation/General/Conceptual/WatchKitProgrammingGuide/Designing-aWatchKitApp.html>, retrieved May 15,2016
- [En89] Englemore, R. “Blackboard Systems”, Addison-Wesley Publishers, 1989

- [FG13] Feldhusen, J. and Grote, K.H. (eds.): “Pahl/Beitz – Konstruktionslehre – Methoden und Anwendung erfolgreicher Produktentwicklung”, 8. Auflage, Springer Vieweg Publishers, Berlin Heidelberg 2013
- [Lu15] Lutze, R., Baldauf, R. and Waldhör, K.: “Dehydration Prevention and Effective Support of Elderly by the Use of Smartwatches”, Proc. IEEE HealthCom Conference, 14-17.10.2015 Boston/USA, pp. 404-409
- [LW16] Lutze, R., and Waldhör, K.: “Integration of Stationary and Wearable Support Services for an Actively Assisted Life of Elderly People: Capabilities, Achievements, Limitations, Prospects – A Case Study”, Proc. Future Living Spaces Conference / Zukunft Lebensräume 2016, 20.-21.4.2016, Frankfurt am Main, VDE Press and Springer Publishers, Berlin Heidelberg (in Press)
- [LW15] Lutze, R., and Waldhör, K.: “A Smartwatch Software Architecture for Health Hazard Handling for Elderly People”, IEEE International Conference on HealthCare Informatics (ICHI), 21.-23.10.2015, Dallas/USA, pp. 356-361
- [Pa72] Parnas, D.L.: „On the Criteria to be used in Decomposing Systems into Modules“, Communications of the ACM 15 (1972), No.12, pp.1053–1058
- [Sa 16] Samsung – Samsung Gear Application Programming Guide, Version 1.0, Samsung Electronics, 2014, http://img-developer.samsung.com/contents/cmm/Samsung_Gear_Application_Programming_Guide_1.0.pdf, retrieved May 15,2016
- [SI14] Swedish Standards Institute: Digital social alarm - Social care alarm internet protocol (SCAIP) – Specification- Svensk Standard SS91100:2014
- [WB15] Waldhör, K. and Baldauf, R.: “Recognizing Trinking ADLs in Real Time using Smartwatches and Data Mining”, in press, Proc. Rapid Miner Wisdom / Europe Conference, 31.8.2015 Ljubljana/Slovenia

Situationsbewusste, patternbasierte Adaption interaktiver Anwendungen durch Auswertung von Emotions- und Biosignal-Daten

Christian Märtin¹, Christian Herdin¹, Sanim Rashid¹

Abstract: Modellbasierte, durch patternbasierte Werkzeuge und Generatoren bereicherte Entwicklungsumgebungen verbessern die Modellierung, den teilautomatischen Entwurf und die Implementierung interaktiver Systeme. Um die situations- und kontextbewusste Adaption der mit solchen Umgebungen erzeugten Software zur Laufzeit zu erlauben, nutzt der vorgestellte Ansatz visuelle Analysemodalitäten wie Eye-Tracking und Blickverfolgung und koppelt diese mit Signalen zur Emotionserkennung. Eine bereits erprobte Systemarchitektur, die zum Emotionstracking die Facereader-Software benutzt, soll durch eine Wearable-Schnittstelle zur Erfassung von nicht-visuellen Biosignalen eine Fusion der Emotionsdaten ermöglichen und eine noch präzisere Adaption der Zielsoftware erlauben.

Keywords: MB-UIDE, HCI-Patterns, Situationsanalyse, Emotionsanalyse, Biosignale, Wearables, Eye-Tracking

1 Einführung und Stand der Forschung

Modellbasierte Entwicklungsumgebungen für interaktive Systeme (MB-UIDEs) sind seit Jahrzehnten in der HCI-Forschung etabliert und können für die systematische Entwicklung qualitativ hochwertiger multimedialer Web- und Desktop-Applikationen oder Apps eingesetzt werden [MC14]. Durch die Integration von Software-Patterns und Pattern-Sprachen in diese Umgebungen konnte in den vergangenen Jahren die Flexibilität der Werkzeugumgebungen erhöht und die Voraussetzung geschaffen werden, um die entstehenden interaktiven Applikationen an wandelnde Umgebungs- und Geräte-Kontexte (Desktop, Tablet, Smartphone, Wearable) und Benutzerbedürfnisse (z.B. barrierefreie Schnittstellen, demographischer Faktor) anzupassen. Mit dem PaMGIS-Framework [EM09], das dem CAMELEON-Referenzmodell [CC02] folgt und in dessen Zentrum ein reichhaltiges Pattern-Repository mit domänen-abhängigen und unabhängigen HCI-Patterns steht, wurde ein leistungsfähiger Ansatz entwickelt, um mit Hilfe der Modellierungssprache PPSL die Struktur, das Verhalten, die Präsentation und die Kontexteinbettung interaktiver Systeme zu spezifizieren und aus den Modellen die Zielsoftware teilautomatisch und mit Werkzeugunterstützung zu generieren [EM16].

¹ Hochschule Augsburg, Fakultät für Informatik, An der Hochschule 1, D86161 Augsburg, Deutschland
{Christian.Maertin, Christian.Herdin, Sanim.Rashid}@hs-augsburg.de

Die zunehmende Dominanz mobiler Systeme für die alltägliche Nutzung und die rapide fortschreitende Marktdurchdringung der Wearables führt durch deren Flexibilität und immer natürlichere Interaktionsfähigkeit mit dem Nutzer zu Möglichkeiten, situations- und kontextabhängige Informationen in die Ausgestaltung der interaktiven Software-Applikationen einzubeziehen. Dadurch ergeben sich neue, teilweise komplexe zusätzliche Anforderungen an die Entwicklungsumgebungen für solche Wearable-Applikationen. Insbesondere sollten geeignete Ansätze neben der Automatisierbarkeit bestimmter Entwicklungsschritte auch die Adaptionfähigkeit der Zielsoftware, also die Anpassung der Softwareeigenschaften der Benutzungsschnittstelle, und über die Anbindung an die Business-Logik auch der inhaltlichen Software-Eigenschaften ermöglichen.

Grundlage dafür sind die simultane Analyse des aufgabenbezogenen, emotionalen, gesundheitlichen und mentalen Benutzerstatus zur Laufzeit und die gleichzeitige Beobachtung weiterer situativer und kontextueller Daten.

In [Ch16] werden die Grundlagen für ein auf der Analyse auch des verborgenen mentalen Zustands basierendes stark benutzerzentriertes und situationsbezogenes Software Engineering gelegt, das als *Situation Analytics* bezeichnet wird. Chang regt an, die Fortschritte der Kognitionswissenschaften und die heute verfügbaren allgegenwärtigen intelligenten Technologien für eine prozessbegleitende ständige Situationsanalyse zu nutzen, die changierende Benutzerwünsche während der Bewältigung der interaktiven Aufgaben erkennt und in Echtzeit zu Änderungen im Softwareverhalten führt. Zu diesem Zweck wurde das Situ-Framework vorgeschlagen, das zur Laufzeit durch Benutzerbeobachtung und kognitive Analyse zur benutzergerechten und situationsbezogenen Anpassung der Software führen soll.

Im Gegensatz zur experimentellen oder klinischen Erfassung und Analyse von Biosignalen zur Patienten-Überwachung würde eine labormäßige und zu technisch/apparatemedizinisch wirkende Gestaltung des User Tracking im alltäglichen Kontext sowohl das Benutzerverhalten als auch die User Experience stark beeinflussen. Daher sind für HCI-Zwecke Beobachtungsmethoden vorzuziehen, die mit geringen Eingriffen in die Bewegungsfreiheit des Benutzers realisierbar sind und während der Beobachtung kaum wahrgenommen werden.

Neben visuell durch Eye- und Gazetracking erfassbaren Fixationspunkten und Blickverläufen [To13] kann Facereading-Software [No15] heute in sehr effektiver Weise visuell die Ausprägungen und zeitlichen Änderungen der Grundemotionen erfassen und in Echtzeit übertragen. Auch die Erfassung und Auswertung von Biosignalen wird im Bereich der Mensch-Computer-Interaktion immer beliebter und handhabbarer [Sc16]. In [MH16] werden beispielsweise Stressmessungen durch kontaktfreie kamerabasierte Messungen im Gesichtsausdruck des Benutzers durchgeführt. Das geschieht durch Vergleich mit vorher vorgenommenen photoplethysmographischen, also auf der Vermessung der Venen-Funktion basierenden Messungen des kognitiven Stress-Levels von Probanden. Dabei kann unter anderem die für die Bestimmung des Stress-Levels signifikante Veränderung der Pulsfrequenz aus den Bilddaten extrahiert werden. Auch die Beobachtung kurzzeitiger Veränderungen der Pulsrate mit in Kleidung integrierten

Wearables kann zur relativ zuverlässigen Detektion einiger Grundemotionen genutzt werden [Gu15]. In [Pi15] wird auf zahlreiche weitere Arbeiten zur Erkennung von Stress, emotionaler Beteiligung und positiven Emotionen verwiesen. Schließlich vergleicht [LW11] systematisch die Fusion mehrere Modalitäten zur Affekterkennung mit der Emotionserkennung durch nur eine Messmethode.

2 Situationsbasierte Software-Adaption

Um mit dem patternbasierten Ansatz interaktive Software wirklich kontext- und situationsbewusst gestalten zu können und eine situative Laufzeitadaption der Zielsoftware zu ermöglichen, wurde der PaMGIS-Ansatz so erweitert, dass mit einem neu entwickelten Auswertungssystem für simultan aufgezeichnete Eye-, Gaze-Tracking- und Facereader-Daten das Benutzerverhalten zur Laufzeit protokolliert wird und ein intelligenter Auswertungsalgorithmus diese Daten verwendet, um in Echtzeit Entscheidungen zur emotionsbasierten Anpassung der Benutzeroberfläche zu treffen.

Die Entscheidungen werden evaluiert und priorisiert und führen zur Auswahl von strukturellen, Verhaltens- und Präsentations-Patterns aus dem PaMGIS-Repository, aus denen dynamisch ein verändertes User Interface generiert wird, sofern erforderlich [MR16]. Die emotionalen Reaktionen auf die veränderte Benutzungsoberfläche werden wiederum protokolliert und genutzt, um über längere Zeitspannen die User Experience auf einem hohen Niveau zu halten. Zur Steigerung der Zuverlässigkeit bei der Gewinnung von Daten zur Emotionsbestimmung soll in die bestehende Umgebung ein Wristband zur simultanen Gewinnung und Übertragung von Biosignalen integriert werden.

3 Situationsanalyse-Testumgebung

Mit Hilfe einer prototypischen Testumgebung (Abb. 1) und für unterschiedliche Anwendungstypen konnte die Leistungsfähigkeit des Ansatzes bereits evaluiert werden. Die Testumgebung wurde mit Hilfe des ASP.NET MVC Web Application Framework von Microsoft entwickelt. Die zu Evaluationszwecken gebauten Demo-Applikationen nutzen das Open Source Java-Script Framework Angular JS und können auf die vom PaMGIS-Framework bereitgestellten Werkzeuge sowie das HCI-Pattern-Repository zugreifen. Die Analysesoftware mit Schnittstellen zum Tobii Eye-Tracker und zur Noldus FaceReader-Software wurde als Windows-Form-Applikation implementiert. Um die Aussagekraft der visuellen Tracking-Daten zu bestätigen und die Adaptionentscheidungen zu verfeinern, wurden Experimente zur Integration verschiedener kommerziell erhältlicher Wearables in das Auswertungssystem durchgeführt. Der Vergleich der am Markt erhältlichen Systeme führte zur Auswahl des Empatica E4-Wristbands [Em16].

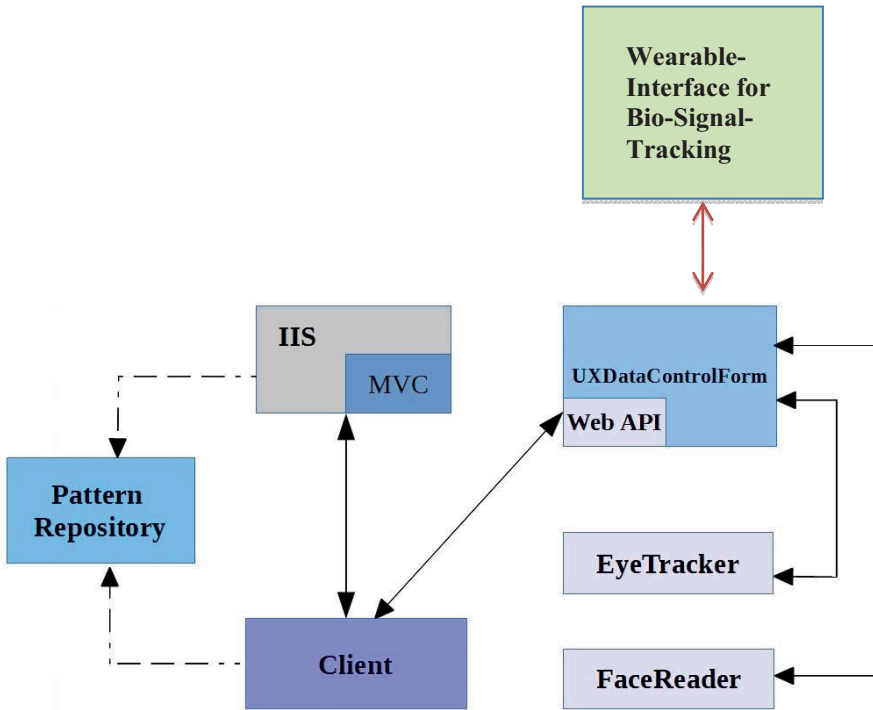


Abb. 1. Software-Architektur der Testumgebung. Die aufgezeichneten Daten werden vom UXDataController gesammelt und ausgewertet. Sowohl Client- als auch Server-Seite der interaktiven Software können durch Patterns aus dem PaMGIS-Repository zur Laufzeit angepasst werden. Als neue Erweiterung der Monitoring- und Auswertungsmöglichkeiten ist eine Wearable-Schnittstelle zur Aufzeichnung von Biosignalen (Puls, Sauerstoffanteil, etc.) in den UXDataController integriert.

Das Wristband verfügt über vier integrierte Sensoren. Einen Photoplethysmographie-Sensor sowie einen elektrodermalen Aktivitätssensor. Ebenfalls sind ein 3-Achsen-Beschleunigungssensor sowie ein Sensor zur Temperaturmessung der Haut integriert.

Folgende Messwerte können ausgewertet werden:

- Pulsrhythmus (Blut-Volumen-basiert) ,@64Hz
- Intervall zwischen Herzschlägen: time, IBI(time) pair
- Elektrodermale Aktivität @4 Hz

- XYZ-Rohbeschleunigung @32Hz
- Hauttemperatur @4Hz*

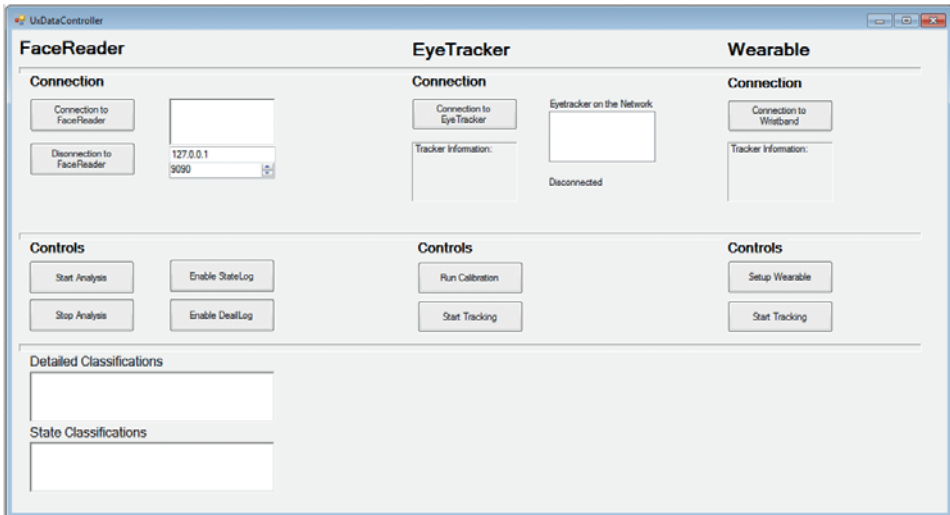


Abb. 2. Das User-Interface der Situationsanalyse-Testumgebung.

Abb. 2 zeigt die in Bezug auf die Wearable-Integration geräteunabhängige Benutzungsschnittstelle zum Auswertungssystem. Über diese wird das E4-Wristband über eine von Empatica mitgelieferte API angeschlossen, um die erfassten Biosignale in Echtzeit auswerten zu können.

Der durch Biosignal-Tracking erweiterte Auswertungsalgorithmus und die damit möglichen dynamischen Laufzeit-Adaptionen der Benutzungsoberfläche sowie die inhaltlichen Auswirkungen auf den Pattern-Vorrat im PaMGIS-Repository sollen in den nächsten Monaten am Beispiel interaktiver Auskunftssysteme für Bahnreisende und an der Benutzungsschnittstelle eines Produktionsplanungssystems für die Kleinserienfertigung erprobt und evaluiert werden.

4 Literatur

- [CC02] Calvary, G., Coutaz, J., Bouillon, L. et al. (2002): “The CAMELEON Reference Framework”. Retrieved April 15, 2015 from <http://giove.isti.cnr.it/projects/cameleon/pdf/CAMELEON%20D1.1RefFramework.pdf>
- [Ch16] Chang, C.K. (2016): “Situation Analytics: A Foundation for a New Software Engineering Paradigm”, COMPUTER, January 2016, pp. 24-33

- [EM09] Engel, J., Märtin, C. (2009): “PaMGIS: Framework for Pattern-based Modeling and Generation of Interactive Systems”, Proc. HCI International '09, San Diego, U.S.A., pp. 826-835
- [EM15] Engel, J., Märtin, C., & Forbrig, P. (2015): „A Concerted Model-driven and Pattern-based Framework for Developing User Interfaces of Interactive Ubiquitous Applications”, Proc. First Int. Workshop on Large-scale and Model-based Interactive Systems, Duisburg, pp. 35-41.
- [Em16] Empatica Inc. Retrieved June 20, 2016 from <https://support.empatica.com/hc/en-us/articles/202581999-E4-wristband-technical-specifications>
- [Gu15] Guo, H.W et al.:”Short-term Analysis of Heart Rate Variability for Emotion Recognition via a Wearable ECG Device”, Proc. ICIIBMS 2015, Okinawa, Japan, IEEE, pp. 262-265
- [LW11] Lingenfels, F., Wagner, J., Andre, E.: “A Systematic Discussion of Fusion Techniques for Multi-Modal Affect Recognition Tasks”, Proc. ICMI '11, Nov. 14-18, 2011, Alicante, Spain, pp. 19-25
- [MC14] Meixner, G., Calvary, G., Coutaz, J. (2014): “Introduction to Model-Based User Interfaces”, W3C Working Group Note 07 January 2014. Retrieved May 27, 2015 from <http://www.w3.org/TR/mbui-intero/>.
- [MH16] McDuff, D.J. et al.: “COGCAM: Contact-free Measurement of Cognitive Stress During Computer Tasks with a Digital Camera”, Proc. CHI '16, May 07-12, 2016, San Jose, CA, USA
- [MR16] Märtin, C., Rashid, S., Herdin, C.: “Designing Responsive Interactive Applications by Emotion-Tracking and Pattern-Based Dynamic User Interface Adaptation”, M. Kurosu (Ed.): HCI 2016, Part III, LNCS 9733, pp. 1-9, 2016
- [No15] Noldus Inc. FaceReader 6 Application Programming Interface, Technical Note.
- [Pi15] Picard, R.:”Recognizing Stress, Engagement, and Positive Emotion”, Prod. IUI 2015, March 29-April 1, 2015, Atlanta, GA, USA, pp. 3-4
- [Sc16] Schmidt, A.: “Biosignals in Human-Computer Interaction”, Interactions Jan-Feb 2016, pp. 76-79
- [To13] Tobii. Tobii Studio SDK. Developer Guide, 8 May 2013

**Seminare,
Symposia,
Gespräche**

Symposium des Fachbereichs Informatik und Gesellschaft

Jens-Martin Loebel¹ und Agata Królikowski²

Zusammenfassung

Informatik bereichert unser Leben: Überall eingesetzt unterstützt sie (idealerweise) den Menschen in seiner Arbeit, rettet Leben, eröffnet neue Möglichkeiten in der Forschung oder dient im einfachsten Fall der Unterhaltung. Dieser technisierte Alltag jedoch hat Auswirkungen, die weitreichender sind, als es möglicherweise auf den ersten Blick erscheint.

Die Mitglieder des Fachbereichs Informatik und Gesellschaft der GI analysieren Voraussetzungen, Wirkungen und Folgen von Informatik, Informationstechnik und Informationsverarbeitung in der Gesellschaft. Wo gibt es Fehlentwicklungen? Wann müssen Informatikerinnen und Informatiker Verantwortung übernehmen? Welche Rolle spielt der niedrige Anteil der Frauen in der Informatik? Wie ist es um unser digitales Kulturerbe bestellt? Wie können Informatiker und Nicht-Informatiker besser zusammenarbeiten? Wie sieht eine sozial verantwortliche Technikgestaltung aus?

Zusammen mit der Österreichischen Computer Gesellschaft (OCG) wurden diese Fragen und weitere Herausforderungen der Informatik im Rahmen des Symposiums in Impulsvorträgen vorgestellt und diskutiert und an einer aktualisierten Fassung der ethischen Leitlinien der Gesellschaft für Informatik gearbeitet.

Die Mitgliederinnen und Mitglieder des Fachbereichs arbeiten an gesellschaftlichen Zielsetzungen für die Informatik und leiten daraus Gestaltungskriterien ab. Sie wollen Fehlentwicklungen in der Informatik aufzeigen und Wege für eine gesellschaftlich und sozial verantwortete Technikgestaltung weisen. Dazu unterstützen sie einschlägige Tätigkeiten in Forschung, Entwicklung, Bildung und anderer beruflicher Praxis.

In dauerhaften Fachgruppen und zeitlich begrenzten Arbeitskreisen werden Schwerpunktbereiche bearbeitet.

Die Fachgruppe *Frauen und Informatik* sieht ihre Aufgabe darin, sich mit der Situation von Frauen als Gestalterinnen und als Betroffene von Informations- und Kommunikationstechnologien auseinanderzusetzen und sich darüber auszutauschen. Es engagieren sich Informatikerinnen und IT-Fachfrauen aus verschiedensten beruflichen Bereichen. Sie sind in der Fachgruppe, in Ausbildung und Beruf, in der Wissenschaft, in

¹ Sprecher des Fachbereichs Informatik und Gesellschaft, Universität Bayreuth, loebel@uni-bayreuth.de.

² Stellv. Sprecherin des Fachbereichs Informatik und Gesellschaft, Interlake Systems GmbH, akrolikowski@acm.org.

der GI, in Politik und der Öffentlichkeit aktiv.

Die Fachgruppe *Informatik- und Computergeschichte* widmet sich historischen Aspekten von Informatik und Gesellschaft. Ihre Aufmerksamkeit gilt insbesondere der Geschichte der Fachdisziplin und Institutionen, der Software und der Programmierung, der Mensch-Computer-Interaktion, der Computernetze und Computerkommunikation, der Sozialgeschichte der EDV- Berufe sowie der Alltagsgeschichte des Computers und der Datenverarbeitung.

Die Fachgruppe *Informatik und Soziale Entwicklung* engagiert sich für den Einsatz von Informatik zur Verbesserung der interkulturellen Kommunikation und untersucht die Auswirkungen der Informationstechnologie auf die wirtschaftliche, soziale und kulturelle Lage der Menschen in Entwicklungsländern. Unterstützt werden auch Informations- und Kommunikationsentwicklungsprogramme und der Einsatz von Informations- und Kommunikationstechnologien in Bildung und Ausbildung.

Die Fachgruppe *Informatik und Ethik* befasst sich mit Aspekten von Ethik in der Informatik und versteht sich als Stachel, um kontroverse Diskussionen zu initiieren. Sie versteht sich nicht als oberste ethische Instanz, sondern will Grundlagen zu Diskussionen erarbeiten. Die Mitglieder entwerfen Lehrmaterialien und erarbeiten Unterlagen zum Thema Plagiat sowie zum Problem wissenschaftlichen Fehlverhaltens. Die Fachgruppe arbeitet an einer Fallsammlung zur Illustration der ethischen Leitlinien der GI.

Die Herausforderung des langfristigen Erhalts und der Nutzung unseres digitalen Kulturerbes bildet das Arbeitsfeld der Fachgruppe *Langzeitarchivierung*. Die Vielzahl verschiedener digitaler Objekte bedarf unterschiedlicher konzeptioneller und technischer Lösungen. Schwerpunkte der FG sind u. a. die Klärung rechtlicher Grundlagen, die Erstellung bzw. Verbreitung von Best-Practice-Beispielen, die Vermittlung von Emulation als Bewahrungskonzept sowie der Dialog mit Politik, Forschung und Wirtschaft.

Ziel der Fachgruppe *Internet und Gesellschaft* ist es, die technischen Entwicklungen des Internets und verwandter Netze auf ihre gesellschaftlichen, politischen, sozialen und kulturellen Auswirkungen hin zu untersuchen und der Öffentlichkeit nahezubringen. Neben den vom Internet induzierten Veränderungen wird die rapide Weiterentwicklung der Technik selbst, insbesondere der Umgang mit den enorm gewachsenen und immer noch wachsenden Informationsbeständen, fokussiert.

Social Media und digitale Medien nehmen einen immer größeren Raum in unserer Kommunikation und sozialen Interaktion ein. Wer nicht auf sozialen Plattformen vertreten ist, verliert den Anschluss oder wird ausgegrenzt. Die Fachgruppe *Informatik und Inklusion* strebt die Anerkennung aller Menschen in ihrer Individualität und die Möglichkeit zur Teilhabe in vollem Umfang an. Die Fachgruppe widmet sich diesen Fragen und Problemen. Dabei gilt das Interesse der Barrierefreiheit von Informations- und Kommunikationstechnik.

Im Fokus des Arbeitskreises *Computational Social Science* stehen Informatikmethoden, wie sie zunehmend in der Modellierung und Analyse gesellschaftlicher Phänomene zum Einsatz kommen. Hierzu gehört die Nutzung von Simulationstechniken zur Untersuchung formaler Modelle sozialen Verhaltens ebenso wie neuartige Data-Mining-Methoden zur Gewinnung sozialwissenschaftlicher Erkenntnisse auf Grundlage großer Datensätze aus sozialen Medien und anderen Plattformen der Digitalwirtschaft. Neben dem unbestreitbaren Potenzial dieser computergestützten Methoden für die Sozialwissenschaften sollen insbesondere die Herausforderungen und Chancen für Informatik und Gesellschaft im Fokus dieses Arbeitskreises stehen. Der Arbeitskreis beschäftigt sich sowohl mit technischen als auch ethischen Fragen, die sich aus der immer stärkeren Verflechtung von informationstechnischen und sozialen Systemen ergeben.

Informatik durchdringt alle Wissenschaftsbereiche, zunehmend auch die Geistes- und Kulturwissenschaften (engl. Humanities). Der Arbeitskreis *Informatik und Digital Humanities* will sich vor allem mit den informatischen Aspekten der verschiedenen Bereiche wie z. B. der Historischen Fachinformatik, Computerlinguistik oder Bildwissenschaft kritisch auseinandersetzen und die rasch fortschreitende Entwicklung in diesem Bereich begleiten.

Der Arbeitskreis *Patente* bietet eine Anlaufstelle mit Informationen sowie wissenschaftlicher und ggf. auch patent- anwaltlicher Betreuung des Themas „Computerimplementierte Erfindungen“. Als Schnittstelle zwischen Patentrecht und Informatik geht es explizit nicht um eine einseitige Förderung von Softwarepatenten. Vielmehr soll die Diskussion insgesamt unabhängig von vertretener Meinung auf eine sachlich fundierte Ebene gehoben werden.

Der Arbeitskreis *Wissenschaftliche Annotationstechniken* dient dem Austausch über fachwissenschaftliches Annotieren als eine wichtige Kulturtechnik, die durch informationstechnische Fortschritte an neuen Möglichkeiten gewinnt. Dabei ist wichtig zu bemerken, dass die Annotation nicht nur eine Technik der Geisteswissenschaften ist, sondern auch Anwendung in den Naturwissenschaften findet (z. B. das Annotieren von technischen Diagrammen, medizinische Aufnahmen, usw.). Ein weiterer Grundpfeiler des Arbeitskreises wird die Vernetzung mit anderen Vorhaben sein, die sich das Vorantreiben von Techniken der digitalen Annotation auf ihre Fahne geschrieben haben. Der Fachbereich Informatik und Gesellschaft kann hier seine Rolle als Vermittler zwischen der Informatik und den Geistes- und Naturwissenschaften nochmals unterstreichen.

In diesem Rahmen werden so beispielsweise die besonderen Belange der Frauen im männlich dominierten IT-Geschäft verfolgt, die Geschichte der Informatik aus der Sicht aktueller Probleme beschrieben, der Digital Divide zwischen Erster und Dritter Welt aufgearbeitet, die informatischen Aspekte in Geistes- und Kulturwissenschaften beleuchtet sowie die medialen Umbrüche der Informationsgesellschaft diskutiert. Es wird zu Themen wie eGovernment, Fragen der langfristigen Bewahrung digitaler Daten,

der Inklusion auf sozialen Plattformen, der Auseinandersetzung mit ethischen Fragestellungen, Softwarepatenten und der Wechselwirkung von digitalen Netzen und Gesellschaft gearbeitet.

Die Veröffentlichung von Arbeitsergebnissen und Mitteilungen der Gliederungen erreicht mit einem E-Mail-Newsletter mehrere Hundert subskribierte Leserinnen und Leser.

Alle Informatikerinnen und Informatiker, die zu Fragen von Informatik in der Gesellschaft arbeiten wollen, sind herzlich eingeladen, sich im Fachbereich (FG) mit dessen Fachgruppen und Arbeitskreisen zu engagieren. Weitere Informationen zum Fachbereich, dessen Gliederungen und den Ergebnissen des Symposiums finden Sie unter <http://fb-iug.gi.de>.

Software as a Medical Device - ein Seminar mit Fallbeispielen zum Mitdiskutieren und Mitarbeiten

Georg Heidenreich¹

Zusammenfassung

Im Zuge der aktuellen Entwicklungen rund um Software als Medizingerät stellen sich bedeutsame Fragen für Anwender, Kliniker und Hersteller. Der Workshop erläutert insbesondere die Konsequenzen der aktuellen europäischen Regelungen für Software aus der Perspektive des Entwicklers und Inverkehrbringers. Der erste Teil des Workshops gibt eine Einführung in den aktuellen europäischen Rechtsrahmen der Qualifikation und Klassifikation von Software als Medizingerät. Ein Vergleich mit dem Rechtsrahmen in den USA wird gefolgt von einer Erläuterung der relevanten Standards und Organisationen dahinter. Nach dieser Vermittlung von Grundlagen können die Teilnehmer an bis zu drei ad-hoc Beispielen praktische Aspekte der Zweckbestimmung, der Qualifikation, der Klassifikation, sowie der Risiko-Analyse durchspielen.

Der zweite Teil des Workshops vermittelt die technischen Aspekte der Entwicklung von Software als Medizingerät. Dabei ist die physische Gefährdungsfreiheit ("Safety") das wesentliche Ziel der Prozesse und Tests. Ausgehend von einer Safety-Klassifikation des Software-Systems, geht es über Traceability und Architektur zu den Regelungen für Entwicklung, Test und Dokumentation. Weitere Ausblicke auf den "Safety Assurance Case" sowie die Umsetzung von agilen Methoden im regulierten Umfeld runden den zweiten Teil ab. Die Umsetzung aller dieser Vorschriften auf die Beispiele des ersten Teils wird dabei beispielhaft skizziert.

- Wann ist Software ein Medizingerät?
- Was bedeutet das für den Hersteller?
- Gibt es Ausnahmen?

Im Hinblick auf eine verantwortungsvolle Auslegung im Sinne der Patientensicherheit ist allerdings festzuhalten, dass Produkte mit medizinischen Zwecken und für die Verwendung durch Heilberufler in Europa reguliert sind und sicher funktionieren müssen !

Die Liste nennt die vier Faktoren, die zusammenkommen müssen, damit ein Produkt in Europa ein Medizingerät ist:

¹ Siemens Healthcare GmbH, Erlangen, georg.heidenreich@siemens.com

1. Erst die Kombination mehrerer Produkteigenschaften macht das Medizingerät aus: Es muss sich um ein Produkt handeln, das eine Zweckbestimmung hat, aus diesem Grund sind allgemeine Bibliotheken, Frameworks und allgemeine Spezifikationen keine Medizingeräte.
2. Das Produkt muss Daten verändern oder komplexe Berechnungen durchführen - dies schließt Suche in medizinischem Wissen, lernende Wissensbasen sowie komplexe Verarbeitung von Suchkriterien ein. Produkte mit ausschließlicher Funktionalität auf den Gebieten "send"/ "store"/ "search" sind ausgenommen.
3. Das Produkt muss unter Bezug auf eine identifizierbare Person arbeiten - das heißt nicht, dass diese Person identifiziert sein muss, aber es muss um eine einzelne bestimmbare Person gehen.
4. Die Zweckbestimmung des Produkts muss Diagnose oder Therapie nach Art 1.2a der Medizingeräterichtlinie beschreiben.

Fehlt auch nur einer dieser Faktoren, dann handelt es sich nicht um ein Medizingerät. Produkte, die ohne diese Zweckbestimmung, jedoch mit der Absicht der medizinischen Nutzung in Verkehr gebracht wurden oder für medizinische Zwecke beworben werden, gelten allerdings als Medizingerät.

Bei Software als Medizingerät (software as a medical device = SAMD) gibt es einige Besonderheiten, die wir hier plakativ anreißen:

- SaMD/Apps gefährden Menschen nicht direkt, sondern nur durch ihre Ausgabe-Daten sowie die jeweilige Performanz (Datendurchsatz, Latenz, etc.). Hardware-Schutzmaßnahmen gegen Software-Fehler gibt es nicht als Teil des SAMD selber.
- Der Schweregrad von Software-Fehlern ist durch den Hersteller der SAMD nicht technisch begrenzt.
- "Spätes Testen" macht die Software auch nicht besser.
- Die klinische Validierung von SAMD - durch den Nachweis einer Wirkung mittels Versuchsreihen - ist (noch) nicht geregelt.

Bei der Gefährdungsanalyse ist zunächst einmal jede Folge eines Softwarefehlers potenziell fatal. Eine genaue Bestimmung der gefährlichen Szenarien ist jedoch nur auf der Basis einer genauen Zweckbestimmung möglich. Der Hersteller haftet dabei nur für diejenigen Szenarien, die laut Zweckbestimmung vorhersehbar sind.

Tatsächlich darf ein Betreiber die SAMD auch mit anderen Medizingeräten so kombinieren, dass durch die Kombination ganz neue medizinische Zwecke unterstützt werden. So etwas kann und muss der Hersteller in einer Gefährdungsanalyse allerdings nicht vorwegnehmen.

SAMD kann alle Fehler aufweisen, die andere Software auch besitzen kann. Hinzu

kommen jedoch auch ein paar zusätzliche typische Fälle:

- Eine große Quelle schwerwiegender Fehler in Software sind Verwechslungen der Patienten-Identität oder die Nichterkennung der Patientenidentität.
- Eine erhebliche und nicht immer erkennbare Gefährdung geht auch von unklarer oder falscher Behandlung physikalischer Einheiten aus.

Schließlich können fehlerhafte Voreinstellungen übersehen werden und fatale Folgen haben. Beispiel Patientengewicht bei der Berechnung der Dosierung einer Medikation.

- SAMD haben normalerweise keine Zweckbestimmung, die mit den Papier-Packungsbeilagen herkömmlicher Medikamente vergleichbare wäre. Allerdings kann nur so eine Zweckbestimmung die Basis für Haftung und Verwendung einer App sein.

Die Rechtslage bezüglich der Darstellungsform solcher umfangreicher Zweckbestimmung ist dabei noch unklar: Faktisch wirkt bei SAMD die ganze Medizingeräte-Gesetzgebung nicht, solange die Zweckbestimmung nicht wirksam erkennbar ist! In der Praxis spielt nämlich für die Bewertung der Zweckbestimmung auch eine Rolle, ob ein "professioneller" Einsatz beschrieben wird, oder die SAMD hauptsächlich im Freizeitbereich eingesetzt werden soll. Oft versuchen Hersteller, per Startbildschirm lediglich "informativ Zwecke" festzulegen und Diagnose / Therapie auszuschließen. Diese Vorgehensweise ist durch einige Mahnungen von Behörden nach und nach abgestellt worden.

Zudem ist für Apps nicht immer leicht zu klären, wer überhaupt als Hersteller fungiert:

- Der Verkäufer ist nicht der haftende "Legal Manufacturer"
- Play Store ? AppStore ? Die (freischaffenden) Entwickler ?
- AppStore -moderiert, vermeidet medizinische Zweckbestimmung. Google PlayStore - vollkommen freie Vermarktung, auffällig viele Freizeit-Apps / Laien-Apps Die EU-Kommission gibt offen zu, dass wegen der Vielzahl von Neu-Erscheinungen eine Durchsetzung der Richtlinie für Apps nicht möglich ist. Es gibt für die "Qualifizierung" von Medizingeräten keine wirksame Internet-Polizei. Als Beispiel sei Deutschland genannt: Die Medizinprodukte-Aufsicht in Deutschland hat das Bundesamt für Arzneimittelsicherheit BfArM, die Exekutive sind die Ordnungsämter auf kommunaler Ebene.

Fazit: Weder die Technik noch die Funktion machen das Medizingerät aus, sondern nur ausschließlich die Zweckbestimmung (und andere Kommunikation des Herstellers zum Produkt).

Die folgenden harmonisierten EU-Normen stellen den anwendbaren Stand der Technik

für die Software-Erstellung in einem/als Medizingerät dar:

- Die Verwendung eines QM-Systems nach ISO 13485, d.h. insbesondere Dokumentenlenkung, Wiederherstellbarkeit und spezifisch für Medizingeräte: systematische Umsetzung der Kundenanforderungen sowie systematische Verbesserung von Fehlern aus dem Feld durch einen Problemlösungsprozess inkl. Feldbeobachtung und Rückrufe . Eine Risiko-Analyse unter Abwägung des Patienten-Nutzens: Auf der Basis der Zweckbestimmung ist eine Gefährdungsanalyse nach ISO 14971 durchzuführen. Jede Beteiligung von Software an Risiko-Szenarien ist als "Gefährdungsschlüssel" in die Anforderungsverwaltung für Software einzubringen.
- Die Produkt-Anforderungen müssen nach IEC 82304 dokumentiert und getestet werden: Diese Norm legt die Dokumentation der Produkthanforderungen und Validierung der Umsetzung fest. Diese Norm betrachtet sowohl Prozesse als auch bestimmte Arten von Anforderungen.
- Der Entwicklungsprozesses nach IEC 62304 schreibt Dokumente und Aktivitäten vor, insbesondere für Software-Anforderungen, Design, Implementierung und den Nachweis der Traceability für alle o.g. ermittelten Gefährdungen.

Laut EU-Dokument ("MEDDEV 2.1/6 Qualification") zur Qualifizierung von Software als Medizingerät kann aus Modulen bestehen, von denen einige Medizingerät sind, jedoch andere nicht. Dieses MEDDEV-Dokument sagt allerdings nicht, was Module sind und wie sie voneinander getrennt werden.

In der Praxis wenden Hersteller die folgenden Methoden an, um der Qualifizierung ihrer Software als Medizingerät zu entgehen:

- In der Praxis erscheint ein Hinweis auf die Verwendung durch Laien ausreichend, um eine Einstufung als Medizingerät zu vermeiden.
- Ein Produkt kann in Module aufgeteilt werden, von denen nur wenige die eigentliche medizinische Funktion in der Zweckbestimmung aufweisen. Diese Module könnte man dann separat in Verkehr bringen.
- Das Produkt kann unter Berufung auf "send, store, search" so entworfen werden, dass es die eigentliche Verarbeitung von Daten nicht in der Zweckbestimmung aufführt . Ein Produkt könnte generische Funktionen anbieten und durch ein Programmiersystem die Umsetzung medizinischer Zwecke dem Anwender überlassen; der dann zum Hersteller wird.
- Ein Produkt könnte die Integration mit anderen IT-Systemen und Medizingeräten im Hinblick auf eine kombinierte Zweckbestimmung technisch unterstützen und genau diese Integration dann in der Zweckbestimmung auslassen, d.h. man kauft eine Komponente, die einen "harmlosen" Zweck festlegt, und integriert sie dann mit anderen Systemen, um einen neuen medizinischen Zweck zu unterstützen.

Vor diesem Hintergrund arbeiten die nationalen Behörden für Medizinproduktezulassungen an einer aktualisierten Umsetzung der Richtlinie und geben Rückmeldung an die "Software / Medical Device"-Arbeitsgruppe der EU-Kommission. Diese Arbeitsgruppe diskutiert dann die aus den Mitgliedsstaaten vorgestellten Fallbeispiele und entscheidet über die Erweiterung und Aktualisierung des Qualifikationspapiers ("MEDDEV Qualification").

*Zur Person: Dr.-Ing. **Georg Heidenreich** ist Vertreter des Herstellerverbands COCIR in der Arbeitsgruppe "Software" der EU-Kommission. Als Angestellter der Siemens Healthcare GmbH ist er für die Koordinierung der Standards im Bereich IT-Security sowie IT-Connectivity zuständig und leitet die deutsche Delegation in ISO TC 215 "Health Informatics".*

Von der Messbarkeit des Wissens - Versuch und Irrtum auf dem Weg in die Wissensgesellschaft

Günter Koch¹

Zusammenfassung

Die österreichische Academia hat in wiederkehrenden Wellen neue Denkschulen hervorgebracht, so z.B. in der Volks- und Betriebswirtschaft (z.B. Hajek, Schumpeter), in der Psychologie (Freud), der Quantenphysik (Schrödinger, Zeilinger) oder von Mitgliedern des sog. Wiener Kreises zur Wissenschaftstheorie, Philosophie und Logik. In jüngerer Zeit emergieren Beiträge sowohl zur noch wenig etablierten Komplexitätsforschung als auch zur Wissensgesellschaft und insbesondere Wissensökonomie.

Ein Ausgangspunkt der letzteren Diskussion war eine gesetzliche Vereinbarung mit den Universitäten, dass diese jährlich eine „Wissensbilanz“ vorzulegen haben, auf Basis derer die resultierende Leistungsvereinbarung zwischen der jeweiligen Universität und dem finanzierenden Wissenschaftsministerium getroffen wird. Modell und Methode der Wissensbilanzierung wurden ursprünglich für die und von der größte/n Forschungseinrichtung Österreichs unter der Regie von Günter Koch entwickelt.

Die Gesprächsrunde diskutierte, ob und wenn ja wie sinnvoll es ist, das „Intellektuelle Kapital“ einer „Denkorganisation“, z.B. eines Softwarehauses, zu qualifizieren und sogar zu quantifizieren.

Günter Koch ist Informatiker der ersten Stunde der Karlsruher Schule (1975) und hat eine bewegte, über 40-jährige Karriere in folgender Abfolge hinter sich: Softwarehaus-Unternehmer in Deutschland, Universitätsrat in Frankreich, Leiter des in Spanien ansässigen „European Software Institute“ (ESI), Top-Consultant für die Java-Evangelisierung bei SUN Microsystems, Vorstandsvorsitzender des Austrian Institute of Technology (AIT - Österreichs größtes Forschungszentrum).

In den letzten Jahren Begründer des „New Club of Paris“, eines Think Tanks für die Wissensgesellschaft, freier Berater in Sachen Wissensmanagement, Aufsichtsrat eines Softwarehauses und Präsident der Think-Tank-Plattform „Humboldt Cosmos Multiversity“ auf Teneriffa.

¹ Mittelgasse 7, 1060 Wien, Österreich, <http://www.execupery.com/>, guenterkochat@gmail.com

Tutorien

Semantic-based Modeling for Information Systems using the SeMFIS Platform

Hans-Georg Fill¹

Abstract: In this paper an outline of semantic-based modeling and the approach of SeMFIS is given. At its core, semantic-based modeling characterizes the annotation of semi-formal conceptual models with concepts from formal semantic schemata such as ontologies. Semantic-based modeling is used for the design and analysis of information systems where it supports the extension of the semantic representation and analysis scope of existing modeling methods. The approach has been previously used for applications in semantic business process management, for benchmarking, semantic-based simulation and several others. With the ADOxx-based SeMFIS platform, which constitutes a flexible engineering platform for realizing semantic annotations of conceptual models, a technical implementation is provided. The SeMFIS platform is offered for free via the Open Models Initiative Laboratory.

Keywords: Conceptual Modeling, Semantic Annotation, Ontology, SeMFIS

1 Introduction

In the area of information systems, modeling methods are today a standard approach for representing and analyzing knowledge and data, for designing and implementing systems, and for interacting with systems at run-time [FJ16, Ka08, BBF09]. For these purposes it can be reverted to a large number of existing modeling methods. Or, if no adequate method is available, new methods can be realized that address domain-specific requirements. Although the implementation of new modeling methods has been considerably simplified with the availability of powerful meta modeling platforms such as ADOxx, Eclipse or MetaEdit, creating new methods from scratch still involves a lot of effort. In addition, once a modeling method has been successfully established in an organization, the wide usage of a method hampers its modification. For example, when thousands of models have been created using a particular modeling language – as it is today the case in large organizations [Ro06] – changes in the modeling language can require considerable adaptation efforts and may lead to unexpected side effects, e.g. regarding the interoperability with run-time systems. Furthermore, regulations may restrict modifications of a modeling method, e.g. as enforced by the Swiss modeling conventions for public administration².

Nevertheless, the representation and analysis scope of modeling methods often needs to be adapted for meeting emerging requirements. This can either result from changes in the

¹University of Vienna, Faculty of Computer Science, Waehringerstrasse 29, 1090 Vienna, hans-georg.fill@univie.ac.at

²See the Guideline eCH-0158: <http://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0158&documentVersion=1.1>

business environment, e.g. due to new legal requirements [Fi07], or also from changes in the technological environment, e.g. with upcoming new technologies such as the internet-of-things [Hu14]. It thus seems desirable to decouple changes in a modeling method from the model instances. In traditional approaches such changes can however not be easily accomplished as the modeling language is strictly tied to its corresponding model instances.

The approach of semantic-based modeling offers a concept for resolving this coupling. In the remainder of this paper we will therefore briefly introduce the concept of semantic-based modeling and its realization in the SeMFIS approach and technical platform in section 2. For illustrating the practical usage of SeMFIS previous applications will be outlined in section 3. The paper will be concluded with an outlook on future developments of the approach in section 4.

2 Semantic-Based Modeling with SeMFIS

For illustrating in detail the approach of semantic-based modeling, the traditional way of extending the scope of a modeling language needs to be described at first. The fundamental relationships of the effects of changing a modeling language are depicted in figure 1.

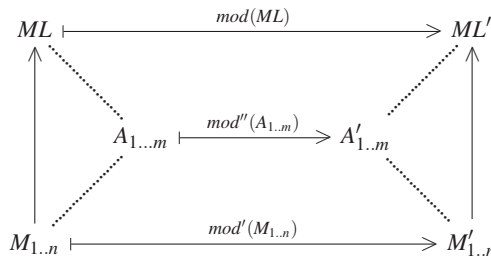


Figure 1: Effects Resulting from Changes in a Modeling Language

We start with the assumption that a number of models $M_{1..n}$ corresponds to a modeling language ML . For brevity we do not make the detailed relationships between modeling languages and models explicit but refer the interested reader to formal specifications e.g. in [FRK12, HR00]. In addition, a number of algorithms $A_{1..m}$ exist that refer to the modeling language ML and the models $M_{1..n}$. The exact nature of these references is intentionally left open as algorithms may have very distinct ways of interacting with the specifications given by the modeling language and the content of models. They encompass for example analysis algorithms for querying model content, process simulation algorithms targeting the behavior of process-like structures or also code generation algorithms.

In case that the modeling language is generally applicable to a problem domain but does not meet some specific requirements for representing information, a modification $mod(ML)$ is applied which results in a new variant of the modeling language ML' . The modification may include the addition, deletion or modification of elements, attributes and relations of the modeling language. Upon these changes, also the models have to be adapted as

expressed by the transition $mod'(M_{1..n})$. Depending on the changes of the modeling language, this may lead to extensions, reductions or modifications of the content of models.

In any case, the resulting models $M'_{1..n}$ have to correspond again to the new variant of the modeling language ML' . At the same time the algorithms need to be adapted. This is expressed by the transition $mod''(A'_{1..m})$. Again, the new variants of the algorithms $A'_{1..m}$ have to refer to ML' and the corresponding models $M'_{1..n}$. From these relationships it follows that changes in the modeling language potentially trigger a large number of changes both in the existing models as well as in the algorithms. It thus needs to be checked at least for each of the n models and each of the m algorithms whether they need to be modified.

2.1 Concept of Semantic-Based Modeling

As an alternative solution, the concept of semantic-based modeling can be applied which uses ontologies for representing new information [FB09]. The combination of conceptual models and ontologies has been discussed in several contexts. The most prominent ones are the area of *semantic business process management* where ontologies have been used primarily to permit the processing of natural language information contained in models and thus support the analysis and execution of processes [TF07, Ho07, He05]. Another direction is found in the context of *semantic lifting* where models are translated into ontologies to enable interoperability [Ka06].

The approach of *semantic-based modeling* presents a third direction. It originated from the goal of meeting three challenges: a. the ex-post representation of new requirements in conceptual models without changing the original modeling language, b. to make knowledge explicit that is implicitly covered in models, and c. to process this explicit, ex-post represented knowledge using algorithms. Semantic-based modeling can be compared to a special variant of what is called *model weaving*, i.e. the linkage between different types of models [De05, DDFV07]. In contrast to model weaving semantic-based modeling however does not aim for the transformation between the linked models but uses the linked models as meta-data for representing additional requirements.

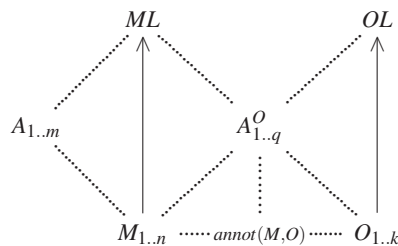


Figure 2: Concept of Semantic-Based Modeling for Representing Information via Annotations

The core idea of the approach is depicted in figure 2. On the left-hand side again the modeling language, models, and algorithms are shown. For adding new representation and

analysis concepts to the existing models, semantic-based modeling does not modify the existing modeling language *ML*. Rather, a second representation pillar is added in the form of an ontology language *OL* and corresponding ontologies $O_{1..k}$. Ontologies are thereby used in their generic sense as formal, i.e. computational, forms of shareable terminologies [Ne91]. They can thus be used for making additional knowledge aspects explicit that can be understood by algorithms and humans alike.

The ontologies are then linked to the conceptual models via annotations $annot(M, O)$. Annotations are a form of meta-data in the sense of a markup that adds additional information to existing data. In the context of semantic-based modeling, annotations add information from ontologies to conceptual models. Therefore, they do not require a modification of the original modeling language but rather act as an additional layer of information on top of existing model content. Based on the annotations, the ontology language, and the ontology instances new algorithms $A_{1..q}^O$ can be designed that process the information in models in new ways. At the same time the existing algorithms $A_{1..m}$ are not affected.

2.2 Conceptual Realization in SeMFIS

The approach of SeMFIS (Semantic-based Modeling Framework for Information Systems) is based on the concept of semantic-based modeling. A particular feature of SeMFIS is the provision of a separate semantic annotation model type – see the meta model in figure 3. Through this model type, the annotation of conceptual models with ontology concepts can be established without modifications in the conceptual modeling language [Fil1a]. This is achieved via *model reference* and *connector reference* objects that are linked to *ontology references*. The linkage between model and ontology references can also be semantically specified via *annotator* objects.

On the ontology side, the SeMFIS approach currently provides three ontology languages in the form of model types: the *term model type* for representing controlled vocabularies, the *frames ontology model type* for representing ontologies in the OKBC frames format as implemented in Protégé, and the *OWL ontology model type* for representing ontologies in OWL format.

2.3 Technical Implementation: The SeMFIS Platform

The SeMFIS approach has been technically implemented and is provided free of charge via the Open Models Initiative [Fi16, Fi12b]³. The SeMFIS platform is based on the ADOxx metamodeling platform [FK13]. ADOxx is an industry-scale platform that offers generic mechanisms for automatically generating model editors from meta model specifications as well as a wide range of additional components and functionalities for interacting with models.

³ See the SeMFIS website on OMiLAB: <http://www.semfis-platform.org> or <http://www.omilab.org/web/semfis>

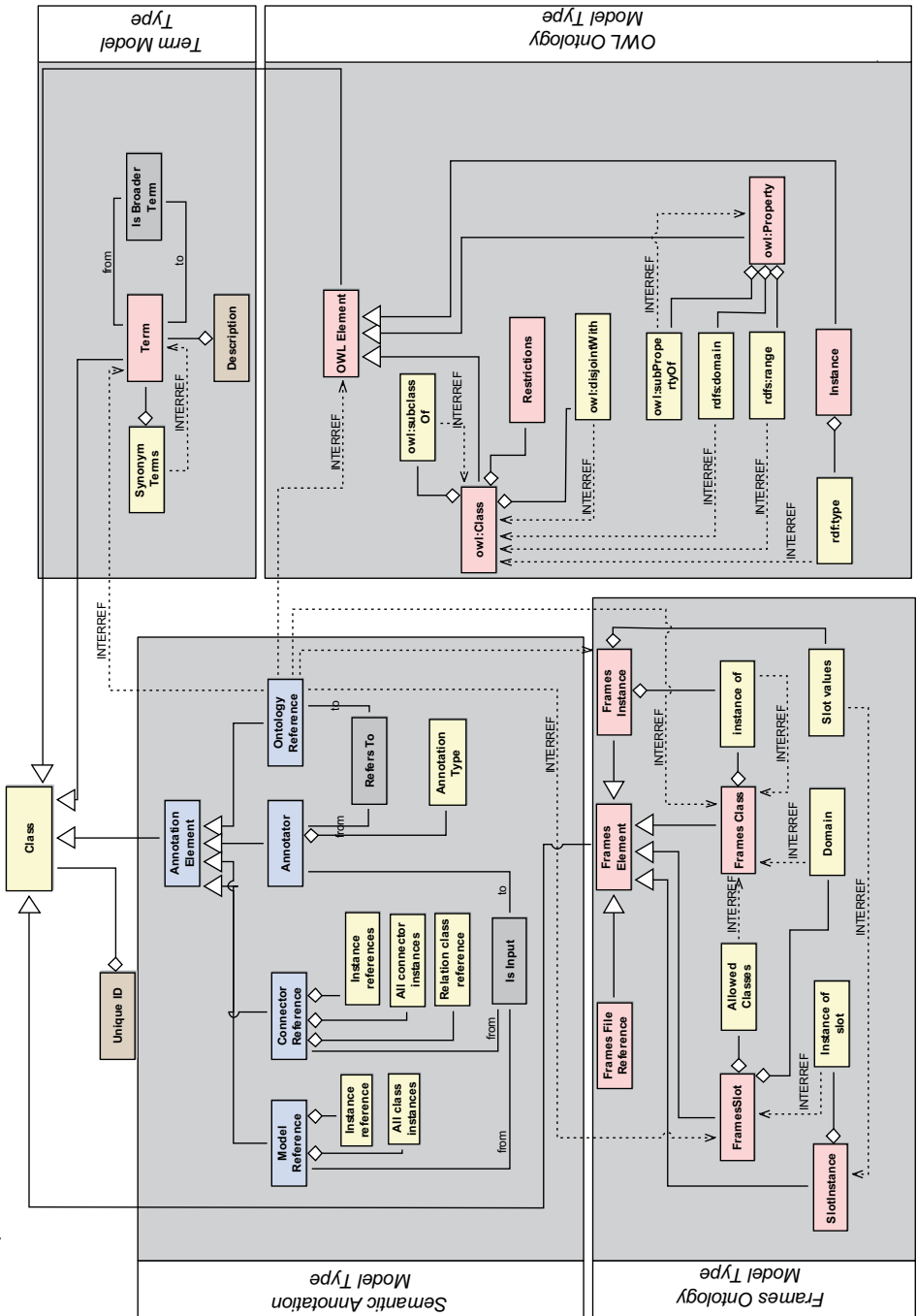


Figure 3: SeMFIS Metamodel [Fi16]

Besides the SeMFIS model types for ontologies and semantic annotations, the implementation also includes ADOxx components for accessing the platform via a SOAP-based web service interface, for processing the models via the ADOscript scripting language, as well as the AQL query component for querying model content in a proprietary query language. In addition, XML import and export interfaces are contained for exchanging model content.

3 Applications of SeMFIS

The approach of SeMFIS has been applied in the past to several scenarios. Some of the most prominent ones include the following. In [Fi12a] it has been described how business process models can be annotated with concepts from a risk ontology. In this way risks in business processes could be represented and subsequently analyzed using rules and simulation algorithms. In [Fi11b] the approach has been used to support the benchmarking of business processes by providing a common reference in the form of ontologies. In [FSK13] parts of the SeMFIS approach together with the web service interface have been used to embed the tool in a service-oriented architecture. This permitted to realize an interface to the Protégé ontology management platform.

Further scenarios that were based on the approach or re-used parts of it included the obfuscation of model information [Fi12c], the use in e-business scenarios [Fi09], for user-specific visualizations of models [FR11] or in the context of social network-based semantic annotations [Fi14].

4 Conclusion and Future Work

In this paper the concept of semantic-based modeling and the SeMFIS approach have been briefly described. For further information on SeMFIS it is referred to the SeMFIS website and the most recent publication on the platform [Fi16]. Future work on SeMFIS will include the further development of the concept and the technical implementation, as well as the application to further scenarios. As next steps it is planned to provide an API to ease the implementation of new algorithms on the SeMFIS platform. In addition, an existing plugin for the Stanford Protégé that permits to export ontologies in the SeMFIS XML format will be further enhanced and extended. Regarding future application scenarios it will be investigated how the approach can be used to represent requirements in the context of internet-of-things applications and social media management.

References

- [BBF09] Blair, G.; Bencomo, N.; France, R.B.: Models@run.time. *IEEE Computer*, 42(10):22–27, 2009. DOI: 10.1109/MC.2009.326.
- [DDFV07] Didonet Del Fabro, M.; Valduriez, P.: Semi-automatic model integration using matching transformations and weaving models. In: *SAC '07 Proceedings of the 2007 ACM symposium on Applied computing*. ACM, 2007.

- [De05] Del Fabro, M. D.; Bezivin, J.; Jouault, F.; Valduriez, P.: Applying Generic Model Management to Data Mapping. In: BDA 2005. 2005.
- [FB09] Fill, H.-G.; Burzynski, P.: Integrating Ontology Models and Conceptual Models using a Meta Modeling Approach. In: 11th International Protégé Conference. 2009.
- [Fi07] Fill, H.-G.; Gericke, A.; Karagiannis, D.; Winter, R.: Modellierung für Integrated Enterprise Balancing. *Wirtschaftsinformatik*, 06/2007:419–429, 2007.
- [Fi09] Fill, H.-G.: Design of Semantic Information Systems using a Model-based Approach. In: AAAI Spring Symposium. AAAI, 2009.
- [Fi11a] Fill, H.-G.: On the Conceptualization of a Modeling Language for Semantic Model Annotations. In (Salinesi, C.; Pastor, O., eds): *Advanced Information Systems Engineering Workshops, CAiSE 2011*, volume LNBIP Vol. 83. Springer, London, UK, pp. 134–148, 2011.
- [Fi11b] Fill, H.-G.: Using Semantically Annotated Models for Supporting Business Process Benchmarking. In (Grabis, J.; Kirikova, M., eds): *10th International Conference on Perspectives in Business Informatics Research. LNBIP 90*. Springer, pp. 29–43, 2011.
- [Fi12a] Fill, H.-G.: An Approach for Analyzing the Effects of Risks on Business Processes Using Semantic Annotations. In: *European Conference on Information Systems 2012. AIS*, 2012.
- [Fi12b] Fill, H.-G.: SeMFIS: A Tool for Managing Semantic Conceptual Models. In (Kern, H.; Tolvanen, J.-P.; Bottoni, P., eds): *Workshop on Graphical Modeling Language Development*. 2012.
- [Fi12c] Fill, H.-G.: Using Obfuscating Transformations for Supporting the Sharing and Analysis of Conceptual Models. In (Robra-Bissantz, S.; Mattfeld, D., eds): *MKW1 2012*. GITO Verlag, 2012.
- [Fi14] Fill, H.-G.: On the Social Network based Semantic Annotation of Conceptual Models. In (Buchmann, R.A.; Kifor, C.V.; Yu, J., eds): *7th International Conference on Knowledge Science, Engineering and Management*. Springer, pp. 138–149, 2014.
- [Fi16] Fill, H.-G.: SeMFIS: A Flexible Engineering Platform for Semantic Annotations of Conceptual Models. accepted for *Semantic Web (SWJ)*, 2016.
- [FJ16] Fill, H.-G.; Johannsen, F.: A Knowledge Perspective on Big Data by Joining Enterprise Modeling and Data Analyses. In: *Proceedings of the 49th Hawaii International Conference on System Sciences*. IEEE, pp. 4052–4061, 2016. DOI: 10.1109/HICSS.2016.503.
- [FK13] Fill, H.-G.; Karagiannis, D.: On the Conceptualisation of Modelling Methods Using the ADOxx Meta Modelling Platform. *Enterprise Modelling and Information Systems Architectures*, 8(1):4–25, 2013.
- [FR11] Fill, H.-G.; Reischl, I.: Stepwise Semantic Enrichment in Health-related Public Management by Using Semantic Information Models. In (Smolnik, S.; Teuteberg, F.; Thomas, O., eds): *Semantic Technologies for Business and Information Systems Engineering: Concepts and Applications*. volume 195–212. IGI Press, 2011.
- [FRK12] Fill, H.-G.; Redmond, T.; Karagiannis, D.: FDMM: A Formalism for Describing ADOxx Meta Models and Models. In (Maciaszek, L.; Cuzzocrea, A.; Cordeiro, J., eds): *Proceedings of ICEIS 2012, Wroclaw, Poland*, volume 3. pp. 133–144, 2012.

- [FSK13] Fill, H.-G.; Schremser, D.; Karagiannis, D.: A Generic Approach for the Semantic Annotation of Conceptual Models using a Service-oriented Architecture. *International Journal of Knowledge Management*, 9(1):76–88, 2013.
- [He05] Hepp, M.; Leymann, F.; Domingue, J.; Wahler, A.; Fensel, D.: Semantic business process management: A vision towards using semantic web services for business process management. In: *IEEE International Conference on e-Business Engineering*, 2005. ICEBE 2005. pp. 535–540, 2005.
- [Ho07] Hoefflerer, P.: Achieving Business Process Model Interoperability Using Metamodels and Ontologies. In (Oesterle, H.; Schelp, J.; Winter, R., eds): *15th European Conference on Information Systems (ECIS2007)*. University of St. Gallen, pp. 1620–1631, 2007.
- [HR00] Harel, D.; Rumpe, B.: *Modeling Languages: Syntax, Semantics and All That Stuff - Part I: The Basic Stuff*. Technical Report MCS00-16, The Weizmann Institute of Science, August 22, 2000 2000.
- [Hu14] Hui, G.: *How the Internet of Things Changes Business Models*. *Harvard Business Review*, July 29, 2014.
- [Ka06] Kappel, G.; Kapsammer, E.; Kargl, H.; Kramler, G.; Reiter, T.; Retschitzegger, W.; Schwinger, W.; Wimmer, M.: On Models and Ontologies - A Layered Approach for Model-based Tool Integration. In (Mayr, H.C.; Breu, R., eds): *Modellierung 2006*. volume GI-Edition-Lecture Notes in Informatics (LNI), P-82, pp. 11–27, 2006.
- [Ka08] Kaschek, R. H.: On the evolution of conceptual modeling. In (Delcambre, L.; Kaschek, R.H.; Mayr, H.C., eds): *Dagstuhl Seminar Proceedings*. volume 08181. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany, 2008.
- [Ne91] Neches, R.; Fikes, R.; Finin, T.; Gruber, T.; Patil, R.; Senator, T.; Swartout, W.R.: *Enabling Technology for Knowledge Sharing*. *AI Magazine*, 12(3):36–56, 1991.
- [Ro06] Rosemann, M.: Potential pitfalls of process modeling: part B. *Business Process Management Journal*, 12(3):377–384, 2006.
- [TF07] Thomas, O.; Fellmann, M.: *Semantic Business Process Management: Ontology-based Process Modeling Using Event-Driven Process Chains*. *IBIS*, 2(1):29–44, 2007.

Informatik hat Geschichte!

Michael Fothe¹

Abstract: Drei Themen aus der Geschichte der Informatik reichen aus, um einen Computer aus historischer Sicht „zusammenzubauen“. Die Themen beziehen sich auf das Rechnen, Programmieren und Kommunizieren und sind für einen Informatikunterricht geeignet, in den sich Schülerinnen und Schüler aktiv einbringen und der zu kreativen Überlegungen anregen soll. Das Vorgehen betrachtet „Geschichte als Steinbruch“; Roland Stowasser tat dies für den Mathematikunterricht bereits in den 1970er-Jahren.

Keywords: Informatikgeschichte, Schulfach Informatik, Kreativität

1 Einleitung

Die Wissenschaft Informatik hat bereits eine (Vor-)Geschichte, die je nach Kriterium auf 50, 75 oder 300 Jahre oder sogar darüber hinaus festzulegen ist. Nachfolgend werden Meilensteine thematisiert, die gut zusammen passen und sich sinnvoll ergänzen. Das Kapitel 2.1 „Die erste Vier-Spezies-Rechenmaschine“ besitzt nicht nur einen Bezug zum Leibniz-Jahr 2016, sondern auch zum Computer. Es beantwortet die Frage, wieso ein Computer überhaupt rechnen kann. Das Kapitel 2.2 „Die erste Konzeption eines frei programmierbaren Computers“ befasst sich mit der Analytical Engine und damit mit Programmierung. Das Kapitel 2.3 „Optische Telegrafie im 19. Jahrhundert“ thematisiert vor allem eine historische Telegrafienlinie in Preußen. Die drei Meilensteine widmen sich mechanischen Verfahren; hinzuweisen ist darauf, dass noch Zuses Z4 (sie ging 1950 an die ETH Zürich) ein mechanisches Speicherwerk hatte [Br15, S. 383].

2 Drei Meilensteine

2.1 Die erste Vier-Spezies-Rechenmaschine

Gottfried Wilhelm Leibniz (1646-1716) war Philosoph, Mathematiker, Politiker, Ratgeber der Fürsten, Wissenschaftsorganisator, Münzsammler aus Leidenschaft, Historiker, Physiker, Techniker, Sprachforscher, Universalgelehrter. Leibniz war auch „der schöpferische Promotor des dualen Zahlensystems“ [Ma00, S. 94]. Leibniz konstruierte prak-

¹ Friedrich-Schiller-Universität Jena, Fakultät für Mathematik und Informatik, Ernst-Abbe-Platz 2, 07743 Jena, michael.fothe@uni-jena.de

tisch sein gesamtes Berufsleben lang Rechenmaschinen. Leibnizens sogenannte jüngere Maschine wurde 1876 in Göttingen wieder aufgefunden. Sie ist die erste Rechenmaschine, die alle vier Grundrechenarten mechanisch ausführt und befindet sich heute in der Gottfried Wilhelm Leibniz Bibliothek in Hannover.

Im Informatikunterricht kann man diese Rechenmaschine thematisieren. Man kann z. B. an den Aufgaben $3 + 4$, $3 + 5$ und $3 + 6$ besprechen, wie sie addiert. Dabei wird herausgearbeitet, dass man ein Zahnrad mit einer flexiblen Anzahl an Zähnen benötigt (in unserem Beispiel müssen 4, 5 oder 6 Zähne einstellbar sein). Damit ist man bei den grundlegenden Erfindungen von Leibniz im Zusammenhang mit Rechenmaschinen: Staffelwalze und Sprossenrad. (Die Leibniz-Rechenmaschine verwendet Staffelwalzen.) Man kann sich im Unterricht auch damit befassen, wie mehrstellige Zahlen addiert werden. Nehmen wir das Beispiel $467 + 234$. In der ersten Phase wird gleichzeitig in allen Stellen addiert: $7 + 4 = 11$, $6 + 3 = 09$ und $4 + 2 = 06$. Der primäre Übertrag, der sich bei der Addition von $7 + 4$ ergibt, wird in der Stellung eines „Fünfhorns“ gespeichert (siehe Abb. 1). In der zweiten Phase werden von rechts nach links eventuelle Überträge bearbeitet. Bei der Einerstelle passiert nichts. Bei der Zehnerstelle wird $9 + 1 = 10$ gerechnet. Dies löst einen sekundären Übertrag aus, der nun seinerseits in einem anderen „Fünfhorn“ gespeichert wird. Abschließend wird bei der Hunderterstelle $6 + 1 = 7$ gerechnet. Primäre und sekundäre Überträge werden gleichartig behandelt, was ein wichtiges Konstruktionsprinzip dieser Rechenmaschine ist.

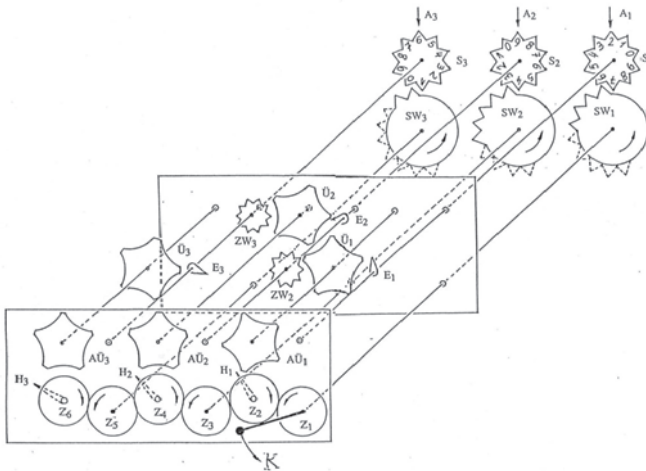


Abb. 1: Die deutlich sichtbaren „Fünfhörner“ sind Speicherelemente für primäre und sekundäre Überträge [Le93, S. 8].

Das Rechenwerk befähigt einen Computer zum Rechnen. Es muss also mindestens über einen Leistungsumfang verfügen, wie ihn die Leibniz-Rechenmaschine hat.

2.2 Die erste Konzeption eines frei programmierbaren Computers

Ada Countess of Lovelace (1815-1852) erkannte bereits Mitte des 19. Jahrhunderts relevante Möglichkeiten und Chancen des Computers – also 100 Jahre bevor der erste Computer fertiggestellt war und zum Einsatz kam [Fo16]. Lovelace übersetzte einen Aufsatz des Luigi Federico Menabrea (1809-1896) aus dem Französischen ins Englische und ergänzte den Text umfangreich mit sieben Anmerkungen. Zur Vorgeschichte: Charles Babbage (1791-1871), der Erfinder der Analytical Engine, nahm im Jahr 1840 an einer Versammlung italienischer Naturforscher in Turin teil und erläuterte dort Grundidee und Details zu deren Konzeption. Auf dieser Grundlage erarbeitete Menabrea seinen Aufsatz von 1842. Dessen englische Übersetzung ergänzt um Lovelaces Anmerkungen erschien dann 1843 und ist die erste vollständige Beschreibung der Analytical Engine und damit die erste Konzeption eines frei programmierbaren Computers [ML96]. Gebaut wurden von der Analytical Engine nur einige Teile. Auf das Computerprogramm zum Berechnen von Bernoulli-Zahlen, das Lovelace in der Anmerkung G entwickelt, ist besonders hinzuweisen. Nach derzeitigem Kenntnisstand handelt es sich um das „erste namhafte Programm“ für einen Computer [Kr15, S. 7].

Die Analytical Engine sollte über mindestens 200 Spalten verfügen. Die Spalten werden Variablen genannt und mit V_0, V_1, V_2, \dots bezeichnet. Es sind auch Dezimalbrüche vorgesehen (Festpunktzahlen). Spezifisches Merkmal der Analytical Engine ist die Verwendung von Lochkarten, mit deren Hilfe die Programmierung erfolgt. Die umfassenden Fähigkeiten der Analytical Engine beruhen auf dem Prinzip, das Joseph-Marie Jacquard (1752-1834) in dem von ihm erfundenen Webstuhl seit 1805 einsetzte. Lovelace schreibt dazu einen viel zitierten Satz [ML96, S. 335]:

Am treffendsten können wir sagen, daß die Analytical Engine *algebraische Muster* webt, gerade so wie der Jacquard-Webstuhl Blätter und Blüten.

Es gibt Operations-, Zulieferungs- und Empfangskarten. Die Zulieferungs- und Empfangskarten sind Variablenkarten, da sie sich auf eine Spalte beziehen. Die *Operatiionskarten* (es gibt solche für die Addition, Multiplikation, Subtraktion und Division) versetzen die Mühle (so heißt das Rechenwerk der Analytical Engine) in den Additionszustand, Multiplikationszustand usw. Lovelace erläutert dazu [ML96, S. 345]:

In jedem dieser Zustände ist der Mechanismus bereit, in der für diesen Zustand charakteristischen Weise zu arbeiten, und zwar mit jedwedem Zahlenpaar, das man in seinen Aktionsradius gelangen läßt.

Eine *Zulieferungskarte* führt dazu, dass die Zahl dieser Variablen in die Mühle übergeht, um dort verarbeitet zu werden. Bei Verwendung einer *erhaltenden Zulieferungskarte* kehrt der Wert, nachdem er benutzt worden ist, von der Mühle in die Variable zurück. Bei Verwendung einer *Null-Zulieferungskarte* wird die Variable auf Null gestellt. Eine *Empfangskarte* führt dazu, dass eine Variable eine Zahl aus der Mühle empfängt. Es kann sich dabei auch um eine Variable handeln, deren Wert zuvor mit einer Zulieferungskarte der Mühle übergeben wurde. Hat man mehrere Operationen der gleichen Art

(z. B. Multiplikationen) unmittelbar nacheinander auszuführen, so muss man nur eine Operationskarte einsetzen. Die Mühle bleibt im Multiplikationszustand. Nach jeweils zwei Zulieferungskarten und mindestens einer Empfangskarte wird eine Multiplikation ausgeführt. Gibt man mehrere Empfangskarten nacheinander an, so wird das Ergebnis der Operation in mehreren Variablen abgelegt. Ein Beispielprogramm (schematisch) enthält Abb. 2.

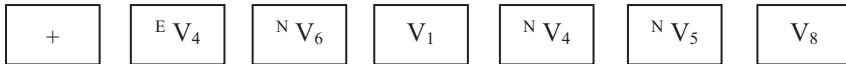


Abb. 2: Die Folge von Lochkarten steht für die Wertzuweisungen $V_1 = V_4 + V_6$, $V_6 = 0$, $V_8 = V_4 + V_5$, $V_4 = 0$, $V_5 = 0$. Die Pluskarte ist eine Operationskarte, die die Mühle in den Additionszustand versetzt. Die Variablenkarte mit dem „E“ ist eine erhaltende Zulieferungskarte, die Variablenkarten mit einem „N“ sind Null-Zulieferungskarten. Bei der ersten Addition ist V_1 , bei der zweiten Addition V_8 die Empfangskarte.

Die Analytical Engine verfügt über die algorithmischen Grundstrukturen Verzweigung und Schleife. Bei der Verzweigung wird mithilfe einer Test-Operation entschieden, welche Lochkarten abgearbeitet werden. Menabrea deutet Schleifen an, indem er ausführt, dass eine Karte einen bestimmten Registrierapparat instruieren wird, eine Variable von n über $n-1$, $n-2$ usw. bis zur 0 herunterzuzählen. Jedes Mal erfolgt eine Multiplikation [ML96, S. 324]. Lovelace formuliert allgemeiner [ML96, S. 347]:

Ziel dieser Erweiterung ist, die Möglichkeit zu eröffnen, zur Lösung eines Problems jede spezielle Karte oder Menge von Karten *beliebig oft hintereinander* einzusetzen.

Solche Schleifen sollten durch ein Rücklaufsystem realisiert werden.

Lovelace definiert [ML96, S. 363]:

Eine Schleife, die n andere Schleifen – *eine in die andere verschachtelt* – umfaßt, wird eine Schleife $(n+1)$ ter Ordnung genannt.

Im Informatikunterricht kann man darüber diskutieren, wie Verzweigung und Schleife mithilfe von Lochkarten realisiert werden könnten.

2.3 Optische Telegrafie im 19. Jahrhundert

Eine optische Telegrafienlinie ist ein einfaches Kommunikationssystem. Das Internet ist erheblich komplizierter, selbst wenn man sich auf Wesentliches konzentriert. Anhand der optischen Telegrafie können im Informatikunterricht Sachverhalte besprochen werden, die auch bei modernen Kommunikationssystemen relevant sind. Man denke an das Codieren, Komprimieren, Verschlüsseln, an das Erkennen und Korrigieren von Übertragungsfehlern, an das Nutzen von verschiedenen Kanälen beim Versenden einer E-Mail,

an das Angeben des vom Sender verwendeten Zeichensatzes und an die Zeitsynchronisation von Rechnern über das Internet. Die Schülerinnen und Schüler können bereits in der Sekundarstufe I auf altersgemäßem Niveau für Fragestellungen sensibilisiert werden, die auch im Internet auftreten, auch wenn zuzeiten der optischen Telegrafie die Paketvermittlung noch gar nicht erfunden war. Die große Zeit der optischen Telegrafie waren die ersten Jahrzehnte des 19. Jahrhunderts. Jürgen Osterhammel weist in seiner Weltgeschichte auf das Ziel der Telekommunikation hin, Informationen schneller reisen zu lassen als Menschen und Objekte, und auch darauf, dass *das* große neue Medium mit globalisierender Wirkung im 19. Jahrhundert der Telegraph war. Er stellt heraus, dass eine telegraphisch vernetzte Welt bereits um 1800, also *vor* der entsprechenden Technik, denkbar geworden war und zwar durch optische Signalübermittlung [Os09, S. 1024 f.]. Dieser Hinweis stellt die optische Telegrafie in eine Entwicklungslinie hinein und macht sie damit bedeutsam.



Abb. 3: Eine Briefmarke mit der Station 2 (St. Annenkirche in Berlin-Dahlem) und zwei Inspektoren, die die Telegrafienlinie beaufsichtigen.

Im Unterricht wird z. B. die optisch-mechanische Telegrafienlinie zwischen Berlin und der Rheinprovinz thematisiert, die das Königreich Preußen von 1833 bis 1849 unterhielt [He78]. Voll ausgebaut umfasste die Telegrafienlinie 62 Stationen (siehe Abb. 3). Sie wurde ausschließlich für staatliche und militärische Nachrichten genutzt. Im Informatikunterricht kann man Antworten zu folgenden Fragen besprechen [Fo10, S. 55-60]:

- Wie ist ein Zeichen aufgebaut?
- Wie lässt sich ein Zeichen beschreiben?
- Wie viele Zeichen gibt es?
- Wie wird eine Nachricht codiert?
- Welchen Informationsgehalt besitzt ein Zeichen?

- Wie groß war die Schrittgeschwindigkeit? Welches Datenvolumen wurde übertragen?
- Wie lange war ein Zeichen von Berlin nach Koblenz unterwegs?
- Wie wurden Übertragungsfehler festgestellt? Wie wurden diese korrigiert?
- Welche Uhrzeit galt auf der Telegrafienlinie?
- Wie wurde eine Depesche von Paris nach Berlin transportiert?

Die optische Telegrafie eignet sich als Gegenstand von Projektarbeiten. Die Schülerinnen und Schüler erhalten die Aufgabe, ein Handbuch für die Kommunikation zu erarbeiten. Wesentlicher Bestandteil des Handbuchs wäre die Beschreibung des Protokolls für die Kommunikation. Dabei hätten die Schülerinnen und Schüler weitere relevante Fragen aufzuwerfen und Antworten zu finden, so u.a.:

- Wie ist mit Eildepeschen umzugehen?
- Wie ist vorzugehen, wenn eine normale Depesche oder sogar eine Eildepesche wegen Nebels nicht weitergeleitet werden kann?
- Wie ist in einer Station zu verfahren, wenn sich zwei Depeschen begegnen – die eine kommt aus Berlin, die andere aus Koblenz? Oder lässt sich diese Situation sogar vermeiden?
- Wie ist dem Empfänger einer Depesche mitzuteilen, ob bei der Codierung das Code-Buch oder das Telegrafistenwörterbuch verwendet wurde?
- Wie erfolgt die Adressierung einer dienstlichen Depesche an eine andere Station?

Bei alledem erfinden die Schülerinnen und Schüler ein Kommunikationssystem teilweise neu. Sie stellen dabei Überlegungen an, die bereits vor 200 Jahren angestellt wurden.

Man kann verschiedene Telegrafienlinien vergleichen, so z. B. hinsichtlich des Informationsgehalts eines Zeichens [BM95]. Ein „preußisches Zeichen“ wurde aus sechs Telegrafienflügeln gebildet (drei Flügelpaare); jeder Flügel konnte vier Stellungen einnehmen. Ein Zeichen hat also 12 Bit Information. In einem englischen System hatte man sechs Klappen, die jeweils zwei Stellungen einnehmen konnten. Der Informationsgehalt eines Zeichens beträgt daher 6 Bit. Mit zwei „englischen Zeichen“ kann man also genau die gleiche Informationsmenge wie mit einem „preußischen Zeichen“ übertragen. Im französischen System hatte man einen langen Balken („Regulator“) und an ihm befestigt zwei kurze Arme („Indikatoren“). Regulator und Indikatoren konnten jeweils verschiedene Stellungen einnehmen (man konnte sie in 45°-Schritten verstellen). Von 256 Zeichen wurden nur 92 als Grundzeichen verwendet; die anderen Zeichen stellten sich als zu schlecht unterscheidbar heraus. Zwei Grundzeichen wurden hintereinander gesendet, sodass man 92² Möglichkeiten hatte (also 8.464). Der Informationsgehalt eines Doppelzeichens beträgt rund 13 Bit.

3 Ausblick

Man kann sicher weitere Arrangements von informatikgeschichtlichen Themen entwickeln und im Unterricht einsetzen. Ein Beispiel dafür wäre das Thematisieren des Kellers mit seinen diversen Anwendungen [FW15].

Literaturverzeichnis

- [BM95] Beyrer, K.; Mathis, B.-S. (Hrsg.): So weit das Auge reicht. Die Geschichte der optischen Telegrafie; eine Publikation des Museums für Post und Kommunikation, Frankfurt am Main, anlässlich der gleichnamigen Ausstellung. Braun, Karlsruhe 1995.
- [Br15] Bruderer, H.: Meilensteine der Rechentechnik. Zur Geschichte der Mathematik und der Informatik. Walter de Gruyter, Berlin/Boston 2015.
- [Fo10] Fothe, M.: Kunterbunte Schulinformatik – Ideen für einen kompetenzorientierten Unterricht in den Sekundarstufen I und II. LOG IN Verlag, Berlin 2010.
- [Fo16] Fothe, M.: Die berühmten Anmerkungen – Zum 200. Geburtstag von Ada Countess of Lovelace. LOG IN, Nr. 183/184, S. 12–18, 2016.
- [FW15] Fothe, M.; Wilke, T. (Hrsg.): Keller, Stack und automatisches Gedächtnis – eine Struktur mit Potenzial. Lecture Notes in Informatics – Thematics. Band T-7, Bonn 2015.
- [He78] Herbarth, D.: Die Entwicklung der optischen Telegrafie in Preussen. Landeskonservator Rheinland. Arbeitsheft 15. Rheinland-Verlag, Köln 1978.
- [Kr15] Krämer, S. (Hrsg.): Ada Lovelace. Die Pionierin der Computertechnik und ihre Nachfolgerinnen. Wilhelm Fink, Paderborn 2015.
- [Le93] Lehmann, N. J.: Neue Erfahrungen zur Funktionsfähigkeit von Leibniz' Rechenmaschine. Studia Leibnitiana, Band XXV/2, S. 1-15, 1993.
- [Ma00] Mackensen, L. v.: Die ersten dekadischen und dualen Rechenmaschinen. In (Popp, K.; Stein, E. Hrsg.): Gottfried Wilhelm Leibniz. Das Wirken des großen Universalgelehrten als Philosoph, Mathematiker, Physiker, Techniker. Universität; Schlütersche, S. 85-100, Hannover 2000.
- [ML96] Menabrea, L. F.; Lovelace, A.: Grundriß der von Charles Babbage erfundenen Analytical Engine. In (Dotzler, B. Hrsg.): Babbages Rechen-Automate. Ausgewählte Schriften. Computerkultur Band VI. Springer, Wien/New York, S. 309-381, 1996.
- [Os09] Osterhammel, J.: Die Verwandlung der Welt. Eine Geschichte des 19. Jahrhunderts. Beck, München 2009.

IoT und Sensornetzwerke: Entwurf und Programmierung von Niedrigstenergiesystemen anhand einer Metaarchitektur

Volker Skwarek¹, Thorsten Kistler², Mark Rawer² und Stefan Schauer³

Abstract: Ein wesentlicher Erfolgsfaktor von Sensornetzwerken wird in deren Autonomie und damit auch dem Energiebedarf liegen. Der Energiebedarf wird aber zunächst wenig vorhersagbar in allen Schichten des Hard- und Softwaredesigns beeinflusst. Daher muss der Energiebedarf für eine Niedrigenergieapplikation das primäre Entwurfsmerkmal sein. In diesem Beitrag wird ein Architekturkonzept vorgestellt, mit dem Energiebedarfe systematisch geplant werden können.

Keywords: Niedrigstenergiesysteme, Sensornetze, Architektur, Entwurfsrichtlinien

1 Einleitung

Die aktuellen Trends der x-4.0-Entwicklungen wie Industrie 4.0, Internet 4.0, o. Ä. setzen einen höheren Informationsgrad durch eine intensivere, vernetztere Auswertung von Informationen um. Diese Informationen werden dann beispielsweise über das Internet-of-Things (IoT) [MF10, S. 1] von *smart objects* eingesammelt werden, die aufgrund ihrer Größe, Beschaffenheit, Vielzahl, Redundanz aber auch aufgrund ihres individuellen Wertes als *smart dust* bezeichnet werden [DP10], [KKP99, S. 8]. Als größte Herausforderung zur Funktion solcher Sensoren und damit auch von deren Akzeptanz wird der Energiebedarf der Sensorknoten genannt:

”A major challenge is to incorporate all these functions while maintaining very low power consumption, thereby maximizing operating life given the limited volume available for energy storage. the dust mote power consumption cannot exceed roughly 10 microwatts. The functionality envisioned for Smart Dust can be achieved only if the total power consumption of a dust mote is limited to microwatt levels, and if careful power management strategies are utilized.“[KKP99, S. 271f.].

Schon in dieser Kernpublikation der drahtlosen Sensornetzwerke Ende der 1990er Jahre besteht also die elementare Anforderung eines sehr restriktiven Energiemanagements. Ob diese Arbeit Ursprung oder Teil eines neuen Trends war, lässt sich nicht eindeutig

¹ Hochschule für Angewandte Wissenschaften Hamburg, Ulmenliet 20, D-21033 Hamburg, volker.skwarek@haw-hamburg.de

² Cypress Semiconductor GmbH, Willy-Brandt-Allee 4, D-81829 München, thorsten.kistler@cypress.com, mark.rawer@cypress.com

³ Texas Instruments Deutschland GmbH, Haggertystr. 1, D-85356 Freising, s-schauer1@ti.com

feststellen. Allerdings ist insbesondere zwischen 2000 und 2010 eine hohe Anzahl von Publikationen zum Energiemanagement und zur Energiebedarfsprädiktion festzustellen. Exemplarisch können folgende Bereiche benannt werden: Chipdesign, Energiespeicherung und -erzeugung, Befehlssätze, Architektur und Programmzyklensimulation oder Kommunikationsprotokolle.

Allen diesen Aspekten ist es gemeinsam, dass der jeweilige Schwerpunkt auf einzelne System Schwerpunkte gelegt wurde, aber nicht auf das gesamte System. Hieraus entsteht dann das Problem, dass eine Energiebedarfsoptimierung immer nur ein Angebot zur Nutzung ist, nie aber verpflichtend, weshalb sie einfach umgangen werden kann. Folglich muss die Nutzung von energieminimierenden Eigenschaften durch das gesamte System hindurch geplant werden.

Dieser Planungsaspekt steht in dieser Publikation im Vordergrund: Anhand eines sehr einfachen Sensor knotens werden exemplarisch hohe Energiebedarfe dargestellt (Kapitel 2.1) und mit Hilfe eines aus vielen Beispielen empirisch entwickelten Energieschichtenmodells unter dem Aspekt der Energiesystemplanung diskutiert (Kapitel 2.2). Im Kapitel 3 wird dann dieser Sensor knoten mit Hilfe des Schichtenmodells in Teilen neu geplant. In diesem Beitrag ist zu berücksichtigen, dass vollständige Anleitung und Umsetzung innerhalb des Workshops erfolgt und hier nur wenige Fokuspunkte betrachtet werden können

2 Schichtenmodell zur Energiebedarfskonzeption

Auf der Modellierungsebene sind bisher Energiemodelle bekannt, die konkrete komplexe Applikationen wie beispielsweise Prozessabläufe in Fabriken auf Systemebene abbilden [Fel+13] oder [SSS13]. Im Gegensatz dazu wird hier ein generisches Modell vorgestellt, das sich an Softwarearchitekturen bzw. dem Softwareentwicklungsprozess orientiert.

2.1 Energiebedarfsermittlung am Beispiel eines Lichtsensors

Zunächst soll als Beispiel eine Energiebedarfsanalyse für eine einfache Applikation herangezogen werden: ein autonomer Lichtsensor. Diese Anwendung ist aufgrund ihrer einfachen Architektur gewählt, da diese keine grundsätzliche Herausforderung an deren Entwicklung und Umsetzung darstellt. Die Architektur dieses Systems ist in der folgenden Abbildung 1 dargestellt. Der Programmablauf umfasst das Messen der Umgebungshelligkeit mit einem Photowiderstand 5 Mal pro Sekunde und das Senden des Messwertes per Bluetooth Low Energy (BLE) an ein zentrales Lichtsteuergerät. Der Betriebszustand wurde mit einer grünen Low-Power-LED angezeigt. Es wurde mit dem

Energy Monitor von EEMBC⁴ der mittlere Strombedarf einzelner Komponenten oder Maßnahmen über 10 Sekunden ermittelt, indem diese sukzessive modifiziert wurden

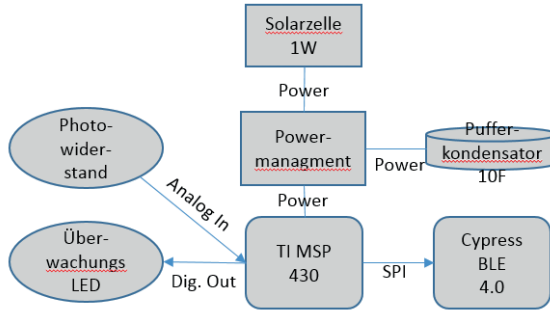


Abb. 1: Architektur eines Beispielsensors..

In Tabelle 1 wird deutlich, wie exemplarische Designmaßnahmen in Code, Hardware und Algorithmus jeweils zu signifikanten Verbesserungen im Energiebedarf des Systems geführt haben. Nach allen - teilweise in der Tabelle auch nicht aufgeführten – Maßnahmen konnte die über 10 Sekunden gemittelte Stromaufnahme von ca. 3mA auf 30µA zuzüglich des BLE-Kommunikationsmoduls gesenkt werden.

Maßnahme	Stromreduktion in µA
Low Power LED	1490
Eingangskonfiguration als Input mit Pullup-Widerstand	2070
Aktivieren des Photowiderstandes über GPIO erst zum Messzeitpunkt	40
Aktivieren des Low-Power-Mode 3 von 4 mit Interruptsteuerung	227

Tab. 1: Mittlerer Strombedarf einzelner Komponenten und Strukturen des Sensors.

2.2 Empirisches Architekturmodell zur Energiebedarfsplanung

Die Energiearchitektur in Abbildung 2 wurde als Strukturmodell zur Energieplanung entworfen. Als Vorbild wurde die ISO/OSI-Kommunikationsarchitektur verwendet [ISO94, Abb. 11].

Die grobe Unterteilung sieht zunächst eine Separation des Problems in software- und hardware-spezifische Anteile vor, was dem oberen und unteren Ende des Modells in Abbildung 2 entspricht. Als Beispiel könnte ein fiktiver Signalvergleich per Korrelation herangezogen werden: Dieser erfolgt in der Regel über analoge Eingangsspiegel. Bei stark verrauschten Signalen würde aber eine Polaritätskorrelation - also eine binäre

⁴ <http://www.eembc.org/ulpbench/about.php>

Korrelation der Vorzeichen in Bezug auf einen Mittelwert - zu vergleichbaren Ergebnissen führen [Pue15, Kapitel 6.3.4]. Selbst wenn an dieser Stelle noch immer eine softwarebasierte Korrelation erfolgt, was nicht zwingend erforderlich wäre, könnte der Energiebedarf dafür schon durch folgende Maßnahmen deutlich gesenkt werden:

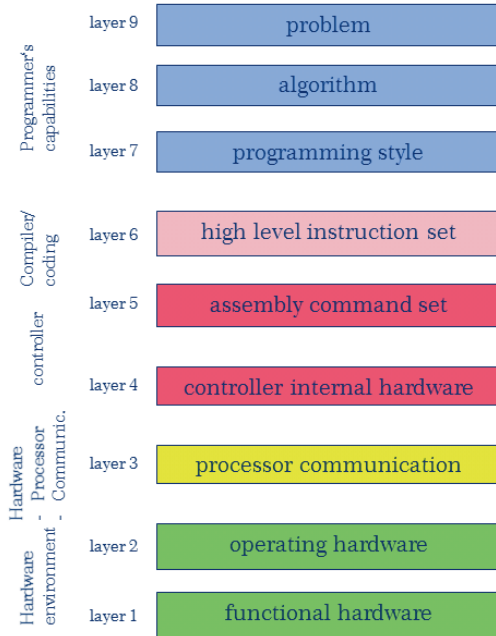


Abb. 2: Modell zu energieoptimierendem Systemdesign

- hardwareseitige Vorverarbeitung der Analo­gsignale in amplitudendiskrete Digital­signale durch einen Schwellwertkomparator.
- Verzicht auf Analog-Digital-Wandler,
- Reduktion der prozessorinternen Berechnungsauf­lö­sung auf Ganz- oder sogar Binärzahlen sowie Verzicht auf Fließkommaoperationen.

Kerngedanke dieser Architektur ist also die genaue Analyse des erforderlichen Algorithmus hinsichtlich der Zielanwendung und -genauigkeit. Im Detail übernehmen die einzelnen Schichten folgende Funktionen, beginnend mit der unteren Schicht der funktionalen Hardware:

Layer 1: functional hardware Im Rahmen der algorithmischen Separation auf Soft- und Hardware wird überlegt, welche Bestandteile des Systems besser hardwareseitig realisiert werden können. Dies sind idealerweise solche, die die Auflösung des Signal zur weiteren Verarbeitung im Prozessor herabsetzen und

damit Rechenaufwand reduzieren, oder solche, die algorithmische Anteile auf einfache physikalische mathematischen Gesetze zurückführen wie Multiplikationen durch Mischer oder Integrationen/Differentiationen auf rückgekoppelte Operationsverstärker. Hier ist allerdings unbedingt zu beachten, dass auch negative Rahmenbedingungen wie sie oft durch Temperaturabhängigkeiten von Hardwarekomponenten auftreten, mit betrachtet werden.

Layer 2: operational hardware beschreibt sämtliche Komponenten zum Betrieb eines Mikrocontrollers sowie dessen geeignete Auswahl selbst. Hier steht die Bewertung im Vordergrund, inwieweit der Controller zur Anforderung passt und welche Peripherie wie beschaltet werden muss. Klassische Auslegungsaspekte wie Speicherbedarf und Taktfrequenz treten hier eher in den Hintergrund. Wichtiger sind Aspekte, die das Sleep- und Wakeup-Timing sowie Umfang und Funktion von Niedrigenergiemodi betreffen. Da das Heruntertakten einen Kernaspekt der Niedrigenergieprogrammierung darstellt, sollten folgende Fragen beantwortet werden: welche Controllerkomponenten erfordern welche Timer zum Betrieb und wie weit lassen diese sich heruntertakten? Welche Komponenten funktionieren noch in welchem Niedrigenergiemodus und wie lang sind deren Wake-up-Zeiten? Ist gegebenenfalls ein kaskadiertes Aufwachkonzept erforderlich, das wiederum zu weiteren Latenzzeiten führt? Hieraus ergibt sich dann die Controllerauswahl sowie die betriebserforderlichen Peripheriekomponenten und Beschaltungen.

Layer 3: processor configuration Unter der Vielzahl der drahtgebunden und drahtlosen Möglichkeiten, Daten aufzunehmen und nach außen zu kommunizieren, muss das der Messaufgabe angemessene energiesparendste Verfahren ausgewählt werden. Es sind dabei nicht nur die vordergründigen Aspekte wie Datenrate und Reichweite zu berücksichtigen, sondern auch das Kommunikationsbasisprotokoll, das die Nettodatenrate und den Overhead definiert, Handshake- und Latenzzeiten nach dem Wake-up aber auch die Konfigurierbarkeit von Protokollen oder Diensten.

Layer 4: controller internal hardware Aus der Selektion des Controllers sowie der Erfassungs- und Verarbeitungsaufgabe ergeben sich meist unterschiedliche Möglichkeiten, diese innerhalb des Controllers umzusetzen. Oft werden unreflektiert Basiskonfigurationen übernommen. Zu ungenutzten Anpassungsmöglichkeiten gehören unter anderem die Herabsetzung der Auflösung von AD-Wandlern auf den minimalerforderlichen Wert aber auch die Konfiguration von Watch-Dogs, Capture/Compare-Strukturen oder das Setzen von ungenutzten Kommunikationsports auf hochohmige Eingänge. Kurz gefasst ist hier das intensive Auseinandersetzen mit der Beschreibung des Controllers erforderlich.

Layer 5: assembly command set Die (genutzte) Befehlsstruktur des Controllers definiert sehr genau, wie welche Komponenten der central processing unit (CPU) bzw. des Controllers verwendet werden. ÜBLICHERWEISEblicherweise kann dies sehr gut durch Compiler-Einstellungen sowie den Programmierstil definiert werden. Mangels einer CompilerEinstellung "energiesparend" ist es hier aber

erforderlich, das Ergebnis der Codegenerierung kritisch zu analysieren und gegebenenfalls in Assembler nachzuoptimieren.

Layer 6: high level instruction set Heutzutage ist die Programmierung eines Mikrocontrollers in Assembler eher selten, sondern es wird auf C zurückgegriffen. Grundsätzlich ist hier aber weiterhin eine systemnahe Programmierung erforderlich, die beispielsweise durch Pointer und Bitmanipulation realisiert werden kann. In modernen Prozessorarchitekturen wird daher eine Abstraktionsschicht angeboten, die diese Bitmanipulation durch Klartextbefehle ersetzt. Hier können aber auch Funktionen zusammengefasst worden sein, die zwar den allgemeinen Umgang mit dem Controller erleichtern, aber nicht im Sinne der Energieeffizienz optimieren. Daher ist es sinnvoll, bei Auffälligkeiten im Energiebedarf diese Routinen zu überprüfen und gegebenenfalls doch die bitbasierte systemnahe Programmierung umzusetzen.

Layer 7: programming style Der Aspekt des Programmierstils ist von dem der Algorithmik zu trennen. Hierzu gehören beispielsweise Aspekte wie Objektorientierung, Verwendung von auf- oder abzählenden Schleifen oder umfangreiche mathematische Operationen. Bei allen diesen Aspekten gilt in erster Näherung als Grundregel: alles, was die Lauf- und Rechenzeit reduziert, optimiert sehr wahrscheinlich auch aufgrund der geringeren Zyklenzahl den Energiebedarf. Natürlich muss auch hier berücksichtigt werden, dass bestimmte Aufgaben von spezialisierten Strukturen auf einem Controller - z. B. digitale Signalprozessoren (DSPs) übernommen werden können, deren Energieeffizienz aber in Frage gestellt werden darf.

Layer 8: algorithm Ein Algorithmus als Kernelement einer rechnerischen Problemlösung entscheidet üblicherweise deutlich über den Rechenaufwand und somit auch über den Energiebedarf. Als Beispiel sei nur die rekursive Programmierung erwähnt: Diese gilt oft als Königsdisziplin der Algorithmik, da ein Problem so beschrieben wird, dass es sich selbst lösen kann. Hieraus entsteht oft ein sehr kurzer, kompakter Code. Der permanente Selbstaufruf hat allerdings im Rechnerkern zur Folge, dass leistungsoptimierende Strukturen wie Pipelines noch unzureichend genutzt werden können. Heap- und Stackspeicher als schnelle Speicher zur Aufnahme von Systemzuständen bei Sprüngen können überlaufen und in langsamere RAM-Speicher verschoben werden. Folglich sind kurze, knappe und einfache Algorithmen in der Regel solchen vorzuziehen, die eine rechnerisch-mathematische Eleganz umzusetzen versuchen.

Layer 9: problem Der Aspekt der Problemformulierung ist elementar für die Systemscheidung. Es existieren beispielsweise derzeit keine praktischen Ansätze, um ein diagnostisches Dauer-Elektrokardiogramm über 24 Stunden mit drahtlosen Sensorsystemen übertragen zu können, ohne auf externe Energiezufuhr zurückgreifen zu müssen (Stichwort: *drahtlose EKG-Elektrode*). Hier sind ca. 4000 Samples pro Sekunde im Millivolt-Bereich mit hoher Auflösung erforderlich. Eine

Randbedingung, die der Leistungsfähigkeit heutiger batteriebetriebener Sensorsysteme widerspricht. Andere Anwendungen wären beispielsweise eine wartungsfreie Überwachung von Umweltparametern mit garantierter Datenverfügbarkeit über einen langen Zeitraum. Verfügbarkeitsgarantien widersprechen heute noch dem Ansatz von Niedrigenergiesystemen und Energy Harvesting, da genau im deren Gegensatz die Philosophie des Designs besteht: Im Zweifel das System schlafen legen, bei Datenaufkommen – gegebenenfalls etwas zu spät - aufwecken und im Zweifel bei Energiemangel solange warten, bis die Speicher wieder aufgeladen sind.

Anhand der vorhergehenden Überlegungen wird deutlich, dass Energiemanagement ein primäres Entwurfsziel sein muss, das anhand des zuvor vorgestellten Vorgehens planerisch in alle Stufen der Architektur und Phasen der Umsetzung integriert werden muss. Dabei kann in jeder Planungsstufe erneut die Entscheidung zwischen Algorithmus und Energieeffizienz erforderlich werden, die zulasten einer der beiden Seiten ausfällt. Im Zweifel kann unter bestimmten Voraussetzungen kein autonomer Sensorknoten entworfen werden oder der Algorithmus/das Anwendungsgebiet muss unter Abstrichen realisiert werden.

3 Architekturbasierte Modellierung eines Sensorknotens

Im folgenden soll kurz dargestellt werden, wie der in Kapitel 2.1 exemplarisch eingeführte Lichtsensor unter Berücksichtigung der Architektur implementiert werden könnte. Dabei wird zunächst die Verarbeitungseinheit von der Kommunikationseinheit getrennt betrachtet, um die Gesamtkomplexität zu reduzieren. Als zentrale Fragestellung wird zunächst *Layer 9 (problem)* betrachtet, ob die Anwendung eines autonomen Lichtsensors mit gegebenenfalls sogar sicherheitsrelevanten Schaltaufgaben einer Niedrigstenergieanwendung widerspricht. Dabei ist zu beantworten, ob für die Anwendung in ausreichend schneller Datenrate Helligkeitsinformation zur Verfügung gestellt werden können und ein Ausbleiben aufgrund durch Energiemangel in Maßen toleriert werden kann. Aufgrund einer erforderlichen Reaktionszeit von 500ms reichen Messungen im Intervall von 200ms, um selbst bei Verlust eines Datenwortes und ungünstiger Intervalllage zumindest einen Datensatz zu übertragen. Weiterhin kann bei Ausbleiben von Daten auf der Empfängerseite ein Fail-Safe-Modus mit dem Standardwert "Licht einschalten" implementiert werden. Auch die Kommunikationsreichweiten zwischen Sender und Empfänger sind mit wenigen Metern für übliche drahtlose Übertragungsverfahren problemlos umsetzbar. Daraus folgt, dass das Problem grundsätzlich für eine Niedrigenergieanwendung geeignet ist.

3.1 Sensorsystem

Die *funktionale Hardware (Layer 1)* kann so gestaltet werden, dass der relevante Helligkeitsbereich zwischen 1000lx (Licht an) und 6000lx (Licht aus) beispielsweise

durch Photowiderstände abgebildet werden kann, die jeweils zum Messzeitpunkt über einen digitalen Eingang aktiviert werden.

Die *operative und prozessorinterne Hardware (Layer 2 und 4)* benötigt daher nur einen AD-Wandler mit einer vergleichsweise geringen Auflösung - z. B. 8 Bit - und Präzision. Diese Aspekte sind deshalb eher unkritisch, da die Präzision wesentlich durch die Qualität und Auslegung der Peripherieelemente beeinflusst wird. Funktionale Randbeschaltungen für den Prozessor sind nicht zu erwarten. Als Taktgeber kann eine interne Uhr mit geringer Präzision und Taktrate (wenige kHz) verwendet werden, da die Echtzeitanprüche der Anwendung niedrig sind und durch einen Fail-Safe-Modus abgefangen werden.

Aufgrund der Einfachheit der Anwendung sind auf den *programmierrelevanten Ebenen (Layer 5 bis 8)* keine besonderen Herausforderungen zu erwarten: die AD-Wandlerwerte können als ganzzahlige *counts* direkt über Integeroperationen auf einen Schaltwert umgerechnet werden. Der Schaltwert für dann zu einer binären Ausgabe des Schaltzustandes auf dem drahtlosen Übertragungskanal.

Als *Programmierstil (Layer 7)* sollte eine interruptbasierte Programmierung vorgesehen werden, um das System zwischen den Messungen in den tiefstmöglichen Sleepmodus zu versetzen.

3.2 Kommunikationssystem

Für die drahtlose Kommunikation wird hier BLE vorgegeben. Mit dem in dieser Applikation separaten Sendesystem wird auf Layer 3 über ein einfaches Protokoll mit langsamer, overhead-ärmer Kommunikation gesendet. Um eine Manipulation der Daten und somit ein fälschliches Ausschalten des Lichtes zu verhindern, sollte kein Broadcast-, sondern ein Peer-to-Peer-Protokoll mit Verschlüsselung eingesetzt werden. Hierdurch entstehen zwar negative Implikationen auf die Energiebilanz erforderliche Sicherheit kann aber hergestellt werden [Tow+14, Kapitel 3]. Oberhalb der Verschlüsselungsebene wird ein sehr einfaches Protokoll verwendet, in dem entweder nur der binäre Schaltzustand oder aber der Messwert selbst kommuniziert wird. Hierzu bietet BLE neben der Verwendung spezialisierter Profile auch die Möglichkeit der Definition eigener Profile, worauf in diesem Fall zurückgegriffen wird [Cyp16b, Kapitel 2.3 und 4.6], [Cyp16a, Kapitel 4].

4 Zusammenfassung

In diesem Beitrag wurde eine generische Architektur für den Entwurf drahtloser Sensorsysteme vorgestellt und in der Funktion der einzelnen Schichten analysiert.

Exemplarisch an dem Entwurfsprozess eines Helligkeitssensors wurde die Anwendung der Architektur vorgestellt. In weiteren Forschungsschritten wird diese Architektur inhaltlich weiter ausgearbeitet und mit Designrichtlinien sowie Entwurfsmustern untermauert.

Literaturverzeichnis

- [Cyp16a] Cypress. Application Note 91162: Creating a BLE Custom Profile. 2. Nov. 2016.
- [Cyp16b] Cypress. Application Note 92584: Designing for Low Power and Estimating Battery Life for BLE Applications. 14. März 2016.
- [DP10] Walteneus Dargie und Christian Poellabauer. Fundamentals of wireless sensor networks: theory and practice. Wiley series on wireless communications and mobile computing. Chichester, West Sussex, U.K. ; Hoboken, NJ: Wiley, 2010. 311 S. Metaarchitektur für IoT und Sensornetzwerke 9
- [Fel+13] S. Feldmann u. a. “Model-Driven Engineering and Semantic Technologies for the Design of Cyber-Physical Systems”. In: IFAC Proceedings Volumes 46.7 (Mai 2013), S. 210–215.
- [ISO94] ISO/IEC. ISO/IEC 7498-1:1994(E): Information technology - Open Systems Interconnection - Basic Reference Model: The Basic Model. Genf/CH: ISO/IEC, 15. Nov. 1994.
- [KKP99] Joseph M. Kahn, Randy H. Katz und Kristofer SJ Pister. “Next century challenges: mobile networking for Smart Dust”. In: Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking. ACM, 1999, S. 271–278.
- [MF10] Friedemann Mattern und Christian Floerkemeier. “From the Internet of Computers to the Internet of Things”. In: From active data management to eventbased systems and more. Springer, 2010, S. 242–259.
- [Pue15] Fernando Puente Leon. Messtechnik. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015.
- [SSS13] A. Smirnov, K. Sandkuhl und N. Shilov. “Multilevel Self-Organisation and Context-Based Knowledge Fusion for Business Model Adaptability in Cyber-Physical Systems”. In: IFAC Proceedings Volumes 46.9 (2013), S. 2045–2050.
- [Tow+14] Kevin Townsend u. a. Getting started with Bluetooth low energy: tools and techniques for low-power networking. Revised First Edition. OCLC: ocn870896083. Sebastopol, CA: O’Reilly, 2014. 164 S.

Unsere Empfehlung für Sie: Präferenzen und Personalisierung in der Informatik

Markus Endres¹ und Andreas Pfandler²

Abstract: Präferenzen spielen im Alltag jedes Menschen eine große Rolle. Einfache Präferenzen wie z.B. „Zum Frühstück mag ich lieber Kaffee als Tee, Schokocroissants sind besser als ein Käsebrof“ oder „Eine Unterkunft in einem 4-Sterne Hotel in der Nähe der Konferenz ist mir wichtiger als eine möglichst günstige Übernachtung“ sind uns allen bekannt. Häufig müssen wir uns zwischen unzähligen verschiedenen Alternativen entscheiden. Systeme, die unsere Präferenzen, also die Benutzerwünsche, einbeziehen, helfen die bestmöglichen Entscheidungen zu treffen. Das Tutorium behandelt verschiedene Aspekte der Modellierung und Verarbeitung von Präferenzen in den Bereichen Datenbanken & Informationssysteme sowie Künstliche Intelligenz. Die Vorstellung von konkreten Einsatzgebieten und Anwendungen von Präferenzen sowie die Darstellung von Risiken, Chancen und Herausforderungen im Umgang mit Personalisierung runden das Tutorium ab.

1 Präferenzen in der Informatik

In der Informatik gewinnen Präferenzen immer mehr an Bedeutung. So haben Präferenzen seit langem Einzug bei Touristikportalen oder Firmen wie Amazon, Google und Facebook gehalten, um personalisierte Produktempfehlungen vorzunehmen. Dabei zeigt sich, dass Kunden sehr viel mehr Wert als früher auf personalisierten Kundenservice anstelle von Standardprodukten legen. Eine Zusammenstellung einer Reise zum Beispiel, einschließlich Flug, Mietwagen, Hotelreservierung, Veranstaltungen und Exkursionen kann nicht mehr als Pauschalangebot verkauft werden, sondern erfordert immer mehr die individuelle Anpassung an die Wünsche der Kunden. Die Bedürfnisse der Benutzer spielen heutzutage insbesondere im Marketing eine bedeutende Rolle. Marketing-Lösungen, die komplett auf den Benutzer zugeschnitten sind, setzen sich durch. Das Stichwort lautet *Personalisierung*.

Aus psychologischer Sicht bevorzugt der Mensch personalisierte Informationen, da er sich damit wertgeschätzt fühlt und Sympathie für sein „Gegenüber“ empfindet. Durch personalisierte Informationen und die auf uns maßgeschneiderten Botschaften entwickelt man unbewusst auch ein Gefühl der Sicherheit. Zudem heben sich personalisierte Nachrichten von dem heutzutage herrschenden Überangebot von Informationen ab. Auch wenn dieser Prozess unbewusst stattfindet, ist er enorm einflussreich und überzeugt ganz unterschwellig. So ist z.B. eine personalisierte E-Mail, in der man mit dem Vornamen angesprochen wird und die einen an die eigenen Interessen angepassten Inhalt hat, zugänglicher. Insbesondere nimmt man sich z.B. mehr Zeit Werbeinformationen zu sichten, wenn der Inhalt den eigenen Interessen entspricht. Dieser Effekt kehrt sich jedoch um, wenn bei der Personalisierung

¹ Universität Augsburg, Universitätsstr. 6a, D 86159 Augsburg, endres@informatik.uni-augsburg.de

² Technische Universität Wien, Favoritenstr. 9, A 1040 Wien und Universität Siegen, Unteres Schloß 3, D 57072 Siegen, pfandler@dbai.tuwien.ac.at

Fehler unterlaufen: Etwa wenn ein Zelt auf einem Campingplatz einem Geschäftskunden als Tagungsraum empfohlen wird oder wenn sich der Benutzer „bespitzt“ fühlt. Es wird klar, dass es sich bei der Personalisierung und der Verarbeitung von Präferenzen um komplexe Aufgaben handelt, die offensichtlich Fingerspitzengefühl erfordern. Ein besonders positiver Eindruck entsteht wiederum, wenn präferierte Alternativen angeboten werden, die der Benutzer nicht bedacht hat, wie etwa günstigere Flugverbindungen zu einem nahe gelegenen anderen Flughafen. Personalisierte Inhalte können somit einen positiven Effekt auf die Marketing-Effizienz haben, da nicht relevante Inhalte aufgrund der vorliegenden Informationsflut ohnehin oft nicht beachtet werden.

Wie oben bereits angedeutet, bedeutet Personalisierung aber nicht nur persönliche Daten in Dokumenten zu nutzen, sondern auch die Präferenzen der Benutzer bei Produktempfehlungen oder bei der Produktsuche zu berücksichtigen. Während bei der erstgenannten Art der Personalisierung meist nur die nüchternen Daten verwendet werden, benötigt die zweitgenannte Art Wissen über die Präferenzen und Methoden, um die richtigen Schlüsse daraus zu ziehen. Damit erhalten Benutzer z.B. bei der Produktsuche auf sie zugeschnittene Antworten und können unter verschiedenen bestmöglichen Alternativen auswählen. Letztendlich profitieren also nicht nur Unternehmen von personalisierten Inhalten, sondern auch die Kunden selbst.

Eine weitere wichtige Bedeutung haben Präferenzen bei der gemeinsamen Entscheidungsfindung etwa innerhalb einer Gruppe oder eines Freundeskreises. Hierbei ist es das Ziel eine Lösung zu finden, welche die Bedürfnisse bzw. Präferenzen der einzelnen beteiligten Personen berücksichtigt und dabei gleichzeitig möglichst fair ist. Ein einfaches Beispiel ist etwa das Finden eines Lokals für ein gemeinsames Mittagessen. Hier geben die Teilnehmer nicht nur an, wann sie Zeit haben, sondern teilen auch ihre Präferenzen bezüglich des Lokals mit. Gesucht ist dann das „passendste“ Lokal. Eine weitere wichtige Frage ist, wie man (komplexe) bedingte Präferenzen behandelt, bei denen die einzelnen Entscheidungsmöglichkeiten nicht unabhängig voneinander sind. So könnte zum Beispiel bei der Frühstückspannung Tee nur dann Kaffee vorgezogen werden, wenn es ein Käsebrot zum Frühstück gibt, ansonsten ist Kaffee die erste Wahl.

Dem Motto der INFORMATIK 2016 „Informatik von Menschen für Menschen“ folgend, werden im Rahmen dieses Tutoriums Methoden vorgestellt, um Informationssysteme der Zukunft zu schaffen, die Präferenzverarbeitung als ein zentrales Konzept der Personalisierung ansehen. Im Folgenden werden wir einen knappen Überblick über die Grundlagen, aktuelle Forschung und praktische Anwendungen von Präferenzen in Datenbankanfragen sowie im Bereich der Künstlichen Intelligenz geben.

2 Präferenzen in Datenbankanfragen

Im ersten Teil des Tutoriums werden wir Präferenzen innerhalb von Datenbanksystemen behandeln, da immer mehr Datenbankhersteller die Verarbeitung von Präferenzen als ein zentrales Konzept der Personalisierung ansehen. So existieren neben den rein akademischen Prototypen wie z.B. FlexPref [LMK10] oder Preference SQL [KEW11] auch kommerzielle

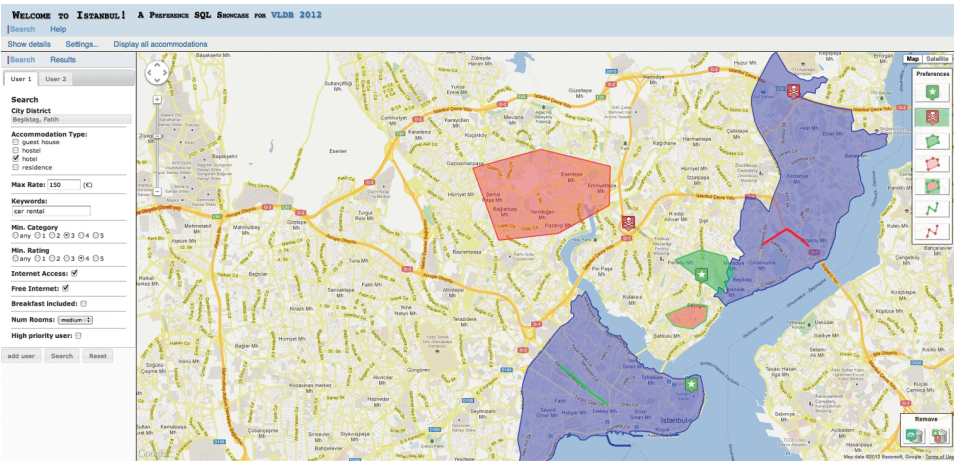


Abbildung 1: Screenshot einer präferenzbasierten Hotel-Suche in Istanbul [We12].

Systeme wie etwa das Skyline Feature in EXASolution [Ma15]. Dass es sich hierbei nicht nur um eine Spielwiese für Wissenschaftler handelt, zeigt das Szenario der Suche nach einem Hotel in einer Stadt, z.B. in Istanbul. Aufgrund der enormen Anzahl von Hotels aller Kategorien in dieser Stadt stellt die Kenntnis der Benutzerwünsche in dieser Anfrage eine große Hilfe dar, um irrelevante Hotels aus dem Datensatz auszublenden und den Fokus ausschließlich auf die relevanten Hotels zu legen. Präferenzen agieren als Filter, welche nur die relevanten Antworten bezüglich der Benutzerpräferenzen liefern.

Abbildung 1 demonstriert solch eine personalisierte ortsbasierte Hotel-Suche³, die es einzelnen Benutzern oder Benutzergruppen erlaubt, Hotels in Istanbul zu finden, wobei sowohl harte Kriterien als auch Benutzerwünsche berücksichtigt werden (vgl. [We12]). So lässt sich nicht nur die Hotel-Kategorie als hartes oder weiches Kriterium spezifizieren, sondern auch die Präferenz bezüglich der Art der Unterkunft oder der Verfügbarkeit von WLAN. Die präferierten Stadtteile können nun etwa als weiches Kriterium formuliert werden, welches (sofern möglich) erfüllt sein sollte. Die Applikation unterstützt sowohl räumliche, numerische und kategorielle als auch komplexe Präferenzanfragen auf eine intuitive Art und Weise, sodass das Ergebnis bestmöglich den Benutzerwünschen entspricht.

Die Verwendung von Präferenzen in Datenbankanfragen wirft allerdings viele komplexe Fragestellungen auf, etwa bzgl. der Darstellung/Modellierung, Auswertung, Sprache, etc., vgl. [SKP11]. Bereits im Jahre 1987 haben Lacroix und Lavency [LL87] sich diese Fragen gestellt und noch heute beschäftigen sich Wissenschaftler auf der ganzen Welt mit diesem Thema.

Kießling und andere haben mit „Preference SQL“ [KEW11] eine Methode zur Repräsentation von Präferenzen in SQL entwickelt. Die Grundlage bilden Striktordnungen (irreflexiv, transitiv, antisymmetrisch), z.B. „Eine Unterkunft im Stadtteil Fatih in Istanbul ist mir wichtiger als ein 4-Sterne Hotel“, welche durch Präferenz-Konstrukturen gebildet werden

³ Web-Applikation zu finden unter <http://www.dbis.informatik.uni-augsburg.de/hotelfinder/>

können. Um Präferenzen in Preference SQL spezifizieren zu können, wurde die Syntax des SQL-Select-Statements um die `PREFERRING`-Klausel erweitert. Preference SQL folgt der „Best Matches Only“-Semantik: Das Ergebnis beinhaltet ausschließlich die besten Treffer bezüglich der gegebenen Präferenzen. Abbildung 2 zeigt eine einfache Preference SQL Anfrage, bei der die ortsbasierte Präferenz durch eine Priorisierung (`PRIOR TO`) mit dem Wunsch nach einem 4-Sterne Hotel verknüpft wird. Dabei wird der Stadtteil Fatih durch einen KML-String⁴ festgelegt, welcher erfüllt sein sollte.

```
SELECT *
FROM accommodation                               -- Unterkünfte in Istanbul
PREFERRING                                       -- Präferenz-Query
  location WITHIN ' <KML-String>'              -- Ortsbasierte Suche
PRIOR TO                                         -- Priorisierung
  category = '4*' AND type = 'Hotel';          -- Pareto-Präferenz
```

Abbildung 2: Preference SQL Anfrage: Eine Unterkunft im Stadtteil Fatih (KML-String) ist wichtiger als ein 4-Sterne Hotel.

Die Firma EXASOL AG⁵ setzt diesen Ansatz in ihrem kommerziellen Datenbanksystem EXASolution um [Ma15]. Außerdem existiert ein präferenzbasiertes Empfehlungssystem auf Grundlage von Preference SQL [SEK06].

Die in Preference SQL vorhandene „Pareto-Präferenz“, bei der verschiedene Attribute als gleich wichtig erachtet werden, ist auch als „Skyline Operator“ [BKS01] bekannt. Chomicki stellt Präferenzen durch logische Ausdrücke dar. Auch in diesem Ansatz werden Präferenzen als Striktordnungen modelliert und zur personalisierten Ergebnisfilterung benutzt. Chomicki’s „Winnow-Operator“ folgt der gleichen Semantik wie „Best Matches Only“ und „Skyline“. Für diese Operatoren existierten zahlreiche Auswertungsalgorithmen, die für (parallele und verteilte) Datenbanksysteme optimiert wurden [CCM13]. Boutilier, Brafman und Domshlak [Bo03] repräsentieren Ordnungen als CP-Netze. Hierbei wird die „Ceteris Paribus“-Semantik verwendet, mit welcher bedingte Präferenzen spezifiziert werden können. So könnte zum Frühstück Tee nur dann Kaffee vorgezogen werden, wenn es ein Käsebrot zum Frühstück gibt, ansonsten ist Kaffee die erste Wahl.

Neben den bereits vorgestellten qualitativen Ansätzen der Präferenzrepräsentation existieren auch quantitative Präferenzen in Datenbankanfragen, vgl. z.B. [HKP01]. In PREFER [HKP01] werden quantitative Präferenzen durch Gewichtung der Attribute festgelegt. Anschließend wird eine gewichtete Aggregatsfunktion verwendet, um einen Score für jedes Tupel zu berechnen. Je höher der Score ist, umso besser ist ein Tupel.

Die Verarbeitung und Verwaltung von Präferenzanfragen in Datenbanken ist im letzten Jahrzehnt zu einer ernstzunehmenden und wichtigen Herausforderung geworden. Im ersten Teil des Tutoriums werden wir deshalb verschiedene Ansätze der Präferenzmodellierung in Datenbanksystemen sowie ihre Vor- und Nachteile im Detail besprechen.

⁴ Keyhole Markup Language (KML) ist eine Auszeichnungssprache zur Beschreibung von Geodaten.

⁵ <http://www.exasol.com/>

3 Präferenzen in der Künstlichen Intelligenz

Im zweiten Teil des Tutoriums werden aktuelle Forschungsthemen (sowie die benötigten Grundlagen) aus Teilbereichen der Künstlichen Intelligenz – insbesondere aus dem Gebiet der Computational Social Choice – behandelt.

Eine einfache Möglichkeit, um Präferenzen über einer Menge von Alternativen bzw. Kandidaten anzugeben, ist mittels einer totalen Ordnung. Die Präferenzen dreier Kollegen bezüglich des Lokals für ein gemeinsames Mittagessen könnten wie folgt lauten:

Anna:	Schnitzel	\succ	Pizza	\succ	Sushi
Berta:	Schnitzel	\succ	Sushi	\succ	Pizza
Christoph:	Sushi	\succ	Pizza	\succ	Schnitzel

Damit wird ausgedrückt, dass Anna ein Lokal, in dem es Schnitzel gibt, am liebsten hat. Die nächstbeste Alternative ist eine Pizzeria, und ein Sushilokal ist nicht gerade ihre favorisierte Option. Die Menge der Kandidaten ist in diesem Fall $C = \{\text{Schnitzel}, \text{Pizza}, \text{Sushi}\}$. Eine wichtige Fragestellung ist nun: Gegeben die obigen Präferenzen, für welches Lokal sollen sich die Kollegen entscheiden, wenn sie die Präferenzen aller Teilnehmer berücksichtigen wollen? Oder anders formuliert: Welche Kandidaten aus C sind die Gewinner der obigen Abstimmung?

Um zu einem Ergebnis zu kommen, fehlt allerdings noch eine *Wahlregel*, mit deren Hilfe eine nichtleere Teilmenge der Kandidaten als Gewinner bestimmt werden kann. Eine sehr einfache und verbreitete Wahlregel ist Pluralität. Bei dieser Wahlregel erhält der erstgereichte Kandidat jeder Stimme einen Punkt. Die Kandidaten mit den meisten Punkten sind die Gewinner. Ein offensichtlicher Nachteil dieser Wahlregel ist, dass sehr viel an Information über die Präferenzen der Wähler ignoriert wird: Die Reihung der Kandidaten an zweiter, dritter, usw. Stelle ist für das Ergebnis irrelevant! Eine andere Wahlregel, die Wahlregel von Borda [Bo84], vergibt bei m Kandidaten pro Stimme $m - 1$ Punkte an den erstgereichten Kandidaten, $m - 2$ Punkte an den zweitgereichten usw. Diese Regel berücksichtigt auch die Präferenzen bezüglich der Kandidaten auf den hinteren Positionen.

Werden die Präferenzen aus dem obigen Beispiel unter der Pluralitätsregel ausgewertet, erhält „Schnitzel“ zwei Punkte, „Sushi“ einen Punkt, und „Pizza“ geht leer aus. Unter der Regel von Borda ergeben sich hingegen vier Punkte für „Schnitzel“, drei Punkte für „Sushi“ und zwei Punkte für „Pizza“. Bei diesem Beispiel gewinnt also in beiden Fällen das Schnitzellokal. Im Allgemeinen können sich die Gewinner aber abhängig von der verwendeten Wahlregel sehr wohl unterscheiden. Wenn man im obigen Beispiel bei Berta und Christoph jeweils „Sushi“ mit „Pizza“ vertauscht, so ändert dies natürlich nichts am Gewinner bezüglich der Pluralitätsregel. Bei Borda führt dies jedoch zu einem Gleichstand von „Schnitzel“ und „Pizza“ mit jeweils vier Punkten – es gibt also zwei Gewinner. Dieses Beispiel soll veranschaulichen, dass selbst in einfachen Präferenzmodellen die Fragestellungen recht vielfältig und komplex werden können. Ähnliche Aufgabenstellungen wie bei der Lokalsuche fürs Mittagessen können natürlich auch in größeren Maßstäben auftreten: Etwa, wenn auf einem Reiseportal eine Gruppenreise basierend auf den Präferenzen der Teilnehmer geplant werden soll.

Im Tutorium werden wir verschiedene Wahlregeln betrachten, die mitunter sehr unterschiedliche Eigenschaften aufweisen. Es wird klar, dass bereits die Auswahl der Wahlregel eine wichtige Entscheidung ist, die das Ergebnis und in der Folge auch die Zufriedenheit der Benutzer stark beeinflussen kann. Ein Beispiel für eine wünschenswerte Eigenschaft einer Wahlregel ist, dass es niemals für einen Wähler von Nachteil sein sollte, wenn er seine Stimme abgibt. Mit Nachteil ist hier gemeint, dass sich das Ergebnis durch die Teilnahme an der Wahl verschlechtert (gemessen an den Präferenzen des jeweiligen Wählers). Auch wenn man vielleicht intuitiv annehmen würde, dass es immer besser ist seine Stimme abzugeben, werden wir sehen, dass ein solches No-show Paradox tatsächlich auftreten kann. Für eine detaillierte Diskussion der behandelten Wahlregeln, deren Eigenschaften und weiterführende Literatur verweisen wir auf [Ro15, Kapitel 4].

Ebenso wünschenswert wäre es, dass Wähler niemals davon profitieren können, wenn sie bei der Angabe ihrer Präferenzen unehrlich sind. Obgleich wünschenswert, wurde von Gibbard und Satterthwaite gezeigt, dass strategisches Wählen (Manipulation) in „realistischen“ Szenarien nicht ausgeschlossen werden kann. Weitere unerwünschte Beeinflussungen sind Bestechung der Wähler (Bribery) und Wahlkontrolle (Control). Bei der Wahlkontrolle versucht ein Vorsitzender durch Strukturänderungen an der Wahl selbst, wie z.B. durch Sperren oder Hinzufügen von Wählern bzw. Kandidaten, das Ergebnis der Wahl (zu seinen Gunsten) zu verändern. Es zeigt sich, dass diese Aspekte unbedingt beim Design eines präferenzbasierten Systems Beachtung finden müssen, da böswillige Benutzer sich derartige Schwachstellen sonst zunutze machen können.

Bei den sogenannten Domain-Restrictions, mit denen wir uns ebenfalls kurz auseinandersetzen werden, ist die grundlegende Annahme, dass die Präferenzen der Benutzer in der Praxis nicht zufällig sind, sondern eine gewisse Struktur aufweisen. Erfüllt eine Instanz eine bestimmte Domain-Restriction, so können nun nicht mehr alle denkbaren Präferenzen auftreten, sondern nur noch jene, die die Kriterien der Domain-Restriction erfüllen. Dies ist besonders spannend, wenn die Struktur der Instanz zu besseren Eigenschaften der Wahlregeln führt oder von effizienten Algorithmen genutzt werden kann. Ein prominentes Beispiel für eine solche Domain-Restriction ist Single-Peakedness [BI48].

Weitere Informationen zu den bisher angesprochenen Themen (und vielen weiteren) sind in zwei aktuellen Büchern zu finden [Ro15, Br16].

In einem weiteren Themenblock werden wir uns mit Präferenzen im Gebiet von Knowledge Representation & Reasoning beschäftigen – genauer gesagt mit Präferenzen im Zusammenspiel mit Answer Set Programming (ASP). Answer Set Programming (für eine Einführung vgl. [EIK09, Ge12]) ist ein erfolgreicher Ansatz der logischen Programmierung, der sich durch eine gut handhabbare, deklarative Semantik und eine Vielzahl an effizienten Solvoren auszeichnet (etwa die Solver der Potassco Familie, sowie DLV und WASP). Ein weiterer Vorteil ist, dass ASP Programme als eine „ausführbare Spezifikation“ gesehen werden können, wodurch man schnell leicht anpassbare und trotzdem brauchbare Prototypen entwickeln kann.

In der Folge wird das Zusammenspiel von ASP und Präferenzen aus zwei Perspektiven beleuchtet. Einerseits werden wir Methoden behandeln, die es erlauben, Präferenzen bezüglich

The screenshot displays the Democratix web interface. At the top, there are navigation links for 'Web interface', 'Examples', 'Downloads', and 'Contact information'. The main content area is divided into three sections:

- Profile:** A text area containing a list of items and their scores:


```
3
1, Siapsons
2, Big Bang Theory
3, IT Crowd
28, 28, 5
11, 2, 3, 1
7, 3, 2, 1
5, 1, 2, 3
3, 3, 1, 2
2, 1, 3, 2
```
- Example instance:** A text area containing the same list of items and scores as the profile, with a note: "For testing purposes, you can copy/paste the following example profile to the profile input area:" and a "[copy to profile]" button.
- Output:** A text area showing the solver's output for the rule 'borda':


```
Evaluating using rule 'borda'
-- STARTING SOLVER (clingos) --
Answer: 1
score(1.17) score(3.33) score(2.34) maxScore(34) winner(2)
SATISFIABLE
Models : 1
Time : 0.000
Prepare : 0.000
Prepro. : 0.000
Solving : 0.000
-- SOLVER TERMINATED (exit code 10) --
Aaaaaaand the winner is: 'Big Bang Theory'
```

On the left side of the interface, there are three speech bubbles with the inequalities $A > B > C$, $A > C > B$, and $B > A > C$. Below them is a 'DEMOCRATIX' logo and a bar chart with three bars labeled B, A, and C, where A is the tallest.

Abbildung 3: Screenshot der Weboberfläche von Democratix.

der Lösungen eines ASP Programms auszudrücken. Das asprin Framework [Br15] wird hier als Beispiel dienen. Andererseits werden wir das Tool Democratix [CP15] vorstellen, mit dem in ASP deklarativ spezifizierte Wahlregeln ausgewertet werden können. Abbildung 3 zeigt einen Screenshot der Democratix Web-Applikation⁶, welche auch im Rahmen des Tutoriums vorgeführt wird.

4 Zusammenfassung

Präferenzen und Personalisierung sind nicht mehr nur von akademischem Interesse, sondern haben weitgehende Auswirkungen auf Industrie und Wirtschaft. Angesichts der stark wachsenden Fülle an Informationen genügt es heute nicht mehr, eine Menge gespeicherter Daten möglichst schnell bei den Nutzern „abzuliefern“. Vielmehr müssen die richtigen Daten in der gegebenen Situation zur rechten Zeit im adäquaten Zusammenhang unter dem relevanten Blickwinkel und in der gewünschten Form zur Verfügung stehen. Die Notwendigkeit und das Potential des präferenzbasierten und personalisierten Datenmanagements wird dadurch deutlich, dass die Ergebnisse auch unmittelbar Anwendung in der Industrie finden. Die Entwicklung passender Methoden zum Erlernen und Verarbeiten von Präferenzinformationen bleibt – insbesondere aufgrund der riesigen Datenmengen – eine Herausforderung für Wissenschaft, Wirtschaft und Gesellschaft.

Literaturverzeichnis

[BKS01] Börzsönyi, S.; Kossmann, D.; Stocker, K.: The Skyline Operator. In: Proceedings of ICDE '01. IEEE, S. 421–430, 2001.

⁶ Webseite von Democratix: <http://democratix.dbai.tuwien.ac.at/>

- [Bl48] Black, D.: On the rationale of group decision making. *Journal of Political Economy*, 56(1):23–34, 1948.
- [Bo84] Borda, J.: *Mémoire sur les élections au scrutin*. Histoire de l'Académie Royale des Sciences, Paris, 1784.
- [Bo03] Boutilier, C.; Brafman, R.; Domshlak, C.; Hoos, H.; Poole, D.: CP-Nets: A Tool for Representing and Reasoning with Conditional Ceteris Paribus Preference Statements. In: *Journal of Artificial Intelligence Research*. Jgg. 21, S. 135–191, 2003.
- [Br15] Brewka, G.; Delgrande, J. P.; Romero, J.; Schaub, T.: asprin: Customizing Answer Set Preferences without a Headache. In: *Proceedings of AAAI 2015*. AAAI Press, S. 1467–1474, 2015.
- [Br16] Brandt, F.; Conitzer, V.; Endriss, U.; Lang, J.; Procaccia, A. D., Hrsg. *Handbook of Computational Social Choice*. Cambridge University Press, 2016.
- [CCM13] Chomicki, J.; Ciaccia, P.; Meneghetti, N.: Skyline Queries, Front and Back. *ACM SIGMOD Record*, 42(3):6–18, 2013.
- [CP15] Charwat, G.; Pfandler, A.: Democratix: A Declarative Approach to Winner Determination. In: *Proceedings of ADT 2015*. Jgg. 9346 in LNCS. Springer, S. 253–269, 2015.
- [EIK09] Eiter, T.; Ianni, G.; Krennwallner, T.: Answer Set Programming: A Primer. In: *Reasoning Web. Semantic Technologies for Information Systems, 5th International Summer School 2009, Tutorial Lectures*. Jgg. 5689 in LNCS. Springer, S. 40–110, 2009.
- [Ge12] Gebser, M.; Kaminski, R.; Kaufmann, B.; Schaub, T.: *Answer Set Solving in Practice. Synthesis Lectures on Artificial Intelligence and Machine Learning*. Morgan & Claypool Publishers, 2012.
- [HKP01] Hristidis, V.; Koudas, N.; Papakonstantinou, Y.: PREFER: A System for the Efficient Execution of Multi-parametric Ranked Queries. In: *Proceedings of SIGMOD '01*. ACM, S. 259–270, 2001.
- [KEW11] Kießling, W.; Endres, M.; Wenzel, F.: The Preference SQL System - An Overview. *Bulletin of the Technical Committee on Data Engineering*, 34(2):11–18, 2011.
- [LL87] Lacroix, M.; Lavency, P.: Preferences: Putting More Knowledge into Queries. In: *Proceedings of VLDB '87*. Morgan Kaufmann Publishers Inc., S. 217–225, 1987.
- [LMK10] Levandoski, J. J.; Mokbel, M. F.; Khalefa, M. E.: FlexPref: A Framework for Extensible Preference Evaluation in Database Systems. In: *Proceedings of ICDE '10*. IEEE, S. 828–839, 2010.
- [Ma15] Mandl, S.; Kozachuk, O.; Endres, M.; Kießling, W.: Preference Analytics in EXASolution. In: *Proceedings of BTW '15*. GI, 2015.
- [Ro15] Rothe, J., Hrsg. *Economics and Computation*. Springer, 2015.
- [SEK06] Satzger, B.; Endres, M.; Kießling, W.: A Preference-Based Recommender System. In: *Proceedings of EC-Web '06*. Jgg. 4082 in LNCS. Springer, S. 31 – 40, 2006.
- [SKP11] Stefanidis, K.; Koutrika, G.; Pitoura, E.: A Survey on Representation, Composition and Application of Preferences in Database Systems. *ACM Transactions on Database Systems*, 36(4), 2011.
- [We12] Wenzel, F.; Endres, M.; Mandl, S.; Kießling, W.: Complex Preference Queries Supporting Spatial Applications for User Groups. In: *Proceedings of VLDB '12*. Jgg. 5. VLDB Endowment, S. 1946–1949, 2012.

D2d — Kreatives Schreiben von XML-codierten Texten

Markus Lepper¹ Baltasar Trancón y Widemann²

1 Anwendungskontext und Motivation

Der XML Standard [Br06] hat sich mittlerweile zur Codierung von formalen oder semi-formalen Texten weithin durchgesetzt. Die Vorteile dieser Standardisierung sind die Wiederverwendbarkeit einer unübersehbaren Fülle von Transformationswerkzeugen und die Unabhängigkeit von proprietären Standards bei Erstellung, Speicherung und Zugriff.

„Mark-Up“ in Texte einzuführen kann in sehr unterschiedlichen Kontexten sinnvoll sein, die weit über das bis dato übliche hinausgehen. Dies aber auf eine jeweils sehr unterschiedliche Art und Weise. So kann es sinnvoll sein, Referenzen auf Personen in mehrbändigen Fantasy-Novellen gleich beim Erstellen, oder in einem ersten redaktionellen Durchgang zu markieren. Ähnlich kann die Markierung von Ortsangaben in Wanderführern, von chemischen Formeln in pharmakologischen Texten, Tempoangaben in Sportreportagen, Tageszeiten in Laborberichten, Zutatenlisten in Kochbüchern etc. den Wert des Elaborates allein durch die mögliche automatische Nach- und Weiterverarbeitung deutlich erhöhen, bis hin zur Präsentation als „dynamisches Dokument“.

Allerdings ist das übliche Serialisierungs-Format von XML keinesfalls geeignet, hingeschrieben zu werden von Autoren im kreativen Fluss des Schreibens. Selbst, ja gerade „syntaxgesteuerte Editoren“ sind nicht in der Lage, Menschen, für die der Akt des Schreibens ein direkter, kontinuierlicher, ja, intimer ist, für Mark-Up zu begeistern.

2 Designprinzipien von D2d

Diese Lücke will D2d schließen. Die Abkürzung steht für „Direct Document Denotation“ oder auch „Directly To Document“ und wird meist „triple-dee“ gesprochen. D2d ermöglicht es einem Autor, mit nur einem ausgezeichneten Befehls-Zeichen im gewohnten Verlauf des Schreibens lesbare aber korrekte, streng typisierte XML-konforme Dokumente zu erzeugen. Insbesondere abstrahiert D2d von dem redundanten Gebrauch von Sonderzeichen und der idiosynkratischen Unterscheidung von „Element“ und „Attribute“ und erlaubt den Permutationsoperator mit beiden.

Der gewohnte Texteditor, die gewohnte Arbeitsumgebung, die gewohnte Methodik kann beibehalten werden. D2d kann mit Kreide und Wandtafel, Papier und Bleistift einfach

¹ semantics GmbH, Berlin, post@markuslepper.eu

² Technische Universität Ilmenau, baltasar.trancon@tu-ilmenau.de

notiert werden und ist geeignet für Sprachein- und Ausgabe, und schon durch die einfache Zeichenstruktur behindertenfreundlich.

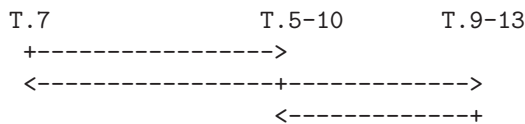
Dass je Arbeitsfeld unterschiedliche Definitionen von Text-Elementen, deren Tags und deren Inhalte getroffen werden müssen, und diese Elemente nicht unmittelbar eine optische Erscheinung ihres Inhaltes bestimmen, sondern vielmehr dessen inhaltliche Rolle im Körper des Gesamttextes, hat dazu geführt, dies „semantisches Mark-Up“ zu nennen, im Gegensatz zum „physikalischen“. Das kann in die Irre führen, denn allemal ist es lediglich *Syntax*, was von D2d erkannt und verwaltet wird. Richtig jedoch ist die wichtige Unterscheidung zwischen dem erstellten Text als seiner „Substanz“ und allen möglichen Verarbeitungen und Darstellungen als seiner „Erscheinung“, – ein im Bereich des Digitalen noch weithin unerforschte Dichotomie.

Die Definition von D2d und die vorliegende Implementierung beinhalten nur einfache bekannte Techniken und kleine, unspektakuläre Erfindungen, – das Zusammenwirken allerdings stellte sich als sehr komplex heraus, so dass fast zehn Jahre Entwicklung und Erprobung in das System geflossen sind. [LTW01, LT11]

Grundlegende Design-Eigenschaften sind, (a) die Verwendung offener Standards, wann immer vorhanden. Ausserdem (b) eine grundlegende Orientierung am Anwender, der oft Fach-Experte und kein Informations-Experte ist. Dies z.B. durch „Erste Klasse“-Unterstützung von polyglottem Dokumentationstext, durch graphische „Syntax-Diagramme“ im WIRTH-Stil, durch extensive Fehlerdiagnose und durch einen (ein- und ausschaltbaren) robusten Übersetzungsmodus, der Vorschläge für fehlende Strukturen ergänzt und überflüssige zu deutlich sichtbarem Kommentar neutralisiert.

3 Beispiel

Ein repräsentatives Beispiel ist eine große musiktheoretische Analysearbeit, die als ein *dynamisches* Dokument vorliegt, und nur als ein solches verlustlos vorliegen kann [Le15]. Darin werden u.a. Takt- und Tonart-Angaben mit Mark-Up versehen. In der Nachbereitung („Rendering“) des eingegebenen Textes werden z.B. Navigationselemente erzeugt, die auf das erste, letzte, vorige und folgende Auftreten derselben Tonart verweisen. Im Falle von Taktangaben werden hingegen auf diese Weise alle Bereiche verbunden, die eine nicht-leere Schnittmenge haben:



Diese Art der Navigation ist die *dieser konkreten Lese-Situation adäquate*, und mit vorgefertigten, festgelegten Standard-Mitteln nicht zu erreichen.

Ein ganz anderes Beispiel ist die Dokumentation von D2d selber, als Teil der Sammlung *metatools*, siehe [met]. Die Quelltexte aller dieser Seiten sind erreichbar durch das al-

lerletzte auf ihnen enthaltene Link. Man beachte besonders die Notenbeispiele, deren verschieden kombinierte Schichten auch durch XML-Mark-Up unterschieden und dann durch XSLT zwecks Rendering rekombiniert werden.

Wir vermuten, dass nur durch einen derart spezialisierten, dezidierten und inhaltlichen Gebrauch des Rechners Fortschritte auf dem Gebiet der „Digital Humanities“ erst wirklich möglich werden.

4 Parsieren von Tags und Character Data

D2d unterscheidet zwei Ebenen von Parsierung:

Auf der groberen Ebene der „Tag-Parsierung“ werden die eröffnenden Tags, eingeleitet durch das eine, erwähnte Kommandozeichen, in den Text eingestreut. Schließende Tags werden durch einfachste LL(1)-Parsierung inferiert, können aber auch explizit gesetzt werden.

Auf der feineren Ebene der „Character-Parsierung“ werden Parser definiert, die auf einzelnen Zeichen operieren. Die dabei gemeinte Tags werden von dem erkannten Parser automatisch in das Parsierungsergebnis eingefügt. Diese Parser-Ebene wird verwendet, um kleine, domänen-spezifische Entitäten ohne syntaktischen Ballast hinschreiben zu können, wie o.e. Takt- und Tonartangaben, oder Aktenzeichen, Kfz-Kennzeichen, Kalenderdaten, Personen-Namen, chemische Formeln, etc.

Die Definitionen und Transformationen sehen dann aus wie ...

```
chars tk = [von (S:digit)~+] ~ ([bis "-" (S:digit)~+] )?
```

```
siehe auch vorher schon #tk 3-5 oder #tk 7
```

```
siehe auch vorher schon
```

```
<tk><von>3</von><bis>5</bis></tk> oder <tk><von>7</von></tk>
```

Die beiden Ebenen von Parsierung funktionieren dezidiert unterschiedlich: die Zeichen-Parser werden nicht-deterministisch ausgewertet und arbeiten „gierig“. Ihre automatische Übersetzung in DTD-Definitionen ist deshalb nicht immer möglich oder zweckmäßig. Ausserdem eignen sich nur für relativ kleine Eingaben. Die in der Standard-Distribution vorgesehene MathML-Unterstützung ist die z.Zt. größte bekannte Anwendung und gerät schnell an die Grenzen der Praktikabilität. Da müssen evtl. im Verlaufe der weiteren Entwicklung intelligentere Parsierungstechniken eingesetzt werden.

5 Moduln, Parametrisierung und Rewriting

Die Notwendigkeit einer je Domäne speziellen Lösung setzt meistens voraus (a) das Vorhandensein einer Grundarchitektur, da ja nicht alle Komponenten eines Textes *ex ovo* neu

definiert werden sollen, und in diesem Falle weiterhin (b) einen Mechanismus zur Parametrisierung zwecks Anpassung dieser Grundarchitektur.

D2d verwendet ein Modul-System. Ein Modul dient zunächst zur Abgrenzung von Gültigkeitsbereichen von Definitionen. Erwähnte Grundarchitektur heißt `d2d_gp`. Sie ist gegeben als Sammlung von Modulen und beschreibt eine Textstruktur, die sich an „`LaTeX article`“, HTML u.ä. orientiert.

Man beachte, dass die gesamten Definitionen ca. 50 Druckseiten beinhalten, darin enthalten der XSLT-Quelltext für die vollständige Übersetzung in das XHTML-Backend, und die (fast !-) vollständige Benutzerdokumentation in Englischer Sprache. (Man vergleiche das mit der Spezifikation „OOXML“ von Microsoft mit ihren angeblich sechstausend Seiten!-)

Zwecks Anpassung kann ein Modul mehrfach importiert werden, wobei es unterschiedlichen *Termersetzungen* unterworfen wird. Die Menge der bei einer zukünftigen Wiederverwendung in einem noch unbekannten Kontext auszutauschenden Definitionen ist (natürlicherweise) nicht bekannt. Deshalb ist in D2d potentiell jede Referenz ein Modulparameter, und ein freies Rewriting kann auf jedes zu importierende Modul angewandt werden und beliebige Referenzen ersetzen, – im Gegensatz zu dynamischer Bindung aber mit kalkulierbaren Auswirkungen.

Die Konsequenzen dieses einfachen Ansatzes sind allerdings recht komplex, und unsere Forschung ist noch im Gange (z.B. bzgl. Terminierung!). Besondere Beachtung verdient dabei der Fall, dass eine Reifikation von transformierten Definitionen notwendig wird. Dies ist z.B. der Fall wenn Benutzerdokumentation generiert werden soll; dann müssen unterschiedliche Parametrisierungen mit ununterscheidbarem Effekt möglichst unifiziert werden, um die Information an den Benutzer so schlank wie möglich zu halten.

6 Dokumentation der Definitionsmoduln

Die Generierung von Dokumentation reist Erster Klasse: mit jeder Definition kann ein Dokumentationstext in verschiedenen Sprachen angegeben werden. Dieser benutzt eine Variante des Standardformates `d2d_gp`, mit allen Möglichkeiten von Verlinkung und Präsentation.

Aus diesen Texten baut die Implementierung einen XHTML-Text zusammen, der ergänzt wird durch den automatisch generierten Benutzungs-Graphen, der wiederum als Liste von Sprungzielen und als interaktive SVG-Graphik präsentiert wird. Aus den textuellen Termen der regulären Ausdrücke werden WIRTH-Diagramme erzeugt, die wiederum „anklickbar“ sind, etc.

Als Beispiel diene die Dokumentation von `d2d_gp` selber, zu finden am Ende der Tool-Dokumentation [met].

7 Unterstützung von XSLT

D2d unterstützt XSLT in zweierlei Hinsicht: einmal gibt es für jedes definierte Eingabeformat „M“ einen „XSLT-Modus“ des Parsierens, in dem nicht Texte des Typs M parsiiert werden, sondern XSLT-Programme, die Texte des Typs M hervorbringen. Der Inferenzmechanismus, der es erlaubt, mit minimalem Mark-Up M-Dokumente hinzuschreiben, wird modifiziert verwendet um zwischen den sog. „Target-Elementen“ und den XSLT-Befehlen zu unterscheiden, – es arbeiten quasi zwei Parser als Ko-Routinen. Dies erlaubt XSLT hinzuschreiben, das schon fast wie eine „richtige Programmiersprache“ aussieht:

```
#template #match a:tableOfContents
  #call makeVisibilitySwitch
    #arg id-switch #xp'_switch_TOC'
    #arg id-table #xp'__div_TOC'
  #p#id __div_TOC
    #a #id TOC #span #class tocTitle #apply #xp a:caption#/a
    #br/
    #choose
      #when /descendant::a:part
        #apply #xp /descendant::a:part #mode maketoc
      #other
        #apply #xp /descendant::a:h1 #mode maketoc
  #/p
```

Nebenbei erreicht diese Eingabemethode mit einfachsten Mitteln auch einen großen Schritt in Richtung typkorrekter XSLT Transformationen, wie beschrieben in [LT15].

Zum zweiten kann gleich bei der Definition eines D2d-Moduls eine Menge von Übersetzungsregeln mitgegeben werden, und zwar für alle Kombinationen von neu definierten Element-Arten und verschiedenen Ausgabeformaten („Backends“). Beim Prozessieren eines Eingabetextes werden dann für jedes ausgewählten Backend diese XSLT Regeln kombiniert und angewandt.

Literatur

- [Br06] Bray, Tim; Paoli, Jean; Sperberg-McQueen, C.M.; Maler, Eve; Yergeau, Francois; Cowan, John: . Extensible Markup Language (XML) 1.1 (Second Edition). W3C, <http://www.w3.org/TR/2006/REC-xml11-20060816/>, 2006.
- [Le15] Lepper, Markus: Gustav Mahler, Dritte Sinfonie, Erste Abtheilung. Senztempo, 2015. http://senztempo.de/mahler/gmahler_sinf3_satz1.html.
- [LT11] Lepper, Markus; Trancón y Widemann, Baltasar: d2d — a Robust Front-End for Prototyping, Authoring and Maintaining XML Encoded Documents by Domain Experts. In (Filipe, Joaquim; J.G.Dietz, Hrsg.): Proceedings of the International Conference on Knowledge Engineering and Ontology Deleopgnt, KEOD 2011. SciTePress, Lisboa, S. 449–456, 2011.

- [LT15] Lepper, Markus; Trancón y Widemann, Baltasar: A Simple and Efficient Step Towards Type-Correct XSLT Transformations. In: Proceedings 26th International Conference on Rewriting Techniques and Applications (RTA 2015). Jgg. 36 in LIPICS. Dagstuhl Publishing, S. 350–364, 2015.
- [LTW01] Lepper, Markus; Trancón y Widemann, Baltasar; Wieland, Jacob: Minimize Mark-Up ! – Natural Writing Should Guide the Design of Textual Modeling Frontends. In: Conceptual Modeling — ER2001. Jgg. 2224 in LNCS. Springer, November 2001.
- [met] The BandM Metatools Homepage. <http://www.bandm.eu/metatools/>.

Planung von Migrationsprojekten

Harry M. Sneed¹

Abstract: Dieses halbtägige Tutorial behandelt die Planung einer Softwaremigration. Viele Unternehmen in Österreich und anderswo haben alte Softwaresysteme die sie gerne in die neue Welt hinüberretten möchten. Es stehen ihnen dazu einige Alternativen mit unterschiedlichen Kosten und Nutzen bevor. Die Alternativen reichen von einer kompletten Neuentwicklung bis zu einer Kapselung der alten Komponente. Die Auswahl der richtigen Alternative ist ausschlaggebend für den Erfolg der Migration. Wählen sie die falsche Alternative wird die Migration scheitern oder wesentlich mehr kosten als geplant. Der Schlüssel zur Auswahl der richtigen Alternative ist eine ausführliche Kosten/Nutzen Analyse. Sie müssen ihre Altsysteme messen um festzustellen wie groß, wie komplex und wie gut sie sind. Davon hängt es ab was sie damit anfangen können. Als nächstes werden die Kosten jeder Alternative auf der Basis der Größe, Komplexität und Qualität geschätzt. Danach wird der Nutzen einer jeden Alternative aufgrund einer Nutzwertanalyse ermittelt. Schließlich wird entschieden welche Alternative man vorzieht. Der Entscheidungsprozess wird erläutert an Hand von Beispielen aus der industriellen Praxis. Die Teilnehmer des Tutoriums erhalten automatisierte Werkzeuge um den Entscheidungsprozess zu unterstützen – Werkzeuge für die Analyse der Altsysteme sowie Werkzeuge für die Kostenschätzung der Migrationsalternative. Die Arbeitsweise und Ergebnisse der Werkzeuge werden in einer Fallstudie demonstriert.

Keywords: Legacy-Systeme, Softwaremigration, Softwaremessung, Aufwandsschätzung, COCOMO, Function-Point, Data-Point, Object-Point, Migrationskostenanalyse, Risikoanalyse, Migrationsnutzen, Migrationsfallstudien.

1 Themenhintergrund

Etliche Anwendungsbetriebe stehen vor dem Problem sich von ihrer Altlasten zu befreien. Es kann sein, dass sie ihre alten Daten migrieren wollen, ihren alten Code ablösen möchten oder wieder von vorne mit neuen Systemen anfangen wollen. Was das zweite Alternativ anbetrifft, haben sie die Wahl zwischen verschiedenen Befreiungsstrategien, darunter:

- Die automatische Konvertierung der alten Codes
- Die manuelle Reimplementierung des alten Codes
- Die Kapselung der alten Code-Komponente und
- Die Neuentwicklung der Codebasis

¹ Technische Universität Dresden, Institut für Softwaretechnik, ZT-Prentner-IT GmbH, Wien-Kagran,
Harry.Sneed@T-Online

Welche Strategie sie wählen, hängt von den Kosten und Nutzen sowie von den Risiken der jeweiligen Strategie ab. D.h. sie müssen im Vorfeld der Migration eine Kosten/Nutzen/Risiko Analyse durchführen. Alle drei Analyseprojekte können neben einander stattfinden. Die Ergebnisse werden am Ende zusammengefasst und mit einander verglichen um zu einer Entscheidung zu kommen welches Alternativ das größte „Return on Investment“ verspricht. Das ROI wird hier mit folgender Wirtschaftlichkeitsrechnung

$$\text{ROI} = (\text{Nutzen} - \text{Kosten}) / (\text{Kosten} * \text{Risiken})$$

errechnet. Irrationale, politische Entscheidungen über die Zukunft der betrieblichen Softwaresysteme sind möglichst zu vermeiden. Daher der quantitative Ansatz. Es geht darum einen rationalen, nachvollziehbaren Entscheidungsprozess einzuführen. In diesem eintägigen Tutorial wird an Hand von Fallbeispielen aus der langjährigen Migrationspraxis des Referenten vermittelt wie man bei einer derartigen Voruntersuchung vorzugehen hat.

2 Tutorium Inhalt

Das Tutorium behandelt die drei Schienen einer Wirtschaftlichkeitsanalyse:

- Risikoanalyse
- Kostenschätzung
- Nutzenanalyse.

2.1 Risikoanalyse

Bei der Risikoanalyse geht es darum die wichtigsten Risiken zu identifizieren und zu bewerten. Dazu sind Standardformulare zu verwenden, in denen die finanzielle Auswirkung und die Wahrscheinlichkeit eines jeden Risikos eingeschätzt werden. Hier werden auch die möglichen Gegenmaßnahmen erfasst, Maßnahmen welche die Risikoauswirkung eindämmen können. Aus ihnen geht der Risikoreduktionsfaktor hervor. Die Risiken werden in diverse Risikoklassen aufgeteilt – z.B. organisatorische Risiken, personale Risiken, technische Risiken und Geschäftsrisiken. Am Ende wird ein Risikofaktor als

$$\text{Risiken} = (\text{Auswirkungsgrad} * \text{Wahrscheinlichkeit}) * (1 - \text{Risikoreduzierungsfaktor})$$

Das Ergebnis – die Risiken – fließt in die Wirtschaftlichkeitsrechnung hinein.

2.2 Kostenschätzung

Die Kostenschätzung einer Migration setzt voraus, dass das Migrationsobjekt, in diesem Falle der Code, vermessen wird. Die Vermessung zielt darauf hin, bestimmte quantitative Größen aus dem Code zu gewinnen, die in Personenaufwand umgesetzt werden können. Solche Größen sind u.a. Codezeilen, Anweisungen, Data-Points, Function-Points und Object-Points. Diese Points werden durch eine automatisierte Analyse des Source-Codes gewonnen, die allenfalls mehrere Stunden dauern sollte. Die gleichen Größen werden auf gleicher Weise aus Systemen abgeleitet, die schon migriert worden – die Benchmark-Systeme. Zu diesen Benchmark-Projekten werden hoffentlich Aufwandsdaten vorliegen. Durch die Gegenüberstellung der Aufwände mit den Systemgrößen wird die Produktivität in diesen Projekten ermittelt. Über die Teilung der Größen des neu zu migrierenden Systems durch die mittlere Produktivität der bereits migrierten Systeme kommt man zum rohen Aufwand.

Der rohe Aufwand wird durch einen Einflussfaktor und möglicherweise durch einen Komplexitäts- und Qualitätsfaktor justiert. Diese Faktoren spiegeln die Differenz zwischen den Benchmark-Systemen und dem Zielsystem wider.

Aufwand = (Systemgröße / Produktivität) * Einflussfaktor * Komplexität * Qualität

Das Ergebnis ist der Personalaufwand, der sich in Personalkosten umsetzen lässt. Die Personalkosten sind durch Overheadkosten, Softwarekosten und Hardwarekosten zu ergänzen. Die ergänzten Kosten geben die Gesamtkosten der Migration.

2.3 Nutzenanalyse

Die Nutzenanalyse einer Migration setzt voraus, dass man quantitative Daten über interne und externe Qualitätseigenschaften hat. Zu den internen Eigenschaften gehören Maßzahlen die aus dem Code ableitbar sind, Eigenschaften wie Wiederverwendbarkeit, Portierbarkeit, Testbarkeit und Wartbarkeit. Hinzu kommen die eigentlichen Wartungsaufwände, Fehlerraten, System Performanzwerte, Operationskosten und Benutzerzufriedenheitsgrade. Diese sind die Istwerte. Die Sollwerte sind die gleichen Maßzahlen multipliziert mit dem jeweiligen Verbesserungsfaktor. Ein Verbesserungsfaktor von 1,2 bedeutet, dass wir uns einen 20% Verbesserung durch die Migration versprechen. Der Gesamtverbesserungsfaktor ist die Summe aller einzelnen Verbesserungsfaktoren. Er wird mit dem gegenwärtigen Nutzen der Zielapplikation multipliziert um den neuen Nutzwert zu errechnen.

Neuer Nutzwert = alter Nutzwert * Gesamtverbesserungsfaktor

Es kann auch zu einem neuen Nutzwert kommen der niedriger als der alte Nutzwert ist, nämlich dann wenn einzelne Verbesserungsfaktoren < 1 ausfallen. Wenn der Verbesserungsfaktor für Wartbarkeit auf 1,2 und der Verbesserungsfaktor für Performanz auf 0,8 geschätzt werden, gleichen sie sich aus. Der

Gesamtverbesserungsfaktor wird 1, d.h. keine Verbesserung. Da einzelne Eigenschaften wie Zuverlässigkeit wichtiger sein können als Andere muss es eine Möglichkeit geben die Eigenschaften zu gewichten. Der Gesamtverbesserungsfaktor ist dann die Summe der gewichteten Verbesserungsfaktoren.

Der Nutzen des Migrationsprojektes ist der neue Nutzwert – der alte Nutzwert. Dies ist der Nutzen der in die Wirtschaftlichkeitsrechnung einfließt.

3 Tutorium Ablauf

Das Tutorium ist eine Mischung aus Vorträgen, Übungen und Demonstrationen. Es beginnt mit einer Einführung in die Migrationswirtschaftlichkeitsanalyse. Danach folgt eine kurze Lektüre über Risikoanalyse, eine Demonstration der Risikoanalyse und eine Übung in Risikoanalyse eines kleinen Migrationsprojektes. Auf die Risikoanalyse folgt die Kostenschätzung mit Vortrag, Demonstration und Übung. Die Teilnehmer bekommen eine Codegröße, einen Komplexitätsmaß, einen Qualitätsmaß und einen Produktivitätsmaß um den Migrationsaufwand zu schätzen. Danach folgen ein Vortrag, eine Demonstration und eine Übung in Migrationsnutzenanalyse. Zum Schluss wird am PC-Arbeitsplatz eine Tool-gestützte Wirtschaftlichkeitsanalyse für ein Musterprojekt vorgeführt.

4 Abbildungen

Die Teilnehmer des Tutoriums erhalten zwei MS-Windows Werkzeuge die sie mitnehmen können:

- SofAudit für die Prüfung und Messung von Legacy Systeme in Assembler, C, COBOL, PL/I, Natural und VisualAge.
- SoftCalc für die Schätzung der Migrationsaufwände nach Anweisungen (COCOMO), Data-Points, Function-Points und Object-Points

Für die Anwendung dieser Tools bekommen die Teilnehmer auch zwei Fallbeispiele aus der IT-Praxis die sie nachvollziehen können.

5 Tutorium Ergebnisse

Nach dem Besuch dieses Tutoriums sind die Teilnehmer in der Lage ihre Legacy Systeme zu messen und die Ablösemöglichkeiten nicht nur zu schätzen sondern auch im Hinblick auf ihren Nutzen zu bewerten. Es geht darum dass sie für jede Situation sie den geeigneten Ausweg finden. Alte Systeme lassen sich neuentwickeln, neu-

implementieren, konvertieren und kapseln. Es komme darauf an die richtige Lösung zu finden. Dies wird in diesem Seminar vermittelt.

Literaturverzeichnis

- [BG99] Bisbal, J.; Grimson, J; Lawless, D.; Wu, B.: Legacy Information Systems: Issues and Directions. IEEE Software, Vol. 16, No. 5, S. 103-111, 1999.
- [Bo99] Boehm, B. W. u.a.: Software Cost Estimation with COCOMO-II, Prentice-Hall, Upper Saddle River, N.J., 1999
- [DF08] De Lucia, A./Francese, R.: "Developing System Migration Methods and Tools for Technology Transfer", in Journal of Systems and Software, Vol. 70, No. 1, 2008, s. 3
- [Di89] Dietrich, W.: "Saving a Legacy System with Objects", Proc. of OOPSLA-88, ACM Press, New York, 1989, p. 5
- [Eb04] Ebert, J.: Software-Reengineering - Umgang mit Software-Altlasten. In: Informatiktage 2003. Konradin-Verlag, Grasbrunn, S. 24-31, 2004.
- [Me07] Meyer, D. : "Modernisierung von Legacy Systemen - Risiken und Kosten senken durch Automation", Objektspektrum, Nr. 5, Sept. 2007, S.33.
- [SP03] Seacord, R./Plakosh, D./Lewis, G. (2003): Software Modernization, Addison-Wesley, Boston, 2003
- [Sn84] Sneed, H.M.: "Software Renewal – A case study", IEEE Software, Vol. 1, No. 3, July, 1984, s. 56
- [Sn91] Sneed, H.M: Softwaresanierung, Rudolf Müller Verlag Köln, 1991
- [Sn91] Sneed, H.M.: "Economics of Software Reengineering", Journal of Software Maintenance, Vol. 3, No. 3, Sept. 1991, s. 163
- [Sn95] Sneed, H.M.: Planning the Reengineering of Legacy Systems. IEEE Software, Vol. 12, No. 1, S. 24-34, Jan. 1995, S. 24.
- [Sn96] Sneed, H.M.: "Encapsulating Legacy Software for Reuse in Client Server Systems" Proc. Of 3rd WCRE, IEEE Computer Society Press, Monterey, CA., Nov., 1996, p. 104
- [Sn97] Sneed, H.M.: „Metriken für die Wiederverwendbarkeit von Softwaresystemen“, in Informatikspektrum, Vol. 6, S. 18-20, 1997.
- [Sn99] Sneed, H. M.: "Risks involved in Reengineering Projects" Proc. of WCRE-1999, IEEE Computer Society Press, Atlanta, Oct. 1999, p. 204-211
- [Sn99] Sneed, H.M.: Objektorientierte Softwaremigration. Addison-Wesley, Bonn, 1999
- [Sn01] Sneed, H.M.: "Wrapping Legacy COBOL programs behind an XML Interface", Proc. of Working Conference on Reverse Eng., IEEE Computer Society Press, Stuttgart, Oct. 2001, p. 189.

- [Sn05a] Sneed, H.M.: Software-Projektkalkulation - Praxiserprobte Methoden der Aufwandsschätzung für verschiedene Projektarten. Hanser, München/Wien, 2005, S. 159.
- [Sn05b] Sneed, H.M.: "Estimating the Costs of a Reengineering Project", IEEE Proc. of 12th WCRE, Computer Society Press, Nov. 2005, Pittsburgh, PA., s. 111
- [SW10] Sneed, H., Wolf, E., Heilmann, H.: "Software Migration in der Praxis", dpunkt.verlag, Heidelberg, 2010.
- [TB02] Tortorella, M./ Bodhum, T./ Albanese, C.: " A Toolkit for Applying a Migration Strategy", IEEE Proc. of 6th CSMR, Computer Society Press, Budapest, March 2002, S. 154.
- [Vo05] Von Hahn (2005): Werterhaltung von Software – Planung und Bewertung von Reengineering Maßnahmen am Beispiel von Standard Software, Deutscher Universitätsverlag, Wiesbaden, 2005, S. 173
- [Wa99] Warren, I.: The Renaissance of Legacy Systems: Method Support for Software-System Evolution. Springer, London, 1999.

Doktoranden-Symposium

Doktoranden–Symposium der Informatik 2016

Ernst Denert¹, Mathias Lux² und Stefan Wagner³

1 Das Symposium

Gedankenaustausch und Disput beleben die Wissenschaft. Gerade auch in einer Promotion ist es hilfreich, sich oft und von verschiedenen Seiten Rückmeldung zum Stand und zur Planung der weiteren Arbeiten zu holen. Das Doktoranden-Symposium der INFORMATIK 2016 bietet Forscherinnen und Forschern die Möglichkeit, ihre laufenden Doktorarbeiten vorzustellen.

Die Promotionen stammen aus allen Gebieten der Informatik. Im Symposium präsentieren die Teilnehmer das Thema und den Forschungsplan in einem 20-minütigen Vortrag vor anderen Doktoranden, vor jungen und erfahrenen Wissenschaftlern, vor Professoren und Praktikern. Zudem können sie in einer Poster-Ausstellung allen Tagungsteilnehmern zeigen, was sich in der Informatik-Forschung tut. Das Symposium ist eine gute Gelegenheit, über den Horizont des eigenen Instituts hinaus zu blicken, sich den Fragen des Publikums zu stellen und Kontakte zu knüpfen. Voraussetzung für die Teilnahme war die prägnante, schriftliche Kurzdarstellung der geplanten Dissertation. Anhand der Kurzdarstellung wurde vom Komitee ob der Einladung zum Symposium entschieden. Ausschlaggebend waren die folgenden Kriterien:

- Motivation, Problemstellung und Lösungsansatz: Ist die Dissertation innovativ, löst sie eine oder mehrere Forschungsfragen und wie tut sie das?
- Verwandte Arbeiten: Wie gut ist die Dissertation in die existierende Forschungslandschaft eingebettet?
- Forschungsplan: Gibt es einen Weg zum Ziel, welche Schritte werden benötigt um die Dissertation abzuschließen?
- Aktueller Stand: Was sind die aktuellen Ergebnisse und wie weit ist die Forschungsarbeit fortgeschritten?
- Lesbarkeit der schriftlichen Einreichung: Wird das Dissertationsvorhaben ausreichend gut und überzeugend schriftlich präsentiert?

¹ TU München & Ernst-Denert-Stiftung für Software-Engineering, München, DE, ernst.denert@web.de

² Alpen-Adria Universität Klagenfurt, Institut für Informationstechnologie, Klagenfurt, AT, mlux@itec.aau.at

³ Universität Stuttgart, Institut für Softwaretechnologie, Stuttgart, DE, stefan.wagner@informatik.uni-stuttgart.de

2 Beiträge

Das Komitee bewertete die Einreichungen und wählte aus 19 Einreichungen die 12 Teilnehmerinnen und Teilnehmer aus.

- Sebastian Bayerl (University of Passau): Managing Distributed Statistical Data in the Semantic Web,
- Timm Caporale (Karlsruher Institut für Technologie KIT): Geschäftsprozessmodellierung mit kontrollierter natürlicher Sprache
- Michael Grunwald (HTWG Konstanz): Wahrnehmungsorientierte optische Inspektion von texturierten Oberflächen
- Daniel Hallmann (Otto-Friedrich-Universität Bamberg): Gemeinsame mentale Modelle in der agilen Softwareentwicklung: Ein Ansatz zur Erstellung von Gestaltungsempfehlungen für „gute“ erfahrungsspezifische User Stories
- Oliver Knodel (Technische Universität Dresden): Virtualisierung rekonfigurierbarer Hardware zur Steigerung von Rechenleistung und Sicherheit in einer flexiblen Cloud Architektur
- Jonas Lehner (Karlsruher Institut für Technologie KIT): Personal BPM – Anwenderorientierte Prozessmodellierung durch persönliches Workflow Management
- Franziska Marx (Institut für Informationsmanagement Bremen GmbH und Universität Bremen): Adaptive Informationsvisualisierung: Der Einsatz von Machine-Learning zur Entscheidungsunterstützung
- Benjamin Saul (Martin-Luther-Universität Halle-Wittenberg): Domänenspezifische Vorabanalyse und geschickte Generierung von Optimierungsproblemen
- Nicolai Spicher (University of Applied Sciences and Arts Dortmund and University Duisburg-Essen): Cardiac activity measurement from video signals of the human skin in ultra-high-field magnetic resonance imaging
- Maciej Sumalvico (Universität Leipzig): Statistisches Lernen wortbasierter Morphologie
- Malte Wattenberg (Fachhochschule Bielefeld): Bestimmung von Erfolgsfaktoren der Unternehmenskommunikation von KMU auf Facebook: Design und Zwischenergebnisse eines Multimethoden-Ansatzes
- Matthäus Zloch (Heinrich Heine Universität): Methods for Automatic Selection of Database Systems for Optimized Query Performance

Managing Distributed Statistical Data in the Semantic Web

Sebastian Bayer¹

Abstract: The RDF Data Cube Vocabulary is the W3C recommendation for publishing multi-dimensional statistical data in a semantic web format. A large and growing number of such data cubes is already available in the linked data cloud, but the data is distributed and isolated in many remote repositories. Having access to the data, applications can now be developed to explore and use it. To handle and correlate the distributed cubes, a process is necessary to merge them. This research aims at developing a homogeneous management process for such cubes reaching from creation to visualization. The focus is put on the merging of RDF data cubes. In order to access and integrate the distributed cubes, a suitable ranking and discovery mechanism is needed. Therefore, a similarity measure for cubes must be developed.

Keywords: Knowledge Discovery, Data Integration, RDF Data Cube Vocabulary, Linked Open Data

1 Motivation

The linked data cloud² offers open access to a vast amount of information. Here, companies, governments and research institutions publish various datasets, like financial statistics or historical data. The core idea of this semantic web is to publish and link different datasets using specialized vocabularies to later gain new insights from the data. The resource description framework (RDF) [WLC14] should be used for the datasets and to connect the data, but there are also lots of datasets available as comma separated values (CSV). The RDF Data Cube Vocabulary [CR14] is the W3C recommended data format for statistical linked data. It is mainly used to publish statistical datasets to the linked data cloud. The cube vocabulary defines a multi-dimensional data cube, which reproduces a data structure - known from traditional Online Analytical Processing (OLAP), namely the OLAP cube [CD97].

This research project targets the management of RDF data cubes in the semantic web to enable an integrated view on the currently distributed and heterogeneous cubes. Hereby, the schema mapping and the data fusion problem must be tackled, which is stated as one of the fundamental research questions for linked data [BHBL09].

Therefore, a homogeneous management process for RDF data cubes will be developed. This process focuses on the discovery and the merging of distributed data cubes, while taking the semantic nature of the datasets into account.

¹ University of Passau, Faculty of Computer Science and Mathematics, Innstraße 43, 94032 Passau, sebastian.bayerl@uni-passau.de

² <http://linkeddata.org>

2 Problem Setting

The process of managing data cubes can be divided into a sequence of tasks, which are shown in Figure 1. This process is based on the RDF Data Cube Vocabulary to employ a homogeneous data structure for statistical data that fosters semantic properties like inter-linked and disambiguated concepts. These properties are an additional value that is mostly not available in traditional data warehousing. This enables new functionality in all steps of the managing process, like the semantic comparison of data.

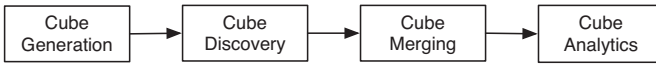


Fig. 1: RDF Data Cube Management Process

To open up this process for other data formats a *cube generation* step is necessary. Hereby, e.g. HTML tables or CSV documents are converted to RDF data cubes. Consequently, the following process can be used with statistical data that is initially not in the RDF data cube format. In order to find and access relevant cubes for the current use case, a suitable *cube discovery* mechanism is needed. Therefore, decentralized and remote repositories must be searched. Also a suitable similarity measures for cubes must be developed to be able to order cubes based on their content or structure.

<table border="1" style="border-collapse: collapse; width: 100%;"> <thead> <tr><th>Year</th><th>Tree</th><th>Count</th></tr> </thead> <tbody> <tr><td>2014</td><td>Birch</td><td>130</td></tr> <tr><td>2015</td><td>Oak</td><td>21</td></tr> <tr><td>...</td><td>...</td><td>...</td></tr> </tbody> </table> <p style="text-align: center; margin-top: 5px;">\mathcal{C}_1: Trees</p>	Year	Tree	Count	2014	Birch	130	2015	Oak	21	+	<table border="1" style="border-collapse: collapse; width: 100%;"> <thead> <tr><th>Year</th><th>Count</th></tr> </thead> <tbody> <tr><td>2014</td><td>72</td></tr> <tr><td>...</td><td>...</td></tr> </tbody> </table> <p style="text-align: center; margin-top: 5px;">\mathcal{C}_2: Pines</p>	Year	Count	2014	72	=	<table border="1" style="border-collapse: collapse; width: 100%;"> <thead> <tr><th>Year</th><th>Tree</th><th>Count</th></tr> </thead> <tbody> <tr><td>2014</td><td>Birch</td><td>130</td></tr> <tr><td>2015</td><td>Oak</td><td>21</td></tr> <tr><td>2014</td><td>Pine</td><td>72</td></tr> <tr><td>...</td><td>...</td><td>...</td></tr> </tbody> </table> <p style="text-align: center; margin-top: 5px;">\mathcal{C}_3: Trees (extended)</p>	Year	Tree	Count	2014	Birch	130	2015	Oak	21	2014	Pine	72
Year	Tree	Count																																			
2014	Birch	130																																			
2015	Oak	21																																			
...																																			
Year	Count																																				
2014	72																																				
...	...																																				
Year	Tree	Count																																			
2014	Birch	130																																			
2015	Oak	21																																			
2014	Pine	72																																			
...																																			

Fig. 2: Cube merging example

Cube merging is the central problem setting while managing data cubes. The main goal of cube merging is to integrate multiple isolated and potentially distributed cubes into a single dataset. Hereby, the validity of the structure and the data must be traced and maintained by finding suitable transformations. The merging process therefore produces an integrated view on a set of cubes. Figure 2 shows the simplified and abstract concept of the cube merging problem. Cube \mathcal{C}_3 contains all information from \mathcal{C}_1 and \mathcal{C}_2 . To do so, it is necessary to introduce the column *Tree* with the default value *Pine* in cube \mathcal{C}_2 . Similar to traditional data warehousing processes, OLAP operations like *roll-up* or *slicing* can now be performed on the integrated data cubes. This *cube analytics* process must efficiently enable the application of these operations. Visualizing the cubes can help to understand the integrated data and to draw new conclusions from it.

3 Approach

Figure 3 shows, how the different steps of the cube management process can be approached. In the following, these steps will be introduced in more detail.

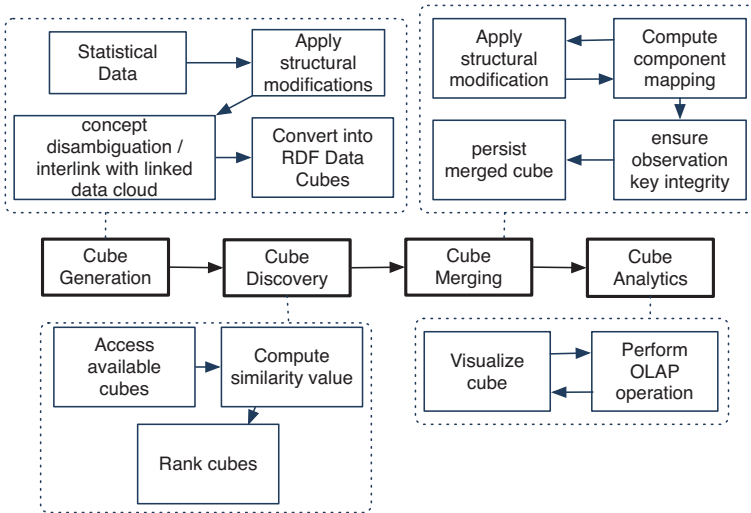


Fig. 3: Detailed Cube Management Process

The main problem of converting statistical data into cubes lies within the normalization of the potentially complex structured data to a homogeneous structure. HTML tables for example are designed to be human-readable but do not reassemble the standardized multi-dimensional data cubes. A simple row-by-row based conversion approach is not sufficient and therefore, the complex structured tables must be validated, modified and transformed, while preserving the meaning of the data. Also, the data values must be disambiguated and interlinked with the linked data cloud to enable the semantic features of the following process. This approach is already implemented and published for a subset of the German Reich Statistics dataset [BG15b].

Cube discovery can be approached as an information retrieval problem, because selecting an appropriately ordered list from all available cubes satisfies an information need. Hereby, syntactic and semantic properties of the cubes can be considered to develop similarity measures for cubes. Currently, measures based on string similarity, concept equality, hierarchical relatedness and semantic similarity are considered good candidates to compare the structure of the cubes. Therefore, a graph-based shortest-path algorithm that utilizes the DBpedia category dataset³ and a Word2Vec [Mi13] model will be implemented. This research will show if these measures or a combination of them are suitable to carry out the discovery process.

To be able to apply the merging process to cubes using the RDF Data Cube Vocabulary, it is necessary to analyze their structure and content to determine the properties of the resulting cube. An iterative process is used to adapt the structure of the cubes until a bijective mapping of the components can be found. This process tackles the following essential problems:

³ <http://wiki.dbpedia.org/services-resources/datasets/dbpedia-datasets>

1. Modify the structure of the input cubes to fit the structure of the resulting cube (Integration on schema level).
2. Detect and handle duplicate observations and therefore maintain the global key integrity. This makes it necessary to develop an update strategy (Integration on instance level).

Beside other metadata about the process, the provenance information is stored with the resulting cube to track modification of the data and to ensure data quality. See [BG15a] for more information on this topic.

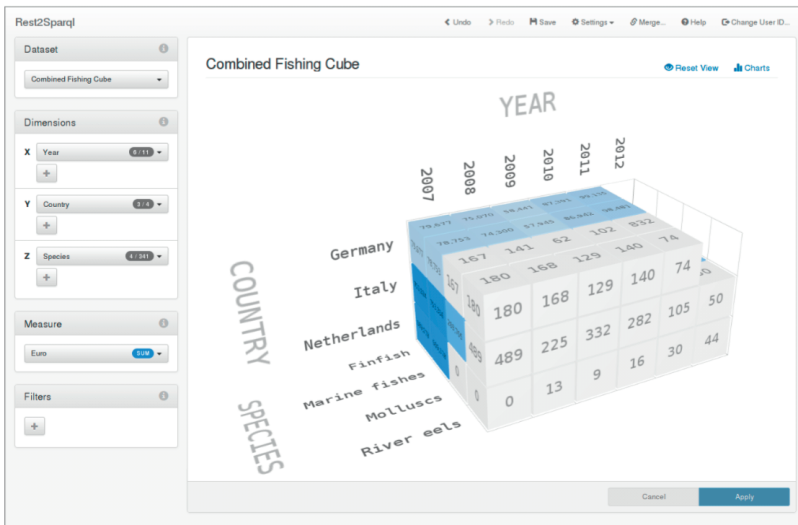


Fig. 4: Cube Visualization GUI

Cube analytics is the last step in the cube management process. A graphical user interface was implemented to visualize the (integrated) data cubes, shown in figure 4. It enables basic OLAP operations to manually browse and aggregate the data, using a 3D-rendered model.

4 Related Work

The presented process is similar to traditional Data Warehousing in various aspects. On the one hand, both deal with the problems of cleaning, transforming, merging and aggregating data [Le03]. On the other hand, handling RDF-based and interlinked open data from the linked data cloud yields advantages and new challenges, like semantic hierarchies and similarities.

There is work describing the process of converting statistical data into RDF data cubes. The *CSV2DataCube* tool converts CSV files [Sa12] and the *Data Extractor* is able to handle HTML and tables, extracted from PDF documents [St12]. These approaches allow basic restructuring, but complex structural modifications cannot be applied to the data.

Multidimensional hierarchical spaces are an essential part in OLAP processes. They are employed for data aggregation, but can also be used to define similarity measures for multidimensional data [BRV11]. Here, the similarity of cubes is computed by utilizing the distances of the cube facts according to the appropriate hierarchies. This approach handles OLAP cubes, but does not consider the specific properties of the RDF Data Cube Vocabulary. Several approaches to determine the similarity of words or linked data concepts have been presented in literature. An extensive description of semantic measures can be found in [Ha13]. Here, general definitions, the basics and a comprehensive classification can be found. The author of [Re99] describes the semantic similarity in a taxonomy in great detail. The survey [EAM14] compares state-of-the-art semantic similarity measures.

The authors of [KSH14] propose an approach to integrate RDF Data Cubes into a global cube based on the Drill-Across operation [ASS02]. If correct mappings for the contained cubes are known, merging and querying operations are possible. The reuse phase of the OpenCube lifecycle [Ka14] proposes approaches, how RDF data cubes can be handled.

5 Conclusion and Future Work

This paper presents a management process for RDF data cubes. This process can be divided into the independent sub-steps creation, discovery, merging and analytics. First results regarding cube creation and cube merging are already published [BG15b, BG15a]. Therefore, research prototypes have been implemented and evaluated. Currently, sophisticated similarity measures for cubes are developed to replace a naive approach, connecting the pool of available cubes with the merging process. This will show, which syntactic or semantic properties of RDF Data Cubes can be reasonably used for cube discovery. This will enable the comparison and ranking of data cubes.

6 Acknowledgement

This research is supervised by Prof. Dr. Michael Granitzer, University of Passau, Faculty of Computer Science and Mathematics / MiCS - Media Computer Science.

References

- [ASS02] Abelló, Alberto; Samos, José; Saltor, Felix: On Relationships Offering New Drill-across Possibilities. In: Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP. ACM, pp. 7–13, 2002.
- [BG15a] Bayerl, Sebastian; Granitzer, Michael; Bacon: Linked Data Integration Based on the RDF Data Cube Vocabulary. In: Proceedings of the 5th International Conference on Web Intelligence, Mining and Semantics. WIMS '15, ACM, New York, NY, USA, pp. 14:1–14:6, 2015.
- [BG15b] Bayerl, Sebastian; Granitzer, Michael: Data-transformation on Historical Data Using the RDF Data Cube Vocabulary. In: Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business. i-KNOW '15, ACM, New York, NY, USA, pp. 15:1–15:8, 2015.

- [BHBL09] Bizer, Christian; Heath, Tom; Berners-Lee, Tim: Linked Data – The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22, 2009.
- [BRV11] Baikousi, Eftychia; Rogkakos, Georgios; Vassiliadis, Panos: Similarity measures for multidimensional data. In: *Data Engineering (ICDE), 2011 IEEE 27th International Conference on*. IEEE, pp. 171–182, 2011.
- [CD97] Chaudhuri, Surajit; Dayal, Umeshwar: An Overview of Data Warehousing and OLAP Technology. *ACM SIGMOD Record*, 26(1):65–74, March 1997.
- [CR14] Cyganiak, Richard; Reynolds, Dave: The RDF Data Cube Vocabulary. W3C recommendation, W3C, January 2014. <http://www.w3.org/TR/2014/REC-vocab-data-cube-20140116/>.
- [EAM14] Elavarasi, S Anitha; Akilandeswari, J; Menaga, K: A Survey on Semantic Similarity Measure. *International Journal of Research in Advent Technology*, 2(3):389–398, March 2014.
- [Ha13] Harispe, Sébastien; Ranwez, Sylvie; Janaqi, Stefan; Montmain, Jacky: Semantic Measures for the Comparison of Units of Language, Concepts or Entities from Text and Knowledge Base Analysis. *CoRR*, 2013.
- [Ka14] Kalampokis, Evangelos; Karamanou, Areti; Nikolov, Andriy; Haase, Peter; Cyganiak, Richard; Roberts, Bill; Hermans, Paul; Tambouris, Efthimios; Tarabanis, Konstantinos: Creating and Utilizing Linked Open Statistical Data for the Development of Advanced Analytics Services. In: *Second International Workshop for Semantic Statistics SemStats*. 2014.
- [KSH14] Kämpgen, Benedikt; Stadtmüller, Steffen; Harth, Andreas: Querying the Global Cube: Integration of Multidimensional Datasets from the Web. In: *Knowledge Engineering and Knowledge Management*, pp. 250–265. Springer, 2014.
- [Le03] Lehner, Wolfgang: *Datenbanktechnologie für Data-Warehouse-Systeme: Konzepte und Methoden*. dpunkt-lehrbuch. dpunkt-Verlag, 2003.
- [Mi13] Mikolov, Tomas; Chen, Kai; Corrado, Greg; Dean, Jeffrey: Efficient Estimation of Word Representations in Vector Space. *CoRR*, 2013.
- [Re99] Resnik, Philip et al.: Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *J. Artif. Intell. Res.(JAIR)*, 11:95–130, 1999.
- [Sa12] Salas, Percy E. Rivera; Martin, Michael; Mota, Fernando Maia Da; Auer, Sören; Britman, Karin; Casanova, Marco A.: Publishing Statistical Data on the Web. In: *Sixth IEEE International Conference on Semantic Computing, ICSC 2012, Palermo, Italy, September 19-21, 2012*. pp. 285–292, 2012.
- [St12] Stegmaier, Florian; Seifert, Christin; Kern, Roman; Höfler, Patrick; Bayerl, Sebastian; Granitzer, Michael; Kosch, Harald; Lindstaedt, Stefanie N.; Mutlu, Belgin; Sabol, Vedran; Schlegel, Kai; Zwicklbauer, Stefan: Unleashing Semantics of Research Data. In: *Specifying Big Data Benchmarks - First Workshop, WBDB 2012, San Jose, CA, USA, May 8-9, 2012, and Second Workshop, WBDB 2012, Pune, India, December 17-18, 2012, Revised Selected Papers*. pp. 103–112, 2012.
- [WLC14] Wood, David; Lanthaler, Markus; Cyganiak, Richard: RDF 1.1 Concepts and Abstract Syntax. W3C recommendation, W3C, February 2014. <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>.

Geschäftsprozessmodellierung mit kontrollierter natürlicher Sprache

Timm Caporale¹

Abstract: Organisationen passen ihre Geschäftsprozesse immer öfter und schneller an die sich kontinuierlich ändernden Rahmenbedingungen an. Dabei werden die Prozesse üblicherweise zunächst modelliert, dann analysiert und dann implementiert. Eine bleibende Herausforderung in diesem Prozess betrifft die angemessene Einbeziehung der Wissensträger bzw. Stakeholder. Im Rahmen dieses Beitrags wird eine Modellierungstechnik vorgestellt, die natürliche Sprache als Grundlage der Geschäftsprozessmodellierung verwendet. Die Technik benutzt Natural Language Processing-Techniken, die durch die Verwendung von Satzschablonen aus kontrollierter natürlicher Sprache Geschäftsprozessmodelle erzeugen. Es wird hierzu eine kontextfreie Grammatik entworfen, ein Parser erzeugt, der erzeugte Parsebaum analysiert und als Geschäftsprozessmodell ausgegeben. Mit diesem Ansatz kann garantiert werden, dass die in natürlicher Sprache vorliegende textuelle Beschreibung exakt die Semantik des Geschäftsprozessmodells repräsentiert. Stakeholder, die am Modellierungsprozess beteiligt sind, aber keine Expertise in der Geschäftsprozessmodellierung besitzen, können mit Hilfe der textuellen Beschreibung die Semantik des Geschäftsprozessmodells erfassen.

Keywords: business process modeling, natural language processing

1 Motivation und Problemstellung

Externe Rahmenbedingungen wie beispielsweise der globale Wettbewerbsdruck, komplexe Geschäftsmodelle und der Wunsch nach einer stärkeren Kundenorientierung führen dazu, dass Organisationen ihre Strategie und damit auch ihre Geschäftsprozesse ständig anpassen müssen. Methoden für diese Anpassung bietet das Geschäftsprozessmanagement (BPM). Der natürlichen Sprache kommt dabei in mehreren Phasen eine große Bedeutung zu. In der Identifikationsphase zum Beispiel werden viele Dokumente in natürlicher Sprache erzeugt und analysiert. Auch in den Interviewtechniken, die zur Identifikation der Geschäftsprozesse benutzt werden, ist die natürliche Sprache Grundlage des Verständnisses. Existierende Arbeiten unterstreichen die Relevanz der natürlichen Sprache für das Verständnis von Geschäftsprozessmodellen. Deshalb empfehlen auch viele Arbeiten sich bei der Formulierung textueller Beschriftungen in Geschäftsprozessen an eine bestimmte Syntax zu halten. Viele Wissensträger haben aber leider keine oder nicht ausreichende Erfahrungen im Bereich BPM, um Geschäftsprozessmodelle eigenständig zu erstellen und zu validieren, so dass das Vorgehensmodell der meisten existierenden Ansätze im

¹ Karlsruher Institut für Technologie (KIT), Institut für Angewandte Informatik und Formale Beschreibungsverfahren (AIFB), Kaiserstr. 89, 76133 Karlsruhe, tim.caporale@kit.edu

Allgemeinen vorsieht, dass in einem ersten Schritt die Geschäftsprozessmodellierer das Prozesswissen einer Organisation durch zum Beispiel (un-)strukturierte Experteninterviews oder Workshops von den Wissensträgern extrahieren und daraus Geschäftsprozessmodelle ableiten. Die Validierung inwiefern die textuellen Beschreibungen mit den erzeugten Geschäftsprozessmodellen übereinstimmen, bleibt dabei für die Wissensträger eine große Herausforderung.

Der Forschungsbereich der Spracherkennung (Natural Language Recognition), der in den letzten Jahren große Fortschritte gemacht hat, entwickelt zunehmend neue Techniken, die eine Verarbeitung natürlicher Sprache durch IT-Systeme ermöglichen (Natural Language Processing). Für das Geschäftsprozessmanagement ergeben sich hieraus neue Möglichkeiten. Dieser Beitrag stellt einen Ansatz zur Geschäftsprozessmodellierung mit kontrollierter natürlicher Sprache vor. Durch die direkte Verbindung zwischen Geschäftsprozessmodell und natürlichsprachlicher (nicht formaler) Beschreibung sollen Wissensträger besser in die Geschäftsprozessmodellierung einbezogen werden und darüber hinaus die Übereinstimmung zwischen Prozessbeschreibung und Prozessmodell garantiert werden.

Der Beitrag ist wie folgt strukturiert: Abschnitt 2 stellt das Lösungskonzept dar, Abschnitt 3 zählt die wichtigsten verwandten Arbeiten auf und schließt mit einem Vergleich. Abschnitt 4 gibt einen Ausblick.

2 Lösungskonzept

Der Ansatz zur Geschäftsprozessmodellierung mit kontrollierter natürlicher Sprache basiert auf der Verwendung eines LL(k)-Parsers für kontextfreie Grammatiken. Der Parser und seine nachgelagerten Analysekomponenten sind für die Verarbeitung der natürlichen Sprache und die Transformation in ein Geschäftsprozessmodell verantwortlich. Der im Rahmen des Ansatzes verwendete Parser wurde mit Hilfe von ANTLR² als LL(*)-Parsergenerator erstellt. ANTLR erwartet eine kontextfreie Grammatik in angepasster erweiterter Backus-Naur-Form. Anforderungen zur Erstellung dieser kontextfreien Grammatik waren, dass alle einfachen Kontrollflussmuster der Workflowpatterns³ integriert werden und der Ausdrucksraum erweiterbar bleiben sollte. Für die englische Sprache wurde eine entsprechende Grammatik entwickelt, die Bedingungen, Aktivitäten und den Kontrollfluss (Sequenz, Parallel Split, Synchronization, Exclusive Choice, Simple Merge) darstellt. Die Analysekomponenten des Parsers erstellen dann auf Basis des aktuellen Parserkontexts eine Vorhersage für das nächste (Nicht-) Terminalsymbol und zeigen es dem Benutzer in Form einer Auto-Vervollständigungsfunktion und als dynamische Satzschablone an.

² <http://www.antlr.org/>

³ <http://www.workflowpatterns.com/patterns/control/>

Satzschablonen dienen dazu, die natürliche Sprache einzuschränken und Formulierungen kontrolliert und bewusst zu verwenden. Mit kontrollierter natürlicher Sprache ist in diesem Zusammenhang die Einschränkung gemeint, dass nur Eingaben akzeptiert werden, die auf den erzeugten Satzschablonen basieren und somit vom Parser auch ausgewertet werden können. (Statische) Satzschablonen werden beispielsweise im Requirements Engineering verwendet, um Softwarefunktionen zu spezifizieren. Während im Requirements Engineering jede Anforderung für sich steht und diese idealerweise auch vollständig beschreibt, ist jeder Teil einer textuellen Beschreibung eines Geschäftsprozesses für sich kontextsensitiv. Die Satzschablonen, die im Rahmen der Modellierung verwendet werden, müssen daher zur Laufzeit dynamisch an den aktuellen Kontext angepasst werden. Parser und Satzschablonen basieren auf folgender Grammatik in eBNF:

```

program: sentence*;
sentence: placestart | transstart;
placestart: plstartlist placeslist ', then I do '
           transitionlist '. ';
plstartlist: 'If ' | 'As soon as ';
placeslist: place | eitherpl place orpl furtherplaces |
           place andpl furtherplaces;
furtherplaces: place | place orpl furtherplaces | place
              andpl furtherplaces;
eitherpl: 'either ';
orpl: ' or ';
andpl: ' and ';
transitionlist: transition | eithertr transition ortr
              furthertransit | transition andtr furthertransit;
furthertransit: transition | transition ortr
              furthertransit | transition andtr furthertransit;
eithertr: 'either ';
ortr: ' or ';
andtr: ' and ';
transstart: lasttrans transfollow '. ' | 'After I did '
           transitionlist ', ' transfollow '. ';
lasttrans: 'After that, ';
transfollow: plfollowlist placeslist | trfollowlist
            transitionlist;
plfollowlist: 'I have ' | 'I get ';
trfollowlist: 'I do ';
place: content;
transition: content;
content: STRING+ ( ' ' | STRING )* | content '.' content;

```


Neben der Vorhersage des nächsten (Nicht-)Terminalsymbols wird im Rahmen der entwickelten Methode der bisher eingegebene Text ausgewertet und in ein einfaches BPMN-Modell transformiert. Abbildung 1 zeigt das Parserkonzept basierend auf der kontextfreien Grammatik und des eingegebenen Textes in kontrollierter natürlicher Sprache. Zur Transformation der natürlichen Sprache in ein grafisches Prozessmodell wird von der Analysekomponente ein Graphviz⁴-konformer Text erstellt. Dieses dient der Komponente als Grundlage zur Erzeugung der Geschäftsprozessmodellgrafik.

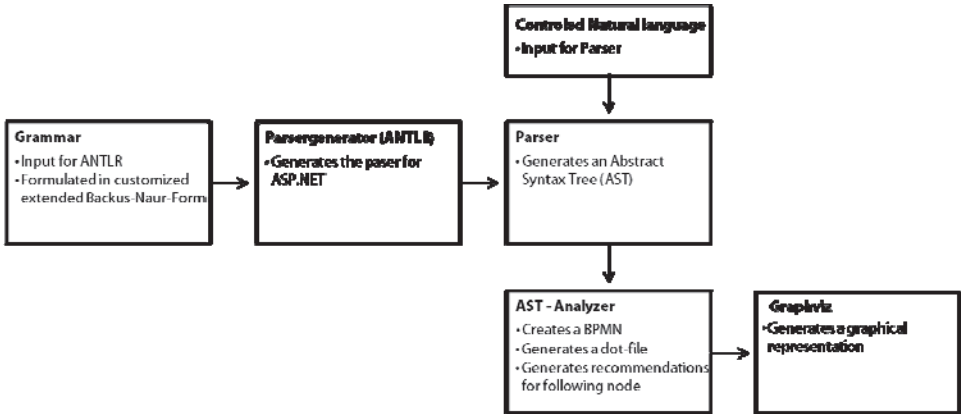


Abb. 1: Transformationskonzept

Zur Veranschaulichung der Funktionalität wurde ein Prototyp als Webapplikation entwickelt⁵ und erste Tests durchgeführt. Die Tests unter Laborbedingungen zeigen eine korrekte Auswertung und Transformation [Ca16]. Die Entwicklung des Prototyps als Webapplikation ermöglicht außerdem, Anwender schnell und einfach in die Modellierungsaktivitäten einzubinden.

3 Verwandte Arbeiten

Die existierenden verwandten Ansätze lassen sich in zwei Kategorien unterteilen: Auf der einen Seite existieren Ansätze, die natürliche Sprache vor oder bei der Modellierung analysieren. Dabei sollen Informationen der natürlichen Sprache verarbeitet werden und bei der Modellierung Berücksichtigung finden. Auf der anderen Seite existieren Ansätze, die die natürliche Sprache in den Modellen analysieren. Dabei handelt es sich meistens um eine Analyse der textuellen Beschriftungen. Ansätze der ersten Kategorie verwenden zumeist NLP-Techniken um aus bereits existierenden Beschreibungen Geschäftsprozessmodelle, ER-Modelle oder UML-Diagramme zu generieren. Alle Arbeiten haben gemeinsam, dass sie Standard-NLP-Tools wie WordNet, Tagger, Parser

⁴ <http://www.graphviz.org>

⁵ <https://bpm.caporale.eu/>

oder existierende Spracherkennungsverfahren verwenden. Für die Erzeugung von Geschäftsprozessmodellen existieren z.B. Ansätze, die mit Hilfe von Parsern Texte analysieren und daraus (halb-) automatisch BPMN Modelle ableiten z.B. [FMP11, GKC07, GSB09, SP10]. Ähnliche Ansätze existieren für Abhängigkeitsdiagramme [Ga01], Entity-Relationship-Modelle [OHA08], UML-Diagramme [DB09, MP12] und Petri-Netze [Ke14]. Einige Autoren beschäftigen sich auch mit abstrakteren Ansätzen, die unabhängig von der gewählten Modellierungssprache sind [FKM07]. Ansätze der zweiten Kategorie analysieren die Beschriftungen auf z.B. terminologische Fragestellungen, um die Gesamtqualität der Modelle zu verbessern und basieren zumeist auf WordNet [BPB09, KHO11]. Einige proprietäre Ansätze gehen über die Analyse syntaktischer Aspekte hinaus und suchen nach semantischen Fehlern [GL11] oder erstellen semantische Annotationen [BSP08]. Um eine Verbindung zwischen Modellen und der natürlichen Sprache herzustellen, realisieren weitere Ansätze die Erzeugung von natürlich sprachlichem Text für UML-Klassendiagramme [MAA08], Objektmodelle [LRR96], Geschäftsprozessmodelle [Ke14] und konzeptuelle Modelle [Da92].

Der im Rahmen dieses Beitrags vorgestellte Ansatz vermischt beide Kategorien. Auf der einen Seite wird ein Parser verwendet, der aus natürlicher Sprache ein Prozessmodell erzeugt (ähnlich zu [Ke14]). Auf der anderen Seite wird die Qualität des erstellten Prozessmodells durch die Verknüpfung zur textuellen Beschreibung sichergestellt. Im Gegensatz zu den meisten existierenden Ansätzen benutzt der Parser keine stochastischen Methoden, um die Semantik des Textes zu erfassen, sondern garantiert durch den festgelegten Eingaberaum (in Form der kontrollierten natürlichen Sprache) eine fehlerfreie und verlustfreie Transformation. Ein Nachteil ist dabei, dass existierende Beschreibungen durch die vorgestellte Technik nicht geparkt werden können.

4 Ausblick

Im Rahmen dieses Beitrags wurde eine Technik vorgestellt, welche natürliche Sprache als Eingabe zur Modellierung von Geschäftsprozessen verwendet. Die so erstellte textuelle Beschreibung repräsentiert exakt die Semantik des Geschäftsprozessmodells. Realisiert wurde die Technik durch den Entwurf einer kontextfreien Grammatik als Parsergrundlage und die Verwendung von Satzschablonen zur Kontrolle der natürlichen Sprache. Der Prototyp zeigt unter Laborbedingungen vielversprechende Resultate.

Die nächsten Schritte im Rahmen der Forschungsarbeit betreffen vor allem die Anpassung der Satzschablonen um weitere und flexiblere Satzstrukturen. Durch den Einsatz moderner Anwendungsoberflächen und die Integration weiterer benutzerunterstützender Komponenten könnte außerdem untersucht werden, inwiefern IT-affine Stakeholder eigenständig Geschäftsprozessmodelle erstellen können. Als Ausblick ist denkbar, dass durch die textuelle Eingabe neue Anwendungsfelder entstehen. Eine Anpassung der Technik zur Modellierung von Anforderungen für Softwaresysteme oder Richtlinien können Anwendungsfälle außerhalb der BPM-Domäne sein.

Literaturverzeichnis

- [BPB09] Breuker, D.; Pfeiffer, D.; Becker, J.: Reducing the Variation in Intra-and Interorganizational Business Process Modeling-An Empirical Evaluation, 2009.
- [BSP08] Bögl, A. et al.: Semantic annotation of epc models in engineering domains to facilitate an automated identification of common modelling practices: Enterprise Information Systems. Springer, 2008; S. 155–171.
- [Ca16] Caporale, T.: A tool for natural language oriented Business Process Modeling. ZEUS, 2016.
- [Da92] Dalianis, H.: A method for validating a conceptual model by natural language discourse generation. Springer, 1992.
- [DB09] Deeptimahanti, D. K.; Babar, M. A.: An automated tool for generating UML models from natural language requirements. IEEE Computer Society, 2009.
- [FKM07] Fliedl, G. et al.: Deriving static and dynamic concepts from software requirements using sophisticated tagging. In Data & Knowledge Engineering, 2007, 61; S. 433–448.
- [FMP11] Friedrich, F.; Mendling, J.; Puhlmann, F.: Process model generation from natural language text. Springer, 2011.
- [Ga01] Gangopadhyay, A.: Conceptual modeling from natural language functional specifications. In Artificial Intelligence in Engineering, 2001, 15; S. 207–218.
- [GKC07] Ghose, A.; Koliadis, G.; Chueng, A.: Process discovery from model and text artefacts. IEEE, 2007.
- [GL11] Gruhn, V.; Laue, R.: Detecting Common Errors in Event-Driven Process Chains by Label Analysis. In EMISA, 2011, 6; S. 3–15.
- [GSB09] Gonçalves, João Carlos de AR; Santoro, F. M.; Baiao, F. A.: Business process mining from group stories. IEEE, 2009.
- [Ke14] Keuter, B.: Bidirektionale Abbildung zwischen Geschäftsprozessmodellen und IT-Kommunikationssystemen. KIT Scientific Publishing, 2014.
- [KHO11] Koschmider, A.; Hornung, T.; Oberweis, A.: Recommendation-based editor for business process modeling. In Data & Knowledge Engineering, 2011, 70; S. 483–503.
- [MAA08] Meziane, F.; Athanasakis, N.; Ananiadou, S.: Generating Natural Language specifications from UML class diagrams. In Requirements Engineering, 2008, 13; S. 1–18.
- [MP12] More, P.; Phalnikar, R.: Generating UML Diagrams from Natural Language Specifications. In International Journal of Applied Information Systems (IJ AIS), Foundation of Computer Science FCS, New York, USA, 2012, 1.
- [OHA08] Omar, N.; Hassan, R.; Arshad, H.; Sahran, S.: Automation of database design through semantic analysis, 2008.
- [SP10] Sinha, A.; Paradkar, A.: Use cases to process specifications in business process modeling notation. IEEE, 2010.

Wahrnehmungsorientierte optische Inspektion von texturierten Oberflächen

Michael Grunwald¹ und Matthias O. Franz²

Abstract: Digital bedruckte Oberflächen müssen strengen funktionalen und ästhetischen Anforderungen genügen. Diese Eigenschaften werden im Rahmen der Qualitätsprüfung kontrolliert. Hierbei wirken sich Oberflächendefekte oftmals erst dann aus, wenn diese auch vom Menschen wahrgenommen werden. Aufgrund der hohen Produktionsgeschwindigkeit kann eine solche Bewertung der Sichtbarkeit von Defekten bisher nur außerhalb des Produktionsflusses durch manuelle – subjektiv geprägte – Inspektion erfolgen. Ziel des Projektes ist (1) die Modellierung von Texturen in einer Form, die an das menschliche visuelle System angepasst ist und (2) die automatisierte Beurteilung der Wahrnehmung von Texturfehlern. Im Rahmen des Projekts wurde ein prototypisches System zur Inline-Erfassung von texturierten Oberflächen entwickelt. Auf Basis von realen Aufnahmen industriell produzierter Holzdekore wurde eine repräsentative Texturdatenbank erstellt. Gezeigt werden erste Resultate im Bereich der Defektdetektion auf Basis von statistischen Merkmalen. Diese Ergebnisse dienen als Grundlage für die spätere wahrnehmungsorientierte Bewertung. Letztlich sollen die im Rahmen des Projekts erlangten Ergebnisse in einen prototypischen Aufbau zur Inspektion von digital bedruckten Dekoren einfließen.

Keywords: Texturmodell, Defektdetektion, wahrnehmungsorientierte optische Inspektion, Filterbank, Digitaldruck

1 Psychopysisches Texturmodell

Moderne digitale Drucktechnologien eröffnen den Herstellern von Dekoroberflächen völlig neuartige Möglichkeiten. In ihrer Funktion als Dekor müssen diese Oberflächen üblicherweise strengen funktionalen und ästhetischen Anforderungen genügen, die wiederum eine geeignete Inspektionstechnologie erfordern. Insbesondere sollte die Oberflächeninspektion zur Vermeidung von Übersortierung im industriellen Produktionsprozess nach Kriterien erfolgen, die der menschlichen Wahrnehmung entsprechen.

Ziel des Projektes ist die Entwicklung eines Systems zur Überwachung und Inspektion von texturierten Oberflächen unter Berücksichtigung der Wahrnehmung des Menschen. Die Basis für die Inspektion bildet ein bildgebendes System, welches innerhalb einer industriellen Produktionsumgebung die Information für eine nachfolgende menschlich inspirierte Beurteilung bereitstellen kann. Neben der Inline-Überwachung erlaubt ein solches System auch die Erfassung von Merkmalen, welche beispielsweise eine visuelle Änderung über

¹ HTWG Konstanz, Institut für Optische System, Brauneggerstr. 55, 78462 Konstanz, m.grunwald@htwg-konstanz.de

² HTWG Konstanz, Fakultät Informatik, Institut für Optische Systeme (AG Kognitive Systeme), Brauneggerstr. 55, 78462 Konstanz, mfranz@htwg-konstanz.de

die Zeit des Produktionsprozesses beschreiben. Rückgekoppelt können diese Prozessdaten zur Prozesssteuerung bzw. Optimierung genutzt werden. Zuletzt soll die Grundlage für eine visuelle Inspektion innerhalb des Produktionsprozesses gelegt werden, auch an Stellen, an denen ein Mensch aus Gründen der Sicherheit nicht inspizieren darf oder dies beispielsweise aufgrund der hohen Produktionsgeschwindigkeit nicht kann.

Die Entwicklung des Inspektionssystems erfordert die Bearbeitung folgender Forschungsfragen:

1. Um die Vielzahl der eingesetzten Dekortexturen mit einem entsprechenden Referenzobjekt vergleichen zu können und eine Aussage über die Ähnlichkeit zu treffen, müssen Texturen entsprechend modelliert bzw. in einer Form, die an das menschliche visuelle System angepasst ist, repräsentiert werden. Neben parametrisierbaren Modellen zur Beschreibung von Texturen soll die Anwendung von künstlichen neuronalen Netzen untersucht werden.
2. Die Interaktion von Textur und Farbe in der menschlichen Wahrnehmung hat in der psychophysischen Forschung bisher nur wenig Beachtung gefunden. Diese Frage spielt aber für das Projekt eine zentrale Rolle, daher sollen hier erste grundlegende Untersuchungen im Bereich der Texturwahrnehmung unternommen werden. Die Ergebnisse sollen in ein psychophysisches Modell einfließen, mit dem die Sichtbarkeit von Texturfehlern beurteilt werden kann.

2 Verwandte Arbeiten

Die statistische Modellierung von Texturen hat eine lange Tradition in der Bildverarbeitung. Dementsprechend gibt es eine Vielzahl unterschiedlicher Ansätze in der Literatur. Allgemein können statistische Texturmodelle grob in zwei Klassen unterteilt werden: (1) Probabilistische Charakterisierung der statistischen Abhängigkeiten innerhalb lokaler Pixelnachbarschaften (Markov-Felder, engl. Markov Random Fields, MRFs) und (2) Anwendung von Filterbänken und anschließende Beschreibung der Filterantworten mit statistischen Kennzahlen. Die große Mehrheit der älteren Ansätze basieren auf MRFs [DJ89, Pa99]. Obwohl hier einige Erfolge erzielt wurden, haben sich MRFs nur als bedingt geeignet für komplexere Texturen erwiesen.

2.1 Filterbänke zur Modellierung von Texturen

Neuere Ansätze verwenden mehrheitlich Filterbänke zur Gewinnung statistischer Deskriptoren für Texturen. Filterbank-Modelle sind an frühe Prozesse des visuellen Systems von Säugetieren angelehnt und eignen sich daher auch zur systematischen Charakterisierung von Texturen im Rahmen psychophysischer Untersuchungen. Die gängigen Filterbank-Modelle wurden bisher nur theoretisch bzw. auf vordefinierten Texturdatenbanken untersucht. Eine Untersuchung ihrer Anwendbarkeit im Hinblick auf industrielle Dekore und die dort anfallenden Inspektionsaufgaben liegt bisher nicht vor.

2.2 Textur in der menschlichen Wahrnehmung

Bereits 1961 Julesz seine Hypothesen [Ju62] über die Bedeutung von Bildstatistiken für das menschliche visuelle System. Arbeiten von Tyler [Ty04a, Ty04b] weisen auf eine Unterscheidbarkeit von binären Texturen mit unterschiedlichen statistischen Eigenschaften höherer Ordnung hin, und Arbeiten von Victor und Kollegen haben gezeigt, dass sich die Detektierbarkeit von lokalen Korrelationen in binären Statistiken von deren Prävalenz in natürlichen Bildern vorhersagen [He14]. Die Anwendung von Filterbänken in psychophysischen Modellen zur Texturwahrnehmung sowie zur Beurteilung der Ähnlichkeit von Texturen wurde für Fragestellungen der optischen Oberflächeninspektion bisher noch nicht im Detail untersucht, vor allem nicht, wenn einfache Filtermodelle mit dem notwendigen, psychophysisch realistischen, nicht-linearen *gain control* kombiniert wird [Go13].

3 Forschungsplan

3.1 Oberflächeninspektion mit Texturmodellen

Wie in Abschnitt 2.1 erläutert, sollen verschiedene Filterbank-Texturmodelle aus der Literatur vergleichend untersucht [Ma89, Po96, PS00] und eigene Ansätze zu Filterbänken entwickelt werden. In einem Filterbank-Texturmodell wird eine bestimmte Dekortextur durch eine typische Häufigkeitsverteilung der einzelnen Filterantworten der Filterbank charakterisiert. Auf der Basis solcher charakteristischen Merkmale sollten sich sichtbare Oberflächendefekte als signifikante Abweichungen zu den Merkmalen des Referenzdekors zeigen.

Im nächsten Schritt werden signifikante Abweichungen in den Texturmerkmalen mit Methoden des maschinellen Lernens automatisch klassifiziert. Hierfür werden sowohl herkömmliche Supportvektormaschinen als auch neuere neuronale Ansätze aus dem Bereich *Deep Learning* zum Einsatz kommen [Sc15]. Zur allgemeinen Texturklassifikation soll ein Vorabtraining von neuronalen Netzen mithilfe von unüberwachten Lernmethoden, beispielsweise mit mehrschichtigen Autoencodern [Vi10], untersucht werden.

Für die Entwicklung und Analyse wird eine entsprechende große Anzahl an Trainings- und Testdaten benötigt. Um schnell eine entsprechende Anzahl an Testdekoren zur Verfügung zu haben, soll ein Defektmodell auf Basis von industriell produzierten Dekormustern zur Generierung synthetischer Daten entwickelt werden. Mithilfe eines prototypischen Aufbaus soll der trainierte Klassifikator mittels realen Dekortexturen verifiziert werden.

3.2 Psychophysische Untersuchung der Sichtbarkeit von Texturdefekten

Neben der Grundlage für die Klassifikation dient das zu entwickelnde Texturmodell als Basis für die Bestimmung der Ähnlichkeit bzw. der Sichtbarkeit von Defekten. Ziel ist es, die durch die Maschine detektierten Defekte um eine oder mehrere Kennzahlen zu

ergänzen, die die Sichtbarkeit der Defekte für menschliche Beobachter quantifizieren. Da der Raum der möglichen Veränderungen oder Abweichungen sehr hoch-dimensional ist – Helligkeit, Farbe, Geometrie – und anzunehmen ist, dass die Sichtbarkeit von Defekten auch von der "Grundtextur" abhängt, wird es notwendig sein, zwei getrennte Ansätze bei den psychophysischen Untersuchungen parallel zu untersuchen: (1) Explorativ: d.h. mit Hilfe von explorativen Methoden wie multi-dimensionaler Skalierung (MDS) wird versucht werden, ob eine geeignete niedrig-dimensionale Repräsentation von Texturen gefunden werden kann. (2) Modellbasiert: hier werden die Filteraktivitäten von einwandfreien und defekten Stimuli in einem Modell der frühen Sehverarbeitung [Go13] verglichen. Die dadurch gewonnenen Hypothesen über die Sichtbarkeit von Defekten können gezielt genutzt werden, um das Modell im Sinne eines generativen Modells zu verwenden, d.h. neue Stimuli werden generiert, um die Hypothesen über deren Wahrnehmungseigenschaften in sich anschließenden psychophysischen Experimenten zu testen.

4 Erste Ergebnisse

Im Allgemeinen sind ästhetische Digitaldruckfehler dadurch gekennzeichnet, dass an fehlerhaften Positionen keine oder unerwünscht Farbe aufgebracht wurde. Experimente mit industriell produzierten Dekoren zeigen, dass der Kontrast eines Fehlers – bezogen auf die Nachbarschaft innerhalb eines lokalen Bereiches und in Abhängigkeit von der Druckfarbe – mittels spektral schmalbandiger Beleuchtung im visuellen Wellenlängenbereich maximiert werden kann. Auf Basis dieser Voruntersuchung wurde eine geblitzte Beleuchtung für das prototypische Aufnahmesystem entwickelt. Für die Erfassung wird ein Verbund aus monochromatischen Kamerasystemen verwendet. Das System wurde mit dem Ziel der Integration in eine Digitaldrucklinie konzipiert und erlaubt die Erfassung von Objekten im Durchlauf. Hierbei wird jede Objektposition unter drei verschiedenen Beleuchtungsbedingungen aufgenommen. Die Auswertung dieser Farbkanäle bzw. deren Kombination ermöglicht die Zuordnung eines detektierten Fehlers zu einer Druckfarbe. Zudem werden die spektralen Kanäle zur Rekonstruktion eines Farbbildes verwendet, vgl. Abb. 1.

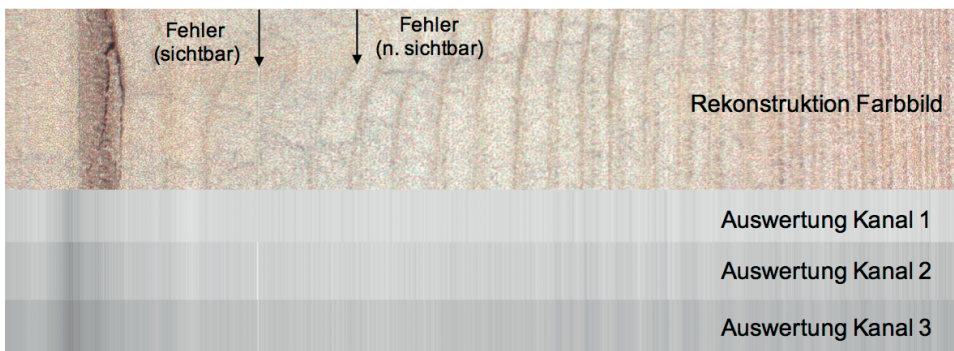


Abb. 1: Düsenfehler – typischer, im Single-Pass-Verfahren auftretender Fehlertyp.

4.1 Textur-Datenbank NWDDS

Um Filterbank-Texturmodelle vergleichend untersuchen und deren Anwendbarkeit auf industrielle Dekore überprüfen zu können, wurde initial eine repräsentative Dekordatenbank (NWDDS) erstellt. Als Basis für den Datensatz dienen Aufnahmen industriell produzierter Dekore. Für die Erfassung wurde das entwickelte Inline-Aufnahmesystem verwendet. Verteilt auf 40 Dekorklassen wurden 800 fehlerfreie Dekorausschnitte³ und 800 Ausschnitte mit markierten Defekten für den Datensatz ausgewählt. Um die Anzahl an Mustern pro Dekorklasse zu erhöhen und später die maximale Variabilität von Texturen einer Klasse untersuchen zu können, wurden pro Klassen weitere 20 Texturen synthetisch generiert (Texturanalyse und -synthese Methode nach [PS00]).

4.2 Filterbank-Texturmodell

Um Oberflächendefekte durch signifikante Abweichungen in Merkmalen erkennen zu können, wurden unterschiedliche Häufigkeitsverteilungen und statistische Kennzahlen untersucht. Als Basis diene hierfür die Veröffentlichung *A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients* von J. Portilla und E. P. Simoncelli, die die Beschreibung von charakteristischen Texturmerkmalen mittels 710 statistischen Parametern zeigt. Im Rahmen der Versuchsreihe „Synthese durch Analyse“ wurde die Anwendbarkeit dieses Modells für die industriellen produzierten Holzdekore des NWDDS Datensatzes untersucht. Wie in Abb. 2a & 2b gezeigt, kann eine visuelle Näherung nach wenigen Iterationen im Syntheseschritt erzielt werden.

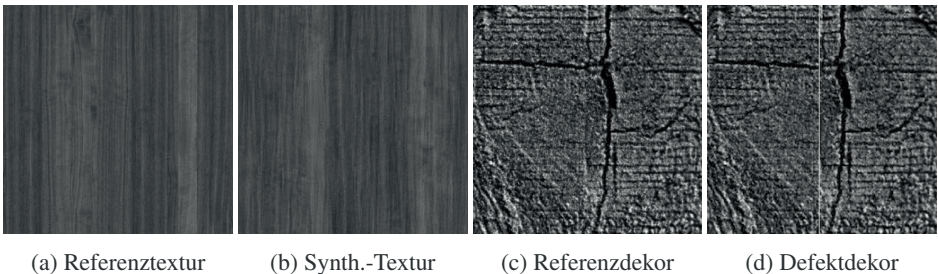


Abb. 2: Auf Basis der in (a) dargestellten Referenztextur wird das Ergebnis der angewandten Synthese durch Analyse Methode nach 12 Iterationsschritten in (b) visualisiert. In (c) bzw. (d) wird beispielhaft ein synthetisch generiertes Defektdekor, welches zur Charakterisierung von Defekten verwendet wurde, dargestellt.

Mit dem Ziel, den Einfluss von Defekten auf die Charakteristik einer gegebenen Referenztextur zu untersuchen, wurden synthetische Defektbilder mit vertikalen linienförmigen Defekten von unterschiedlicher Breite generiert, vgl. Abb. 2c & 2d. Die natürliche Varianz aus Produktionsprozess und Aufnahmesystem wurde durch additives Gaußsches Rauschen

³ Unter Dekorausschnitt werden die drei Aufnahmen der Einzelkanäle sowie das rekonstruierte Farbbild verstanden. Fehlerfrei bedeutet, dass die Aufnahmen keine erkennbaren Abweichungen zur jeweiligen digitalen Referenz aufweisen.

simuliert. Erste Resultate zeigen, dass die Anzahl der benötigten statistischen Merkmale zur Defekterkennung texturabhängig reduziert werden kann. Von besonderer Bedeutung ist hierbei die Defektregion, d.h. die räumliche Defektposition einschließlich deren Umgebung.

Literaturverzeichnis

- [DJ89] Dubes, R. C.; Jain, A. K.: Random field models in image analysis. *Journal of Applied Statistics*, 16(2):131–164, 1989.
- [Go13] Goris, R. L. T.; Putzeys, T.; Wagemans, J.; Wichmann, F. A.: A neural population model for pattern detection. *Psychological Review*, 120(3):472–496, 2013.
- [He14] Hermundstad, A. M.; Briguglio, J. J.; Conte, M. M.; Victor, J. D.; Balasubramanian, V.; Tkačik, G.: Variance predicts salience in central sensory processing. *eLife*, 3:1–28, 2014.
- [Ju62] Julesz, B.: Visual pattern discrimination. *Information Theory, IRE Transactions on*, 49:41–46, 1962.
- [Ma89] Mallat, S. G.: Theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.
- [Pa99] Paget, R. D.: Nonparametric Markov Random Field Models for Natural Texture Images. Dissertation, The University of Queensland, 1999.
- [Po96] Portilla, J.: Texture synthesis-by-analysis method based on a multiscale early-vision model. *Optical Engineering*, 35(8):2403, 1996.
- [PS00] Portilla, J; Simoncelli, E P: Aparametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40(1):49–71, 2000.
- [Sc15] Schmidhuber, J: Deep Learning in Neural Networks: An Overview. *Neural Networks*, 61:85–117, 2015.
- [Ty04a] Tyler, C. W.: Beyond fourth-order texture discrimination: generation of extreme-order and statistically-balanced textures. *Vision Research*, 44(18):2187–2199, 2004.
- [Ty04b] Tyler, C. W.: Theory of texture discrimination of based on higher-order perturbations in individual texture samples. *Vision Research*, 44(18):2179–2186, 2004.
- [Vi10] Vincent, P.; Larochele, H.; Lajoie, I.; Bengio, Y.; Manzagol, P.-A.: Stacked denoising auto-encoders: learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11(3):3371–3408, 2010.

Gemeinsame mentale Modelle in der agilen Softwareentwicklung: Ein Ansatz zur Erstellung von Gestaltungsempfehlungen für „gute“ erfahrungsspezifische User Stories

Daniel Hallmann¹, Ute Schmid² und Rüdiger von der Weth³

Abstract: User Stories sind ein probates Mittel zum Austausch von Anforderungen in der agilen Softwareentwicklung. Die Anwendung ist jedoch kritisch, da sich immer wieder Fehler in der Beschreibung einzelner Formularfelder wie den Akzeptanzkriterien einschleichen. Mit diesen fehlerhaften User Stories müssen Personen mit individuellen Erfahrungen ein gemeinsames mentales Modell konstruieren. Häufig ist der Aufbau jedoch nicht möglich und es kommt zu negativen Reaktionen wie Frustration bei fehlenden Erfahrungen, langen Diskussionen bei unterschiedlichen Vorstellungen und teuren Nacharbeiten bei missverstandenen Inhalten. Aufgrund dieser Effekte sollen mit einer Ex-post-Analyse, einer Feldstudie und einem Experiment die Zusammenhänge von fehlerhaften User Stories, individuellen Erfahrungen und gemeinsamen mentalen Modellen untersucht werden. Mit dieser Sicht auf die Verarbeitung von User Stories wird das Forschungsinteresse für kognitionspsychologische Prozesse in der agilen Softwareentwicklung unterstützt. Außerdem liefert die Arbeit für die Praxis Gestaltungsempfehlungen für „gute“ erfahrungsspezifische User Stories. Mit dieser Strategie sollen frühzeitig die Kosten in der agilen Softwareentwicklung reduziert werden.

Keywords: Agile Softwareentwicklung, Qualität von User Stories, Kognitive Psychologie, Erfahrungen, Gemeinsame Mentale Modelle

1 Einleitung

Für den Austausch von Anforderungen zwischen einem Product Owner, dem Interessenvertreter des Kunden und den Entwicklern nehmen User Stories in der agilen Softwareentwicklung eine zentrale Position ein. Bei der Erstellung durch den Product Owner schleichen sich jedoch immer wieder Fehler ein. Danach werden die *Felder* im Formular: *Titel - Als Rolle möchte ich Ziel, um Nutzen - Akzeptanzkriterien* [Co04] entweder (i) falsch ausgefüllt oder (ii) gar nicht angegeben [Gi10, Ka11, Kl13, NS10, Vi11, Wa12] Neben einer falschen Beschreibung von Titel, Rolle und Ziel fehlen bei einigen User Stories gänzlich die Angabe des Nutzens und der Akzeptanzkriterien. Außerdem ist es in der agilen Softwareentwicklung nicht unüblich, dass Entwickler mit unterschiedli-

¹ Otto-Friedrich-Universität Bamberg, Lehrstuhl für Kognitive Systeme, An der Weberei 5, 96047 Bamberg, daniel.hallmann@uni-bamberg.de

² Otto-Friedrich-Universität Bamberg, Lehrstuhl für Kognitive Systeme, An der Weberei 5, 96047 Bamberg, ute.schmid@uni-bamberg.de

³ Hochschule für Technik und Wirtschaft Dresden, Professur Betriebswirtschaftslehre / Personalwirtschaft und Arbeitswissenschaften, Friedrich-List-Platz 1, 01069 Dresden, weth@htw-dresden.de

chen technischen und kundenspezifischen Erfahrungen zusammenarbeiten [Bo81, Es07]. Die Kenntnisse sind dabei abhängig vom persönlichen Werdegang und variieren bei Fragen zu Entwicklungswerkzeugen oder zum Fachbereich des Kunden.

Trotz dieser mangelhaften Qualität und individuellen Erfahrungen der Entwickler müssen sich der Product Owner und die Programmierer auf ein gemeinsames Bild der Anforderung festlegen. Häufig ist eine Einigung problematisch, da bei fehlerhaften User Stories beispielsweise neue Entwickler in Softwareprojekten mit Unverständnis reagieren. Aussagen wie:

„Damit kann ich nichts anfangen“ (O-Ton eines Entwicklers)

zeigen den Unmut bezüglich verminderter Qualität sowie unzureichender Erfahrungen. Hinzu kommen möglicherweise teure Nacharbeiten bei missverstandenen Inhalten und lange Diskussionen bei ungleichen Vorstellungen. Qualitativ schlechte User Stories und falsch eingeschätzte Erfahrungen können damit nicht nur Frustration hervorrufen, sondern sich auch erheblich auf die Wirtschaftlichkeit des Softwareprojektes für den Kunden durch einen erhöhten Korrektur- und Abstimmungsaufwand auswirken.

Ziel der Arbeit ist es die Zusammenhänge zu verstehen, die zwischen fehlerhaften User Stories, individuellen Erfahrungen sowie dem gemeinsamen Bild der Anforderung bestehen. Die Auswertung der Ergebnisse soll zur Erstellung von Gestaltungsempfehlungen für „gute“ erfahrungsspezifische User Stories genutzt werden. Dazu wird im Folgenden die Literatur erörtert, Punkt 3 beschreibt die Ziele, den Inhalt und die Voraussetzungen von Gestaltungsempfehlungen, Punkt 4 erläutert den Forschungsplan, Punkt 5 geht auf den Stand der Arbeit ein und Punkt 6 stellt den wissenschaftlichen und praktischen Nutzen vor.

2 Bisherige Arbeiten

Empirische Untersuchungen zu User Stories im Bereich der agilen Softwareentwicklung konzentrieren sich auf die Analyse von Praktiken und Problemstellungen zu Anforderungsdokumenten in Projektteams [dAO05, SRP09, RCB10, SK13]. Parallel werden in theoretischen Abhandlungen neue Techniken für die Verwendung von User Stories ausgearbeitet [MBA09, CD13].

Sowohl die empirischen wie auch die theoretischen Untersuchungen betrachten User Stories auf einer organisatorischen und technischen Ebene. Auf den kognitionspsychologischen Verstehensprozess zwischen dem Product Owner und den Entwicklern wird dagegen nicht eingegangen. Unbeachtet bleibt der Prozess zum Aufbau eines gemeinsamen mentalen Modelles [CCS93] zu der User Story, mit deren Hilfe die Personen eine kollektive Beschreibung und Erwartungshaltung der Aufgabe formulieren. Außerdem fehlen Erkenntnisse zu Einflussgrößen auf den Prozessverlauf. Danach ist die beiderseitige Wirkung von mangelhaften Formularfeldern in der User Story wie beispielsweise

fehlerhafte Akzeptanzkriterien sowie individuelle Erfahrungen wie etwa technische Programmierkenntnisse und kundenspezifische Produktionsabläufe auf das gemeinsame mentale Modell unbekannt. Um diese Forschungslücke zu schließen, wird daher folgende Forschungsfrage gestellt:

Wie beeinflussen:

- (i) die *Qualität* einer User Story US_1 und
- (ii) individuelle *Erfahrungen* der Entwickler
- (iii) das *gemeinsame mentale Modell* von US_1 zwischen dem Product Owner und den Entwicklern?

3 Idee

Nach dem ersten Lesen einer User Story durch die Entwickler sollte sich zwischen dem Product Owner und den Entwicklern ein gemeinsames Bild zum Inhalt der Anforderung einstellen. Ein falsch ausgefülltes Formular und unbeachtete Erfahrungen bei den Entwicklern dürfen nicht mehr zu Frustration, einer differenzierten Vorstellung der Aufgabe und weiterführend zu teuren Nacharbeiten und zeitraubenden Diskussionen führen. Vielmehr soll ein Product Owner, als Autor von User Stories, aktiv den gemeinsamen Verstehensprozess zum Inhalt einer User Story positiv beeinflussen können. Dazu braucht es Hilfestellungen für einen gezielten Eingriff in den Vorgang mit Aussicht auf eine Verbesserung.

Ein Vorgehen kann hier die Erstellung von Gestaltungsempfehlungen für „gute“ erfahrungsspezifische User Stories sein. Die Hinweise sollen sich dabei an der Wirkung von Fehlern und individuellen Erfahrungen auf das gemeinsame mentale Modell orientieren. Die Idee beinhaltet zwei Gestaltungsdimensionen: (i) Eine qualitative Stabilisierung berücksichtigt die Einhaltung der formalen Kriterien. Alle Formularfelder sollen „vorhanden“, „korrekt“ und „aktuell“ geschrieben sein. (ii) Eine erfahrungsspezifische Individualisierung umfasst die Anpassung des Detailgrads der User Story anhand vorhandener Erfahrungen. Ziel ist die Erstellung einer optimierten User Story, bei der die Entwickler nicht mehr durch Fehler irritiert und mit dem Inhalt über- oder unterfordert sind. Qualitativ stabilisierte und individualisierte User Stories sind damit eine wichtige Voraussetzung für ein gemeinsames mentales Modell zwischen dem Product Owner und den Entwicklern.

Für den Aufbau der Empfehlungen müssen der Inhalt und die Struktur sowie die Zusammenhänge der Fehler und die individuellen Erfahrungen bekannt sein. Dazu gehören eine fertige Kategorisierung der Problemstellen in den Formularfeldern und ein Konzept zu unterschiedlichen Ausprägungen von Erfahrungen der Entwickler sowie eine inhaltliche Klärung der wechselseitigen Beziehungen. Der Detailbetrachtung von fehlerhaften User Stories und Erfahrungen von Entwicklern mit beiderseitiger Wirkung muss eine Sicht auf die Effekte von Fehlern in User Stories und individuellen Erfahrungen auf das gemeinsame mentale Modell folgen.

4 Forschungsplan

Ein Strukturgleichungsmodell [BS10] formt die Basis für das Forschungsmodell. Es beschreibt die Beziehungen zwischen den latenten und nicht direkt beobachtbaren Variablen Qualität der User Story, Erfahrungen der Entwickler und gemeinsames mentales Modell. Im Strukturgleichungsmodell werden auch die Beziehungen der latenten Variablen zu den messbaren Indikatoren definiert. Folgende Schritte zeigen den Weg für die Bearbeitung der Forschungsfrage anhand des Strukturgleichungsmodells. Eine Ex-post-Analyse, eine Feldstudie und ein Experiment bilden den empirischen Kern der Arbeit.

- Themenfindung, Aufstellen der Forschungsfrage und Überführung der Fragestellung in ein Strukturgleichungsmodell im 1. Jahr.
- Indikatorbildung für das Strukturgleichungsmodell auf Basis einer Ex-post-Analyse und einer Studie der COCOMO-Ansätze [Bo81, BR87, Bo95] im 2. Jahr.
- Erhebung von empirischen Daten für die Modellauswertung in einer Feldstudie in Softwareprojekten [LWW01] im 3. Jahr.
- Betrachtung des Strukturgleichungsmodells unter der Kontrolle von Störgrößen über ein Experiment [HSE09] im 4. Jahr.
- Konzeption der Gestaltungsempfehlungen für User Stories und Abgabe der Dissertationsschrift im 5. Jahr.

5 Stand der Arbeit

Die Forschungsarbeit befindet sich im 2. Jahr. Abgeschlossen sind die Themenfindung, die Formulierung der Forschungsfrage und die Übersetzung der Forschungsfrage in ein Strukturgleichungsmodell. Das im Strukturgleichungsmodell enthaltene Strukturmodell besteht aus den latenten Variablen Qualität der User Stories, Erfahrungen der Entwickler und dem gemeinsamen mentalen Modell zwischen Product Owner und Entwicklern.

Aktuell in Bearbeitung befindet sich die Konzeption des Messmodells für das Strukturgleichungsmodell. Dazu werden manifeste Indikatoren für die latenten Variablen bestimmt. Geplant ist die Durchführung einer Ex-post-Analyse auf Basis von abgeschlossenen User Stories. Ergänzt wird die Operationalisierung der latenten Variablen, um eine Studie mit Schwerpunkt auf die COCOMO-Modelle [Bo81, BR87, Bo95] im Licht der agilen Softwareentwicklung.

Es folgen Arbeiten zur Auswertung des Strukturgleichungsmodells. Das Vorgehen umfasst die Erhebung der messbaren Indikatoren in einer Feldstudie in unterschiedlichen Softwareprojekten. Möglich wird damit die Berechnung der Korrelationen im Strukturgleichungsmodell. Den Abschluss bildet eine experimentelle Phase. Der Schritt dient der objektiven Überprüfung der Kausalbeziehungen unter Kontrolle von Störvariablen.

6 Nutzen

Die Evaluation der Ergebnisse präsentiert wissenschaftliche Erkenntnisse über Zusammenhänge von fehlerhaften User Stories, individuellen Erfahrungen und gemeinsamen mentalen Modellen. Mit dieser Sicht auf die Verarbeitung von User Stories wird das Forschungsinteresse für kognitionspsychologische Prozesse in der agilen Softwareentwicklung [SRP09] unterstützt.

Unter Einhaltung der Gestaltungsanweisungen in der Praxis ergeben sich motivationale Verbesserungen und zeitliche sowie finanzielle Einsparungen für die agile Softwareentwicklung. Bei der Vermeidung von Irritation durch Fehler und der Abschaffung von Unter- und Überforderung wird eine motivierte Arbeitseinstellung bei Entwicklern gefördert. Außerdem wirken die Anweisungen gegen unbemerkte Abweichungen im gemeinsamen Bild zur User Story im Projektteam, was eine Senkung von teuren Nacharbeiten bereits in einer frühen Phase des Entwicklungsprozesses ermöglicht. Hinzu kommt mit der Reduzierung von langen Diskussionen eine Einsparung von Entwicklungszeit, was sich ebenfalls positiv auf die Wirtschaftlichkeit auswirkt.

Literaturverzeichnis

- [Bo81] Boehm, B. W.: Software Engineering Economics. Prentice-Hall, N.J., 1981.
- [Bo95] Boehm, B. et al.: Cost Models for Future Software Life Cycle Processes: COCOMO 2.0. In *Annals of Software Engineering*, 1995, 1; S. 57–94.
- [BR87] Boehm, B.; Royce, W.: Ada COCOMO and the Ada Process Model. In *Proceedings. Third COCOMO Users Group Meeting*, SEI, 1987; S. 1–34.
- [BS10] Bortz, J.; Schuster, C.: Statistik für Human- und Sozialwissenschaftler. Springer Berlin Heidelberg, Heidelberg, 2010.
- [CCS93] Converse, S.; Cannon-Bowers, J. A.; Salas, E.: Shared Mental Models in Expert Team Decision Making. In *Individual and Group Decision Making: Current Issues*, 1993; S. 221–246.
- [CD13] Chugh, N.; Dev Mishra, A.: Assimilation of Four Layered Approach to NFR in Agile Requirement Engineering. In *International Journal of Computer Applications*, 2013, 78; S. 25–28.
- [Co04] Cohn, M.: User Stories Applied: For Agile Software Development. Addison-Wesley Professional, 2004.
- [dAO05] de Souza, Sergio Cozzetti B; Anquetil, N.; Oliveira, K. M. de: A Study of the Documentation Essential to Software Maintenance: Proceedings of the 23rd Annual International Conference on Design of Communication: Documenting & Designing for Pervasive Information. ACM, New York, NY, USA, 2005; S. 68–75.
- [Es07] Espinosa, J. A. et al.: Familiarity, Complexity, and Team Performance in Geographically Distributed Software Development. In *Organization Science*, 2007, S. 613–630.

- [Gi10] Gill, W.: Better Scrum User Stories: Save the Solution for the Spec. <http://williamgill.de/2010/12/16/better-scrum-user-stories-save-the-solution-for-the-spec/>, 14.04.2016.
- [HSE09] Hussy, W.; Schreier, M.; Echterhoff, G.: *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor*. Springer-Verlag, 2009.
- [Ka11] Kaczor, K.: 5 Common Mistakes We Make Writing User Stories. <http://www.scrumalliance.org/community/articles/2011/august/5-common-mistakes-we-make-writing-user-stories>, 14.04.2016.
- [Kl13] Klement, A.: The Problem with User Stories and What's Better. <http://alanklement.blogspot.de/2013/03/the-problem-with-user-stories-and-whats.html>, 14.04.2016.
- [LWW01] Levesque, L. L.; Wilson, J. M.; Wholey, D. R.: Cognitive Divergence and Shared Mental Models in Software Development Project Teams. In *Journal of Organizational Behavior*, 2001, 22; S. 135–144.
- [MBA09] Miranda, E.; Bourque, P.; Abran, A.: Sizing User Stories Using Paired Comparisons. In *Information and Software Technology*, 2009, 51; S. 1327–1337.
- [NS10] Nazzaro, W.; Suscheck, C.: New to User Stories? <http://www.scrumalliance.org/community/articles/2010/april/new-to-user-stories>, 14.04.2016.
- [RCB10] Ramesh, B.; Cao, L.; Baskerville, R.: Agile Requirements Engineering Practices and Challenges: An Empirical Study. In *Information Systems Journal*, 2010, S. 449–480.
- [SK13] Stettina, C. J.; Kroon, E.: Is There an Agile Handover? An Empirical Study of Documentation and Project Handover Practices Across Agile Software Teams: 19th ICE & IEEEITMC International Conference, The Hague, Netherlands, 2013; S. 1–12.
- [SRP09] Sharp, H.; Robinson, H.; Petre, M.: The Role of Physical Artefacts in Agile Software Development: Two Complementary Perspectives. In *Interacting with Computers*, 2009, 21; S. 108–116.
- [Vi11] Vin: User Stories: Focus on the Problem Not the Solution. <http://brainslink.com/2011/02/user-stories-must-focus-on-the-problem-not-the-solution/>, 14.04.2016.
- [Wa12] Wake, B.: “As a Developer...” Is Not a User Story. <http://www.industriallogic.com/blog/as-a-developer-is-not-a-user-story/>, 14.04.2016.

Virtualisierung rekonfigurierbarer Hardware zur Steigerung der Rechenleistung und Sicherheit in einer flexiblen Cloud-Architektur

Oliver Knodel¹

Abstract: Rechenleistung und Skalierbarkeit sind essentielle Bestandteile moderner Rechenzentren. Spezielle rekonfigurierbare Hardwarebeschleuniger wie FPGAs (Field Programmable Gate Arrays) stellen eine Möglichkeit dar, um sowohl Rechenleistung, als auch die Energieeffizienz zu steigern, da die Architektur direkt an die Problemstellung angepasst werden kann. Insbesondere bei der Hintergrundbeschleunigung von Cloud-Anwendungen können durch den Einsatz von FPGAs die Betriebskosten eines Rechenzentrums deutlich verringert werden. Eine weitere Einsatzmöglichkeit ist die Erhöhung der Sicherheit von Anwendungen in der Cloud. Dieser Beitrag erläutert, wie rekonfigurierbare Hardware flexibel in eine Cloud-Umgebung eingebettet werden kann. Die Zielstellung besteht darin, die *Ressource FPGA* einer breiten Nutzergruppe für unterschiedlichste Services bereitzustellen. Neben der Verwaltung von Ressourcen wird insbesondere die Möglichkeit einer Virtualisierung der Hardware diskutiert.

Keywords: Rekonfigurierbare Hardware, Cloud Computing, Virtualisierung, Ressourcenverwaltung

1 Motivation und Problemstellung

Das Konzept des Cloud Computing basiert darauf, dass Nutzer Zugang zu gemeinsamen Ressourcen oder Diensten erhalten, die nach Belieben allokiert und wieder freigegeben werden können. Eine möglichst geringe Interaktion mit dem Anbieter der Cloud oder dem Betreiber des Rechenzentrums stellt dabei ein wesentliches Kriterium dar [MG11]. Dieses Konzept der Elastizität in der Cloud ist ein essentieller Unterschied zu traditionellen Rechenzentren, bei denen der Nutzer eine feste Anzahl von Ressourcen mietet.

Ein Hauptproblem bei großen Rechenzentren, welche Clouds bereitstellen, besteht in ihrem enormen Energieverbrauch. Eine deutliche Steigerung der Rechenleistung bei zusätzlicher Einsparung von Energie kann durch heterogene Systeme erreicht werden, in denen rechenintensive Aufgaben auf spezielle Koprozessoren oder Hardwarebeschleuniger wie FPGAs ausgelagert werden [Pu14]. Eine tiefgreifende und flexible Integration von FPGAs in ein skalierbares Rechenzentrum, welches die Cloud Charakteristiken *on demand self service, broad network access, resource pooling, rapid elasticity* und *measured service* erfüllt [MG11], bildet daher ein Forschungsfeld, welches an Bedeutung gewinnt. Um eine solche Integration zu erreichen ist neben einer Ressourcenverwaltung für spezielle Hardwarekomponenten eine Virtualisierung des FPGAs selbst unumgänglich.

¹ Technische Universität Dresden, Fakultät Informatik, 01069 Dresden, oliver.knodel@tu-dresden.de

Im Folgenden wird in Abschnitt 2 ein Überblick zur existierenden Literatur und verwandten Arbeiten gegeben. Danach werden in Abschnitt 3 zunächst Sichtweisen sowie mögliche Nutzer einer Cloud mit integrierten FPGAs diskutiert, bevor ein Entwurf einer flexiblen Cloud-Architektur auf Hardware- und Softwarebasis aufgezeigt wird. Hierbei spielen sowohl die Möglichkeit eines sicheren Zuganges zur Cloud über FPGAs, als auch eine Virtualisierung der Hardware selbst zur Erhöhung ihrer Auslastung eine wichtige Rolle. Nach diesem kurzen Abriss des Forschungsgegenstandes werden in Abschnitt 4 der aktuelle Stand der Arbeit, sowie erste Ergebnisse und ein Ausblick präsentiert.

2 Literaturüberblick und verwandte Arbeiten

In Cloud-Systemen werden FPGAs derzeit hauptsächlich zur Hintergrundbeschleunigung von Kernanwendungen statisch eingesetzt. Beispiele sind die Anonymisierung von eingehenden Nutzeranfragen, bevor diese an die eigentlichen Cloud-Dienste weitergereicht werden [EV12]. Ebenso ist die Erhöhung der Sicherheit bei Datenbanken und sicherheitskritischen Teilanwendungen Gegenstand der Forschung [Mo11]. Beiträge mit einem stärkeren Fokus auf austauschbaren rekonfigurierbaren Rechenkernen und einer flexiblen Bereitstellung der Ressource für eine breite Nutzergruppe gewinnen aber an Bedeutung [Fa15]. Neben einer Steigerung der Rechenleistung kann insbesondere auch die Sicherheit durch den Einsatz rekonfigurierbarer Hardware erhöht werden [EV12].

Entscheidend dabei ist der Begriff der *Virtualisierung*, welcher im Bereich der rekonfigurierbaren Hardware in unterschiedlichsten Zusammenhängen genutzt wird. Eine tiefgreifende Virtualisierung stellt pvFPGA [Wa13] mit einer Integration eines FPGAs in eine paravirtualisierte Xen-Umgebung dar. Ein Framework, welches rekonfigurierbare Hardware direkt in eine Cloud-Architektur integriert, wird in der Arbeit von Byma et. al [By14] vorgestellt. Das System ermöglicht die Beschleunigung von Anwendungen mit nutzerspezifischen Rechenkernen. Ein Ansatz mit mehreren unterschiedlichen Rechenkernen auf demselben physischen FPGA wird von Fahmy et. al [Fa15] und Weerasinghe et. al [We15] vorgestellt.

In der Literatur sind ebenso Beiträge zu finden, welche sich mit dem Speichern und dem Wiederherstellen des Kontextes eines Hardwaredesigns auf einem FPGA beschäftigen. In [Jo10] stellen Jozwick et al. ein System vor, welches die Konfiguration der internen Registerzustände aus einem FPGA ausliest und diesen Kontext zu einem späteren Zeitpunkt wiederherstellt, wobei jedoch zusätzliche Logik erforderlich ist. Happe et al. [Ha15] zeigen in ihrer Arbeit einen Kontextwechsel ohne zusätzliche Logik über das Auslesen und Modifizieren der Konfigurationsdatei.

3 Einsatzmöglichkeiten rekonfigurierbarer Hardware in einer Cloud

Das Ziel der Arbeit besteht darin, aufbauend auf der aktuellen Forschung (Abschnitt 2) die Einsatzmöglichkeiten rekonfigurierbare Hardwarekomponenten in Cloud-Architekturen zu untersuchen. Die grundlegenden Forschungsfragen dabei sind:

- Einsatzmöglichkeiten für rekonfigurierbare Hardware und deren Verwaltung in einer universellen Cloud-Architektur.
- Hintergrundbeschleunigung von Anwendungen und optimale Auslastung durch Virtualisierung der rekonfigurierbaren Hardware.
- Steigerung der Sicherheit von Daten und dem Zugang zur Cloud durch homomorphe Verschlüsselung [Ge09].

Die Bereitstellung von rekonfigurierbarer Hardware in einer Cloud oder einem Mehrbenutzersystem erfordert zunächst die in Abschnitt 3.1 beschriebene Analyse der unterschiedlichen Nutzer und die Festlegung auf entsprechende Servicemodelle, bevor Architektur und Komponenten in Abschnitt 3.2 aufgezeigt werden. Schließlich wird in Abschnitt 3.3 eine mögliche Virtualisierung erläutert, um den FPGA in einen *virtualisierten* FPGA (vFPGA) zu überführen.

3.1 Nutzer und Interaktionsmöglichkeiten

Der Einsatz von FPGAs zur Beschleunigung von Anwendungen und Services in Rechenzentren ist aufgrund der komplexen Integration und Bereitstellung die Ausnahme. In den meisten Fällen werden mittels FPGAs Dienste in einer Cloud beschleunigt, welche direkt vom Eigentümer des Rechenzentrums bereitgestellt werden, wie beispielsweise die Suchmaschine *Bing* von Microsoft [Pu14]. Hier ist ein vollständiger Zugriff auf die Ressource FPGA aufgrund des Zugangs zum Gesamtsystem möglich. Sollen jedoch Dienste von Anbietern beschleunigt werden, die selbst nur Nutzer einer Cloud-Infrastruktur sind, wie in Abb. 1(a) gezeigt, ist eine Virtualisierung der FPGAs erforderlich.

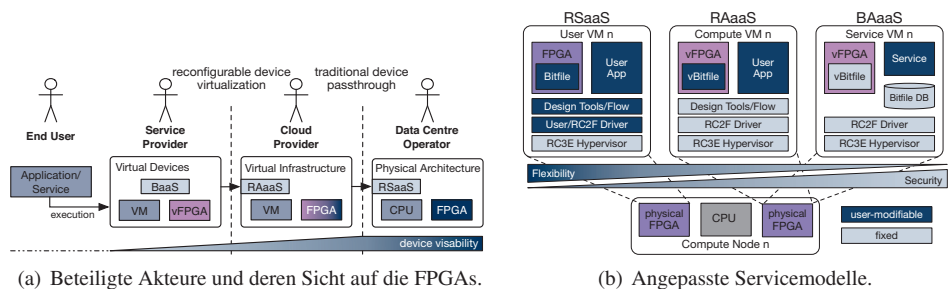


Abb. 1: Akteure und Interessengruppen sowie angepasste Servicemodelle in einer FPGA-Cloud.

Neben der Virtualisierung und der Bereitstellung eines Zugangs sind aber auch entsprechende Servicemodelle und eine flexible Einbettung von FPGAs in eine Cloud-Architektur erforderlich. Eine Ressourcenverwaltung, welche die Rechenlast auf die unterschiedlichen Knoten verteilt und den Zugang zu virtuellen FPGA-Ressourcen bereitstellt, ist das **Reconfigurable Common Cloud Computing Environment – RC3E**, welches im Rahmen der Promotion entstanden ist.

Die Ressourcenverwaltung basiert auf drei Servicemodellen, welche vom vollständigen Zugriff auf einen physischen FPGA bis hin zur Hintergrundbeschleunigung auf vFPGAs unterschiedlichste Stufen von Sicherheit und Flexibilität bieten (siehe Abb. 1(b)). Die Modelle sind im Einzelnen:

Reconfigurable Silicon as a Service – RSaaS: Voller Zugriff auf den gesamten physischen FPGA mit selbst definiertem Entwurfsablauf und sämtlichen Schnittstellen.

Reconfigurable Accelerators as a Service – RAaaS: Generieren einer eigenen Konfiguration auf einem vFPGA. Die Schnittstellen nach außen und der Entwurfsablauf sind vordefiniert.

Background Acceleration as a Service – BAaaS: Nur vordefinierte Konfigurationen auf einem vFPGA sind möglich, sodass eine sichere Hintergrundbeschleunigung von Anwendungen und Services umgesetzt werden kann.

Die FPGA-spezifischen Komponenten, wie das Hardwaredesign, welches die vFPGAs (siehe Abb. 2(c)) und die Softwareschnittstellen bereitstellt, bilden das ebenfalls während der Arbeit entstandene **Reconfigurable Cloud Computing Framework – RC2F**. Das Framework unterstützt die Möglichkeit der Virtualisierung von FPGAs, welche in Abschnitt 3.3 näher erläutert wird. Ergänzt wird das Virtualisierungskonzept durch die Nutzung des FPGAs als sicheren Zugang zur Cloud mittels des **Reconfigurable Secured Cloud Access – RSCA** in Abschnitt 3.4.

3.2 Entwurf einer flexiblen Cloud-Architektur mit sicherem Zugang

Eine Architektur, welche den unterschiedlichen Modellen gerecht wird, und somit unterschiedlichste Möglichkeiten der Interaktion mit der rekonfigurierbaren Hardware ermöglicht, bildet die Grundlage des RC3E. Des Weiteren ist der direkte Zugang zur Hardware ein wesentliches Kriterium, um einerseits die Rechenlast verteilen zu können und andererseits die Möglichkeit eines direkten Zugriffes in Form einer verschlüsselten Verbindung herzustellen.

Der Zugriff auf die Cloud wird über einen Managementknoten ermöglicht, der über Virtuelle Maschinen (VM) auf einem Rechenknoten den Zugang zu den vFPGAs bereitstellt, wie Abb. 2(a) zeigt. Jeder Rechenknoten besteht aus einem Host-System mit zwei mittels PCIe verbundenen FPGAs (siehe Abb. 2(b)). Die Allokation eines einzelnen FPGAs oder eines vFPGAs mit vordefiniertem RC2F Design (siehe Abb.2(c)) wird durch das Servicemodell RAaaS ermöglicht, welches für den Entwurf neuer Kerne zur Beschleunigung von Anwendungen vorgesehen ist.

3.3 Virtualisierung der FPGAs

Um eine dynamische Interaktion mit rekonfigurierbarer Hardware in eine derartige Cloud-Architektur zu ermöglichen, ist eine tiefgreifende Virtualisierung der FPGAs erforderlich.

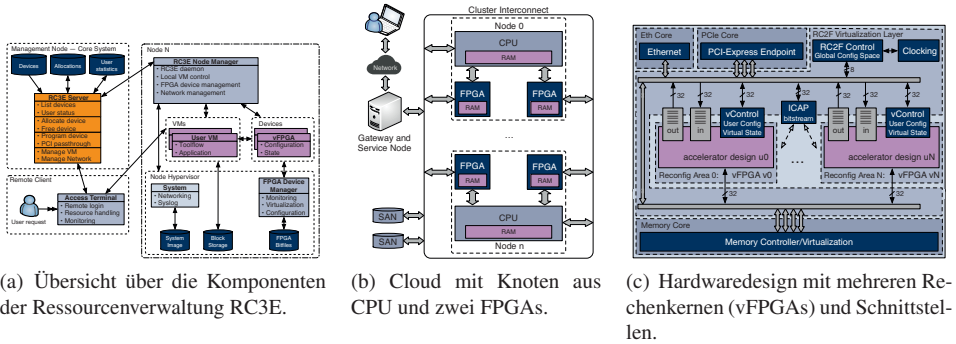


Abb. 2: Aufbau der Ressourcenverwaltung RC3E, der Knoten und des RC2F Hardwaredesigns.

Ziel ist es dabei, mit unterschiedlichsten Nutzern auf demselben physischen FPGA die Auslastung der Ressourcen zu maximieren. Wie bereits in Abschnitt 2 gezeigt, wird der Begriff Virtualisierung im Kontext von FPGAs für unterschiedliche Konzepte genutzt. Ein Grundgedanke des gewählten Ansatzes besteht darin, das Hardwaredesign innerhalb eines vFPGAs auf einem physischen FPGA auszuführen (analog Bare-Metal Virtualisierung).

3.4 Sicherheitskonzept unter Verwendung einer homomorphen Verschlüsselung

Die Architektur aus 3.2 bietet des Weiteren die Möglichkeit, mittels des internen Cloud-Verbindungsnetzwerks direkt auf die vFPGAs zuzugreifen. Auf diese Weise kann über den FPGA ein unmittelbarer Zugang zum Cloud-System ermöglicht werden. Durch Berechnungen mit Klartextdaten auf dem FPGA kann die Sicherheit durch eine Anonymisierung der Nutzerdaten ähnlich wie in [EV12] erreicht werden. Das Prinzip der homomorphen Verschlüsselung [Ge09] kann des Weiteren genutzt werden um Anwendungsteile auf den Host auszulagern.

4 Stand der Arbeit, erste Ergebnisse und Ausblick

Der Stand der Forschung auf dem Gebiet rekonfigurierbarer Hardware und mögliche Einsatzfelder von FPGAs in einer Cloud-Umgebung wurden bereits analysiert, wobei energieeffiziente Hintergrundbeschleunigung und Erhöhung der Sicherheit im Vordergrund standen. Essentiell bei der Entwicklung der flexiblen Cloud-Architektur ist eine Virtualisierung der rekonfigurierbaren Hardware, um Auslastung und Effizienz des Systems zu maximieren, eine Ressourcenverwaltung sowie ein Sicherheitskonzept. Die bisherigen Forschungsergebnisse können wie folgt zusammengefasst werden:

- Exemplarische Implementierung einer Cloud-Verwaltung (RC3E), in welche rekonfigurierbare Hardware mit unterschiedlichsten Servicemodellen integriert werden kann [KLS16].

- Erstellen eines Konzepts zur Virtualisierung von FPGAs sowie dessen Einbettung in ein Framework (RC2F) zur Beschleunigung von Hintergrundanwendungen [Kn15, KGS16] mit der Möglichkeit zur Migration eines Hardwaredesigns.
- Entwurf einer Architektur mit flexiblen Zugangsmöglichkeiten zur Cloud, welche die Sicherheit innerhalb des Systems erhöht [KLS16], indem der Zugang zum Prozessor eines Knotens über den FPGA erfolgen kann.

Literaturverzeichnis

- [By14] Byma, Stuart et al.: FPGAs in the Cloud: Booting Virtualized Hardware Accelerators with OpenStack. In: Field-Programmable Custom Computing Machines (FCCM), 22nd Int'l Symp. on. IEEE, 2014.
- [EV12] Eguro, Ken; Venkatesan, Ramarathnam: FPGAs for trusted cloud computing. In: Field Programmable Logic and Applications (FPL), 22nd Int'l Conf. on. IEEE, S. 63–70, 2012.
- [Fa15] Fahmy, Suhaib A et al.: Virtualized FPGA accelerators for efficient cloud computing. In: Cloud Computing Technology and Science (CloudCom), Int'l Conf. on. IEEE, 2015.
- [Ge09] Gentry, Craig et al.: Fully homomorphic encryption using ideal lattices. In: STOC. Jgg. 9, S. 169–178, 2009.
- [Ha15] Happe, Markus et al.: Preemptive Hardware Multitasking in ReconOS. In: Applied Reconfigurable Computing. Springer, 2015.
- [Jo10] Jozwik, Krzysztof et al.: A novel mechanism for effective hardware task preemption in dynamically reconfigurable systems. In: Field Programmable Logic and Applications (FPL). Int'l Conf. on. IEEE, 2010.
- [KGS16] Knodel, Oliver; Genßler, Paul; Spallek, Rainer: Migration of long-running Tasks between Reconfigurable Resources using Virtualization. In: ACM SIGARCH Computer Architecture News Volume 44 - HEART '16. ACM, 2016.
- [KLS16] Knodel, Oliver; Lehmann, Patrick; Spallek, Rainer: RC3E: Reconfigurable Accelerators in Data Centres and their Provision by Adapted Service Models. In: Cloud Computing, 9th Int'l Conf. on. IEEE, 2016.
- [Kn15] Knodel, Oliver et al.: Computing Framework for Dynamic Integration of Reconfigurable Resources in a Cloud. In: Digital System Design, Euromicro Conf. on. IEEE, 2015.
- [MG11] Mell, Peter; Grance, Timothy: The NIST definition of cloud computing. National Institute of Standards and Technology Gaithersburg, S. 23–27, 2011.
- [Mo11] Mondol, J-AM: Cloud security solutions using FPGA. In: Communications, Computers and Signal Processing (PacRim), Pacific Rim Conf. on. IEEE, S. 747–752, 2011.
- [Pu14] Putnam, Andrew et al.: A reconfigurable fabric for accelerating large-scale datacenter services. In: Computer Architecture (ISCA), 41st Int'l Symp. on. IEEE, 2014.
- [Wa13] Wang, Wei et al.: pvFPGA: Accessing an FPGA-based hardware accelerator in a paravirtualized environment. Hardware/Software Codesign and System Synthesis (CODES+ISSS), Int'l Conf. on, 2013.
- [We15] Weerasinghe, Jagath et al.: Enabling FPGAs in Hyperscale Data Centers. In: Cloud and Big Data Computing (CBDC), Int'l Conf. on. IEEE, 2015.

Personal BPM – Anwenderorientierte Prozessmodellierung durch persönliches Workflow Management

Jonas Lehner¹

Abstract: Der Vorgang der Modellierung von Geschäftsprozessen in Unternehmen ist mitunter aufwändig und fehleranfällig. Normalerweise werden die Modelle vom Management oder Modellierungsexperten unter Einsatz von Workshops und Experteninterviews mithilfe spezieller Modellierungssprachen erstellt. Da den Modellierungsexperten das Wissen der Mitarbeiter nur indirekt zur Verfügung steht, entstehen dabei häufig Fehler. Um diese zu verhindern, wird hier vorgeschlagen, Prozessmodellfragmente direkt durch einzelne Mitarbeiter erzeugen zu lassen, die anschließend automatisiert zu Gesamtmodellen zusammengesetzt werden. Die Motivation des Mitarbeiters zur Modellierung seiner eigenen Abläufe besteht in der Möglichkeit, durch die Automatisierung seiner Aufgaben bei der Arbeit entlastet zu werden. Dazu wird in diesem Forschungsvorhaben eine Modellierungssprache entwickelt, die einfach genug ist, um von ungeschulten Mitarbeitern verstanden zu werden und die aufbauend auf einem Workflowmanagementsystem die Automatisierung von Arbeitsabläufen ermöglicht. Desweiteren wird eine Möglichkeit untersucht, die Modellfragmente der einzelnen Mitarbeiter zu einem Gesamtprozessmodell automatisiert zusammenzusetzen.

Keywords: Personal BPM, Personal Workflow Management, Petri-Netze

1 Einleitung

Bei der Modellierung von Geschäftsprozessen als Teil des Geschäftsprozessmanagements (BPM) gibt es einige strukturelle Probleme: Prozessmodelle werden häufig durch das Management entwickelt (Top-Down-Ansatz). Die Modellierungsexperten, die diese Modellierung durchführen, sind im Normalfall spezialisierte Berater, die nicht mit dem Domänenwissen der den Prozess ausführenden Mitarbeiter ausgestattet sind. Das führt dazu, dass das Wissen über die tatsächlichen Abläufe beispielsweise durch Workshops oder Experteninterviews erst erhoben werden muss. Dabei besteht das Problem, dass das implizite Prozesswissen der Mitarbeiter externalisiert und mithilfe des Modellierungsexperten in eine spezielle, meist grafische, Modellierungssprache übersetzt werden muss. Diese ist für den Mitarbeiter nur schwer verständlich (z.B. [FL15]). Da das gegenseitige Verständnis dabei nicht immer gegeben ist oder die Modellierer nicht die nötige Erfahrung haben, können bei diesem Vorgang Fehler passieren, die dazu führen, dass die Prozessmodelle nicht die tatsächlichen betrieblichen Abläufe repräsentieren (z.B. [Ba07, KW10]).

Werden die fehlerhaften Prozessmodelle in IT-Systemen abgebildet, führt das zu einer Situation, in der die Mitarbeiter bei der Ausführung ihrer Aufgaben nicht durch diese unterstützt oder sogar behindert werden [GKW08]. Darüber hinaus kann vorhandenes Automatisierungspotential nicht ausgeschöpft werden, was ein ökonomisches Risiko birgt.

¹ Karlsruher Institut für Technologie (KIT), Institut für Angewandte Informatik und Formale Beschreibungsverfahren (AIFB), Kaiserstr. 89, 76133 Karlsruhe, Deutschland, jonas.lehner@kit.edu

Eine Möglichkeit, dem Problem der Modellierung zu begegnen, könnte es sein, die Erstellung von Prozessmodellen auf Mitarbeiterebene zu ermöglichen (Bottom-Up-Ansatz). Einzelne Mitarbeiter modellieren dabei einen Ablauf jeweils aus ihrer eigenen Perspektive und erzeugen dadurch subjektive Modellfragmente. Diese einzelnen Fragmente können anschließend – möglichst automatisiert – zu einem Gesamtmodell zusammengesetzt werden. Ein Anreiz, die eigenen Abläufe zu modellieren, kann dem Mitarbeiter dadurch gegeben werden, dass durch die Modellierung ausführbare Workflows erzeugt werden, die den Mitarbeiter durch Automatisierung von Aufgaben entlasten und ihn bei der Erfüllung seiner Arbeit unterstützen.

Ein weiterer Vorteil dieses Mitarbeiter-zentrierten Ansatzes ist die Möglichkeit, flexibel Anpassungen an bestehenden Prozessen vorzunehmen. Im Gegensatz zum traditionellen Vorgehen, bei dem eine Änderung zunächst modelliert und anschließend im IT-System abgebildet werden muss, kann hier die Anpassung direkt vom betroffenen Mitarbeiter umgesetzt werden. Diese Probleme mit der Anpassung von Prozessen an neue Situationen wird z.B. von Indulska et al. beschrieben [In06].

2 Lösungsansatz

In dieser Arbeit wird *Personal BPM* als Konzept eingeführt, um den in der Einleitung beschriebenen Problemen durch den Einsatz einer Anwender-zentrierten Perspektive auf das Geschäftsprozessmanagement und die Ausführung von Workflows zu begegnen. Als Anwender wird hierbei ein Domänenexperte verstanden, der nicht notwendigerweise über Modellierungserfahrung verfügen muss und dessen Tagesgeschäft zumindest teilweise durch sich stark ähnelnde Abläufe am Computer bestimmt wird. Beispiele hierfür sind Vertriebsmitarbeiter bei einer Versicherung, Unternehmensberater oder wissenschaftliche Mitarbeiter in einer Universität.

Geschäftsprozessmanagement wird in Unternehmen vorangetrieben, kommt aber, wie gezeigt wurde, nicht immer auf der operativen Ebene, also beim einzelnen Mitarbeiter, an oder basiert aufgrund der Barriere zwischen Modellierer und Mitarbeiter nicht unbedingt auf den real existierenden betrieblichen Abläufen. Im Rahmen dieser Arbeit ergeben sich dadurch folgende Forschungsfragen:

- FF 1** *Wie können die positiven Aspekte des Geschäftsprozessmanagements auf die operative Ebene einzelner Mitarbeiter übertragen werden?*
- FF 2** *Wie können einzelne Prozessfragmente auf Mitarbeiterebene automatisiert zu einem Gesamtprozess verbunden werden?*

Diese Forschungsfragen können in weitere Fragen untergliedert werden:

- FF 1.1** *Welche Sprache eignet sich für die Modellierung von Geschäftsprozessen auf Mitarbeiterebene?*

Normalerweise werden Geschäftsprozesse mit grafischen Modellierungssprachen, wie *Business Process Model and Notation (BPMN)*, *ereignisgesteuerten Prozessketten (EPK)* oder *Petri-Netzen*, modelliert. Während diese Sprachen für Modellierungsexperten sinnvoll sind, haben die meisten Mitarbeiter Probleme, diese zu verstehen oder zu benutzen. Es ist also notwendig, eine Sprache zu finden, die von ungeschulten Anwendern verwendet werden kann, aber trotzdem ausreichend formal ist, um als Grundlage für das Geschäftsprozessmanagement eines Unternehmens zu dienen.

FF 1.2 *Wie muss ein Workflowmanagementsystem in diesem Kontext beschaffen sein?*

Workflowmanagementsysteme sind normalerweise komplexe Softwareprodukte, die von IT-Experten bedient und konfiguriert werden. In diesem Kontext sollen sie jedoch durch normale Mitarbeiter verwendet werden können. Diese Frage ist eng mit der vorherigen (FF 1.1) verbunden, da die Modellierungssprache maßgeblich die Erscheinung des Workflowmanagementsystems beeinflusst.

FF 1.3 *Wie verändert sich die Produktivität einzelner Mitarbeiter durch den Einsatz von Personal BPM?*

Es muss untersucht werden, ob die Automatisierung von Arbeitsabläufen in diesem Bereich positive Effekte auf die Produktivität der jeweiligen Mitarbeiter hat oder ob der Aufwand so groß ist, dass diese Effekte kompensiert oder sogar ins Negative gekehrt werden. Diese Forschungsfrage dient der Evaluation der Fragen FF 1.1 und FF 1.2.

FF 2.1 *Wie lassen sich zusammengehörige Prozessfragmente identifizieren?*

Die Fragmente, die durch die einzelnen Mitarbeiter erzeugt wurden, stehen zunächst nur für sich allein. Es müssen anschließend Schnittstellen zu den Prozessfragmenten anderer Mitarbeiter identifiziert werden. Anhand dieser Schnittstellen könnten zusammengehörige Fragmente – möglicherweise automatisch – erkannt und verbunden werden.

FF 2.2 *Wie lässt sich das erzeugte Prozessmodell validieren?*

Während der Entstehung des Prozessmodells können zahlreiche Fehler passieren (einzelne Fragmente fehlerhaft modelliert, korrekte Fragmente falsch zusammengesetzt usw.). Deswegen muss eine Form der Validierung geschaffen werden, durch die sichergestellt wird, dass das erzeugte Modell den tatsächlichen betrieblichen Ablauf repräsentiert.

3 Aktueller Stand

Es wurde zunächst mit der Auswahl einer für den Anwender geeigneten Modellierungssprache begonnen (Forschungsfrage FI.1). Um auf bereits erfolgte Forschung zurückgreifen zu können, wurde hierbei zweistufig vorgegangen: Zunächst wurde eine geeignete etablierte Sprache ausgewählt und diese um eine Abstraktionsschicht ergänzt, die eine leicht verständliche grafische Repräsentation für den Anwender bietet [Ko15].

Zur Auswahl standen hier BPMN, Petri-Netze, ereignisgesteuerte Prozessketten (EPK) und UML Aktivitätsdiagramme. Da das Ziel eine einfach zu verstehende Sprache war, wurden Petri-Netze als Grundlage ausgewählt. Diese erlauben es, auch komplexe Abläufe anschaulich darzustellen und bieten aufgrund der mathematischen Definiertheit die Möglichkeit zur Analyse und Simulation. Außerdem existieren zahlreiche Verfahren und Werkzeuge zur Arbeit mit Petri-Netzen.

Eine genaue Analyse ergab, dass Low-Level-Netze, im Speziellen Bedingungs/Ereignis-Netze (*B/E-Netze*), für den Anwendungsfall ausreichend sind. Ein Bedingungs/Ereignis-Netz ist ein Tupel $BEN = (S, T, F, M_0)$, für das gilt:

1. S ist eine Menge von Stellen (Bedingungen).
2. T ist eine Menge von Transitionen (Ereignissen).
3. $S \cap T = \emptyset$
4. $F \subseteq (S \times T) \cup (T \times S)$ ist die Flussrelation.
5. $M : S \rightarrow \{0, 1\}$ ist eine Abbildung, die jeder Stelle des Netzes eine Markierung zuweist. M_0 ist die Startmarkierung.

Mit B/E-Netzen lassen sich Geschäftsprozesse als bipartiter Graph darstellen. Dabei wird durch die aktuelle Markierung festgelegt, welche Transition als nächstes schalten kann, also welche Aktion im Prozess als nächstes durchgeführt wird.

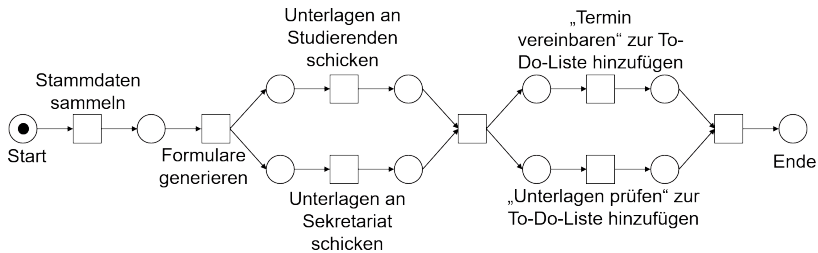


Abb. 1: B/E-Netz-Darstellung eines Prozessfragments zur Anmeldung einer Abschlussarbeit an einer Universität aus Sicht eines wissenschaftlichen Mitarbeiters.

Der Dual Coding-Theorie folgend, die auch neurowissenschaftlich untersucht wurde, ist eine Kombination aus Text und grafischen Elementen kognitiv am effektivsten [Pa86]. Außerdem sind Flussdiagramme bei Laien die beliebtesten und gleichzeitig am wenigsten fehleranfälligen Darstellungsformen [RSR12]. Insofern sind B/E-Netze als Grundlage in diesem Kontext gut geeignet.

Aufbauend auf den B/E-Netzen wurde eine Abstraktionsschicht entwickelt, die für den Anwender leichter verständlich ist. Dabei wurden Erkenntnisse aus den Kognitionswissenschaften (benutzerbezogene Einflussfaktoren, z.B. [Ma05, Cu84]) ebenso berücksichtigt, wie modellbezogene Einflussfaktoren (z.B. [RSR12]). Diese Darstellung ist bezüglich der

von Koschmider et al. in [Ko15] beschriebenen visuellen Variablen (u.a. Position, Form, Farbe, Größe) an die Anforderungen von ungeschulten Mitarbeitern angepasst.

Die in der neu entwickelten Sprache modellierten Prozessmodellefragmente sind aufgrund ihrer Ableitung von B/E-Netzen automatisch in diese überführbar, wodurch die weitere Verarbeitung ermöglicht wird.

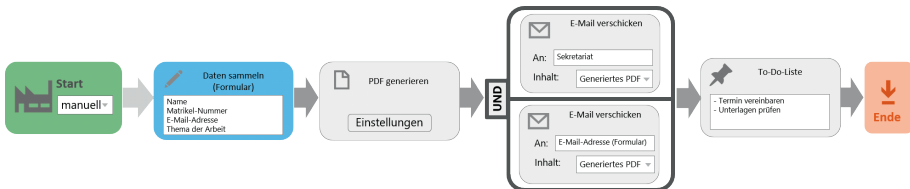


Abb. 2: Darstellung des gleichen Prozesses in der neu entwickelten Sprache

Bezüglich der Forschungsfrage *FF 2.1* erscheinen Bezüge zum subjektorientierten Geschäftsprozessmanagement (S-BPM), insbesondere zum Subjekt-Interaktions-Diagramm (SID) des Parallel Activity Specification Schema (PASS) vielversprechend [F196]. Diese werden aktuell untersucht.

4 Verwandte Ansätze

Die Idee, Petri-Netze für die Modellierung von Geschäftsprozessen zu verwenden, ist häufig beschrieben (z.B. [va98]). Dem Problem der teilweise schwierigen Verständlichkeit bei Laien wird auch durch den Vorschlag von Assistenzsystemen begegnet [Fe15]. Obwohl dadurch die Qualität von Prozessmodellen erhöht werden kann, ist auch hier Vorwissen nötig.

Eine Variante, die Barriere zwischen Modellierer und Wissensträger zu schließen, wird von Luebbe und Weske [LW11] beschrieben. Es werden Methoden des Design Thinking verwendet, um Prozessmodellierung für den Anwender leichter verständlich zu machen. Dazu wird das “Tangible Business Process Modelling” beschrieben, bei dem mit Plastikscheiben, die den Elementen von BPMN nachempfunden sind, gemeinsam durch Modellierer und Wissensträger modelliert.

Weitere verwandte Ansätze existieren auf dem Bereich der persönlichen Aufgabenverwaltung [St08]. Dort ist jedoch die Automatisierbarkeit nicht gegeben und es fehlt der Bezug zu den Prozessfragmenten anderer Mitarbeiter.

Des weiteren gibt es Webdienste, die sehr einfache Workflows für den persönlichen Bereich ermöglichen, z.B. IFTTT², und Applikationen für Smartphones, wie “workflow”³. Diese Dienste erlauben es dem Nutzer, einfache Wenn-Dann-Regeln für Social-Media-Dienste zu erstellen. Ein ähnlicher Dienst von Microsoft ist aktuell im Beta-Stadium⁴.

² <http://ifttt.com>, zuletzt abgerufen am 13.05.2016

³ <http://workflow.is>, zuletzt abgerufen am 13.05.2016

⁴ <https://flow.microsoft.com>

Literatur

- [Ba07] Bandara, Wasana; Indulska, Marta; Chong, S.; Sadiq, Shazia: Major Issues in Business Process Management: An Expert Perspective. ECIS 2007 Proceedings, 2007.
- [Cu84] Curtis, Bill: Fifteen Years of Psychology in Software Engineering: Individual Differences and Cognitive Science. In: Proceedings of the 7th International Conference on Software Engineering. ICSE '84, IEEE Press, Piscataway, NJ, USA, S. 97–106, 1984.
- [Fe15] Fellmann, Michael; Zarvic, Novica; Metzger, Dirk; Koschmider, Agnes: Requirements Catalog for Business Process Modeling Recommender Systems. In: Wirtschaftsinformatik. S. 393–407, 2015.
- [Fl96] Fleischmann, Albert: Distributed systems: software design and implementation. Springer Science & Business Media, 1996.
- [FL15] Figl, Kathrin; Laue, Ralf: Influence factors for local comprehensibility of process models. International Journal of Human-Computer Studies, 82:96 – 110, 2015.
- [GKW08] Gschwind, Thomas; Koehler, Jana; Wong, Janette: In: Business Process Management: 6th International Conference, BPM 2008, Milan, Italy, September 2-4, 2008. Proceedings. Springer Berlin Heidelberg, S. 4–19, 2008.
- [In06] Indulska, Marta; Chong, Sandy; Bandara, Wasana; Sadiq, Shazia; Rosemann, Michael: Major Issues in Business Process Management: An Australian Perspective. ACIS 2006 Proceedings, 2006.
- [Ko15] Koschmider, Agnes; Caporale, Timm; Fellmann, Michael; Lehner, Jonas; Oberweis, Andreas: Business process modeling support by depictive and descriptive diagrams. In: 6th International Workshop on Enterprise Modelling and Information Systems Architectures (EMISA), Innsbruck, Austria. 2015.
- [KW10] Karagiannis, Dimitris; Woitsch, Robert: Knowledge engineering in business process management. In: Handbook on Business Process Management 2, S. 463–485. Springer, 2010.
- [LW11] Luebbe, Alexander; Weske, Mathias: Bringing Design Thinking to Business Process Modeling. In (Meinel, Christoph; Leifer, Larry; Plattner, Hasso, Hrsg.): Design Thinking, Understanding Innovation, S. 181–195. Springer Berlin Heidelberg, 2011.
- [Ma05] Mayer, Richard E: The Cambridge handbook of multimedia learning. Cambridge University Press, 2005.
- [Pa86] Pavilion, A: Mental representation: A dual-coding approach. New York: Oxford University Press, 1986.
- [RSR12] Recker, Jan; Safrudin, Norizan; Rosemann, Michael: How Novices Design Business Processes. Inf. Syst., 37(6):557–573, September 2012.
- [St08] Stoitsev, Todor; Scheidl, Stefan; Flentge, Felix; Mühlhäuser, Max: From Personal Task Management to End-User Driven Business Process Modeling. In: Business Process Management, Jgg. 5240 in Lecture Notes in Computer Science, S. 84–99. Springer Berlin Heidelberg, 2008.
- [va98] van der Aalst, W. M. P.: The Application of Petri Nets to Workflow Management. Journal of circuits, systems, and computers, 08(01):21–66, 1998.

Adaptive Informationsvisualisierung: Der Einsatz von Machine-Learning zur Entscheidungsunterstützung

Franziska Marx ^{1, 2}

Abstract: In meiner Dissertation beschäftige ich mich mit der Frage, inwiefern die für die datengestützte Entscheidungsfindung relevanten Informationsvisualisierungen durch Machine-Learning-Algorithmen optimiert werden können. Ausgehend von der Problematik, dass es vielen Entscheidungsträgern schwerfällt, statistische Daten korrekt zu lesen, ist es notwendig, für die Erfüllung der Aufgabe relevante Informationen optimal zu visualisieren. Dabei muss sowohl das Wissen des Entscheidungsträgers als auch der Kontext der dargestellten Informationen berücksichtigt werden. Durch den Einsatz von Machine-Learning-Konzepten sollen die daraus gewonnenen Erkenntnisse sukzessive zur Verbesserung des Visualisierungssystems beitragen und den Entscheidungsträger gezielt unterstützen. Durch die Teilnahme am Doktoranden-Symposium erhoffe ich mir einen intensiven Austausch zur besseren Verankerung meiner interdisziplinären Fragestellung sowie die Möglichkeit der Weitergabe von bereits gewonnenen Erkenntnissen.

Keywords: Informationsvisualisierung, Decision Support Systems, Machine Learning, Statistical Literacy

1 Forschungsvorhaben

1.1 Motivation

Durch die im Zuge der *Datafication* vorhandenen Fülle an digitalen Daten, deren hohe Verfügbarkeit sowie die weltweite Vernetzung ist es heute ein Leichtes, Daten miteinander zu verknüpfen und in einem Decision Support System (DSS) einzusetzen. Diese Systeme haben zum Ziel, für die Entscheidungsfindung notwendige Daten effizient analysieren zu können und häufig durch geeignete Visualisierung zu präsentieren, sodass eine schnelle, flexible und einfache Einsicht in die Daten erlangt werden kann. Während Unternehmen Decision Support Systems zur Steigerung der Wettbewerbsfähigkeit einsetzen, soll beispielsweise durch den Einsatz in Krankenhäusern die Wartezeit für Patienten³ verkürzt werden oder in Schulen die individuelle Förderung einzelner Schüler ermöglicht werden. Die Einbettung in grafische Benutzungsoberflächen oder Web-Applikationen ermöglicht eine Interaktion mit den Daten, wodurch diese individuell selektierbar werden. Allerdings kann ein solches System nur soweit unterstützend wirken, wie die für den Benutzer sichtbaren Daten zugänglich und verständlich sind. Zur Interpretation und späteren Entschei-

¹ Institut für Informationsmanagement Bremen GmbH (ifib), Am Fallturm 1, 28359 Bremen, fmarx@ifib.de

² Universität Bremen, Bibliothekstraße 1, 28359 Bremen, franziska.marx@uni-bremen.de

³ Aus Gründen des besseren Leseflusses wird im vorliegenden Beitrag die weibliche Formen nicht explizit angeführt. An dieser Stelle wird jedoch ausdrücklich darauf hingewiesen, dass sich alle personenbezogenen Formulierungen grundsätzlich gleichermaßen auf Frauen und Männer beziehen.

dungsfindung ist es somit von Belang, inwiefern der Nutzer mit dem Kontext der Daten vertraut ist, aber auch, ob er diese interpretieren kann.

1.2 Fragestellung

In meiner Dissertation möchte ich mich der Frage widmen, inwiefern Machine-Learning-Algorithmen zur Optimierung von Decision Support Systems beitragen können, sodass die Darstellung von Informationen den Nutzer in seiner Entscheidung bestmöglich unterstützt. Dazu soll zunächst das Themengebiet "Entscheidungsfindung" aus verschiedenen Perspektiven und Disziplinen, wie Psychologie oder Betriebswirtschaftslehre, betrachtet werden, um herauszufinden, inwiefern die Messung der Qualität einer Entscheidung möglich ist und wie diese unterstützt werden kann. Daneben soll durch die Analyse von Machine-Learning-Konzepten die Frage beantwortet werden, welche Algorithmen für die Unterstützung von Entscheidungsfindung am Besten geeignet sind.

2 Stand der Forschung

Zur Beantwortung meiner Forschungsfrage erscheint es mir zunächst notwendig, das Themengebiet von Decision Support Systems näher zu beleuchten. Die Forschung in diesem Bereich hat in den letzten Jahren verstärkt die Benutzbarkeit und Verständlichkeit solcher Systeme in den Vordergrund gerückt, wodurch die Darstellungsform der Informationen diskutiert wurde. Auf der anderen Seite spielt der Einsatz von künstlichen Intelligenzen (*Artificial Intelligence*) und selbst lernenden Systemen (*Machine Learning*) eine immer größere Rolle, wodurch ein DSS komplexe Handlungen erfassen kann und aus vergangenen Entscheidungen automatisch lernt. Eine Herausforderung stellt hierbei die Fähigkeit des Entscheidungsträgers dar, statistische Daten zu lesen und auszuwerten, die so genannte *Statistical Literacy*. Forschungsbedarf sehe ich in der Verknüpfung dieser Themengebiete, da sich die bisherigen Forschungen entweder auf die Generierung von verständlichen Visualisierungen im Rahmen von DSS konzentrieren oder aber auf den Einsatz von Machine-Learning-Konzepten in diesem Feld, wobei die adaptive Gestaltung hier auf die Prozesse und nicht auf die Eigenschaften des Entscheidungsträger bezogen ist. In meiner Dissertation werde ich eine Brücke zwischen den einzelnen Themenbereichen schlagen und die *Statistical Literacy* des Nutzers sowie den Kontext der Entscheidung fokussieren, um daraus die bestmögliche grafische Darstellung zu gewinnen. Im Folgenden wird daher der aktuelle Stand der Forschung meiner thematischen Schwerpunkte kurz erläutert.

2.1 Decision Support Systems

Little beschreibt 1970 [Li70] mit dem Begriff des "Decision Calculus" ein "set of procedures for processing data and judgments to assist a manager in his decision making" und argumentiert, dass ein solches System erfolgreich ist, wenn es simpel, robust, einfach zu bedienen, adaptiv, in den wichtigsten Punkten vollständig und die Kommunikation mit

dem System einfach ist. 1971 etabliert Morton den Begriff des "Decision Support Systems", das mit Computern und analytischen Modellen Managern helfen soll, Probleme zu lösen [Mo71]. Keen betont, dass "a final system can be developed only through an adaptive process of learning and evolution" und fordert eine stärkere Verknüpfung von System, Benutzer und Entwickler, während der Nutzer den adaptiven Designprozess bestimmt [Ke80]. Um die Entscheidungsfindung zu erleichtern, wird zunehmend die Ausgestaltung der Datenpräsentation in Form von Visualisierungen untersucht (vgl. [Te99] [SHS02]). Yu et al. fokussieren vor allem die Möglichkeit der Interaktion des Nutzers mit den dargestellten Informationen [YDZ09], während Zhang die Visualisierung durch die Anwendung von abstrakten Regeln optimiert [Zh98]. Tsoukalas et al. stellen die Einbindung von Machine-Learning-Algorithmen zur Verbesserung von DSS vor [TAT15].

Diese Entwicklung zeigt, wie sich das Verständnis des Begriffs Decision Support System über die Jahre vom reinen Management-Tool für einzelne Prozesse zu einem vielschichtigen Werkzeug gewandelt hat, bei dem der Nutzer immer mehr in den Fokus des Systemdesigns rückt. Meine Arbeit soll diesen Trend aufgreifen und durch die Verknüpfung mit weiteren Techniken zugunsten der Benutzbarkeit und Verständlichkeit weiter entwickeln, sodass ein DSS entsteht, welches das Vorwissen und die Fähigkeiten des Benutzers berücksichtigt.

2.2 Machine Learning

Machine Learning, die künstliche Generierung von Wissen aus Erfahrung, findet seit Jahren verstärkt Anwendung in DSS, um dort bestimmte Muster zu erkennen und zu beurteilen (vgl. Turban, Part IV[TA05]). Die Grundlagenforschung des Machine Learning entwickelt dabei stets neue Algorithmen, die sich für die Lösung von Problemen unterschiedlich gut eignen. Auch in diesem Forschungsbereich wird der Nutzen für eine verbesserte Nutzerinteraktion in den letzten Jahren verstärkt erforscht, z. B. in Bezug auf die Verarbeitung natürlicher Sprache [ZL08] oder die Visualisierung von Daten [Ar15]. Eine Herausforderung stellt hierbei die Wahl des richtigen Algorithmus sowie dessen Integration in das DSS dar [MN06]. Durch die Kooperation mit Wissenschaftlern aus dem Bereich der Grundlagenforschung im Machine Learning möchte ich einen für mein Vorhaben optimalen Algorithmus auswählen und in das DSS integrieren.

2.3 Informationsvisualisierung

Die Visualisierung von Informationen zur besseren Verständlichkeit ist keineswegs eine Erfindung der modernen Gesellschaft, wie Landkarten aus der Steinzeit beweisen. Die menschliche Wahrnehmung in Bezug auf visuelle Informationen ist dabei gut erforscht (z. B. Few [Fe09], Miller [Mi56]), wird aber häufig nicht berücksichtigt. Tufte prangert an, dass bei der grafischen Darstellung von quantitativen Daten oftmals gestalterische Aspekte im Vordergrund stehen, welche die Korrektheit der Daten - teils unbeabsichtigt - verfälschen und die Aussage damit zur "Lüge" abwandeln [Tu01]. Er entwickelte daraufhin Grundregeln für die Visualisierung von Informationen, insbesondere von statisti-

schen Daten, die ich in meiner Dissertation mit den für das Lesen statistischer Daten notwendigen Kompetenzen verknüpfen möchte. Durch die Implementierung eines Machine-Learning-Systems soll dies sukzessive ausgebaut werden.

2.4 Statistical Literacy

Bereits 1951 machte Walker auf die gesellschaftliche Problematik aufmerksam, welche die Unfähigkeit statistische Daten zu lesen mit sich bringt [Wa51]. Die durch Sozialforscher oder Organisationen erhobenen Daten zur Erlangung von mehr Transparenz und Veranschaulichung von Sachverhalten wurden missverstanden oder durch deren Autoren falsch eingesetzt, sodass sie oftmals das Gegenteil bewirkten. Wallmann betont, dass Statistical Literacy keineswegs nur im wissenschaftlichen oder politischen Kontext relevant ist, sondern für alle Teile einer Gesellschaft im täglichen Leben eine bedeutende Rolle spielt [Wa93]. Die fehlende Statistical Literacy beschränkt sich dabei nicht etwa auf bildungsferne Bevölkerungsschichten, sondern ist auch in Berufsgruppen wie Ärzten [We15] oder Lehrern [PCG13] zu finden. Um dem entgegenzuwirken untersuchten Chick und Pierce die konkreten Kompetenzen, die eine Lehrkraft braucht, um Schul-Statistiken zu lesen und formulierten daraus Empfehlungen für gezielte Schulungen[CP13], während Carvalho und Solomon die Notwendigkeit der Individualisierung von Statistiken und die Berücksichtigung des Kontexts des Lesers herausstellen [CS12]. Ausgehend von dem Wissen über die Visualisierung von Informationen kann hier angesetzt werden, um fehlende Kompetenzen im Umgang mit Statistiken kontextabhängig zu unterstützen. Dabei ist noch zu untersuchen, wie die Statistical Literacy der Zielgruppe des zu entwickelnden DSS, also u.a. Hochschulleitungen oder Studieninteressierte, eingeschätzt werden kann.

3 Forschungsplan

Im Zuge meiner Dissertation werde ich zunächst auf Basis vorhandener Forschungsergebnisse die grundlegenden Anforderungen an das zu entwickelnde System herausarbeiten, die sich auf die zu unterstützenden Kompetenzen und die Darstellungsform der Informationen beziehen. Im zweiten Schritt werde ich die Besonderheiten benennen, die der gewählte Informationskontext mit sich bringt. Aktuell sieht mein Forschungsvorhaben vor, Hochschul-Ranking-Daten zu verwenden, da hier eine große Anzahl von Datensätzen vorhanden ist, zwischen denen interessante Wechselwirkungen auftreten und die das notwendige Maß an Komplexität aufweisen. Meine bisherigen Recherchen haben zudem ergeben, dass es im deutschsprachigen Raum keine Anwendung gibt, die diese Daten unter Berücksichtigung des Nutzers (z. B. ob Hochschul-Direktor oder Studienanfänger) zur Verfügung stellt, obwohl ein DSS im Hochschulbereich beispielsweise für die Studiengangswahl oder organisatorische Entscheidungen von Vorteil wäre. Anschließend werde ich verschiedene Konzepte des Machine Learning betrachten sowie deren wesentliche Gemeinsamkeiten und Unterschiede herausarbeiten. Daraus soll abgeleitet werden, welcher Algorithmus am besten für das geplante System geeignet ist. Im nächsten Schritt werde ich ein DSS auf Basis der vorherigen Erkenntnisse entwickeln.

Daneben muss eine Metrik ausgewählt oder angepasst werden, mit der die Qualität der Entscheidung bestimmt werden kann. Hierbei ist insbesondere eine Analyse der aus Psychologie oder BWL bereits bestehenden Metriken erforderlich.

Abschließend soll das entwickelte System mittels vergleichender Nutzertests in zwei Schritten evaluiert werden. Im ersten Schritt werde ich die Statistical Literacy von Testnutzern sowie deren Hintergrundwissen in einem bestimmten Bereich durch einen Fragebogen erheben. Anschließend werden jeder Testperson Visualisierungen von Informationen gezeigt, auf Basis derer sie Entscheidungen treffen muss. Die Ergebnisse aus diesem Schritt werden als Trainingsdaten in das Machine Learning System übergeben. Im zweiten Schritt werden eine weitere Testgruppe sowie eine Kontrollgruppe bestimmt. Beide Gruppen durchlaufen denselben Testzyklus wie in Schritt 1, wobei die Testgruppe angepasste Visualisierungen auf Basis der Erkenntnisse aus Schritt 1 gezeigt bekommt, während die Kontrollgruppe auf Basis der unangepassten Visualisierungen Entscheidungen treffen soll. Die Ergebnisse der beiden Gruppen werden anschließend gegenüber gestellt.

4 Stand der Arbeit

Seit Anfang dieses Jahres beschäftige ich mich mit der Konkretisierung meines Dissertationsthemas. Zu diesem Zweck habe ich begonnen, die Schnittmengen und Abgrenzungen der vorhergehend aufgeführten Themen herauszuarbeiten. Hier gilt es besonders zu berücksichtigen, dass eine Entscheidung meist abhängig von dem Themengebiet getroffen wird, in dem sie angesiedelt ist. Daraus ergibt sich die Akquise von Daten aus einem konkreten Anwendungsfeld, aus dem Trainingsdaten für die initiale Eingabe in das Machine-Learning-System gewonnen werden können. Daneben möchte ich eine Kooperation mit Wissenschaftlern erreichen, die sich mit der Grundlagenforschung von Machine-Learning-Algorithmen beschäftigen und deren Vorgehen für die Auswahl eines bestimmten Algorithmus näher beleuchten. Von dem Doktoranden-Symposium der INFORMATIK 2016 erhoffe ich mir einen intensiven Austausch mit Doktoranden, deren Forschungsvorhaben in einem oder mehreren der oben genannten Bereiche angesiedelt ist sowie die Exploration neuer Schnittmengen mit Themengebieten, denen ich bisher weniger Beachtung geschenkt habe. Zudem würde ich meine geplante methodische Vorgehensweise und bisherigen Erfahrungen gerne mit weiteren jungen Wissenschaftlern teilen und zur Diskussion stellen, da dies in der interdisziplinären Forschung eine Besondere Herausforderung darstellt.

Literaturverzeichnis

- [Ar15] Archambault, Daniel; Bunte, Kerstin; Carreira-Perpiñán, Miguel; Ebert, David; Ertl, Thomas; Zupan, Blaz: Machine Learning Meets Visualization: A Roadmap for Scalable Data Analytics. In: Bridging Information Visualization with Machine Learning. Dagstuhl, S. 7–12, 2015.
- [CP13] Chick, Helen; Pierce, Robyn: The Statistical Literacy Needed to Interpret School Assessment Data. *Mathematics Teacher Education and Development*, 15(2), 2013.
- [CS12] Carvalho, Carolina; Solomon, Yvette: Supporting statistical literacy: What do culturally relevant/realistic tasks show us about the nature of pupil engagement with statistics? *International Journal of Educational Research*, 55:57–65, 2012.

- [Fe09] Few, Stephen: Now you see it: simple visualization techniques for quantitative analysis. Analytics Press, Oakland, 2009.
- [Ke80] Keen, Peter G. W.: Adaptive Design for Decision Support Systems. ACM SIGOA Newsletter, 1(4-5):15–25, September 1980.
- [Li70] Little, John D. C.: Models and Managers: The Concept of a Decision Calculus. The Institute of Management Sciences, 16(8):466 – 485, April 1970.
- [Mi56] Miller, George A.: The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological Review, 63(2):81–97, 1956.
- [MN06] Maniyar, Dharmesh M.; Nabney, Ian T.: Visual Data Mining: Integrating Machine Learning with Information Visualization. ACM, Philadelphia, 2006.
- [Mo71] Morton, Michael S. Scott: Management Decision Systems: Computer-Based Support of Decision Making. Harvard University Press, Cambridge, 1971.
- [PCG13] Pierce, Robyn; Chick, Helen; Gordon, Ian: Teachers’ perceptions of the factors influencing their engagement with statistical reports on student achievement data. Australian Journal of Education, 57(3):237–255, November 2013.
- [SHS02] Shen-Hsieh, Angela; Schindl, Mark: Data Visualization for Strategic Decision Making. In: Case Studies of the CHI2002, AIGA Experience Design FORUM. CHI ’02, ACM, New York, S. 1–17, 2002.
- [TA05] Turban, Efraim; Aronson, Jay E.: Decision Support Systems and Intelligent Systems. Pearson/Prentice Hall, 2005.
- [TAT15] Tsoukalas, Athanasios; Albertson, Timothy; Tagkopoulos, Ilias: From Data to Optimal Decision Making: A Data-Driven, Probabilistic Machine Learning Approach to Decision Support for Patients With Sepsis. JMIR Medical Informatics, 3(1):e11, Februar 2015.
- [Te99] Tegarden, David P.: Business Information Visualization. Communications of the Association for Information Systems, 1(1):2–38, Januar 1999.
- [Tu01] Tufte, Edward R.: The visual display of quantitative information. Graphics Press, Cheshire, 2nd ed. Auflage, 2001.
- [Wa51] Walker, Helen M.: Statistical Literacy in the Social Sciences. The American Statistician, 5(1):6–12, 1951.
- [Wa93] Wallman, Katherine K.: Enhancing Statistical Literacy: Enriching Our Society. Journal of the American Statistical Association, 88(421):1–8, 1993.
- [We15] Wegwarth, Odette: Statistical literacy in medicine. Dissertation, Freie Universität Berlin, Freie Universität Berlin, Germany, 2015.
- [YDZ09] Yu, Shidong; Deng, Liguog; Zhang, Yong: Visualization User Interface for Decision Support Systems. In: Ninth International Conference on Hybrid Intelligent Systems, 2009. HIS ’09. Jgg. 1, Shenyang, S. 63–66, August 2009.
- [Zh98] Zhang, Ping: An image construction method for visualizing managerial data. Decision Support Systems, 23(4):371–387, Oktober 1998.
- [ZL08] Zhao, Li; Li, F.: Statistical Machine Learning in Natural Language Understanding: Object Constraint Language Translator for Business Process. In: IEEE International Symposium on Knowledge Acquisition and Modeling Workshop, 2008. KAM Workshop 2008. S. 1056–1059, Dezember 2008.

Domänenspezifische Vorabanalyse und geschickte Generierung von Optimierungsproblemen

Benjamin Saul¹ (Betreuer: Wolf Zimmermann)

Abstract: Optimierungsprobleme lassen sich mit domänenspezifischen Sprachen beschreiben, aus denen dann z. B. Gemischt-Ganzzahlige Lineare Programme (MILP) erstellt werden können. Bei der Auswertung dieser durch entsprechende Löser erhält man jedoch nur unspezifische Fehlermeldungen. Um die Anwendbarkeit der domänenspezifischen Sprache zu fördern, soll durch entsprechende Analysen die Lösbarkeit des Problems untersucht und entsprechende Fehlermeldungen ausgegeben werden. Zusätzlich können die ermittelten Informationen bei der Generierung des MILPs angewandt werden, um dessen weitere Verarbeitung zu begünstigen.

1 Motivation

Diese Forschungsarbeit wird im Rahmen des Projektes „Hocheffiziente Pumpensysteme“ durchgeführt. Ein Pumpensystem kann vereinfacht als Graph von Komponenten, also Pumpen, Ventilen und Rohren, aufgefasst werden. Abbildung 1 zeigt ein solches Pumpensystem.

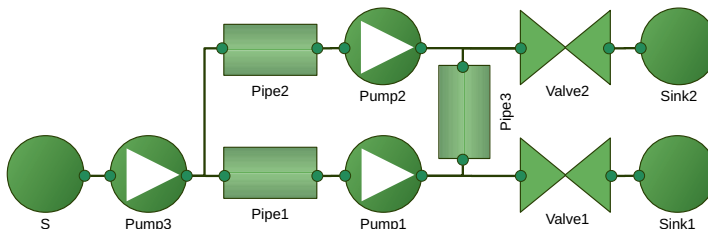


Abb. 1: Beispiel eines Pumpensystems mit Pumpen, Rohren und Ventilen.

Die Aufgabe des Projektes besteht darin, die Rahmenbedingungen eines Pumpensystems zu modellieren und daraus die energie- bzw. kostengünstigste Konfiguration der Komponenten zu bestimmen. Für die Lösung der Aufgabe wird ein Gemischt-Ganzzahliges Lineares Programm (MILP) generiert und durch Standard-Löser verarbeitet. Die entstandene Lösung wird anschließend ausgewertet, indem beispielsweise Simulationsmodelle generiert werden. Das Verfahren wird in Abbildung 2 dargestellt.

Die Modellierung von Pumpensystemen als Gemischt-Ganzzahlige Lineare Programme wird in [Pö15] vorgestellt und u. a. am Beispiel des Darmstadtiums in Darmstadt verwendet. Hier erfolgte die Modellierung des linearen Programmes jedoch von Hand.

¹ Martin-Luther-Universität Halle-Wittenberg, Institut für Informatik, 06120 Halle (Saale), Deutschland, saul@informatik.uni-halle.de

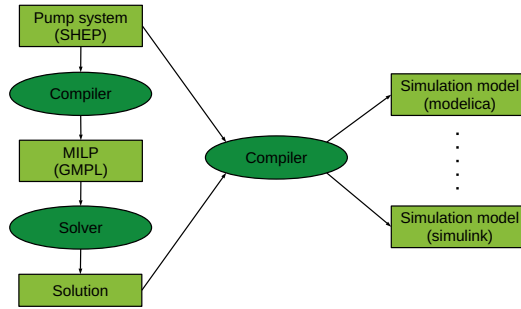


Abb. 2: Übersicht über die Werkzeugkette zur Erstellung optimaler Pumpensysteme.

Domänenspezifische Sprachen erlauben eine effiziente Beschreibung von Problemstellungen durch entsprechende Fachkräfte. Auf dieser Spezifikation erfolgen dann Berechnungen, die das Problem analysieren oder sogar lösen. Aus dem domänenspezifischen Modell kann dann z. B. das entsprechende Optimierungsproblem generiert werden.

Die Löser von Optimierungsproblemen verarbeiten nur das Ungleichungssystem, wodurch domänenspezifische Informationen verloren gehen. Gibt es keine Lösung, so ist es dadurch nur schwer nachvollziehbar, welche Ungleichungen nicht erfüllbar sind. Insbesondere, wenn die Gleichungen generiert wurden, ist dies nur durch fundierte Kenntnisse des entsprechenden Gleichungssystems zu bewältigen. Außerdem kann die Lösungszeit einen erheblichen Zeitaufwand bedeuten, welcher nur schwer vorher einschätzbar ist.

In dieser Arbeit soll daher das Optimierungsproblem in einer domänenspezifischen Beschreibung analysiert werden. Diese Analysen ermöglichen es, entsprechende Widersprüche im Modell aufzudecken, bevor diese im Löser zu Fehlern führen. Dadurch erhält der Anwender ein besseres Feedback und kann mögliche lange Lösungszeiten einsparen.

Als zusätzlicher Effekt können die ermittelten Informationen dabei helfen, Variablen und Gleichungen einzusparen, was das erzeugte Optimierungsproblem verkleinert und damit Lösungszeit einsparen kann. Es ergeben sich die folgenden Forschungsfragen:

- Lassen sich domänenspezifische Informationen verwenden, um auf die Lösbarkeit des entsprechenden Optimierungsproblems schließen zu können?
- Sind diese Methoden besser als die eines Löser für Lineare Optimierungsprobleme?
- Können durch domänenspezifische Analysen Variablen und Restriktionen eingespart werden, sodass die gesamte Lösungszeit verringert wird?

In den folgenden Abschnitten werden zunächst ähnliche Arbeiten aufgeführt. Darunter zählen insbesondere Arbeiten zur Generierung von Optimierungsproblemen sowie Arbeiten über deren Analyse. Anschließend wird der Forschungsplan vorgestellt, der die bewältigten und anstehenden Schritte der Forschungsarbeit zusammenfasst. Der Beitrag schließt mit den bisher erzielten Ergebnissen und einer Zusammenfassung.

2 Verwandte Arbeiten

Die automatische Erzeugung linearer und nichtlinearer Probleme ist notwendig, um auch große Systeme zu modellieren. Übersichtliche Formulierungen bietet die Sprache AMPL [FGK93]. Mit dieser ist es möglich, lineare Probleme in einer verkürzten Form zu beschreiben, indem z. B. Gleichungen und Daten getrennt voneinander angegeben werden können.

Auf so beschriebene Ungleichungssysteme können nur Methoden der allgemeinen Optimierung, wie in [Pa13] beschrieben, angewendet werden. Das heißt, es werden z. B. Variablen eliminiert, indem die zugehörigen Gleichungssysteme weitestgehend gelöst werden. Allerdings kann dieses Vorgehen das weitere Wissen über eventuelle Verknüpfungen von Variablen nicht verwenden, die nicht unmittelbar aus dem Ungleichungssystem hervorgehen.

Andere Ansätze bieten ein komplettes Framework an, bei dem die Modellierung in einer eigenen Sprache erfolgt, damit diese dann direkt weiterverarbeitet werden kann. Ein Beispiel hierfür ist [PJM12], bei dem eine Python-Anwendung entwickelt wurde. Hier werden Optimierungsmodelle mit bereits vorhandenen Bibliotheksfunktionen von Python gelöst. Dabei können dann direkt in der Eingabesprache Vorschläge gemacht werden, welche Lösungsalgorithmen verwendet werden sollten.

Gemischt-Ganzzahlige Lineare und Nichtlineare Probleme besitzen oft alternative Formulierungen. In [SA13] werden einige äquivalente Linearisierungen vorgestellt. Durch den Austausch der Formulierungen kann die Lösungszeit zu Gunsten von Genauigkeit verringert werden.

Für die Generierung von Optimierungsproblemen bietet es sich an, verschiedene Formulierungen für lineare und nichtlineare Zusammenhänge auswählbar zu machen. Ein Beispiel für eine neue Formulierung des Scheduling-Problems findet sich in [OAV12]. Dort werden u. a. die Gleichungen für die benötigten Laufzeiten umformuliert, was eine Verringerung der Lösungszeit mit sich brachte.

Die Lösung der Optimierungsaufgabe kann auch ohne die Generierung eines MILPs erfolgen. In [Su11] wird eine domänenspezifische Sprache vorgestellt, mit dem Quellcode für parallel rechnende Grafikkartenprozessoren beschrieben werden kann. Dieser wird dann auf die jeweilige Architektur angepasst, sodass die Hardware möglichst gut ausgenutzt wird. Die Optimierung selbst erfolgt über eigene Algorithmen innerhalb des Übersetzers.

Ein ähnliches Forschungsfeld ist die Constraint-Basierte Programmierung. Dieses Paradigma erlaubt die Formulierung von Restriktionen, welche dann Einfluss auf die davon betroffenen Variablen haben [RVBW06]. Eine Forschungsarbeit für die Anwendung der Constraint-basierten Programmierung ist [Ma12]. Hier wurde ein Verfahren vorgestellt, welches den Constraint-Ansatz für Produktionslinienmodelle nutzt. Dieser ist deutlich vielseitiger und umfangreicher nutzbar als die grafische Modellierung von Produktionslinien.

Zusammenfassend lässt sich sagen, dass für spezielle Domänen bzw. Gleichungen besonders gute Formulierungen als MILP gefunden wurden, welche so auch generiert werden sollten. Für die domänenspezifischen Analysemethoden können Methoden der Constraint-Programmierung verwendet werden, um Widersprüche zu aufzudecken. Dabei kann die Verknüpfung von Constraint-basierten Modellierungsmethoden und den äquivalenten Umformungen innerhalb eines Optimierungsproblems zu beiderseitigem Vorteil genutzt werden. Die Constraint-Programmierung erhält neue Lösungsmethoden und die Optimierungsprobleme neue Analysemethoden.

3 Aktueller Stand

Die domänenspezifische Sprache SHEP zur Beschreibung der Rahmenbedingungen von Pumpensystemen wurde im Rahmen des Projektes zusammen mit der Werkzeugkette umgesetzt. Die Sprache ermöglicht die Spezifikation von Komponententypen, z. B. Pumpen, deren Zusammenstellung in einem Pumpensystem (vgl. Abb. 1), sowie der geforderten Werte an Druck und Durchfluss an verschiedenen Stellen im System. Auf Basis dieser Sprache und insbesondere am Beispiel von Pumpensystemen sollen die weiteren Verfahren untersucht werden.

Einfache Tests auf Lösbarkeit wurden bereits umgesetzt. So werden die an verschiedenen Stellen des Modells spezifizierten Einschränkungen an Attribute geprüft. Codeabschnitt 1 zeigt einen entsprechenden Ausschnitt mitsamt den generierten Fehlermeldungen.

```
pipe Steelpipe
    length = {[20,35], [40,60], [80,100]};
end

system
    Steelpipe P1(length = [20,30]);// -> ok
    Steelpipe P2(length = [65,75]);// -> ERROR "Empty resulting interval"
    Steelpipe P3(length = [30,50]);// -> WARNING "Impossible values specified"
end
```

Code 1: Auszug einer Systemspezifikation mit drei Rohren unterschiedlicher Länge und der daraus resultierenden Fehlermeldungen.

Im Code wird zunächst der Typ `Steelpipe` definiert, dessen Länge entweder im Intervall $[20, 35]$, $[40, 60]$ oder $[80, 100]$ liegen kann. Anschließend werden im System drei dieser Rohre benutzt und deren Längen bestimmt. Dabei liegt die Länge von P1 komplett im möglichen Intervall und erzeugt somit keinen Fehler. Die Länge des Rohres P2 soll im Intervall $[65, 75]$ liegen, und ist daher mit diesem Typ nicht möglich zu erreichen. Die letzte Länge liegt genau zwischen zwei möglichen Längen, die der Typ erlaubt. Daher wurden sowohl gültige als auch ungültige Werte für die Rohrlänge spezifiziert, was auf einen möglichen Modellierungsfehler hinweist. Daher sollte eine Warnung erzeugt werden.

Das zu generierende lineare Programm wird unter anderem in [Pö15] vorgestellt und wird im Rahmen dieser Arbeit als Beispiel eines Optimierungsproblems verwendet.

4 Forschungsplan

Die folgenden Schritte sind aktuell geplant und werden demnächst umgesetzt. Dabei wird zunächst eine Lösbarkeitsanalyse auf dem domänenspezifischen Modell entwickelt, mit deren Informationen dann versucht wird, das erstellte lineare Programm zu vereinfachen.

4.1 Bestimmung der Lösbarkeit

Ein lineares Programm ist lösbar, wenn es eine Belegung der Variablen gibt, sodass alle Ungleichungen erfüllt sind. In einem Pumpensystem sind die gültigen Belegungen gerade die Betriebspunkte, die die Anlage bedienen kann. Das Berechnen von erreichbaren Betriebspunkten erfordert die Modellierung des Verhaltens der einzelnen Komponenten, um dieses dann über den Systemgraphen zu transportieren. Die Verhaltensfunktionen der Komponenten sowie die Verteilungen der propagierten Werte sind durch die Sprache bzw. das Modell definiert. Erreichbare Zustände müssen abgeschätzt werden. Dazu ist ein Algorithmus in Planung, welcher den Systemgraphen iterativ traversiert und die jeweiligen Betriebspunkte aktualisiert.

Durch die Abschätzung kann angegeben werden, ob die geforderten Werte überhaupt erreicht werden könnten. Ist dies nicht möglich, so muss aus der Analyse ersichtlich sein, welche Komponenten die geforderten Werte nicht unterstützen.

Bei dem entwickelten Verfahren muss per Beweis die Aussagefähigkeit festgestellt werden, also welche Fehler entdeckt werden können und welche nicht. Weiterhin müssen die Fehlermeldungen nach ihrer Verwertbarkeit untersucht werden. Hierzu sollten Testversuche auch durch die endgültigen Anwender erfolgen und ausgewertet werden.

4.2 Bestimmung redundanter Restriktionen und konstanter Variablen

Mit den Informationen über die erreichbaren Zustände können die Verhaltensfunktionen einzelner Komponenten angepasst werden. So muss z. B. nicht abgebildet werden, wie sich eine Pumpe für hohe Durchflüsse verhält, wenn diese dort gar nicht auftreten können. Entsprechende Intervalle sollten angepasst bzw. Stützstellen zur Interpolation weggelassen werden.

Die Untersuchung hierzu ist zweigeteilt. Zum einen kann die Anzahl der entfernten Variablen gezählt werden. Zum anderen muss der Einfluss dieser auf das Lösungsverhalten untersucht werden. Es besteht die Möglichkeit, dass der Löser die nicht benötigten Variablen schneller entdeckt, als dies durch domänenspezifische Verfahren möglich ist.

5 Zusammenfassung

Ziel dieser Arbeit ist die Entwicklung und Untersuchung von Methoden zur Analyse von Optimierungsproblemen, die sich als lineare Probleme formulieren lassen und in einer

domänenspezifischen Sprache vorliegen. Dies geschieht am Beispiel von Pumpensystemen im Rahmen des Projektes "Hocheffiziente Pumpensysteme".

Die zu entwickelnden Methoden sollen das Modell auf Lösbarkeit hin untersuchen und im Falle von Widersprüchen diese konkret durch Fehlermeldungen aufzeigen. Als Zusatz soll untersucht werden, ob durch die ermittelten Informationen das Optimierungsproblem besser formuliert werden kann, also z. B. durch Einsparung ganzzahliger Variablen.

Die zusammengetragenen Methoden sollen beispielhaft implementiert und ausgewertet werden. Dabei soll insbesondere untersucht werden, wie viel Berechnungszeit durch Voranalysen eingespart werden kann.

Falls möglich soll von der aktuellen Domäne abstrahiert werden. Dazu sollen auch andere Domänen betrachtet werden und Gemeinsamkeiten und Unterschiede herausgefunden werden. Die Verfahren können dann helfen, Fehler im Modell frühzeitig zu erkennen.

Literaturverzeichnis

- [FGK93] Fourer, Robert; Gay, David; Kernighan, Brian: *Ampl*, Jgg. 119. Boyd & Fraser, 1993.
- [Ma12] Mazo, Raúl; Salinesi, Camille; Djebbi, Olfa; Diaz, Daniel; Lora-Michiels, Alberto: Constraints: The heart of domain and application engineering in the product lines engineering strategy. *International Journal of Information System Modeling and Design IJISMD*, 3(2):50, 2012.
- [OAV12] Ostrowski, James; Anjos, Miguel F; Vannelli, Anthony: Tight mixed integer linear programming formulations for the unit commitment problem. *IEEE Transactions on Power Systems*, 27(1):39, 2012.
- [Pa13] Padberg, Manfred: *Linear optimization and extensions*, Jgg. 12. Springer Science & Business Media, 2013.
- [PJM12] Perez, Ruben E; Jansen, Peter W; Martins, Joaquim RRA: pyOpt: a Python-based object-oriented framework for nonlinear constrained optimization. *Structural and Multidisciplinary Optimization*, 45(1):101–118, 2012.
- [Pö15] Pöttgen, Philipp; Ederer, Thorsten; Altherr, Lena; Lorenz, Ulf; Pelz, Peter F: Examination and Optimization of a Heating Circuit for Energy-Efficient Buildings. *Energy Technology*, 2015.
- [RVBW06] Rossi, Francesca; Van Beek, Peter; Walsh, Toby: *Handbook of constraint programming*. Elsevier, 2006.
- [SA13] Serali, Hanif D; Adams, Warren P: A reformulation-linearization technique for solving discrete and continuous nonconvex problems, Jgg. 31. Springer Science & Business Media, 2013.
- [Su11] Sujeeth, Arvind; Lee, HyoukJoong; Brown, Kevin; Rompf, Tiark; Chafi, Hassan; Wu, Michael; Atreya, Anand; Odersky, Martin; Olukotun, Kunle: OptiML: an implicitly parallel domain-specific language for machine learning. In: *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. S. 609–616, 2011.

Cardiac activity measurement from video signals of the human skin in ultra-high-field magnetic resonance imaging

Nicolai Spicher^{1,2}

Abstract: Ultra-high-field magnetic resonance imaging (MRI) with field strengths ≥ 7 Tesla comes with many benefits concerning image quality such as increased signal-to-noise ratio (SNR) as well as high spatial and temporal resolution. However, there are still many technical challenges encountered, impeding its full potential. For example, conventional contact-based methods for cardiac monitoring and triggering such as pulse oximetry and electrocardiography are limited by an increased error rate at higher field strengths. In this paper, we give an overview on our works in developing video-based, contact-free, real-time methods to overcome these limitations based on recent findings in remote vital sign measurement.

Keywords: Biomedical engineering, biosignal and image processing, remote vital sign measurement, magnetic resonance imaging, pulse oximetry, photoplethysmogram, real-time applications

1 Motivation and problem statement

MRI is an established medical imaging modality that allows for obtaining an accurate depiction from the inside of the human body without the application of ionizing radiation. If not handled correctly, physiological motion from cardiac activity can reduce image quality significantly. In clinical practice, the conventional way to reduce these effects is by incorporating pulse oximetry (PO) or electrocardiography (ECG) for image acquisition ("triggering") according to the heart beat of the patient: Images are acquired at the same phase of consecutive cardiac cycles and therefore the object in the image (e.g. an artery) is always at an approximately similar position and the image is free of motion artefacts.

ECG uses electrodes attached to the skin for measuring the electrical activity of the heart. In contrast, PO measures the cardiac activity indirectly by detecting changes in blood volume. The probe passes light throughout the finger or earlobe and measures the transmitted light intensity, resulting in a photoplethysmogram (PPG). However, both contact-based approaches are limited by several constraints: The PO probe is susceptible to motion artefacts, its application areas are limited, and during long-time examinations the perfusion of the hands can decrease so far that the signal is lost. ECG is more robust to patient movement, but their placement requires medical staff. In addition, magnetohydrodynamic effects falsify ECG measurement, especially during ultra-high-field MRI [Sn09]. In this context, a 12-lead ECG in combination with independent component analysis [Kr13], and

¹ University of Applied Sciences and Arts Dortmund, Department of Computer Science, Emil-Figge-Str. 42, 44227 Dortmund, nicolai.spicher@fh-dortmund.de

² University Duisburg-Essen, Erwin L. Hahn Institute for Magnetic Resonance Imaging, Kokereiallee 7, 45141 Essen

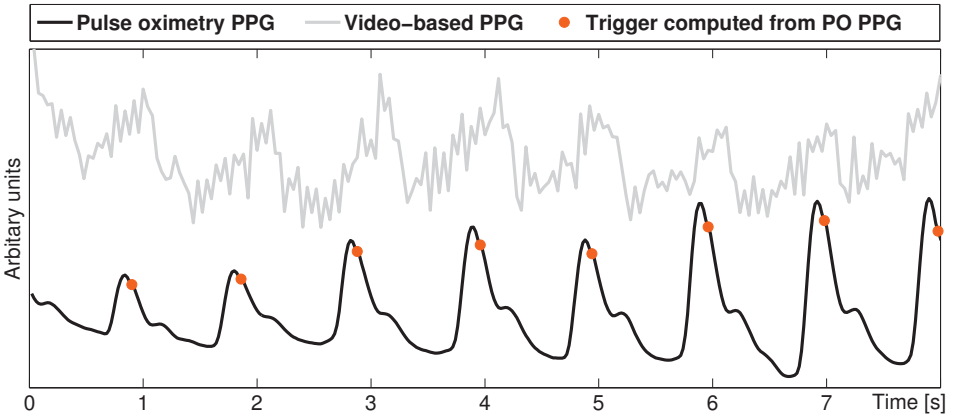


Fig. 1: Cardiac activity measured from a subject using a conventional PO probe (black curve) and a video signal recorded with a MR-compatible camera (gray curve). The latter is obtained by computing the mean pixel intensity in a ROI positioned on a volunteers forehead over time. Red dots mark trigger points that were determined from PO by the MR vendor-provided software and were sent to the scanner for image acquisition.

an optical microphone [Fr10] have been proposed which achieve promising results, but also increase patient preparation time significantly.

2 Problem solving approach

Recent findings in remote vital sign monitoring have shown that the cardiac activity can be estimated from videos of the human skin recorded with video cameras under ambient illumination [ST16]. The measured intensity variations of the skin contain a subtle signal, resulting from dermis deformation caused by transmural arterial pressure changes, which is similar to a PPG measured by PO [Ka15]. Due to the subtleness of this video-based PPG (vPPG), which can not be perceived by naked eye, it is common to increase SNR by examining a spatially pooled region-of-interest (ROI) instead of single pixels. Fig. 1 displays this kind of signal and a PPG obtained in parallel by conventional finger PO. As can be seen, both signals are similar although the SNR of the vPPG is significantly lower.

We pursue to transfer this video-based and contact-free technique for cardiac activity estimation and apply it to subjects undergoing ultra-high-field MRI examinations with the aim to overcome the described limitations of contact-based hardware. However, many methods from literature cannot be used without adjustments in this context because (1) MR-compatible cameras have a significantly lower performance than most off-the-shelf digital cameras, (2) real-time processing of the video signal is required, (3) patients undergoing examinations are potentially not cooperative or not able to avoid movement, and (4) the illumination conditions inside the MR bore are poor which additionally reduces video quality.

3 Related work

To the knowledge of the authors, in 2005 Wieringa et al. were the first that acquired a remote vPPG using a camera and designated LED illumination [WMv05]. Subsequently, Takano [TO06] and Verkruyse [VSN08] were the first that by made use of this signal for vital sign monitoring with ambient illumination only by analyzing pooled pixel intensities in a ROI over time and performing spectral analysis for heart rate (HR) measurement. Considering the visualization of the vPPG signal, Kamshilin et al. proposed an algorithm for the pseudo-colors visualization of the blood perfusion [Ka11] and Wu et al. proposed an algorithm called "Eulerian Video Magnification" (EVM) for the amplification of subtle motions in videos that can be used to photo-realistically visualize the flow of blood [Wu12]. Recently, elaborate algorithms for HR estimation from vPPG signals recorded with color cameras in fitness [dJ13] or office settings [MGP14, WSd15] have been presented.

Due to the highly specialised character of applying this methods in MRI, the body of research is rather small in this field. Yang et al. estimated the HR of a subject in a mock MRI scanner but applied a conventional digital camera and did not measure a ground truth modality as reference [Ya14]. Maclaren et al. applied an MR-compatible camera inside the MR bore with additional illumination from an LED. They processed the obtained videos offline by using frequency filtering in order to obtain the vPPG signal as well as respiratory information [MAB15] [Ma14]. Pulse oximetry and a respiratory belt were used as ground truth and closely resembled the filtered vPPG signals.

4 Research plan

In the following, we denote a signal, obtained over time t by mean pixel computation in a ROI centered on human skin, $vPPG$. It exhibits subtle color intensity variations that can be roughly approximated as a cosine wave with frequency f , associated to the HR, and phase ϕ . This sinusoid allows to approximate the stages of the cardiac cycle from the outflow of blood (low mean value in ROI) to the influx of blood (high mean value), and the reoccurring outflow of blood (low mean value). However, due to the subtleness of this signal, it is considerably degraded by noise $N(t)$ introduced by the patient (e.g. motion artefacts) and noise $M(t)$ introduced by the limitations of the used camera:

$$vPPG(t) \approx \cos(2\pi ft + \phi) + N(t) + M(t) \quad (1)$$

We seek to estimate f and ϕ and use them for two tasks: HR monitoring of subjects undergoing MRI examination and computation of trigger points for MR image acquisition.

5 Preliminary results and current status

Inspired by the works from Wu et al. [Wu12], initial experiments began in 2014. Since then, the experimental set-up as well as the algorithms have been revised extensively:

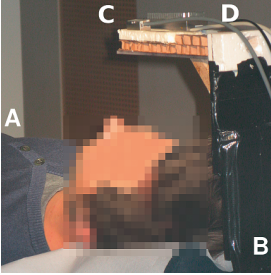


Fig. 2: Current camera set-up applied to a volunteer (A) outside the MR bore with the patient table in home position. The custom-build stand (B) holds the low-speed (C) and high-speed (D) MR-compatible cameras. The set-up can be applied inside the bore without modification.

Set-up As initially no MR-compatible camera was available at the 7T MRI site, preliminary experiments were conducted using an off-the-shelf camera (RGB, 640x480, 30 FPS) outside the 7T scanner bore and using a MR-compatible camera (B/W, 720x576, 25 FPS) during a routine 3T MRI examination on a volunteer. In both cases, the ROI was centered on a the volunteers' finger and no particular illumination instead of room lighting was used [Sp14]. We observed that (1) the SNR of $vPPG$ is rather low on the finger, (2) the low illumination inside the MR bore reduces SNR additionally, and (3) the subjects tend to move the finger.

Hence, when the MR-compatible camera (B/W, 720x576, 25 FPS) was available at the 7T site, we built a stand that was used to install the camera above the well-perfused subject's head. Additionally, a video projector was placed at the end of the MR bore and used to provide illumination [Sp15a] [Sp15b]. Recently, we added a high-speed MR-compatible camera prototype (up to 1076 FPS) to the set-up which allows us to evaluate the performance increase [Sp16]. Fig. 2 shows the current experimental set-up.

Algorithm Since beginning of the project, algorithms were developed using C++11 as glue code, OpenCV¹ for image processing, and ROOT² for mathematical computations.

We began our research using the EVM algorithm, which allows to magnify a certain frequency range in videos [Wu12], as a starting point. We developed a real-time feasible implementation that first estimates the subjects' HR frequency and then magnifies it in the videos using EVM. Videos were processed by our implementation, the $vPPG$ signals obtained from the magnified videos were compared to PO PPG acquired with the MR-compatible probe and we observed that both signals correlated well [Sp14]. However, our further research revealed that the EVM algorithm is occasionally vulnerable to Gibbs phenomenon when using short video signals which prohibits to obtain accurate results for our application that depends on current cardiac information.

Therefore, we decided to develop a more lightweight algorithm: A real-time algorithm processing raw $vPPG$ signals was developed that estimates f of the subject by detecting the peak in the Fourier spectrum associated with the cardiac activity and then estimates ϕ of this component for trigger point computation [Sp15a] [Sp15b].

¹ <http://www.opencv.org/>

² <https://root.cern.ch/>

Results Using this algorithm based on spectral analysis and the set-up shown in Fig. 2, we first conducted a study with eight subjects in- and outside the MR bore using the algorithm for HR estimation based on f . We compared its performance to ECG and PO which suggests that the video-based approach is feasible but still inferior in accuracy compared to contact-based methods. As results outside the bore achieved significantly better results, we assumed that the illumination conditions inside the bore still pose a challenge for accurate $vPPG$ acquisition although we increased illumination by using a video projector [Sp15b].

For one volunteer, we used our algorithm for trigger point computation based on ϕ and showed that our approach outperforms PO triggering in case of PO interference caused by gradient vibrations during 7T MRI. However, $vPPG$ SNR decreases during head motion and requires additional measures (e.g. motion correction) to increase performance [Sp15a]. Recently, we applied the high-speed camera and investigated how accurately one can approximate the PO PPG from $vPPG$. We observed that using a simple filtering technique, physiological peaks that are clearly visible in PO PPG (Fig. 1 low amplitude peaks following the high amplitude peaks) but not in the raw $vPPG$ can be made visible [Sp16].

Our current aim is to apply our video-based MRI triggering technique to a larger group of subjects and compare its performance to contact-based methods.

Acknowledgments

This project is performed in collaboration with *Prof. Dr. Markus Kukuk*¹, *Dr. Stefan Maderwald*², and *Prof. Dr. Mark E. Ladd*^{2,3}, and the author thanks all three of them for fruitful discussions and their valuable advices.

References

- [dJ13] de Haan, Gerard; Jeanne, Vincent: Robust Pulse Rate From Chrominance-Based rPPG. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013.
- [Fr10] Frauenrath, Tobias; Hezel, Fabian; Renz, Wolfgang; de Geyer d’Orth, Thibaut; Dieringer, Matthias; von Knobelsdorff-Brenkenhoff, Florian; Prothmann, Marcel; Schulz Menger, Jeanette; Niendorf, Thoralf: Acoustic cardiac triggering: a practical solution for synchronization and gating of cardiovascular magnetic resonance at 7 Tesla. *Journal of Cardiovascular Magnetic Resonance*, 12(67), 2010.
- [Ka11] Kamshilin, Alexei A.; Miridonov, Serguei; Teplov, Victor; Saarenheimo, Riku; Nippolainen, Ervin: Photoplethysmographic imaging of high spatial resolution. *Biomedical Optical Express*, 2(4):996–1006, 2011.
- [Ka15] Kamshilin, Alexei A.; Nippolainen, Ervin; Sidorov, Igor S.; Vasilev, Petr V.; Erofeev, Nikolai; Podolian, Natalia P.; Romashko, Roman: A new look at the essence of the imaging photoplethysmography. *Scientific Reports*, 5(10494), 2015.
- [Kr13] Krug, Johannes W.; Rose, Georg; Clifford, Gari D.; Oster, Julien: ECG-based gating in ultra high field cardiovascular magnetic resonance using an independent component analysis approach. *Journal of Cardiovascular Magnetic Resonance*, 15(104), 2013.

³ Deutsches Krebsforschungszentrum (German Cancer Research Center, DKFZ), Division of Medical Physics in Radiology, Im Neuenheimer Feld 280, 69120 Heidelberg

- [Ma14] Maclaren, Julian; Aksoy, Murat; Ehrl, Jakob; Saranathan, Manojkumar; Bammer, Roland: Simultaneous monitoring of cardiac and respiratory signals using a markerless optical system. In: Proceedings of the 22nd Annual Meeting of the ISMRM. 2014.
- [MAB15] Maclaren, Julian; Aksoy, Murat; Bammer, Roland: Contact-Free Physiological Monitoring Using a Markerless Optical System. *Magnetic Resonance in Medicine*, 74:571–577, 2015.
- [MGP14] McDuff, Daniel J.; Gontarekand, Sarah; Picard, Rosalind W.: Improvements in Remote Cardiopulmonary Measurement Using a Five Band Digital Camera. *IEEE Transactions on Biomedical Engineering*, 61(10):2593–2601, 2014.
- [Sn09] Snyder, Carl J.; DelaBarre, Lance; Metzger, Gregory J.; van de Moortele, Pierre-Francois; Akgun, Can; Ugurbil, Kamil; Vaughan, John T.: Initial Results of Cardiac Imaging at 7T. *Magnetic Resonance in Medicine*, 61(3):517–524, 2009.
- [Sp14] Spicher, Nicolai; Brumann, Christopher; Kukuk, Markus; Ladd, Mark E.; Maderwald, Stefan: Eulerian Video Magnification for Heart Pulse Measurement in MRI Scanners. In: Proceedings of the 22nd Annual Meeting of the ISMRM. 2014.
- [Sp15a] Spicher, Nicolai; Kukuk, Markus; Ladd, Mark E.; Maderwald, Stefan: In vivo 7T MR imaging triggered by phase information obtained from video signals of the human skin. In: Proceedings of the 23rd Annual Meeting of the ISMRM. 2015.
- [Sp15b] Spicher, Nicolai; Maderwald, Stefan; Ladd, Mark E.; Kukuk, Markus: Heart rate monitoring in ultra-high-field MRI using frequency information obtained from video signals of the human skin compared to electrocardiography and pulse oximetry. In: *Current Directions in Biomedical Engineering*. volume 1, pp. 69–72, 2015.
- [Sp16] Spicher, Nicolai; Maderwald, Stefan; Ladd, Mark E.; Kukuk, Markus: High-speed, contact-free measurement of the photoplethysmography waveform for MRI triggering. In: Proceedings of the 24th Annual Meeting of the ISMRM. 2016.
- [ST16] Sun, Yu; Thakor, Nitish: Photoplethysmography Revisited: From Contact to Noncontact, From Point to Imaging. *IEEE Transactions on Biomedical Engineering*, 63(3):463–477, 2016.
- [TO06] Takano, Chihiro; Ohta, Yuji: Heart rate measurement based on a time-lapse image. *Medical Engineering & Physics*, 29:853–857, 2006.
- [VSN08] Verkruyse, Wim; Svaasand, Lars O.; Nelson, J. Stuart: Remote plethysmographic imaging using ambient light. *Optics Express*, 16(26):21434–21445, 2008.
- [WMv05] Wieringa, Fokko P.; Mastik, Frits; van der Steen, Antonius F. W.: Contactless multiple wavelength photoplethysmographic imaging: a first step toward "SpO2 camera" technology. *Annals of Biomedical Engineering*, 33:1034–41, 2005.
- [WSd15] Wang, Wenjin; Stuijk, Sander; de Haan, Gerard: Exploiting Spatial Redundancy of Image Sensor for Motion Robust rPPG. *IEEE Transactions on Biomedical Engineering*, 62(2):415–425, 2015.
- [Wu12] Wu, Hao-Yu; Rubinstein, Michael; Shih, Eugene; Guttag, John; Durand, Frédo; Freeman, William T.: Eulerian Video Magnification for Revealing Subtle Changes in the World. *ACM Transactions on Graphics*, 31(4), 2012.
- [Ya14] Yang, Shang-Yi; Huang, Hsiao-Hui; Liang, Chi-Wei; Tsai, Shang-Yueh; Huang, Teng-Yi: Noncontact physiological measurements using video recording inside an MRI scanner. In: Proceedings of the 22nd Annual Meeting of the ISMRM. 2014.

Statistisches Lernen wortbasierter Morphologie

Maciej Sumalvico¹

Abstract: Im Rahmen der Promotion wird ein Ansatz zum automatischen Lernen von Sprachmorphologie entwickelt. Dabei wird auf die Zerlegung von Wörtern in kleinere strukturelle Elemente (die s.g. *Morpheme*) verzichtet und stattdessen mit Transformationsregeln gearbeitet, die an ganzen Wörtern operieren. Das Lernen wird mithilfe eines probabilistischen Modelles realisiert, das mit dem EM-Algorithmus trainiert wird. Die gelernten morphologischen Regeln können bei verschiedenen praktischen Problemen der Sprachtechnologie angewandt werden, um den Umgang mit unbekanntem Wörtern zu verbessern.

Keywords: Natural Language Processing, NLP, Sprachtechnologie, Grammatik, Morphologie, maschinelles Lernen, statistische Modellierung.

1 Einleitung

Die Morphologie ist ein Bereich der Sprachgrammatik, der strukturelle Zusammenhänge zwischen Wörtern beschreibt. Dabei wird die *morphologische Analyse* eines Wortes üblicherweise als die Zerlegung in minimale bedeutungs- oder funktionstragende Einheiten, die s.g. *Morpheme*, verstanden [AF11]. Zum Beispiel besteht das im Titel dieses Artikels enthaltene Wort *wortbasierter* aus folgenden Morphemen: *wort-bas-ier-t-er*. Darunter sind neben den Morphemen, die für die Kernbedeutung verantwortlich sind (*wort*, *bas*), auch zwei derivationale Affixe (*-ier*, *-t*) zu finden, sowie die Flexionsendung *-er*.

In der automatischen Sprachverarbeitung ist ein morphologischer Analysierer ein häufiger Bestandteil von Prozessketten. Da es unmöglich ist, ein vollständiges Lexikon von Wörtern einer Sprache zu erstellen, müssen die Systeme der Sprachtechnologie immer auch mit unbekanntem Wörtern umgehen können. Durch morphologische Analyse können die Merkmale eines unbekanntem Wortes durch seine strukturelle Beziehung zu bekannten Wörtern erraten werden.

Üblicherweise werden morphologische Analysierer als handgeschriebene Grammatiken und Lexika erstellt und anschließend zu endlichen Automaten kompiliert, die eine performante Durchsuchung ermöglichen. Dieser Ansatz, bekannt als *Zwei-Ebenen-Morphologie*, geht auf die Arbeit von Koskenniemi [Ko83] zurück. Formalismen wie XFST [BK03] ermöglichen die Kompilierung von komplexen Grammatiken und die Optimierung von resultierenden Automaten. Dieser Ansatz zeichnet sich zwar durch hohe Effizienz und Genauigkeit aus, erfordert aber erheblichen Aufwand bei der manuellen Erstellung der Grammatiken und Lexika. Deshalb sind die Methoden des automatischen Lernens von Morphologie ein aktuelles Forschungsthema.

¹ Universität Leipzig, Abteilung Automatische Sprachverarbeitung, Augustusplatz 10, 04109 Leipzig, sumalvico@informatik.uni-leipzig.de

Obwohl fast alle vorhandenen Ansätze zum Lernen von Morphologie auf die Zerlegung von Wörtern in Morpheme zielen [HB11], ist das eigentliche Ziel die Vorhersage von Merkmalen unbekannter Wörter. Die Zerlegung ist dafür weder ausreichend noch notwendig. Darüber hinaus wurde der Begriff des Morphems auch in der Linguistik kritisiert, was zur Formulierung von „morphemlosen“ Morphologietheorien geführt hat [An92, FSM97]. Manche Kritikpunkte sind auch für das automatische Lernen von Morphologie relevant: z.B. hat die morphembasierte Analyse Schwierigkeiten mit nichtkonkatenativen Operationen (*Haus:Häuser*), sie führt Entitäten ein, die in der Oberflächenform des Wortes nicht zu erkennen sind („Null-Affixe“), die Morphemgrenzen werden durch phonologische Prozesse verwischt, der Sinn der Zerlegung von manchen abgeleiteten Wörtern ist durch verwischte Beziehung zum Grundwort (z.B. *gehören* < *ge-hören*) oder sogar das Fehlen vom Grundwort (*vergessen* < *ver-*gessen*) fraglich. Aus diesen Gründen stellt sich die hier vorgestellte Promotion das Ziel, eine Methode für das automatische Lernen von Morphologie zu entwickeln, die direkt auf die Ableitung neuer Wörter und ihrer Eigenschaften zielt, ohne die Wörter in Morpheme zu zerlegen. Dabei wird insbesondere die *Whole Word Morphology* von [FSM97] als linguistische Grundlage verwendet.

2 Stand der Forschung und aktuelle Trends

Die Aufgabe des automatischen Lernens von Morphologie ist in der Literatur seit langem bekannt: Der erste Ansatz wurde bereits in den 50er Jahren vorgestellt [Ha55]. Lange wurde die Aufgabe vor allem durch heuristische Ansätze gelöst, wie die von Harris eingeführte *Letter Successor Variety* (LSV) [Go06, Bo08] oder die Entdeckung von morphologisch verwandten Wörtern mittels verschiedener Ähnlichkeitsmaße [YW00, BMT02, Ki13]. Ein heuristischer Ansatz, der auf Ganzwortmorphologie (*Whole Word Morphology*) basiert, wurde von Neuvel und Fulop [NF02] vorgeschlagen. Für das überwachte Lernen wurden hingegen die üblichen Methoden des maschinellen Lernens verwendet, wie u.a. Conditional Random Fields [Ru13], Pair Hidden Markov Models [Cl02] oder Maximum Entropy Classifiers [CDvG08].

In den letzten Jahren gewinnen probabilistische Modelle auch beim unüberwachten Lernen immer mehr an Bedeutung [CL05, PCT09, Ca11]. Ihr großer Vorteil ist, dass das gleiche Modell oft sowohl überwacht, als auch unüberwacht trainiert werden kann. Eine zweite wichtige Entwicklung ist die Einbeziehung vom Wortkontext in das Lernen von Morphologie. In früheren Arbeiten wurde dies durch kookkurenzbasierte Ähnlichkeitsmaße erreicht [BMT02, Bo08]. In der letzten Zeit ist mit der Veröffentlichung von *word2vec* [Mi13, MYZ13] ein mächtiges Werkzeug für die numerische Beschreibung von Wortbedeutung verfügbar geworden. Seine Nützlichkeit bei der Aufgabe des Morphologielernens wurde bereits nachgewiesen [SO15].

3 Ein probabilistisches Modell für die Ganzwortmorphologie

Im Rahmen der hier vorgestellten Promotion wird ein generatives probabilistisches Modell vom Sprachlexikon entwickelt, in dem morphologische Zusammenhänge zwischen Wör-

tern mittels Transformationsregeln, die an ganzen Wörtern operieren, ausgedrückt werden. Zum Beispiel wird die Regel, die aus *Haus Häuser* ableitet, folgendermaßen dargestellt:²

$$/X_1 a X_2 /_{N.SG} \rightarrow /X_1 ä X_2 er /_{N.PL} \tag{1}$$

Die Objekte innerhalb von $/ \cdot /$ sind ganze, existierende Wörter. X_1 und X_2 sind variable Elementen, die mit einer beliebigen Kette von Phonemen (bzw. Buchstaben) instantiiert werden können und von der Regel unverändert bleiben. Zusätzlich zur phonologischen oder orthographischen Repräsentation von Wörtern kann die Regel auch an verschiedenen anderen Merkmalen operieren, wie syntaktische Merkmale (Wortart, Flexionsmerkmale) oder Bedeutung.

Das probabilistische Modell nach aktuellem Stand wurde in [Ja15] detailliert vorgestellt. Es betrachtet Lexika als Graphen (genauer: gerichtete Wälder) von Wörtern mit Ableitungsrelation und schreibt ihnen Wahrscheinlichkeiten zu (Abb. 1). Die Wahrscheinlichkeit eines Wortes, das keine eingehende Kante besitzt, wird aus einer unigrammbasierten Verteilung über alle Buchstabenketten ausgerechnet (Produkt von Häufigkeiten von Buchstaben). Sie ist typischerweise sehr klein und nimmt mit der Länge des Wortes stark ab. Die Wahrscheinlichkeit von Wörtern, die durch Regeln abgeleitet werden (d.h. eine eingehende Kante besitzen), wird für jede Regel festgelegt und als Modellparameter behandelt. So erhalten diejenigen Lexika eine hohe Wahrscheinlichkeit, die oft wiederkehrende strukturelle Zusammenhänge nutzen, um Wörter abzuleiten und die Anzahl von Wurzelknoten zu verringern.

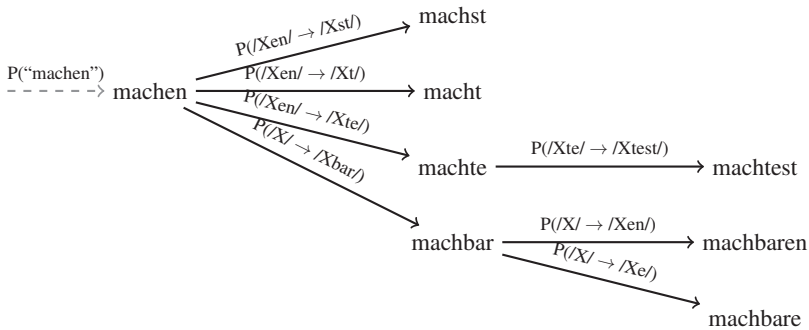


Abb. 1: Fragment eines möglichen deutschen Lexikons.

Das Modell wird im Paradigma der bayesschen Statistik erstellt: Die beobachteten Daten sind das Vokabular V und das zu optimierende Parameter die Regelmenge R . Bei der Berechnung von Wahrscheinlichkeiten spielt auch die unbekannte Menge von Kanten E eine Rolle. Der zu optimierende Wert ist die a-posteriori Wahrscheinlichkeit der Regelmenge

² In der ursprünglichen Theorie von [FSM97] sind die Regeln immer bidirektional: sie drücken keine Ableitung aus, sondern eine ungerichtete Relation. In dieser Arbeit wird mit Hinblick auf das probabilistische Modell eine Ableitungsrichtung eingeführt.

gegeben Daten: $P(R|V)$, die folgendermaßen transformiert wird:

$$\begin{aligned}\arg \max_R P(R|V) &= \arg \max_R \frac{P(V|R)P(R)}{P(V)} = \arg \max_R P(V|R)P(R) \\ &= \arg \max_R \sum_E P(V,E|R)P(R)\end{aligned}\tag{2}$$

Neben der Wahrscheinlichkeit von Graphen $P(V,E|R)$, deren Berechnung auf der Abb. 1 schematisch gezeigt wird, wird also auch die Komplexität der gelernten Grammatik durch die a-priori Wahrscheinlichkeit $P(R)$ kontrolliert. Für die Suche nach der optimalen Menge von Regeln können zwei verschiedene Varianten des Expectation-Maximization-Algorithmus [DLR77, SCR12] angewandt werden: entweder wird abwechselnd der Graph und die Regelmenge optimiert („hard EM“), oder aber wird nach einer Regelmenge gesucht, die den Erwartungswert des Log-Likelihoods über alle möglichen Kantenstrukturen maximiert („soft EM“). Die letztere Methode ist im Moment noch in der Entwicklungsphase. Es wird geplant, für die Berechnung von Erwartungswerten Markov Chain Monte Carlo (MCMC) Methoden zu verwenden [RC05]. Das überwachte Trainieren vom Modell ist ebenfalls möglich – in dem Fall wird die Suche nach der optimalen Regelmenge vereinfacht, da die Kantenmenge E bekannt ist.

Der Formalismus lässt sich leicht erweitern, indem Wörter nicht nur als Zeichenketten betrachtet werden, sondern als Vektoren von Merkmalen, die außer orthographischer Repräsentation auch andere Informationen enthalten. Dafür müssen lediglich Wahrscheinlichkeitsverteilungen für die Merkmalswerte der Wurzelwörter und die von Regeln durchgeführten Transformationen festgelegt werden. Zum Beispiel hat sich Worthäufigkeit als ein nützliches Merkmal erwiesen. Das am vielversprechendste Merkmal sind die von `word2vec` berechneten Vektoren, die den Kontext des Wortes – und damit seine Bedeutung – als Koordinaten in einem kontinuierlichen Raum erfassen. Ihre Integration in das Modell wird für die nahe Zukunft geplant.

4 Arbeitsplan

Das Projekt ist auf drei Jahre angelegt. Der Arbeitsplan ist in drei Phasen gegliedert, die jeweils ein Jahr dauern:

Jahr 1. Entwurf des probabilistischen Modells, Auswahl von Methoden und Werkzeugen, Prototyp-Implementierung, erste Auswertungsexperimente. Diese Phase ist fast abgeschlossen.

Jahr 2. Tiefere Einsicht in die verwendeten Algorithmen (z.B. Konvergenzanalyse von MCMC-Sampling), Erweiterung des Modells um weitere Merkmale und Kompositionsregeln. Performanzoptimierung und Fehleranalyse.

Jahr 3. Anwendung und Evaluierung des Modells bei praktischen Aufgaben der Sprachtechnologie: OCR-Nachkorrektur, Lemmatisierung, Tagging. Untersuchung zu weiteren Anwendungsmöglichkeiten. Umfassende Evaluierung auf Daten aus verschiedenen Sprachen. Veröffentlichung der Dissertation.

Finanzierung

Die Promotion wird aus den Mitteln des Europäischen Sozialfonds (ESF) der Europäischen Union (EU) und des Freistaates Sachsen im Rahmen des ESF-Projekts Nr. 100234741 „Landesinnovationspromotionen“ finanziert. Projektzeitraum: 01.09.2015–31.08.2018.

Literaturverzeichnis

- [AF11] Aronoff, Mark; Fudeman, Kirsten Anne: What is morphology? Fundamentals of Linguistics. Wiley, 2011.
- [An92] Anderson, Stephen R.: A-Morphous Morphology. 1992.
- [BK03] Beesley, Kenneth R.; Karttunen, Lauri: Finite State Morphology. Center for the Study of Language and Information, 2003.
- [BMT02] Baroni, Marco; Matiasek, Johannes; Trost, Harald: Unsupervised discovery of morphologically related words based on orthographic and semantic similarity. In: Proceedings of the 6th Workshop of the ACL Special Interest Group on Phonology. Jgg. 6, S. 48–57, 2002.
- [Bo08] Bordag, Stefan: Unsupervised and Knowledge-free morpheme segmentation and analysis. Lecture Notes in Computer Science, 5152 LNCS:881–891, 2008.
- [Ca11] Can, Burcu: Statistical Models for Unsupervised Learning of Morphology and POS Tagging. Dissertation, University of York, 2011.
- [CDvG08] Chrupała, Grzegorz; Dinu, Georgiana; van Genabith, Josef: Learning morphology with morfette. In: Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC '08. S. 2362–2367, 2008.
- [CI02] Clark, Alexander: Memory-Based Learning of Morphology with Stochastic Transducers. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL). S. 513–520, 2002.
- [CL05] Creutz, Mathias; Lagus, Krista: Unsupervised Morpheme Segmentation and Morphology Induction from Text Corpora Using Morfessor 1.0. Bericht, 2005.
- [DLR77] Dempster, A. P.; Laird, N. M.; Rubin, Donald B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, 39(1):1–38, 1977.
- [FSM97] Ford, Alan; Singh, Rajendra; Martohardjono, Gita: Pace Pāṇini: Towards a word-based theory of morphology. American University Studies. Series XIII, Linguistics, Vol. 34. Peter Lang Publishing, Incorporated, 1997.
- [Go06] Goldsmith, John: An algorithm for the unsupervised learning of morphology. Natural Language Engineering, 12(1):353, 2006.

- [Ha55] Harris, Zellig S.: From phoneme to morpheme, 1955.
- [HB11] Hammarström, Harald; Borin, Lars: Unsupervised Learning of Morphology. *Computational Linguistics*, 37(2):309–350, 2011.
- [Ja15] Janicki, Maciej: A Multi-purpose Bayesian Model for Word-Based Morphology. In (Mahlow, Cerstin; Piotrowski, Michael, Hrsg.): *Systems and Frameworks for Computational Morphology – Fourth International Workshop, SFCM 2015*. Springer, 2015.
- [Ki13] Kirschenbaum, Amit: Unsupervised Segmentation for Different Types of Morphological Processes Using Multiple Sequence Alignment. In: *1st International Conference on Statistical Language and Speech Processing, SLSP*. Tarragona, Spain, S. 152–163, 2013.
- [Ko83] Koskenniemi, Kimmo: *Two-Level Morphology: A General Computational Model for Word-Form Recognition and Production*. Dissertation, University of Helsinki, 1983.
- [Mi13] Mikolov, Tomas; Corrado, Greg; Chen, Kai; Dean, Jeffrey: Efficient Estimation of Word Representations in Vector Space. In: *Proceedings of the International Conference on Learning Representations (ICLR 2013)*. S. 1–12, 2013.
- [MYZ13] Mikolov, Tomas; Yih, Wen-tau; Zweig, Geoffrey: Linguistic regularities in continuous space word representations. In: *Proceedings of NAACL-HLT*. S. 746–751, 2013.
- [NF02] Neuvel, Sylvain; Fulop, Sean A.: Unsupervised Learning of Morphology without Morphemes. In: *Proceedings of the 6th Workshop of the ACL Special Interest Group in Computational Phonology (SIGPHON)*. S. 31–40, 2002.
- [PCT09] Poon, Hoifung; Cherry, Colin; Toutanova, Kristina: Unsupervised morphological segmentation with log-linear models. June, S. 209, 2009.
- [RC05] Robert, Christian P.; Casella, George: *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Springer-Verlag New York, Inc., 2005.
- [Ru13] Ruokolainen, Teemu; Kohonen, Oskar; Virpioja, Sami; Kurimo, Mikko: Supervised Morphological Segmentation in a Low-Resource Learning Setting using Conditional Random Fields. In: *Proceedings of the Seventeenth Conference on Computational Natural Language Learning (CoNLL)*. Sofia, Bulgaria, S. 29–37, 2013.
- [SCR12] Samdani, Rajhans; Chang, Ming-Wei; Roth, Dan: Unified Expectation Maximization. In: *2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. S. 688–698, 2012.
- [SO15] Soricut, Radu; Och, Franz Josef: Unsupervised Morphology Induction Using Word Embeddings. In: *NAACL 2015*. S. 1626–1636, 2015.
- [YW00] Yarowsky, David; Wicentowski, Richard: Minimally Supervised Morphological Analysis by Multimodal Alignment. In: *ACL '00*. S. 207–216, 2000.

Bestimmung von Erfolgsfaktoren der Unternehmenskommunikation von KMU auf Facebook: Design und Zwischenergebnisse eines Multimethoden-Ansatzes

Malte Wattenberg¹

Abstract: Der Beitrag skizziert ein laufendes Dissertationsprojekt zum Thema Erfolgsfaktoren der Unternehmenskommunikation von kleinen und mittelständischen Unternehmen (KMU) in Social Media am Beispiel Facebook. Dazu werden die Notwendigkeit und der Einsatz eines Methodenmix aus qualitativen und quantitativen Forschungsmethoden aus der Perspektive der Nutzer und Unternehmen vorgestellt sowie Ergebnisse abgeschlossener Teilstudien diskutiert. Anschließend wird ein Ausblick auf zukünftige Arbeiten zum Abschluss des Projekts gegeben.

Keywords: Social Media, Social Network Sites, SNS, Facebook, Social Media Kommunikation, Erfolgsfaktoren in Social Media, erfolgreiche Unternehmenskommunikation auf Facebook, Kommunikation von KMU auf Facebook

1 Einleitung

Der Einsatz von Social Media in Unternehmen erfreut sich nach wie vor großer Beliebtheit. Laut einer repräsentativen Umfrage des Branchenverbandes BITKOM [BI15] nutzen 75% aller deutschen Unternehmen Social Media Kanäle für ihre externe oder interne Kommunikation. Hauptgründe der Unternehmen im Bereich B2C sind die Nutzung als Marketingplattform, engerer Kontakt zum Kunden und die Erhöhung des Bekanntheitsgrades [BB15]. Facebook stellt dabei mit knapp 1,7 Mrd. monatlich aktiven Nutzern weltweit im 1. Quartal 2016 [FB16] das größte Soziale Netzwerk (SNS) dar und bietet demnach Unternehmen ein nach wie vor hohes Potenzial als Kommunikationsinstrument.

Der Begriff Unternehmenskommunikation umfasst dabei nach Zerfaß „alle Kommunikationsprozesse, mit denen ein Beitrag zur Aufgabendefinition und -erfüllung in gewinnorientierten Wirtschaftseinheiten geleistet wird und die insbesondere zur internen und externen Handlungskoordination sowie Interessenklärung zwischen Unternehmen und ihren Bezugsgruppen (Stakeholdern) beitragen“ [Za14]. Weiter kann unterschieden werden in interne Kommunikation, transaktionsorientierter Marktkommunikation und interaktionsorientierten Public Relations zum Beziehungsaufbau bzw. -pflege [Za14, Ma13, MHG05]. Meffert, Burmann und Kirchgeorg hingegen unterteilen in die interne und externe Markenkommunikation und ordnen sowohl Social Media Kommunikation als

¹ Fachhochschule Bielefeld, Fachbereich Wirtschaft und Gesundheit - Schwerpunkt Wirtschaftsinformatik, Interaktion 1, 33619 Bielefeld, malte.wattenberg@fh-bielefeld.de, malte.wattenberg@t-online.de

auch Public Relations der Kommunikationspolitik im klassischen Marketing-Mix der 4P's zu [MBK15]. Im Rahmen dieser Forschungsarbeit wird Unternehmenskommunikation als ganzheitlich marktorientierte Kommunikation begriffen, die neben den wirtschaftlichen und sozialen auch die strukturellen Aspekte der Plattform berücksichtigt.

Soziale Medien stellen durch die neuen Vernetzungsmöglichkeiten eine sehr starke Herausforderung für die Kommunikationsarbeit von Unternehmen dar, da zwar neue Chancen durch das Erhalten von Feedback, der direkten Kundenansprache und der Verbreitung von Informationen entstehen, gleichzeitig aber auch Risiken durch die Eigendynamik der Kommunikationsvorgänge, der hohen Dialogintensität sowie der Öffnung der Kommunikationsräume und dem damit einhergehenden Kontrollverlust [Ma13, As14].

2 Fragestellung und Zielsetzung

Die Einordnung von Social Media in den organisationalen Unternehmenskontext erfolgt hinsichtlich der inhaltlichen Betreuung der verschiedenen Kanäle häufig im Bereich Marketing. Begleitend von Corporate Social Media Guidelines und Redaktionsplänen werden Inhalte von idealerweise fachübergreifenden Teams anhand der festgelegten strategischen Zielsetzung erstellt. Insbesondere in KMU stehen jedoch die dafür notwendigen personellen als auch finanziellen Ressourcen nicht zur Verfügung. Es stellen sich folgende Fragen:

Wie können KMU in Sozialen Netzwerken erfolgreich kommunizieren? Wann führen also Beiträge im unternehmerischen Sinn zu relevanten Anschlüssen und welche Faktoren außerhalb der Beiträge gilt es zu beachten?

Bisherige Studien betrachten aus Anwendersicht dabei insbesondere die Motive, das Sozialkapital, die Selbstdarstellung und Privatsphäre der Nutzer [Wi12, WL11]. Im Businessbereich stehen primär Großunternehmen mit Darstellungen von Best-Practices sowie statistische Auswertungen im Fokus.

Zielsetzung dieser Forschungsarbeit ist nun die Beantwortung der Fragen durch den Einsatz eines Methodenmix, der im folgenden Abschnitt erläutert wird.

3 Lösungsansatz

Lösungsansätze zu komplexen betriebswirtschaftlichen Fragestellungen, wie im Bereich der Unternehmenskommunikation in Social Media, benötigen einerseits den Einbezug relevanter Schwerpunkte wie die Wirtschaftsinformatik und das (Online)Marketing, andererseits Erkenntnisse aus den Schnittstellen ergänzender Fachgebiete wie die Linguistik bzw. Kommunikationswissenschaften. Als Ergebnis wird deutlich, dass auch in Hinblick auf die Forschungsmethodik ein umfassender Methodenmix aus qualitativen

und quantitativen Erhebungsphasen Anwendung finden muss [Br11]. Die Erkenntnisse einer Methode beschränken sich so nicht auf einzelne Forschungsfragen, sondern können kritisch mit einer anderen Methode hinterfragt und kombiniert werden [Br11]. Die folgende Abbildung gibt einen Überblick über das dieser Arbeit zugrundeliegende Forschungsdesign:

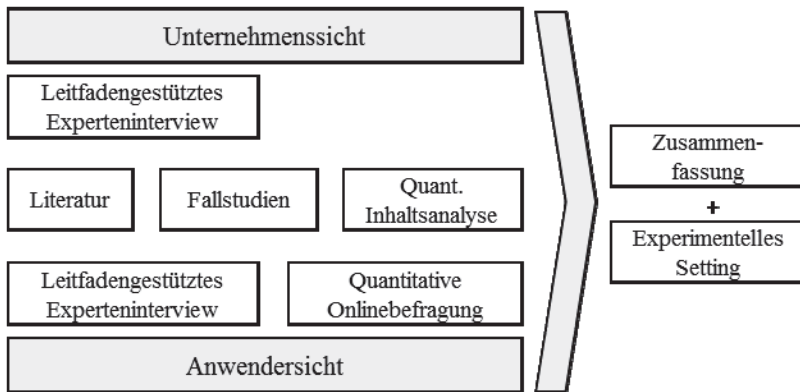


Abb. 1: Forschungsdesign zur Bestimmung von Erfolgsfaktoren der Unternehmenskommunikation von KMU auf Facebook

So wird in Hinblick auf die Methoden in zwei Sichtweisen unterschieden: die Unternehmenssicht und die Anwendersicht.

Neben einer Literaturrecherche zu Erfolgsfaktoren und -messung wurden dazu in beiden Bereichen leitfadengestützte Experteninterviews (je 10) durchgeführt und ausgewertet. Dabei wurde bei der Stichprobenauswahl auf ein möglichst heterogenes Untersuchungsfeld geachtet, um eine größtmögliche Verallgemeinerung der Fälle zu gewährleisten. Zur grundlegenden Erkenntnisgenerierung aus den Interviews bot sich dann eine klassische Inhaltsanalyse nach Mayring an [Ma02, Kr11]. Mayring beschreibt ein Prozessmodell mit regelgeleitetem Vorgehen, bei dem nach vorheriger Festlegung einer Kategoriendefinition anhand des Leitfadens ein Materialdurchgang stattfindet und durch Subsumption eine Neu- bzw. Reformulierung der Kategorien vollzogen wird. Bei der Auswertung der Interviews wurde das Verfahren demnach sowohl induktiv als auch deduktiv angewendet. Auf Anwenderseite wurde zudem eine quantitative Onlineumfrage mit 601 Teilnehmern, davon 58,5% weiblich und einem Durchschnittsalter von 29,2 Jahren (MD = 26, SD=10,2), durchgeführt. Die Umfrage enthielt 28 meist geschlossene Fragen, welche sowohl aus Einfach- als auch aus Mehrfachauswahlen bestanden und ein breites Spektrum an Faktoren abfragte. Methodisch in Planung sind zudem Fallstudien zu ausgewählten Beiträgen bestimmter KMU sowie quantitative Inhaltsanalysen anhand eines über die Facebook API generierten Datensatzes. Abschluss des Projektes soll die Überprüfung der Ergebnisse und Hypothesen in einem (Semi-)experimentellen Setting sein.

Nachfolgend werden die Ergebnisse der Kategorisierung und den Teilstudien der qualitativen Experteninterviews und der quantitativen Umfrage dargelegt.

4 Ergebnisse

4.1 Kategorisierung von Erfolgsfaktoren

Durch die Entwicklung und Strukturierung der Leitfäden als auch durch die Auswertung der Experteninterviews mit den Anwendern aus Unternehmens- und Nutzersicht konnten folgende Kategorien eruiert werden: Organisation/Strategie, Struktur, allgemeine Nutzung, Beitragsgestaltung und Interaktion sowie Umwelt. Die folgende Tabelle stellt diese Kategorien den weiteren Methoden des Forschungsdesigns gegenüber und verdeutlicht sowohl die Anwendbarkeit als auch den bzw. die Schwerpunkte (grauer Hintergrund) der jeweiligen Methode:

Kategorie \ Anwendbarkeit und Fokus	Interview mit Unternehmen	Interview mit Nutzern	Quant. Nutzerbefragung	Grundlagen zur Erfolgsmessung	Fallstudien	Quantitative Inhaltsanalyse
Organisation/Strategie	√	-	-	√	-	-
Struktur	√	√	√	√	√	√
Nutzung	√	√	√	√	√	√
Beitragsgestalt. und Interaktion	√	√	√	√	√	√
Umwelt	√	√	-	-	-	-

Tab. 1: Matrix zur Kategorisierung anhand Methoden

Es lassen sich zwei Erkenntnisse ableiten: 1. Jede der Erhebungsmethoden lässt sich in der Auswertung auf die zentrale Forschungsfrage zur Beitragsgestaltung fokussieren. Dies erscheint durch die Planung des Forschungsdesigns wenig verwunderlich, ist aber von eminenter Bedeutung. 2. Es wird deutlich, dass die Kategorien Struktur, Nutzung sowie Beitragsgestaltung und Interaktion eine allgemeingültige Untergliederung darstellen. Die Interaktion lässt sich dabei nicht von den Inhalten trennen, da eine gegenseitige Abhängigkeit besteht. Ebenso wird aber auch erkennbar, dass nicht jede Methode in den weiteren Kategorien Organisation/Strategie und Umwelt tiefergehende Ergebnisse erwarten lässt – so werden bspw. Nutzer nur wenig über die Unternehmensorganisation der besuchten Seiten zu berichten wissen. Zur Identifikation von weiteren Erfolgsfaktoren stellen diese jedoch eine unverzichtbare Erweiterung dar.

4.2 Erkenntnisse aus den Teilstudien

Befragte KMU gaben mehrheitlich an, dass hinsichtlich ihrer **Organisation und Strategie** keine oder nur wenig Vorarbeit geleistet wurde. Lediglich die Festlegung, welche Person im Unternehmen Beiträge auf Facebook schreibt wurde berücksichtigt. Dies sind

entweder der Geschäftsführer sowie Familienmitglieder oder, sofern vorhanden, der/die Verantwortliche aus dem Marketing. Ziel sei zunächst eine Präsenz an sich zu haben und grundsätzlich im Gespräch zu sein. Generell fehle aber die Zeit zur weiteren Präzisierung: „Man kann noch so viel mehr machen“, so ein Experte. Diese Praxis steht demnach im Widerspruch mit der in der Literatur empfohlenen umfassenden Strategiefestlegung, welche im Vorfeld der Aktivitäten als zentrales Element angesehen wird [Fi12].

Die Kategorie **Struktur** beschreibt die Gesamtzahl aller Möglichkeiten auf der Plattform unternehmensbezogene Informationen zu hinterlegen, wie bspw. eine Beschreibung des Unternehmens, Produktbeschreibungen oder Kontaktdaten. Befragte Experten in den KMU gaben an, lediglich das Aussehen der Seite sei bisher im Fokus ihrer Überlegung gewesen. Die Ergebnisse der Onlineumfrage ergibt nun ein klares Bild der Rangfolge der Relevanz dieser Informationen aus Nutzersicht. Auf die Frage, welche Seiteninformation wie wichtig ist (Likert-Skala, 1 sehr wichtig – 5 unwichtig), stehen an vorderster Stelle der vollständige Name des Unternehmens ($M=1,65$), ein Link zur Unternehmenswebseite ($M=1,66$) sowie eine Beschreibung des Unternehmens ($M=1,80$). In den Anwenderinterviews wurde zudem erwähnt, dass zur Kontaktaufnahme aktiv nach der Emailadresse in den Seiteninformationen gesucht wird. Hier ist also zu vermuten, dass Anwendern die erkennbare Echtheit der Seite wichtig ist. In der Literatur wird zudem auf weitere Aspekte verwiesen: So stellen die generellen Seiteninformationen einen deutlichen Mehrwert in Hinblick auf Suchmaschinenoptimierung und Auffindbarkeit dar [AKM15, Ad09]. Zusammenfassend muss also festgehalten werden, dass das Ausfüllen dieser Informationen ein Erfolgsfaktor darstellt.

Hinsichtlich der allgemeinen **Nutzung** lautet das Ergebnis aus Sicht der KMU, dass eine stete Beobachtung der Veränderungen durch die Benachrichtigungsfunktion von Facebook per Email stattfindet. Die eigentliche Nutzung innerhalb der Plattform falle hingegen kurz aus. Seitens der Nutzer wurde in den Interviews deutlich, dass über den Tag mehrere, wenige Minuten andauernde, Logins stattfinden. Die Interviews lassen vermuten, dass der jeweilige Zeitpunkt stark mit dem persönlichen Lebensstatus wie Schüler/Student oder Berufstätigkeit zusammenhängt. Ergebnisse der Onlineumfrage [WN16] quantifizieren die Ergebnisse: So nutzen 83% der befragten Anwender Facebook täglich. Die verbleibenden 17% tendieren leicht zur Nutzung am Wochenende. In Bezug auf die Häufigkeit gaben 62% an, im Durchschnitt zwischen 1-5-mal am Tag Facebook zu besuchen. Akkumuliert steigt dieser Wert bei 91% der Anwender auf 1-15-mal. Die Hauptnutzungszeit liegt dabei zwischen 18:00 und 21:00 Uhr. Eine Studie von Buzzsumo untersuchte ebenfalls die Nutzung durch die statistische Auswertung von 1 Mrd. Beiträgen großer Unternehmen [Bu16]. Als Ergebnis wurde hier deutlich, dass auf Facebook seitens der Anwender am Wochenende eine deutlich höhere Interaktion stattfindet, zudem zwischen 22:00 und 24:00 Uhr. Es ist demnach zu vermuten, dass Anwender wochentags eher passiv agieren und sich die Intensität der Nutzung am Wochenende durch das Liken von Beiträgen, Kommentieren und Teilen erhöht.

In Bezug auf die **Beitragsgestaltung und Interaktion** gaben die Experten aus den KMU an, dass grundsätzlich alle Themen, die aus dem Unternehmen kommen und aktuell sind,

relevant seien. Allerdings müssten diese häufig aktiv gesucht werden. Beiträge sollten in betrieblichem Ton, jedoch locker und witzig geschrieben sein. Zudem knapp formuliert, mit Bildinhalten, persönlicher Ansprache und Aufforderung zum Kommentar. Auf Fragen der Nutzer sollte grundsätzlich eingegangen werden, ein Lob hingegen kann unkommentiert bleiben. Auch habe die Häufigkeit hohen Einfluss auf den Erfolg: Zumindest 2-mal sollte in der Woche ein Beitrag eingestellt werden, idealerweise morgens, mittags oder abends. Als Grad zur Erfolgsmessung wurde dabei die Reichweite sowie die, in der Literatur [Fi12] stark kritisierte, Anzahl an Kommentaren und Likes angegeben. In der Nutzerbefragung stellte sich heraus, dass der Newsfeed auf der Startseite der User aktiv hinsichtlich Neuigkeiten beobachtet wird. Hierbei wurden ebenfalls die Anzahl der Likes, Shares und Kommentare durch Freunde als hochrelevant eingestuft. Hauptgrund für das Folgen von Unternehmensseiten sei der Wunsch die persönliche Unterstützung zu zeigen sowie der Erhalt von aktuellen Informationen. Diese müssten jedoch inhaltlich auf einen Blick erkennbar sein und einen deutlichen Mehrwert bieten. Reine Werbebeiträge oder Clickbaiting, also reißerische Überschriften ohne Information, sollten nicht vorkommen. Hinsichtlich des Sprachstils gaben die Anwender an, über dieses Thema noch nicht reflektiert zu haben und wenn, dann ausschließlich einen schlechten Stil zu erkennen. Kontaktaufnahme über Facebook zum Unternehmen fand nur selten statt und es wird eine Reaktion innerhalb eines Tages erwartet, jedoch nicht zwangsläufig im gleichen Medium, sondern bspw. auch per Mail oder telefonisch. In der Onlineumfrage kamen nun folgende Ergebnisse zum Vorschein: 61% der Nutzer folgen einer Unternehmensseite, um aktuelle Informationen über Termine und Daten zu erhalten, 44% erwarten Neuigkeiten aus der Branche, 42% neue Produktinformationen. Hauptgrund für das Liken eines Beitrags ist informativer (81%) oder lustiger Inhalt (56%), der sich thematisch an Produkttests, Termine und Events, Produktinnovationen, Angebote sowie den Blick hinter die Kulissen orientiert. Ebenso gaben 69% an, dass der Inhalt sofort erkennbar sein sollte und optimaler Weise ein Bild enthält (63%). Kommentare verfassen die befragten Nutzer um ihre Meinung zu teilen (50%), eine Frage an das Unternehmen zu richten oder Bezug zu einem vorigen Kommentar zu nehmen. Hinsichtlich der Kontaktaufnahme über Kommentare, Nachrichten oder Chatfunktion gaben jedoch 69% an, diese Möglichkeiten noch nicht genutzt zu haben. Mehr als die Hälfte (55%) der Nutzer erwarten jedoch eine Reaktion innerhalb eines Tages, 31% innerhalb von sechs Stunden.

Die Kategorie **Umwelt** konnte insbesondere durch Ergebnisse der Interviews mit Unternehmen und Anwendern identifiziert werden. Sie beschreibt die Zusammenhänge der Unternehmensseiten auf Facebook mit der Offline-Welt. So tätigte ein Interviewpartner aus den KMU die Aussage, Erfolg sei, wenn er „persönlich darauf angesprochen werde“, und meint damit die Ansprache auf die Facebook-Seite im Privatleben. Aus Sicht der Nutzer wurde bspw. erwähnt, dass das Folgen eines Unternehmens auf Facebook allein die Unterstützung desselbigen als Ziel hatte und begründet in dem vorherigen Kontakt offline liegt. Somit wird deutlich, dass keine eindeutige Trennung der Kundenbeziehung zwischen on- und offline getätigt werden darf.

5 Fazit und Ausblick

Der vorliegende Beitrag erläutert einerseits das Design und die Notwendigkeit eines Methodenmix zur Bestimmung von Erfolgsfaktoren der Unternehmenskommunikation von KMU auf Facebook. Anhand eines entwickelten Kategoriensystems können diese Faktoren gegliedert und anhand der Methoden untersucht werden. Andererseits wird deutlich, dass für eine erfolgreiche Kommunikation und damit für das Unternehmen relevante Anschlüsse, Faktoren zur Nutzung sowie inhaltlichen Gestaltung der Seiten und Beiträge sicherlich eine zentrale Rolle spielen, jedoch nicht allein betrachtet werden dürfen. So haben auch organisatorische wie strukturelle Faktoren eine zusätzliche, nicht zu unterschätzende Bedeutung. Gerade für KMU ist es entscheidend, diese zu kennen, Maßnahmen abzuleiten und nicht zuletzt die Trennlinie der Kundenbeziehung zwischen on- und offline nicht zu scharf zu ziehen. Dies kann somit bedeuten, dass der Wert eines einzelnen Kunden als Fan und Multiplikator der Marke deutlich höher als bei großen Unternehmensseiten einzuschätzen ist und entsprechend begleitende Maßnahmen zur Erhöhung der Kundenbindung offline entscheidend zum Erfolg der Präsenz in SNS beitragen.

Zukünftige und laufende Arbeiten dieses Projektes werden die deskriptiven Ergebnisse in den Kategorien durch Inhaltsanalysen vervollständigen. Diese bestehen aus der Untersuchung von Fallstudien erfolgreicher Beiträge von KMU sowie der quantitativen Beschreibung eines umfangreichen Datensatzes, welcher die gesamten Inhalte ausgewählter KMU aus unterschiedlichen Branchen beinhaltet. Schließlich werden anhand der Gemeinsamkeiten der Studien endgültige Hypothesen gebildet und im letzten Projektschritt anhand eines zu entwickelnden (Semi-)experimentellen Settings, welches ich ausdrücklich zur Diskussion vorschlage, verifiziert.

Literaturverzeichnis

- [Ad09] Adweek: 10 Key SEO Strategies Every Facebook Page Owner Should Know. <http://www.adweek.com/socialtimes/seo-facebook-pages-10-key-strategies/224907?red=if>, Stand: 25.04.2016
- [AKM15] Alpar, A.; Koczy, M.; Metzen, M.: SEO – Strategie, Taktik und Technik, Online Marketing mittels effektiver Suchmaschinenoptimierung. Wiesbaden, 2015
- [As14] Aßmann, S.; Pleil, T.: Social Media Monitoring: Grundlagen und Zielsetzungen. In (Zerfaß, A.; Piwinger, M. Hrsg.): Handbuch Unternehmenskommunikation: Strategie, Management, Wertschöpfung. 2. Auflage, Wiesbaden, S. 585-604, 2014
- [BB15] bevh Boniversum: Was ist der Hauptgrund für die Nutzung von Social Media in Ihrem Unternehmen?. In Statista, <http://de.statista.com/statistik/daten/studie/214159/umfrage/gruende-fuer-die-nutzung-von-social-media-durch-unternehmen/>, Stand 30.04.2016.
- [BI15] BITKOM: Drei von vier Unternehmen nutzen Social Media. <https://www.bitkom.org/Presse/Presseinformation/Drei-von-vier-Unternehmen-nutzen->

Social-Media.html, Stand 22.02.2016.

- [Br11] Bryman, A.: Mixed Methods Research in Organizational Research. In (Buchanan, D.; Bryman, A. Hrsg.): The Sage Handbook of Organizational Research Methods. Los Angeles et al., S. 516-531, 2011
- [FB16] Facebook: Anzahl der monatlich aktiven Facebook Nutzer weltweit vom 3. Quartal 2008 bis zum 1. Quartal 2016 (in Millionen). In Statista, <http://de.statista.com/statistik/daten/studie/37545/umfrage/anzahl-der-aktiven-nutzer-von-facebook/>, Stand 30.04.2016.
- [Fi12] Fiege, R.: Social Media Balanced Scorecard. Erfolgreiche Social Media-Strategien in der Praxis, Wiesbaden, 2012
- [Kr11] Kruse, J.: Reader „Einführung in die Qualitative Interviewforschung“. Freiburg, 2011
- [Ma02] Mayring, P.: Einführung in die qualitative Sozialforschung. Weinheim, 2002
- [Ma13] Mast, C.: Unternehmenskommunikation. 5. Auflage, Konstanz, 2013
- [MBK15] Meffert, H.; Burmann, C.; Kirchgeorg, M.: Marketing. Grundlagen marktorientierter Unternehmensführung, 12. Aufl., Wiesbaden, 2015
- [MHG05] Mast, C.; Huck, S.; Güller, K.: Kundenkommunikation. Ein Leitfaden, Stuttgart, 2005
- [WN16] Wattenberg, M.; Niggemeier, M.: The Key Factors of Success of Corporate Communications in Facebook: Results of a Quantitative Online Survey. In (Stützer, C. et al. Hrsg.): 18th General Online Research Conference. Proceedings, Dresden, S. 49, 2016
- [WL11] Weissensteiner, E.; Leiner, D.: Facebook in der Wissenschaft: Forschung zu sozialen Onlinenetzwerken. Medien & Kommunikationswissenschaft, Jg. 59 H. 4, S. 526-544, 2011
- [Wi12] Wilson, R. et al: A Review of Facebook Research in the Social Sciences. Perspectives on Psychological Science 7(3), S. 203-220, 2012
- [Za14] Zerfaß, A.: Unternehmenskommunikation und Kommunikationsmanagement: Strategie, Management und Controlling. In (Zerfaß, A.; Piwinger, M. Hrsg.): Handbuch Unternehmenskommunikation: Strategie, Management, Wertschöpfung. 2. Auflage, Wiesbaden, S. 21-79, 2014

Methods for Automatic Selection of Database Systems for Optimized Query Performance

Matthäus Zloch ¹

Abstract: On a database system query runtime performance is, besides the data volume, mainly affected by the data model and the index structures that one database system utilizes to store its data in. There are lots of different implementations of the traditional relational model, triples stores, and yet more with the rising NoSQL technologies. For a developer who needs to choose one database system, it is key to understand how the data, that one application will store, looks like (data model and schema implementation) and what typical use-cases the system will be addressed with (queries), in order to make sure the system is truly suited to reach optimal overall query performance. This thesis will introduce methods and applications, in order to detect the optimal database system for a given data model and its queries.

Keywords: database systems, data modeling, machine learning, software architecture

1 Motivation and Problem Statement

Relational databases have been around for several decades. It has been a first choice technology for developers, when it comes to the decision of how to store the data of an application. In the past years, developers tend to move to technologies outside the relational model, that comprise different solutions and ways of storing data. The core assumption in this thesis is that a developer cannot know in advance, which of the available database systems are suited best for the data and the queries of an application. If the decision shall be made with care, an evaluation must be made each time an application is developed. In this thesis a closer look at the structures of data models and queries is taken, in order to find the database technology that is suited for the data of an application.

A data model is an extremely organized and selective abstraction of knowledge [Ev03] and with rising size and complexity, it tends to have different structural characteristics. For instance, there are parts with lots of many-to-many relationships and tree-like structures, which are likely to be seen in graph-based models. Those models can be found mostly in applications which need a high interconnected-ness of their entity types, e.g. social media platforms. On the other hand, we can find models which store pure numerical or string-based data in a plain tabular format, e.g. an address book or a financial application. Every database system utilizes a different kind of internal data structure that deals with the data it stores. A data model that is composed mainly of characteristics that are typical for graph-structures, a graph-database might be suited best. However, since each query addresses

¹ Heinrich-Heine Universität, Lehrstuhl für Datenbanken und Informationssysteme, Universitätsstrasse 1, 40589 Düsseldorf, matthaeus.zloch@hhu.de

different parts of a data model in the database, a developer cannot tell in advance which database system is suited best for an application.

The research idea presented in this thesis is to look at data models prior to implementation of the application and to utilize their structure, queries and data, in order to predict a suitable database system. A suitable database system is one that serves with as best performance as possible every query of an application, compared to other database systems. In order to predict the runtime of some queries which address a schema, an algorithm by means of a cost-function has to be developed, which takes a number of features into account. The features need to be identified on the schema level, query level, and, if available, on the data level. The algorithm is supposed to compute a value which will represent a degree of suitability of the given set of queries addressing the given schema, to a list of a fixed set of database systems of different types. It is planned to use a method from the area of machine learning to solve that problem.

From this aim arise several research questions that should be answered throughout the thesis, which are among others:

- Is it reasonable to study the structure of the data model and the queries before making a decision about the database system?
- What makes a data model be graph-based? What makes it be not graph-based?
- How "good" can the runtime of a query be predicted?
- What are the factors that influence a decision about a database?

2 Related Work

Related work is to be found from the time when triple stores were introduced to store RDF data [MMM04]. Research on alternative database systems is as old as database systems itself. With the growing number of database systems available and the recent change in pragmatism (from full ACID to BASE, i.e. from traditional to new models), researchers start to think about which database system is the "right" one for their first of application. Usually, comparison papers look within a group of similar database types, e.g. the group of NoSQL databases or RDF data stores, [Ro07][An12][JV13][Ga14]. In the area of RDF data stores, recent works reach out to compare among database systems of other types, like relational databases and NoSQL databases in theory and applied informatics [AG05][Vi10][Cu13]. *polyglot persistence*[Fo11] is a term which came up in the context of applied NoSQL databases. It describes the fact that different kinds of data is stored in different types of database systems. Related work and similar methods is also expected in this area.

A *recommender* for database systems according to a given data model and a set of queries, what this thesis is trying to accomplish, is not known to us. There are lots of one-to-one comparisons of features [Ga12] of database systems and articles which document experiences of usage [Ri16]. However, most likely the developer is left with picking one

system according to the application needs. This decision is rather backed up by experience and not by characteristic numbers.

Other related work is expected in the area of database systems and query evaluation plans. We expect to learn from that area of research.

3 Research Plan

Related Work, from now. Find more similar works in this area.

Evaluation, by mid 2016. First, a motivating paper shall show that it is feasible to investigate in further research. It will cover an evaluation with different types of database systems on one specific data model and example data from a real-life application, together with a fixed set of use-case queries for that application and data model. By the time of writing this paper, there are already results available, which will be described in more detail in Section 4.

Implementation Work, from mid 2016. Regarding the cost-function, in order to learn from as much queries and query runtimes for all database systems considered as possible, it is either (1) aimed to use a data generator that creates as much as realistic test data as possible for any arbitrary data model, or (2) use benchmarks with test data and queries for that. Since benchmarks are designed to be run on a specific database system, it has to be evaluated, if their use is reasonable at all. A data generator would suite for any data model and generate data with different complexity. For instance, it might be interesting how different probabilities and relations influence the cost-function's output. The method and the algorithm will constitute the second paper of the thesis.

Implementation Work, by mid 2017. The implemented algorithm shall be integrated in an application that is more user-friendly. A lightweight web application is considered. Also considerable is an integration into another already existing application. As input data, the user would pass a data model in a specific format and a set of queries that are likely to be executed on the system. This implementation will wrap up the implementation work for this thesis. It is planned to publish a comprehensive article about the final product.

4 Current Status and Results

The evaluation paper has started to be written and some results of the evaluation are still to be interpreted.

In preparation for the first paper the evaluation data has been extracted from a real-life application called MISSY², which is a web application that presents official statistics data about European and national surveys. MISSY implements the DDI-RDF Discovery Vocabulary (DISCO in short) [Bo13] as its native data model. DISCO is of sufficient complexity

² The Microdata Information System MISSY - <http://www.gesis.org/missy>

for the evaluation, since it consists of graph-like data (highly connected social science survey information) on one hand and of plain tabular data (statistics data) on the other.

For evaluation there are currently four systems taken into consideration:

- Relational database MySQL Community Edition - an open source database of wide acceptance. The original data is stored in a MySQL database and this implementation forms the baseline.
- Triple store Virtuoso OpenSource³ - a famous triple store implementation in Java. Since the original model is supposed to be a linked data model, this database has been chosen because of its capability to store RDF data.
- Graph databases: Since graph databases make a dominant part of NoSQL technologies, two different candidates were chosen from this category: (1) Neo4j Community Edition⁴ - a native graph database with Cypher as its own query language and (2) Stardog⁵ - a graph database that allows to query with SPARQL.

The queries were composed from three different sources (around 40 in total): the project's web application, official usage-cases document available for DISCO[Vo15], and validation queries for DISCO for functional data integrity[Ha15]. Since all types of database systems listed above allow different query languages, they had to be translated into the query language each system supports. The queries were executed according to two different approaches: (1) equal distribution, where each query is executed 25 times within a shuffled list of queries (leading to 1000 executions overall), and (2) probability distribution, where each query is assigned to one (out of four in total) application usage scenarios according to a specific weight. In approach (2) some queries will be executed more often than others in one usage scenario, respectively. Like in approach (1), the total number of executions is set to 1000.

An example of results can be seen in Figure 1. It shows the results for an evaluation run with different distributions of queries on usage-scenario 1. Usage-scenario 1 contains queries which relate to a user's navigation behaviour of the web application MISSY. There are five different weights per usage scenario, here 50/50, 60/40, up to 90/10, where 50/50 means that 50% of the queries are specific to usage-scenario 1 and 50% are not. In 60/40 there are 60% of the queries specific to usage-scenario 1 and 40% are not, and so on. From usage scenario 1 we can see in Figure 1 that the higher the frequency of returning queries is, the faster is MySQL database and slower Neo4j. Else is with Stardog and Virtuoso: Stardog's performance is slightly worse, where Virtuoso's merely differ.

The results remain to be interpreted, but are expected to contribute on different levels. For approach (1) we want to learn from each query individually: how does it look like? Which concepts does it use (aggregation-functions, relations covered, size of where-clause, etc.)? Which parts of the data model does it address, and according to that, on which database

³ <https://github.com/openlink/virtuoso-opensource>

⁴ <http://neo4j.com>

⁵ <http://stardog.com/>

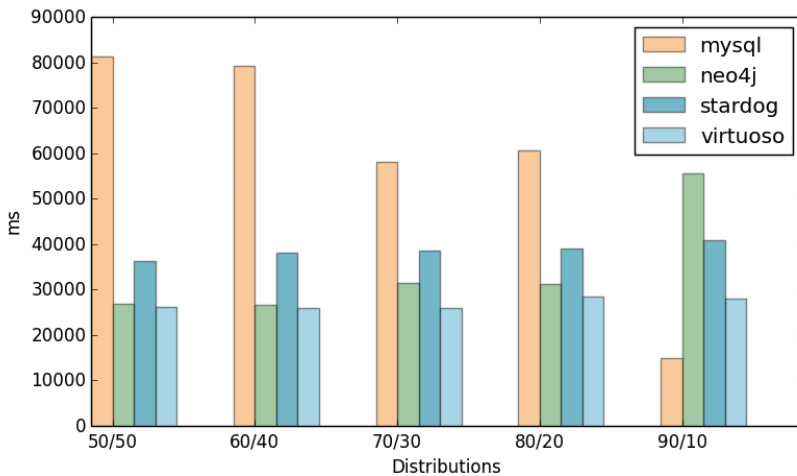


Abb. 1: Increasing the frequency of queries in usage-scenario 1 results in better overall query runtime for MySQL and worse performance for Neo4j. Stardog and Virtuoso perform slightly worse.

system is it executed faster? For approach (2) we will look at overall query runtime performance to motivate that it is reasonable to look at the data model and the queries used in advance.

References

- [AG05] Angles, Renzo; Gutierrez, Claudio: Querying RDF Data from a Graph Database Perspective. In: Proc. of 2nd European Semantic Web conference. ESWC, Crete, pp. 346–360, 2005.
- [An12] Angles, Renzo: A Comparison of Current Graph Database Models. In: ICDE Workshops. IEEE Computer Society, pp. 171–177, 2012.
- [Bo13] Bosch, Thomas; Cyganiak, Richard; Gregory, Arofan; Wackerow, Joachim: DDI-RDF Discovery Vocabulary: A Metadata Vocabulary for Documenting Research and Survey Data. In: Proceedings of the 6th Linked Data on the Web (LDOW) Workshop at the World Wide Web Conference (WWW). CEUR Workshop Proceedings. CEUR-WS.org, 2013.
- [Cu13] Cudr-Mauroux, Philippe; Enchev, Iliya; Fundatureanu, Sever; Groth, Paul T.; Haque, Albert; Harth, Andreas; Keppmann, Felix Leif; Miranker, Daniel P.; Sequeda, Juan; Wylot, Marcin: NoSQL Databases for RDF: An Empirical Evaluation. In: International Semantic Web Conference (2). Lecture Notes in Computer Science. Springer International Publishing, pp. 310–325, 2013.
- [Ev03] Evans, Eric: Domain-driven Design: Tackling Complexity in the Heart of Software. Addison-Wesley Longman Publishing Co., Inc., 2003.
- [Fo11] Fowler, Martin: , Polyglot Persistence. Available from <http://martinfowler.com/bliki/PolyglotPersistence.html>, 2011.

- [Ga12] Garshol, Lars Marius: , RDF triple stores - an overview. Available from <http://www.garshol.priv.no/blog/>, September 2012.
- [Ga14] Gandini, Andrea; Gribaudo, Marco; Knottenbelt, WilliamJ.; Osman, Rasha; Piazzolla, Pietro: Performance Evaluation of NoSQL Databases. In: Computer Performance Engineering. Lecture Notes in Computer Science. Springer International Publishing, pp. 16–29, 2014.
- [Ha15] Hartmann, Thomas; Zopilko, Benjamin; Wackerow, Joachim; Eckert, Kai: Constraints to Validate RDF Data Quality on Common Vocabularies in the Social, Behavioral, and Economic Sciences. Computing Research Repository (CoRR), abs/1504.04479, 2015. <http://arxiv.org/abs/1504.04479>.
- [JV13] Jouili, Salim; Vansteenbergh, Valentin: An Empirical Comparison of Graph Databases. In: SocialCom. IEEE Computer Society, pp. 708–715, 2013.
- [MMM04] Manola, Frank; Miller, Eric; McBride, Brian: , RDF Primer. Available from <https://www.w3.org/TR/rdf-primer/>, 2004.
- [Ri16] Richey, Clark: , From Good to Graph: Choosing the Right Database. Available from <http://neo4j.com/blog/good-to-graph-right-database>, 2016.
- [Ro07] Rohloff, Kurt; Dean, Mike; Emmons, Ian; Ryder, Dorene; Sumner, John: An Evaluation of Triple-Store Technologies for Large Data Stores. In: OTM 2007 Workshop SSWS. pp. 1105–1114, 2007.
- [Vi10] Vicknair, Chad; Macias, Michael; Zhao, Zhendong; Nan, Xiaofei; Chen, Yixin; Wilkins, Dawn: A comparison of a graph database and a relational database: a data provenance perspective. In: ACM Southeast Regional Conference. ACM, p. 42, 2010.
- [Vo15] Vompras, Johanna; Gregory, Arofan; Bosch, Thomas; Wackerow, Joachim: Scenarios for the DDI-RDF Discovery Vocabulary. DDI Working Paper Series, 2015. <http://dx.doi.org/10.3886/DDISemanticWeb02>.

**Studierendenkonferenz
Informatik 2016
(SKILL 2016)**

Studierendenkonferenz Informatik 2016 (SKILL 2016)

Michael Becker,¹ Agnes Koschmider,² Judith Michael³ und Thomas Riechert⁴

Mit der diesjährigen Studierendenkonferenz Informatik (SKILL) wird die seit 2014 bestehende Zusammenarbeit mit der Gesellschaft für Informatik e.V. weiter vertieft. Bereits zum dritten Mal findet die SKILL im Rahmen der GI-Jahrestagung INFORMATIK statt.

Insgesamt wurden 47 Beiträge als Full oder Short Paper eingereicht und begutachtet. Auf der Konferenz haben die Verfasserinnen und Verfasser von 18 Beiträgen (5 Short und 13 Full Paper) die Möglichkeit, ihre Arbeit zu präsentieren. Die Mitglieder des Organisationskomitees der SKILL 2016 bedanken sich zunächst bei den Autorinnen und Autoren, ohne deren qualitativ hochwertiger Beiträge die Konferenz nicht möglich wäre. Das Themenspektrum eingereicherter Beiträge reicht dabei von theoretisch-mathematischen Grundlagenarbeiten über Beiträge, die hardwareseitige Aspekte diskutieren, bis hin zur Anwendung von Erkenntnissen der Informatik in anderen Fachbereichen. Anhand der Vielzahl verteilter Themen zeigt sich die große Vielfalt der Informatik sowie angrenzender Disziplinen im deutschsprachigen Raum.

Wir freuen uns darüber hinaus, dass wir auch in diesem Jahr wieder namhafte Gutachterinnen und Gutachter gewinnen konnten, die den Studierenden mit hilfreichen Kommentaren zu ihren Arbeiten zur Seite standen. Durch deren Hilfe konnten alle eingereichten Beiträge von mindestens zwei ausgewiesenen Expertinnen und Experten begutachtet werden.

Organisationskomitee der SKILL 2016

- Michael Becker, Universität Leipzig
- Agnes Koschmider, Karlsruher Institut für Technologie
- Judith Michael, Alpen-Adria-Universität Klagenfurt
- Thomas Riechert, Hochschule für Technik, Wirtschaft und Kultur Leipzig

¹ Universität Leipzig, Institut für Informatik, Hainstr. 11, 04109 Leipzig, mbecker@informatik.uni-leipzig.de

² Karlsruher Institut für Technologie, Institut AIFB, 76128 Karlsruhe, agnes.koschmider@kit.edu

³ Alpen-Adria-Universität Klagenfurt, Institut für Angewandte Informatik, Universitätsstr. 65-67, A-9020 Klagenfurt am Wörthersee, Judith.Michael@aau.at

⁴ HTWK Leipzig, Fakultät Informatik, Mathematik und Naturwissenschaften, Gustav-Freytag-Str. 42A, 04277 Leipzig, thomas.riechert@htwk-leipzig.de

Gutachterinnen und Gutachter der SKILL 2016

- Ernst Althaus, Johannes Gutenberg-Universität Mainz
- Thomas Bartz-Beielstein, Fachhochschule Köln
- Eugen Brenner, TU Graz
- Ruth Breu, Universität Innsbruck
- Karin Harbusch, Universität Koblenz-Landau
- Hermann Hellwagner, Alpen-Adria-Universität Klagenfurt
- Andreas Henrich, Otto-Friedrich-Universität Bamberg
- Dimitris Karagiannis, Universität Wien
- Paul Molitor, Martin-Luther-Universität Halle
- Wolfgang Mulzer, Freie Universität Berlin
- Dirk Nowotka, Christian-Albrechts-Universität zu Kiel
- Andreas Nüchter, Julius-Maximilians-University Würzburg
- Johannes Oberzaucher, Fachhochschule Kärnten
- Gabriele Peters, FernUniversität Hagen
- Klaus Pohl, Universität Duisburg-Essen
- Franz-Josef Radermacher, Universität Ulm
- Stefan Rass, Alpen-Adria-Universität Klagenfurt
- Ulrich Rückert, Universität Bielefeld
- Gunter Saake, Otto-von-Guericke-University Magdeburg
- Thomas C. Schmidt, Hochschule für Angewandte Wissenschaften Hamburg
- Ingo Scholtes, ETH Zürich
- Björn Schuller, University of Passau
- Claudia Steinberger, Alpen-Adria-Universität Klagenfurt
- Sebastian Thöne, Fachhochschule Münster
- Dietmar Tutsch, Bergische Universität Wuppertal
- Hubert Zangl, Alpen-Adria-Universität Klagenfurt
- Katharina Zweig, TU Kaiserslautern

Das Ende der Miniaturisierung

Tom Standfuß¹

Abstract: In dieser Arbeit wird eine Zeitspanne berechnet, die durch statistische Analysen des Technologieknotenverlaufes der letzten 30 Jahre das besagte Ende der Miniaturisierung, bei gleichbleibenden Umwelteinflüssen, präzise vorhersagt.

Keywords: Ende der Miniaturisierung, Moore'sches Gesetz, Technologieknoten, Quantenmechanischer Tunneleffekt, Nachweis des kleinstmöglichen Abstandes zwischen Leiterbahnen und dessen Eintrittsdatum

1 Einleitung

Diese Arbeit stellt primär eine Ergänzung der, von Herrn Gordon Moore gehaltenen, Rede aus dem Jahr 1975 [GM75], die er beziehend auf seine wissenschaftliche Arbeit aus dem Jahr 1965 [GM65] gehalten hatte, dar. Die Erweiterung bezieht sich dabei auf das Ende der Miniaturisierung in der heutigen Schaltkreisentwicklung. Ebenfalls wird die Abbildung 3 in dem White Paper von der International Technology Roadmap for Semiconductors (ITRS) auf der Seite 8 [MtM] präzisiert, welche die Entwicklung neuer Technologien in Abhängigkeit von der Baugröße von Transistoren darstellt.

2 Grundbegriffe

2.1 Moore'sches Gesetz

Gordon Moore beschrieb erstmals 1965 [GM65] den zukünftigen Verlauf der vorhandenen Transistoren auf einem Computerchip und formulierte eine bis heute zutreffende Gesetzmäßigkeit, die besagt, dass sich ca. alle 12 bis 24 Monate die Transistoranzahl und damit die Integrationsdichte auf einem Computerchip verdoppelt [GM75]. Hierbei handelt es sich jedoch nicht um ein Gesetz, sondern lediglich um eine Vorhersage, die durchaus kleineren Schwankungen unterliegt. Ein essentieller Faktor bei dieser exponentiellen Entwicklung ist die Miniaturisierung der Strukturen und in diesem Zusammenhang auch der Technologieknotenverlauf. Ein Technologieknoten bezeichnet einen Zeitpunkt, bei dem erstmals der kleinste Abstand zwischen zwei benachbarten Leiterbahnen (engl. half pitch) [PLHT] in einem Computerchip erreicht wurde.

¹ Hochschule für Telekommunikation Leipzig, Gustav-Freytag-Straße 43-45, 04277 Leipzig, Tom.Standfuss@hft-leipzig.de

2.2 Quantenmechanischer Tunneleffekt

Der Tunneleffekt ist eine quantenmechanische Erscheinung, bei der ein Durchgang eines Teilchens durch einen Potentialwall möglich ist, auch wenn die kinetische Energie des Teilchens geringer ist als die, für die Überwindung des Potentialwalls, benötigte Energie [GQ11]. Die Länge des Potentialwalls wäre in diesem Fall der zu überwindende Half Pitch-Abstand und die Höhe wäre die Gesamtenergie, die ein Elektron benötigt, um den Potentialwall auch ohne den Tunneleffekt vollständig überwinden zu können.

3 Nachweis des kleinstmöglichen Abstandes

Die Wahrscheinlichkeit, dass ein Elektron durch den Isolationsbereich zwischen den Leiterbahnen (engl. space) tunnelt, wird mit der Transmissionswahrscheinlichkeit T angegeben. Diese Transmissionswahrscheinlichkeit für ein Elektron ist von der Höhe des Potentialwalls V_0 , dessen Breite L und der Anfangsenergie E und der Masse m des Elektrons abhängig [GQ11]:

$$T = \frac{16}{3 + \frac{V_0}{E}} \cdot e^{-2L \cdot \frac{\sqrt{2m \cdot (V_0 - E)}}{\hbar}} \approx e^{-2L \cdot \frac{\sqrt{2m \cdot W_A}}{\hbar}} \quad (1)$$

Die Austrittsarbeit W_A ($W_A = V_0 - E$) ist die Energie, die einem Silicium-Atom zugeführt werden muss, damit ein Elektron dieses Atoms den Potentialwall überwindet (ohne Eintreten des Tunneleffekts). Für ein Silicium-Atom beträgt diese Energie ungefähr 4,60eV bis 4,85eV [WIKIa]. Hier wird von dem Mittelwert beider Grenzen ausgegangen und dadurch beträgt $W_A = 4,725\text{eV}$.

Nun kann eine Funktion für die Transmissionswahrscheinlichkeit T in Abhängigkeit von dem Abstand zwischen zwei nebeneinanderliegenden Leiterbahnen aufgestellt werden:

$$T(L) \approx e^{-2L \cdot \frac{\sqrt{2(9,11 \cdot 10^{-31} \text{kg}) \cdot 4,725 \text{eV}}}{1,055 \cdot 10^{-24} \text{J} \cdot \text{s}}} \quad | \quad 1 \hbar = 1,055 \cdot 10^{-24} \text{J} \cdot \text{s} \quad (2)$$

$$T(L) \approx e^{-2L \cdot \frac{\sqrt{2 \cdot (9,11 \cdot 10^{-31}) \cdot (4,725 \cdot 1,6 \cdot 10^{-19})}}{1,055 \cdot 10^{-24}} \cdot \frac{\sqrt{\text{kg} \cdot \frac{\text{kg} \cdot \text{m}^2}{\text{s}^2}}}{\text{J} \cdot \text{s}}} \quad | \quad 1 \text{eV} = 1,6 \cdot 10^{-19} \cdot \frac{\text{kg} \cdot \text{m}^2}{\text{s}^2} \quad (3)$$

$$T(L) \approx e^{-2L \frac{\sqrt{1,377432 \cdot 10^{-48}}}{1,055 \cdot 10^{-24}} \sqrt{\frac{kg^2 \cdot m^2}{s^2}} \frac{kg \cdot m^2}{s^2}} \quad | \quad 1J = 1 \cdot \frac{kg \cdot m^2}{s^2} \quad (4)$$

$$T(L) \approx e^{-2L \frac{\sqrt{1,377432 \cdot 10^{-48}}}{1,055 \cdot 10^{-24}} \frac{kg \cdot m}{s}} \frac{kg \cdot m^2}{s} \quad (5)$$

$$T(L) \approx e^{-2L \frac{\sqrt{1,377432 \cdot 10^{-48}}}{1,055 \cdot 10^{-24}} \frac{1}{m}} \quad \forall L > 0 \quad (6)$$

$$T_2(L) \approx e^{-2(L \cdot 10^{-9} \cdot m) \cdot \frac{1}{m} \frac{\sqrt{1,377432 \cdot 10^{-48}}}{1,055 \cdot 10^{-24}}} \quad (7)$$

$$T_2(L) \approx e^{-2,2249 \cdot 10^{10} \cdot L \cdot 10^{-9}} \quad | \quad \text{Für } L \text{ in nm} \quad (8)$$

Die Funktion $T_2(L)$ gibt nun an, wie viele Elektronen, in Abhängigkeit von dem Abstand (von insgesamt einem Elektron), durch diesen Potentialwall tunneln (s. Anhang, Abb. 1). Diese Wahrscheinlichkeit ist für ein einziges Elektron sehr gering, allerdings muss es in der Gesamtheit für alle betroffenen Valenzelektronen² im Schaltkreis bestimmt werden. Dazu benötigt man noch die Anzahl aller vorhandenen Silicium-Atome $n_{SI-Atome}$ in einem Chip, die man mithilfe des Molaren Volumens V_m von Silicium [WIKIb] und dem Volumen des Chips V_{Chip} berechnet (hier am Beispiel der Intel Core i7-4770-Serie mit einer Herstellhöhe von 500µm eines Wafers [ITW]):

$$V_{Chip} = (177 \cdot 10^{-3})m^2 \cdot (500 \cdot 10^{-6})m = 8,85 \cdot 10^{-5}m^3 \quad (9)$$

$$n_{SI-Atome} = \frac{V_{Chip} \cdot N_A}{V_m} = \frac{(8,85 \cdot 10^{-5}m^3)(6 \cdot 10^{23} \cdot \frac{1}{mol})}{12,06 \cdot 10^{-6} \cdot \frac{m^3}{mol}} \approx 4,4 \cdot 10^{24} \quad (10)$$

Daher wird im Folgenden nur von 5% aller Atome im Schaltkreis ausgegangen, die den Tunneleffekt auslösen können. Selbst große Abweichungen von $\pm 50\%$ bei dieser Schätzung haben fast keinen Einfluss auf den späteren Verlauf der Funktion, da dieser Koeffizient nicht in dem Exponenten steht. Aus diesem Grund wurde ebenfalls für die Bestimmung der Transmissionswahrscheinlichkeit eine vereinfachte Formel benutzt, die das Ergebnis nicht erkennbar beeinflusst, da der Koeffizient vor dem Exponenten deutlich kleiner ist als die Anzahl der prädestinierten Elektronen. Es sollte hier lediglich auf dessen

² Valenzelektronen bezeichnen alle Elektronen auf der äußersten Schale eines Atoms.

Beachtung hingewiesen werden.

Silicium befindet sich in der vierten Hauptgruppe des Periodensystems und besitzt deshalb 4 Valenzelektronen, die den Tunneleffekt auslösen könnten und daher in die Berechnung mit einfließen müssen. Aus den oben genannten Gründen ist die Anzahl aller prädestinierten Elektronen n_{PE} :

$$n_{PE} = n_{SI-Atome} \cdot 0,05 \cdot 4 = 8,8 \cdot 10^{23} \quad (11)$$

Nun kann die endgültige Funktion für die Transmissionswahrscheinlichkeit T in Abhängigkeit von dem Minimalabstand zweier benachbarter Leiterbahnen und ausschließlich bezogen auf die prädestinierten Elektronen aufgestellt werden:

$$T_3(L) \approx n_{PE} \cdot e^{-2,2249 \cdot 10^{10} \cdot L \cdot 10^{-9}} \quad | \text{Für } L \text{ in nm} \quad (12)$$

Die Funktion $T_3(L)$ gibt nun an, wie viele Elektronen (von allen prädestinierten Elektronen) durch einen Potentialwall der Länge L tunneln (s. Anhang, Abb. 1).

Man erkennt, dass die Steigung der Funktion $T_3(L)$ ungefähr ab $L > 2,5$ Nanometer stark abfällt und gegen null konvergiert. Beeinflussungen eines Schaltkreises durch den Tunneleffekt entstehen demnach erst bei Prozessoren, deren Abstände zwischen den Leiterbahnen kleiner als 2,5 Nanometer werden.

Mit einem Sicherheitsabstand zu der 2,5-Nanometer-Grenze, könnte man sagen, dass der minimale Abstand zwischen den Leiterbahnen mindestens 4 Nanometer betragen sollte, um Fehlfunktionen durch den Tunneleffekt zu vermeiden.

Zusätzlich wird noch die Aussage von Herrn Dr. Robert Colwell für eine technische Einschätzung des Sachverhaltes mit in die Berechnung einbezogen. Er war Hauptarchitekt für die P6-Familie von Intel und besitzt damit praktische Erfahrung von dem Bau von Prozessoren [WIKIc]:

„I pick about 2020 as the earliest thing where I think we can call [Moore's Law] dead and it's only seven years away. I am picking 7 nm - [maybe] you can talk me into 2022 and you might even be able to talk me into 5 nm – I don't know, but you're not gonna talk me into 1 nm. I think physics dictates against that“ [HC13]

Aus dem Statement von Herrn Dr. Colwell und der vorher berechneten Grenze, wird in der späteren Betrachtung immer von dem Mittelwert beider Grenzen ausgegangen: Es sollten keine kleineren Abstände zwischen Leiterbahnen als 4,5 Nanometer in der Prozessorentwicklung zustande kommen, da sonst eine hohe Wahrscheinlichkeit besteht, dass diese von dem Tunneleffekt betroffen sein werden.

4 Datenbeschaffung

Für eine Analyse des Miniaturisierungsverlaufes werden die Daten von Prozessoren benötigt, die einen neuen Technologieknoten erreicht haben.

Jeder der unten aufgelisteten Prozessoren hat den nächst-kleinere(n) Technologieknoten durchbrochen und wurde von der Firma Intel produziert. Alle weiteren Spalten, außer Prozessornamen, Veröffentlichungsdatum und erreichter Technologieknoten werden ab dem Kapitel 5 besprochen und dienen einer besseren Nachvollziehbarkeit der Analyse.

<i>Prozessor</i>	<i>Zeitpunkt der Veröffentlich.</i>	<i>TK³ in nm (oder f(x))</i>	<i>x</i>	<i>b</i>	<i>f_{ideal}(x)</i>	<i>Abweichung g</i>
<i>Intel 80286</i>	1982	1500	0		1500,00	0,00
<i>Intel 80386</i>	1985	1000	3	0,1352	1039,29	-39,29
<i>Intel 80486</i>	1989	800	7	0,0898	637,19	162,81
<i>Intel 486DX4</i>	1994	600	12	0,0764	345,68	254,32
<i>Intel Pentium Pro</i>	1995	350	13	0,1119	305,89	44,11
<i>Intel Pentium II (Deschutes)</i>	1998	250	16	0,1120	211,94	38,06
<i>Intel Pentium III (Coppermine)</i>	1999	180	17	0,1247	187,54	-7,54
<i>Intel Pentium 4 (Northwood)</i>	2002	130	20	0,1223	129,94	0,06
<i>Intel Pentium 4 (Prescott)</i>	2004	90	22	0,1279	101,74	-11,74
<i>Intel Core 2 (Conroe)</i>	2006	65	24	0,1308	79,67	-14,67
<i>Intel Core 2 (Penryn)</i>	2008	45	26	0,1349	62,38	-17,38
<i>Intel Core iX-Reihe (Westmere)</i>	2010	32	28	0,1374	48,84	-16,84
<i>Intel Core iX-Reihe (Ivy Bridge)</i>	2012	22	30	0,1407	38,24	-16,24
<i>Intel Core iX-Reihe (Broadwell)</i>	2014	14	32	0,1461	29,95	-15,95

Tab. 1: Prozessoren und Technologieknoten

Alle Elemente einer Spalte s werden im Folgenden als Datenbasis _{s} bezeichnet:

³ Technologieknoten

$$\text{Datenbasis}_x = \{0,3,7,12,13,16,17,20,22,24,26,28,30,32\} \quad (13)$$

5 Analyse des Datenbestandes

Das Analyseverfahren wurde so gewählt, dass es allgemein auf eine endliche Anzahl von gegebenen Technologieknotenpunkten anwendbar ist und somit auch für weiterführende Analyse Zwecke verwendet werden kann. Aus diesem Grund werden augenscheinliche erkennbare Tatsachen noch einmal mathematisch nachgewiesen.

5.1 Approximation der Idealfunktion

Dem ersten Anschein nach sieht diese Anordnung der Punkte nach einer e-Funktion mit negativem Exponenten aus, dessen Funktionswerte gegen null konvergieren (s. Anhang, Abb. 2). Durch eine Beschränkung des Betrachtungszeitraumes auf die Jahre nach 2004 ($x \geq 22$), könnte man sogar von einer linearen Abnahme ausgehen, jedoch soll in dem Rahmen dieser Arbeit der Technologieknotenverlauf über eine längere Zeit analysiert werden, um präzisere Mittelwerte zu erhalten, weshalb dieser Ansatz nicht zielführend wäre. Die allgemeine Formulierung einer abfallenden e-Funktion sieht wie folgt aus:

$$f(x) = a \cdot e^{-bx} = y \quad (14)$$

Der Ordinatenabschnitt wird durch die Variable a gekennzeichnet und das Krümmungsverhalten der Funktion f durch die Variable b . Den Ordinatenabschnitt bestimmt man mithilfe des Funktionswertes für $x = 0$. Demnach gilt für a :

$$a = f(0) = 1500 \quad (15)$$

Um nun für alle Funktionswerte $f(x)$ einen durchschnittlichen Krümmungskoeffizienten b bestimmen zu können, stellt man die allgemeine e-Funktion $f(x)$ nach b für alle $x \in \text{Datenbasis}_x$ um.

$$a \cdot e^{-bx} = y \quad | : a \quad (16)$$

$$e^{-bx} = \frac{y}{a} \quad | \ln(\cdot) \quad (17)$$

$$\ln(e^{-bx}) = \ln\left(\frac{y}{a}\right) \quad | \log_a(a^b) = b \cdot \log_a(a) = b \quad (18)$$

$$-bx = \ln\left(\frac{y}{a}\right) \quad | : (-x) \quad (19)$$

$$b_x = -\frac{\ln\left(\frac{y}{a}\right)}{x} \quad \forall a, x \neq 0 \quad (20)$$

Im Folgenden bestimmt man jedes b für $x \in \text{Datenbasis}_x$. Durch die Bestimmung des arithmetischen Mittels von b erhält man den Krümmungskoeffizienten, der durch den Punkt $P_1(0|1500)$ und so nahe wie möglich an den zusätzlichen Punkten verläuft.

$$\bar{b} = \frac{1}{n} \sum_{i=1}^n b_i \approx 0,1223 \quad (21)$$

Diese Funktion nennen wir $f_{\text{ideal}}(x)$, da sie ohne die Ungenauigkeiten und statistische Anomalien in den Daten zu beachten, auf dem Graphen verläuft. Sie lautet (s. Anhang, Abb. 2, grüne Markierung):

$$f_{\text{ideal}}(x) = 1500 \cdot e^{-0,1223x} \quad (22)$$

5.2 Festlegung eines Toleranzbereiches

Um noch aussagekräftigere Ergebnisse zu erhalten, legt man nun einen Toleranzbereich fest, indem man zwei weitere Funktionen berechnet, die alle Punkte einmal oberhalb und einmal unterhalb abgrenzen. Dazu bestimmt man zuvor noch den Mittelwert m und die Standardabweichung s aller nominalen Abweichungen (Abweichung = $f(x) - f_{\text{ideal}}(x)$), getrennt nach positiven und negativen Werten, um später statistische Anomalien zu erkennen.

$$\bar{m}_+ = \frac{1}{n} \sum_{i=1}^n m_i = 83,23\text{nm} \quad m_i \in \text{Datenbasis}_{\text{Abweichung}}, m_i \geq 0 \quad (23)$$

$$\bar{m}_- = \frac{1}{n} \sum_{i=1}^n m_i = -15,52\text{nm} \quad m_i \in \text{Datenbasis}_{\text{Abweichung}}, m_i \leq 0 \quad (24)$$

Der Mittelwert gibt Aufschluss darüber, wie sehr ein neu hinzugefügter Wert im Mittel von der Funktion $f_{\text{ideal}}(x)$ abweichen kann.

$$s_+ = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z})^2} \approx 103,02 \text{ nm} \quad z_i \in \text{Datenbasis}_{\text{Abweichung}}, z_i \geq 0 \quad (25)$$

$$s_- = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z})^2} \approx 9,63 \text{ nm} \quad z_i \in \text{Datenbasis}_{\text{Abweichung}}, z_i \leq 0 \quad (26)$$

Die Standardabweichung ist ein Maß für die durchschnittliche Streuung von neu hinzugefügten Werten um den Mittelwert der Abweichung.

Im durchschnittlich schlechtesten Fall, kann dadurch ein neuer Punkt P_{neu} mit positiver Abweichung um weitere 186,25 nm in positiver Richtung von $f_{\text{ideal}}(x)$ abweichen. Alle weiteren Punkte, die nicht mehr in diesem maximal erreichbaren Toleranzbereich liegen, sind statistische Anomalien und werden im Folgenden nicht mehr für die Ermittlung der Toleranzfunktionen beachtet (s. Anhang, Abb. 2, rot markiert). Da der Punkt $P_2(7|800)$ nur 14 nm von dieser Toleranzgrenze entfernt ist, wird er ebenfalls nicht mehr betrachtet.

5.3 Berechnung der Toleranzfunktion

Nun berechnet man die Toleranzfunktionen nach dem in 4.1 beschriebenen Schema, jedoch getrennt nach positiv und negativ abweichenden Werten, um die Ober- und Untergrenze bestimmen zu können. Demnach bestehen die Datenbasen $\text{Datenbasis}_{x\pm}$ nur noch aus folgenden Werten:

$$\text{Datenbasis}_{x_+} = \{0,13,16,20\} \quad (27)$$

$$\text{Datenbasis}_{x_-} = \{0,17,22,24,26,28,30,32\} \quad (28)$$

Nun wird wieder das arithmetische Mittel aus allen Krümmungskoeffizienten b für jedes $x \in \text{Datenbasis}_{x\pm}$ bestimmt und nach $x > 0$, $x < 0$ getrennt:

$$\bar{b}_+ = \frac{1}{n} \sum_{i=1}^n b_i \approx 0,1154 \quad b_i \in \text{Datenbasis}_{x_+} \quad (29)$$

$$\bar{b}_- = \frac{1}{n} \sum_{i=1}^n b_i \approx 0,1346 \quad b_i \in \text{Datenbasis}_{x_-} \quad (30)$$

Daraus können nun die Funktionen $f_{OG}(x)$ für die Obergrenze und $f_{UG}(x)$ für die Untergrenze gebildet werden, die wie folgt lauten (s. Anhang, Abb. 2, gelbe Markierung):

$$f_{OG}(x) = 1500 \cdot e^{-0,1154x} \quad (31)$$

$$f_{UG}(x) = 1500 \cdot e^{-0,1346x} \quad (32)$$

Zusammen mit der Idealfunktion ist man nun in der Lage, einen Näherungswert für die Erreichung des wahrscheinlich letzten Technologieknotens vorherzusagen zu können.

5.4 Berechnung des Schnittpunktes

Die Berechnung der Schnittpunkte der Funktionen mit der Geraden g der Höhe h (hier: Abstand zwischen der x -Achse und den Funktionswerten von g), die parallel zur Abszisse verläuft, ist für diese Arbeit der wichtigste Analysebestandteil.

Die Höhe von der Geraden ist dabei der in Abschnitt 2.3 berechnete, minimale Abstand, der benötigt wird, um bei Prozessoren keine Fehlfunktionen durch den Tunneleffekt hervorzurufen.

Zuerst stellt man die allgemeine Funktion für exponentielle Funktionen so um, dass man zu einem gegebenen Funktionswert und Krümmungskoeffizienten den Zeitpunkt x auf der Abszisse bestimmen kann. Dazu geht man ähnlich, wie im Punkt 4.1 beschrieben, vor:

$$a \cdot e^{-bx} = y \quad | : a \quad (33)$$

$$e^{-bx} = \frac{y}{a} \quad | \ln(\cdot) \quad (34)$$

$$\ln(e^{-bx}) = \ln\left(\frac{y}{a}\right) \quad | \log_a(a^b) = b \cdot \log_a(a) = b \quad (35)$$

$$-bx = \ln\left(\frac{y}{a}\right) \quad | : (-b) \quad (36)$$

$$x = -\frac{\ln\left(\frac{y}{a}\right)}{b} \quad \forall a, b \neq 0 \quad (37)$$

Nun bestimmt man für alle drei gegebenen Funktionen $f_{ideal}(x)$, $f_{OG}(x)$ und $f_{UG}(x)$ dessen

Schnittpunkte:

$$f_{ideal}(x): x_{ideal} = -\frac{\ln\left(\frac{4,5}{1500}\right)}{0,1223} \approx 47 \quad (38)$$

$$f_{OG}(x): x_{OG} = -\frac{\ln\left(\frac{4,5}{1500}\right)}{0,1154} \approx 50 \quad (39)$$

$$f_{UG}(x): x_{UG} = -\frac{\ln\left(\frac{4,5}{1500}\right)}{0,1346} \approx 43 \quad (40)$$

Das Zeitfenster kann damit terminiert werden:

$$Eintrittsjahr_{ideal} = 1982 + x_{ideal} = 2029 \quad (41)$$

$$Eintrittsjahr_{OG} = 1982 + x_{OG} = 2032 \quad (42)$$

$$Eintrittsjahr_{UG} = 1982 + x_{UG} = 2025 \quad (43)$$

Die Miniaturisierung wird zwischen den Jahren 2025 und 2032 anfangen zu stagnieren. Es zeichnet sich jedoch eine Tendenz in Richtung des Jahres 2025 ab, da sich viele der aktuellen Technologieknoten unterhalb der durchschnittlichen Abweichung befinden und damit das Ereignis früher eintreten lassen können.

Die in Abschnitt 4 berechneten Schnittpunkte könnte man bei einer weiteren Analyse mithilfe von genaueren Daten und einer besseren Methode zur Bestimmung der Approximationsfunktionen optimieren. Des Weiteren sind hier gezeigte Berechnungen nur bei einer gleichbleibenden Umwelt – also keiner künstlich verlangsamten Entwicklung seitens der Unternehmen, die dadurch mehr Umsatz mit älteren Prozessormodellen erzielen können, gültig [HC13].

6 Auswertung

Das Ende der Miniaturisierung wird eintreten – es ist nur eine Frage der Zeit. Wie in Abschnitt 4.3 der Analyse berechnet, wird es zwischen den Jahren 2025 und 2032 zu einer Stagnation der Miniaturisierung kommen, da dort die physikalische Grenze der derzeitigen Bauweise von Prozessoren erreicht sein wird. Diese Zeitspanne deckt sich ebenfalls mit der Aussage von Herrn Dr. Robert Colwell aus dem Abschnitt 3 [HC13].

Um zukünftig weiterhin einen technischen Fortschritt erzielen zu können, muss an neuen Technologien geforscht werden, die das Problem einer minimalen Strukturgröße von Bauteilen umgehen.

Beispielhaft kann dabei die Forschung an den Quantencomputern, den neuromorphen Chips oder den Bau von dreidimensionalen Prozessoren genannt werden. Um den technischen Fortschritt weiter aufrechterhalten zu können, ist die Forschung in diesen Bereichen sehr wichtig.

Literaturverzeichnis

- [GM65] G. E. Moore, „Cramming more components onto integrated circuits,“ 19. April 1965. [Online]. Available: http://www.monolithic3d.com/uploads/6/0/5/5/6055488/gordon_moore_1965_article.pdf. [Stand: 14. April 2016].
- [GM75] G. E. Moore, „Progress In Digital Integrated Electronics,“ 1975. [Online]. Available: http://www.eng.auburn.edu/~agrawvd/COURSE/E7770_Spr07/READ/Gordon_Moore_1975_Speech.pdf. [Stand: 6. April 2016].
- [GQ11] Prof. A. Walraff, „Grundlagen der Quantenmechanik Teil 2“, 2011. [Online]. Available: <http://qudev.ethz.ch/content/courses/phys4/index.html>; http://qudev.ethz.ch/content/courses/phys4/phys4_fs11/phys4_11_Kapitel9_Teil2.pdf. [Stand: 25. April 2016]
- [HC13] „The Chip Design Game at the End of Moore’s Law“, 26. August 2013 (25. Hot Chips), Online. Available: http://www.hotchips.org/wp-content/uploads/hc_archives/hc25/HC25.15-keynote1-Chipdesign-epub/HC25.26.190-Keynote1-ChipDesignGame-Colwell-DARPA.pdf; <https://www.youtube.com/watch?v=JpgV6rCn5-g> [Stand: 4. April 2016]
- [ITW] „IT-Wissen Lexikon: Wafer“ <http://www.itwissen.info/definition/lexikon/Wafer-wafer.html> [Stand: 28. April 2016]
- [MtM] W. Arden, M. Brillouët, P. Coge, M. Graef, B. Huizing und R. Mahnkopf, „More-than-Moore“, S. 8 [Online]. Available: http://www.itrs2.net/uploads/4/9/7/7/49775221/irc-itrs-mtm-v2_3.pdf. [Stand: 16 April 2016].
- [PLHT] P. Laube, „Halbleitertechnologie von A bis Z,“ [Online]. Available: <https://www.halbleiter.org/lexikon/H/Half-Pitch/>. [Stand: 23. April 2016].
- [WIKIa] „Austrittsarbeit von Silicium-Elektronen“, [Online]. Available: https://en.wikipedia.org/wiki/Work_function [Stand: 25. April 2016].
- [WIKIb] „Molares Volumen von Silicium“, [Online]. Available: <https://de.wikipedia.org/wiki/Silicium> [Stand: 25. April 2016].
- [WIKIc] „Dr. Robert P. Colwell Biografie“, [Online]. Available: https://en.wikipedia.org/wiki/Bob_Colwell [Stand: 25. April 2016].

Anhang

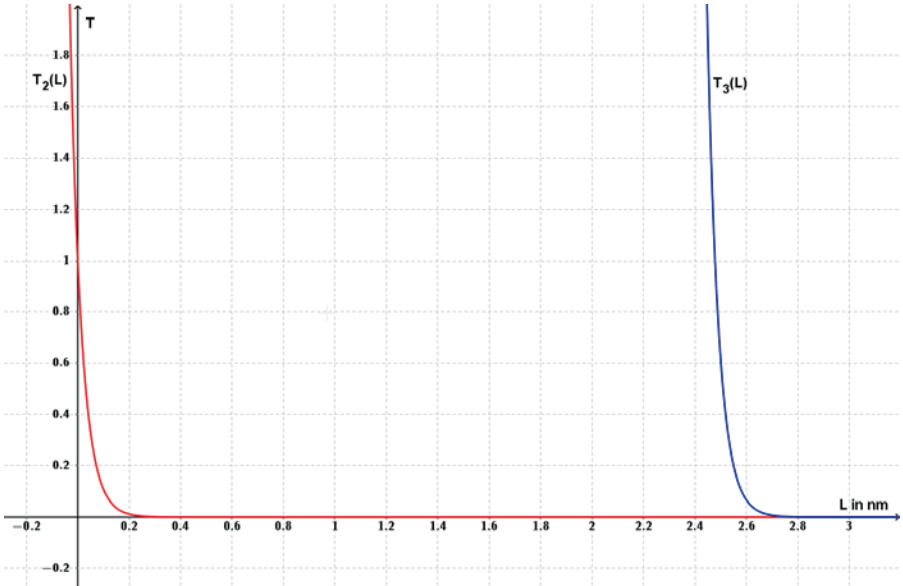


Abb. 1: Transmissionswahrscheinlichkeit in Abhängigkeit von dem Abstand L zwischen den Leiterbahnen

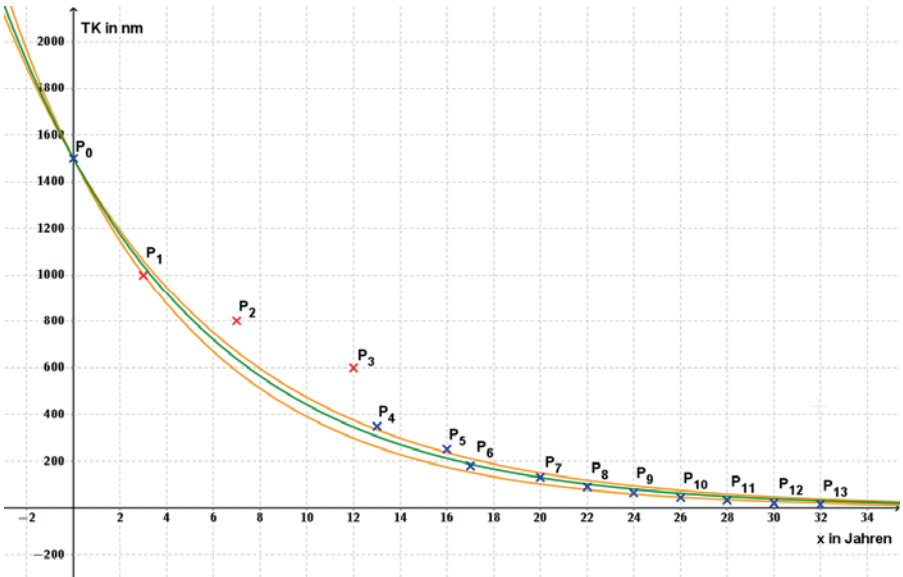


Abb. 2: Technologieknotenverlauf mit Ober-, Untergrenze und Idealfunktion

Domain-specific recommendation based on deep understanding of text

Sigurd Sippel¹

Abstract: This paper considers the process of development for a domain-specific recommender system that uses the domain of cocktail recipes as a example for experiments. Based on ontology a deep understanding of text is created — recipes are considered. The ontology is designed by basic categories to extract features such as ingredients. Ingredients are modeled by flavors for comparability. The process of data processing along with the recommendation extract over 2.000 recipes based on a ontology with over 1.000 ingredients. The key of the recommendation is based on domain-specific distance functions. A nearest-neighbor approach is used to classify recommendations for a given favorite. Validation is considered based on the acceptability of domain experts.

Keywords: Content-based recommender systems, Data Mining, Deep Understanding, Feature extraction, Ontology, Basic Categories, Validation, Domain experts

1 Introduction

In order to understand the recommendation process, a specific domain is used for experiments that are focused on deep understanding of text. Deep understanding [ASdB08] leads to a rich semantic representation of data, which is necessary for content-based recommendation. As an example of a specific domain, the domain of cocktails is chosen because it is definite and documented by bartending manuals and books of cocktail recipes written by domain experts. The deep understanding such as flavors of ingredients enriches the recommendation in the perspective of perception. Domain experts are interviewed to get feedback on the recommendation quality.

Section 2 considers the objectives. In section 3 it follows the related work. To achieve the objectives following four challenges are considered: In section 4 a domain-specific survey with domain experts is used to understand the field of cocktail recipes (challenge one) to process a huge volume of recipes (challenge two). The aim is to learn how recipes depends on recommendation. An ontology is designed to store the features such as ingredients in hierarchy. For challenge three section 5 describes domain-specific distances between classic recipes. The last experiment in section 6 considers a validation of nearest-neighbor recommendation (challenge four). The last section 7 considers the conclusion and future work.

¹ Hamburg University of Applied Sciences, Department of Computer Science, Berliner Tor 7, 20099 Hamburg, sigurd.sippel@haw-hamburg.de

2 Objective

Cocktails are written down as cocktail recipes that contain a name, ingredients including quantity, and partially short information about preferred glassware and preparation.

Manhattan Cocktail ²
1 dash of gum syrup, very carefully;
1 dash of bitters (orange bitters);
1 dash of curacao, if required;
1/2 wine glass of whiskey;
1/2 wine glass of sweet vermouth;
stir up well; strain into a fancy cocktail glass;

These recipes are available in cocktail books³, blogs⁴, or cocktail databases⁵. The sources present a huge volume of data, which is already available and increases with time.

There are different types of cocktails: Besides classic cocktails such as a Manhattan, which are cold and contain only liquid ingredients, there are hot cocktails and molecular recipes containing drops or foams. This approach focuses on the classic recipes with two or more recipes that contain partially a cherry, a zest, or mint but are basically liquid. If it is liquid, the result is a mixture containing all ingredients of this recipe. This approach assumes that a recipe results from a single mixture and each necessary ingredient is already prepared.

Cocktail recipes contain relevant information to prepare a specific cocktail. Partially a longer descriptive text is available, but the main information is written down in a short, compressed style of language.

This paper considers a recommendation is based on content-specific features such as ingredients and their characteristics. These features are extracted from cocktail recipes. Implicit personalization is modeled with the help of an exemplary favorite which tells something about the characteristic. It contains quantities, which put different ingredients in relation. This information is used to recommend cocktails.

A recommendation has to be appropriate for the guest; therefore, it has to capture the interest of the guest. It has to combine what he likes — and implicitly knows — as well as something new. Something he likes or is new could be a ingredient, a combination of ingredients or a specific flavor.

The main question is as follows: Does a knowledge-based distance function present a sufficient precision for a cocktail recommendation? A given recommendation is subjective therefore a recommendation for a specific domain — in this case, cocktails — can only be validated by acceptability survey of domain experts such as bartenders.

² 1882 Harry Johnson, Bartenders Manual p. 162

³ euvs-vintage-cocktail-books.cld.bz

⁴ www.winebags.com/50-Top-Cocktail-Blogs-of-2015/2910.htm

⁵ www.kindredcocktails.com

3 Related work

Domain-specific knowledge is necessary for deeper understanding of a domain to improve content-based recommendations [LdGS11]. Based on semantic modeled knowledge in ontology a deep understanding of text is possible [ASdB08].

A ingredient substitution recommendation based on a ingredient ontology with factors of perception is validated by domain expert [Bo14]. A graph-based recommendation approach is used for cooking recipes, which is focused on ingredients as nodes and preparation as edges [Wa08]. Based on recipe recommendations, the following step is to recommend a complete meal [Ku12]. This is called meal planning, which is usable for daily dinners or holiday events to obtain a meal including salads, appetizers, main dishes, and desserts. Another modeling approach is on the basis of nutritional balance [KF10]. The goal is to generate healthy meal plans. The user can get a completely auto-generated meal plan and can choose favorites, including self-monitoring of balance changes.

The acceptability factors of sensation of food include the following: Appearance, flavor, and texture [Bo02]. Jelinek's odor effects diagram describe flavors, which contains four main categories — acid, sweet, bitter, and animalic [Je97]. In a study uses odor databases [ZS09] that describe either semantically by a list of similar words or map a numeric value of an odor to reference materials. The result represents a database of similarities, which is called odorant object space. Besides the challenge to understand what a name of odor semantically represent, they show a high accordance between odorant object spaces and expert models such as Jelinek's.

4 Understanding the field of cocktail recipes (challenge one and two)

In the domain of cocktails, explicit assured knowledge about cocktails and the recommendation of cocktails is missing. There are manuals and cocktail recipe books, but the apprenticeship is based on voluntarism provided by accomplished bartenders who have written the books. There is no related research. Therefore, at first knowledge has to be received to find appropriate recommendations (challenge one). Domain experts are asked in a survey which parts of a cocktail recipe and which information about the guest are necessary for cocktail recommendations [Si16].

The target group comprises domain experts such as bartenders, bar owners, connoisseurs, and interested guests, who are invited to participate in the survey through online communities and social media portals such as Twitter. Twenty domain experts aged between 22 and 48 years answer all questions of the survey. Three people claim to work in a bar or own a bar. The rest consider themselves as connoisseurs or guests in a bar. Most of them have experiences in the domain of cocktails of about 3–10 years.

This qualitative survey shows which information a cocktail recommendation system can use to get an appropriate recommendation. The focus of a cocktail recipe is on the ingredients with their quantities. Preparation, glassware, and ice are not in focus, because this information can be derived from ingredients, opinion, and context. The recommendation

for a guest needs to be appropriate to their preferences. Ingredients and in particular their flavors are useful features to describe these preferences. Using a given favorite, the interviewed person recommends either with a focus on the ingredients of the favorite or with a focus on the flavors such as sourness and sweetness and alcohol ratio. These experiments give the first understanding of domain in the first challenge, which forms the basis to extract features out of recipes.

The aim of the experiment is to learn how a cocktail recipe is constructed and which information is extractable for further experiments. A library of 2,155 recipes extracted successfully. The detailed process of parsing is described in [Si16, p. 60]. Following assumptions how recipes works are core aspects of the parsing process (challenge two):

- Recipes contain many different spellings such as *sugar syrup* or *simple syrup*, as well as singular and plural words. These spellings are persistent in the ontology as alternative synonyms. If the spellings differs in clause position a rule is needed to convert the spellings.
- Recipes contain the known default names of ingredients. Since recipes need to be short, ingredient names are as short as possible. The problem is that the names are not distinct. *Chartreuse* is a company, but usually the product *Chartreuse Verte* is meant. The *vermouth* is a category, but *red vermouth* is meant; therefore, *vermouth* is a *superordinate* and also *vermouth* is added to basic category *red vermouth* as a synonym. The most concrete item have to be chosen by entity recognition.
- Recipes contain numbers and fractions as words such as *one-third*. It needs synonyms of numbers or fractions in the ontology. A conversion to digits is necessary. Recipes contain ranges of quantities. It often means seasoning an ingredient.
- Recipes also contain fillers such as *soda*, which are ingredients without a concrete quantity. However, that does not mean a *dash* or a *splash*, which is always a small quantity. A filler could be about 10 *cl* and therefore it is tendentially the most important ingredient. The chosen concrete quantity must be realistic in terms of the glassware.
- Recipes in historic books contain or-relations such as *bourbon* or *rye*. For example, either bourbon or rye has to be used, not both. Recipes also contain optional ingredients.
- Recipes contain solid ingredients. The mapping of solids to liquids allows one to find better similarities with other recipes. Converting the measurements is not enough, because it is necessary to combine a qualitative unit such as *half* with an ingredient such as *lemon*. The ontology has to know that one *lemon* contains about 5 *cl*, in order to convert this correctly. The conversion is declared in ontology.
- Quantities are implicit if they are usual (Egg is shortened form of one piece of egg). Items of preparation such as stir or shake, drinking glass, preparation glass, or preferred ice contain many recipes, but every type of item could be missing.

- Ingredients are known by names. If a name is a universal one, which is contained in dictionaries or is a public brand, the ingredient is understandable by every domain expert. If it is a very special name, a recipe for the ingredient is necessary. For this approach, ingredients are assumed to have universal names. It is also assumed that the recipes are thoroughly mixed.

The target structure is the result of manual extraction by a domain expert and describes one cocktail recipe. A flexible structure is required to extract different styles of cocktail recipes. The extracted features represent the internal representation (Equation 1). It is a technical presentation that is necessary for the recommendation.

$$\textit{trait Item}\{val\ uri : String\} \quad (1)$$

The URI guarantees unique identification. Different spellings, which are extracted to the same identifier, could be interpreted as the same. The user needs to understand and classify the extra information attached to the recipe such as the name, the original spelling of an assignment, and meta-information about the book and the author. The representation, which contains information for the user, is the external representation (Equation 2). The result is one data structure that represents the internal and external data.

$$\textit{trait ValueItem}\{val\ i : Item, val\ name : String\} \quad (2)$$

The assignment list contains a sequence of items and a quantity. The sequence shows that only one has to be chosen. This sequence is defined as a *or – relation* of items. Allowed items are touchable such as ingredients, glassware, or ice. Preparations cannot be an assignment. The cocktail data structure combined all information about a cocktail. A cocktail needs a name, but all other values such as assignments are optional. Preparation, glassware and ice are subtypes of item, which represents one taxonomy in the ontology.

5 Distances between classic recipes (challenge three)

For recommendation a distance measurement is considered in this experiment. It is assumed that classic recipes have been known for a long time, because they contain a characteristic that isolates them from each other. 52 recipes are clustered by domain experts into 19 clusters [Si16, p. 130] and extracted to measure how well the distances work. This is the first step to get an idea about how distances work.

The similarity between items is defined by shared categories in the ontology. The type is referred to the imaginable class. All classes that do not present superordinates are subclasses of the imaginable class. The basic ingredient categories and ingredient subordinates are subclasses that represent basic categories such as *gin* and subordinates such as *London dry gin*. The superordinates such as *spirits* are clearly excluded, because the shared properties between two *spirits* such as *absinthe* and *gin* are too low.

The core ontology for more abstract categories contains more than 200 ingredients and the extended ontology contains over 1.000 ingredients.

The result presents a list of ingredients showing the ingredient path in the ingredient tree. The searched ingredient is always the first item in the path. In the example (Figure 1), there is a subordinate ingredient *Plymouth*, which has a parent *gin* as a basic category of ingredients, as well as a superordinate *spirits*, which is not declared as an imaginable ingredient.

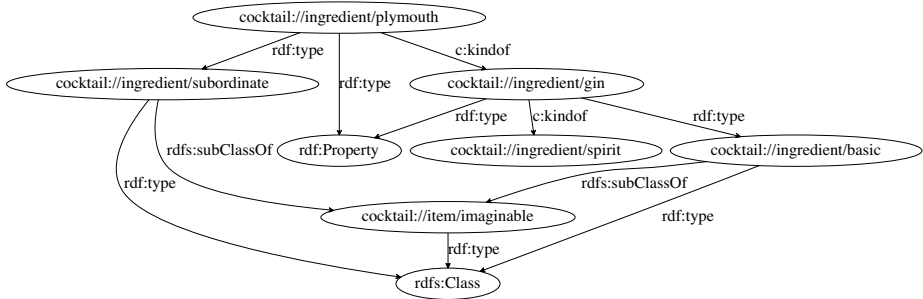


Fig. 1: Example of a ingredient categorization

The path of *Plymouth* contains itself and the parent *gin* (Equation 3). The superordinate is ignored and the types are represented by the chosen data structure such as *BasicIngredient*.

$$\begin{aligned}
 path_1(Plymouth) = & \text{SubordinateIngredient}(\text{cocktail://ingredient/plymouth}) :: \quad (3) \\
 & \text{BasicIngredient}(\text{cocktail://ingredient/gin}) :: Nil
 \end{aligned}$$

In addition the weight of an ingredient for distance function is defined by ratio referred to the total volume, therefore the used quantity have to be extracted. For a comparable quantity, the unit has to be normalized. The main task of the unit in the ontology is to identify measurement units and to convert them into the standard unit *cl*. This conversion normalizes the quantity. The convertible measurement units are separated into quantitative and qualitative units. Quantitative units such as *cl* are scalable, while qualitative units such as *dash* are not. There are metric units such as *ml* and American or British units such as *ounce*. For non-metric units, there are synonyms like singular and plural words. Pairs of ingredients and units such as *splash champagne* have default values, because these pairs are imprecise, therefore these pairs substituted [Si16, p. 78] into metric and quantitative units.

5.1 Balance

The balance is an abstract perspective on the cocktail which leans on Jelineks odor model. The result of the survey based on appropriate features for recommendation are flavors (see survey), a extract of the most important ones to describe the classic recipes are chosen: The cocktail balance represents six pieces of information — the amounts of sweet, sour, water, cream, bitter, and alcohol. These features are developed by describing classic recipes

by domain experts and are qualitatively determined information. Alcohol is an exception because the ratio is available. It is necessary to get these six features for every ingredient. However, not all of this information is always available and the ontology does not contain all the information. Therefore, it needs a default logic approach. For example, the ontology does not contain balance information for a concrete gin product, but the balance of the gin prototype is known. Besides, the balance information of gin has to be used.

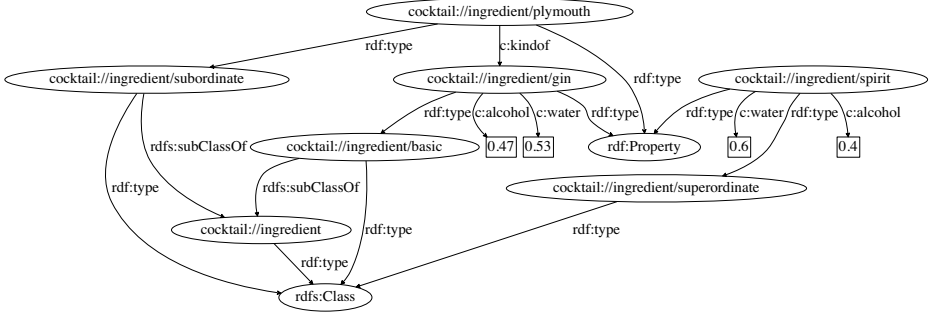


Fig. 2: Balance in ingredient categories

In this example, the given ingredient *Plymouth* does not have balance information. The basic category *gin* has alcohol and water in the proportion of 0.47 and 0.53, respectively. The superordinate has alcohol and water in the proportion of 0.4 and 0.6, respectively. As sweetness is not declared, the default value of the balance property, which is not found, stands at 0.

The path contains the balance information of all the single ingredients — first *Plymouth*, then *gin*, and finally *spirits* (Equation 4). The question mark is used as a symbol to indicate that the information remains unknown.

$$\begin{aligned}
 & bal(\text{water}, \text{alcohol}, \text{sweet}, \text{sour}, \text{cream}, \text{bitter}) \quad (4) \\
 path_B(\text{Plymouth}) &= (?, ?, ?, ?, ?, ?) :: (0.53, 0.47, ?, ?, ?, ?) :: (0.6, 0.4, ?, ?, ?, ?) :: Nil \\
 bal(\text{Plymouth}) &= (0.53, 0.47, 0, 0, 0, 0)
 \end{aligned}$$

5.1.1 Ingredient distance

The distance of a ingredient pair (I_a, I_b) is a path distance (Equation 5), which uses a declared path of two ingredients in the ontology. A quantity weighting is added because the quantity tells something about the importance. 6 *cl gin* are more important than 1 *cl sugar syrup*. The weight is the quantity in relation to the volume of the cocktail. The volume is the sum of quantities of all quantitatively measured ingredients. All quantities are transformed into the standard unit *cl*.

$$d_{DPI}(a, b) = stepDistance(I_a, I_b) \cdot \frac{quantity(I_a)}{volume(a)} \quad (5)$$

The distance of steps has the lowest value 0 if both ingredients remain the same but the quantity is different the quantity-based distance function (Equation 6) is used, which calculates a normalization related to the volume. The DPQ needs a weight w to prevent too high distances compared to DPI , because this is only used for equal ingredients. A proper weight based on the experiment is 0.25.

$$d_{DPQ}(a, b) = \left| \frac{quan(I_a)}{vol(a)} - \frac{quan(I_b)}{vol(b)} \right| \cdot w \quad (6)$$

The distance of a ingredient pair is dependent on the distance of steps (Equation 7).

$$d_{DP}(a, b) = if(stepDistance == 0) d_{DPQ}(a, b) else d_{DPI}(a, b) \quad (7)$$

A cocktail recipe contains a list of ingredients. The order must not affect the distance, because the order could be different and don't change the recipe. If there is an ingredient I_a of the cocktail a , the aim would be to find the most similar ingredient to I_a in the ingredients of cocktail b . The number of ingredients of a are n . The number of ingredients of b are m .

The distance d_I (Equation 8) between ingredients of recipe a and the ingredients of b represents the ingredient distance between two recipes. It uses the distance d_{DP} , which maps an ingredient to another ingredient. A mapping is not completely accurate, the distance must be calculated in both directions to catch all the ingredients in the distance. The distance d_I sums up all minimum d_{DP} distances in both directions.

$$d_I(a, b) = \frac{\sum_{i=1}^n \arg \min(d_{DP}(I_{a_i}, I_{b_j})) + \sum_{j=1}^m \arg \min(d_{DP}(I_{b_j}, I_{a_i}))}{2} \quad (8)$$

5.1.2 Balance distance

The balance distance shows how different recipes are with respect to balance. The aim is to find cocktails with the same characteristics. Every ingredient has a balance. The balance of a cocktail is the sum of balances of n ingredients (Equation 9).

$$bal(c) = \sum_{i=1}^n bal_i(water, alcohol, sweet, sour, bitter, cream) \cdot \frac{quan(I_i)}{vol(c)} \quad (9)$$

$$d_B(bal) = water + alcohol + sweet + sour + bitter + cream \quad (10)$$

$$d_B(c_a, c_b) = d_B(|bal(c_a) - bal(c_b)|) \quad (11)$$

The difference between two balances (Equation 10) is a balance having a difference in each component, such as *sour*. The balance distance is the difference between the final balance of c_a and c_b (Equation 11). All components will be added up to a scalar distance.

6 Validation by domain-experts (challenge four)

Based on existing distance measurement this experiment validates the recommendation by domain experts (challenge four). The last experiment uses the extracted recipes represented in resulted target structure. An ingredient-based distance as well as a balance-based distance function is defined based on the extracted features by a huge number of recipes. The ingredient distance demonstrates the uniqueness of classic recipes, while the balance show a similar characteristic, which is an example of a good recommendation. This last experiment combines these results to get a working recommendation system.

Classic recipes are the popular ones. Therefore, it is assumed that these are preferred examples of recommendation. The results of recommendation are validated by domain experts to get feedback on the results. A recommendation needs to combine something known with something new, in context of the given distance functions there are two approaches of recommendation — the first is used to get recipes with the same balance but different ingredients and the second is used to get recipes of the same ingredients but with a different balance.

The recommendation approach uses the nearest-neighbor classification kNN of a given favorite. A analysis of coherence and distinction of classic recipe clusters results a empiric value of distance, which separates the distances into too near distance and distances which shows significant differences [Si16, p. 99]. In the first instance, called focus on balance, the nearest neighbors have an ingredient distance d_I higher than $t_I = 0.3$ and a balance distance lower than $t_B = 0.3$. Too low distances of ingredients are too similar while too high distances of balance are too different. The recommendation r gives a list of cocktail recipes. This is ordered increasingly according to ingredient distances. The first $k = 10$ elements are considered as the most important and are used for recommendation. The second instance, called focus on ingredients, uses $t_I = 0.4$ as the maximum threshold of ingredient distance and $t_B = 0.4$ as the minimum threshold of distance of balance. If the focus is on balance, the balance distance has to be very low, because balance distance does not show which component of balance such as *sweet* is different. If the distance is caused in only one component, the change is higher than it is distributed on all components. The focus on ingredient approach needs an higher threshold because it is more differences between the recipes necessary to get enough results.

The offline experiments with a static testing set and feedback by domain experts is used to test whether a recommendation is appropriate. A specific group of domain experts — such as bartenders or connoisseurs — was offered the examples and a list of recommendations. The domain experts rated the validity of each recommendation on a numeric scale (Equation 12). This scale is designed to present how acceptable a recommendation is.

$$\left[\begin{array}{cccccc} -2 & -1 & 0 & 1 & 2 \\ \text{(unacceptable)} & \text{(slightly similar)} & \text{(obviously)} & \text{(rather appropriate)} & \text{(appropriate)} \end{array} \right] \quad (12)$$

19 classic recipes [Si16, Appendix B] used as a favorite to calculate the recommendations. These process is either done for the a focus on balance approach (in total 181 recommen-

dations) and for the focus on ingredient approach (in total 141 recommendations). The pairs of favorite recipe and recommended recipe are rated by the domain experts.

6.1 Acceptability of domain-experts

Four domain experts are interviewed for validation, three are independent and additionally one is dependent to development, who rated in total 1288 pairs of favorites and recommendations (extract in [Si16, p. 145]). The independents are briefed shortly, which is the idea behind the two approaches of recommendation. They are supposed to use the same numeric scale while creating their own validation criteria. If they use the given criteria, they are not independent.

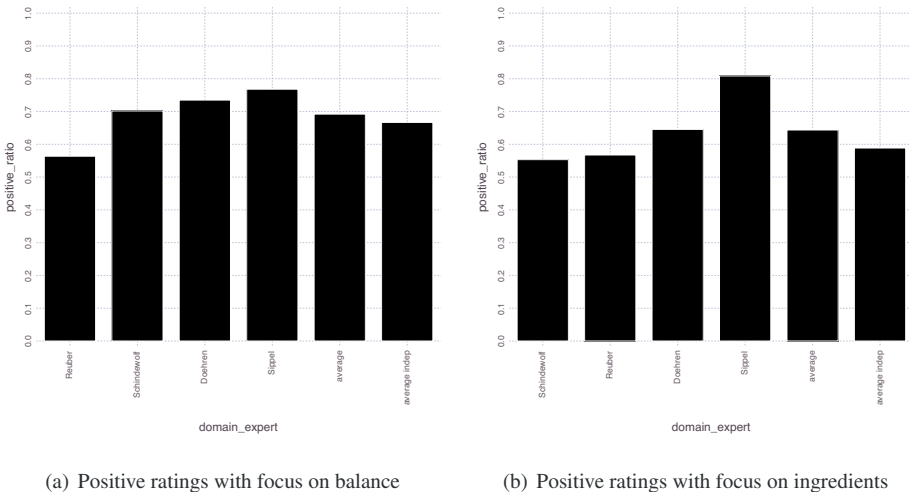


Fig. 3: Positive ratings of domain experts

The rate of positive ratings of focus on balance (Figure 3(a)) is on average about 69 %. The dependent domain expert gives the highest ratings, but the independent average value has a value of about 67 %, which is very close to that. The ratio of positive rating in the focus of ingredients (Figure 3(b)) is about 64 %. However, the independent average is only 59 %. This shows less acceptance of this approach as well as fewer objective ratings of the dependent one.

The domain experts need 3–4 hours to fill the rating sheet, which shows how time consuming the knowledge elicitation of domain expert is. The domain experts give feedback that their recommendations which are appropriate to the favorite, but they would not recommend that because the recipe itself was not persuasive for their expectations of quality. Therefore, a quality measurement is necessary to increase the precision of recommendation.

This result qualifies the recommendation focused on balance for validation with more domain experts such as in an online study. The recommendation focused on ingredients needs higher precision. A replacement of a favorite recipe with favorite ingredient is an opportunity that needs to be proved.

7 Conclusion and future work

The experiments uses a semi-automated and domain-specific process for recommendation which shows first acceptable results. Deep understanding is possible because it is used a limited domain with available background knowledge.

In review of this experiments the ground truth about a domain-specific recommender system is that the main interest of the user has to be in focus: To arouse the user's interest, it is necessary to find something known such as parts of a defined favorite and understand it in deeper way. The understanding is used to find something new. The modeling for such interests has to be according to the domain. Interviewing domain experts is a necessary precondition for extracting an abstract model. The extraction process is done with a huge volume of recipes. These have to be proceeded successfully before a validation of the recommendation by domain experts. A validation needs a lot of feedback from domain experts but it shows how acceptable this recommendation is. The personal opinion has to be dismissed to get a useful result, therefore domain expert have to evaluate the acceptance and not whether it is equal to its own chosen recommendation. If this steps are performed, then the validation will give a meaningful measurement of the quality of recommendation. The validation shows that the used process is functional.

For optimizations also the combination with contextualization and individualization should be considered. In perspective of individualization the user model is extensible with further favorites or dislikes, in order to get a higher precision of recommendation. Assumed a huge database of recipes is given, the contextualization such as changes in process of time should be considered. This is a basis for analysis of which kinds of ingredient or recipe will be the trend of tomorrow. Assuming precise recommendations are available, a kind of meal planning is a research opportunity: The transferability of meal planning of cooking recipes to cocktail recipes should be proved, which means recommending a follower of a given drink to plan the time of a guest in bar.

For specific domains such as news deep understanding could be working, therefore it is a possible research question to prove how it is possible to integrate several domain-specific recommender systems in a bigger recommender system, which classify automatically which specialized recommender system is qualified for a specific query of recommendation.

Bibliography

- [ASdB08] Allen, James F.; Swift, Mary; de Beaumont, Will: Deep Semantic Analysis of Text. In: Proceedings of the 2008 Conference on Semantics in Text Processing. STEP '08, Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 343–354, 2008.

- [Bo02] Bourne, Malcolm: Food texture and viscosity: concept and measurement. Academic press, 2002.
- [Bo14] Boscarino, Corrado; Nedović, Vladimir; Koenderink, Nicole J. J. P.; Top, Jan L.: Automatic Extraction of Ingredient's Substitutes. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. UbiComp '14 Adjunct, ACM, New York, NY, USA, pp. 559–564, 2014.
- [Je97] Jellinek, Paul: Perfumery and eroticism. In (Jellinek, J.Stephan, ed.): The Psychological Basis of Perfumery, pp. 3–3. Springer Netherlands, 1997.
- [KF10] Karikome, Shihono; Fujii, Atsushi: A System for Supporting Dietary Habits: Planning Menus and Visualizing Nutritional Intake Balance. In: Proceedings of the 4th International Conference on Uniquitous Information Management and Communication. ICUIMC '10, ACM, New York, NY, USA, pp. 56:1–56:6, 2010.
- [Ku12] Kuo, Fang-Fei; Li, Cheng-Te; Shan, Man-Kwan; Lee, Suh-Yin: Intelligent Menu Planning: Recommending Set of Recipes by Ingredients. In: Proceedings of the ACM Multimedia 2012 Workshop on Multimedia for Cooking and Eating Activities. CEA '12, ACM, New York, NY, USA, pp. 1–6, 2012.
- [LdGS11] Lops, Pasquale; de Gemmis, Marco; Semeraro, Giovanni: Recommender Systems Handbook. Springer US, Boston, MA, chapter Content-based Recommender Systems: State of the Art and Trends, pp. 73–105, 2011.
- [Si16] Sippel, Sigurd: Domain-specific recommendation based on deep understanding of text <http://users.informatik.haw-hamburg.de/~ubicomp/arbeiten/master/sippel.pdf>. 2016.
- [Wa08] Wang, Liping; Li, Qing; Li, Na; Dong, Guozhu; Yang, Yu: Substructure similarity measurement in chinese recipes. In: Proceedings of the 17th international conference on World Wide Web. ACM, pp. 979–988, 2008.
- [ZS09] Zarzo, Manuel; Stanton, DavidT.: Understanding the underlying dimensions in perfumers odor perception space as a basis for developing meaningful odor maps. Attention, Perception, Psychophysics, 71(2):225–247, 2009.

A Parallel Non-Hydrostatic Shallow Water Model on Adaptive Triangular Meshes in sam(oa)²

Philipp Samfaß¹ and Raphael Schaller²

Abstract: Even with current extreme scale systems, the accurate simulation of tsunamis continues to be a challenging problem. One commonly used model for this task are the *hydrostatic* shallow water equations which, however, are not able to represent all relevant physical effects of tsunamis. In this paper, we therefore show how to solve the *non-hydrostatic* shallow water equations in parallel within the partial differential equation framework sam(oa)² by extending the existing finite volume method for solving the hydrostatic shallow water equations. We present an element-oriented dual grid discretization of the equation for the non-hydrostatic pressure on triangular adaptive meshes which allows for a matrix-free conjugate gradient solver for the corresponding system of linear equations. In addition, the validity of the new model is shown based on two common numerical experiments for non-hydrostatic shallow water models. Because we aim at solving large-scale problems on high performance computing architectures, we demonstrate our distributed memory parallelization which resulted in a parallel efficiency of 90.1% from 1 to 8192 cores for weak scaling and 96.8% from 16 to 512 cores for strong scaling.

Keywords: Adaptive triangular mesh, conservation law, dual grid, element-oriented method, finite volume method, high performance computing, matrix-free solver, non-hydrostatic shallow water equations, parallelization, tsunami

1 Introduction

During the past few decades, computation resources on extreme scale systems have significantly increased. This has resulted in an increasing interest of scientists in using such high performance machines in order to simulate important problems stemming from natural sciences and engineering. One of these problems is the modelling and prediction of tsunamis. Considering the devastating consequences of a tsunami such as the one caused by the Tōhoku earthquake in Japan in 2011, the critical need for accurate computational models becomes evident. Valid mathematical and physical models should be efficiently implemented on current high performance systems. This is a requirement to obtain meaningful insights at scale into the development and propagation of tsunamis.

Even with current supercomputers having their peak performance in the petaflop regime, a full three-dimensional model for the simulation of a tsunami might still not be computationally feasible. Instead, the so-called shallow water equations provide a reasonable two-dimensional approximation by essentially depth-averaging the full three-dimensional governing equations. A model that is commonly used—the hydrostatic shallow water model—neglects the impact of the vertical velocity and assumes that the pressure is given by the

¹ Technical University of Munich, Department of Informatics, samfass@in.tum.de

² Technical University of Munich, Department of Informatics, schaller@in.tum.de

hydrostatic pressure (the result of gravity). This applies to (long) shallow water waves that are characterized by $h/\lambda \ll 1$, where h is the height of the water column and λ the wave length [LN08]. However, the corresponding hydrostatic shallow water equations do not model the dispersion of deeper water waves that might e.g. occur near the coast [LN08].

Therefore, efforts have been made to extend existing hydrostatic solvers towards a modified non-hydrostatic shallow water model (e.g. in [Cu13], [Fu13], [SZ03] or [Wa05]). Here, the general approach is to correct the solution produced by the hydrostatic solver in each time step while still retaining a 2D depth-averaged model. As will be discussed later, this correction requires to solve a system of linear equations in each time step.

In this paper, we first show how to adapt this scheme to solving the non-hydrostatic shallow water equations using a conservation-law-based finite volume method for solving the hydrostatic shallow water equations (Sec. 2.1). In Sec. 2.2, we present an element-oriented approach towards solving the system of linear equations for the non-hydrostatic pressure using a dual-grid discretization. Our extension has been integrated into `sam(oa)2` [MRB11] which is a parallel framework for solving partial differential equations (PDE) on adaptive triangular meshes using space-filling curves for cache and memory efficient grid traversals. Based on some common experiments for non-hydrostatic models, we show the validity of the new model (Sec. 4.1) and investigate the feasibility of using a conjugate gradient solver for the system of linear equations (Sec. 4.2).

Since we aim at solving large-scale problems on high performance computing (HPC) systems, we furthermore discuss our parallelization strategy for the model which is based on `sam(oa)2`'s parallelization interface (Sec. 3). Finally, strong and weak scaling experiments verify a successful parallelization (Sec. 4.3).

2 Numerical Approach

In this section, we briefly introduce the non-hydrostatic shallow water equations, the assumptions on which they are based and the general approach towards solving them. Next, we present our element-oriented spatial discretization on triangular adaptive meshes in `sam(oa)2` that allows for a matrix-free linear solver.

2.1 Finite Volume Formulation of the Non-Hydrostatic Shallow Water Equations

The subsequently described numerical approach for the non-hydrostatic extension closely follows the one presented in [Cu13] and [Fu13]. Special attention is paid to the modifications required to combine this approach with our conservation-law-based finite volume solver (cf. [Le02]) for the hydrostatic shallow water equations.

As a basis for the extension, we use the non-hydrostatic shallow water equations in conservation law form (see [Sa14] for a derivation based on [Ma]):

$$\left[\begin{array}{c} h \\ hU \\ hV \\ hW \end{array} \right]_t + \left[\begin{array}{c} hU \\ hU^2 + \frac{1}{2}gh^2 \\ hUV \\ hUW \end{array} \right]_x + \left[\begin{array}{c} hV \\ hUV \\ hV^2 + \frac{1}{2}gh^2 \\ hVW \end{array} \right]_y = \left[\begin{array}{c} 0 \\ -ghb_x - \left(\left[\frac{h\hat{q}}{2} \right]_x + \hat{q}b_x \right) \\ -ghb_y - \left(\left[\frac{h\hat{q}}{2} \right]_y + \hat{q}b_y \right) \\ \hat{q} \end{array} \right] \quad (1)$$

Here, $h(x,y) = \eta(x,y) - b(x,y)$ denotes the height of the water column, $\eta(x,y)$ the free surface elevation above the mean sea level, $b(x,y)$ the bathymetry (elevation of the sea floor), (U, V, W) the depth-averaged velocity vector, g the gravity of earth and a subscript the partial derivative with respect to the given coordinate. Further, $q(x,y,z)$ represents the non-hydrostatic portion of the pressure and $\hat{q} = [q]_{z=b}$ the non-hydrostatic pressure at the sea bottom. The hydrostatic shallow water equations are obtained as a special case by setting $q = 0$ and neglecting the vertical momentum equation.

The above equations include the assumption that the pressure can be decomposed into a hydrostatic part p_H and a non-hydrostatic portion q as introduced in [CS98]. We further assume a linear vertical distribution of the non-hydrostatic pressure (decreases to zero at the surface, maximum at the bottom) and the vertical velocity component (maximum at the water surface) [Wa05]. Since bathymetry is represented as a step function in our model, the vertical velocity w is zero at the sea bottom [Cu13]. Under these assumptions, $W = 0.5[w]_{z=\eta}$ holds.

We employ a fractional step scheme towards extending the previously existing hydrostatic solver similar to the one used in [Cu13]: in each time step, the quantities of the hydrostatic solution will be computed and then corrected with the numerical non-hydrostatic pressure parameter \hat{q} at the sea bottom—the new unknown in the non-hydrostatic model. This gives rise to the need of solving a system of discretized Poisson-like equations for \hat{q} . The system is obtained using the pressure projection method ([Ch68]) that is e.g. commonly used to solve the Navier-Stokes equations [GDN95]. Our correction formulas for the discharges hU, hV computed by the hydrostatic solver follow from an explicit Euler time discretization of the non-hydrostatic equations (1). For correcting the vertical velocity field, we follow [Cu13] and neglect the non-linear terms in the vertical momentum equation. In summary, the main steps in a non-hydrostatic timestep are (in this order): compute a timestep with the hydrostatic solver (resulting in an intermediate solution), solve a linear system of equations for \hat{q} and correct the intermediate solution with \hat{q} .

2.2 Dual Grid Discretization on Triangular Meshes in sam(oa)²

Sam(oa)²'s particular setup allows efficient access only to element-local data in each grid traversal step such that an element-oriented and matrix-free assembly of the system of linear equations for \hat{q} is required. Instead of collocating the control volumes for computing the non-hydrostatic pressure with the finite volume cells, we therefore propose a dual grid arrangement as shown in Fig. 1.

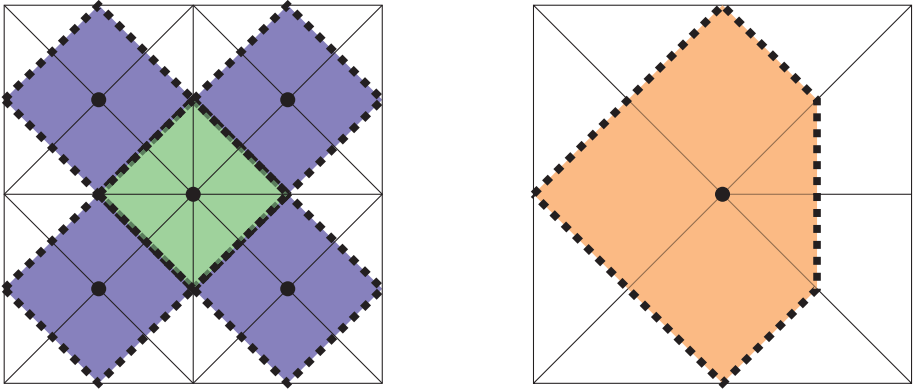


Fig. 1: Control volumes/dual grid cells (dashed) for the non-hydrostatic pressure: the left picture shows the two kinds (green and blue) of dual grid cells that might occur in a grid of uniform refinement depth. On the right, a control volume for the non-uniform case is depicted. Note that an element will contribute to three different control volumes. The points denote the nodes on which the unknowns \hat{q}_d and the vertical velocity w_d for a dual grid cell d are located. For each \hat{q}_d , there is exactly one corresponding control volume/dual cell d .

Our finite volume method for solving the hydrostatic shallow water equations is basically a Galerkin method with piecewise constant ansatz and test functions that determines for each primary cell an average $Q_i^n = (h_i^n, [hU]_i^n, [hV]_i^n)^T$. For compatibility reasons, the same choice of basis and ansatz functions has also been made for discretizing \hat{q} and w .

With these remarks, the basis for deriving the system of linear equations for the non-hydrostatic pressure will be the following weak form of the continuity equation given for a dual grid control volume:

$$\int_{\Gamma_d} n_x hU \, d\Gamma + \int_{\Gamma_d} n_y hV \, d\Gamma + \int_{\Omega_d} w \, d\Omega = 0 \tag{2}$$

Here, Γ_d denotes the boundary, $\mathbf{n} = (n_x, n_y)$ its outward pointing normal unit vector and Ω_d the region of the dual cell d . Note that some terms occurring in the derivation of this formula had to be neglected since they cannot be directly obtained from the hydrostatic quantities computed in our model³.

Let N_d denote the set of primary elements that contribute to the dual cell d , $\Gamma_{e,d}$ the part of the boundary of d that lies in the primary element e and $\Omega_{e,d}$ the region of the dual cell that lies within e (compare Fig. 1). Then, an element-wise assembly of the equation as a sum of element contributions is reflected in the following equivalent form of (2):

$$\sum_{e \in N_d} \left(\int_{\Gamma_{e,d}} n_x hU \, d\Gamma + \int_{\Gamma_{e,d}} n_y hV \, d\Gamma + \int_{\Omega_{e,d}} w \, d\Omega \right) = 0. \tag{3}$$

³ For instance, $\int_{V_d} \frac{\partial u}{\partial x} \, dV = \int_{\Gamma_d} n_x hU \, d\Gamma + \int_{\Omega_d} u_{z=b} \frac{\partial b}{\partial x} - u_{z=\eta} \frac{\partial \eta}{\partial x} \, d\Omega$ holds and the second and third summands had to be neglected.

Plugging the correction formulas for the discharges hU, hV and the vertical velocity w into the summand of (3) gives the following dual cell equation contribution:

$$\int_{\Gamma_{e,d}} \mathbf{n} \cdot \left[\begin{pmatrix} \widetilde{hU}^{(n+1)} \\ \widetilde{hV}^{(n+1)} \end{pmatrix} - \Delta t \frac{h^{(n)}}{2} \nabla \hat{q} \right] d\Gamma + \int_{\Omega_{e,d}} \widetilde{w}^{(n+1)} + 2\Delta t \frac{\hat{q}}{h^{(n+1)}} d\Omega = 0. \quad (4)$$

In contrast to a collocated scheme, our dual grid arrangement avoids having to evaluate derivatives of hydrostatic quantities at primary cell boundaries (the location of a discontinuity between two cell averages) which would be particularly problematic in the adaptive case. The gradient of the non-hydrostatic pressure term $\nabla \hat{q}$ at the dual cell boundaries is approximated using central finite differences. Both the normal vectors and the non-hydrostatic pressure gradient have to be rotated appropriately from a canonical standard reference element orientation into the actual orientation of the considered element.

In our implementation, during a non-hydrostatic grid traversal (after the hydrostatic step), all three dual cell equation contributions (4) will be computed yielding 3×3 local element matrices for the unknowns on the three nodes of a primary element. In another traversal, the linear solver determines the local residuals for the dual cell equations using these element matrices. Finally, the unknowns \hat{q} on the nodes are updated according to the chosen iterative method for solving the system. In another traversal, the obtained solution will be used to correct the cell averages and the averaged vertical velocity.

3 Parallelization

In order to be able to solve large-scale problems, the non-hydrostatic extension is intended to being run on current HPC systems. While the linear solver and the hydrostatic component had already been parallelized in previous work, our non-hydrostatic extension required a solid parallelization strategy as presented subsequently.

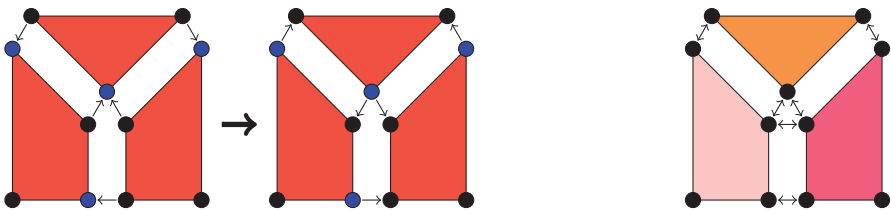


Fig. 2: Merging process. Left two pictures: Merging sections owned by one process. Rightmost picture: Merging sections owned by multiple processes. Color of sections denote their process affiliation. The blue marked nodes are master nodes. The figure is based on [MB15].

3.1 Parallelization in sam(oa)²

Sam(oa)² provides solid groundwork for parallelization: by ordering elements along a space-filling curve and then cutting sequentialized elements into sections, the grid is automatically split into atomic work units which are distributed to the processes. In each time

step, $\text{sam}(\text{oa})^2$ also transparently handles load balancing. When splitting the grid, border nodes belonging to multiple sections are duplicated (see Fig. 2). Therefore, only partial results are stored on the border nodes which requires a merging of the nodes. Merging is performed in different ways depending on whether the border nodes' sections belong to one process (shared memory) or to multiple processes (distributed memory), as depicted in Fig. 2. For distributed memory, a peer-to-peer merging is conducted, so each node is merged with each other node. For shared memory, $\text{sam}(\text{oa})^2$ merges the nodes master-slave-wise: first, all slave nodes are merged into one master node, then the master's result is copied back to the slaves.

3.2 Merge Operators

Merge operators can be used to define how $\text{sam}(\text{oa})^2$ merges the nodes. A merge operator takes two nodes, the local node and the neighbor node, with the latter being read only. The merge operator then shall merge the neighbor node into the local node.

Here, two merge operators are required: one after setting up the system of equations and one for the correction traversal. In the former, the partial right hand sides of the system of linear equations are merged. In the latter, we merge two helper variables which are required for the correction. For all these cases, we can abstract the problem to a sum being computed on the node, which sums information stored on the node's adjacent cells:

$$\hat{p}_n = \sum_{\forall m \in M_n} p_{m;n}, \quad (5)$$

where \hat{p}_n is the result to be stored on node n , M_n is the set of cells adjacent to this node and $p_{m;n}$ is the information on the cell m for this node. If node n is a border node, the adjacent cells are a subset of M_n , so only a partial sum is computed. Since $M_{n,1} \cup M_{n,2} \cup \dots \cup M_{n,k} = M_n$, where $M_{n,i}$ denotes the set of adjacent cells of node n in section i , we get

$$\hat{p}_n = \sum_{\forall m \in M_{n,1}} p_{m;n} + \sum_{\forall m \in M_{n,2}} p_{m;n} + \dots + \sum_{\forall m \in M_{n,k}} p_{m;n}. \quad (6)$$

Thus, the merge operators need to compute the sum of the partial results stored on the nodes. For merging the right hand side, it generally looks like this:

```
local_node.rhs := local_node.rhs + neighbor_node.rhs
```

4 Results

4.1 Physical Validation

In order to validate the new model, we conducted different common experiments for non-hydrostatic models. To reliably ensure convergence we used for these experiments a simple Jacobi solver with a maximum absolute local residual threshold of $\varepsilon = 0.1$. However, as we show in Sec. 4.2, the CG method is suitable for solving the system, too.

Standing Wave

As a first experiment, we simulated a standing wave in a closed (wall boundaries) cuboidal basin of quadratic shape with length 10m. This experiment has been conducted in other related work such as [Cu13], [Fu13] and [SZ03]. For details on the experimental setup and the analytical solution, please see [Fu13]. The constant depth of the basin d is incrementally increased yielding a higher ratio d/λ such that a hydrostatic model fails to approximate the correct propagation speed as can be seen in Fig. 3.

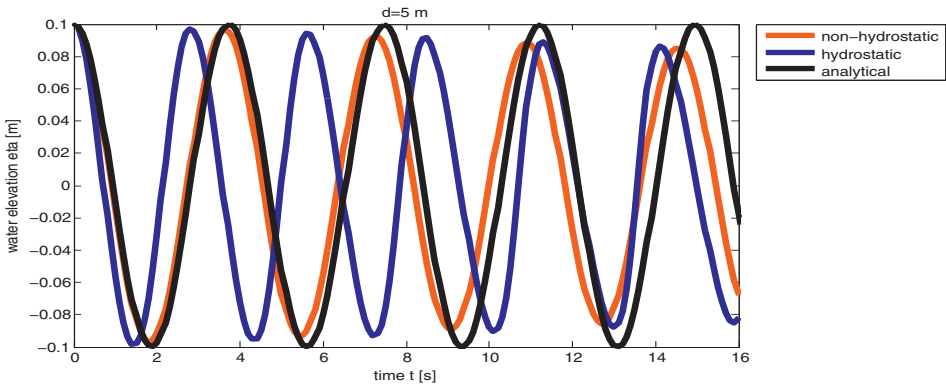


Fig. 3: Standing wave: plot of the water elevation at $(x, y) = (10\text{ m}, 5\text{ m})$ over time. 32768 cells were used for the uniform grid.

Notice the numerical damping for both the hydrostatic and the non-hydrostatic model. This is a common phenomenon especially for low-order discretizations like the one we employ here.

Solitary Wave

In the next test case, the propagation of a solitary wave in a closed channel of width 2m and depth $d = 10\text{ m}$ is examined—another standard test case for non-hydrostatic models (cf. e.g. [Cu13], [SZ03], [Wa05] and [WJ00]). Since viscosity and friction are absent in our model, the solitary wave should not deform while propagating. As depicted in Fig. 4, the non-hydrostatic solution keeps the correct water level over time with some small trailing waves. These trailing waves have been observed in other non-hydrostatic models such as [SZ03]. They are likely due to the spatial discretization's inherent failure of resolving large eigenmodes of the problem correctly. Fig. 5 shows that the hydrostatic model produces an overall wrong sawtooth shaped water surface profile.

4.2 Solver

Even if the physical validation experiments were successfully performed by using the Jacobi method, we intended to apply conjugate gradient (CG) to benefit from faster con-

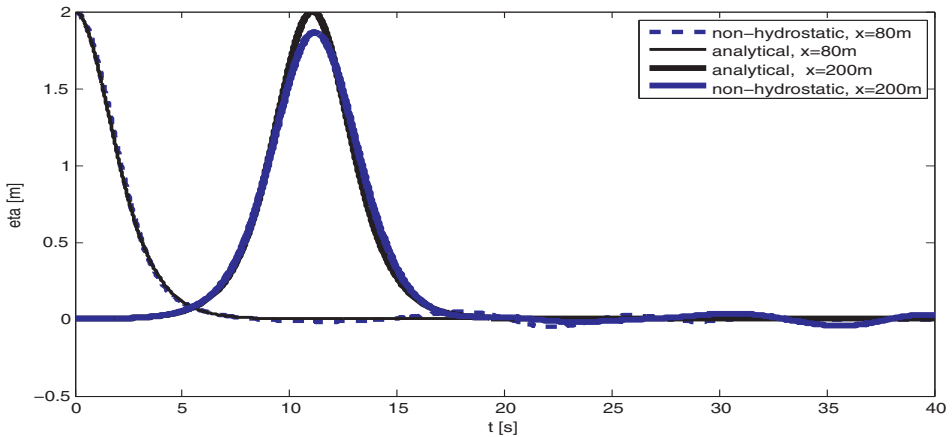


Fig. 4: Solitary wave (#cells = 38400): water elevation over time at the points $(x_1, y_1) = (80\text{m}, 0.25\text{m})$ and $(x_2, y_2) = (200\text{m}, 0.25\text{m})$. For the sake of clarity, the hydrostatic solution is omitted in this plot.

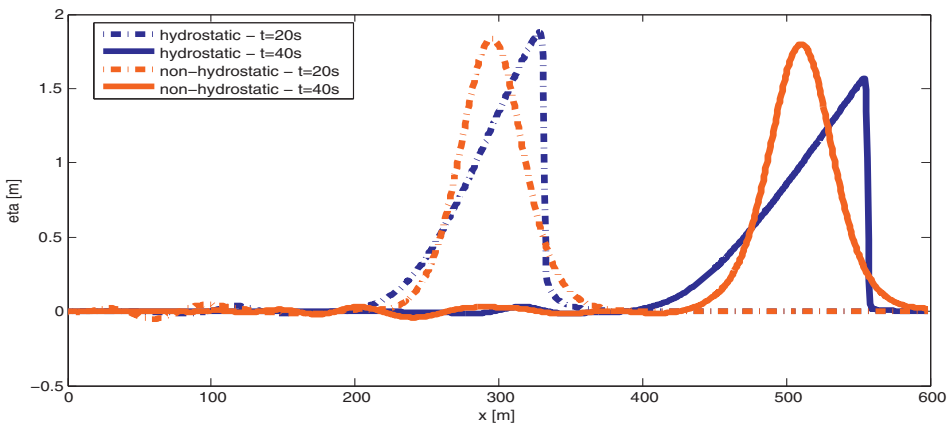


Fig. 5: Solitary wave (#cells = 38400): water level at two different points in time.

vergence. However, CG is only defined for symmetric positive definite system matrices [HS52]. Although we expected the matrix to be symmetric positive definite for constant bathymetry, we conducted tests by simulating several scenarios and checking the properties of the global matrix at every time step. We only found symmetric positive definite system matrices during these tests. Certainly, our tests did not cover all possible scenarios, so even if the matrix was symmetric positive definite in our case, these results do not prove that the matrix will also be symmetric positive definite for other scenarios.

The blue line in Fig. 6 displays the average number of CG iterations for the first 0.05s of simulation time of the standing wave scenario for different grid resolutions. Starting with 58 iterations for 32k cells, we ended at 812 iterations for 256M cells. The line is neatly fitted by the function

$$y = 1.2712x^{0.3258}, \quad (7)$$

where x is the number of cells and y the number of iterations. It is depicted in Fig. 6 as a dashed line. Thus, for this scenario, the number of iterations increases by a factor of about $2^{0.3258} = 1.2534$ if the number of cells is doubled. So even with perfect weak scaling (see Sec. 4.3), one has to expect at least an 25% increase in simulation run time each time the number of cells is doubled.

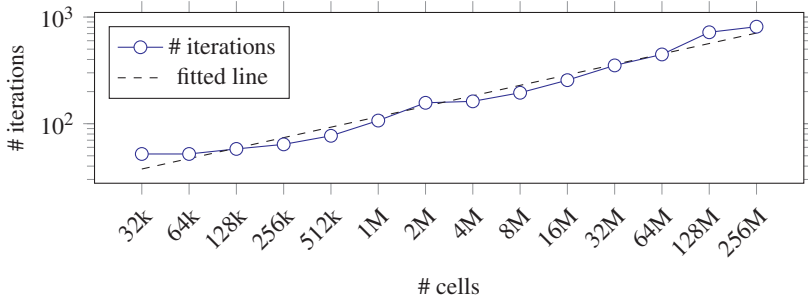


Fig. 6: Average number of CG iterations for the first 0.05s seconds of simulation time of the standing wave scenario for different grid resolutions. The dashed line shows the fitted function as denoted in Eq. 7.

4.3 Scaling

All measurements were performed on the SuperMUC⁴ Thin Nodes, a supercomputer operated by the Leibniz Supercomputing Center near Munich. Each node is equipped with two eight-core Xeon E5-2680 processors. One island incorporates 512 nodes, resulting in 8192 cores per island. The nodes are connected via Infiniband FDR10. We used the standing wave scenario and always placed one MPI process and one section on each physical core with section splitting⁵ turned on. We employed *PipeCG* as solver which is a specialized version of CG. While being equivalent to CG up to rounding errors, it reduces the global communication (see [GV14] for more information).

We measured performance in *element throughput*, which is defined as the number of elements which are handled per second per core. Thus, since this is a measure of parallel efficiency, a constant element throughput indicates perfect scaling.

Weak Scaling

The results of the weak scaling measurements are displayed in Fig. 7. Each measurement ran for 0.05 s of simulation time. The scaling was conducted for one up to 8192 cores with

⁴ <http://www.lrz.de/services/compute/supermuc/systemdescription/>

⁵ Normally, sam(oa)² treats sections as atomic work units (see Sec. 3). For load balancing, however, sections can be split and partially transferred to other processes.

a grid resolution of 32k cells up to 256M cells. Furthermore, a run on 16384 cores and a grid resolution of 512M cells was performed, but for 0.0002 s of simulation time due to limited CPU time budget. We achieved a parallel efficiency of 90.1% for 8192 cores.

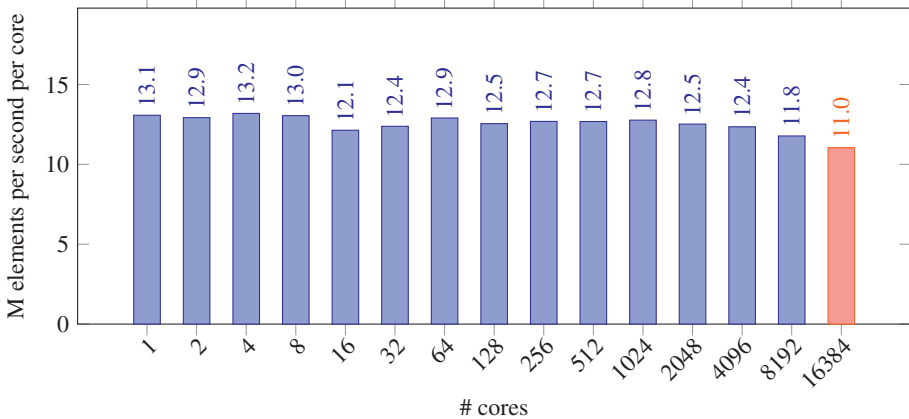


Fig. 7: Weak scaling of the standing wave scenario on 1 up to 16384 cores. While the blue bars ran for 0.05 s of simulation time, the red bar ran for 0.0002 s.

Strong Scaling

The strong scaling results are shown in Fig. 8. We performed measurements on 16 up to 512 cores, each for 100 time steps and with a grid resolution of 2M, 4M and 8M cells. On 512 cores with a grid resolution of 8M cells, we reached a parallel efficiency of 96.8%. On 128 cores with the same grid resolution, we even achieved super-linear speedup (101.0%). However, due to increased overhead, the lower the grid resolution is, the worse is the parallel efficiency. This effect appears for a grid resolution of 2M cells: on 512 cores, a parallel efficiency of only 46.8% is reached, since the number of cells per core (here 2048) is too low.

5 Conclusion & Outlook

Starting with a finite-volume conservation-law formulation of the non-hydrostatic shallow water equations, we have shown how the equation for the non-hydrostatic correction parameter can be discretized on triangular meshes. This allows for adaptivity and a matrix-free approach towards solving the corresponding system. This was attained by using a dual grid which resolved the drawbacks of a collocated discretization. Our results obtained with common test cases demonstrate the success of our numerical model to more accurately resolve the dispersion relation of a standing and a solitary wave compared to the previously existing hydrostatic solver. We furthermore have shown feasibility of conjugate gradient for our test cases. Our parallelization strategy yielded solid strong and weak scaling results which allows to solve large problems on HPC systems.

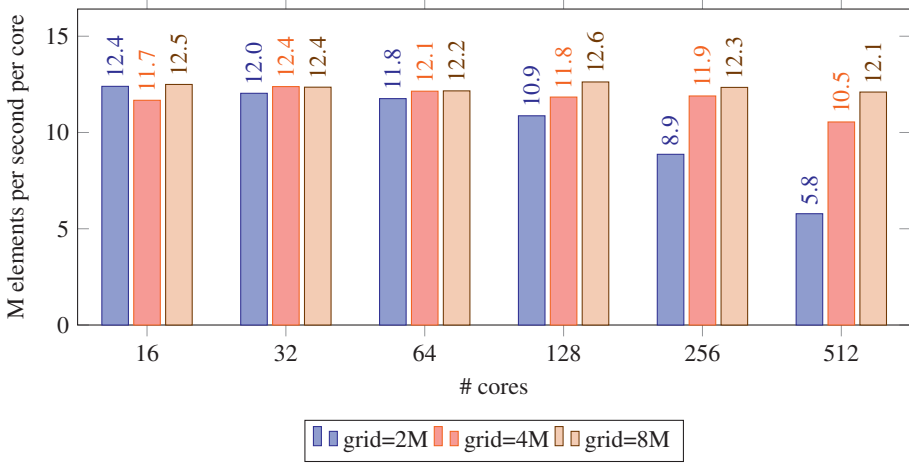


Fig. 8: Strong scaling of the standing wave scenario for 16 up to 512 cores for a grid resolution of 2M, 4M and 8M cells.

Future work will have to study the non-hydrostatic model's behavior on more complex and realistic test cases. Although the terms that had to be neglected in the continuity equation did not seem to adversely affect the results we obtained in the validation test cases, including them into the model certainly leaves room for improvement. Moreover, further research is required for applying faster solvers.

Acknowledgments

We would like to particularly thank Michael Bader, Oliver Meister and Kaveh Rahnema for giving very useful hints and directions on this work. Special thanks also go to Kyle Mandli (Columbia University) who provided an internal report that has been very helpful for deriving the non-hydrostatic shallow water equations in conservation-law form.

References

- [Ch68] Chorin, Alexandre J.: Numerical Solution of the Navier-Stokes Equations. *Mathematics of Computation*, 22(104):745–762, 1968.
- [CS98] Casulli, Vincenzo; Stelling, Guus S.: Numerical Simulation of 3D Quasi-Hydrostatic, Free-Surface Flows. *Journal of Hydraulic Engineering*, 124(7):678–686, 1998.
- [Cu13] Cui, Haiyang: A New Numerical Model for Simulating the Propagation of and Inundation by Tsunami Waves. Dissertation, Delft University of Technology, 2013.
- [Fu13] Fuchs, Annika: Effiziente parallele Verfahren zur Lösung verteilter, dünnbesetzter Gleichungssysteme eines nichthydrostatischen Tsunamimodells. PhD thesis, Staats- und Universitätsbibliothek Bremen, 2013.
- [GDN95] Griebel, Michael; Dornseifer, Thomas; Neunhoffer, Tilman: *Numerical Simulation in Fluid Dynamics*. Vieweg, 1995.
- [GV14] Ghysels, Peter; Vanroose, Wim: Hiding global synchronization latency in the preconditioned Conjugate Gradient algorithm. *Parallel Computing*, 40(7):224 – 238, 2014.
- [HS52] Hestenes, Magnus R.; Stiefel, Eduard: *Methods of Conjugate Gradients for Solving Linear Systems*. *Journal of Research of the National Bureau of Standards*, 49(6):409–436, 1952.
- [Le02] LeVeque, Randall J.: *Finite Volume Methods for Hyperbolic Problems*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2002.
- [LN08] Levin, Boris; Nosov, Mikhail: *Physics of Tsunamis*. Springer Science & Business Media, 2008.
- [Ma] Mandli, Kyle T.: Derivation of Depth Averaged Flow. Private report.
- [MB15] Meister, Oliver; Bader, Michael: 2D Adaptivity for 3D Problems: Parallel SPE10 Reservoir Simulation on Dynamically Adaptive Prism Grids. *Journal of Computational Science*, 9:101–106, May 2015.
- [MRB11] Meister, Oliver; Rahnama, Kaveh; Bader, Michael: A Software Concept for Cache-Efficient Simulation on Dynamically Adaptive Structured Triangular Grids. In: PARCO. pp. 251–260, 2011.
- [Sa14] Samfass, Philipp J.: Extension of the Finite Volume Solver SWE towards the Non-Hydrostatic Shallow Water Equations. Bachelor's thesis, Institut für Informatik, Technische Universität München, September 2014.
- [SZ03] Stelling, Guus; Zijlema, Marcel: An accurate and efficient finite-difference algorithm for non-hydrostatic free-surface flow with application to wave propagation. *International Journal for Numerical Methods in Fluids*, 43(1):1–23, 2003.
- [Wa05] Walters, Roy A.: A semi-implicit finite element model for non-hydrostatic (dispersive) surface waves. *International Journal for Numerical Methods in Fluids*, 49(7):721–737, 2005.
- [WJ00] Weilbeer, H; Jankowski, JA: A Three-Dimensional Non-Hydrostatic Model for Free Surface Flows – Development, Verification and Limitations. In: Proceedings of the 6th International Conference on Estuarine and Coastal Modeling. pp. 162–177, 2000.

Überführung von arithmetischen Ausdrücken in ein normalisiertes Polynom mittels Baumtransformation

Felix Knispel ¹

Abstract: In dieser Arbeit soll eine Technik vorgestellt werden, arithmetische Terme in eine normalisierte Polynomialform mittels Baumtransformationen zu überführen. Die Manipulation arithmetischer Terme ist eine Grundfunktion heutiger Computeralgebrasysteme. Diese verwenden dabei verschiedenste Ansätze. Ziel dieser Arbeit ist die Entwicklung eines Ansatzes zur Termumformung unter Verwendung von Übersetzerbautechnologien. Eine solche Technologie sind *bottom-up rewrite systems*, die die technologische Grundlage der Baumtransformation in dieser Arbeit darstellen. Es wird das Werkzeug `poly` vorgestellt, welches eine Technik implementiert, die Baumtransformation iterativ und phasenweise durchzuführen, um eine Aufteilung der Problematik in Teilprobleme zu ermöglichen.

Keywords: Baumtransformation, Termersetzung, Normalformen, Übersetzerbau

1 Problemstellung

In modernen Computeralgebrasystemen, wie *MATLAB* [Ma16a] und *Maxima* [Ma16b], ist die Transformation von arithmetischen Ausdrücken eine bedeutsame Grundfunktion. Es gibt eine Vielzahl von Techniken zur symbolischen Verarbeitung von arithmetischen Termen, wie beispielsweise die funktionale Programmierung oder Listen-Datenstrukturen. Insbesondere die Verwendung von Listen ist im Bereich der Computeralgebrasysteme ein häufiges Mittel [Fa01]. Die Wahl der zu verwendenden Technik ist anwendungsabhängig. Im Projekt „Entwicklung hocheffizienter Pumpensysteme“ [Pr16], in dessen Rahmen diese Arbeit entstand, werden unter Verwendung von Übersetzerbautechnologien aus einer entworfenen domänenspezifischen Sprache Beschreibungen von MILP-Problemen (engl. *mixed integer linear programming*) generiert. MILP-Problemlöser, wie z. B. *SCIP* [Ac04], geben dabei eine Syntax vor, in der das Problem formuliert werden muss. So wird im Falle von *SCIP* auch verlangt, dass Nebenbedingungen in einer Polynomialform aufgestellt werden müssen. Da die Integrierung von Computeralgebrasystem und deren Ergebnisse in Übersetzerbautechnologien sich als umständlich erwies, ergab sich in diesem Zusammenhang die Problemstellung arithmetische Ausdrücke in eine normalisierte Polynomialform unter der Anforderung der Verwendung von Übersetzerbautechnologien zu überführen.

Im Feld des Übersetzerbaus gibt es im Bereich der Codeerzeugung verschiedene Technologien zur Generierung von optimalem Code. Eine solche Technologie ist die BURS-Theorie (engl. *bottom-up rewrite systems*) von Pelegrí-Llopart und Graham [PLG88]. BURS bietet durch auf dynamischer Programmierung basierten Kostenanalyse eine Möglichkeit

¹ Martin-Luther-Universität Halle-Wittenberg, Institut für Informatik, Von-Seckendorff-Platz 1, 06120 Halle

durch Transformation von Ableitungsbäumen (kosten-)optimalem Code in einer Zielsprache zu generieren. Die Technologie wird in dieser Arbeit genutzt, um arithmetische Ausdrücke, die als Ableitungsbäume dargestellt sind, in eine normalisierte Polynomform zu überführen, indem diese Bäume transformiert werden. Emmelmann [Em94] konnte bereits zeigen, dass Baumtransformation zur Manipulation arithmetischer Terme genutzt werden kann.

Das in dieser Arbeit entwickelte Werkzeug `poly` bekommt einen arithmetischen Ausdruck als Eingabe, überführt diesen Ausdruck in ein normalisiertes Polynom und gibt dieses aus. Mit einer gegebenen Bezeichnermenge $V = \{v_1, \dots, v_k\}$ sei die Normalform eines gültigen Ergebnispolynoms P gegeben durch

$$P = \sum_{i=0}^n \sum_{\substack{j_1+\dots+j_k=i \\ 0 \leq j_1, \dots, j_k \leq n}} c_{j_1, \dots, j_k} \cdot v_1^{j_1} \cdot \dots \cdot v_k^{j_k}, \quad c_{j_1, \dots, j_k} \in \mathbb{R}. \quad (1)$$

Das Problem der schrittweisen Überführung in ein normalisiertes Polynom wird dabei in drei Teilprobleme unterteilt. Im ersten Abschnitt sollen alle Klammern im Ausdruck durch Ausmultiplizieren aufgelöst werden. Das Ergebnis ist für die meisten MILP-Problemlöser bereits ein gültiges Polynom. Dieses entspricht in der Regel jedoch nicht obiger Normalform, bestehend aus Monomen der Form $c_{j_1, \dots, j_k} \cdot v_1^{j_1} \cdot \dots \cdot v_k^{j_k}$, da das Ausmultiplizieren nicht zwangsläufig eine feste Bezeichnerordnung innerhalb der Monome gewährleistet und gegebenenfalls noch Konstanten zusammengefasst werden müssen. Die weitere Überführung des Polynoms in die Normalform erfüllt den Zweck, das nach dem Ausklammern resultierende Polynom weiter zu verkürzen, mit dem Ziel die Laufzeit des MILP-Problemlösers durch Linearisierung zu verbessern. Im folgenden Schritt sollen innerhalb der Monome gleiche Faktoren zu Potenzen zusammengefasst werden, sodass sich innerhalb der Monome keine Bezeichner wiederholen. Der letzte Schritt hat das Ziel das Polynom in ein gültiges Ergebnispolynom gemäß obiger Normalform zu überführen. Hierfür müssen Monome, die dieselben Bezeichner mit jeweils den gleichen Exponenten aufweisen, zusammengefasst werden.

Technische Grundlage für die Implementierung des Werkzeugs ist das Übersetzerbauwerkzeugkasten *Eli* [Gr92, El16]. Diese Entwicklungsumgebung stellt unter anderem ein Baumersetzungswerkzeug zur Verfügung.

2 Grundlagen

2.1 Bottom-up rewrite systems

Bottom-up rewrite systems (BURS) werden im Bereich der optimalen Codegenerierung verwendet. Ausgehend von einem als Ableitungsbaum dargestellten Programm, wird in diesem Baum nach Mustern gesucht, um optimal Code in einer Zielsprache zu generieren. Zu optimierende Eigenschaften können beispielsweise die Anzahl der Anweisungen oder die Anzahl der Prozessorinstruktionen sein. BURS bietet dabei die Möglichkeit Ableitungsbäume zu *transformieren*, d. h. Teilbäume durch andere Bäume zu ersetzen. In dieser

Arbeit wird diese Technik genutzt, um arithmetische Terme, dargestellt in einer Baumstruktur, durch Transformation ihrer Ableitungsbäume umzustellen.

Bäume werden in Präfixnotation oder als Graph, wie im Beispiel unten, dargestellt. So stellt beispielsweise $o(t_1, t_2)$ einen Baum dar, der in der Wurzel mit einem Symbol o beschriftet ist, deren Kinder zwei Unterbäume t_1 und t_2 sind. Die Knotenbeschriftungen werden einem *Operatoralphabet* Σ entnommen. Symbole weisen eine Stelligkeit $n \in \mathbb{N}$ auf, die der Anzahl der Kinder entspricht. *Baummuster* sind Bäume über diesem Alphabet, erweitert um Symbole der Stelligkeit 0, die *Variablen* genannt werden. Variablen einer Variablenmenge Var sind Statthalter für gültige Bäume aus dem Operatoralphabet. In einem Baummuster dürfen ausschließlich Variablen als Blätter vorkommen. $\sigma(m)$ beschreibt die Zuweisung von Bäumen an Variablen eines Baummusters m . Ein Baummuster m passt auf einen Baum t , wenn es eine Variablenzuweisung σ gibt, sodass $\sigma(m) = t$. *Baumersetzungs-* oder *Baumtransformationsregeln* überführen Baummuster und haben die Form $\alpha \rightarrow \beta$, wobei alle Variablen aus β in α vorkommen müssen. Ein *Baumersetzungs-system* ist eine Menge von Baumersetzungsregeln. Die *Position* eines Knotens in einem Baum wird als Folge von natürlichen Zahlen dargestellt. Die Position der Wurzel eines Baumes t wird durch die leere Folge ε angegeben. Der Unterbaum von t an einer Position l wird durch $t_{@l}$ notiert. Jede natürliche Zahl in einer Folge beschreibt den Index eines Kindknotens von links nach rechts, beginnend bei 1. Ist ks eine Folge bestehend aus $k \in \mathbb{N}$ und einer Folge s und $=$ ist zu verstehen als „ist definiert als“, so hängen Positionen und Unterbäume folgendermaßen zusammen:

$$t_{@\varepsilon} = t$$

$$\text{op}(t_1, \dots, t_n)_{@ks} = (t_k)_{@s}, 1 \leq k \leq n, \text{op} \in \Sigma_n.$$

Eine Baumersetzungsregel $r : \alpha \rightarrow \beta$ ist auf einen Baum t an der Position p *anwendbar*, wenn α auf $t_{@p}$ passt. Die *Anwendung* von r auf t ergibt einen neuen Baum, der t gleicht, nur dass der Unterbaum $t_{@p}$ durch $\sigma(\beta)$ ersetzt wird.

Beispiel:

Gegeben sei ein Operatoralphabet $\Sigma = \Sigma_0 \cup \Sigma_2$. Es seien $+, * \in \Sigma_2$ zweistellige Symbole für die entsprechenden arithmetischen Operationen. $a, 0, 5 \in \Sigma$ sind nullstellige Symbole und damit gültige Blätter im Baum. Der Baum $t = *(+(5, 0), a)$ ist somit ein zulässiger Baum und soll den Term $(5 + 0) \cdot a$ darstellen.



Das Baummuster $m = *(+(X, Y), Z)$ passt auf t mit $\sigma = \{X = 5, Y = 0, Z = a\}$ für Variablen $X, Y, Z \in Var$ mit $Var \subset \Sigma_0$. Die Baumersetzungsregel $r : +(X, Y) \rightarrow +(Y, X)$ ist auf t anwendbar, da das Baummuster der linken Seite auf $t_{@1}$ passt. Die Anwendung von r auf t ergibt $t = *(+(0, 5), a)$.

2.2 Baumtransformationsregeln

In dem folgenden Kapitel werden die verwendeten Baumtransformationsregeln aufgezeigt. Auf der linken Seite einer Regel steht das zu ersetzende Baummuster. Auf der rechten Seite steht das Baummuster, welches das Muster auf der linken Seite ersetzt. Die Muster bestehen aus den Symbolen Add (Addition), Sub (Subtraktion), Mul (Multiplikation), Div (Division), Pot (Potenzierung) und Neg (Vorzeichen-Minus). Die Klammerung ist implizit durch den Aufbau des Baumes gegeben. Nichtterminale beginnen mit einem c bei Konstanten, einem v bei Bezeichnern und einem t bei arithmetischen Termen. In den Transformationsregeln sind t_i somit Variablen im Sinne von BURS. Konstanten und Bezeichner sind gültige Terme und können somit auch auf Terme passen, aber nicht umgekehrt. Abbildung 1 zeigt eine Baumtransformationsregel für das Distributivgesetz und den zu ersetzenden sowie den resultierenden Unterbaum.

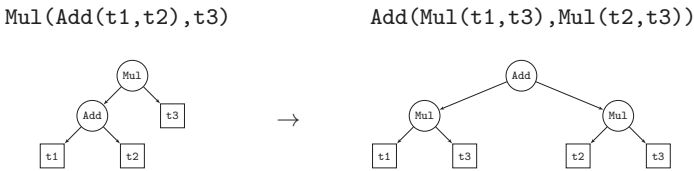


Abb. 1: Beispiel einer Baumtransformationsregel mit entsprechendem zu ersetzenden und resultierenden Baum

Die Operatoren „+“, „*“ und „/“ in Baumtransformationsregeln ergeben konstante Ergebnisse. Beispielsweise ergibt c_1+c_2 das Ergebnis der Addition zweier Konstanten c_1 und c_2 als Konstante.

3 poly

3.1 Eingabeausdrücke

poly erwartet, dass der einzugebende arithmetische Term sich an eine bestimmte Form hält. Diese Form wird durch folgende kontextfreie Grammatik festgelegt. Kleingeschriebene Worte sind Nichtterminale. Operatoren sind Terminale und werden in einfachen Anführungszeichen angegeben. ID und CONST sind Terminale, die Bezeichner bzw. Konstanten repräsentieren. „|“ grenzt Auswahlmöglichkeiten voneinander ab.

```

expr      : sum .
sum       : sum '+' term |
          : sum '-' term |
          : term .
term      : term '*' factor |
          : term '/' constant |
          : factor .
    
```

```

factor   : ident '^' CONST |
          constant '^' CONST |
          primary .
primary  : ID | CONST | '-' primary | '(' sum ')' .
ident    : ID | '-' ident .
constant: CONST | '-' constant .

```

Es darf nur durch Konstanten dividiert werden. Potenzen dürfen entweder eine Konstante oder einen Bezeichner zur Basis haben. Als Exponenten sind ausschließlich natürliche Konstanten zulässig. Erlaubt sind somit folgende Operatoren mit der angegebenen Priorität.

	Klammerung ()		
>	Vorzeichen-Minus -	(Symbol: Neg)	
>	Potenzierung ^	(Symbol: Pot)	
>	Multiplikation *	(Symbol: Mul) =	Division / (Symbol: Div)
>	Addition +	(Symbol: Add) =	Subtraktion - (Symbol: Sub)

3.2 Iterative Baumtransformation in Phasen

Für die Implementierung von `poly` wurde der Übersetzerbauwerkzeugkasten *Eli* verwendet. *Eli* stellt unter anderem ein Werkzeug zur Baumtransformation zur Verfügung. Nach der Angabe von Baumersetzungsregeln kann man diese auf Bäume einer internen Datenstruktur anwenden. Im Kontext der Transformation von arithmetischen Termen kann es aber vorkommen, dass die Anwendung einer Regel Muster im Baum hervorbringen kann, auf die wiederum Regeln anwendbar sind. Aus diesem Grund ist es nötig die Transformation der Terme in Iterationen ablaufen zu lassen, bis ein Fixpunkt erreicht ist, in dem der Term sich nicht mehr ändert. Folgender Algorithmus zeigt dieses Vorgehen.

```

repeat
  vergleich = term
  term = Baumtransformation(term)
until vergleich = term
return term

```

Die Grundidee zur Lösung des Problems der Überführung von arithmetischen Ausdrücken in ein normalisiertes Polynom ist die Unterteilung in Teilprobleme, die nacheinander gelöst werden sollen. Es ist also nötig verschiedene Baumersetzungssysteme zu implementieren und getrennt von einander in Phasen ablaufen zu lassen. In *Eli* kann man jedoch nur ein Baumersetzungssystem angeben. Dabei kann jedoch das Problem auftreten, dass mehrere Baumtransmutationsregeln verschiedener Phasen um ein Muster im Baum konkurrieren. Somit musste eine Methode entwickelt werden die Baumtransformation in aufeinanderfolgenden Phasen unter Angabe eines Baumersetzungssystems ablaufen zu lassen. Damit Muster verschiedener Transformationsphasen nicht um dieselben Teilbäume konkurrieren,

werden zusätzliche Operatorsymbole eingeführt. Es werden die Symbole der entsprechenden Operationen mit einer Ziffer der zugehörigen Transformationsphase versehen. Für die Addition ergeben sich somit die Operationssymbole Add1 für die erste Transformationsphase, Add2 für die zweite Transformationsphase und Add3 für die dritte Transformationsphase. Obiger Algorithmus muss für alle drei Transformationsphasen nacheinander angewandt werden. Die Knoten des Baumes werden zunächst mit den Operatorsymbolen der ersten Transformationsphase versehen und das Baumersetzungssystem bis zum Erreichen eines Fixpunktes angewandt. Ist der Fixpunkt erreicht, wird der resultierende Baum mit den Operatorsymbolen der zweiten Transformationsphase versehen und das Verfahren wird erneut angewandt. Selbiges schließlich auch im dritten Transformationsschritt.

3.3 Überführung in ein normalisiertes Polynom

Der Ablauf der kompletten Umformung in ein Polynom wird in drei Schritte unterteilt. Jeder dieser Schritte hat eine bestimmte Polynomialform als Ergebnis, welche für den folgenden Schritt als Ausgang dient. Diese Unterteilung hat zudem den Effekt, dass das Verfahren übersichtlich geordnet abläuft und die Auswahl der anzuwendenden Transformationsregeln in jedem Schritt eingeschränkt wird. Abbildung 2 stellt das strategische Vorgehen beim Aufruf einer Transformation dar.

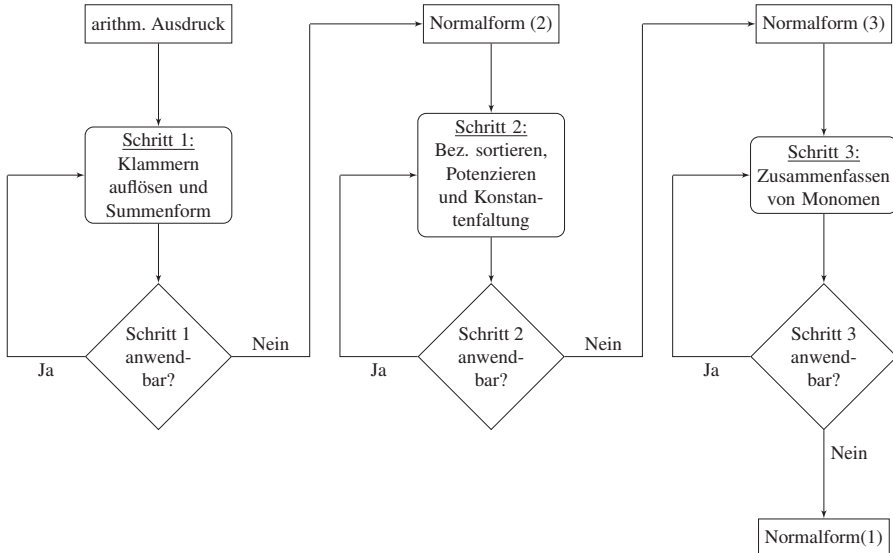


Abb. 2: Schematische Darstellung der Überführung von arithmetischen Ausdrücken in ein normalisiertes Polynom

Klammern auflösen und Überführung in Summenform

Im ersten Schritt soll der eingegebene arithmetische Term in eine klammerfreie Summe von Monomen überführt werden. Das Ergebnis dieses Schrittes soll mit einer Bezeichner-

menge V folgender Polynomialform genügen.

$$\sum v_1 \cdot \dots \cdot v_k, k \in \mathbb{N}, v_1, \dots, v_k \in V \cup \mathbb{R} \tag{2}$$

Die Monome des Ergebnispolynoms dieses Schritts sind ein beliebiges Produkt von Konstanten und Bezeichnern. Konstanten und Bezeichner können in einem Monom mehrfach vorkommen. Abbildung 3 listet alle hierfür nötigen Transformationsregeln auf.

(A) Assoziativgesetz		
$\text{Add1}(t_1, \text{Add1}(t_2, t_3))$	\rightarrow	$\text{Add1}(\text{Add1}(t_1, t_2), t_3)$ (1)
(B) Distributivgesetz		
$\text{Mul1}(\text{Add1}(t_1, t_2), t_3)$	\rightarrow	$\text{Add1}(\text{Mul1}(t_1, t_3), \text{Mul1}(t_2, t_3))$ (2)
$\text{Mul1}(t_1, \text{Add1}(t_2, t_3))$	\rightarrow	$\text{Add1}(\text{Mul1}(t_1, t_2), \text{Mul1}(t_1, t_3))$ (3)
(C) Eliminierung von Subtraktion und Division		
$\text{Div1}(c_1, c_2)$	\rightarrow	c_1/c_2 (4)
$\text{Div1}(t, c)$	\rightarrow	$\text{Mul1}(1/c, t)$ (5)
$\text{Neg1}(t_1)$	\rightarrow	$\text{Mul1}(-1, t_1)$ (6)
$\text{Sub1}(t_1, t_2)$	\rightarrow	$\text{Add1}(t_1, \text{Mul1}(-1, t_2))$ (7)

Abb. 3: Baumtransformationsregeln, um einen arithmetischen Ausdruck in einen Ausdruck ohne Klammerung und Subtraktion zu überführen.

Das Distributivgesetz findet Anwendung, um alle Summen innerhalb von Multiplikationen auszuklammern. Um die in der Normalform für Ergebnispolynome geforderte Summenform zu gewährleisten, werden Subtraktionen eliminiert, indem sie in eine Addition und eine Multiplikation des zweiten Operanden mit der Konstante -1 umgeformt werden. Die Anwendung des Distributivgesetzes und der Eliminierung der Subtraktion ermöglichen zusammen eine resultierende Summenform nach diesem Schritt. Vorbereitend, um den dritten Schritt zu erleichtern, wird das Assoziativgesetz für die Addition angewandt, um den Baum der Summe nach links abzuleiten.

Sortierung der Bezeichner, Potenzieren und Konstantenfaltung

Der erste Transformationsschritt stellt die klammerfreie Summenform der Ergebnispolynome her. Der folgende zweite Schritt überführt die Monome innerhalb der Summe in die gewünschte Form. Das Ergebnis dieses Schrittes genügt mit einer Bezeichnermenge V folgender Polynomialform.

$$\sum_{i=0}^n \sum_{\substack{j_1 + \dots + j_k = i \\ 0 \leq j_1, \dots, j_k \leq n}} c \cdot v_1^{j_1} \cdot \dots \cdot v_k^{j_k}, c \in \mathbb{R}, v_1, \dots, v_k \in V \tag{3}$$

Um die Transformation zu erleichtern, werden die Monome nach links abgeleitet. Dies verringert die benötigten Transformationsregeln. Hierfür ist eine Realisierung des Assoziativgesetzes für die Multiplikation nötig. Eine Anforderung der Normalform für Ergebnispolynome ist, dass höchstens eine Konstante in einem Monom vorkommt, die zudem im

Monom ganz links stehen soll. Dies wird in diesem Schritt realisiert, indem Konstanten im nach links abgeleiteten Unterbaum des Terms nach unten gereicht werden, bis diese im unteren linken Blatt des Unterbaums stehen. Kommen mehrere Konstanten in einem Monom vor, so werden sie solange nach unten gereicht, bis zwei Konstanten Kinder des untersten Multiplikationsknotens sind. Dieses Muster wird durch einen einzelnen Konstantenknoten mit dem Ergebnis der Multiplikation beider Konstanten ersetzt.

(A) Assoziativgesetz		
$Mul2(t1, Mul2(t2, t3))$	\rightarrow	$Mul2(Mul2(t1, t2), t3)$ (1)
(B) Konstantenfaltung		
$Mul2(c1, c2)$	\rightarrow	$c1*c2$ (2)
$Mul2(t, c)$	\rightarrow	$Mul2(c, t)$ (3)
(C) Bezeichner sortieren und Potenzieren		
$Mul2(c, v)$	\rightarrow	$Mul2(c, Pot2(v, 1))$ (4)
$Mul2(v1, v2)$	\rightarrow	wenn $v1 = v2$: $Pot2(v1, 2)$ wenn $v1 < v2$: $Mul2(Pot2(v1, 1), Pot2(v2, 1))$ wenn $v2 < v1$: $Mul2(Pot2(v2, 1), Pot2(v1, 1))$ (5)
$Mul2(Pot2(v1, c), v2)$	\rightarrow	wenn $v1 = v2$: $Pot2(v1, c+1)$ wenn $v1 < v2$: $Mul2(Pot2(v1, c), Pot2(v2, 1))$ wenn $v2 < v1$: $Mul2(Pot2(v2, 1), Pot2(v1, c))$ (6)
$Mul2(Mul2(t, Pot2(v1, c)), v2)$	\rightarrow	wenn $v1 = v2$: $Mul2(t, Pot2(v1, c+1))$ wenn $v1 < v2$: $Mul2(Mul2(t, Pot(v1, c)), Pot2(v2, 1))$ wenn $v2 < v1$: $Mul2(Mul2(t, v2), Pot2(v1, c))$ (7)

Abb. 4: Baumtransformationsregeln um arithmetische Ausdrücke weiter zu vereinfachen. Es werden Konstanten gefaltet sowie Bezeichner nach einer Bezeichnerordnung sortiert und potenziert. Die Vergleichsoperatoren = und < beziehen sich auf eine Bezeichnerordnung.

Weiterhin verlangt die Normalform, dass innerhalb der Monome gleiche Bezeichner zu Potenzen zusammengefasst und die Bezeichner nach einer Bezeichnerordnung sortiert sind. Die in Abbildung 4 aufgelisteten Baumtransformationsregeln realisieren die Potenzierung von Bezeichnern mittels einer zeitgleich stattfindenden Sortierung eben dieser. Damit das folgende Verfahren funktioniert, ist gegebenenfalls eine Vorverarbeitung des Termes nötig. So werden beim Einlesen der Eingabe alle Potenzen in Multiplikationen umgeformt. Ausgangspunkt für die Sortierung und Potenzierung ist nun ein nach links abgeleiteter Unterbaum, in dem gegebenenfalls im untersten linken Blatt eine Konstante steht. Bezeichner, die nun Kinder der untersten Multiplikation sind, werden durch Potenzen mit dem Exponenten 1 ersetzt. Potenzen befinden sich während dieses Verfahrens prinzipiell stets in den unteren Blättern des Baumes. Anschließend wird der Baum nach dem Muster $Mul(Mul(Term, Potenz), Bezeichner)$ abgesucht. Ein Bezeichner wird nun entsprechend der Bezeichnerordnung im Baum nach unten gereicht, wenn der Bezeichner entsprechend der Ordnung kleiner als der Bezeichner der benachbarten Potenz

ist. Sobald dies nicht mehr möglich ist, wird der Bezeichner in eine Potenz umgeformt. Hier sind zwei Fälle zu beachten. (1) Der linke Nachbar ist eine Potenz, dessen Bezeichner gemäß der Ordnung kleiner ist. In diesem Falle wird der Bezeichner in eine Potenz mit dem Exponenten 1 umgeformt. (2) Der linke Nachbar ist eine Potenz, dessen Bezeichner der gleiche ist. In diesem Falle werden beide zu einer Potenz zusammengefasst, indem der Exponent des linken Nachbarn um eins erhöht wird. Das Verfahren arbeitet sich so Iteration für Iteration im Baum nach oben, solange bis alle Bezeichner in eine Potenz aufgenommen wurden. Am Ende dieses Schrittes bleibt die aus dem ersten Schritt resultierende Summenform erhalten. In diesem Schritt wird lediglich die Struktur der Monome verändert.

Zusammenfassen von Monomen

Im dritten und letzten Schritt der Transformation soll die Normalform für Ergebnispolynome

$$P = \sum_{i=0}^n \sum_{\substack{j_1+\dots+j_k=i \\ 0 \leq j_1, \dots, j_k \leq n}} c_{j_1, \dots, j_k} \cdot v_1^{j_1} \cdot \dots \cdot v_k^{j_k}, \quad c_{j_1, \dots, j_k} \in \mathbb{R}, \quad n \in \mathbb{N} \quad (1)$$

hergestellt werden. Ausgehend vom Ergebnis des zweiten Schritts ist es nun noch nötig Monome mit gleichen Bezeichnern zusammenzufassen, d. h. zwei Monome im Polynom weisen dieselben Bezeichner mit den gleichen Exponenten auf, so müssen sie zu einem Monom mit gleichen Bezeichnern zusammengefasst und die vorkommenden Konstanten aufsummiert werden. Grundlegende Idee dies zu realisieren ist es, im Baum nach benachbarten Monomen zu suchen, die die gleiche Bezeichnermenge aufweisen. Es ist also zunächst nötig, die Reihenfolge der Monome im Baum zu verändern. Hierfür wird die Vorgehensweise des zweiten Schritts, um Bezeichner in Monomen zu sortieren, aufgegriffen und auf die Sortierung von Monomen angepasst. Im zweiten Schritt wurden Bezeichner im Baum des entsprechenden Monoms entsprechend einer Bezeichnerordnung nach unten gereicht. Befanden sie sich an der richtigen Stelle, wurden sie „markiert“, indem sie mit einem Exponenten versehen wurden. Die Markierung war nötig, damit der Baum von unten nach oben in jeder Iteration abgesucht werden konnte, damit so sortierte Bezeichner von noch nicht einsortierten Bezeichnern unterschieden werden konnten. Für das Zusammenfassen von Monomen ist es somit ebenfalls nötig eine Ordnung auf Monome festzulegen, nach der sortiert werden kann. Sind zwei Monome gleich, so müssen sie zusammengefasst werden, d. h. ihre Konstanten müssen summiert werden. Für ein Monom m ermittle $k(m)$ die darin vorkommende Konstante.

$$k(m) = \begin{cases} c, & \exists c \in m, c \in \mathbb{R} \\ 1, & \text{sonst} \end{cases} \quad (4)$$

Für die Monome m_1 und m_2 mit gleicher Bezeichnermenge fasse $\text{merge}(m_1, m_2)$ die Monome zusammen.

$$\text{merge}(m_1, m_2) = (m_1 \setminus k(m_1)) \cup \{k(m_1) + k(m_2)\} \quad (5)$$

Eine geschickte Markierung von bereits einsortierten Monomen ist nötig. Es empfiehlt sich den obersten Multiplikationsknoten im Unterbaum, an dem ein ganzes Monom hängt, zu markieren, damit dieser in der Mustersuche innerhalb der aus dem ersten Schritt der Transformation nach links abgeleiteten Summe berücksichtigt werden kann. Es wird ein neues Symbol Mon3 (kurz für *Monom*) eingeführt, welches sich mathematisch nicht von dem Multiplikationssymbol Mul3 unterscheidet. Ist der oberste Multiplikationsknoten eines Monoms mit Mon3 markiert, so soll das bedeuten, dass das Monom an seiner endgültigen Stelle im Polynom einsortiert wurde. Am Ende dieses Transformationsschritts sind also nur Konstanten, Add3 - oder Mon3 -Knoten Kinder eines Add3 -Knotens. Die in Abbildung 5 aufgelisteten Baumtransformationsregeln realisieren den letzten Transformationsschritt.

(A) Assoziativgesetz		
$\text{Add3}(t_1, \text{Add3}(t_2, t_3))$	\rightarrow	$\text{Add3}(\text{Add3}(t_1, t_2), t_3)$ (1)
(B) Konstantenfaltung		
$\text{Add3}(c_1, c_2)$	\rightarrow	c_1+c_2 (2)
$\text{Add3}(t, c)$	\rightarrow	$\text{Add3}(c, t)$ (3)
(C) Sortieren und Zusammenfassen von Monomen		
$\text{Add3}(t, v)$	\rightarrow	$\text{Add3}(t, \text{Mul3}(1, \text{Pot3}(v, 1)))$ (4)
$\text{Add3}(v, t)$	\rightarrow	$\text{Add3}(\text{Mul3}(1, \text{Pot3}(v, 1)), t)$ (5)
$\text{Add3}(t, \text{Pot3}(v, c))$	\rightarrow	$\text{Add3}(t, \text{Mul3}(1, \text{Pot3}(v, c)))$ (6)
$\text{Add3}(\text{Pot3}(v, c), t)$	\rightarrow	$\text{Add3}(\text{Mul3}(1, \text{Pot3}(v, c)), t)$ (7)
$\text{Add3}(c, \text{Mul3}(t_1, t_2))$	\rightarrow	$\text{Add3}(c, \text{Mon3}(t_1, t_2))$ (8)
$\text{Add3}(\underbrace{\text{Mul3}(t_1, t_2)}_{m_1}, \underbrace{\text{Mul3}(t_3, t_4)}_{m_2})$	\rightarrow	wenn $m_1 = m_2$: $\text{merge}(m_1, m_2)$ wenn $m_1 < m_2$: $\text{Add3}(\text{Mon3}(t_1, t_2), \text{Mon3}(t_3, t_4))$ wenn $m_1 > m_2$: $\text{Add3}(\text{Mon3}(t_3, t_4), \text{Mon3}(t_1, t_2))$ (9)
$\text{Add3}(\underbrace{\text{Mon3}(t_1, t_2)}_{m_1}, \underbrace{\text{Mul3}(t_3, t_4)}_{m_2})$	\rightarrow	wenn $m_1 = m_2$: $\text{merge}(m_1, m_2)$ wenn $m_1 < m_2$: $\text{Add3}(m_1, \text{Mon3}(t_3, t_4))$ wenn $m_1 > m_2$: $\text{Add3}(\text{Mon3}(t_3, t_4), m_1)$ (10)
$\text{Add3}(\text{Add3}(t_1, \underbrace{\text{Mon3}(t_2, t_3)}_{m_1}), \underbrace{\text{Mul3}(t_4, t_5)}_{m_2})$	\rightarrow	wenn $m_1 = m_2$: $\text{Add3}(t_1, \text{merge}(m_1, m_2))$ wenn $m_1 < m_2$: $\text{Add3}(\text{Add3}(t_1, m_1), \text{Mon3}(t_4, t_5))$ wenn $m_1 > m_2$: $\text{Add3}(\text{Add3}(t_1, m_2), m_1)$ (11)

Abb. 5: Baumtransformationsregeln für den letzten Transformationsschritt. Es werden Konstanten gefaltet und Monome mit gleichen Bezeichnern und zugehörigen gleichen Exponenten zusammengefasst. m_1 und m_2 stehen für die gekennzeichneten Monome.

Bereits im ersten Transformationsschritt wird die Polynomsumme nach links abgeleitet. Dies hat, wie bereits im zweiten Transformationsschritt, den Vorteil, dass so die Anzahl der nötigen Transformationsregeln minimiert wird. In dem nach links abgeleiteten Baum werden durch Anwendung des Assoziativgesetzes Konstanten nach unten gereicht. Benachbarte Konstanten werden schließlich durch Addition zusammengefasst. Monome innerhalb des Polynoms sind immer mit einem Multiplikationsknoten Mul3 gewurzelt. Durch Erkennen dieser Knoten bei der Baumustersuche lassen sich alle Monome innerhalb der

Summe finden. Die an diesen Knoten hängenden Terme werden nun entsprechend einer Ordnung auf Monomen verglichen, um die Monome zu sortieren. Ähnlich wie im zweiten Transformationsschritt werden so entsprechend der Ordnung kleinere Monome im Polynom nach links und somit im nach links abgeleiteten Baum der Summe nach unten gereicht. Werden dabei zwei benachbarte Monome als gleich befunden, indem sie dieselben Bezeichner mit gleichen Exponenten aufweisen, so werden sie durch `merge()` zusammengefasst. Kann ein Monom aufgrund der Ordnung nicht weiter im Baum nach unten gereicht werden, wird es als einsortiert markiert, indem es mit einem neuen Symbol `Mon3` versehen wird. Das Monom ist strukturell anschließend dasselbe wie zuvor, es ist lediglich in `Mon3` gewurzelt. Somit lassen sich in der Baummustererkennung bereits einsortierte Monome von noch nicht einsortierten Monomen unterscheiden, sodass die Sortierung sich Iteration für Iteration im Baum des Polynoms nach oben arbeitet. So wird jedes Monom einsortiert und gegebenenfalls mit anderen Monomen zusammengefasst. Einzelne Bezeichner oder potenzierte Bezeichner sind ebenfalls gültige Monome. Damit auch diese korrekt einsortiert werden können, sind die Regeln (3) bis (7) nötig. Hier werden diese Monome durch eine Multiplikation mit 1 mit `Mu13` gewurzelt.

4 Zusammenfassung

Ergebnis dieser Arbeit ist, dass es im Bereich des Übersetzerbaus eine praktikable Möglichkeit ist Übersetzerbautechnologien zu verwenden, um arithmetische Terme umzuformen, insofern eine Normalform als Ziel vorgegeben wird. Die Überführung von arithmetischen Ausdrücken in ein normalisiertes Polynom mittels Baumtransformation konnte durch eine Zerlegung in kleinere Teilprobleme realisiert werden. Hierfür wurde eine Technik entwickelt, wie man durch Angabe eines einzigen Baumersetzungssystems die Baumtransformation in Phasen ablaufen lassen kann. Da bei der Termumformung nach einem Umformungsschritt nicht immer die gewünschte Normalform resultiert, wurde zudem eine Möglichkeit gezeigt die Baumtransformation iterativ durchzuführen.

Für die Korrektheit des dargelegten Baumersetzungssystems wird auf die Masterarbeit, in dessen Rahmen diese Arbeit entstand, verwiesen [Kn15]. Es war hierfür nötig zu zeigen, dass die Baumersetzungssysteme der jeweiligen Transformationsschritte sowohl noethersch als auch konfluent sind und stets in den angegebenen Normalformen resultieren.

Es stellt sich jedoch die Frage, ob die in dieser Arbeit entwickelte Technik verbessert werden kann. So ist beispielsweise die Konstantenfaltung zum Zusammenfassen von Bezeichnern gemessen an ihrer mathematischen Einfachheit eine recht aufwendige Operation, da beispielsweise bei dem in dieser Arbeit vorgestellten Verfahren in bestimmten Fällen viele Konstanten anfallen können, die gefaltet werden müssen.

Die Laufzeit des entwickelten Programms war im Rahmen dieser Arbeit nicht von primärer Wichtigkeit. Laufzeituntersuchungen waren somit nicht Bestandteil dieser Arbeit. Deren Ergebnisse wären jedoch hilfreich zur Bewertung der Praktikabilität der verwendeten Technik der Baumtransformation hinsichtlich des Kontextes dieser Arbeit. Oben genannter Ansatzpunkt könnte die Laufzeit verbessern. Ein weiterer Ansatzpunkt hinsichtlich

der Laufzeit und Effizienz wären Untersuchungen, die den Umfang des zu verwendenden Transformationsregelsatzes betrachtet. Es zeigte sich, dass die nötigen Transformationsregeln für das Ziel der Arbeit hinsichtlich ihrer Anzahl überschaubar waren. Durch das Vergrößern des Regelsatzes wäre es möglich größere Muster abzufangen und somit Aktionen mehrerer Transformationsiterationen in einer Aktion zu vereinen. Der Preis wäre jedoch eine höhere Anzahl von im Baum zu suchenden Mustern. Es stellt sich also die Frage, ob die Verwendung weniger Regeln mit weniger abzusuchenden Mustern und dafür vielen Transformationsiterationen performanter ist, als die Einsparung von Iterationen durch die Verwendung mehrerer Regeln und somit vielen abzusuchenden Mustern.

Die BURS-Theorie bietet durch Kostenoptimierung die Möglichkeit die Anwendung von Transformationsregeln z. B. in der Reihenfolge zu beeinflussen. Die Ziele dieser Arbeit konnten mit einheitlichen Kosten erreicht werden. Wiederum wären in Hinblick auf die Performanz Untersuchungen interessant, wie man durch Kostensteuerung das bestehende System gegebenenfalls verbessern könnte.

In dieser Arbeit sind Eingabeausdrücke noch eingeschränkt. Die entwickelte Technik müsste noch um Transformationsregeln oder gar Transformationsschritte erweitert werden, um beliebige arithmetische Ausdrücke als Eingabe zuzulassen.

Literaturverzeichnis

- [Ac04] Achterberg, Tobias: SCIP - a framework to integrate Constraint and Mixed Integer Programming. Bericht 04-19, ZIB, 2004.
- [Eli16] Eli. <http://eli-project.sourceforge.net/>, Stand: Feb. 2016.
- [Em94] Emmelmann, H.: Codeselektion mit regulär gesteuerter Termersetzung. Berichte der Gesellschaft für Mathematik und Datenverarbeitung - 241, 1994.
- [Fa01] Fateman, Richard J.: A review of Macsyma. IEEE Trans. on Knowledge and Data Eng, 1:133–145, 1982–1984, 2001.
- [Gr92] Gray, Robert W.; Levi, Steven P.; Heuring, Vincent P.; Sloane, Anthony M.; Waite, William M.: Eli: A Complete, Flexible Compiler Construction System. Commun. ACM, 35(2):121–130, Februar 1992.
- [Kn15] Knispel, F.: Untersuchungen von Baumersetzungsstrategien zur Überführung von arithmetischen Ausdrücken in ein Polynom. Masterarbeit, 2015.
- [Ma16a] MathWorks. <http://de.mathworks.com/products/matlab/>, Stand: Feb. 2016.
- [Ma16b] Maxima. <http://maxima.sourceforge.net/>, Stand: Feb. 2016.
- [PLG88] Pelegrí-Llopert, E.; Graham, S. L.: Optimal Code Generation for Expression Trees: An Application BURS Theory. In: Proceedings of the 15th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages. POPL '88, ACM, New York, NY, USA, S. 294–308, 1988.
- [Pr16] Projekt - Entwicklung hocheffizienter Pumpensysteme. http://swt.informatik.uni-halle.de/projekte/46534_2710967/, Stand: Feb. 2016.

Sigreturn Oriented Programming is a real Threat

Rémi Mabon¹

Abstract: This paper shows that Sigreturn Oriented Programming (SROP), which consists of using calls to *sigreturn* to execute arbitrary code, is a powerful method for the development of exploits. This is demonstrated by developing two different kinds of SROP based exploits, one asterisk exploit which was already portrayed in the paper presenting SROP, and one novel exploit for a recently disclosed bug in the DNS address resolution of the default GNU C library. Taking advantage of the fact, that these exploits have very few dependencies on the program being exploited, a library is implemented to automate wide parts of SROP exploit creation. This highlights the potential of SROP in respect to reusable and portable exploit code which strongly supports the conclusion of the original paper: SROP is a real threat!

1 Introduction

In the world of security, there is a constant cat and mouse game between programmers, which want to make software safer and attackers, which try to exploit vulnerabilities. This paper focuses on the Sigreturn Oriented Programming (SROP) technique presented by Erik Bosman and Herbert Bos in 2014 [BB14].

First, the different preconditions and ways to exploit SROP are shown in section 2. To demonstrate the use of both SROP exploitation techniques on real targets, section 3 focuses on the implementation of the SROP Chain based Asterisk exploit mentioned in [BB14] and the following section 4 demonstrates the use of an SROP based attack using a known address on a target which resolves DNS addresses with *glibc*. The reusability of SROP attacks is then demonstrated in section 5 by analyzing how much of the code used in the previous exploits can be reused to exploit a self made exploitable binary. This results in a reusable library for SROP attacks. Following this, section 6 gives an estimate on how much some defence mechanisms mitigate the danger of SROP attacks. Finally, a conclusion is given in section 7.

2 Sigreturn Oriented Programming

Sigreturn calls exist in UNIX to restore the user context after a program is interrupted by a signal. The signal handling procedure as seen by the program looks as follows [BB14]:

¹Furtwangen University, Robert-Gerwig-Platz 1, 78120 Furtwangen im Schwarzwald, remi.mabon@hs-furtwangen.de

- A signal frame is added to the stack. This frame contains the current value of all registers.
- A new return address is added to the top of the stack. This return address simply leads to a small stub which executes the *sigreturn* system call.
- The signal handler is called. What the signal handler does depends on the signal received.
- After the signal handler is done, the return address is used to execute the *sigreturn* system call if the program wasn't terminated.
- The *sigreturn* call uses the signal frame to restore all registers to their previous state.
- Finally, the execution of the program is continued exactly as before.

SROP exploits the fact, that UNIX systems don't check if a signal was sent when a program executes the *sigreturn* system call. Since the *sigreturn* call restores all registers from stack, this can be extremely powerful when the attacker has control of the stack as all registers can be set to attacker controlled value with one single gadget: *syscall* and *return*. Furthermore, this gadget can always be found at the same locations on some systems, e.g. Linux with a kernel older than 3.3. A list of systems where this is the case is provided in [BB14]. The attack is further facilitated by the fact that the *sigreturn* call is executed by using the processor instruction *syscall* which can also execute all other system calls depending only on the value of the *AX/EAX/RAX* register (the name of the register varies depending on the architecture). This way, by chaining *sigreturn* and other system calls, arbitrary code can be executed with only this one gadget. As demonstrated in [BB14], such a chain is even Turing complete without using any shellcode.

Depending on the concrete situation, there are at least two possible sets of preconditions needed to execute arbitrary code using *sigreturn*. The first one can be used if the attacker knows an address with user controlled data and is shown in subsection 2.1. The second one bypasses this need by using an SROP chain but requires a known file descriptor to which the attacker can send data. This is shown in subsection 2.2.

2.1 Known address

To execute arbitrary code using a known address with user controlled content, the following preconditions have to be met [BB14]:

1. The attacker needs a way to control the instruction pointer, e.g. through a *return* instruction on stack overwritten by the attacker.
2. The stack pointer has to be located on data controlled by the attacker where adding *NULL* bytes is possible.
3. The attacker has to know the address of a piece of data he controls. This can either be on the stack or in any other user controlled location e.g. some buffer.

4. Either the address of a *sigreturn* gadget or the address of a *syscall* gadget and a way to set the *AX/EAX/RAX* to the right value have to be known.

The preconditions 1 and 2 can be met with a simple stack overflow, provided that the vulnerable code allows *NULL* bytes. The precondition 3 can be met easily in most cases if the exact version of the program is known and the application is not compiled as position independent executable, since the location of buffers and other data pieces can be looked up by debugging the exact same version of the program locally. If this is not possible or the program was compiled using the *PIE* flag, a memory disclosure vulnerability may be needed to get a valid address. As mentioned earlier, precondition 4 can be met trivially for systems with *sigreturn* gadgets at fixed locations. On other systems, the same restrictions as for precondition 3 apply.

The attack flow with this flavor of SROP looks as follows [BB14]:

1. The stack is overwritten by the attacker using a vulnerability.
2. When the program executes a *return* instruction, the program flow is redirected to the *sigreturn* gadget by an overwritten instruction pointer stored on the stack.
3. The *sigreturn* call loads all register values from the stack. This way, the registers can be prepared for using an *execve* system call with pointers to the attacker controlled data as arguments.
4. The application selected by the attacker can then be executed by using the gadget to place another system call. This way, the attacker can for example spawn a shell.

If the attacker wants to perform more complex operations than spawning a shell or starting some other program, step 3 can be changed to call *mprotect* to make the user controlled data executable. By setting the instruction pointer in the signal frame to the user controlled data, arbitrary code can then be executed.

2.2 SROP Chain

This attack is far more complicated, but has other preconditions which can make it more suitable for some situations. Because of differing system call numbers, it only works as depicted with x86_64 Linux versions. The attack should also work with small modifications on x86, OSX BSD and ARM in thumb mode but additionally needs a way to set the appropriate register (*EAX/RAX/R7*) to 3 for a read call or to the correct value for a *sigreturn* call, either with a system call or an additional gadget [W3]. The exact reason for this will be made clear following the presentation of the attack. The preconditions are as follows [BB14]:

1. The attacker needs a way to control the instruction pointer, e.g. through a *return* instruction on stack overwritten by the attacker.

2. The stack pointer has to be located on data controlled by the attacker where adding *NULL* bytes is possible.
3. The address of a *syscall* and *return* gadget has to be known.
4. The *RAX* register has to be set to 15 at the beginning.
5. The location of any writable page in memory has to be known.
6. The data sent to a pipe, file or socket with a known file descriptor number is controlled by the attacker.

For the preconditions 1 to 3, the same restrictions as in the previous section apply but the precondition 4 is far easier to match. Since the content and location of this page don't matter, any random page in the data section of the program can be picked. Even if the program changes and the data section moves a bit, the attack is not affected as long as the selected page is still a valid writable page. This way, the exact same attack can work for different versions of a program and can in some cases even work for other programs, as long as there is a writable page at the same place. Precondition 5 can be more complicated to meet depending on the situation. In cases where the attacker has access to the standard input stream of the program, for example when exploiting a binary with a *setuid* bit to escalate privileges, this is trivial. On the other hand, when exploiting a remote code execution vulnerability on a web server which may have hundreds of open connections, this can be far more difficult. The difficulty is mitigated a little by the fact that file descriptor numbers are always allocated in sequence and that the data can be sent over multiple connections. For example, if it is known that probably less than 50 users are connected to a web service, one could pick the file descriptor number 50 and send the needed data over 50 connections. If the initial assumption is right, one of the connections should get the file descriptor number 50 which would lead to success.

To perform this attack, a chain of *sigreturn* and other system calls is combined with user supplied data read from the file descriptor to execute arbitrary code. The concrete steps for this are as follows, a graphical overview can be seen in [BB14].

1. Using a stack overflow vulnerability, the return pointer of a function is replaced with a pointer to the *sigreturn* gadget. If only a *syscall* gadget is available, the *RAX* register has to be set to 15 so that the *sigreturn* system call is executed. A signal frame filled only with zero values except the following content has to be placed right beneath it.
 - a) Stack pointer(*RSP*): Address of the writable page.
 - b) Instruction pointer(*RIP*): Address of the *syscall* and *return* gadget.
 - c) Akkumulator register(*RAX*): System call number for *read*.
 - d) *RDI* register: File descriptor number, where to read from.
 - e) *RSI* register: Address of the writable page, where to write to.
 - f) *RDX* register: 306, number of bytes to read.

2. After the *sigreturn* gadget is executed, the *syscall* and *return* gadget is executed again because of the instruction pointer set in the frame. Because of the values in the registers, the system call *read* is executed which can write up to 306 bytes from the data received on the file descriptor to the writable page. The second part of the exploit is now sent over the file descriptor. This part contains the address of the *syscall* and *return* gadget three times, then a signal frame and following this, one can include arbitrary code or information. The amount of data received has to be 306 bytes for the next steps to work.
3. After having received the data, the *read* function returns with the number of bytes read in the *RAX* register. Our writable page is populated with the data that was sent and the stack pointer points to the beginning of our writable page. This leads to another call to the *syscall* function. Because of the value 306 returned by the *read* function, the system call *syncfs* is executed. This system call should always return with 0, the *syscall* number of *read* on x86_64, in the *RAX* register, as long as the file descriptor is still valid.
4. Another system call is executed due to the pointer on the stack. The value 0 in the *RAX* register leads to another *read* call. This time, the data sent is unimportant, the only thing that matters is that exactly 15 bytes are sent. This leads to *RAX* containing the value 15, the system call number for *sigreturn*.
5. Our last pointer initiates yet another system call. This time, the value 15 leads to another *sigreturn* call. The content of the following signal frame can be chosen according to ones need. One good option is executing an *execve* call with pointers to the data below the signal frame as arguments, another interesting option is calling *mprotect* to make the data below the frame executable.
6. The selected system call is executed. Either we are already done because we initiated *execve* directly, or we now have attacker controlled executable code at a known location which makes further exploitation trivial.

As could be seen, this technique uses *syncfs* to set the *RAX* register to 0 for a *read* call which finally enables the attacker to perform a *sigreturn* call. Therefore, this method is more difficult on other systems, since the *read* system call is executed with the value 3 on most. Of course if a *sigreturn* gadget is available, the *syncfs* and *read* call can be skipped altogether and the rest of this technique should work on all common systems. This is especially interesting on Android versions below 5.0 since they mostly use the ARM Linux kernel versions below 3.11 which maps a *sigreturn* call at a known address for all processes [BB14, An].

2.3 Conclusion

As could be seen in this section, using *sigreturn* during attacks is quite powerful. With the use of a single gadget, arbitrary code can be executed with preconditions which can

be fulfilled easily in some cases. What makes it especially interesting is, that as long as the preconditions are met, the internal structure of an application almost doesn't matter for SROP based attacks.

3 Asterisk Exploit

The Asterisk Exploit is made possible by a stack consumption vulnerability with the CVE number "2012-5976" [Vua], for which a ROP remote execution attack was presented in [As]. The vulnerability being used lies in the HTTP Asterisk Management Interface and is due to an *alloca* call with a remotely supplied Content-Length value when receiving a HTTP POST header. The difficulty in exploiting it relies in the fact, that the allocated buffer is filled with the *fgets* method, which makes it impossible to wrap the stack pointer around to a location higher on the stack, since *fgets* doesn't allow values greater than *0x7FFFFFFF*. This means, that it isn't possible to alter the stack of the thread which created the buffer. The way around this is to create a second connection to the management interface to start an additional thread. As long as no other thread was created in between, the distance between the stack of the first connection to the stack of the second connection is defined in *AST_STACKSIZE* and is exactly *0x3C000*. By using a greater value, one can create a buffer on the first connection which contains the stack of the second connection. The steps to overwrite the return instruction pointer and other parts of the stack are therefore as follows [As]:

1. Creating two remote connections to the Asterisk Management Interface.
2. Sending a POST header with the Content-Length set to *0x3C900* to the first connection.
3. Sending approximately *0x700* bytes of data to the first connection.
4. Sending a POST header with the Content-Length set to *0x200* bytes to the second connection.

By sending the right value of bytes, the buffer pointer of the first thread can be placed directly above the return instruction pointer of the *_IO_getline_info* method of the second thread. Placing it there is essential, since the *RAX* register then contains the number of bytes read when our payload is executed, which is necessary to fulfill precondition 4 mentioned in section 2.2. Now it is time to send the first SROP payload:

3.1 First SROP Payload

To create the first SROP payload, we first need to know the address of a *syscall* and *return* gadget, the location of a writable page in memory and the number of a file descriptor on which data can be sent.

Since Asterisk was deployed on a server with Debian 7.9, getting the address of a *syscall*

and *return* gadget is trivial, as this distribution still uses a *syscall* page to provide some system functions [BB14]. If this wasn't the case, we would have to find a *syscall* and *return* gadget in the binary. This can be done with tools such as ROPgadget [Gi].

To find a writable page, we just have to use the *readelf* utility on the Asterisk binary [re]. This way, the different sections of the binary are displayed. Any page in the *.data* and *.bss* section can be used.

The final requirement is a valid file descriptor number to which we can write. While no usable file descriptor number is known, we can work around this as was described in section 2.2 by using many connections to the server. These connections should be opened before executing the first payload as the *read* call will return with an error if it is applied to a file descriptor number which isn't opened yet.

The first payload can now be created as was explained in section 2.2 with the values we selected before. To create the concrete signal frame, one can either use the SROP library referred to in [MP15] or, since it is the same as a context frame, one can use the *ucontext* struct which is defined in *ucontext.h*.

After having sent the first payload to the first connection, the first payload can be executed by sending 15 bytes to the second connection and closing it.

3.2 Second Payload

The second thread should now be waiting on 306 bytes of data on the connection to which the selected file descriptor number belongs. To continue the exploit, we have to create the second payload as described in section 2.2. The values for the signal frame and the data below it can be chosen freely to execute any system call we like. For further exploitation, using *mprotect* to make the writable page executable was chosen.

One problem building the second payload is, that only 306 bytes are available. This is critical, since the *ucontext* struct alone has a size of 944 bytes. This problem can be solved by building a struct which only includes *uc_flags*, **uc_link*, *uc_stack*, *gregs* and *fpregs* from the *ucontext* struct since the rest of the struct is not needed for this application. This gives us a size of 232 bytes for the signal frame. Together with the 3 pointers to the *syscall* and *return* gadget and a pointer to our own data to execute it, this leaves us with 42 bytes for our own code. Since this isn't enough for most applications, a shellcode was added which simply reads from the file descriptor number again and appends the result to the already existing code. This way, an arbitrary amount of code can be executed as long as it fits in the *.data* or *.bss* section of the binary.

After the second payload was constructed, it still has to be sent to the right connection and be read by the command issued by the first payload. One thing which makes this more difficult is, that each connection we created spawned an own thread which tries to read from it already and which closes the connection as soon as it received a request and responded to it. One option to get a window of opportunity in which our payload can read from the connection without being disturbed is to send a login request to the owner of the connection. While the original owner checks if the account information is valid, one has enough time to communicate with our payload.

To finish our exploit, we have to go through all connections we opened and send every

one a login request followed by our second payload, fifteen random bytes and finally the shellcode we want to execute - in this case a reverse shell connection to another server taken from [Li].

4 Glibc Exploit

In early 2016, a stack based overflow vulnerability with the CVE number “2015-7547” concerning DNS address resolution in the widely used glibc was disclosed [Vub]. The exploitation path for this vulnerability was outlined in the patch notes on a glibc mailing list and looks as follows [PA]:

- The attacker needs to be able to control the DNS answers received by the program.
- The program to be exploited must send out two DNS queries at once, e.g. for the A and AAAA records.
- The attacker needs to respond to both DNS queries. The first response needs to be a valid DNS response with a length of 2048 bytes to completely fill the allocated stack buffer. The second response now triggers a malloc call to create a new buffer with 65535 bytes since the response buffer is already full. To trigger the bug, the second response has to be flawed in a way that gets glibc to retry the query.
- Glibc then sends two more requests and reuses the 2048 bytes big buffer placed on the stack to save the corresponding responses. Due to a bug, the size of the buffer is now wrongly specified as the size of the heap buffer, 65535 bytes.
- The attacker can now send an overlong response to the program which leads to an overflow of the stack buffer.

This way, a Denial of Service attack against any target still using a vulnerable version of glibc is easily created.

Remote code execution, as tested on Debian 7.9 using the newest glibc package before the bug fix, isn't quite as simple. This is due to the fact that the address of the stack buffer used to save the result is saved on a part of the stack which is overwritten. Setting this address to *NULL* leads to an assert failure and writing any value differing from the initial one to it leads to a free call on this address, which causes an exception if the address doesn't point to a valid freeable memory block. Therefore, a stack memory disclosure vulnerability is also needed to be able to enable code execution. This means that the only additional preconditions needed for an address based SROP attack are a method to set *RAX* to 15 and a *syscall* gadget. These can be fulfilled as described in section 2.1.

To create a remote code execution exploit, a simple exploitable program was implemented. This program leaks a stack address to the attacker and then tries to resolve an address using the *getaddrinfo* method of glibc as shown in [fj]. To make the exploitation easier, an attacker controlled server is set as the primary DNS server of the server which is to be attacked although this could be bypassed if the attacker is able to intercept the messages sent by the vulnerable program through some other way. The attacker can now trigger the

vulnerability as previously described. The only restriction is, that the memory location containing the address of the initial response buffer has to be filled with the same value as before. This can be done by calculating the address using the leaked stacked address and some fixed offset. Finally, arbitrary code can be executed using SROP as shown in section 2.1.

5 Reusability of SROP Exploits

To demonstrate the reusability of SROP exploits, two simple exploitable programs were created, one for address based exploits and one for SROP chain exploits. These programs were built in a way to fulfill all conditions required for the specific exploit.

5.1 SROP Chain based Exploit

This program contains a large global char array, so that enough writable space is available. After starting, it simply executes a function with a return value of type long. This function first allocates a variable from type long and a char array of size 10 on the stack, then asks how many bytes it should receive and uses *read* to write the number of bytes specified to the array. Since the number of bytes that should be written is provided by the user and isn't checked by the program, a stack overflow can be caused trivially by sending it a number greater than the size of the array. Finally, the function returns the value of the previously allocated long variable.

After taking a look at the previous Asterisk exploit, it was determined that the two SROP Chain payloads only depend on three parameters:

- The address of the *syscall* and *return* gadget.
- The address of the writable page.
- The file descriptor number used to communicate with the exploit.

The SROP payload creation instructions were therefor extracted from the Asterisk exploit to a library which then contains two simple to use functions to create the first two payloads with the aforementioned parameters. Since the final shellcode also doesn't depend on the application, this part could also be copied without altering it.

The address of the *syscall* and *return* gadget and of the writable page were acquired in the same way as for the Asterisk exploit. Since the exploit communicates with the exploitable program over the standard input stream, the file descriptor number was simply set to *STDIN_FILENO*. This information was used to create the two SROP payloads.

The only thing left was adding some bytes before the first SROP payload, so that the payload starts at the return address of the function, while taking care that the long variable on the stack is overwritten by the value 15 to initiate a *sigreturn* call.

The exploit therefor only has to do the following steps:

1. Generate the two SROP payloads using the SROP library and the supplied addresses and file descriptor number.
2. Start the exploitable program.
3. Send the first payload together with the additional bytes.
4. Send the second payload.
5. Send 15 random bytes.
6. Send the final shellcode.

This shows that quite a lot of the code of a SROP chain exploit can be reused from one application to another. While, depending on the vulnerability, overwriting the return instruction pointer and the part below it on the stack may still be quite complicated, the rest of the exploit is easy, provided we have the following conditions:

- One can communicate freely with a known file descriptor number
- Finding the two required addresses isn't complicated by some defence mechanism.
- There is a way to set the *RAX* value to 15, either with a gadget or by exploiting a function which returns a value from the stack using this register.

5.2 SROP Exploit with known address

The same program was reused to demonstrate an SROP exploit with a known address with the only difference being, that the global buffer was removed and a stack memory address disclosure was added. As with the Chain based attack, all the SROP specific code from the *glibc* attack was exported to the library, since it only depends on the location of a *syscall* gadget and the address of the stack at which the exploit is located. Reusing the known address based SROP code from the *glibc* exploit worked without any modifications and due to the few restrictions, this should be the case for all vulnerabilities which give the same preconditions.

5.3 Conclusion

Compared to ROP based exploits, where, after overwriting the stack, one still has to create a ROP chain, which varies strongly from one program to another and which can be quite complicated and long depending on the available gadgets, using SROP is far simpler. When the preconditions of one of the two SROP flavors are met, the entire ROP Chain can be replaced by some reusable SROP based code which can be automatically generated by a library.

6 Impact of Defence Mechanisms

This section will examine how critical some defence mechanisms are for the usage of SROP exploits. The examined mechanisms are the default implementation of address space layout randomization (ASLR) on Linux, position independent executables (PIE) and Oxymoron.

6.1 Address Space Layout Randomization

ASLR is a technique to map each part of a program to a random address on start. It is activated by default on Linux since the kernel version 2.6.12 released in 2005. The problem with the Linux implementation is, that for normal executables, while some parts of the program, like external libraries, are added at random addresses, the program code and the data sections still always are at the same location [Ho]. Since the gadget requirements of SROP are quite low and the data sections can be used as writable pages, ASLR shouldn't affect SROP Chain exploits in most cases.

6.2 Position Independent Executables

By compiling an application as a PIE, one can get Linux to put every section of the program at a random address. This makes SROP attacks far more difficult, since the addresses of writable pages and gadgets are not fixed. But SROP attacks, like ROP attacks, might still be possible if the program also has a memory disclosure vulnerability [As].

6.3 Oxymoron

Oxymoron improves on ASLR by splitting every part of the program at the page level and placing every page at a random location. With Oxymoron, finding out an address through a memory disclosure only helps locating one single page of a section, which makes ROP attacks very difficult, since in most cases, a single page won't contain all the gadgets needed to mount an attack. Because SROP attacks require only one or two gadgets and the only other address information needed is the location of some writable page, an SROP attack on Oxymoron protected executables should still be possible in some cases [MP15].

7 Conclusion

This paper demonstrates how SROP exploits can be created and that SROP makes stack overflow based attacks easier compared to ROP attacks when some conditions are met. This is especially true since large parts of the exploit may be reused between different applications which is shown by creating an application independent SROP library for the

x86_64 architecture.

While some existing protection mechanisms can make SROP exploitation far more difficult to perform, they still impede SROP attacks less than ROP attacks because of the lower gadget requirement. This makes SROP a real threat and highlights the need for SROP mitigation techniques for Unix systems like the ones proposed in [BB14].

References

- [An] Android Versions - eLinux.org. http://elinux.org/Android_Versions. Accessed: 2016-02-16.
- [As] Asterisk Exploit — Exodus Intelligence. <http://blog.exodusintel.com/tag/asterisk-exploit/>. Accessed: 2016-02-15.
- [BB14] Bosman, Erik; Bos, Herbert: Framing Signals â A Return to Portable Shellcode. IEEE Symposium on Security and Privacy, pp. 243–258, 2014.
- [fj] fjserna/CVE-2015-7547. <https://github.com/fjserna/CVE-2015-7547/blob/master/CVE-2015-7547-client.c>. Accessed: 2016-04-25.
- [Gi] Github JonathanSalwan/ROPgadget. <https://github.com/JonathanSalwan/ROPgadget/>. Accessed: 2016-02-15.
- [Ho] How Effective is ASLR on Linux Systems? <http://securityetalii.es/2013/02/03/how-effective-is-aslr-on-linux-systems/>. Accessed: 2016-02-15.
- [Li] Linux/x86-64 - Reverse TCP shell - 118 bytes. <http://shell-storm.org/shellcode/files/shellcode-857.php>. Accessed: 2016-02-15.
- [MP15] Mithra, Zubin; P., Vipin: Evaluating the theoretical feasibility of an SROP attack against Ozymoron. International Conference on Advances in Computing, Communications and Informatics, pp. 1872–1876, 2015.
- [PA] PATCH CVE-2015-7547 — glibc getaddrinfo() stack-based buffer overflow. <https://sourceware.org/ml/libc-alpha/2016-02/msg00416.html>. Accessed: 2016-04-25.
- [re] readelf - GNU Binary Utilities. <https://sourceware.org/binutils/docs/binutils/readelf.html>. Accessed: 2016-02-15.
- [Vua] Vulnerability Summary for CVE-2012-5976. <https://web.nvd.nist.gov/view/vuln/detail?vulnId=CVE-2012-5976>. Accessed: 2016-02-15.
- [Vub] Vulnerability Summary for CVE-2015-7547. <https://web.nvd.nist.gov/view/vuln/detail?vulnId=CVE-2015-7547>. Accessed: 2016-04-25.
- [W3] W3Challs Syscall tables. <https://w3challs.com/syscalls/>. Accessed: 2016-02-16.

Nutzerzentrierte Konzeption, Entwicklung und Evaluation: Informatik beGreifen - Schülerinnen und Schüler mit Lego Mindstorms und Raspberry Pi für die Informatik begeistern

Isabel Hagen¹

Abstract: In der vorliegenden Arbeit wird die Anwendung „Informatik beGreifen“ konzipiert, entwickelt und evaluiert. Mit Informatik beGreifen soll für Schülerinnen und Schüler die Informatik greifbar, also auch begreifbar gemacht und somit diese für die Informatik begeistert werden. Informatik beGreifen umfasst einen Roboter, eine entsprechende Programmieranleitung sowie die Gestaltung eines Standplatzes für die Präsentation und die Durchführung auf Veranstaltungen. Die vorliegende Arbeit zeigt, wie Informatik beGreifen in seiner Gesamtheit gestaltet werden muss, um Schülerinnen und Schüler für die Informatik zu begeistern. Dabei bedient sich die Arbeit dem Vorgehen des Human Centered Design Process. Die Ideenfindung und die nutzergerechte Umsetzung wird mit Hilfe einer verteilten Fokusgruppe durchgeführt. Als Ergebnis entsteht ein Weck-Roboter, basierend auf Lego Mindstorms und Raspberry Pi. Die Programmieranleitung ist nutzergerecht aufgebaut und leitet Schülerinnen und Schüler an, den Roboter mit Scratch/BYOB selbst zu programmieren. Durch eine summative Evaluation wird gezeigt, dass Informatik beGreifen Ängste sowie Vorurteile gegenüber der Informatik abbaut und dass Informatik beGreifen Schülerinnen und Schüler für die Informatik begeistert.

Keywords: Mensch-Maschine Interaktion, Human Centered Design Process, Design Thinking, Fokusgruppe, Lego Mindstorms, Raspberry Pi, BrickPi, domänenspezifische Sprache, Scratch, BYOB

1 Einleitung

Informatik ist aus dem täglichen Leben nicht mehr wegzudenken. Sie ist in nahezu allen technischen Geräten, in Unternehmen, Organisationen und Maschinen enthalten. Deshalb ist die Informatik ein wichtiges Berufsfeld und es besteht ein hoher Bedarf an Informatikern. Nach dem Arbeitsmarktbericht für IT-Fachleute der Bundesagentur für Arbeit vom Mai 2015 wurden in den 33.100 ausgeschriebenen Stellen im Jahr 2014 zum großen Teil nach IT-Spezialisten gesucht [Bu15]. Um dauerhaft dem hohen Bedarf an Informatikern zu begegnen ist es wichtig, dass sich Studieninteressierte für ein Informatikstudium entscheiden. Nach einer Untersuchung der Technischen Universität München ist ein Hauptgrund gegen das Informatikstudium, dass sich die Studieninteressierten das Studienfach nicht zutrauen [ELK08, S. 37]. Daraus erschließt sich die Notwendigkeit, schon bei Schülerinnen und Schülern ein Verständnis für die Informatik zu schaffen, um Vorurteile und Ängste abzubauen. In dieser Arbeit wird deshalb für Schülerinnen und Schüler der 7. bis 12. Klasse

¹ Hochschule Reutlingen, Fakultät Informatik, Alteburgstraße 150, 72762 Reutlingen, isabel.hagen@student.reutlingen-university.de

„Informatik beGreifen“ konzipiert, entwickelt und evaluiert. Informatik beGreifen stellt ein Wortspiel dar und bedeutet, dass die Informatik sowohl für Schülerinnen und Schüler greifbar, als auch begreifbar gemacht wird. Dadurch sollen mit Informatik beGreifen die Ängste und Vorurteile gegenüber der Informatik abgebaut, sowie eine Begeisterung für die Informatik erzeugt werden. Die Informatik soll in Informatik beGreifen anhand eines Roboters, bestehend aus Lego Mindstorms und einem Raspberry Pi sowie der Roboterprogrammierung, greif- und begreifbar gemacht werden. Die Schülerinnen und Schüler sollen mit Informatik beGreifen an öffentlichen Veranstaltungen, wie dem „Studieninformationstag“ und dem „Tag der offenen Tür“ angesprochen werden. Daraus ergeben sich die Forschungsfragen:

1. Wie kann Informatik beGreifen benutzergerecht entwickelt werden [F1]?
2. Welcher Roboter ist für die Ansprache der Zielgruppe geeignet [F2]?
3. Wie kann der Roboter mit Lego Mindstorms und Raspberry Pi sowie die Roboterprogrammierung benutzergerecht realisiert werden [F3]?
4. Wie kann benutzergerecht die Programmierung des Roboters erklärt und unterstützt werden [F4]?
5. Wie kann Informatik beGreifen im Rahmen der Veranstaltungen „Studieninformationstag“ und „Tag der offenen Tür“ durchgeführt werden [F5]?

Nachfolgend werden diese Forschungsfragen in vier Abschnitten behandelt. Der erste Abschnitt umfasst das Konzept für die Entwicklung und die Evaluation von Informatik beGreifen. In diesem wird das Vorgehen nach dem Human Centered Design Process sowie die Ideenentwicklung und Evaluation durch eine verteilte Fokusgruppe beschrieben. Im Abschnitt Kontextanalyse und Ideenfindung bildet die Kontextanalyse die Grundlage für alle weiteren Anforderungen in der Entwicklung und der Evaluation. Die Ideenfindung beschreibt dabei die genaue Vorgehensweise der Ideenentwicklung mit Hilfe einer verteilten Fokusgruppe. Der Abschnitt Umsetzung beinhaltet die technische Umsetzung der Roboteridee sowie die Erstellung der Anleitung für die Programmierung und die Gestaltung des Standplatzes. Im Abschnitt Evaluation werden die summative Evaluation und die Ergebnisse von Informatik beGreifen dargelegt. Im letzten Abschnitt erfolgt das Fazit mit der zusammenfassenden Beantwortung der Forschungsfragen.

2 Konzeption von Informatik beGreifen

Die benutzergerechte Entwicklung von Informatik beGreifen basiert auf dem Human Centered Design Process. Bei diesem wird der Benutzer zu einem möglichst frühen Zeitpunkt in den Gestaltungs- und Entwicklungsprozess mit einbezogen [De10, S.9]. Der Gestaltungsprozess zeichnet sich durch seinen iterativen Aufbau aus, da es zu Beginn des Prozesses nicht möglich ist, alle Aspekte der späteren Benutzerinteraktion zu erfassen [De10, S.9 und S.11]. Der Human Centered Design Process beginnt mit der Planung des menschenzentrierten Gestaltungsprozesses und umfasst vier iterative Phasen [De10, S.15]. Anhand des

Human Centered Design Process wird nachfolgend die Forschungsfrage [F1] beantwortet. Dabei erfolgt zuerst die Vorstellung der Fokusgruppe und im Anschluss eine Beschreibung der iterativen Vorgehensweise.

2.1 Verteilte Fokusgruppe

Für das Vorgehen nach dem Human Centered Design Process ist es wichtig, den Nutzer zu einem möglichst frühen Zeitpunkt in den Gestaltungs- und Entwicklungsprozess mit einzubeziehen [De10, S.9] und in den Mittelpunkt zu stellen. Um diese Ziele zu erreichen, eignen sich besonders Fokusgruppen.

Mit Hilfe einer Fokusgruppe können moderierte Diskussionen zu einem bestimmten Thema geführt werden [TS08, S.268]. Fokusgruppen bestehen aus drei bis zwölf Mitgliedern, welche potenzielle Nutzer aus der Zielgruppe darstellen [KLB04, S.2] [TS08, S.269].

Die Fokusgruppe wird von einem Moderator gelenkt, welcher bspw. mit Fragen, einem Vortrag oder einer Präsentation einen Stimulus für eine Diskussion innerhalb der Gruppe schafft [TS08, S.269]. Das Ziel einer Diskussion innerhalb der Fokusgruppe ist, die „Einstellungen, Gefühle, Vorstellungen und Ideen der Teilnehmer zu einem bestimmten Themengebiet zu erfahren“ [TS08, S.269]. Für die Ideenentwicklung, die Umsetzung und die formative Evaluation der Anwendung Informatik beGreifen wird eine Fokusgruppe von sechs zufällig ausgewählten Schülerinnen und Schülern eingesetzt. Diese sind zwischen 12 und 17 Jahre alt³.

Bei der Mitgestaltung von Informatik beGreifen können die Mitglieder der Fokusgruppe verschiedene Rollen einnehmen: Designpartner, Informant, Tester und Nutzer [Dr02, S.4]. In der Rolle des Designpartners sind die Mitglieder der Fokusgruppe gleichberechtigte Projektpartner und können von Beginn an ihre eigenen Ideen einbringen [Dr02, S.19f].

In der Rolle des Informanten informieren die Mitglieder der Fokusgruppe den Entwickler im Design-Prozess. Dafür werden die Mitglieder bereits vor der Entwicklung nach ihren Vorstellungen und Ideen gefragt. Zudem sollen sie auch während des Entwicklungsprozesses Anregungen und Feedback geben. [Dr02, S.15]

Die Rolle des Testers beginnt in dem Moment, in dem der erste Prototyp existiert. In der Rolle des Testers testen die Mitglieder der Fokusgruppe Prototypen der zu entwickelnden Anwendung. Die Kernaufgabe dieser Rolle ist, die Anwendung mit zu formen, bevor diese zum Einsatz mit der Zielgruppe kommt. Damit soll eine höhere Akzeptanz bei der Zielgruppe erreicht werden. [Dr02, S.10]

In der Rolle des Nutzers beschäftigen sich die Mitglieder der Fokusgruppe mit einer fertigen Anwendung oder einem weit fortgeschrittenen Prototypen. Dabei sollen Schwächen und Verbesserungspotentiale erkannt werden. [Dr02, S.10]

Die Mitglieder der Fokusgruppe befinden sich für den gegenseitigen Austausch jedoch nicht an einem gemeinsamen Ort. Daher handelt es sich bei der Fokusgruppe im Projekt Informatik beGreifen um eine verteilte Fokusgruppe. Daraus ergibt sich das Problem,

³ Die Alters- und Schulstruktur setzt sich wie folgt zusammen: Eine Schülerin im Alter von 12 Jahren mit der Schulart Realschule, zwei Schüler im Alter von 13 Jahren mit den Schularten Realschule und Gymnasium sowie je eine Schülerin im Alter von 14, 15 und 17 Jahren mit der Schulart Gymnasium.

wie eine Kommunikation zwischen den Mitgliedern der Fokusgruppe und dem Moderator gewährleistet werden kann. Daher wird eine Kommunikationsplattform benötigt, welche die Mitglieder der Fokusgruppe nutzen und die es ermöglicht, Prototypen oder auch die fertige Anwendung der Fokusgruppe zu zeigen. Außerdem muss das Setzen eines Stimulus auf dieser Plattform möglich sein, um bspw. kurze Filme der Fokusgruppe zu zeigen. Laut einer Studie von Bitkom aus dem Jahr 2014 liegt WhatsApp in der Zwischenzeit vor Facebook als beliebtestes soziales Netzwerk [bi14, S. 28]. Der Vorteil von WhatsApp für die Fokusgruppe ist, dass hier ein müheloser Austausch von Informationen, Videos und Bildern sowie Diskussionen möglich sind. Daher wird WhatsApp für die Ideenfindung und Entwicklung verwendet. Um die Gruppenmitglieder für die aktive Mitarbeit an der Ideenfindung zu motivieren, wird ein geeigneter Stimulus benötigt [TS08, S.269]. Als Stimulus wird ein selbstgedrehtes Video gewählt und als Kreativitätstechnik zur Generierung von Ideen, die Technik des Brainstormings angewandt. Das Video eignet sich gut als Stimulus, da es sowohl Bild als auch Ton überträgt. Zudem können kurze Filmsequenzen zur Unterstreichung von Erklärungen eingebunden werden. Der gewählte Stimulus soll damit eine ähnliche Wirkung auf die Fokusgruppe erzielen, wie die persönliche Ansprache durch eine Präsentation [TS08, S.269]. Während der 1,5 wöchigen Phase des Brainstormings müssen die unterschiedlichen Rollen des Aufgabenstellers, des Protokollanten und des Moderators besetzt werden [Sc99, S.40f.]. Die Rolle des Aufgabenstellers wird anhand des Videos durch den Entwickler übernommen. Dieser übernimmt zudem die Rolle des Moderators, um die Gruppenmitglieder mit ihren Ideen zu bestärken, unverständliche Beiträge zu erklären oder den Kreativitätsprozess vor dem Versiegen erneut anzustoßen [Sc99, S.41]. Die Rolle des Protokollanten übernimmt während dem Brainstorming WhatsApp, da alle Beiträge der Gruppenmitglieder in einem Chatverlauf für alle sichtbar festgehalten werden [Sc99, S.40].

2.2 Iteratives Vorgehen

Im Rahmen dieser Arbeit wurde Informatik beGreifen nach dem Vorgehensmodell des Human Centered Design Process entwickelt. Das Vorgehen gliedert sich in die vier, aufeinanderfolgende Schritte Kontextanalyse, Ideenfindung, Umsetzung und Evaluation. Im ersten Schritt „Kontextanalyse“ wird der Nutzungskontext von Informatik beGreifen verstanden und festgelegt. Da sich der Nutzungskontext im Verlauf des Projektes nicht ändert, bildet dieser Schritt die Grundlage für die Anforderungen in den nachfolgenden Schritten.

Im zweiten Schritt „Ideenfindung“ wird iterativ eine konkrete Roboteridee ermittelt.

Im dritten Schritt „Umsetzung“ erfolgt eine iterative Umsetzung der Roboteridee, eine benutzergerechte Umsetzung der Programmieranleitung und eine Gestaltung des Standplatzes für die Ansprache der Zielgruppe. Jeder Schritt wird dabei in einer formativen Evaluation von der Fokusgruppe evaluiert.

Im vierten und letzten Schritt „Evaluation“ erfolgt eine summative Evaluation, in der die Idee und die Umsetzung von Informatik beGreifen in seiner Gesamtheit von der Zielgruppe evaluiert wird. Die vier vorgestellten Schritte gliedern sich jeweils in eine unterschiedliche Zahl an Iterationen.

3 Kontextanalyse und Ideenfindung

Damit die richtigen Handlungsmaßnahmen bezüglich der Anwendungsanforderungen, der Gestaltung und der Evaluierung der Anwendung getroffen werden können, muss der Nutzungskontext genau definiert werden. Der Nutzungskontext ermöglicht genaue Erkenntnisse über die Benutzer, die Anwendungsziele, die Aufgaben der Benutzer sowie die Systemumgebung [De10, S.16f.].

3.1 Nutzungskontext der Benutzer

Im ersten Schritt wird die relevante Zielgruppe identifiziert und auf deren Merkmale und Ziele eingegangen. Die Zielgruppe kann nach der DIN EN ISO 9241-210:2010 sowohl ein oder mehrere Benutzer, als auch weitere Interessengruppen umfassen [De10, S. 16]. Die Anwendung soll für Schülerinnen und Schüler zwischen der 7. und der 12. Klasse konzipiert werden. Um mehr Informationen über die Zielgruppe zu erhalten und ein Verständnis für diese aufzubauen, dienen Empathie-Interviews. Durch diese Interviews sollen einerseits Informationen über Hobbies und Interessen der Zielgruppe erhalten und andererseits eine Beziehung zu den Interviewten aufgebaut werden, um von deren Zweifel und Wünschen zu erfahren. Dadurch können deren Emotionen zu bestimmten Themen erkundet werden. [d., S. 10]

Um die Zielgruppe besser zu verstehen, wurden mit insgesamt vier Schülerinnen und Schülern der Zielgruppe Empathie-Interviews durchgeführt. Der Leitfaden für diese Interviews orientiert sich an dem Verlauf für die Interviews for Empathy [d., S. 10].

Um anschließend die Informationen aus den Empathie-Interviews zusammen zu fassen, wird der Point of View erstellt [Sa14, S.32]. Dadurch soll eine einheitliche Wissensbasis bezüglich den Aussagen der Interviewpartner und den Beobachtungen des Interviewers geschaffen werden [Sa14, S.32].

Als Point of View der Empathie-Interviews lassen sich die nachfolgenden Inhalte zusammenfassen: Die vier Interviewpartner sind Schülerinnen und Schüler zwischen 12 und 17 Jahren. Bei allen Interviewten ist sehr wenig Erfahrung mit Technik oder Informatik vorhanden. In der Freizeit verwenden die Interviewten ihre technischen Geräte zum Spielen, für die eigene Unterhaltung oder für die Kommunikation mit Freunden. Von allen Interviewpartnern wird Programmieren als komplex und schwierig angesehen. Da alle Interviewten viel Spaß beim Spielen mit dem mitgebrachten Lego Mindstorms Roboter hatten, kann von einem grundlegenden Interesse für Technik ausgegangen werden. Zudem kann davon ausgegangen werden, dass Spielen unabhängig vom Alter sehr bedeutend ist.

Aus den Informationen des Point of View werden fiktive Personen, die Personas erstellt. Eine Persona wird geschaffen, um eine konkrete Vorstellung von der Zielgruppe zu erhalten [To11, S.38]. Die Persona stellt dabei den typischen Nutzer dar [Gü14, S.48].

3.2 Nutzungskontext der Anwendungsumgebung

Damit die richtigen Handlungsmaßnahmen ergriffen werden, muss neben dem Nutzungskontext der Benutzer, auch die Anwendungsumgebung untersucht werden [De10, S.18f.].

Die Anwendung Informatik beGreifen soll an zwei Veranstaltungen der Hochschule, dem Tag der offenen Türe und dem Studieninformationstag, an einem Standplatz der Fakultät Informatik präsentiert werden, um Schülerinnen und Schüler für die Informatik zu begeistern. Dabei nehmen meist Schülerinnen und Schüler der Oberstufe, an diesen Veranstaltungen teil, um sich über die Angebote der Hochschule zu informieren [MfW15][Re15]. Als Teil der technischen Umgebung wird der Einplatinencomputer Raspberry Pi als Hardwarekomponente verwendet. Um den Roboter zu bauen wird als zusätzliche Hardwarekomponente Lego Mindstorms verwendet. Für die Kommunikation zwischen dem Raspberry Pi und den Lego Mindstorms wird zudem ein BrickPi benötigt. Der BrickPi ist ein Erweiterungsboard für den Raspberry Pi, welcher Schnittstellen für die Motoren und Sensoren von Lego Mindstorms bietet [WM13]. Als Software wird die visuelle Programmiersprache Scratch/BYOB verwendet.

3.3 Ideenfindung

Für Informatik beGreifen soll ein Roboter entwickelt werden. Die Ideen für den Roboter werden dabei mit Hilfe einer verteilten Fokusgruppe über WhatsApp entwickelt. Um die Gruppenmitglieder für die aktive Mitarbeit an der Ideenfindung zu motivieren, wird ein geeigneter Stimulus benötigt [TS08, S.269]. Als Stimulus wird ein selbstgedrehtes Video gewählt und als Kreativitätstechnik zur Generierung von Ideen, die Technik des Brainstormings angewandt. Damit die Anwendung optimal den Erfordernissen der Benutzer und der Anwendungsumgebung entspricht, werden auf Basis des Nutzungskontextes Anforderungen spezifiziert [De10, S.17]. Die Ideen der Fokusgruppe werden auf den Erfüllungsgrad der Anforderungen untersucht. Dafür werden diese Anforderungen als Kriterien einer Nutzerwertanalyse verwendet [Kü14, S.7ff.]. Zusätzlich werden die Kriterien gemäß ihrer Wichtigkeit für die Entscheidungsfindung gewichtet. Laut Nutzwertanalyse und einer anschließenden Priorisierung durch einen Fragebogen, erfüllt der WakeMeBot als Weckroboter am besten die gestellten Anforderungen.

4 Umsetzung von Informatik beGreifen

Für die Umsetzung von Informatik beGreifen wurden mehrere Iterationen benötigt. Die Umsetzung besteht dabei aus der technischen Umsetzung und der benutzergerechten Umsetzung. In der technischen Umsetzung wird die Forschungsfrage [F3] nach der benutzergerechten Umsetzung des Roboters mit Raspberry Pi und Lego Mindstorms sowie der Programmierung, beantwortet. In der benutzergerechten Umsetzung wird zuerst die Forschungsfrage [F4], nach der benutzergerechten Programmierung des Roboters und die Forschungsfrage [F5], nach der benutzergerechten Umsetzung im Rahmen der öffentlichen Veranstaltungen der Hochschule, beantwortet.

4.1 Technische Umsetzung

Um den WakeMeBot umzusetzen müssen die Anforderungen von Informatik beGreifen um die technischen Anforderungen des WakeMeBots erweitert werden.

Die Anforderungen basieren dabei auf der Idee des WakeMeBots. Der WakeMeBot ist ein Weckroboter. Als Anforderung soll dieser eine schlafende Person nach einer festgelegten Zeitspanne wecken. Die Anforderungen im Weckvorgang sind, dass der WakeMeBot Musik spielt, leuchtet und davonfährt. Dabei bewegt er sich als eine Anforderung selbstständig durch den Raum. Um den Alarm zu beenden muss aufgestanden und der Roboter gesucht werden. Wenn der Schüler den Alarm beendet, bekommt er von dem WakeMeBot einen Müsli-Riegel als schnelles Frühstück. Daher muss der WakeMeBot als Anforderungen einen Ausschalter besitzen, um den Weckvorgang zu beenden und eine Vorrichtung besitzen, um die Müsli-Riegel anzubringen.

Um die Anforderungen an die technische Umsetzung zu erfüllen werden Gestaltungslösungen für die Umsetzung des WakeMeBots mit Lego Mindstorms gesucht. Der Wake-

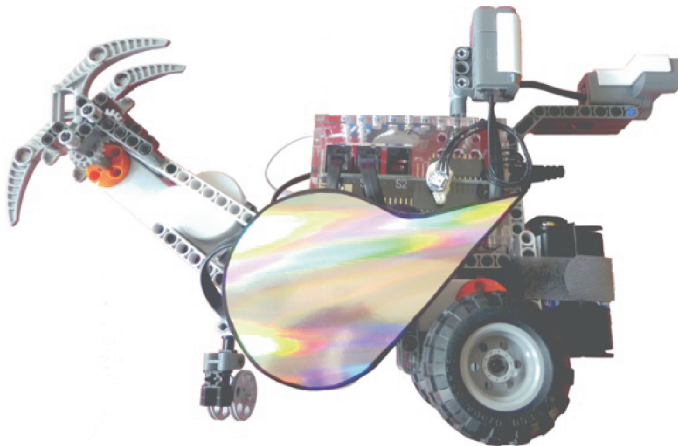


Abb. 1: Der WakeMeBot

MeBot besteht aus einem Fahrgestell, Sensoren, Greifer, Seitenverkleidung, Batteriepack sowie einem Case für BrickPi und Raspberry Pi, die über einen Grundrahmen miteinander verbunden sind (vgl. Abb. 1). Das Fahrgestell besteht aus einer Drei-Rad-Konstruktion, wie sie oft bei Lego Mindstorms Robotern zu finden ist [Pa15]. Um Hindernisse zu erkennen, ist ein Ultraschall-Sensor auf der Vorderseite des WakeMeBots angebracht. Für eine direkte Interaktion zwischen dem Benutzer und dem WakeMeBot ist ein Tastsensor an der Seite des Cases von BrickPi und Raspberry Pi montiert. Um einen Müsli-Riegel zu halten ist ein Greifer an der Rückseite des Grundrahmens angebracht. Für eine gleichmäßige Gewichtsverteilung ist das Batteriepack an der Vorderseite des WakeMeBots angebracht. Da der WakeMeBot leuchten soll, wurde als Gestaltungslösung eine LED-Erweiterung mit WS2812b NeoPixel umgesetzt. Aufgrund der elektrischen Eigenschaften kann der NeoPixel WS2812b direkt am Raspberry Pi über drei Leitungen angeschlossen werden. Die Datenübertragung an den NeoPixel WS2812b erfordert ein exaktes Timing der Signale. Daher ist eine einfache Steuerung über die GPIO-Pins des Raspberry Pis nicht möglich.

Für die Datenübertagung wird deshalb die C und Python Bibliothek `jgarff/rpi_ws281x` [gi15] verwendet, welche die Datensignale über spezielle Hardware des Raspberry Pi Controllers [Di15] erzeugt. Da der WakeMeBot Musik spielen soll, wird eine Schnittstelle für einen Lautsprecheranschluss benötigt. Bei der Verwendung des 3,5mm Audioausgang besteht jedoch ein schlechtes Audiosignal. Ein weiteres Problem ist, dass der Audioausgang bei gleichzeitiger Verwendung der NeoPixel nicht nutzbar ist. Daher wird eine externe USB-Soundkarte verwendet.

Um mit Softwarekomponenten durch die Programmiersprache Scratch/BYOB außerhalb der Scratchumgebung zu kommunizieren, enthält Scratch verschiedene Broadcast-Funktionen. Mit diesen können beliebige Zeichen über eine Remote-Verbindung gesendet und empfangen werden [Sc]. Für den WakeMeBot wurde deshalb in Python ein Translator-Skript entwickelt. Dieses verwendet die Python-Bibliotheken `BrickPi_Python` [Gia] und `rpi_ws281x` [gi15] um den BrickPi und die NeoPixel der LED-Erweiterung anzusteuern. Des Weiteren verwendet dieses neue Translator-Skript die Python-Bibliothek `scratchpy` [Gib], um von Scratch Broadcasts zu empfangen und zurück zu senden. Daher wurde in der Umsetzung des Translator-Skripts, die Ports der Sensoren fest initialisiert. Dadurch ist eine Initialisierung in Scratch/BYOB nicht notwendig. Da Translator-Skript und Scratch asynchron laufen, werden die Broadcast-Befehle in einem extra Thread abgefragt und in einem FiFo zwischengespeichert. Die Bearbeitung der im FiFo abgelegten Befehle sowie die Aktualisierung der Sensorwerte erfolgen in einer Endlosschleife. Zu Beginn jeder Schleife wird eine Pause von 60ms eingelegt, damit der Mikrocontroller des BrickPis genügend Zeit für die Verarbeitung hat. Die LEDs wurden in einem extra Thread im Translator-Skript realisiert. Dies führt zu einer Entlastung des Scratch/BYOB-Programms. Die Befehle des Translator-Skripts sind für die Zielgruppe jedoch ungeeignet, da sie für diese ohne Programmiererfahrung nicht verständlich sind. Um dieses Problem zu lösen wird eine domänenspezifische Sprache erstellt. Sie erlaubt die kompakte Lösung einer Problemstellung in der Domäne [Kr10, S.8]. Informatik beGreifen und der darin enthaltene WakeMeBot stellen eine Anwendungsdomäne dar. Für einen schülergerechten Abstraktionsgrad und eine schülergerechte Notation von Scratch, müssen „WakeMeBot-Funktionen“ implementiert werden. Da Scratch mit den WakeMeBot-Funktionen erweitert wird, entsteht eine interne DSL. Als Beispiel dieser internen DSL ist in Abb. 2 die Funktion „WB fahre geradeaus“ dargestellt. WB steht für WakeMeBot.

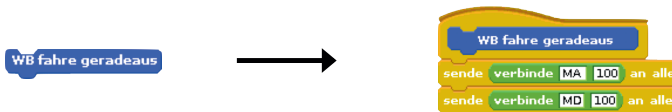


Abb. 2: DSL-Funktion „WB fahre geradeaus“

4.2 Benutzergerechte Umsetzung

Um Informatik beGreifen benutzergerecht umzusetzen ist es wichtig, die Anforderungen um das Nutzererlebnis zu erweitern. Dafür steht eine Anleitung für Scratch/BYOB zur Programmierung des Roboters zur Verfügung. In der Anleitung wird Schritt für Schritt der

Programmablauf des WakeMeBots erklärt. Der WakeMeBot soll nach einer vorher festgelegten Zeitspanne mit dem Weckvorgang beginnen. Anschließend soll er Musik spielen und bunt leuchten. Dabei fährt er so lange geradeaus, bis er mit dem Ultraschallsensor ein Hindernis erkennt. Falls er auf ein Hindernis trifft, soll er ein Stück rückwärts fahren und dann nach links wenden. Anschließend fährt er wieder geradeaus. Der WakeMeBot beendet den Weckvorgang, sobald der Tastsensor gedrückt wurde. Ist dieser gedrückt worden, wird die Musik beendet, er hält an, das Licht geht aus und der Greifer öffnet sich mit dem Müsli-Riegel. Zusätzlich werden Schülerinnen und Schüler während der Programmierung des Roboters von dem Standbetreuer der Veranstaltung unterstützt. Der Standplatz ist so gestaltet, dass dem Nutzer eine entspannte und angenehme Atmosphäre geboten wird. Dies beinhaltet einen Film über den WakeMeBot, Plakate und Lego-Bausteine für die spielerische Heranführung an den Roboter. Zudem wird den Schülerinnen und Schülern am Ende der Programmierung eine Urkunde und ein Müsli-Riegel übergeben.

5 Evaluation und Ergebnisse von Informatik beGreifen

Die summative Evaluation findet am Studieninformationstag und am Tag der offenen Tür an der Hochschule Reutlingen statt. Die summative Evaluation bewertet Informatik beGreifen in der Gesamtheit. Ziele der Evaluation sind die Einstellung der Probanden zur Programmierung zu erfahren, die Meinung der Probanden über die Anwendung zu erhalten, die Selbsteinschätzung der Probanden zu ihren eigenen Programmierkenntnissen zu erhalten, den Spaßfaktor von Informatik beGreifen zu erfahren und den Schwierigkeitsgrad der Aufgaben zu ermitteln.

An diesen zwei Tagen der Evaluation nahmen insgesamt 28 Probanden an der Evaluation teil. Davon sind 43% weiblich und 57% männlich. Von den 28 Probanden besuchen 57% das Gymnasium, gefolgt von 29% welche eine berufliche Schule besuchen. Die restlichen Probanden gehen auf die Grundschule, Haupt- oder Werkrealschule, Realschule oder Gemeinschaftsschule. Von den Befragten sind 57% der Meinung, dass Programmieren ein wenig oder sehr kompliziert ist. Lediglich 43% der Probanden sind der Meinung, dass Programmieren nicht kompliziert, einfach oder sogar sehr einfach ist. Zudem geben 50% der Probanden an, bereits über Programmiererfahrung zu verfügen. Als Programmiererfahrung gelten hierbei bereits rudimentäre Erfahrungen mit der Programmierung. Dementsprechend verfügen auch 50% der Probanden über keine Programmiererfahrung. 79% von den Probanden ohne Programmiererfahrung sind der Meinung, dass Programmieren kompliziert ist. Lediglich 21% der Probanden ohne Programmiererfahrung haben nicht diese Voreinstellung gegenüber der Programmierung. Vor der Programmierung des WakeMeBots geben die Probanden mehrheitlich an, dass sie überhaupt nicht, nur ein wenig oder weniger gut Programmieren können. Nach der Programmierung des WakeMeBots war dagegen die Mehrheit der Probanden der Meinung, dass sie ein wenig oder gut Programmieren können. Signifikant ist, dass nach der Programmierung des WakeMeBots keiner der Probanden der Meinung ist, nicht programmieren zu können. Dies zeigt, dass Informatik beGreifen eine geeignete Anwendung ist, um Ängsten und Vorurteilen gegenüber der Informatik zu begegnen.

Die Schwierigkeit der Aufgaben wurde als leicht und sehr leicht bewertet. Die Idee des

WakeMeBots gefällt bei der Evaluation 86% der Befragten. Lediglich 11% der Probanden stehen dem WakeMeBot neutral gegenüber und nur 3% gefällt die Idee nicht.

Das Feedback zum Spaßfaktor ist durchweg gut oder sehr gut. Keiner der Probanden gibt an keinen oder wenig Spaß gehabt zu haben. Bei der Befragung geben 3% an, dass sie Spaß, 36% viel Spaß und 61% sehr viel Spaß hatten.

Durch die Evaluation werden viele Beobachtungen, welche bereits bei den Interviews for Empathy gemacht wurden, bestätigt.

Die Zielgruppe hat, durch die wenigen Berührungspunkte mit der Informatik und den fehlenden Programmiererfahrungen, viele Vorurteile gegenüber der Programmierung und damit auch gegenüber der Informatik. Diese Vorurteile lassen sich wie bei der Evaluation zu beobachten, durch eigene Erfahrungen mit der Programmierung beheben. Daher ist Informatik beGreifen eine geeignete Anwendung, um Ängste und Vorurteile gegenüber der Informatik zu begegnen und diese abzubauen. Der WakeMeBot als zentraler Teil der Anwendung scheint Begeisterung auszulösen und ist deshalb ein wesentliches Element zur Zielerreichung. Da die Besucher Spaß bei der Programmierung haben und Begeisterung zeigen, kann daraus gefolgert werden, dass die Nutzer insgesamt ein positives Gesamterlebnis haben.

6 Fazit

Die Informatik ist ein wichtiges Berufsfeld, in welchem Spezialisten mit einem Hochschulabschluss gesucht werden. Gleichzeitig trauen sich aber viele Schulabgänger kein Informatikstudium zu. Daraus ergibt sich die Notwendigkeit, schon Schülerinnen und Schüler der Mittel- und Oberstufe für die Informatik zu begeistern und damit frühzeitig Vorurteilen und Ängsten gegenüber der Informatik zu begegnen. Informatik beGreifen soll die Informatik für die Schülerinnen und Schüler sowohl greifbar, als auch begreifbar machen. Dabei soll Informatik beGreifen die Ängste und Vorurteile gegenüber der Informatik abbauen und eine Begeisterung für die Informatik erzeugen. In dieser Arbeit wird Informatik beGreifen konzipiert, entwickelt und evaluiert.

Die Forschungsfrage [F1] nach der benutzergerechten Entwicklung wird durch die Anwendung des Human Centered Design Process in Verbindung mit einer Fokusgruppe beantwortet. Der Human Centered Design Process basiert auf einem guten Verständnis für den Nutzer. Dabei wird der Nutzer in den Mittelpunkt des Gestaltungs- und Entwicklungsprozesses gestellt. Als Nutzer agiert in dieser Arbeit eine verteilte Fokusgruppe. Die Kommunikation erfolgt über WhatsApp.

Die Forschungsfrage [F2] nach einem für die Zielgruppe geeigneten Roboter wird durch den WakeMeBot beantwortet. Die Eignung des WakeMeBots wurde im Rahmen der summarischen Evaluation bestätigt. Bei dieser gefielen 86% der 28 Befragten die Idee, sich von einem WakeMeBot wecken zu lassen. Daher ist der WakeMeBot für die Ansprache der Zielgruppe von Informatik beGreifen geeignet.

Für die benutzergerechte Realisierung [F3] des Roboters wurde die Konstruktion entsprechend der Idee des WakeMeBots gewählt. Dafür hat der WakeMeBot drei Rädern für maximale Beweglichkeit, einen Greifer für den Müsli-Riegel, einen Tastsensor zum Beenden

des Weckvorgangs, einen Ultraschallsensor zum Erkennen von Hindernissen, eine externe USB-Soundkarte zum Anschluss des Lautsprechers, bunt blinkende LEDs und eine bunte Seitenverkleidung. Für die einfache Programmierung des WakeMeBots durch Schülerinnen und Schüler wurde die Programmiersprache Scratch/BYOB verwendet. Um eine Kommunikation zwischen Scratch/BYOB und dem BrickPi zu realisieren, wurde ein Translator-Skript geschrieben. Durch dieses Translator-Skript kann über Befehle der WakeMeBot gesteuert werden. Für die schülergerechte Programmierung wurde Scratch/BYOB durch Funktionsblöcke erweitert, wodurch eine interne DSL entstand.

Um die benutzergerechte Programmierung des Roboters zu erklären und zu unterstützen [F4] wurde eine Schritt für Schritt Anleitung entworfen. Dadurch wird sichergestellt, dass Schülerinnen und Schüler den WakeMeBot erfolgreich programmieren. Am Standplatz unterstützte zudem der Standbetreuer, um Problemen bei der Programmierung entgegen zu wirken.

Informatik beGreifen kann im Rahmen des Studieninformationstags und dem Tag der offenen Tür [F5] durch eine gute Gestaltung des Standplatzes durchgeführt werden. Die Gestaltung sorgt dafür, dass dem Besucher ein positives Nutzererlebnis geboten wird. Dafür wurde anhand von Plakaten, Legobausteinen zum Spielen, einem Film und dem WakeMeBot selbst, Interesse bei den Besuchern erzeugt. Zudem wurde den Besuchern nach der Programmierung ein Müsli-Riegel und eine Urkunde ausgehändigt. Diese Dinge sollen als mitzunehmende Erinnerung das positive Gesamterlebnis des Schülers stärken.

Durch die Beantwortung der Forschungsfragen [F1] bis [F5] konnten alle gesetzten Ziele in dieser Arbeit erreicht werden. Dadurch ist es möglich, durch Informatik beGreifen die Informatik für Schülerinnen und Schüler sowohl greifbar, als auch begreifbar zu machen. Die summative Evaluation hat zudem gezeigt, dass durch Informatik beGreifen Ängste bei den Schülerinnen und Schülern genommen werden und diese sich selbst mehr in der Programmierung zutrauen. Den Vorurteilen der Schülerinnen und Schüler gegenüber der Informatik wird somit begegnet.

Weiterführende Arbeiten könnten Verbesserungen in der Oberfläche von Scratch/BYOB für eine bessere Usability vornehmen. Ungeklärt bleibt der Einfluss des Standbetreuers auf den Erfolg von Informatik beGreifen. Deshalb wäre es sinnvoll, diesen Einfluss in einer weiterführenden Arbeit zu untersuchen.

Für die Zukunft könnte Informatik beGreifen so erweitert werden, dass eine Durchführung auch ohne Standbetreuer möglich wäre. Dadurch könnte Informatik beGreifen auch für Unterrichtseinheiten an Schulen und bei anderen Veranstaltungen genutzt werden.

Literaturverzeichnis

- [bi14] bitkom: , Jung und vernetzt, Kinder und Jugendliche in der digitalen Gesellschaft, 2014. https://www.bitkom.org/Publikationen/2014/Studien/Jung-und-vernetzt-Kinder-und-Jugendliche-in-der-digitalen-Gesellschaft/BITKOM_Studie_Jung_und_vernetzt_2014.pdf.
- [Bu15] Bundesagentur für Arbeit Statistik und Arbeitsmarktbersichterstattung: , Der Arbeitsmarkt für IT-Fachleute in Deutschland, Mai 2015. <https://statistik.arbeitsagentur.de/Statischer-Content/Arbeitsmarktberichte/Akademiker/generische-Publikationen/Broschuere-Informatik-2014.pdf>.

- [d.] d.School: , Bootcamp bootleg. <http://dschool.stanford.edu/wp-content/uploads/2011/03/BootcampBootleg2010v2SLIM.pdf>.
- [De10] Deutsches Institut für Normen e.V.: , Ergonomie der Mensch-System-Interaktion - Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme (ISO 9241-210:2010), Oktober 2010.
- [Di15] Dicola, Tony: , NeoPixels on Raspberry Pi, Overview, December 2015. <https://learn.adafruit.com/neopixels-on-raspberry-pi/overview>.
- [Dr02] Druin, Allison, Hrsg. The role of children in the design of new technology. LNI 999. University of Maryland, <http://hci12.cs.umd.edu/trs/99-23/99-23.pdf>, 2002.
- [ELK08] Engeser, Stefan; Limbert, Nina; Kehr, Hugo: Studienwahl Informatik. Abschlussbericht zur Untersuchung. Technische Universität München, 2008.
- [Gia] GitHub: , Github repository of BrickPi_Python. https://github.com/DexterInd/BrickPi_Python.
- [Gib] GitHub: , Github repository of scratchpy. <https://github.com/pilliq/scratchpy>.
- [gi15] github: , Userspace Raspberry Pi PWM library for WS281X LEDs, December 2015. https://github.com/jgarff/rpi_ws281x/tree/rpi2.
- [Gü14] Gürtler, Jochen und Meyer, Johannes: , Design Thinking, 2014.
- [KLB04] Kontio, Jyrki; Lehtola, Laura; Bragge, Johanna: , Using the Focus Group Method in Software Engineering: Obtaining Practitioner and User Experiences, 2004.
- [Kr10] Krahn, Holger: , Monticore: Agile Entwicklung von domänenspezifischen Sprachen im Software-Engineering, 2010.
- [Kü14] Kühnapfel, J. B.: , Nutzwertanalysen in Marketing und Vertrieb, 2014.
- [MfW15] Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg: , Studieninformationstag 2015, December 2015. <http://www.studieninfotag.de/>.
- [Pa15] Parker, Dave: , Projects for NXT 1.0, December 2015. <http://www.nxtprograms.com/index1.html>.
- [Re15] Reutlingen, Hochschule: , Tag der offenen Tür, December 2015. <https://www.reutlingen-university.de/aktuelles/termine/detail/article/tag-der-offenen-tuer.html>.
- [Sa14] Sauvonnnet, Emmanuel und Blatt, Markus, Hrsg. Wo ist das Problem? Design Thinking als neues Management-Paradigma. Books on Demand, 2014.
- [Sc] Scratch: , Broadcast. <http://wiki.scratch.mit.edu/wiki/Broadcast>.
- [Sc99] Schlicksupp, Helmut: , 30 Minuten mehr Kreativität, 1999.
- [To11] Tonhauser, Pauline: , Business Model Box, der siebte Schritt von Design Thinking, August 2011.
- [TS08] Töpfer, Armin; Silbermann, Steffen: , Einsatz von Kunden-Fokusgruppen, DOI 10.1007/978-3-540-49924-4 2008.
- [WM13] Walker-Morgan, DJ.: , BrickPi: Raspberry Pi und Lego Mindstorms, vereinigt euch!, June 2013. <http://www.heise.de/make/meldung/BrickPi-Raspberry-Pi-und-Lego-Mindstorms-vereinigt-euch-1887544.html>.

Type Inference on Wikipedia List Pages

Patrick Kuhn, Sven Mischkewitz, Nico Ring and Fabian Windheuser¹

Abstract: The extraction of information from Wikipedia has led to a huge amount of knowledge made widely available by projects like the DBpedia². So far, most effort is put into extracting explicitly encoded information e.g. infoboxes. However, Wikipedia also contains a huge amount of implicit knowledge. One example for an untouched source of implicit knowledge are Wikipedia's *List of* pages, in which multiple entities with a common type are collected. If this common type is known, it can be added to all entities of the list. Moreover, entities which are part of this list but not yet presented in the DBpedia can be added. This offers a huge potential for extending the DBpedia by adding missing type information. This paper proposes an approach to extract the shared types of a list using statistical methods and natural language processing. For a list entity, it was possible to infer new types with a precision of 86%.

Keywords: Linked Data, Ontology Enrichment, DBpedia, Wikipedia

1 Introduction

In the year 2006 Tim Berners-Lee coined the term linked data, which describes a set of best practices on how to expose and connect data from different sources [BHBL09]. Since then the linked data movement has experienced a remarkable growth. Over the years multiple classical datasets have been integrated. One of the most important techniques for representing and linking entities is the *Resource Description Framework* (RDF) [GB14]. RDF uses subject-predicate-object triples to describe relationships between entities and make them available for automatic interpretation.

Maybe the most important source of information is Wikipedia³, one of the world's most visited websites. With around 5 million articles in the English version alone⁴, it is the largest encyclopedia available. The DBpedia project⁵ is one approach to extract the information from Wikipedia and make it accessible in an abstracted form. DBpedia is a community driven database, which aims at extracting structural information from Wikipedia and making it publicly available using the RDF model. Structural information includes "infobox templates, categorisation information, images, geo-coordinates, links to external web pages and links across different language editions of Wikipedia" [Au07]. Currently the English version of DBpedia contains around 3.6 million entities including around 763,000 persons, 572,000 places, and 192,000 organizations. On average an entity has four types associated with it [DB16a].

¹Hasso-Plattner-Institute, 14482 Potsdam, Germany, {patrick.kuhn, sven.mischkewitz, nico.ring, fabian.windheuser}@student.hpi.de

²<http://dbpedia.org/>

³<https://www.wikipedia.org/>

⁴https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia/, April 2016

⁵<http://wiki.dbpedia.org/>

The great majority of extraction has been focused on information, which is explicitly stated in Wikipedia e.g. infoboxes. However there is also information, which can be concluded by studying the relations and structures of resources (implicit knowledge). One example of a currently untouched source of information are Wikipedia's *List of pages*⁶. *List of pages* are collections of multiple entities grouped together under a shared set of types.

After the types describing a list have been identified they can be added to every member of the list. As there are more than 350 thousand *list of pages*⁷ in the English Wikipedia this approach offers a huge opportunity to enrich the DBpedia. This information is particularly important because "proper classification of entities into types is indispensable for any Information Extraction (IE) system" [Po12].

Table 1 shows four entities from the *list of German scientists*⁸. The row *parsed types* shows an excerpt of the associated DBpedia types. As *Franz Aepik* is not represented in the DBpedia, he does not have any type information. *Scientist*, *Person*, *Agent*, and *Thing* can be determined as the shared types for the list and added to every member. In this case the type *dbo:Scientist* can be added to the entity *Roland Benz*. Additionally it is possible to add missing entities such as *Franz Aepik* with all associated types to the DBpedia.

	Leonhard Euler	Carl Friedrich Gauss	Roland Benz	Franz Aepik
Parsed types	dbo:Scientist dbo:Person dbo:Agent owl:Thing	dbo:Scientist dbo:Person dbo:Agent owl:Thing	dbo:Person dbo:Agent owl:Thing	missing in DBpedia
Missing types			dbo:Scientist	dbo:Scientist dbo:Person dbo:Agent owl:Thing

Tab. 1: List of German scientists

This paper proposes an approach to automatically extract type information using the information in Wikipedia's *list of pages* combined with the information from DBpedia. This type information can be used to extend the DBpedia by adding missing types and even adding new entities.

This paper is structured as follows: Section 2 discusses previous work. The approach to extract common types is presented in Section 3. After the entities belonging to a list have been extracted from Wikipedia as described in Section 3.1, candidate types are proposed. These types are then ranked using statistical methods and natural language processing as described in Section 3.3. To evaluate the accuracy of the proposed types a test set has been created as described in Section 4. This test set is used to evaluate the approach as discussed in Section 5.

⁶ https://en.wikipedia.org/wiki/List_of_lists_of_lists/

⁷ Titles starting with "List.of" in the title dump: <http://wiki.dbpedia.org/Downloads2015-04#titles/>

⁸ https://en.wikipedia.org/wiki/List_of_German_scientists/

2 Related Work

The goal of this work is to increase the type coverage of DBpedia entities. In linked data, type information is represented by the *rdf:type* predicate. This predicate associates an entity with an ontology. In the case of linked data the term ontology most often refers to a set of classes, their properties, and relations between the class members. The ontology provided by DBpedia is a "shallow, cross-domain ontology" [DB16b], which covers 685 classes. The DBpedia ontology is pretty small and not all concepts can be expressed by only using classes from it. Therefore multiple other ontologies such as YAGO [SKW07] have been developed. In contrast to DBpedia, the YAGO ontology consists of more than 350,000 classes. These classes are derived from WordNet [Mi95] and Wikipedia categories. The problem with the YAGO ontology is that its "granularity is often too high" [Ga12]. A result of a larger ontology is that the overlap of types is smaller. Since our algorithm uses this overlap to infer new types, better results can be achieved using the DBpedia ontology.

In general, approaches can be divided into methods which work directly on the information of the Wikipedia page and methods which work on the information provided by linked data.

Tipalo [Nu13] falls into the former category and tries to extract defining statements (e.g. "XX is a YY") of an entity from the abstract of its corresponding Wikipedia page. The type information of these statements are extracted and matched against WordNet types. The approach reaches an overall recall of 74% with a precision of 76%. This proves that the abstract can contain useful type information. Because generating information from natural languages is always error prone, we only use the content of Wikipedia pages to further confirm already extracted types instead of generating new ones.

Giovanni et al. introduced a method [Nu12], which makes use of the links between Wikipedia pages to infer type information. The authors report a recall of 86% and a precision of 52%. However the approach is limited to only predict one of the nine top level classes, whereas our approach can be used to infer types from the complete DBpedia ontology.

One of the first paper which only uses linked data to extract new information is discussed by Neville et al. [NJ00]. They train a model on already labeled data and iteratively apply this model to find new types for unlabeled entities. This yields an accuracy of 82%.

Another approach which makes use of the relations of existing linked data sources is presented in [PB13]. The underlying idea of this approach is that each relation is connected to a specific type with a given probability. This can be used to infer new types using a weighted voting approach. This leads to F-measure of 88.5%. In contrast to these approaches which only rely on the linked data itself, our approach combines information from Wikipedia and linked data.

This paper extends the approach of Paulheim et al. [PP13], which proposes a two staged process to extract type information from Wikipedia's *list of* pages. In the first step candidate classes are identified based on statistical methods. Second, these classes are ranked

according to information gained by natural language processing frameworks. Paulheim et al. presents this method from a theoretical point of view and list a set of problems, such as fail-safe extraction and suitable scoring functions [PP13], which have to be solved in order to actually implement this process.

3 Extracting Common Types

Given all entities belonging to a *list of* page the main goal is to find a set of DBpedia types which describes the entire list. Figure 1 shows the stages of the approach. In order to achieve this, the first step is to parse and translate the Wikipedia article into an abstract representation. From this representation the actual list members and their types have to be determined. Once this is done the types are ranked according to their relevance for the list. From this ranking a set of fitting types is extracted.

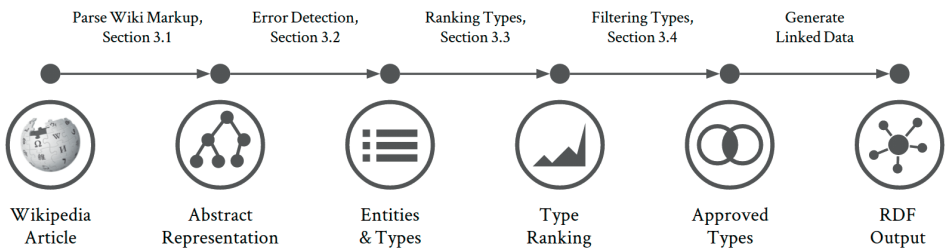


Fig. 1: The linear process from Wikipedia articles to mined data.

3.1 Parsing Wikipedia

Multiple solutions already exist for converting Wiki markup into abstract representations. A flexible solution is provided with the *json-wikipedia parser*⁹. The *json-wikipedia parser* utilises the *TU-Darmstadt parser*¹⁰ to convert Wiki markup text from a XML dump [Wi16a] into JSON¹¹. To satisfy all needs for the extraction of list pages, adaptations¹² have been made to *json-wikipedia* and the *TU-Darmstadt parser*. To filter unimportant sections like *External Links*, knowledge about the entity links in the respective section and paragraph on the Wikipedia page has to be preserved. Besides that, there are constructs in the Wiki markup syntax which are not recognized by the original *TU-Darmstadt parser* e.g. lists composed of multiple columns.

The analysis of 2000 *List_of* pages¹³ showed, that lists can be categorized into three major groups:

⁹ <https://github.com/Wikilist-Extraction/json-wikipedia/>

¹⁰ <http://mvnrepository.com/artifact/de.tudarmstadt.ukp.wikipedia/>

¹¹ <http://json.org>

¹² <https://github.com/Wikilist-Extraction/json-wikipedia/>

¹³ <http://windheuser.com/p/random2000.zip/>

- **Bullet point lists:** Lists that consist of a sequence of bullet points followed by the entity.
- **Table lists:** Lists containing entities represented in one column of a table.
- **Mixed lists:** Lists without a clear structure or with multiple different representations.

We concentrate on the extraction of knowledge from bullet point and table lists, since they exist more often in Wikipedia. Furthermore bullet point and table lists provide some kind of structure which makes extraction less error prone.

When dealing with bullet point lists the greatest difficulty in parsing arises from nested lists. Furthermore other edge cases have to be considered like multiple entities behind a bullet point.

When extracting table lists, it is often unclear which column is the main column and which are only supportive columns describing the main column. There are several factors affecting the probability of a column being the main one. Features for identifying the main column are its index and the ratio of unique entities. Moreover the list entities in the main column are often described by the other columns. Table 2 shows the list of NBA champions with the respective number of their titles and their division. The entities in the club column are described by the other columns using the *dbp:division* and *dbo:title* predicate. Many relations to another column is a clear indication, that this column contains the list entities. These relations can be extracted using the DBpedia.

This feature is converted to a numerical value by calculating the average number of connections from an entity to the other columns. Once all these features have been normalized between zero and one, the column with the highest weighted sum of these features is chosen as the main column. The weights have been determined using grid search.

Rank	Club	Division	Title
1	Boston Celtics	Atlantic	17
2	Los Angeles Lakers	Pacific	16
3	Chicago Bulls	Central	6
4	San Antonio Spurs	Southwest	5

Tab. 2: List of NBA-Champions¹⁴

3.2 Error detection

One central problem is, that not every concept, can be mapped to a DBpedia type. The *list of dance pop artists*¹⁵ is an example for a list without a direct counterpart in the DBpedia ontology. The types *dbo:MusicalArtist* and *dbo:Band* both describe the content of the list

¹⁴ https://de.wikipedia.org/wiki/National_Basketball_Association/

¹⁵ https://en.wikipedia.org/wiki/List_of_dance-pop_artists

pretty well, but are distinct on a very top level. Musical Artists is a subclass of Person, whereas Band is a subclass of Organisation.

Another problem is that different writing styles in combination with the loose structured syntax make it impossible to take all special cases into account. As a consequence of this a criteria to detect errors is needed. One clear indicator that the wrong entities have been extracted is a great variety of not closely related types in the result set. To detect this a notion to quantify semantic relatedness between two types is needed. For this, an adapted version of the *Leacock and Chodorow similarity* [LC98] is used. The original Leacock and Chodorow similarity is defined by

$$Sim(T1, T2) = -\log\left(\frac{distance(T1, T2)}{2 * depth}\right)$$

where *distance* is defined by the length of the shortest path between two types in the ontology tree and *depth* being the maximum depth of the ontology.

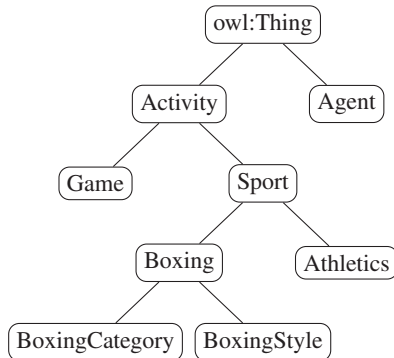


Fig. 2: DBpedia ontology

Figure 2 shows an extract of the DBpedia ontology tree. From this diagram it can be seen that the similarity between two terms is highly dependent on the level of a node. For example *dbo:BoxingCategory* and *dbo:BoxingStyle* are much closer related than *dbo:Activity* and *dbo:Agent*, even though the Leacock Chodorow similarity is the same. For this reason we use the sum of the depths instead of the taxonomy depth for normalization.

$$Sim(T1, T2) = -\log\left(\frac{distance(T1, T2)}{depth(T1) + depth(T2)}\right)$$

3.3 Ranking Types

The open question is how to select a set of type candidates relevant for the whole list. Therefore identification of types specific to the entire list is required. As proposed by Paulheim this can be achieved by statistical means and textual evidence [PP13].

Statistical means: Many resources are annotated with generic types like *owl:Thing*. If a type occurs with a much higher frequency in a list than in the DBpedia, it is likely that this type is significant for this list. For example having multiple entities with the type *dbo:Scientist* attached is far more expressive than the the type *owl:Thing*. This can be expressed with the TF-IDF statistic. TF-IDF indicates how relevant a single term is to a document in a corpus. In the scope of type properties of a DBpedia entity this means types should occur frequently in the generated type set (term frequency) but should be specific to the considered list as well (inverse document frequency) [PP13]. Term frequency in our case denotes the relative frequency that a type occurs in all list members. Inverse term frequency refers to how common the type is in the whole DBpedia. The TF-IDF weight is computed as follows:

$$tfidf(type) = tf(type) * idf(type)$$

$$tf(type) = \frac{countInList(type)}{\#entitiesInList}$$

$$idf(type) = \log\left(\frac{\#entitiesInDBpedia}{countInDBpedia(type)}\right)$$

countInList(type) is the number of entities in the list with the given type. *entitiesInList* states how many entites where found in the list. Analogously, *countInDBpedia(type)* and *entitiesInDBpedia* counts the entities and types in the entire DBpedia.

Textual evidence: The statistical measure is further supported by searching for textual appearances of the type in the Wikipedia articles. The list page and the wikipages of the respective entities are used to scan for type labels. Title, abstract, and categories are considered as they are the most accurate description of an entity. Abstracts in Wikipedia articles tend to describe more than the reason why an entity is a member of a specific list. Therefore matches in the title and categories are weighted stronger than matches in the abstract. To improve matching for word variations in the article, the Porter Stemming algorithm [Po80] is used. Each match contributes to the textual evidence score.

Statistical means and textual evidence produce two independent rankings of types. In order to compare the two rankings, they have to be normalized to a scale between zero and one. A combined ranking is obtained by multiplying both scores for each type in the list. A weighting of 85:15 between statistical means and textual evidence is used, based on empirical results.

3.4 Filtering Types

To produce a set of fitting type candidates the ranked list of types has to be filtered according to the computed score. Cutting off at a fixed value does not always fit the underlying data. In the manual analysis of a test set of 2000 lists, we observed that quite often the transition between fitting and non-fitting types is marked by a significant score drop. To

take this into account, the last value before the score drop gets determined as a threshold. This divides the ordered list of types into two cluster with the approved and declined types. This approach adapts more flexible to the respective lists.

3.5 Architecture of the Implementation

The previously described approach has been implemented in Scala and Java and is publicly available on Github¹⁶. Apache Jena, a Java framework¹⁷ for building Semantic Web and Linked Data applications, is used to connect the application to RDF datasources. To process massive data asynchronously, Reactive Streams¹⁸ are used.

Querying the RDF datasets is the slowest task of the application. Jena TDB¹⁹ is used to minimize the impact of data access to the overall performance. Jena TDB is a native high performance triple store, which supports the full range of Jena APIs. It is possible to use TDB with a multiple reader or single writer policy for concurrent access [Fo16]. This enables parallel queries to the database. Thus slow queries can be compensated by concurrency. Additionally TDB is accessed asynchronously to avoid blocking database queries. Therefor an asynchronous wrapper for Jena SPARQL queries based on Scala Futures²⁰ was build. The wrapper can be used with SPARQL endpoints, Linked Data Fragments²¹ endpoints, and TDB. By preventing busy waiting the pipeline can process 2000 list pages in about 85 seconds²².

4 Evaluation

The quality of the results is normally evaluated using existing type information by checking whether this information can be reproduced by the system. This was not suitable for this approach, because the aim of this work is to infer new type information and in most cases only top level types are present in DBpedia. As a result a ground truth dataset had to be developed. A total of 400 randomly chosen lists have been annotated by four different persons. For each list all types of the parsed entities have been annotated as approved or declined. In addition lists could be tagged as incorrectly parsed. The evaluation set is publicly available [Wi16b].

There are multiple cases in which it is hard to decide if a type is appropriate for a certain list. For example one could argue that it would be correct to label Shakespeare as an artist, although in the DBpedia class hierarchy *dbo:Writer* is not a subclass of *dbo:Artist*. In this cases agreements have been made between the annotators.

¹⁶ <https://github.com/Wikilist-Extraction/wikipedia-list-extraction>

¹⁷ <https://jena.apache.org/index.html>

¹⁸ <http://doc.akka.io/docs/akka-stream-and-http-experimental/current/scala.html>

¹⁹ <https://jena.apache.org/documentation/tdb>

²⁰ <http://docs.scala-lang.org/overviews/core/futures.html>

²¹ <http://linkeddatafragments.org/>

²² Run on a 2 GHz Intel Core i7 Processor and 8 GB 1600 MHz DDR3 Memory

5 Results and Discussion

As a result of the evaluation a precision of 86.19% with a recall of 33.13% has been achieved (F1 score of 0.48). A total number of 60,786 new RDF triples were produced based on the 400 annotated list pages.

Since our approach concentrates on finding the most specific types for a list, more generic types are left out. The recall decreases since these generic types are still correct for most lists. Since DBpedia types are arranged in a hierarchy one could also propose all types of a higher level. This can be done in a post-processing step by also adding all super types and would increase the recall.

The main limitation of the algorithm is still fail safe parsing. When regarding the performance of the algorithm on correctly parsed lists, the precision of our algorithm increases to 94.97%.

Difficulties occur, when the list entities do not contain the shared types. One example for such a list is the *list of German expressions in English*²³. In this case the common type for the list is *German expressions*. The list entities however have a wide variety of types such as *dbo:Food*, *dbo:MusicalGenre* or *dbo:Weapon*. Since the type *German* or *expression* does not occur it can not be proposed. This reduces the recall, since these lists are filtered out using the error detection mechanism as described in section 3.2.

6 Conclusion and Outlook

In this paper we proposed an approach to extract new type information by combining linked data and natural language processing techniques. On a randomly chosen sample of 2000 randomly chosen lists a total number of 303,934 new type triples have been computed in 85.56 seconds computation time²⁴.

As solutions to the problems discussed in [PP13] this paper described a reliable method to extract the correct entities from wikipedia list pages by improving the parsing of Wiki markup and introducing a quality measure for the parsed entities. For specifying the correct types, the weighted TF-IDF and textual evidence scores are combined. Moreover, a method for determining the shared types from the obtained results was developed. As the evaluation has shown, this method is able to produce high quality type information (86.19% precision), while still having a feasible run time.

While the focus of this paper is adding DBpedia types to entities, there are a lot of cases where the common type of a list is not contained in DBpedia. For example, there is a list of hills in the Scottish lowlands over 2000 feet, which are called Donalds²⁵. To get more specific types for these lists, it would be possible to generate a new type. In this case the type *dbo:Donald* could be proposed. This could be achieved using the list name and an

²³ https://en.wikipedia.org/wiki/List_of_German_expressions_in_English/

²⁴ Run on a 2 GHz Intel Core i7 Processor and 8 GB 1600 MHz DDR3 Memory

²⁵ https://en.wikipedia.org/wiki/List_of_Donalds/

advanced textual evidence approach with entity recognition based on the abstracts of the entities.

Another way to extend our approach is to not only extract the type of a list but to also extract its describing axioms. At the moment only the type *scientist* is extracted for the *list of german scientist*. The describing property *German* is not inferred. This could be achieved by extending the statistical ranking proposed in this paper to also include the properties of the entities.

7 Acknowledgment

This work was created based on the seminar *Knowledge Mining* at the Hasso-Plattner-Institute. The authors thank Harald Sack, Jörg Waitelonis and Magnus Knuth for their support.

References

- [Au07] Auer, Sören; Bizer, Christian; Kobilarov, Georgi; Lehmann, Jens; Cyganiak, Richard; Ives, Zachary: DBpedia: A Nucleus for a Web of Open Data. In: In Proceedings of the 6th International Semantic Web Conference. Springer, 2007.
- [BHBL09] Bizer, Christian; Heath, Tom; Berners-Lee, Tim: Linked data-the story so far. *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, pp. 205–227, 2009.
- [DB16a] DBpedia Data Set Statistics, <http://wiki.dbpedia.org/services-resources/datasets/data-set-38/data-set-statistics/>.
- [DB16b] DBpedia Ontology, <http://wiki.dbpedia.org/services-resources/ontology/>.
- [Fo16] TDB Java API - Concurrency, https://jena.apache.org/documentation/tdb/java_api.html#concurrency/.
- [Ga12] Gangemi, Aldo; Nuzzolese, Andrea Giovanni; Presutti, Valentina; Draicchio, Francesco; Musetti, Alberto; Ciancarini, Paolo: Automatic Typing of DBpedia Entities. In: *Proceedings of the 11th International Conference on The Semantic Web - Volume Part I. ISWC'12*, Springer-Verlag, Berlin, Heidelberg, pp. 65–81, 2012.
- [GB14] Guha, Ramanathan; Brickley, Dan: RDF Schema 1.1. W3C recommendation, W3C, February 2014. <http://www.w3.org/TR/2014/REC-rdf-schema-20140225/>.
- [LC98] Leacock, Claudia; Chodorow, Martin: Combining local context and WordNet similarity for word sense identification. *WordNet: An electronic lexical database*, 49(2):265–283, 1998.
- [Mi95] Miller, George A: WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41, 1995.
- [NJ00] Neville, Jennifer; Jensen, David: Iterative classification in relational data. In: *Proc. AAAI-2000 Workshop on Learning Statistical Models from Relational Data*. pp. 13–20, 2000.

-
- [Nu12] Nuzzolese, Andrea Giovanni; Gangemi, Aldo; Presutti, Valentina; Ciancarini, Paolo: Type inference through the analysis of Wikipedia links. In: LDOW. 2012.
- [Nu13] Nuzzolese, Andrea Giovanni; Gangemi, Aldo; Presutti, Valentina; Draicchio, Francesco; Musetti, Alberto; Ciancarini, Paolo: Tipalo: A Tool for Automatic Typing of DBpedia Entities. In: The Semantic Web: ESWC 2013 Satellite Events, pp. 253–257. Springer, 2013.
- [PB13] Paulheim, Heiko; Bizer, Christian: Type inference on noisy rdf data. In: The Semantic Web–ISWC 2013, pp. 510–525. Springer, 2013.
- [Po80] Porter, Martin F: An algorithm for suffix stripping. *Program*, 14(3):130–137, 1980.
- [Po12] Pohl, Aleksander: Classifying the Wikipedia articles into the OpenCyc taxonomy. In: Proceedings of the Web of Linked Entities Workshop in conjunction with the 11th International Semantic Web Conference. volume 5, p. 16, 2012.
- [PP13] Paulheim, Heiko; Ponzetto, Simone Paolo: Extending DBpedia with Wikipedia List Pages. In: NLP-DBPEDIA@ISWC'13. pp. –1–1, 2013.
- [SKW07] Suchanek, Fabian M; Kasneci, Gjergji; Weikum, Gerhard: Yago: a core of semantic knowledge. In: Proceedings of the 16th international conference on World Wide Web. ACM, pp. 697–706, 2007.
- [Wi16a] Wikipedia: Database Download, https://en.wikipedia.org/wiki/Wikipedia:Database_download/.
- [Wi16b] Evaluated List Types, <https://github.com/Wikilist-Extraction/annotated-lists/>.

new/s/leak - Anforderungsanalyse einer interaktiven Visualisierung für Data-Driven Journalism

Franziska Lehmann¹

Abstract: Für Datenjournalisten² wird es immer schwieriger und zeitaufwändiger, die steigende Menge an Dokumenten zu analysieren und die für sie relevanten Informationen zu extrahieren. Mit Hilfe von Natural Language Processing können die in den Dokumenten vorkommenden Personen, Organisationen und Orte, sogenannte Entitäten, sowie deren Verbindungen identifiziert werden. Die Exploration der extrahierten Entitäten wird durch interaktive Visualisierungen unterstützt. Diese Idee wird bereits durch das Tool "Netzwerk des Tages" umgesetzt. Allerdings benötigt das Programm wesentliche Verbesserungen, um die journalistischen Bedürfnisse zu erfüllen. Auf dieser Basis erfolgte im Rahmen eines Projektes für das Tool "new/s/leak" eine Anforderungsanalyse mit Experten eines großen, deutschen Verlagshauses. Dieser Beitrag behandelt die in den Interviews identifizierten wesentlichen Anforderungen.

Keywords: Data-driven Journalism, Anforderungsanalyse, Informationsvisualisierung, Visual Analytics, Experteninterviews

1 Einleitung

"Now that we're deep into the information age, it's time for everyone to accept that the amount of information in our lives is only going to keep growing" (Mark Briggs [Br09])

Im Zeitalter der Globalisierung und Digitalisierung entstehen immer größer werdende Datenbestände. Das Ziel von investigativen Journalisten ist die Identifizierung bisher unbekannter, bedeutender Muster und Beziehungen innerhalb dieser Datensammlungen sowie deren Veröffentlichung [Lu07]. Die enormen Datensammlungen stellen hierbei eine große Herausforderung dar. Denn neben strukturierten Daten beinhalten die Dokumentensammlungen normalerweise vor allem unstrukturierte Textdokumente, wie im Falle von Wikileaks [Gy14] oder der Panama Papers. Die Exploration der Daten sowie die Identifizierung wichtiger Dokumente ist schwierig und zeitaufwändig. Dies ist insbesondere der Fall, wenn Journalisten im Voraus nicht wissen, wo sie ihre Recherche beginnen sollen, was in den Dokumenten wichtig sein könnte oder wie Ereignisse im Zusammenhang stehen [Br14, Gö13]. Zudem müssen die Texte aufgrund des Wettbewerbs zwischen Medienhäusern unter Zeitdruck analysiert werden [UK15]. Somit steigt der Bedarf an explorativen Datenanalysetools zur Unterstützung der journalistischen Recherche.

¹ Technische Universität Darmstadt, f.lehmann@stud.tu-darmstadt.de

² In dieser Arbeit wurde immer die männliche Form von Personen benutzt, damit der Lesefluss nicht gestört wird. Die Personen umfassen dementsprechend auch weibliche Personen.

Die Erstellung eines journalistischen Artikels basiert auf der Beantwortung der sog. fünf W-Fragen der journalistischen Recherche: 'Wer?', 'Was?', 'Wo?', 'Wann?' und 'Warum?' [Zh13]. Das Untersuchen und Entdecken der in den Dokumenten enthaltenen Themen, Inhalte sowie das Verständnis von Verbindungen und Beziehungen zwischen den darin genannten Personen, Organisationen und Orten (Entitäten) ist somit ein wichtiger Bestandteil des investigativen Prozesses [Gö13]. Allerdings kann dies aufgrund der meist enormen Menge an Dokumenten ohne computergestützte Analyse zeitaufwändig sein [Br14, Gö13].

Zur Lösung dieser Herausforderung können zum einen auf Basis einer **computergestützten Analyse** mit Hilfe von Data-Mining Methoden Modelle zur Charakterisierung der Daten berechnet werden, zum anderen können Journalisten mittels **Informationsvisualisierung** direkt mit der visuellen Schnittstelle interagieren und den Datensatz analysieren [Gö13, Su13]. Die Kombination und Interaktion visueller und automatischer Analyse ist Gegenstand von Visual Analytics, einem interdisziplinären Ansatz, der die Vorteile aus den beiden Forschungsgebieten verbindet [Su13]. Neben den unzähligen Tools zur Analyse von strukturierten Daten [Jä15], existieren nur wenige Tools für die Analyse von unstrukturierten Daten, die Entitätenextraktion mit visuellen Interaktionsmöglichkeiten bieten [Gö13, Su16]. Diese weisen allerdings verschiedene Nachteile auf (siehe hierzu Abschnitt 2.2), repräsentieren bzw. beantworten die fünf W-Fragen unzureichend und funktionieren bisher nur mit englischsprachigen Texten.

An der Schnittstelle von visueller und automatischer Analyse setzt das Projekt "Data Extraction and Interactive Visualization of unexplored Textual Datasets for Investigative Data-Driven Journalistic (DIVID-DJ)" an. Dies ist ein interdisziplinäres Projekt des Fachgebiets Graphisch-Interaktive Systeme mit dem Fachgebiet Sprachtechnologie der TU Darmstadt und wird in Zusammenarbeit mit einem großen, deutschen Verlagshaus durchgeführt. Das Ziel des genannten Projektes ist die Entwicklung eines Programms zur Unterstützung von Datenjournalisten. Das Tool "network of searchable leaks" (new/s/leak) soll dem Journalisten als Hilfsmittel dienen, um in zeitkritischen Situationen tiefe Einblicke in neu gewonnene Textdokumente zu erhalten.

Die Basis des Tools bildet das System "Netzwerk des Tages" [Fa14, KLB14], das aus einem vorigen Projekt der genannten Fachgebiete resultierte. Mit diesem Tool ist schon jetzt eine Exploration von Entitätenbeziehungen möglich. Ein beispielhafter Netzwerkausschnitt ist in Abb. 1 dargestellt.

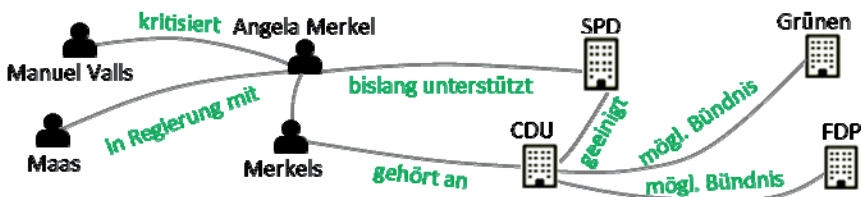


Abb. 1: Ausschnitt eines Entitätennetzwerks aus dem "Netzwerk des Tages"

Bereits der Projektantrag enthielt Darstellungselemente, sogenannte Views, die für das Tool new/s/leak geplant sind, wie bspw. einer Zeitleiste, einer aggregierten Sicht der Entitäten sowie einer Dokumentenansicht. Damit das Programm entsprechend der journalistischen Bedürfnisse entwickelt werden kann, müssen zum einen die geplanten Views geprüft und konkretisiert, sowie weitere Anforderungen innerhalb einer Anforderungsanalyse identifiziert werden. Der vorliegende Beitrag greift daher die weiter oben aufgezeigte Forschungslücke auf und präsentiert die Analyse und die identifizierten Anforderungen bzgl. der interaktiven Visualisierung des Tools.

Der Beitrag ist wie folgt gegliedert: In Abschnitt 2 werden die relevanten journalistischen Begrifflichkeiten definiert sowie die in der Literatur identifizierten Visual Analytics Tools in Bezug auf die vorliegende Problemstellung vorgestellt. Abschnitt 3 enthält die Beschreibung der Vorgehensweise der Anforderungsanalyse. In Abschnitt 4 erfolgt die Beschreibung der wichtigsten Anforderungen, extrahiert aus den Interviewergebnissen. Der Beitrag endet mit einem Fazit und Ausblick in Abschnitt 5.

2 Journalistischer Kontext und verfügbare Tools

Dieser Abschnitt behandelt die relevante journalistische Terminologie sowie relevante Tools aus dem Bereich Visual Analytics.

2.1 Datenjournalismus

Datenjournalismus wird als die Kombination eines Recherche-Ansatzes und einer Veröffentlichungsform herausgestellt, indem maschinenlesbare Datensätze durch die Verwendung von Software miteinander verschränkt und analysiert werden [Ma10]. Die Analyse der Daten erfolgt anhand des experimentellen Einsatzes von Algorithmen, Daten und sozialwissenschaftlichen Methoden [Au15]. "[Damit] wird ein schlüssiger und vorher nicht ersichtlicher informativer Mehrwert gewonnen. Diese Information wird in statischen oder interaktiven Visualisierungen angeboten und mit Erläuterungen zum Kontext, Angaben zur Datenquelle (bestenfalls wird der Datensatz mit veröffentlicht) versehen." [Ma10] Hierbei deckt Datenjournalismus den journalistischen Arbeitsprozess, von der Datensammlung, Datenanalyse und Filterung, Visualisierung sowie Berichterstattung, ab und ermöglicht dem Leser den Einstieg in große Datensätze [Au15].

2.2 Verfügbare Tools

In der Literatur zu Visual Analytics wird eine Vielzahl an Tools beschrieben, die sich auf einen bestimmten Aspekt der in diesem Beitrag behandelten Problemstellung fokussieren.

Die Tools **PaperLens** [Le05] und **NetLens** [Ka06] unterstützen die Filterung der Dokumente nach Metadaten in Form von Balkendiagrammen, Listen, Grafiken und textbasierten Visualisierungen zur Darstellung von Autor, Thema von Dokumenten sowie Zitationsdaten. Mit Hilfe der Identifikation der Themenbereiche erfolgt ebenso eine Gruppierung der Dokumente. Aufgrund der fehlenden Entitätenerkennung sowie deren Verbindungen müssten Journalisten immer noch einen Großteil der Dokumente lesen.

Die folgende Gruppe von Tools fokussiert sich insbesondere auf die Darstellung inhaltlicher Themengebiete der vorliegenden Dokumentsammlung. Während **Overview** [Br14] lediglich den Inhalt der Dokumente als eine Menge von Schlüsselworten visualisiert, stellen **ThemeRiver** [Ha00] sowie **Parallel Topics** [Do11] zusätzlich die zeitliche Veränderung von Themen, extrahiert aus der Dokumentensammlung, dar. Diese Tools bieten allerdings keine Entitätenerkennung an, weshalb sie angesichts des Mangels an intuitiver und präziser Visualisierung zur Untersuchung der vorliegenden Problemstellung ungeeignet scheinen.

Die Erkennung und Visualisierung von Entitäten wird u.a. durch die nachfolgenden ausgewählten Tools unterstützt. **Open Calais** [Ga13] unterstützt die Zusammenfassungen (Aggregationen) von Entitäten und markiert sie im Text. **BiSet** [Su16] und **Jigsaw** [Gö13] visualisieren die extrahierten Entitäten in einer Listenansicht und stellen Beziehungen zwischen Entitäten mit Verbindungslinien dar. Hingegen visualisiert das **Netzwerk des Tages** [Fa14, KLB14] die extrahierten Entitäten als Node-Link-Diagramm. Diese Tools bieten entweder keine Zeitleiste als Visualisierungsmöglichkeit an oder erfassen nur Zeitpunkte, benötigen einen zu hohen Schulungsaufwand bzgl. der Nutzung der Tools und/oder beinhalten keine bzw. geringe Filtermöglichkeiten. In der nachfolgenden Tabelle sind die möglichen Nachteile der Tools aufgeführt.

Tool	Mögliche Nachteile
PaperLens/NetLens	keine Entitätenerkennung, großer Leseaufwand
Overview	keine Entitätenerkennung, fehlende Zeitleiste
ThemeRiver/Parallel Topics	keine Entitätenerkennung sowie Verbindung derer
Open Calais	fehlende Zeitleiste, Schulungsaufwand
BiSet	fehlende Zeitleiste
Jigsaw	fehlende Filtermöglichkeiten
Netzwerk des Tages	Anzeige Netzwerk nur für einen Zeitpunkt

Tab. 1: Zusammenfassung möglicher Nachteile der Tools

3 Vorgehen der Anforderungsanalyse

Um eine möglichst vollständige Liste der Anforderungen an das Tool "new/s/leak" zu

erhalten, wird die Designstudienmethodik (engl. Design Study Methodology) als Grundgerüst für ein strukturelles Vorgehen zur Beantwortung der Forschungsfrage genutzt. Eine Designstudie ist ein Projekt, in dem ein Forscher ein reales Problem, dem ein Domänenexperte gegenübersteht, analysiert, zur Problemlösung ein Visualisierungssystem entwirft, diesen Entwurf validiert und die gewonnenen Erkenntnisse zur Verfeinerung der Designrichtlinien reflektiert und wird durch einen 9-stufigen Rahmen praktisch angeleitet [SMM12]. Dieser Beitrag konzentriert sich auf die Stufe 'Erkenntnisphase'.

Die '**Erkenntnisphase**' schließt die Anforderungsanalyse ein [SMM12]. Das wichtigste Hilfsmittel von Anforderungsanalysten zur Anforderungsermittlung sind Interviews mit Nutzern und Stakeholdern. Das Ziel des Gesprächs ist die Erkennung der spezifischen Bedürfnisse von Nutzern an das Tool. Dies umfasst ebenso zu verstehen, welche Bedingungen erfüllt sein müssen, um das Tool für die tägliche Arbeit zu nutzen [SG06].

Die Analyse der Anforderungen erfolgt anhand nichtstandardisierter Leitfadeninterviews mit Experten [MN09]. Leitfadeninterviews enthalten vorgegebene Fragen, deren Reihenfolge und Formulierung unverbindlich gestaltet ist. Hierdurch können mit einem möglichst natürlichen Gesprächsverlauf gezielt Themen angesprochen und neue Gesichtspunkte identifiziert werden [GL10]. Bei Experteninterviews steht zudem nicht der Befragte als Person, sondern seine Erfahrungen und Interpretationen bzgl. des entsprechenden Forschungsthemas im Vordergrund [BG09]. Demnach verfügen Experten über Spezialwissen, das nicht Jedem in dem interessierenden Handlungsfeld zugänglich ist [MN09].

Als Grundlage der Anforderungsanalyse existiert bereits ein geplanter, grober Rahmen, anhand dessen die Anforderungen abgeleitet werden können. Der Rahmen beinhaltet verschiedene Sichten (Views) der Dokumente: eine Übersicht, eine Zeitleiste, ein Netzwerk sowie eine Dokumentenansicht. Durch die Interviews werden die Views überprüft, konkretisiert und entsprechende Anforderungen abgeleitet. Um die Anforderungsanalyse zu erleichtern und Gedanken zu konkretisieren wird zudem die Methode Sketching der Five Design Sheet (FDS) Methodik verwendet [RHR16]. Die FdS Methodik entspricht einer Komplettlösung, die divergentes Denken fördert, auf dem Skizzieren basiert und den Fokus auf das Ziel des zu entwerfenden Tools, seine Operationen und Interaktionen lenkt.

Zur Erstellung der Fragen des Interviewleitfadens wurden Informationen aus dem Projektantrag sowie Informationen über artverwandte Tools aus der Literatur genutzt. Der Leitfaden enthält zwei größere Abschnitte. Der erste Teil beinhaltet Fragen zur derzeitigen journalistischen Arbeitsweise. Der zweite Abschnitt enthält Fragen, welche direkt das Tool betreffen. Der grundsätzliche Aufbau des zweiten Teils orientiert sich an den Darstellungselementen (Views), die im Projektantrag genannt wurden: Übersicht, Netzwerk, Suche, Zeitinformationen/Zeitleiste, Dokumentenansicht, Annotationen. Hinzu kamen die Kategorien Geo-Visualisierung, Vergleiche sowie Anzeigedauer. Die Interviews fanden mit fünf Datenjournalisten in drei separaten Interviews statt.

Die Datenanalyse und Auswertung orientiert sich an dem von Gläser und Laudel (2010) vorgestellten Verfahren, der qualitativen Inhaltsanalyse. Hierbei erfolgt eine Extraktion von Informationen aus den Interviews hinsichtlich eines systematischen Schemas. Dadurch wird eine frühzeitige Separation vom Ausgangstext erreicht, die nur Informationen enthält, welche für die Anforderungsanalyse von Interesse sind [GL10].

4 Ergebnisse: relevante Anforderungen

Im folgenden Abschnitt erfolgt die Vorstellung der relevanten Anforderungen, die sich aus der Anforderungsanalyse ergeben haben. In Tabelle 2 sind die wesentlichen Anforderungen aufgeführt. Die letzte Spalte der Tabelle enthält die Nennung der Anforderungen innerhalb den Interviews (eine Nennung pro Interviewteilnehmer möglich). Die erste Zahl beschreibt die Anzahl der Nennungen und die zweite Zahl, wie viele Interviewteilnehmer, bedingt durch den Leitfaden, die Anforderung genannt haben. Die Beschreibung der Anforderungen findet nach der Tabelle statt.

ID	Anforderung	Kurzbeschreibung	Ø
A1	Suche	Funktion, um bestimmte Begriffe in den Dokumenten zu finden	5/5
A2	Tagging	Hinzufügen von Metainformationen an Dokumente	5/5
A3	Filterung	Filterung der Dokumente nach bspw. Personen, Schlüsselwörtern oder Orten	4/4
A4	Interessantheitseinstellung	Einstellbarkeit der angezeigten Entitäten nach der Häufigkeit	4/4
A5	Interaktion zwischen Views	Durch Veränderung von Informationen einer View verändern sich ebenso die angezeigten Informationen anderer Views	4/4
A6	Export Dokumente	Export von interessanten Dokumenten als bspw. Archivdatei	4/4
A7	Annotationen	Funktion, um im Netzwerk/in Dokumenten Kommentare und Markierungen zu machen	4/4
A8	Bearbeitungsoptionen der Netzwerkkomponenten	Manuelle Beseitigung von Fehlern im Netzwerk durch Zusammenfassung, Löschung, Bearbeitung und Hinzufügen von Entitäten	1/1
A9	Übersichtsansicht	Anzeige der Metadaten der Dokumente, häufigste Entitäten und Schlüsselworte	5/5
A10	Netzwerk	Anzeige von Entitäten, deren Verbindungslinien und -kontexte & Netzwerkkenzahlen	3/5
A11	Ausschaltbare Netzwerkelemente	Ausblendbare Informationen innerhalb des Netzwerkes	4/4
A12	Zeitleiste	Anzeige von Dokumentenhäufigkeit,	4/4

		Entitäten, Beziehungen, Schlüsselwörter über Zeit	
A13	Dokumentenanzeige	Direkter Zugriff auf Dokumente durch Anzeige von Überschrift plus Textausschnitt, Metadaten und Schlüsselwörter der Dokumente	5/5
A14	Karte	Visualisierung von Orten auf der Karte sowie Karte als Filterungsmöglichkeit nach Orten	4/4
A15	Not getting lost	Gestaltung der interaktiven Visualisierung, dass Journalisten nicht den Überblick verlieren	3/3

Tab. 2: Übersicht der wichtigsten Anforderungen

Suche: Die Suche ist mitunter eine der wichtigsten Anforderungen an das Tool und wurde innerhalb der Interviews von jedem Teilnehmer angesprochen. Die Suche nach Namen, Ländern oder Schlüsselwörtern ist eine gängige Vorgehensweise bei der Recherche innerhalb einer großen Menge an Dokumenten, wie bspw. im Falle der Wikileaks. Die durch die Suche getätigte Filterung ermöglicht einen wertvollen Informationsgewinn, z.B. wie sich Häufigkeitsdiagramme von Entitäten oder die Zeitleiste ändern. Informationen, die zusätzlich angezeigt werden sollten, sind u.a. die Häufigkeit des Suchbegriffs (verbleibende Anzahl der Dokumente nach Filterung) oder wie oft der Suchbegriff in Verbindung mit einem Namen in den Dokumenten vorkommt.

Tagging: Auch der Bedarf einer Tagging-Funktion konnte aus den Interviews (in-)direkt abgeleitet werden. Mit Hilfe der Tagging-Funktion können Journalisten Metainformationen an Dokumente anhängen, um diese bspw. in unterschiedliche Themenbereiche einzusortieren. Weiterhin ist damit die Möglichkeit verbunden, Dokumente oder relevante Teile des Netzwerkes mit einem Kollegen zu teilen. Da diese Funktion auch genutzt werden soll, um interessante Dokumente mit anderen Kollegen teilen zu können, ist es von Vorteil Notizen an die Tagging-Gruppe anhängen zu können.

Filterung: Eine der wichtigsten Anforderungen, die aus allen Interviews hervorgeht, betrifft die Filterung der Dokumente nach unterschiedlichen Kriterien. Die Visualisierung wird als ein einfach zu bedienender Filter für die Suche nach berichtenswerten Dokumenten benötigt. Durch die Anwendung von Filtern kann die Menge an Dokumenten reduziert werden. Dabei ist es vor allem wesentlich, Wichtiges von Unwichtigem zu trennen und einen Überblick zu bekommen, was wirklich relevant ist. Im Verlauf der einzelnen Interviews konnten mehrere Kriterien identifiziert werden, nach denen eine Filterung der Dokumente möglich sein sollte. Nachfolgend ist eine Auswahl der identifizierten Kriterien aufgeführt: aus den Dokumenten extrahierten Personen, Organisationen und Orte (Entitäten); Schlüsselwörter, die besonders oft in den Dokumenten vorkommen; Zeitspanne/Zeitpunkt; Dateiformate; Dateigröße sowie eine durch den Journalisten vergebene Metainformation (Tag).

Interessantheitseinstellung: Was für Journalisten innerhalb einer Dokumentensammlung interessant ist, fällt je nach Dokumentensammlung und welche Fragen Journalisten beantwortet haben wollen, unterschiedlich aus. Dementsprechend

gibt es Fälle, in denen häufig auftretende Entitäten interessant sein können sowie Fälle, in denen selten auftretende Entitäten interessant sind. Um die unterschiedlichen Betrachtungsweisen zu unterstützen, ist die Einstellbarkeit der angezeigten Entitäten nach der Häufigkeit wichtig.

Interaktionen zwischen Views: Ebenso wurde die Interaktion zwischen Views in den Interviews mehrfach genannt. Die Interaktion zwischen den Views meint hierbei, dass bspw. durch das Setzen einer Filterung sich das Netzwerk ändert. Ein weiteres Beispiel ist, dass beim Anwählen eines Knoten innerhalb des Netzwerkes zusätzlich die Zeitleiste des Knotens angezeigt werden soll. Dementsprechend wird die Interaktion der Filter mit den Views und den darin enthaltenen Elementen benötigt. Dies umschließt u.a. die Zeitleiste, die angezeigten Häufigkeitsdiagramme, das Netzwerk, die Schlüsselwörter und weitere Informationen, die für die getätigte Filterung bestehen.

Export Dokumente: Bei der Zusammenarbeit von Journalisten mit anderen Fachkollegen, werden relevante Dokumente untereinander ausgetauscht. Aufgrund dieser Gegebenheit kam der Wunsch auf, Dokumente exportieren zu können. Daher müssen Dokumente im ersten Schritt auswählbar gemacht werden, um diese im zweiten Schritt bspw. als Archiv exportieren zu können.

Annotationen: Damit Journalisten Verknüpfungen zwischen Dokumenten herstellen, bestimmte Passagen im Dokument optisch hervorheben (highlighten) oder einzelne Textpassagen mit ausführlichen Kommentaren versehen zu können werden Annotationsmöglichkeiten benötigt. Hierbei sollen Annotationen nicht nur in Dokumenten, sondern auch im Netzwerk möglich sein.

Bearbeiten von Netzwerkkomponenten: Die automatische Extraktion von Entitäten inkl. deren Verbindung mit Hilfe von Natural Language Processing ist mit Unsicherheit behaftet und somit fehleranfällig. Somit können Entitäten im Dokument nicht erkannt werden, der Typ einer Entität wird falsch erkannt oder unterschiedlich formulierte Entitäten, die jedoch die gleiche Bedeutung besitzen, werden nicht als eine einzige Entität extrahiert. Zur Beseitigung dieser Fehler sind daher verschiedene Bearbeitungsfunktionen notwendig, wie die Zusammenfassung zweier Entitäten mit der gleichen Bedeutung, das Löschen von Entitäten, das Hinzufügen von Entitäten zum Netzwerk oder der Bearbeitung einer Entität (Beschriftung sowie Metadaten wie bspw. der Typ einer Entität).

Informationen zur Übersicht: Damit der Journalist einen Einblick bekommt, welche Themen in Dokumentensammlungen enthalten sind, welche Personen vorkommen und welche Zeitspanne die Dokumentensammlung beinhaltet, werden zunächst die wichtigsten Informationen aus den Dokumenten benötigt. Die Anzeige von Häufigkeitsdiagrammen von bspw. Personen, Organisationen, Orten, Schlüsselwörtern ist somit eine zentrale Anforderung an das Tool. Als Visualisierung der häufigsten Entitäten werden schlichte Balkendiagramme präferiert. Zudem stellte sich die Anzeige der Metadaten der Dokumente innerhalb aller Interviews als bedeutende Information über den Datensatz heraus. Die in Dokumenten vorkommende Metadaten sind u.a. die

Quelle, von der die Dokumente stammen; das Sicherheitslevel der Dokumente; der Dateiname des Dokuments; das Dateiformat der Dokumente; die Dateigrößen; wie viele Dokumente in der Dokumentensammlung enthalten sind sowie dem Autor der Dokumente.

Netzwerk: Das Netzwerk stellt eine zentrale View des Tools 'new/s/leak' dar. Darin sind die aus den Dokumenten extrahierten Entitäten aufgeführt. Neben den Entitäten wird die Anzeige der Verbindungslinien zwischen zwei Entitäten benötigt. Die Verbindungslinie signalisiert hierbei die Existenz von Beziehungen zwischen Entitäten, extrahiert aus den darunter liegenden Dokumenten. Allerdings ist neben der Verbindungslinie auch der Inhalt der Verbindungslinie für die Journalisten von Interesse und verleiht der Verbindungslinie mehr Aussagekraft. Durch die Darstellung der Entitäten und deren Verbindungen mittels eines Netzwerkes können zudem weitere Informationen berechnet und extrahiert werden. So wurden in den Interviews Kennzahlen genannt, die Fragen abbilden wie wer die meisten Verbindungen hat oder wo im Netzwerk dichte Gruppen existieren. Die Anzeige und Filterung nach bspw. denjenigen Entitäten, die die meisten Verbindungen zu anderen Entitäten aufweisen, sollte daher durch geeignete Kennzahlen umgesetzt werden.

Ausschaltbare Netzwerkelemente: Werden alle Elemente von Anfang an angezeigt, kann dies zur Überladung des Netzwerkes führen. Da die Journalisten unterschiedliche Informationen zu verschiedenen Zeitpunkten angezeigt bekommen möchten, ist es notwendig, Netzwerkelemente ausschalten zu können. Hierbei bilden die extrahierten Entitäten die Basis des Netzwerkes. Die weiteren Netzwerkinformationen wie die Verbindungslinien, der Inhalt einer Verbindungslinie oder die Gewichtung von Knoten sollen als hinzufügbare Elemente auswählbar sein.

Zeitleiste: Die Zeit ist eine der am häufigsten genannten Information innerhalb der Interviews. Mit Hilfe der Zeitleiste sehen Journalisten auf einen Blick den zeitlichen Umfang der Dokumente (definiert über die Erstellungsdaten des ältesten und jüngsten Dokuments der Sammlung) sowie die Verteilung von Dokumenten über die Zeit. Anhand der Zeitleiste lassen sich zu verschiedenen Zeitpunkten der Recherche unterschiedliche Informationen darstellen, die im Zusammenspiel mit dem Netzwerk und der Dokumentenansicht interessante Einblicke in die Daten geben können. Informationen, die entlang der Zeitachse für die Journalisten von Bedeutung sind, ist die Dokumentenhäufigkeit, das Vorkommen der Entitäten über die Zeit und deren Verbindungen sowie die Anzeige der häufigsten Schlüsselwörter über die Zeit.

Dokumentenansicht: Eine weitere View, die Journalisten benötigen, ist die Dokumentenansicht. Hierdurch wird sichergestellt, dass die Journalisten immer direkten Zugriff auf die Dokumente der Dokumentensammlung haben und sehen können, welche Dokumente nach getätigter Filterung noch übrig bleiben. In der Dokumentenansicht werden Informationen über die Dokumente der Dokumentensammlung benötigt. Eine der wichtigsten Informationen, ist die Anzeige der Überschrift der Dokumente inklusive der ersten Zeilen. Mittels des Anrisses der Dokumente verschafft sich ein Journalist

einen Überblick über die Dokumente und muss nicht jedes einzelne durchlesen. Daneben sind die Metadaten der Dokumente ein wichtiger Bestandteil der Dokumentenübersicht. Die Metadaten können die Informationen Dateiname, Dateiformat, Dateigröße sowie Autor umfassen. Aber auch die Schlüsselworte eines Dokumentes wird für die Dokumentenansicht benötigt, weil die Journalisten hierdurch schnell einen Überblick bekommen, welche Dokumente spannend und welche nicht spannend sein könnten.

Karte: Einer geographischen Karte wurde innerhalb der Interviews eine eher nachrangige Bedeutung beigemessen. Hierdurch könnte die Karte als "nice-to-have" Funktion gesehen werden. Aufgrund der wenigen durchgeführten Interviews und dem Wissen, dass außerhalb der Interviews nach einer Karte als Visualisierung explizit gefragt wurde, wird deren Visualisierung trotzdem als Anforderung aufgenommen. Einerseits können anhand der Karte Orte und Datenpunkte visualisiert werden, die innerhalb der Dokumente vorkommen. Andererseits wird die Option benötigt, um nach Ländern und Städten filtern zu können.

Not getting lost: Eine nicht-funktionale Anforderung betrifft den Anspruch, in der Visualisierung nicht den Überblick zu verlieren. Im Journalismus wird in der Recherche ständig der Blickwinkel gewechselt, von der Metaebene in die Mikroebene und wieder zurück. Daher ist es wichtig den Prozess des investigativen Journalismus so gut wie möglich zu unterstützen.

In Abb. 2 sind die meisten der vorgestellten Anforderungen in einem Designvorschlag des Tools umgesetzt. Die roten Kreise beinhalten die IDs der jeweiligen Anforderungen.

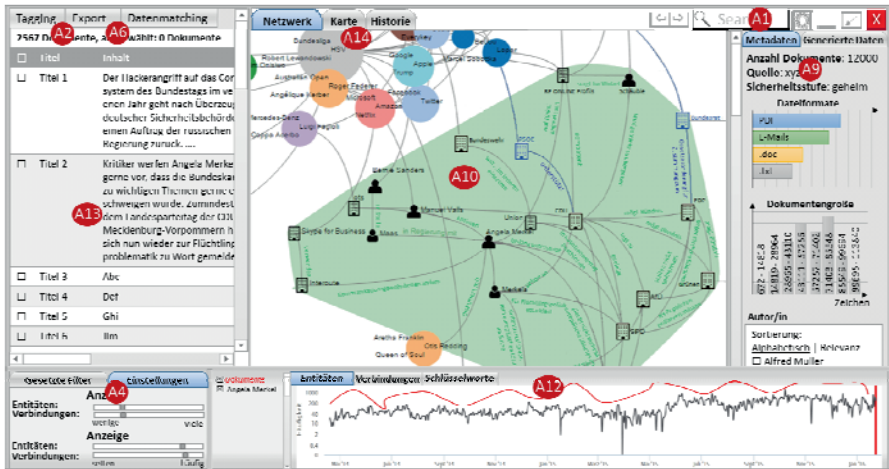


Abb. 2: Umsetzungsbeispiel der Anforderungen

5 Fazit

New/s/leak ist ein Visual Analytics Tool zur Unterstützung von Datenjournalisten bei der Exploration großer Datensammlungen. Der vorliegende Beitrag behandelt die relevanten Anforderungen bzgl. der interaktiven Visualisierung des Tools, identifiziert aus den durchgeführten Experteninterviews. Die Filterung der Dokumentensammlung anhand von Metadaten, Zeit, Entitäten und Schlüsselwörtern ist wichtig, um die Menge an Dokumenten in kürzester Zeit zu minimieren. Allerdings darf zwischen dem Journalist und der Dokumentensammlung keine Barriere vorhanden sein, sodass eine permanente Darstellung der Dokumentenanzeige benötigt wird. Da die Exploration von Datensammlungen im Team stattfindet, ist das einfache Kennzeichnen von relevanten Dokumenten durch Vergabe von Tags eine weitere der von Journalisten benötigten Funktionen. Aufgrund der automatisierten Entitätenerkennung können die Netzwerke Fehler enthalten. Daher ist es sinnvoll, verschiedene Bearbeitungsfunktionen des Netzwerks anzubieten, wie das Löschen von Entitäten. Zudem interessierten sich die interviewten Datenjournalisten außerdem für die Berechnung von Netzwerk Kennzahlen. Dadurch lässt sich das Netzwerk bspw. nach denjenigen Entitäten filtern, die die meisten Verbindungen zu anderen Entitäten aufweisen.

Die nächsten Schritte des Projektes beinhalten die Implementierung der Anforderungen im bisher existierenden Tool sowie die Evaluierung des bisherigen Prototyps durch Nutzerstudien. Da sich die herausgearbeiteten Anforderungen insbesondere auf derzeitige Abläufe sowie bekannte Visualisierungen zurückführen lassen und zukunftsweisende, innovative Gestaltungselemente nicht ausreichend berücksichtigt werden, können im nächsten Schritt mögliche Visualisierungen entwickelt und getestet werden, um die Anwendbarkeit neuartiger Darstellungen im Journalismus zu überprüfen.

Danksagung

An dieser Stelle möchte ich mich bei Dr. Tatiana von Landesberger sowie Prof. Dr. Chris Biemann bedanken, die mit ihrem konstruktiven Feedback bei der Erstellung dieses Beitrags mitgewirkt haben. Besonderer Dank geht an Kathrin Ballweg, welche den Prozess der Anforderungsanalyse tatkräftig unterstützt hat.

Literaturverzeichnis

- [Au15] Ausserhofer, J.: "Die Methode liegt im Code": Routinen und digitale Methoden im Datenjournalismus. In (A. Maireder et al., Hrsg.): Digitale Methoden in der Kommunikationswissenschaft, Berlin, 87-111, 2015.
- [BG09] Borchardt, A.; Göthlich, S.: Erkenntnisgewinnung durch Fallstudien. In (S. Albers et al., Hrsg.): Methodik der empirischen Forschung. Gabler Verlag, 33-48, 2009.
- [Br14] Brehmer, M. et al.: Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool for Investigative Journalists. IEEE Transactions on

- Visualization and Computer Graphics, 20(12), 2271-2280, 2014.
- [Br09] Briggs, M.: Data-driven journalism and digitizing your life. <http://www.journalism20.com/blog/2009/07/14/data-driven-journalism-and-digitizing-your-life/>, Stand: 29.04.2016.
- [Do11] Dou, W. et al.: ParallelTopics: A probabilistic approach to exploring document collections. IEEE Conference on Visual Analytics Science and Technology, 2011.
- [Fa14] Fahrer, U. et al.: Network of the Day: Interactive Visualization of Time-Dependent Entity Relation Networks. Darmstadt, Germany, Vision Modeling and Visualization Workshop, 2014.
- [Ga13] Gangemi, A.: A Comparison of Knowledge Extraction Tools for the Semantic Web. In (P. Cimiano et al., Hrsg.): The Semantic Web: Semantics and Big Data: 10th International Conference. Springer Berlin Heidelberg, Berlin, Heidelberg, 351-366, 2013.
- [GL10] Gläser, J.; Laudel, G.: Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen. 4. Aufl., Wiesbaden, VS-Verl., 2010.
- [Gö13] Görg, C. et al.: Combining Computational Analyses and Interactive Visualization for Document Exploration and Sensemaking in Jigsaw. IEEE Transactions on Visualization and Computer Graphics, 19(10), 1646-1663, 2013.
- [Gy14] Gynnild, A.: Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets. Journalism, 15(6), 713-730, 2014.
- [Ha00] Havre, S. et al.: ThemeRiver: visualizing thematic changes in large document collections. IEEE Transactions on Visualization and Computer Graphics, 8(1), 9-20, 2002.
- [Jä15] Jänicke, S. et al.: On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges. in Eurographics Conference on Visualization, R. Borgo et al., Hrsg.: The Eurographics Association, 2015.
- [Ka06] Kang, H. et al.: NetLens: Iterative Exploration of Content-Actor Network Data. IEEE Symposium On Visual Analytics Science And Technology, 2006.
- [KLB14] Kochtchi, A. et al.: Networks of Names: Visual Exploration and Semi-Automatic Tagging of Social Networks from Newspaper Articles. Comput. Graph. Forum, 33(3), 211-220, 2014.
- [Le05] Lee, B. et al.: Understanding research trends in conferences using paperLens. in Extended Abstracts on Human Factors in Computing Systems. Portland, OR, USA: ACM, 1969-1972, 2005.
- [Lu07] Ludwig, J.: Investigativer Journalismus. 2. überarb. Aufl., Konstanz, UVK-Verl.-Ges., 2007.
- [Ma10] Matzat, L.: Data Driven Journalism: Versuch einer Definition. <http://datenjournalist.de/data-driven-journalism-versuch-einer-definition/>, Stand: 29.04.2016.
- [MN09] Meuser, M.; Nagel, U.: Das Experteninterview - konzeptionelle Grundlagen und methodische Anlage. In (S. Pickel et al., Hrsg.): Methoden der vergleichenden Politik- und Sozialwissenschaft. Verlag für Sozialwissenschaften, 465-479, 2009.

- [RHR16] Roberts, J. C. et al.: Sketching Designs Using the Five Design-Sheet Methodology. *IEEE Transactions on Visualization and Computer Graphics*, 22(1), 419-428, 2016.
- [SMM12] Sedlmair, M. et al.: Design Study Methodology: Reflections from the Trenches and the Stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2431-2440, 2012.
- [SG06] Stellman, A.; Greene, J.: *Applied Software Project Management*. O'Reilly Media, 2006.
- [Su13] Sun, G.-D. et al.: A Survey of Visual Analytics Techniques and Applications: State-of-the-Art Research and Future Challenges. *Journal of Computer Science and Technology*, 28(5), 852-867, 2013.
- [Su16] Sun, M. et al.: BiSet: Semantic Edge Bundling with Biclusters for Sensemaking. *IEEE Transactions on Visualization and Computer Graphics*, 22(1), 310-319, 2016.
- [UK15] Uskali, Turo; Kuutti, Heikki: Models and Streams of Data Journalism. *The Journal of Media Innovations*, 77-88, 2015.
- [Zh13] Zhang, Z. et al.: The five Ws for information visualization with application to healthcare informatics. *IEEE Transactions on Visualization and Computer Graphics*, 19(11), 1895-1910, 2013.

μ -Force Control - A Device for Controlling Power Wheelchairs for Severely Mobility Impaired Persons

Christoph Dobler¹

Abstract: A new device for the control of power wheelchairs is introduced. Besides a speech recognition module its main component consists in two force sensors, shaped to match the hands of the user. The sensors are able to measure slightest pressures executed by the thumbs. Commonly, mobility impaired persons control a power wheelchair via a joystick. This requires the ability to execute distinct movements of the fingers or hands. In case the impairments are more severe and affect also the upper limbs, the possibility of wheelchair control via a joystick might not exist. Such persons in mind, a concept has been developed in which minimal muscular forces suffice to control a power wheelchair. A prototype of the device has been constructed that measures and combines the pressures from both thumbs and translates them to direction and speed of the wheelchair. It has been tested under reality conditions in various indoor and outdoor scenarios.

Keywords: biomedical equipment, assistive devices, wheelchairs

1 Introduction

There are many different forms of diseases that lead to movement-limiting disabilities with similar endings: the persons cannot move individual body parts or even the whole body. Finally, they are confined to a wheelchair. Among common causation of such disabilities are types of muscular dystrophy and neuromuscular diseases. Muscular dystrophy is a group of inherited diseases that are characterized by weakness and wasting away of muscle tissue. They are commonly caused by mutations in the genetic material. The most well known of the muscular dystrophies are the Duchenne muscular dystrophy (DMD) and the Becker muscular dystrophy (BMD). Neuromuscular diseases comprise a large number of different ailments that impair directly or indirectly nerves or neuromuscular junctions. While these diseases may have many different reasons, the result is often the same: the persons end up in a wheelchair.

Mobility is essential for handicapped people. It allows them to take part in social life. A power wheelchair is an important tool to provide handicapped people with some autonomy. In early stages of neuromuscular diseases a regular joystick is often sufficient to control the wheelchair. But with progressive weakening of muscles it becomes increasingly difficult to move the joystick. In this case a mini joystick is commonly the only option to retain control. It needs a force of 10 grams and the ability to move the thumb and the pointer finger in a small radius of approximately 5 – 10 mm. If the force and the freedom of movement is decreasing further, the impaired person will be able to use the joystick

¹FernUniversität in Hagen – University of Hagen, Faculty of Mathematics and Computer Science, Human-Computer Interaction, 58084 Hagen, Germany, christoph.dobler@studium.fernuni-hagen.de

for fewer directions (e. g., only forward). Afterwards, this input device is getting more and more useless. For keeping the ability of driving power wheelchairs in this situation, an alternative control approach is needed. Without it, affected people would not be able to go to school or university, to work or to meet other people.

2 Related Work

Commercially available input devices for power wheelchairs are mostly proportional devices like the mini joystick. They need distinct movements of limbs in order to operate. Further common input devices are shown in [WR15]. Examples are joysticks, touchpads, tablets, foot controls, head arrays and sip/puff controls. Future directions are going to isometric input devices and smart wheelchairs. Isometric joysticks [DCC10] are alternatives to traditional proportional joysticks. They sense force exerted on them and do not change position perceptively when a subject applies force. Smart Wheelchairs [Ta13, Tr13] consist of a power wheelchair and a personal computer with sensors. The personal computer interacts with the user and allows to control the wheelchair functions. Based on smart wheelchairs, there are projects of alternatives using brain computer interfaces [In11], myoelectric signals [OOH10], eye tracking [EGP16, Fo11, WSP10] as well as pressure sensitive touchpads [CHK14]. However, many of these systems are very expensive or their development is time-consuming.

3 μ -Force Control

The μ -Force Control is a new approach keeping the ability of driving power wheelchairs even with the a very small residual muscle strength. In comparison to other existing systems it has two essential advantages: it only requires minimum muscle strength, and unlike the interaction with a mini-joystick, the fingers do not have to be moved.

3.1 Hardware

Figure 1 illustrates the components of the proposed system in comparison to system utilizing a mini joystick. The μ -Force Control combines isometric input devices [DCC10] and smart wheelchairs [Ta13, Tr13] and needs no distinct movements of limbs but only minimal force of the thumbs. It consists of two force sensors (each one per thumb). The 3-D form of the sensors has been designed to match the hand form of the user. Of course every person has different requirements to the 3-D hand form of the sensors. The sensors' shape of the μ -Force Control can be customized to meet these requirements. In case of the presented prototype, the 3-D hand forms were modeled with a 3-D software and created with a 3-D printer.

The sensors are connected to an Intel Nuc personal computer running Microsoft Windows 10. For connecting the personal computer to the control of the power wheelchair, a serial

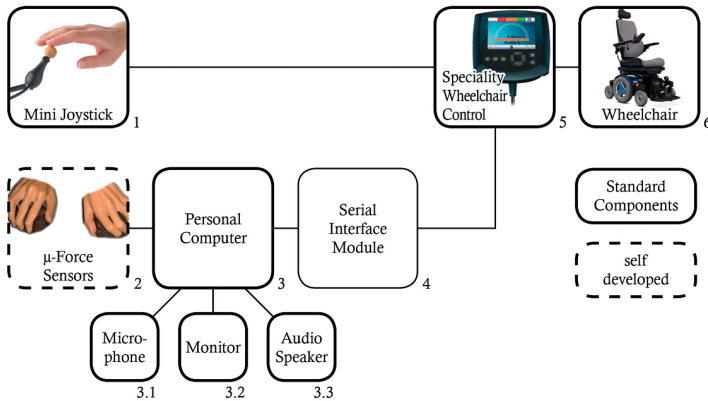


Fig. 1: Comparison of components of a system with a mini joystick (1,5,6) and the components of the μ -Force Control (2-6). The complete system is more complex than described in this paper. For example, it comprises a microphone for interaction via speech recognition. (Image sources: Christoffer Steen^a, mini joystick / Curtiss-Wright Industrial Group^b, R-net Omni speciality wheelchair control / Permobil^c, wheelchair.)

^a <http://kristoffersteen.dk>

^b <http://www.cv-industrialgroup.com/Products/Mobility-Vehicle-Solutions/R-net/Omni-Control-Interface.aspx>

^c http://countries.permobil.com/Austria/Produkte/Alle_modelle/M400-3G-corpus/

interface module is used [In11]. The fully functional prototype of the μ -Force Control is attached on a *Permobil m400* wheelchair with an *R-Net Omni* speciality wheelchair control (see Figure 3). The μ -Force Control application reads the sensor data and calculates the output for the wheelchair. The software allows to calibrate the sensors according to the force of the user.

3.2 Interaction

Pressing the left thumb turns the wheelchair right and vice versa. A light pressure with both thumbs selects the driving direction forward or backward and pressing both thumbs with a slightly stronger pressure accelerates the wheelchair in the current direction. However, controlling the wheelchair solely based on the pressure of the finger is – regardless of the calibration – not accurate enough. Therefore, a 7 inch touch-sensitive USB-screen with a resolution of 1024×600 px is used to generate an additional visual feedback. With the help of visual feedback through the driving vector (see Figure 2), a suitable coordination of the pressure of both thumbs is possible. Additionally, auditive feedback is emitted via speaker. This is done for important events, e. g., when the driving direction is changed, a calibration is started or done, or if any of the connected devices (microphone, force sensors, etc.) loses connection. Outside, if strong sunlight makes the display difficult to read, an auditory feedback is quite reasonable in addition to the visual feedback.

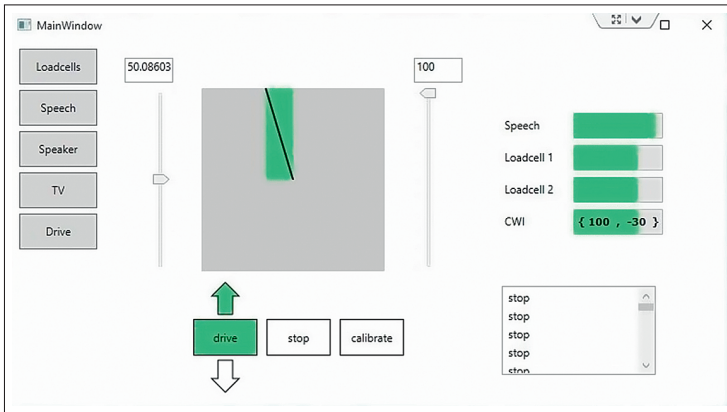


Fig. 2: This figure shows a screenshot of the graphical user interface with the driving vector indication (shown in the gray square at the center), the driving direction (bottom left), the system status (top right), and a speech input history of control commands (bottom right).

3.3 Software

Figure 2 shows a screenshot of the graphical user interface which can be controlled via speech recognition. The sliders on the left and right of the gray area show the measured force of the sensors. The driving vector shown in this box (black line in green rectangle) is calculated from these two values. Visible changes of the driving vector occur only when exceeding a certain value, e. g., 20% of the maximal pressure. The moving direction (bottom left on Figure 2) is flipped by pressing (and releasing) both sensors with a power below this threshold. If *drive* is activated, the values of the driving vector are sent to the wheelchair. The wheelchair is turned to the right/left by pressing the opposite thumb. When pressing both thumbs, the wheelchair is accelerated. If *stop* is activated, no control commands will be sent to the wheelchair. Instead, the force values will be used to control the mouse of the computer. The calibration procedure is started by activating the button *calibrate*. Then the sensors must be pushed with minimal and maximal force within 10 seconds. The bar indicators shown at the top right of Figure 2 give an overview of the current measured pressure of the sensors (*Loadcell 1/2*) and feedback on the status of the speech recognition and the computer wheelchair interface (CWI). Finally, the text-box at the bottom right on Figure 2 gives feedback on already used commands.

In combination with the other menus on the left side, it is possible to modify parameters of sensors (Figure 2), configure the speech recognition and the speaker. Further, the software allows the user to control all wheelchair functions including seat actuators and light.

4 Evaluation

The first author of this paper suffers from Duchenne muscular dystrophy and has developed the described prototype according to his own needs. In a period of continuous developing

and testing he was able to operate Y-Force Control in various indoor and outdoor scenarios under realtime conditions. The insights gained from these field studies have been incorporated in the development of the control device. Following real world scenarios have been tested:



Fig. 3: Different indoor and outdoor operation scenarios for the μ -Force Control.

- rooms in the author home (see Figure 3). In this environment, the doors have a standard width of about 80cm. A video showing the accuracy of control in this environment is located here:
<http://www.fernuni-hagen.de/mci/resources/software/video1.wmv>.
- at work and in public buildings such as shopping malls, restaurants, cafés, or cinemas. Moreover, also elevators in public buildings have been used.
- outdoor scenes. This includes passing through a meadow, covering larger distances (1.5km) on paved roads (see Figure 3) or shorter distances on cobblestone, and driving onto the ramp of a car loading system. The latter can be watched in a video which is located here:
<http://www.fernuni-hagen.de/mci/resources/software/video2.wmv>.

Turning around, braking, accelerating, or driving the wheelchair backwards are possible without any major problems. The tested top speed is currently at approximately 6 km/h. Depending on the pavement, more force is needed to compensate for the vibration. Short rest periods or a repositioning of the hands may occasionally be required by an accompanying person. Though a systematic evaluation remains to be done there is already ample evidence, that the prototype of μ -Force Control introduced in this paper represents a significant relief in the everyday life of severely mobility impaired persons.

5 Conclusion

In this paper a new approach to control power wheelchairs for severely mobility impaired persons was proposed. Compared to the common approach of using a joystick, the pro-

posed μ -Force Control approach requires no distinct movement of the fingers but only a minimal force of the thumbs. Thus, finding a good position of the hands is much easier in contrast to systems using a joystick. The latter requires the hand to be positioned precisely. If the hand is too much ahead (behind), a user of the target group is not able to drive backward (forward). In addition, the utilization of standard computer components offers much flexibility. The driving algorithm can be changed to meet special user demands and the application can be extended with eye tracking. In addition, using both hands supports a better sitting posture than one-sided controlled joysticks.

Summarizing, the proposed control device for mobility impaired persons has proven to enhance mobility in various indoor and outdoor scenarios and thus to improve participation in social life.

References

- [CHK14] Carrington, Patrick; Hurst, Amy; Kane, Shaun K.: The gest-rest: a pressure-sensitive chairable input pad for power wheelchair armrests. In: Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility, ASSETS '14, Rochester, NY, USA, October 20-22, 2014. pp. 201–208, 2014.
- [DCC10] Dicianno, Brad E.; Cooper, Rory A.; Coltellaro, John: Joystick Control for Powered Mobility: Current State of Technology and Future Directions. *Physical Medicine and Rehabilitation Clinics of North America*, 21(1):79 – 86, 2010.
- [EGP16] Eidam, Simone; Garstka, Jens; Peters, Gabriele: Towards Regaining Mobility Through Virtual Presence for Patients with Locked-in Syndrome. In: Proceedings of the 8th International Conference on Advanced Cognitive Technologies and Applications. Rome, Italy, pp. 120–123, 2016.
- [Fo11] Fondation Suisse pour les Téléthèses (FST): Computer Wheelchair Interface. <http://www.fstlab.ch/site/index.php/produits/cwi>, 2011. Accessed: 2016-04-15.
- [In11] Instructables: Brain-Controlled Wheelchair. <http://www.instructables.com/id/Brain-Controlled-Wheelchair/>, 2011. Accessed: 2016-04-15.
- [OOH10] Oonishi, Yuusuke; Oh, Sehoon; Hori, Yoichi: A New Control Method for Power-Assisted Wheelchair Based on the Surface Myoelectric Signal. *IEEE Trans. Industrial Electronics*, 57(9):3191–3196, 2010.
- [Ta13] Tavares, João; Barbosa, Jorge L. V.; da Costa, Cristiano André; Yamin, Adenauer C.; Real, Rodrigo Araújo: A smart wheelchair based on ubiquitous computing. In: The 6th International Conference on Pervasive Technologies Related to Assistive Environments, PETRA '13, Island of Rhodes, Greece, May 29-31, 2013. pp. 1:1–1:4, 2013.
- [Tr13] Trivedi, Amit Ranjan; Singh, Abhash Kumar; Digumarti, Sundara Tejaswi; Fulwani, Deepak; Kumar, Swagat: Design and Implementation of a Smart Wheelchair. In: Advances In Robotics 2013, AIR '13, Pune, India, July 4-6, 2013. pp. 96:1–96:6, 2013.
- [WR15] Walls, Ginger; Rolt, Russ: Evaluation for Alternative Input Devices for Power Wheelchairs – Steps for Successful Assessment, Documentation, and Delivery. In: RESNA - Annual Conference on Rehabilitation Engineering and Assistive Technology Society of North America. 2015.
- [WSP10] Wästlund, Erik; Sponseller, Kay; Pettersson, Ola: What you see is where you go: testing a gaze-driven power wheelchair for individuals with severe multiple disabilities. In: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, ETRA 2010, Austin, Texas, USA, March 22-24, 2010. pp. 133–136, 2010.

Twistor – Simulation des Twitterstroms für Evaluationszwecke

Harry Schilling¹

Abstract: Twitter ist ein Mikroblogging-Dienst, in dem aktuelle Ereignisse und Geschehnisse diskutiert werden. Es existiert eine Reihe von Algorithmen, um solche Ereignisse aus Twitterdaten extrahieren zu können. Die Evaluation dieser Verfahren gestaltet sich schwierig. Es stellt sich die Frage, ob ein gefundenes Ereignis auch ein Ereignis aus der Realität repräsentiert. Des Weiteren ist auch nicht bekannt, ob das Verfahren alle Ereignisse findet, da die Informationen fehlen, wie viele Ereignisse in den Twitterdaten wirklich vorhanden sind. Auch ist die Weitergabe von Twitterdaten verboten, sodass jeder Forscher gezwungen ist, seine eigenen Daten zu sammeln, die sich hinsichtlich der Qualität unterscheiden. Aufgrund dessen sind die Evaluationen von Ereigniserkennungsalgorithmen oft sehr heterogen und schlecht miteinander vergleichbar. Eine Möglichkeit diese Probleme anzugehen, bietet Twistor (**Twitter Stream Simulator**). Twistor ist ein Verfahren, um den Twitterstrom (mit dazugehörigen Ereignissen) zu simulieren. Somit ist immer eine konsistente Datenbasis und die Information über die zu suchenden Ereignisse vorhanden, um die Ergebnisse des Ereigniserkennungsalgorithmus bewerten und so einer einheitlichen Evaluation unterziehen zu können.

Keywords: Twitter Social Media Stream, Event Detection, Evaluation, Simulation

1 Einführung

Ein Mikroblogging-Dienst wie Twitter ist ein Medium, das Benutzern erlaubt, sehr kleine, digitale Inhalte wie z.B. kleine Texte mit anderen Benutzern zu tauschen. Dabei zeichnet sich Twitter vor allem dadurch aus, dass kurze Nachrichten schnell verarbeitet und geteilt werden können. So ist es möglich, Informationen in Echtzeit zu verbreiten und so aktuelle Geschehnisse und Ereignisse zu diskutieren oder zu kommentieren. Ereignisse wie z.B. Naturkatastrophen verbreiten sich mittels Twitter innerhalb kürzester Zeit [SOM10]. Es gibt eine Reihe von Algorithmen, die versuchen diese Ereignisse zu erfassen. Ein Problem hierbei ist, dass Twitter zwar unendlich viele und sehr schnell Informationen liefert, aber dass Twitternachrichten oft wenig gehaltvoll sind [KH11].

Neben der Ereigniserkennung an sich bereitet auch die Evaluation der Ergebnisse eines Ereigniserkennungsalgorithmus Schwierigkeiten. So muss z.B. sichergestellt werden, ob ein gefundenes Ereignis auch ein tatsächliches Ereignis repräsentiert. Um die Qualität der Ergebnisse richtig einordnen zu können, müssen neben der Verifikation eines Ereignisses auch alle in den Daten vorhandene Ereignisse bekannt sein. Hierzu können z.B. Nachrichtenseiten wie Reuters² hinzugezogen werden. Das ist deshalb problematisch, da die

¹ Universität Konstanz, Fachbereich Informatik und Informationswissenschaft, P.O. Box 188, 78457 Konstanz, Germany, harry.schilling@uni-konstanz.de

² <http://www.reuters.com/>

in den Daten vorhandenen Ereignisse nicht unbedingt die Ereignisse der hier beispielhaft aufgeführten Nachrichtenseite Reuters widerspiegeln. Infolgedessen sind viele Evaluationen von Ereigniserkennungsalgorithmen oft sehr unterschiedlich gestaltet, was zu einer schlechten Vergleichbarkeit zwischen den verschiedenen Ereigniserkennungsverfahren führt. Hinzu kommt, dass die Weitergabe von Twitterdaten von Twitter verboten wurde. Somit ist auch die Weitergabe der Evaluation zugrunde liegenden Daten untersagt. Es muss folglich jeder Forscher seine eigenen Twitterdaten sammeln, die sich in Qualität und Quantität unterscheiden. Die unterschiedliche Quantität und Qualität ist auch auf die Wahl der verschiedenen Twitter APIs, die benutzt werden, um die Daten zu sammeln, zurückzuführen. Möchte man mehr als 1 % der Twitterdaten erfassen, müssen verschiedene öffentlich zugängliche APIs miteinander kombiniert werden. Eine API, die einen speziellen Zugang erfordert, liefert 10 % des Stroms (Gardenhose). Gegen eine Bezahlung ist es möglich, 100 % des Stroms zu erfassen (Firehose). Auch die auf verschiedene Arten gesammelten Daten führen dazu, dass sich die Evaluationsergebnisse zwischen verschiedenen Ereigniserkennungsverfahren schlecht vergleichen lassen.

Es bietet sich somit an, ein standardisiertes Evaluationsverfahren zu verwenden, das sich in der Quantität und Qualität der zugrundeliegenden Daten nicht unterscheidet und vorgibt, wie viele Ereignisse vorhanden sind und welche Ereignisse gefunden werden können. Die Idee ist hierbei, den Twitterstrom zu simulieren. Als Grundlage dient hier der Gardenhose Zugang. In diesem erzeugten Twitterstrom werden dann künstlich Ereignisse integriert. Somit sind die Ereignisse, die der Ereigniserkennungsalgorithmus erkennen muss, bekannt und die Datenbasis weist im Hinblick auf die Größe und Qualität eine konsistente Struktur auf. Dieses Verfahren eignet sich somit als einheitliche Evaluationsmethode, um verschiedene Ereigniserkennungsalgorithmen miteinander vergleichen zu können. Das hier vorgestellte Verfahren, um den Twitterstrom zu simulieren, wird Twistor (**Twitter Stream Simulator**) genannt.

2 Verwandte Arbeiten

Da es eine Fülle von verschiedenen Evaluationsmethoden für die Ergebnisse von Ereigniserkennungsalgorithmen gibt, wird hier nur eine kleine Übersicht gegeben. Die Evaluationsmethoden werden in vier Gruppen aufgeteilt.

Eine von diesen Gruppen sind Fallstudien. So führte Corney [CMG14] eine Studie über die Ereignisse (z.B. Treffer erzielt) während eines Fußballspiels durch. Eine andere Gruppe sind einzelne Evaluierungen, die durchgeführt wurden, damit verschiedene Parameter angepasst werden können, um so bessere Ergebnisse zu erzielen. Ein Beispiel hierfür stammt von Ifrim [GI14]. Hier war ein Bestandteil der Evaluation zu untersuchen, wie sich die Veränderung der Parameter auf die Ereigniserkennung auswirkt. Die einzelnen Verfahren in dieser Gruppe lassen sich nur schlecht miteinander vergleichen. Die dritte Gruppe bilden komparative Evaluationen. So verglichen z.B. Weng und Lee [WL11] ihr Verfahren mit einer LDA [BNJ03]. Eine andere Gruppe von Evaluationen basiert auf Benutzerstudien. Hier werden die Resultate von Menschen bewertet. Eine Benutzerstudie zur Evaluation von Ergebnissen führte z.B. Thapen [TSH15] durch.

Die Heterogenität der verschiedenen Evaluationsmethoden macht es schwierig, die ver-

schiedenen Ereigniserkennungsalgorithmen miteinander zu vergleichen. Um dieses Problem zu lösen, implementierte Weiler [WGS15] verschiedene Algorithmen zur Ereigniserkennung in das gleiche Framework (Niagarino³) und unterzog die einzelnen Verfahren einer einheitlichen Evaluation.

3 Twistor – Simulation des Twitterstroms

3.1 Analyse des Twitterstroms

Da ein Großteil der Twitternachrichten oft wenig gehaltvoll ist [KH11], wird der Twisterstrom vor allem durch ein Grundrauschen (zur Ereigniserkennung nicht verwendbare Twitternachrichten) charakterisiert. Nach der Analyse der Twitterdaten, die mittels des Gardenhose-Zugang gesammelt wurden, kann festgestellt werden, dass die Verteilung der Wörter in den Daten innerhalb eines Tages einem ähnlichem Muster folgt.

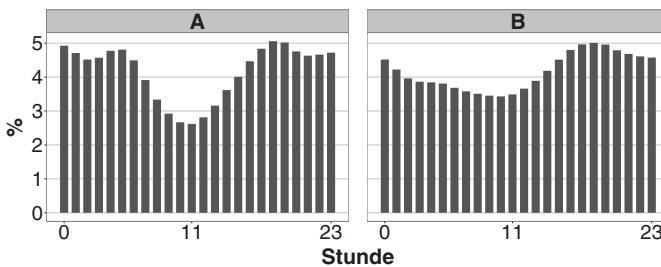


Abb. 1: Ein Beispiel für eine typische Verteilung der Wörteranzahl über einen Tag.

In Abbildung 1 ist auf der linken Seite (A) die relative (zu allen Wörtern pro Tag) Anzahl aller Wörter pro Stunde abgetragen. Hier ist zu erkennen, dass die Anzahl der Wörter zur Mitte des Tages abnimmt. Im weiteren Verlauf erhöht sich dann die Anzahl der Wörter wieder. Auf der rechten Seite (B) ist die relative (zu allen einmaligen Wörtern pro Tag) Anzahl an einmaligen Wörtern pro Stunde gegeben. Bei den einmaligen Wörtern (B) ist ein ähnlicher Effekt wie bei (A) auszumachen, der aber nicht ganz so stark ausgeprägt ist.

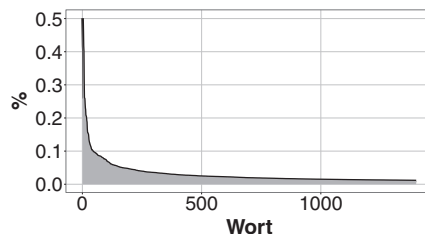


Abb. 2: Eine typische Verteilung der Worthäufigkeit über einen Tag betrachtet.

In Abbildung 2 ist die relative (zu allen Wörtern pro Tag) Häufigkeit aller Wörter über den ganzen Tag zu sehen. Da es für einen Tag über 3,5 Millionen Wörter gibt, sind hier

³ <http://www.informatik.uni-konstanz.de/grossniklaus/software/niagarino/>

nur die Werte, die über 0,01 % liegen, abgetragen. Dies entspricht 1400 Wörtern. Es ist zu erkennen, dass einige wenige Wörter im Vergleich zu allen anderen Wörtern relativ häufig vorkommen (linke Seite der Verteilung). Der Großteil der Wörter tritt aber im Vergleich zu allen Wörtern eher selten auf.

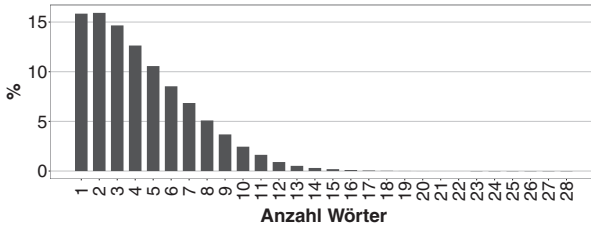


Abb. 3: Eine typische Verteilung der Wortanzahl in Twiternachrichten über einen ganzen Tag.

In Abbildung 3 ist die relative (zu allen Twiternachrichten pro Tag) Anzahl der Wörter in Twiternachrichten über einen ganzen Tag zu erkennen. Erwähnt werden muss, dass sog. Stoppwörter wie z.B. „the“, „also“ usw. oder auch Wörter, die nur aus Zahlen bestehen, URLs etc. gefiltert werden, da diese Wörter von den Ereigniserkennungsverfahren nicht verwendet werden. Es ist zu sehen, dass Twiternachrichten mit einem und zwei Wörter am häufigsten vorkommen (ca. 15 % aller Twiternachrichten). Je mehr Wörter in den Twiternachrichten vorkommen, desto geringer wird auch die Anzahl dieser Twiternachrichten.

Das letztendliche Ziel ist es, das Grundrauschen zu simulieren, indem eine ähnliche Verteilung der Wörter erzeugt wird.

3.2 Erzeugung des Twitterstroms

3.2.1 Überblick

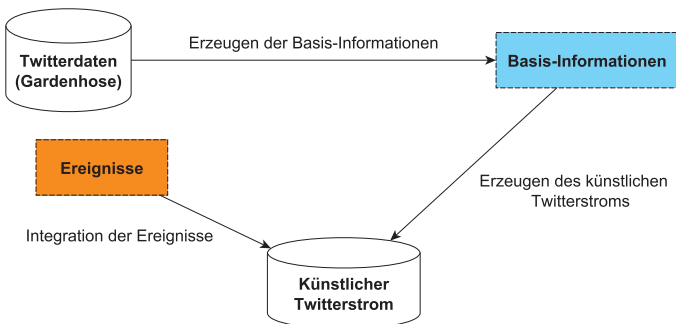


Abb. 4: Eine Übersicht über den Ablauf von Twistor.

Zuerst werden die Basis-Informationen aus den Twitterdaten ermittelt. Die Basis-Informationen geben Auskunft über die Verteilung der Twitterdaten und müssen nur einmal erzeugt

werden. Dann muss ausgewählt werden, welche Ereignisse in den künstlichen Twitterstrom integriert werden sollen. Die Ereignisse basieren auf realen Gegebenheiten und werden anhand von Wörtern, die subjektiv passend zum Ereignis sind, identifiziert. Von diesen Wörtern ist die Information vorhanden, wie häufig diese in den künstlichen Twitterdaten auftreten müssen. Die Ereignisse werden in den künstlichen Twitterstrom integriert, indem die Häufigkeiten der jeweiligen Wörter in den Strom abgebildet werden.

3.2.2 Ermitteln der Basis-Informationen

Um den Twitterstrom zu verarbeiten, wird dieser vom Ereigniserkennungsverfahren in Zeitfenster aufgeteilt. Alle Twitternachrichten, die innerhalb des Zeitfensters enthalten sind, werden gesammelt und verarbeitet. Dies wird für alle aufeinanderfolgenden Zeitfenster so durchgeführt. Die Größe des Zeitfensters kann beliebig gewählt werden. Liegen z.B. insgesamt Twitterdaten von einer Stunde vor und als Zeitfenster wird 15 Minuten gewählt, ergeben sich vier Zeitfenster. Bei der Erzeugung des Twitterstroms ist wichtig zu beachten, dass ein Wort z.B. im ersten und vierten Zeitfenster vorkommen kann, dazwischen aber nicht unbedingt.

Als Basis für die Erzeugung des künstlichen Twitterstroms dienen Daten, die sich über einen zufällig ausgewählten Tag (24 Stunden) erstrecken. Diese Daten wurden anhand des Gardenhose-Zugang gesammelt und repräsentieren 10 % des Twitterstroms. Die 24 Stunden dieser Daten werden in 1-Minuten-Zeitfenster aufgeteilt. Für jedes 1-Minuten-Zeitfenster werden zwei charakteristische Merkmale erfasst. Erstens wird ermittelt, wie häufig jedes Wortes pro 1-Minuten-Zeitfenster vorkommt. Zweitens wird die Verteilung der Anzahl der Wörter in den Twitternachrichten pro 1-Minuten-Zeitfenster festgestellt (z.B. Twitternachrichten mit einem Wort kommen 20 Mal vor, mit vier Wörtern 60 Mal und so weiter). Nach Abarbeitung aller 1-Minuten-Zeitfenster wird jedes vorhandene Wort durch ein künstlich erzeugtes Wort ersetzt. Die so erzeugten Informationen werden abgespeichert und dienen als Basis-Informationen. Da die Basis-Informationen abgespeichert werden, müssen diese auch nur einmal erzeugt werden. Möchte man einen anderen Tag als Grundlage für die Basis-Informationen hinzuziehen, so müssen diese neu erzeugt werden.

3.2.3 Erzeugen des künstlichen Twitterstroms aus den Basis-Informationen

Aufbauend auf den Basis-Informationen kann nun der künstliche Twitterstrom erzeugt werden. Hierbei muss zunächst angegeben werden, welche Größe das Zeitfenster haben soll. Die Größe des Zeitfensters sollte der Einstellung des Ereigniserkennungsverfahrens entsprechen. Die Mindestgröße für das Zeitfenster beträgt eine Minute. Außerdem muss angegeben werden, wie viel Stunden an Daten erzeugt werden sollen. Der maximale Wert hierfür ist 24 Stunden.

Bevor eine genaue Beschreibung des Algorithmus gegeben werden kann, müssen einige Definitionen getroffen werden.

Sei $W = \{w_1, \dots, w_n\}$ alle Wörter aus den Basis-Informationen. $I = \{i_1, \dots, i_m\}$ entspricht allen vorhandenen 1-Minuten-Zeitfenstern. Die Häufigkeit eines Wortes in einem gegebenen Zeitfenster wird mit $\theta_{k,l}$ angegeben, wobei k der Index für das Wort und l der Index für das Zeitfenster ist. Es gilt $1 \leq k \leq |W|$ und $1 \leq l \leq |I|$.

Es sei mit $C = \{c_1, \dots, c_p\}$ die Menge gegeben, bei der jedes Element eine mögliche Wortanzahl einer Twitternachricht repräsentiert. Mit $v_{q,t}$ wird angegeben, wie oft eine Twitternachricht mit q Wörtern im Zeitfenster t vorkommt, wobei $1 \leq q \leq \max(C)$ und $1 \leq t \leq |I|$ gilt.

Die Beschreibung des Algorithmus zur Erzeugung des künstlichen Twitterstroms soll nachfolgend in Pseudocode gegeben werden.

Algorithmus 1: Erzeugung des künstlichen Twitterstroms

Eingabe: $W, I, \theta_{k,l}, v_{q,t}$

```

foreach  $i \in I$  do
  tweets  $\leftarrow$  List()
  foreach  $v \in v_{q,i}$  do
    wordAmount  $\leftarrow$   $q$ 
    tweetAmount  $\leftarrow$   $v$ 
    for  $j \leftarrow 1$  to tweetAmount do
      words  $\leftarrow$  List()
      for  $l \leftarrow 1$  to wordAmount do
        word  $\leftarrow$  ExtractWord( $W, \theta_{k,i}, \text{words}$ )
        words  $\leftarrow$  Insert(word)
      tweets  $\leftarrow$  Insert(words) // mehrere Wörter repräsentieren einen Tweet
  Output(tweets) // schreibe oder streame Tweets

```

Um das passende Wort aus der Menge von Wörtern (W) für eine Twitternachricht auszuwählen, wird folgender Algorithmus verwendet:

Algorithmus 2: ExtractWord

Eingabe: $W, \theta_{k,l}, \text{words}$

```

word  $\leftarrow$  null
for  $i \leftarrow 1$  to  $|W|$  do
  if  $\theta_{i,l} > 0$  and  $W[i]$  IsNotInList(words) then
    word  $\leftarrow$   $W[i]$ 
     $\theta_{i,l} \leftarrow \theta_{i,l} - 1$ 
    break
/* word kann nicht null werden, da die Verteilung der Wörter so angelegt ist,
   dass immer ein Wort gefunden wird */
return word

```

Zu beachten ist, dass das in Algorithmus 1 angegebene Verfahren in der dargestellten Form nur für 1-Minuten-Zeitfenster definiert ist. Möchte man z.B. 5-Minuten-Zeitfenster verwenden, so müssen immer zuerst fünf 1-Minuten-Zeitfenster zusammengefasst werden, bevor der Algorithmus angewendet werden kann.

3.3 Definition und Identifikation von Ereignissen

Neben dem Erzeugen des Grundrauschens des Twitterstroms müssen Ereignisse künstlich in diesen integriert werden. Um dieses umsetzen zu können, muss ein Ereignis erst einmal charakterisiert werden. Vielen Ereigniserkennungsverfahren wie [WL11, Co12, KI02, Fu05] liegt zugrunde, dass diese an- und absteigende Häufigkeiten von Wörtern in den Twitternachrichten verwenden bzw. weiterverarbeiten, um Ereignisse ausfindig zu machen. Um diese Häufigkeiten zu repräsentieren, wird das IDF-Signal [SB88] der Wörter (engl. *Inverse Document Frequency*) verwendet. Dies hat den Vorteil, dass Wörter, die in fast jeder Twitternachricht oder nur selten vorkommen, eine geringere Bedeutung beigemessen wird als Wörtern, die nur in wenigen Twitternachrichten oft auftreten. Wörter, die in wenigen Twitternachrichten häufig vorkommen, können ein gerade stattfindendes Ereignis beschreiben. Noch zu erwähnen ist, dass je öfter ein Wort auftritt, desto kleiner sein IDF-Wert wird (und nicht größer).

Um eine möglichst genaue Definition von Ereignissen wie z.B. Naturkatastrophen oder Sportveranstaltungen anhand des IDF-Signals treffen zu können, wurden reale Ereignisse herangezogen und näher analysiert. Es folgen nun zwei ausgewählte Beispiele. Die Wörter, von denen das IDF-Signal erfasst wurde, wurden subjektiv passend zum Ereignis ausgewählt.

3.3.1 Beispiele

Das erste Ereignis ist das Tor von Mario Götze im Finale der Weltmeisterschaft von 2014. Dieses ist am 13.07.14 um 21:24 Uhr (GMT) gefallen. Als passende Wörter für dieses Ereignis wurden „goal“, „goetze“ und „scored“ ausgewählt. Abbildung 5 zeigt den Verlauf der IDF-Werte für die drei Wörter. Es ist gut zu erkennen, dass ab ca. 21:25 Uhr die IDF-Werte für alle drei Wörter sehr viel kleiner werden. Dies bedeutet, dass die Wörter häufiger genannt werden.

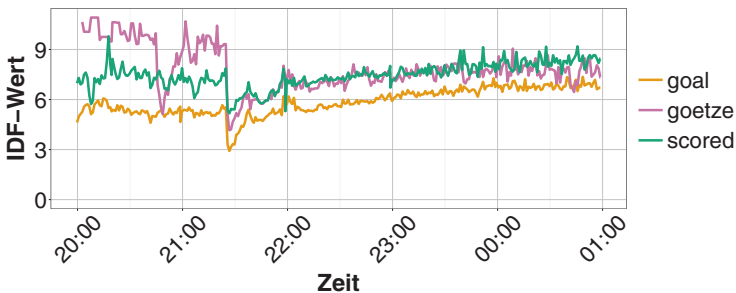


Abb. 5: Die IDF-Werte für das Tor im Finale der Weltmeisterschaft von 2014.

In den fortlaufenden Stunden steigen die IDF-Werte dann langsam an. Die drei Wörter werden somit immer weniger genannt, sodass dieses Ereignis an Bedeutung verliert.

Das zweite Ereignis ist die Papst-Wahl am 13.03.2013 um 18:06 Uhr (GMT). Als Wörter, die das Ereignis beschreiben, wurden „habemus“ und „papam“ ausgewählt.

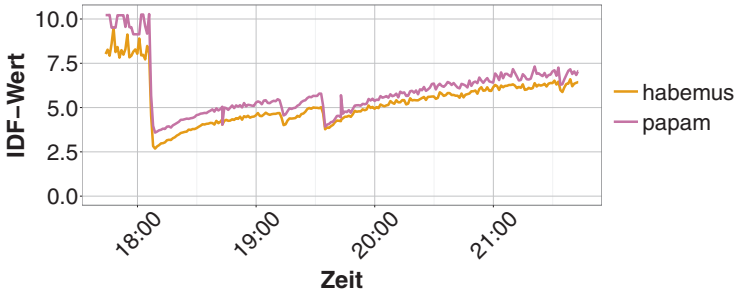


Abb. 6: Die IDF-Werte für die Papst-Wahl von 2013.

In Abbildung 6 ist das IDF-Signal für die beiden Wörter abgebildet. Es zeigt sich, dass es ab ca. 18:07 Uhr zu einem Abfall der IDF-Werte kommt. Die beiden Wörter werden zu diesem Zeitpunkt deutlich häufiger genannt als vorher. Im weiteren Verlauf steigen die IDF-Werte langsam an. Die zwei Wörter treten immer seltener auf und die Relevanz des Ereignisses sinkt.

Es zeigt sich, dass Ereignisse in den Twitterdaten sich dadurch auszeichnen, dass der Verlauf der IDF-Werte der zum Ereignis passenden Wörter einen starken „Knick“ (s. Abbildung 5 ca. 21:25 Uhr und Abbildung 6 ca. 18:07 Uhr) aufweist.

3.4 Integration der Ereignisse

Um Ereignisse in den künstlichen Twitterstrom zu integrieren, werden die Daten von realen Ereignissen herangezogen. Das bedeutet, dass der Verlauf der IDF-Werte von Wörtern, die für das Ereignis relevant sind, in den künstlichen Twitterstrom abgebildet wird. Auf diese Weise entsteht in dem künstlichen Twitterstrom ein reales Abbild eines Ereignisses. Um den IDF-Verlauf von diesen Wörtern zu imitieren, muss die Anzahl an Twitternachrichten und die Auftretenshäufigkeit der Wörter in ein bestimmtes Verhältnis gesetzt werden.

Zunächst die Definition des IDF-Wertes:

$$\text{idf}(w) = \log \left(\frac{N}{n_w} \right) \quad (1)$$

N steht für die Anzahl aller Twitternachrichten, w entspricht dem Wort, von dem der IDF-Wert berechnet wird und n_w ist die Anzahl aller Twitternachrichten, die w enthalten. Da der IDF-Wert bekannt ist (dieser soll imitiert werden) und auch die Anzahl der Twitternachrichten durch den künstlichen Twitterstrom vorgegeben wird, muss die Anzahl an

Twitternachrichten, die w enthalten, berechnet werden. Umstellen der Gleichung 1 nach n_w :

$$n_w = \frac{N}{e^{\text{idf}(w)}} \quad (2)$$

Durch Berechnen von n_w kann nun angegeben werden, wie oft das Wort w in den Twitterdaten auftreten muss, damit der entsprechende IDF-Wert erreicht wird.

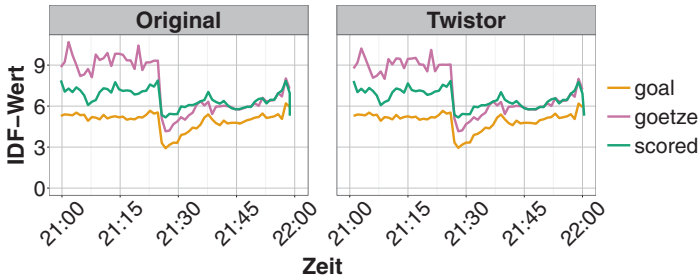


Abb. 7: Die IDF-Werte für das WM-Tor im Finale von 2014.

In Abbildung 7 ist beides mal das Ereignis des WM-Tors im Finale von 2014 mit den entsprechenden IDF-Werten zu sehen. Links (Original) ist der Verlauf der IDF-Werte in den originalen Twitterdaten dargestellt. Rechts (Twistor) ist das IDF-Signal abgebildet, nachdem es in den künstlichen Twitterstrom integriert wurde. Vergleicht man alle IDF-Werte des Originals mit den Twistor Werten so beträgt die durchschnittliche Abweichung gerundet 0.04. Die Unterschiede sind also minimal.

Auch lassen sich mehrere Ereignisse mit den entsprechenden IDF-Signalen innerhalb des künstlichen Twitterstroms an verschiedenen Zeitpunkten abbilden.

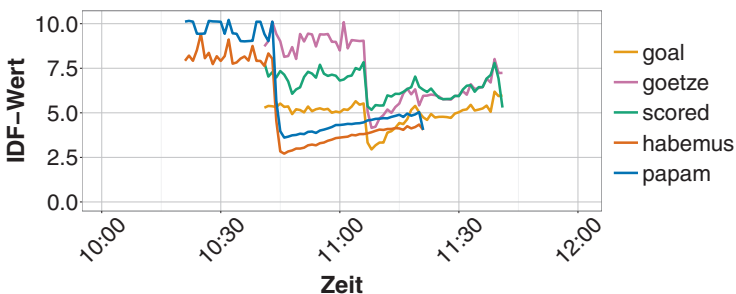


Abb. 8: Das Ereignis des WM-Tors im Finale von 2014 und der Papst-Wahl von 2013.

In Abbildung 8 wurde das Ereignis der Papst-Wahl von 2013 („habemus“ und „papam“) um 20 Minuten und das Ereignis des WM-Tors im Finale von 2014 („goal“, „goetze“ und „scored“) um 40 Minuten verschoben.

Die IDF-Signale der jeweils zum Ereignis passenden Wörter sind gespeichert und müssen nur einmal gebildet werden.

4 Abschluss und Ausblick

Mit Twistor lässt sich die Evaluation von Algorithmen zur Ereigniserkennung besser bewältigen, da die Ereignisse, die zu finden sind, vorgegeben werden können. Es bietet somit eine zuverlässigere Variante als z.B. das Abgleichen der Ergebnisse des Ereigniserkennungsverfahrens mit einer Nachrichtenseite wie Reuters. Es lässt sich damit eine standardisierte Evaluation durchführen.

Mittlerweile untersagt Twitter die Weitergabe von Twitterdaten, was dazu führt, dass jeder Forscher seine eigenen Twitterdaten sammeln muss. So benutzen alle ihre eigenen Daten, die sich von der Qualität und Größe unterscheiden. Dieser Umstand führt dazu, dass die Vergleichbarkeit zwischen den Ergebnissen der Ereigniserkennungsalgorithmen nicht gewährleistet ist. Auch hier kann Twistor helfen, da keine Daten weitergegeben werden müssten.

Außerdem kann Twistor z.B. auch direkt als Input-Quelle in einem Datenstrommanagementsystem genutzt werden. Üblicherweise werden in einem Datenstrommanagementsystem zur Analyse des Twitterstroms nicht der Twitterstrom an sich, sondern Textdateien benutzt. Um den Twitterstrom zu simulieren, kann Twistor benutzt werden.

Das Projekt ist in dieser Form noch nicht abgeschlossen. Als Grundlage für die Erzeugung der Basis-Informationen dient der Gardenhose-Zugang. Dieser liefert 10 % des gesamten Twitterstroms. Somit produziert auch Twistor 10 % des realen Twitterstroms. Zukünftig ist geplant, dass die Anzahl der Twitternachrichten z.B. auf 100 % hochskaliert werden kann. Bei einer Hochskalierung der Anzahl der Twitternachrichten bleibt die Verteilung der Wörter auf dem Stand von 10 %, da keine weiteren Informationen über die Verteilung der Wörter vorliegen.

Weiterhin ist auch geplant, dass Ereignisse, die in den künstlichen Twitterstrom integriert werden, „abgeschwächt“ werden können. Dies bedeutet, dass der charakteristische „Knick“ im Verlauf der IDF-Werte eines Ereignisses in die Länge gezogen wird, sodass der drastische Abfall der IDF-Werte reduziert wird.

Außerdem soll auch ein synthetisches Ereignis, das keinem realen Ereignis nachempfunden ist, definiert und in den künstlichen Twitterstrom integriert werden können.

Trotz aller Vorteile, die Twistor gegenüber den schon vorhandenen Evaluationsmethoden liefern würde, muss natürlich die Qualität und Brauchbarkeit des simulierten Stroms überprüft werden. Dies soll geschehen, indem die Ergebnisse der Ereigniserkennungsalgorithmen, die einmal durch den künstlichen Twitterstrom und ein anderes Mal durch die Originaldaten (Gardenhose) entstanden sind, miteinander verglichen werden. Zur Evaluation sollen verschiedene Ereigniserkennungsverfahren hinzugezogen werden. Wie die Evaluation im Detail aussehen soll, steht zu diesem Zeitpunkt aber noch nicht fest.

Danksagung

Einen großen Dank gebührt Michael Grossniklaus und Andreas Weiler für die Betreuung und Unterstützung bei der Anfertigung dieser Arbeit. Die Forschung zu Erlangung der in diesem Beitrag präsentierten Forschungsergebnisse wird teilweise gefördert von der Deutschen Forschungsgemeinschaft (DFG), Grant No. GR 4497/4: Adaptive and Scalable Event Detection Techniques for Twitter Data Streams.

Literaturverzeichnis

- [BNJ03] Blei, David M.; Ng, Andrew Y.; Jordan, Michael I.: Latent Dirichlet Allocation. *J. Mach. Learn. Res.*, 3:993–1022, 2003.
- [CMG14] Corney, David; Martin, Carlos; Göker, Ayse: Spot the Ball: Detecting Sports Events on Twitter. In (de Rijke, Maarten; Kenter, Tom; de Vries, Arjen P.; Zhai, ChengXiang; de Jong, Franciska; Radinsky, Kira; Hofmann, Katja, Hrsg.): *Advances in Information Retrieval*, Jgg. 8416 in *Lecture Notes in Computer Science*, S. 449–454. Springer International Publishing, 2014.
- [Co12] Cordeiro, Mário: Twitter Event Detection: Combining Wavelet Analysis and Topic Inference Summarization. In: *Proc. Doctoral Symposium on Informatics Engineering (DSIE)*. S. 123–138, 2012.
- [Fu05] Fung, Gabriel Pui Cheong; Yu, Jeffrey Xu; Yu, Philip S.; Lu, Hongjun: Parameter Free Bursty Events Detection in Text Streams. In: *Proc. Intl. Conf. on Very Large Data Bases (VLDB)*. S. 181–192, 2005.
- [GI14] Georgiana Ifrim, Bichen Shi, Igor Brigadir: Event Detection in Twitter using Aggressive Filtering and Hierarchical Tweet Clustering. In: *Proc. Workshop on Social News on the Web (SNOW) in conjunction with Intl. Conf. Companion on World Wide Web (WWW)*. S. 33–40, 2014.
- [KH11] Kaplan, Andreas M.; Haenlein, Michael: The Early Bird Catches the News: Nine Things You Should Know about Micro-Blogging. *Business Horizons*, 54(2):105–113, 2011.
- [KI02] Kleinberg, Jon: Bursty and Hierarchical Structure in Streams. In: *Proc. Intl. Conf. on Knowledge Discovery and Data Mining (KDD)*. S. 91–101, 2002.
- [SB88] Salton, Gerard; Buckley, Christopher: Term-Weighting Approaches in Automatic Text Retrieval. *Inf. Process. Manage.*, 24(5):513–523, 1988.
- [SOM10] Sakaki, Takeshi; Okazaki, Makoto; Matsuo, Yutaka: Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. In: *Proc. Intl. Conf. on World Wide Web (WWW)*. S. 851–860, 2010.
- [TSH15] Thapen, Nicholas A.; Simmie, Donal Stephen; Hankin, Chris: The Early Bird Catches the Term: Combining Twitter and News Data for Event Detection and Situational Awareness. *CoRR*, abs/1504.02335, 2015.
- [WGS15] Weiler, Andreas; Grossniklaus, Michael; Scholl, Marc H.: Evaluation Measures for Event Detection Techniques on Twitter Data Streams. In: *Proc. British Intl. Conf. on Databases (BICOD)*. S. 108–119, 2015.
- [WL11] Weng, Jianshu; Lee, Bu-Sung: Event Detection in Twitter. In: *Proc. Intl. Conf on Weblogs and Social Media (ICWSM)*. S. 401–408, 2011.

A survey on approaches to anonymity in Bitcoin and other cryptocurrencies

Felix Konstantin Maurer¹

Abstract: Bitcoin is a crypto currency with several advantages over previous approaches. Transactions are confirmed and stored by a peer-to-peer network in a blockchain. Therefore, all transactions are public and soon solutions were designed to increase privacy in Bitcoin. Many come with downsides, like requiring a trusted third-party or requiring modifications to Bitcoin. In this paper, we compare these approaches according to several criteria. Based on our findings, CoinJoin emerges as the best approach for anonymizing Bitcoins today.

Keywords: Bitcoin, cryptocurrencies, coin mixing, anonymity, transaction linkability

1 Introduction

Bitcoin [Na08] is a new cryptocurrency with several advantages over previous approaches. A peer-to-peer network is used to confirm the validity of transactions. However, the network stores all valid transactions which are therefore always public. Even though Bitcoin uses pseudonyms, it does not provide anonymity. Each transaction is linked to previous transactions and thus only one pseudonym must be known to infer other pseudonyms.

Consequently, employers paying in bitcoins might be able to track your spending and stores, landlords or anyone receiving payments could be able to know your balance. In currently used monetary systems this is not possible. Therefore, anonymity is going to be an requirement for any crypto currency in the future that tries to replace existing systems.

For Bitcoin, there already exist services that allow performing transactions through a third party. These are called mixers as they try to conceal a transaction in a large amount of unrelated transactions. Depending on the design, several problems can arise. For example, the mixing service might learn which addresses are connected. Therefore, other concepts were developed, even new cryptocurrencies which provide more privacy than Bitcoin.

2 Related Work

In Bitcoin, transactions are confirmed and preserved by being inserted into a chain of transaction blocks. As part of the chain they can not be modified, as the blocks are linked to their precursor by embedding its cryptographic hash. This block chain forms the public ledger of the Bitcoin network and represents the consensus about all performed transactions. Each transaction can consist of many inputs and outputs. An output is an amount

¹ Karlsruhe Institute for Technology, Institute of Telematics, felix.maurer@student.kit.edu

of bitcoins and a small program, called output script, that is used to verify if a person can spend the coins. Each input references an output and provides the input for the script. Usually, the output script verifies that the input was created with a specific private key.

Regal Reid and Martin Harrigan [RH13] are able to demonstrate that multiple pseudonymous addresses can be linked to a single user. They construct a graph of Bitcoin transactions (T) and a graph of Bitcoin addresses (U). Assuming that all inputs of a transaction belong to the same user, they then contract the graph nodes of U by merging addresses that appear in the inputs of a single transaction. Furthermore, they include temporal and external information to link more addresses to real identities.

Florian Tschorsch and Björn Scheuermann [BdL13] extensively discuss Bitcoin. They provide a section on enabling privacy where they present several approaches, also discussed in our work. However, they do not compare them or provide recommendations.

Bonneau et al. [Bo15] also cover various aspects of Bitcoin. They discuss privacy and anonymity in a short section including a comparative table. The approaches are divided into peer-to-peer mixing protocols, distributed mix networks and altcoins. As the first survey, they do not draw conclusions from their comparison.

3 Taxonomy

In this paper, we assume an honest but curious adversary. It could be for example a credit institute, that wants to learn about the spending habits of its clients, or a landlord that would like to know whether her tenant is financially stable. The adversary will conduct transactions with the user and therefore knows at least one of her pseudonyms addresses. To protect the privacy of the user, public information of the cryptocurrency should not allow the adversary to infer other pseudonyms or transactions of the user. We assume that he or she will not try to gain additional information that is not part of the blockchain.

We compare existing work based on several criteria. Most important to us is *Bitcoin compatibility* meaning whether the Bitcoin protocol would have to be modified or not. Right now, Bitcoin is the largest crypto currency¹ by value and transaction volume and is actively developed and well understood. Therefore, it is likely that Bitcoin remains the dominant cryptocurrency and an incompatible mixing approach might not be adopted. The approach should make *theft impossible* as loss of reputation might be acceptable for a mix as long as it gains enough Bitcoins. Different protocols use different *architectures* for mixing coins. Like Bonneau et al. [Bo15], we will distinguish between peer-to-peer mixing protocols, mixing services and altcoins. Furthermore, most approaches need more than a *single transaction* to anonymously send an arbitrary amount of Bitcoins. This increases the time it takes to complete the mix, the amount of transaction fees paid and the energy needed by the network. In cases of peer-to-peer mixing protocols or services, *anonymity against the mixer* will be compared. We also distinguish whether it is *reliant on new cryptographic methods* not used in Bitcoin as it might lessen the confidence in the solution.

¹ <https://coinmarketcap.com/>

4 Survey

As we can not cover all existing approaches, we chose relevant and representative ones.

4.1 Mixcoin and Blindcoin

Mixcoin [Bo14] is a mixing service with accountability features. Bitcoin users negotiate a set of parameters with the service, including the address where the coins should be sent to. To provide anonymity, all users must use the same amount when mixing and multiple users must use the service at the same time. The service will provide a signed warranty that can be published in case the service steals the coins.

Mixcoin is compatible with Bitcoin and does not require new cryptographic methods. As a central mixing service, it is easier to protect against DoS attacks by single users compared to p2p mixing protocols. However, the mixer will learn the connection between the input and output address. To protect against this, different providers can be used in sequence. This further increases the number of Bitcoin transaction needed and total transaction costs.

Blindcoin [VR15] improves on Mixcoin by using blind signatures to ensure that the mix can't link the input and output address. Nevertheless, the amount that can be mixed is still fixed and the anonymity depends on the number of simultaneous users. Also, the user must be able to anonymously publish the output address to a public log which might result into a bootstrapping problem. Furthermore, while theft will be detected and can be proven, it is not prevented and Bitcoins might still be "lost"².

4.2 CoinJoin

CoinJoin [Ma13a] is a concept by Gregory Maxwell of mixing transactions by joining them into a larger transaction. It exploits the fact that a transaction can have multiple inputs and outputs that do not need to belong to the same person. This increases the anonymity of a single transaction, but also can increase the anonymity of Bitcoin in general. As these join transactions are in principle indistinguishable from other transactions, the assumption that inputs of a transaction belong to a single person does no longer hold.

However, an actual implementation still has to overcome some challenges. First, linking of inputs and outputs might still be possible, based on the value. This can be fixed by requiring each user to transmit the same amount. However, this is unpractical and might make join transactions distinguishable from regular transactions. Then, depending on how the transactions is constructed, all participants learn the connection between input and output addresses. If the transaction joining is performed by a service, it might be required to trust a third-party. On the other hand, if the transaction is created in peer-to-peer network, it might be susceptible to DoS attacks. In any case, it works without modifications to Bitcoin and without necessarily relying on new crypto-graphic methods.

² https://en.wikipedia.org/wiki/Mt._Gox

A popular centralized implementation is the **SharedCoin**³ service. While it can not steal coins from users, it can link the input and output addresses. Thus, if the service is compromised, all anonymity gains are lost.

CoinShuffle [RMSK14] provides a decentralized solution based on a peer-to-peer network. Similar to Tor⁴, users encrypt their output address multiple times with the keys of their participants. This ensures, that no one learns the connection between input and output addresses. An optional blame phase is used to protect against DoS attacks.

4.3 CoinSwap

CoinSwap [Ma13b] is another proposal of Gregory Maxwell to perform a transaction through a third party. Instead of Alice transferring coins directly to Bob, she sends the coins to Carol who in turn sends them to Bob. The transactions between Alice and Carol and Carol and Bob are escrow transactions that can be spent with a redeeming transaction that is protected by a hash-lock. This ensures that neither Alice nor Carol can steal coins.

CoinSwap is usable on Bitcoin today. It can even be used to perform transactions across different chains. However, the anonymity does depend on all 2of2 escrow transactions going on at the same time. Furthermore, it increases the number of needed transactions.

4.4 CryptoNote and Monero

CryptoNote [vS13] describes a new crypto currency concept. While the basic structure of transactions and the block-chain is the same as in Bitcoin, address derivation and signature generation make use of new cryptographic methods. When transferring coins, the sender A calculates a new receiver address based on the public key B of the receiving party. The matching private key can only be calculated by the owner of the private key B. To spent coins, the transaction output is signed with a one-time ring signatures. These signatures can be verified against a set of public keys without revealing the actually used private key. The most successful implementation to date is **Monero**⁵.

CryptoNote provides anonymity for the sender and the receiver. As it is not a mixing service but a completely new currency, it is not susceptible to DoS attacks. However, it is not compatible with Bitcoin without introducing breaking changes. Furthermore, it relies on new cryptographic methods like one-time key pairs and one-time ring signatures.

4.5 Zerocoin and Zerocash

Zerocoin [Mi13] implements a new crypto currency atop of Bitcoin. It extends Bitcoin by new transaction types, that mint and spent a new sort of coins. The spending of these new

³ <https://sharedcoin.com/>

⁴ <https://www.torproject.org/>

⁵ <https://getmonero.org/>

	Bitcoin compatible architecture		Bitcoin cryptography theft impossible single transaction remarks			
Mixcoin	✓	service	✓	✗	✗	anonymity depends on simultaneous users, mixer can link addresses
Blindcoin	✓	service	✓	✗	✗	anonymity depends on simultaneous users
CoinJoin	✓	-	-	✓	✗	details left to implementation, improves overall anonymity in Bitcoin
SharedCoin	✓	service	✓	✓	✗	mixer can link addresses
CoinShuffle	✓	p2p	✓	✓	✗	-
CoinSwap	✓	p2p	✓	✓	✗	allows transactions across chains/altcoins
CryptoNote	✗	altcoin	✗	✓	✓	needs transactions with same amount
Zerocoin	✗	altcoin	✗	✓	✗	large transaction overhead compared to Bitcoin
Zerocash	✗	altcoin	✗	✓	✓	also anonymizes the transaction amount

Tab. 1: Comparison of existing approaches

coins can not be linked to the minting and thus provides anonymity. To prevent double spending, an accumulator of commitments is used. When a coin is spent, a non-interactive zero knowledge proof is used to prove that one such commitment is known. A serial number linked to the commitment ensures that each commitment can only be spent once.

While Zerocoin intended to be used with Bitcoin, it would require breaking changes to Bitcoin. Furthermore, the generated transactions are quite large ($>10\text{KB}$) compared to traditional transactions. One instance of Zerocoin can also only support coins of one value. To support more than one value, several instances would have to be run on simultaneously.

Zerocash [Sa14] improves on Zerocoin by allowing any amount. It is able to hide the origin, destination and amount of a transaction. Compared to Zerocoin it also performs better by reducing the transaction size and time spend on verification. However, it still requires breaking changes to Bitcoin and similar to Zerocoin needs a trusted party to setup public parameters of the protocol.

5 Conclusion

Bitcoin is a new successful approach to crypto currency but does not guarantee anonymity. Figure 1 highlights the relevant properties of current approaches for anonymizing Bitcoins.

Services like Mixcoin and Blindcoin do not require modifications to Bitcoin and are easier to implement than decentralized approaches. However, they do not prevent theft like CoinJoin implementations or CoinSwap. All of them are unable to hide the transaction amount and therefore require extra transactions with fixed amounts. This increases transaction delays and costs. It may also require more blocks in the chain and thus raise the

energy needed by the Bitcoin network. More recent altcoins provide higher anonymity but are incompatible with Bitcoin and introduce overhead. Furthermore, they also rely on new implementations of new cryptographic methods, that might not be trusted by everybody.

For current usage with Bitcoin, an implementation of the CoinJoin concept is the most promising approach. A peer-to-peer implementation like CoinShuffle can be added to existing Bitcoin wallets and used opportunistically. This will increase the anonymity of participants and other Bitcoin users by breaking the assumption in [RH13].

We think that future research on CoinJoin transactions with arbitrary values and whether they can increase anonymity is needed. This would allow making payments while simultaneously mixing and therefore reduce the number of overall transactions, fees payed and energy consumed. One such approach could be “confidential transactions”⁶ that hide the amount but can currently not be implemented in Bitcoin in a backwards compatible way.

References

- [BdL13] Bergstra, Jan A; de Leeuw, Karl: Bitcoin and beyond: exclusively informational monies. arXiv preprint arXiv:1304.4758, 2013.
- [Bo14] Bonneau, Joseph; Narayanan, Arvind; Miller, Andrew; Clark, Jeremy; Kroll, Joshua A; Felten, Edward W: Mixcoin: Anonymity for Bitcoin with accountable mixes. In: Financial Cryptography and Data Security, pp. 486–504. Springer, 2014.
- [Bo15] Bonneau, Joseph; Miller, Andrew; Clark, Jeremy; Narayanan, Arvind; Kroll, Joshua A; Felten, Edward W: Research Perspectives and Challenges for Bitcoin and Cryptocurrencies. Technical report, Cryptology ePrint Archive, Report 2015/452, 2015.
- [Ma13a] Maxwell, Gregory: , CoinJoin: Bitcoin privacy for the real world, 2013.
- [Ma13b] Maxwell, Gregory: CoinSwap: transaction graph disjoint trustless trading. CoinSwap: Transactiongraphdisjointtrustlesstrading, 2013.
- [Mi13] Miers, Ian; Garman, Christina; Green, Matthew; Rubin, Aviel D: Zerocoin: Anonymous distributed e-cash from bitcoin. In: Security and Privacy (SP), 2013 IEEE Symposium on. IEEE, pp. 397–411, 2013.
- [Na08] Nakamoto, Satoshi: , Bitcoin: A peer-to-peer electronic cash system, 2008.
- [RH13] Reid, Fergal; Harrigan, Martin: An analysis of anonymity in the bitcoin system. Springer, 2013.
- [RMSK14] Ruffing, Tim; Moreno-Sanchez, Pedro; Kate, Aniket: CoinShuffle: Practical decentralized coin mixing for Bitcoin. In: Computer Security-ESORICS 2014. Springer, 2014.
- [Sa14] Sasson, E. B.; Chiesa, A.; Garman, C.; Green, M.; Miers, I.; Tromer, E.; Virza, M.: Zerocash: Decentralized Anonymous Payments from Bitcoin. In: 2014 IEEE Symposium on Security and Privacy. 2014.
- [VR15] Valenta, Luke; Rowan, Brendan: Blindcoin: Blinded, accountable mixes for bitcoin. In: Financial Cryptography and Data Security, pp. 112–126. Springer, 2015.
- [vS13] van Saberhagen, Nicolas: Cryptonote v 2.0. 2013.

⁶ https://people.xiph.org/~greg/confidential_values.txt

Ein Eclipse-Plugin zur kontrollierten Schema-Evolution im NoSQL Datenbanksystem MongoDB

Dennis Schmidt¹

Abstract: Schemafreie NoSQL-Datenbanken wie MongoDB bieten in der Softwareentwicklung große Flexibilität. Dies ist vor allem in der agilen Softwareentwicklung sehr nützlich, bei der die Software früh veröffentlicht und in kurzen Abständen aktualisiert wird. Des Weiteren werden in der professionellen Anwendungsentwicklung gerne Objektmapper als Bindestück zwischen einer objektorientierten Sprache und dem Datenbankmodell verwendet. Durch einen Objektmapper wird über die Klassendeklaration der zu speichernden Objekte ein Schema in die eigentlich schemafreie NoSQL-Datenbank impliziert. Durch häufige Aktualisierungen der Software ergeben sich große Probleme beim Verändern des Schemas bzw. der Klassendeklarationen (Schema-Evolution): Während sich die Klassendeklaration verändert, behalten bereits gespeicherte Objekte ihr Schema bei und werden anschließend über die neue Klassendeklaration geladen. Dadurch kann es zu Datenverlust oder gar Laufzeitausnahmen kommen.

In diesem Beitrag wird ein Eclipse-Plugin vorgestellt, das den Entwickler bei der kontrollierten Schema-Evolution mit dem Objektmapper Morphia für MongoDB unterstützt. Veränderungen an Klassendeklarationen werden mit früheren Veröffentlichungen einer Software verglichen. Dabei werden mögliche Probleme erkannt und passende Lösungsvorschläge angeboten.

Keywords: Schemafreie NoSQL-Datenbanken, MongoDB, Objektmapper, Schema-Evolution

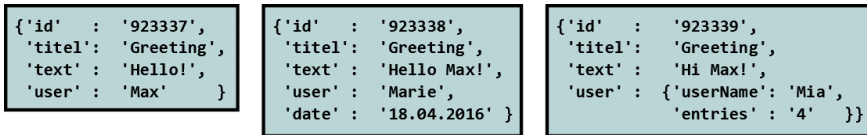
1 Einleitung

NoSQL-Datenbanken wie MongoDB verzichten im Gegensatz zu relationalen Datenbanken auf ein festes Schema für die gespeicherten Daten. Durch diese Flexibilität werden sie vor allem in der agilen Softwareentwicklung gerne eingesetzt. In einem solchen Entwicklungsprozess wird die Software in einem frühen Stadium veröffentlicht und dann in kurzen Abständen aktualisiert [Hü07]. Durch die häufigen Aktualisierungen ist es schwer zu Beginn ein Schema zu definieren, das in den folgenden Versionen gleich bleiben kann. MongoDB gehört zu den dokumentenorientierten Datenbanktypen, wobei *JSON* ähnliche Dokumente verwaltet werden. Ein Eintrag in MongoDB heißt Dokument und ist eine Datenstruktur aus Feld-Wert Paaren [RW12]. In MongoDB *Collections* werden üblicherweise Dokumente gruppiert, die denselben oder einen ähnlichen Zweck erfüllen. Dabei schreiben die *Collections* kein festes Schema für die gespeicherten Dokumente vor.

Die Abbildung 1 gibt im *JSON*-Format drei unterschiedlich strukturierte Dokumente in derselben *Collection* an. Durch die Schemafreiheit von MongoDB ist es kein Problem, unterschiedlich strukturierte Dokumente in derselben *Collection* zu speichern und zu laden. In der professionellen Anwendungsentwicklung werden gerne Objektmapper verwendet,

¹ Ostbayerische Technische Hochschule Regensburg, Fakultät für Informatik und Mathematik, dennis.schmidt@st.oth-regensburg.de

um ein Objekt einer Programmiersprache bequem in der Datenbank zu speichern und zu laden. Der Objektmapper übernimmt dabei das Erstellen eines passenden Datenbankobjekts und das Abbilden der Attribute auf dieses. Morphia ist ein Objektmapper für MongoDB, der ein gewöhnliches Java-Objekt auf ein Dokument für MongoDB abbildet [Mo15].



```

{ 'id' : '923337',
  'titel': 'Greeting',
  'text' : 'Hello!',
  'user' : 'Max' }

{ 'id' : '923338',
  'titel': 'Greeting',
  'text' : 'Hello Max!',
  'user' : 'Marie',
  'date' : '18.04.2016' }

{ 'id' : '923339',
  'titel': 'Greeting',
  'text' : 'Hi Max!',
  'user' : { 'userName' : 'Mia',
             'entries' : '4' } }

```

Abb. 1: Drei unterschiedlich strukturierte Dokumente in derselben *Collection*

Im Folgenden wird eine Gästebuch-Software betrachtet, die in der agilen Softwareentwicklung mit MongoDB als NoSQL-Datenbank und Morphia als Objektmapper realisiert wird (Abbildung 2). Im oberen Bereich der Abbildung wird die Deklaration der Klasse „*Guestbook*“ in der Entwicklungsumgebung dargestellt. Mit *@Entity* wird die Klasse zum Speichern markiert und mit *@Id* wird das eindeutige Id-Feld des Dokuments markiert. Beim Speichern eines Objekts dieser Klasse erstellt der Objektmapper ein passendes Dokument, indem es die Attribute auf Felder des Dokuments abbildet. Der Klassename wird dabei zum Namen der *Collection* und die Namen der Attribute werden zu den Namen der Felder innerhalb des Dokuments. Die gespeicherten Dokumente werden im JSON-Format im unteren Bereich der Abbildung in der Produktionsumgebung dargestellt. Durch die agile Softwareentwicklung wird die erste Version früh veröffentlicht. Der erste Nutzer „*Max*“ verfasst einen Eintrag mit der Gästebuch-Software. Dabei wird zuerst ein Objekt der Klasse „*Guestbook*“ erzeugt und dann mit dem Morphia-Aufruf *save* abgespeichert. Dies wird mit einem Pfeil von der ersten Version der Klasse „*Guestbook*“ nach dem erzeugten Dokument dargestellt. Wird dieses Dokument mit Morphia geladen, wird ein Objekt der Klasse „*Guestbook*“ erstellt und die Felder werden auf Attribute abgebildet. Nun wird eine zweite Version der Software entworfen und anschließend veröffentlicht. Dabei hat der Entwickler die Klassendefinition verändert: Das Attribut „*text*“ wurde auf „*content*“ umbenannt. Werden nun Objekte dieser Klasse abgespeichert, haben diese das Feld „*content*“. Neu abgespeicherte Dokumente haben jetzt die Struktur der zweiten Klassendeklaration und werden über diese erfolgreich geladen. Durch die Schemaflexibilität von MongoDB können die neu erzeugten Dokumente in derselben *Collection* gespeichert werden, obwohl sich die Struktur von denen der älteren Dokumente unterscheidet. Das Problem ist nun, dass auch ältere Dokumente, die bereits in der Datenbank gespeichert wurden, mit der neuen Klassendeklaration geladen werden. Dies wird mit einem Pfeil zwischen dem Dokument mit der alten Struktur und der zweiten Version der Klassendeklaration dargestellt. Zum Laden des Dokuments wurde zum Beispiel der Morphia-Aufruf *get* verwendet. Dabei erstellt der Objektmapper ein Java-Objekt, wobei das Feld „*text*“ nicht mehr zugeordnet werden kann. Das Attribut „*content*“ hat kein passendes Feld in dem geladenen Dokument und wird deshalb ohne Wert initialisiert. Der Wert von „*text*“ ist also nicht mehr abrufbar. Wird dieses geladene Dokument gespeichert, wird das alte Dokument überschrieben und der Wert von „*text*“ ist verloren. Dies wird mit einem Pfeil zwischen der zweiten Version der Klassendeklaration und dem Dokument rechts unten dargestellt. Das Dokument wird mit der neuen Klassendefinition abgespeichert, das heißt, das spei-

cherte Dokument besitzt nun statt „text“ das Feld „content“ ohne Wert.

Bei einem Objektmapper ist also die Klassendefinition der zu speichernden Objekte wichtig, da diese das Schema der Datenbankobjekte bestimmen. Nach dem Speichern werden die Datenbankobjekte über diese Klassendeklaration geladen. Obwohl viele NoSQL-Datenbanken kein festes Schema besitzen, implizieren Objektmapper also ein Schema über die Klassendeklaration der zu speichernden Objekte. Beim Verändern der Klassendeklarationen, auch genannt Schema-Evolution, kann es zu Probleme beim Laden älterer Datenbank-Objekte kommen.

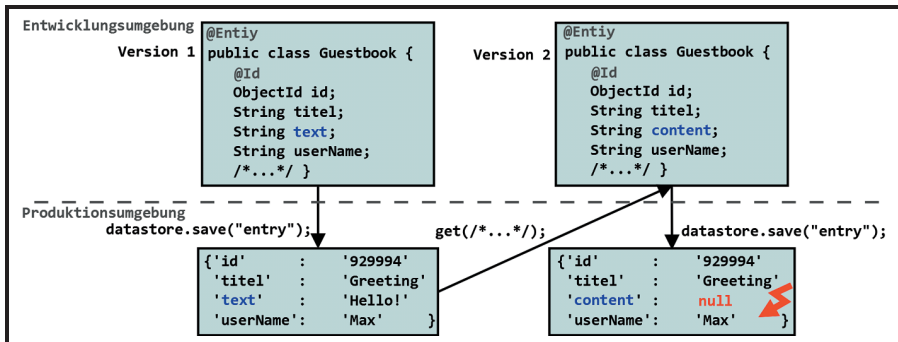


Abb. 2: Beispiel einer unkontrollierten Schema-Evolution: Umbenennung führt zu Datenverlust

Contributions:

1. In diesem Beitrag wird das erste Eclipse-Plugin vorgestellt, das Probleme bei der Schema-Evolution in MongoDB frühzeitig erkennt und über das *Quickfix*-Fenster passende Lösungsvorschläge anbietet.
2. Das Plugin wurde ursprünglich für den Objektmapper Objectify für den Google Datastore implementiert und wurde nun in seinem Funktionsumfang auf den Objektmapper Morphia und MongoDB, einer der bekanntesten NoSQL-Datenbanken, erweitert [CCS15, DB16].

Aufbau: In Kapitel 2 wird gezeigt, wie das Eclipse-Plugin das in der Einleitung angegebene Problem erkennt und Lösungsvorschläge anbietet. In Kapitel 3 wird das Konzept des Eclipse-Plugins beschrieben.

2 Die Plugin gestützte Schema-Evolution mit Morphia Annotationen

Im Folgenden wird gezeigt, wie das Eclipse-Plugin den Entwickler bei der kontrollierten Schema-Evolution unterstützt. Dabei wird auf das in Kapitel 1 gezeigte Problem eingegangen. Viele Objektmapper wie Morphia oder Objectify bieten spezielle Annotationen an, mit denen angegeben werden kann, wie auch ältere Objekte über eine neue Klassendeklaration verlustfrei geladen werden können. Diese werden in der jeweiligen Klassendeklaration der Datenbankobjekte platziert. Somit wird die Schema-Evolution den Softwareentwicklern überlassen. Diese haben möglicherweise wenig Erfahrung im Bereich Datenbanken, wodurch sich viele Fehlerquellen ergeben können. In Kapitel 1 wird beschrieben,

wie ein Entwickler in einer Gästebuch-Software ein Attribut ohne weiteres umbenennet und deshalb Datenverlust auftritt. Wird der Typ eines Attributs verändert, kann sogar eine Laufzeitausnahme auftreten, z.B. wenn der Typ von *String* auf *Integer* verändert wird.

Mit dem Plugin kann der Entwickler die erste Version der Gästebuch-Software als Veröffentlichung kennzeichnen. Wenn er anschließend das Attribut „*text*“ auf „*content*“ umbenennet, erkennt das Plugin, dass Datenverlust beim Laden älterer Datenbank-Objekte über diese Klassendeklaration auftreten kann und erstellt eine Markierung. In Bild 3 werden die Lösungsvorschläge angegeben. Das Plugin erkennt, welcher Objektmapper benutzt wurde und bietet deshalb Lösungsvorschläge mit Morphia Annotationen an, welche bei Auswahl automatisch so platziert werden, dass das jeweilige Problem gelöst wird. Wählt der Entwickler die Lösung mit der Annotation `@AlsoLoad("text")`, wird das Attribut „*content*“ damit markiert (Abbildung 4). Dokumente mit dem Feld „*content*“ werden

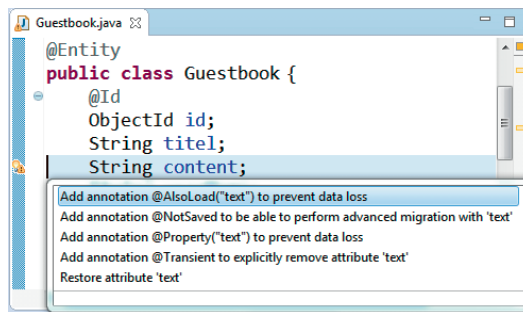
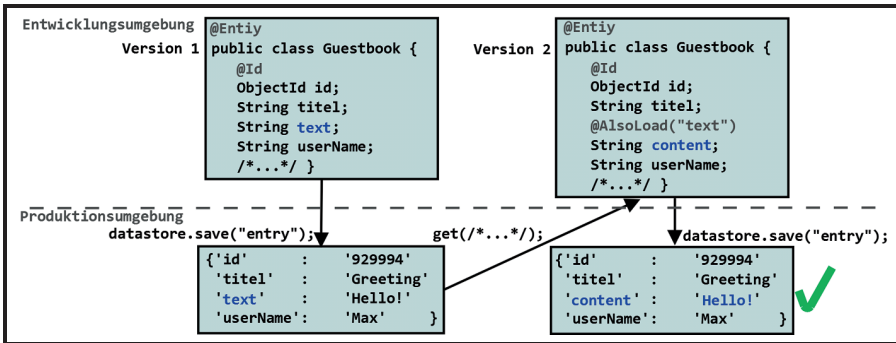


Abb. 3: Das *Quickfix*-Fenster bietet nach dem Umbenennen eines Attributs Lösungsvorschläge an

wie bisher geladen und gespeichert. Aber nun werden auch ältere Dokumente, die das Feld „*text*“ statt dem Feld „*content*“ besitzen verlustfrei geladen. Wird ein Dokument mit dem Feld „*text*“ geladen, wird der Wert von „*text*“ nach „*content*“ kopiert. Beim anschließenden Speichern dieses Dokuments wird „*text*“ verworfen, da es in der Klassendeklaration nicht mehr vorhanden ist. Dafür wird das Attribut „*content*“ mit dem Wert von „*text*“ gespeichert. Der Entwickler hat außerdem die Möglichkeit, das Attribut „*content*“ mit `@Property("text")` zu markieren. Damit wird das Attribut „*content*“ unter dem Bezeichner „*text*“ gespeichert und geladen. Das heißt, das Datenbank-Objekt behält das Feld „*text*“ bei, während in der Entwicklungsumgebung das Attribut „*content*“ oder ein anderer beliebiger Bezeichner benutzt werden kann. Außerdem gibt es die Möglichkeit, das Umbenennen rückgängig zu machen oder den alten Bezeichner als eigenes Attribut mit der Annotation `@Transient` bzw. `@NotSaved` explizit zu entfernen. Bei `@Transient` wird das Attribut nicht mehr gespeichert oder geladen. Bei `@NotSaved` wird das Attribut weiterhin geladen, was für eine komplexe Migration nützlich ist. Dafür kann eine Methode mit der Annotation `@PostLoad` definiert werden, wodurch diese nach dem Laden eines Objekts dieser Klasse ausgeführt wird. In dieser kann beispielsweise überprüft werden, ob ein Wert bei „*text*“ geladen wurde, um diesen nach einer Typkonvertierung nach „*content*“ zu kopieren. Wird „*content*“ auf „*post*“ umbenannt, kann überprüft werden ob ein Wert bei „*text*“ oder „*content*“ geladen wurde, um diesen nach „*post*“ zu kopieren, da mehrere `@AlsoLoad` Annotationen bei einem Attribut nicht kombiniert werden können.

Abb. 4: Beispiel einer kontrollierten Schema-Evolution mit der Annotation `@AlsoLoad`

3 Das Konzept des Plugins zur kontrollierten Schema-Evolution

Damit das Plugin mögliche Probleme erkennen kann, speichert es alle Attribute von Klassen, die in die Datenbank abgebildet werden, zusammen mit der Information über eventuell vorhandene Annotationen. Das Plugin erkennt anhand der vorhandenen Bibliotheken, welcher Objektmapper verwendet wurde. Sind Morphia-Bibliotheken vorhanden, benutzt das Plugin Morphia-spezifische Typprüfungsregeln, die definieren, welche Veränderungen an Klassendeklarationen erlaubt sind [Sc16]. Dabei wird die aktuelle Version der Deklaration eines Attributs paarweise mit allen älteren Deklarationen verglichen. Es gibt eine schreibende Klassendeklaration C_W und eine lesende Klassendeklaration C_R , die ohne Datenverlust oder Laufzeitausnahme lesen kann. Beispielsweise gibt es die Regel 1, die definiert, dass der Typ eines Attributs, welches mit `@Transient` markiert ist, verändert werden darf, da dieses Attribut nicht mehr geladen oder gespeichert wird.

$$\frac{C_W \vdash \text{type}_1 \text{ att};}{C_R \circ C_W \vdash @\text{Transient type}_2 \text{ att}; \text{ OK}} \quad (1)$$

Weitere Regeln geben an, dass jedes Attribut, das jemals in die Datenbank abgebildet wurde entweder als eigenes Attribut, explizit mit `@NotSaved` bzw. `@Transient` entfernt oder innerhalb einer `@Property` oder `@AlsoLoad` Annotation eines Attributs vorhanden sein muss. Das Plugin unterscheidet zwischen drei Problemkategorien: Das Umbenennen, das Entfernen und das Ändern in einen inkompatiblen Typ. In Bild 3 werden die Lösungen für das Umbenennen gezeigt. Wurde beim Umbenennen nicht das `Refactoring`-Werkzeug von Eclipse verwendet, erkennt das Plugin diese Änderung nur als Löschen und neu Hinzufügen eines Attributs. Genauso verhält es sich, wenn bei einem Attribut, welches umbenannt wird, bereits eine `@AlsoLoad` vorhanden ist, da mehrere `@AlsoLoad` Annotationen nicht miteinander verknüpft werden können. Wurde ein Attribut gelöscht, kann der Entwickler dieses wiederherstellen bzw. mit `@Transient` oder `@NotSaved` explizit entfernen. Wurde der Typ in einen inkompatiblen Typ geändert, hat der Entwickler die Möglichkeit die Änderung rückgängig zu machen oder dieses mit `@Transient` auszuschließen (Regel 1). Außerdem kann er das ursprüngliche Attribut mit `@NotSaved` markieren und ein Methoden-Gerüst zur Typkonvertierung einfügen. Dadurch, dass das Plugin die aktuelle

Version mit allen älteren Versionen vergleicht, erkennt es auch, ob ein Attribut in einer früheren Version gespeichert, anschließend über mehrere Versionen mit *@Transient* entfernt und danach mit einem inkompatiblen Typ wieder hinzugefügt wurde.

Interaktive Demo

1. Es wird eine Gästebuch-Software mit MongoDB als NoSQL-Datenbank vorgestellt. Diese besitzt die in Abbildung 2 links oben angegebene Klassendeklaration.
2. Nun kennzeichnen wir mit dem Plugin die Software als Veröffentlichungsversion. Anschließend speichern wir ein Objekt der Klasse mit dem Objektmapper Morphia in der Datenbank und laden dieses erfolgreich.
3. Danach nennen wir ein Attribut der Klasse um. Dabei erzeugt das Plugin eine Warnung. Nun laden wir das vorhin gespeicherte Objekt, ohne die Warnung zu beachten. Dabei können wir beobachten, dass das veränderte Attribut aus der Datenbank nicht mehr geladen wird und ein Datenverlust auftritt.
4. Das Plugin bietet mehrere Lösungsvorschläge an. Wir wählen den Lösungsvorschlag mit der Annotation *@AlsoLoad*. Danach laden wir das Objekt ein weiteres Mal. Dabei können wir beobachten, dass das Objekt jetzt erfolgreich geladen wird.

Zusammenfassung

Viele Anwendungen nutzen die Flexibilität, die NoSQL-Datenbanken durch ihre Schemafreiheit bieten. Doch werden Anwendungen zusammen mit Objektmappern entwickelt, kann es zu Probleme bei der Schema-Evolution kommen. Um Datenverlust und Laufzeitausnahmen beim Laden älterer Objekte zu verhindern, müssen Annotationen richtig angewendet werden. In diesem Beitrag wird ein Eclipse-Plugin vorgestellt, das Probleme frühzeitig erkennt und passende Lösungsvorschläge anbietet.

Danksagung: Ein besonderer Dank gilt Prof. Dr. Stefanie Scherzinger, welche die zugehörige Bachelorarbeit betreute [Sc16].

Literaturverzeichnis

- [CCS15] Cerqueus, T.; Cunha de Almeida, E.; Scherzinger, S.: Safely Managing Data Variety in Big Data Software Development. In: Proc. BIGDSE'15. 2015.
- [DB16] DB-Engines Ranking, <http://db-engines.com/de/ranking>, Stand: 02.05.2016.
- [Hü07] Hüttermann, M.: Agile Java-Entwicklung in der Praxis. O'Reilly, 2007.
- [Mo15] Morphia, <http://mongodb.github.io/morphia/>, Stand: 02.05.2016.
- [RW12] Redmond, E.; Wilson, R.: Seven Databases in Seven Weeks: A Guide to Modern Databases and the NoSQL Movement. Pragmatic Bookshelf, 2012.
- [Sc16] Schmidt, D.: Ein Eclipse-Plugin zur kontrollierten Schema-Evolution im NoSQL Datenbanksystem MongoDB. Bachelorarbeit an der Ostbayerischen Technischen Hochschule Regensburg, 2016.

A Learning Algorithm Based on λ -Policy Iteration and Its Application to the Video Game "Tetris Attack"

Le Hoang Thanh¹

Abstract: We present an application of the λ -policy iteration, an algorithm based on neuro-dynamic programming (described by Bertsekas and Tsitsiklis [BT96]) to the video game Tetris Attack in the form of an automated player. To this end, we first introduce the theoretical foundations underlying the method and model the game as a dynamic programming problem. Afterwards, we perform multiple experiments using an approximate version with different methods and analyze their results.

Keywords: Neuro-Dynamic Programming, Reinforcement Learning, Artificial Intelligence in Video Games

1 Introduction to Tetris Attack



Fig. 1: Stack with colored blocks

Tetris Attack is a puzzle video game released for the Super Nintendo Entertainment System. The game is played with a stack of colored blocks (see figure 1) where new blocks steadily appear at the bottom of the screen. These can be cleared by switching blocks with a cursor and arranging them in a way where at least three same-colored blocks are next to each other.

¹Hochschule Anhalt, Fachbereich Informatik und Sprachen, Lohmannstraße 23, 06366, Köthen, thanh_le_hoang@web.de

The goal of this game is to maximize the points gained from clearing blocks until the round ends which happens in this game's Time Attack mode when the stack reaches the upper border of the screen (called "game over") or when a time limit of two minutes runs out ("time over"). Additional points are awarded when a greater number of blocks disappears simultaneously ("combo") or in succession ("chain"). Due to this, reaching a high score requires planning in order to facilitate the application of those advanced techniques.

2 Basic Concepts of Dynamic Programming

The following introduction is based on "Neuro-Dynamic Programming" (Bertsekas, Tsitsiklis) [BT96] where the following concepts are described in greater detail.

A *dynamic programming problem* (short: DP problem) involves states $i \in \{1, 2, \dots, n\}$ (here we assume that the state space is discrete and finite) and controls $u \in U(i)$ where $U(i)$ stands for the set of permitted controls in state i . Choosing a control u in state i leads to a successor state j (which does not necessarily have to be different from i). In the case that the underlying system is nondeterministic (i.e. it is possible that the same control in the same state leads to different successor states), we refer to $p_{ij}(u)$ as the probability that the control u in state i leads to state j . Each state transition $i \rightarrow j$ using the control u incurs a certain cost, noted by $g(i, u, j)$.

A special type of DP problem is the *stochastic shortest path problem* where the state space contains an absorbing state 0 with the properties

$$p_{00}(u) = 1 \quad \text{and} \quad g(0, u, 0) = 0$$

for every control u . Based on stochastic shortest path problems, we define a policy μ as a mapping that assigns a control $\mu(i) \in U(i)$ to each state i . For a given policy μ , we define the *cost-to-go* $J^\mu(i)$ as

$$J^\mu(i) = E \left[\sum_{k=0}^{\infty} g(i_k, \mu(i_k), i_{k+1}) \mid i_0 = i \right] \quad (1)$$

with the convention $J^\mu(0) = 0$. This value can be interpreted as the expected cost incurred from state i onwards if μ is used as the policy. The vector $J^\mu = (J^\mu(1), J^\mu(2), \dots, J^\mu(n))$ is referred to as the *cost-to-go vector*. Even though the above summation is infinite, the cost-to-go can be finite if the state 0 is reached along the trajectory (i_0, i_1, i_2, \dots) since the terms summed after this state are $g(0, u, 0) = 0$. Equation (1) can also be written in a recursive form known as the *Bellman equation*:

$$\begin{aligned} J^\mu(i) &= E [g(i, \mu(i), j) + J^\mu(j)] \\ &= \sum_{j=0}^n p_{ij}(\mu(i)) (g(i, \mu(i), j) + J^\mu(j)) \end{aligned}$$

The goal of dynamic programming is to find an *optimal policy* μ^* which chooses controls u for each state i in a way that minimizes the expected cost starting from i . Its cost-to-go vector J^* is the only vector [BT91] that satisfies

$$J^*(i) = \min_{u \in U(i)} E[g(i, u, j) + J^*(j)] = \min_{u \in U(i)} \sum_{j=0}^n p_{ij}(u) (g(i, u, j) + J^*(j)). \quad (2)$$

If the optimal costs-to-go J^* are known for every state, the optimal policy $\mu^*(i)$ can be calculated as the control u that minimizes the right-hand side of equation (2). There are different ways to determine J^* . The approach used here is called λ -policy iteration which is described in the following section.

3 λ -Policy Iteration

Similar to the normal policy iteration, which calculates a sequence of policies μ_1, μ_2, \dots and their corresponding cost-to-go values $J^{\mu_1}, J^{\mu_2}, \dots$, the λ -policy iteration also generates a sequence of policies, but replaces the exact cost-to-go values with estimates $\hat{J}^{\mu_1}, \hat{J}^{\mu_2}, \dots$. To describe the algorithm, we use *temporal differences*

$$d(i, j) = g(i, \mu(i), j) + J(j) - J(i),$$

which can be interpreted as the difference (or "error") between the cost-to-go estimate for state i and the sum of the actually incurred cost $g(i, \mu(i), j)$ and the estimate $J(j)$ for the successor state j .

The λ -policy iteration for a given policy μ and an estimate of its cost-to-go J consists of two steps [BT96]:

1. *Policy evaluation*: Calculate the cost-to-go estimate $\hat{J}^\mu(i)$:

$$\hat{J}^\mu(i) = J(i) + \sum_{m=0}^{\infty} E[\lambda^m d(i_m, i_{m+1}) \mid i_0 = i] \quad (3)$$

This equation can be interpreted as a correction of the given cost-to-go values J using the term $\sum_{m=0}^{\infty} E[\lambda^m d(i_m, i_{m+1}) \mid i_0 = i]$ which stands for the expected total estimation error. Here we incorporate a factor $\lambda \in [0, 1]$ that discounts future temporal differences. It can be shown [BT96] that this step is identical to the policy evaluation step of the normal policy iteration if $\lambda = 1$.

2. *Policy improvement*: Generate a new policy $\bar{\mu}$:

$$\bar{\mu}(i) = \arg \min_{u \in U(i)} \sum_{j=0}^n p_{ij}(u) (g(i, u, j) + \hat{J}^\mu(j))$$

This step is very similar to the policy improvement step of the normal policy iteration (where J^μ is used instead of \hat{J}^μ).

It can be shown that the sequences μ_1, μ_2, \dots and $\hat{J}^{\mu_1}, \hat{J}^{\mu_2}, \dots$ converge to the optimal policy μ^* and its cost-to-go vector J^* [Bel2].

4 Application to Tetris Attack

The modeling of the video game as a stochastic shortest path problem is as follows: The state (a stack of colored blocks) is represented by a (6×11) -matrix, while a control u represents a switch at a position (x, y) . The absorbing state 0 corresponds to the blocks touching the top of the screen or the time limit running out. The goal of minimizing cost is replaced by maximizing the reward (making J^μ the *reward-to-go*). The one-step reward $g(i, u, j)$ here is calculated beforehand by replicating the game's scoring system while the modeling of the state transitions is based on the game's rules.

Similar to the "Tetris" case study in [BT96], we implemented an approximate version of the λ -policy iteration (also called *approximate λ -policy iteration*). The algorithm is derived by reformulating (3) as a minimization problem and replacing the cost-to-go values $\hat{f}^\mu(i)$ (which so far have been assumed to be represented as a vector) by a *cost-to-go function* $\tilde{J}(i, \underline{r})$ that is characterized by a set of parameters \underline{r} and calculates a value for every state i using a compact representation. The policy is then chosen with respect to this function:

$$\bar{\mu}(i) = \arg \max_{u \in U(i)} \sum_{j=0}^n p_{ij}(u) (g(i, u, j) + \tilde{J}(j, \underline{r})) \quad (4)$$

The parameters \underline{r} of $\tilde{J}(\cdot, \underline{r})$ can be learned through systems simulation which means in this context that learning data is generated by letting the algorithm play the game. To describe this as an algorithm, let K be the number of games played with (old) parameters \underline{r} . Note that all games have finite length (with N_k being the length for the k th game and $i_{k, N_k} = 0$ for $k \in \{1, 2, \dots, K\}$) due to the nature of the Tetris Attack Time Attack mode (in the context of dynamic programming this means that all policies are proper). Let $i_{k, m}$ denote the m th state visited in the k th game. Based on the changes described above, the resulting minimization problem to calculate the new parameters \underline{r}' is

$$\underline{r}' = \arg \min_{\underline{r}'} \frac{1}{K} \sum_{k=1}^K \left\| \underline{y}_k - \underline{z}_k(\underline{r}') \right\|_2^2 \quad (5)$$

with

$$\underline{y}_k = \begin{pmatrix} \tilde{J}(i_{k,0}, \underline{r}) \\ \tilde{J}(i_{k,1}, \underline{r}) \\ \vdots \\ \tilde{J}(i_{k, N_k-1}, \underline{r}) \end{pmatrix} + \begin{pmatrix} \sum_{m=0}^{N_k-1} \lambda^{m-0} \cdot d(i_{k,m}, i_{k,m+1}) \\ \sum_{m=1}^{N_k-1} \lambda^{m-1} \cdot d(i_{k,m}, i_{k,m+1}) \\ \vdots \\ \sum_{m=N_k-1}^{N_k-1} \lambda^{m-(N_k-1)} \cdot d(i_{k,m}, i_{k,m+1}) \end{pmatrix}$$

and

$$\underline{z}_k(\underline{r}') = \begin{pmatrix} \tilde{J}(i_{k,0}, \underline{r}') \\ \tilde{J}(i_{k,1}, \underline{r}') \\ \vdots \\ \tilde{J}(i_{k, N_k-1}, \underline{r}') \end{pmatrix}.$$

Solving this problem corresponds to the policy evaluation step (3) while the equation (4) corresponds to the policy improvement after updating the cost-to-go function $\tilde{J}(i, \underline{r})$.

Based on the aforementioned "Tetris" case study, we used a linear architecture, i.e. $\tilde{J}(i, \underline{r}) = \underline{\phi}^T(i) \cdot \underline{r}$ with feature extractions $\underline{\phi}(i)$, in which case the problem (5) becomes a linear least squares problem. We implemented two algorithms for solving this least squares problem, the first being an iterative linear least squares method based on the Kalman filter and the second being the commonly used TD(λ) described in [BT96]. Note that the TD(λ) update

$$\underline{r}' = \underline{r} + \gamma \cdot \sum_{k=1}^K \sum_{m=0}^{N-1} \nabla \tilde{J}(i_{k,m}, \underline{r}) \sum_{s=m}^{N-1} \lambda^{s-m} d(i_{k,s}, i_{k,s+1}),$$

which is obtained by applying the steepest descent method to (5), can also be applied if $\tilde{J}(i, \underline{r})$ uses a nonlinear architecture. However, parameter convergence is only established for linear architectures [BT96].

5 Results

Using the methods described above, 1000 games were played multiple times with different values for λ and a two-step lookahead policy. An approximate λ -policy iteration was performed after K games with K being 100 or 200. The features $\underline{\phi}(i)$ used for the linear architecture were based on the "Tetris" case study in [BT96] as well as personal experience from the author with the game and involve information about stack column heights, smoothness at the top of the stack and different measures for the closeness of same-colored block structures.

As an example, figure 2 shows the results for $\lambda = 0.8$ and $K = 100$. The single (red) line stands for the percentage of games lost (i.e. ended by "game over"), the highest (blue) line for the average score over 100 games (ignoring lost games) while the lower (black) line shows the average score over 100 games where lost games are added with a score of 0. Aside from the average scores, the 95% confidence interval is also shown. The best performance regarding the average score is reached with the initial parameters \underline{r}_0 chosen from the author's intuition. Afterwards, the performance drops until the second iteration and slightly increases without reaching the initial parameters' score. After the fourth λ -policy iteration, the percentage of lost games increases which leads to a decrease in the average score. On the other hand, the average score without lost games slightly increases until the seventh iteration. Coupled with the increasing of lost games this could be interpreted as the agent learning to play in a riskier way that yields many points, but losing more often. Using different values for λ and replacing the method with TD(λ) yielded similar results. A demonstration of the Tetris Attack player can be found at

<https://youtu.be/8PvAQj9c3Tc>

As for the future of this project, improvements can be made in the choice of features $\underline{\phi}(i)$: This aspect provides a large potential for the recognition of complex block structures that can be cleared with techniques used by advanced players (especially combos or chains). The usage of other algorithms based on neuro-dynamic programming also opens up further possibilities for future research.

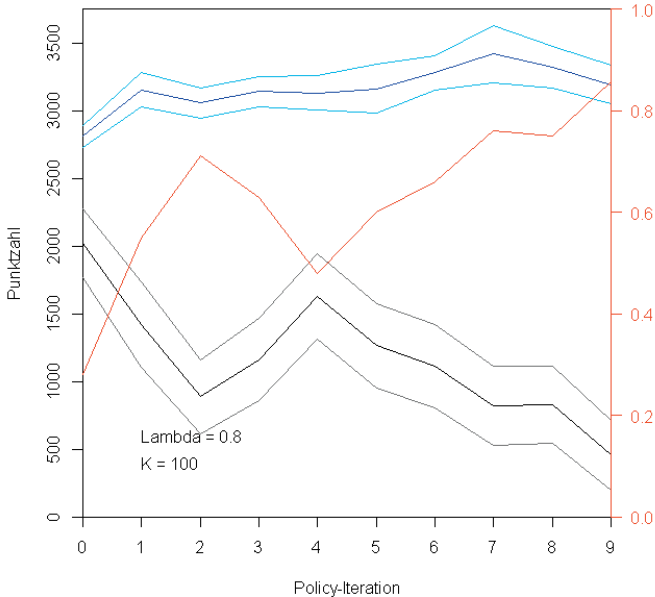


Fig. 2: Results of the approximate λ -policy iteration with the linear least squares method [R 14]. Every data point representing a score is averaged over 100 games and shown with its 95 % confidence interval [Ar12].

Acknowledgments

I would like to thank Bernd Krause for his patience and understanding regarding the supervision of my bachelor thesis this paper is based on. I also thank my co-supervisor Gunther Schwenzfeger for his support during and after the submission of the thesis.

References

- [Ar12] Arnholt, A.: BSDA: Basic Statistics and Data Analysis, R package version 1.01. <http://CRAN.R-project.org/package=BSDA>, 2012.
- [Be12] Bertsekas, D.: Lambda-Policy Iteration: A Review and a New Implementation, http://www.mit.edu/~dimitrib/Lambda_PI.pdf, 2012.
- [BT91] Bertsekas, D.; Tsitsiklis, J.: An Analysis of Stochastic Shortest Path Problems. *Mathematics of Operations Research*, pp. 580–595, 1991.
- [BT96] Bertsekas, D.; Tsitsiklis, J.: *Neuro-Dynamic Programming*. Athena Scientific, Belmont, Massachusetts, 1996.
- [R 14] R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>, 2014.

Overview of the current state of research on parallelisation of evolutionary algorithms on graphic cards

Paul Jähne¹

Abstract: Evolutionary algorithms (EA) have a lot of potential for parallelisation, which can be used by graphics processing units (GPU). They present an available, cheap and energy-efficient alternative to computer clusters. There are already several publications on using GPUs for EAs. This paper presents selected publications and discusses important implementation details of them. Hence recommendations are derived on how to efficiently implement EAs on GPUs. Thereby it's important to take the architecture of GPUs into consideration. Furthermore it's shown that GPUs can only be used profitably when a certain problem complexity is reached. In addition the speed-up to expect is critically scrutinised and it's explained, that it's impossible to reach a speed-up of 100 or even more by means of a fair comparison of GPU and CPU implementations.

Keywords: meta-paper, evolutionary algorithm, GPGPU, CUDA

1 Motivation

Often there are no efficient algorithms known for difficult problems or even impossible to construct due to their complexity. Then only general search methods remain to get a good approximate solution within acceptable time. Evolutionary algorithms (EA) are one option to construct such domain independent searches. Yet also these reach their limits at a certain problem size. Fortunately they have big potential for parallelisation. This can be utilised through the use of multiple processors (CPU) or computers. However clusters out of several computers are harder to use and maintain and also not generally available.

An alternative to this is the use of graphics processing units (GPU). They evolved from specialised graphic accelerators to flexible, highly parallel coprocessors and are available in every modern computer. Therefore they were recently used for many applications aside from graphics processing with partly remarkable results [Ow05][Nv16]. Since EAs have a lot of parallelisation potential, there were also efforts to use graphic cards in this domain. This paper presents selected works and gives an overview of the different approaches.

2 Evolutionary Algorithms

Evolutionary algorithms are powerful, domain independent search methods derived from the evolutionary theory. Their search strategy is based on stochastic methods. Because of this there is no guarantee to find the optimal solution, but in most cases sufficient approximate solutions are found. The process is as follows. At the beginning a random population

¹ Hochschule für Technik, Wirtschaft und Kultur, Fakultät Informatik, Mathematik und Naturwissenschaften, Karl-Liebknecht-Straße 132, 04277 Leipzig, paul.jaehne@gmx.de

is generated. This consists of possible solutions for the given problem, referred to as individuals. Based on this initial population new generations are created by repeating evaluation, selection, crossover and mutation. This process iterates until a sufficient solution is found or if the maximum number of generations is exceeded. The evaluation assigns a fitness value to all individuals, which describes the quality of the solution. Thereby a search direction is given. The selection chooses promising individuals for the next generation based on their fitness value. The crossover uses the chosen individuals to create new ones through mixing. The mutation alters some of the resulting individuals to retain the populations diversity and open up new areas of the search space.

Based on this principle the four following classical variants were developed: genetic algorithms (GA), evolutionary strategies (ES), genetic programming (GP) and evolutionary programming (EP). They differ in the type of individuals and the used operators. Exact differentiations are difficult, as the approaches are interacting and influencing each other [WW09]. All have in common, that calculation effort increases with increasing number of individuals, generations and operators. Yet the operators work only with few individuals at the same time and the evaluation can be executed independently for each individual. Therefore EAs have a lot of potential for parallelisation.

3 Programming graphic cards

Through the popularity of 3D consumer applications like video games a market for affordable hardware accelerators was established. They implemented parts of the rendering pipeline of application programming interfaces (API) like Open Graphics Library (OpenGL) and DirectX. Graphics calculations need to do a series of operations for every pixel on the screen. Therefore large computing power is required, but each pixel can be calculated independently in parallel. Thus graphic cards evolved into powerful and highly parallel coprocessors, while their price remained in the consumer segment [SK10].

This development attracted the attention of some scientists. Since GPUs calculate colour values out of textures, objects and additional information, which are specified by the programmer, any data is possible. Only a mapping between the problem, it's solution algorithm and graphics calculation is required and the result has to be interpreted accordingly. Therefore GPUs could be used for non-graphic computations. This is called general-purpose computing on graphics processing units (GPGPU) [SK10]. Yet the only way to interact with a GPU were graphics APIs in which the actual calculation has to be packaged.

In the year 2004 Brook² was a first attempt to solve these issues and to make GPGPU accessible for more people. Brook is an extension for the programming language C. It abstracts the graphics APIs and presents the programmer an environment, which he's used to from the CPU. Still various constraints existed on the GPU. These were resolved through architectural evolution of the GPU. In addition in the year 2007 CUDA³ and Open Computing Language⁴ (OpenCL) were released. These are low-level APIs specifically designed

² <http://graphics.stanford.edu/projects/brookgpu/>

³ <https://developer.nvidia.com/content/cuda-10>

⁴ https://www.khronos.org/news/press/the_khronos_group_releases_opengl_1.0_specification

for GPGPU. An high-level alternative emerged in 2011 with OpenACC⁵. OpenACC uses compiler directives to annotate parallelisable code segments for the GPU. These are processed by the compiler. So there is no need to rewrite existing code and the programmer doesn't has to bother with details. Also Open Multi-Processing⁶ (OpenMP) got support for coprocessors in 2013 with version 4.0. OpenMP is as well based on compiler directives, but was originally developed for multiprocessor systems with shared memory.

The rise of language extensions and the related simplification increased the adoption of GPGPU. The development of systems with coprocessors in the 500 most powerful computer systems, which are listed in TOP500, proves this. The development over time of systems using coprocessors in this list can be seen in figure 1.

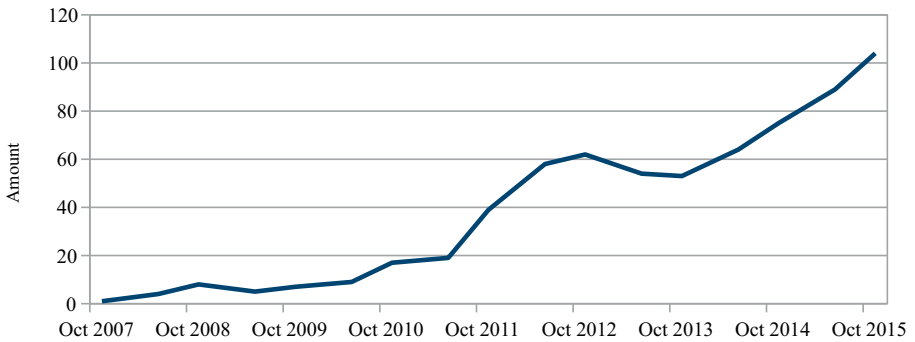


Fig. 1: Temporal development of the amount of systems using coprocessors in TOP500 as per [TO15]

The main limitations for systems in the TOP500 are space and energy consumption [Be08]. This is where graphic cards provide advantages compared to CPUs besides there computing power. Table 1 shows important features of a high-end server CPU and GPU with comparable release dates. It can be seen that the GPU has the eightfold computing power at twice the power consumption. Thus it has the fourfold efficiency. To compensate for the differences in price, power consumption and space, one could compare a graphic card with two such CPUs. Then a fourfold advantage in computing power remains.

Feature	Intel Xeon E5 2697 v2 ^a	Nvidia Tesla K40 ^b
Release date	September 10, 2013	November 18, 2013
Manufacturing process	22 nm	28 nm
Release price	2614 \$	5499 \$
Power dissipation	130 W	235 W
Computing power ^c	0.52 Tflop/s	4.29 Tflop/s
Processing power ^c /Watt	3.98 Gflop/Ws	18.3 Gflop/Ws

^a <http://ark.intel.com/products/75283>

^b http://www.nvidia.com/content/PDF/kepler/Tesla-K40-PCIe-Passive-Board-Spec-BD-06902-001_v05-en.pdf

^c For calculations in single precision

Tab. 1: Exemplary comparison of important features of CPUs and GPUs

⁵ <http://www.nvidia.co.uk/object/openacc-parallel-computing-standard-20111114-uk.html>

⁶ <http://openmp.org/wp/openmp-40-api-released/>

4 Early days

Until 2007 there were no programming languages or frameworks specifically designed for GPGPU. So graphic APIs were the only way to interact with a GPU. Because they were designed for graphic applications, the calculations had to be transformed into graphic calculations to satisfy their input and output format. Therefore several limitations existed like a maximum number of arrays and elements within these through the size and number of textures. Due to such restrictions GPUs were difficult to use for general computation. But when the high initial hurdles were overcome one could get a considerable speed-up compared to a CPU implementation, which showed the potential of GPUs [Ow05]. Hereafter are several papers from that period presented.

Chitty presents in [Ch07] an implementation of genetic programming with OpenGL and C for Graphics (Cg). Therein the fitness evaluation was offloaded to the GPU. GP isn't based on the principle Single Instruction, Multiple Data (SIMD) like the GPU, but rather on Multiple Instruction, Single Data (MISD). The problem is solved by running the fitness function on different example data in parallel. For comparison an Intel Pentium 4 and a Nvidia GeForce 6400 Go is used. The results show that the CPU is faster for small problems. Beginning with 200 samples the GPU outperforms the CPU. With a sufficiently large data set and complex examples a speed-up between 12 and 30 is reached.

Fok, Wong and Wong implement evolutionary programming on the GPU in [FWW07]. They chose EP, because crossover is not used therein and thus it is easier to implement. Random number generation and selection is done on the CPU and the results are provided as a texture to the GPU. The used operators are tournament selection and Cauchy mutation. The selection of individuals isn't done by sorting, but by searching the median of the tournament victories. Subsequently the population is traversed and all individuals with greater or equal amount of wins are chosen. For comparison an Intel Pentium 4 and a Nvidia GeForce 6800 Ultra is used. The GPU implementation is up to five times faster for different test functions with more than 800 individuals. The runtime increase is sublinear. This is an indication for underutilisation of the GPU. By increasing the problem size more computing units can be used. Hence the increase in runtime is sublinear.

Wong and Wong present a genetic algorithm based on [FWW07] in [WW09] which is executed on the GPU except for the random number generation. The used operators are tournament selection, single point crossover and Cauchy mutation. Different functions are used to test on an AMD Athlon 64 and a Nvidia GeForce 6800 Ultra. The GPU implementation is up to five times faster depending on complexity of the function and number of individuals.

Li et al. present in [Li07] a genetic algorithm, which runs completely on the graphic card. The used operators are single point crossover, single point mutation and local roulette wheel selection. Different test functions are used for comparison. The tests are carried out on an Intel Pentium 4 and a Nvidia GeForce 6800 LE. The GPU implementation reaches a speed-up up to 74 for large population sizes and complex functions. Also for small populations a small speed-up is achieved.

5 Modern era

The emergence of frameworks, language extensions and development tools specifically designed for GPGPU lowered the initial hurdles and therefore made it accessible to a wider audience. Furthermore graphic cards developed from fixed-function devices into flexible, highly parallel coprocessors, which resolved several limitations. This facilitated the implementation of more complex algorithms.

5.1 Fitness evaluation on the GPU

The simplest use of graphic cards is the implementation of the fitness evaluation on the GPU. This can be done entirely independent for each individual and is therefore suited for parallelisation according to the SIMD architecture of GPUs. An easy implementation can be seen in listing 1. There is an easy and general pattern for parallelisation of loops shown. `evaluateKernel` is a GPU function, which is also called kernel. Additionally the arrays `population` and `fitness` have to be transferred to the GPU for processing and back afterwards. This is not included in the following example.

```
C:
...
for (int i = 0; i < populationLength; i++) {
    fitness[i] = evaluate(population[i]);
}
...

CUDA:
__global__ void evaluateKernel(float *population,
    float *fitness, int populationLength) {
    for (int i = blockIdx.x * blockDim.x + threadIdx.x;
        i < populationLength; i += blockDim.x *
            gridDim.x) {
        fitness[i] = evaluate(population[i]);
    }
}
...
evaluateKernel<<<BLOCKS, THREADS>>>(population,
    fitness, populationlength);
...
```

List. 1: Parallel fitness evaluation in CUDA

The described method is easy to implement and can achieve a reasonable speed-up with sufficient individuals and complex fitness functions. If these conditions aren't fulfilled, the GPU isn't fully utilised or the data transfer takes more time, than what is saved through parallel execution.

An example for this strategy present Cavouti et al. in [Ca12]. They extend an existing implementation of a genetic algorithm for machine learning with CUDA and offload the creation of the initial population and the fitness evaluation to the GPU. For comparison they used an Intel Core i7-2630QM and a Nvidia GeForce GT540M. The classification results are the same, but the GPU reaches a speed-up of 75. However the CPU implementation only used one core out of four. Therefore the comparison is unfair in favour of the GPU and the achieved speed-up should be divided by four.

If one population isn't enough to keep the graphic card busy, parallel evaluation of multiple populations through streams is an option to provide enough work. Streams are one way to do different tasks in parallel on the GPU. All commands placed into one stream are executed sequentially while different streams run in parallel. So it is possible to run multiple small kernels on the GPU or to interleave copy processes with computations. Most CUDA functions provide an additional stream parameter for this purpose. An example for a kernel launch can be seen in listing 2.

```

...
    cudaStream_t stream1;
    cudaStreamCreate(&stream1);
    evaluateKernel<<<BLOCKS, THREADS, 0, stream1>>>(
        population, fitness, populationlength);
    cudaStreamDestroy(&stream1);
...

```

List. 2: Use of CUDA-streams on kernel call

The parallel evaluation of several small populations is used by Robilliard, Marion-Poty and Fonlupt in [RMPF08]. They notice that many typical problems are too small to keep a GPU busy. Therefore multiple small populations are evaluated in parallel. They don't use the previously shown approach with streams, but an explicit implementation within the kernel. Thereby the fitness evaluation is implemented on the GPU. The remaining operations are executed on the CPU with the ECJ⁷ library. The implementation is compared to a serial one using the ECJ library for all parts. The tests show a speed-up between 8 and 80 in the evaluation phase. This leads to a reduction of the overall runtime by a factor between 5 and 45. No speed-up is achieved for examples with many conditionals in the fitness function. This is due to the SIMD architecture of GPUs, which is unsuitable for branching.

5.2 Complete GPU-implementation

Even though the parallelisation of fitness evaluation on the GPU is profitable under certain circumstances, it is limited by the bandwidth of the interconnection between CPU and GPU, because the population has to be copied back and forth for each evaluation. One solution for this problem is a complete GPU implementation, which eliminates the need for frequent copies. Even if individual steps are slower on the GPU, the overall speed-up can

⁷ <http://cs.gmu.edu/~eclab/projects/ecj/>

be increased by removing the slow transfers. However a complete GPU implementation is more complex, because the algorithm has to be fitted to the GPU architecture, to use the full potential.

Debattisti et al. present a complete implementation of a genetic algorithm in CUDA in [De09]. The test problem is maximizing the number of bits set to one, which isn't computationally demanding. Therefore it is limited by the available bandwidth. The operators are tournament selection, two-point crossover and mutation via bitwise exclusive OR with a template. Random numbers are generated with the Mersenne Twister implementation out of the CUDA SDK examples. The implementation runs on a Nvidia GeForce 8800 GT and is compared with a sequential implementation, created with TinyGA on an Intel Core2Duo. A speed-up of 20 is reached for larger populations and individuals.

A critical examination to parallelisation of evolutionary algorithms on GPUs can be found in [JP12] by Jaroš and Pospíchal. They notice, that comparisons of GPU and CPU implementations often use less optimised or sequential versions for the CPU, although they have multiple cores available. Because of this they present and compare one implementation specifically tuned for each architecture. This means using vector instructions and multiple cores on the CPU and efficient memory access and low branching on the GPU. The paper describes the memory organisation and all operators in detail. The operators are tournament selection, uniform crossover and mutation via one bit flip. Additionally several statistics are gathered. Random numbers are generated via an algorithm based on hash functions from Salomon et al. published in [Sa11]. Both implementations are available on GitHub⁸. Large instances of the knapsack problem (10,000 items, 12,000 individuals) are used for comparison, to achieve maximal utilisation. They use an Intel Xeon X5650 and a Nvidia GTX 580 for their tests. The tests and there parameters are also described in detail. They additionally implement the same algorithm with the GALib⁹ library, which is often used for comparison in other papers. The results show that the GPU implementation reaches a speed-up of 375 compared to GALib. The speed-up was reduced to 221 when GALib is compiled with optimisations. The GPU reached a speed-up of 68 in comparison to the proposed CPU implementation with one thread. This value corresponds to the speed-up values reported in other papers, which as well use only one core. The speed-up decreases to 12 if all cores are used. This complies with the ratio of the theoretical peak performance for the CPU and GPU. Compared to two CPUs a speed-up of 6 is reached. Additionally the cache hit ratio is analysed. GALib reaches a hit ratio of 18 % and the proposed CPU implementation achieves 98 %. This shows how well they utilise the CPU architecture. Overall the paper shows that a GPU can't achieve a speed-up of 100 or even more, if an equally optimised CPU implementation is considered.

5.2.1 Cellular EA

In contrast to standard EAs cellular EAs use operators, which work locally. This means, that individuals are arranged in some structure and the operators apply to a certain neigh-

⁸ <https://github.com/jarosjir/GPU-GA-Knapsack> und <https://github.com/jarosjir/MPI-GA-Knapsack>

⁹ <http://lancet.mit.edu/ga/>

bourhood within this structure. Costly global communication and dependencies between threads are thereby reduced. An example is the Moore neighbourhood, where the eight surrounding cells in a 2D grid are considered as neighbours. A schematic representation of the field of action for the Moore neighbourhood can be seen in figure 2.

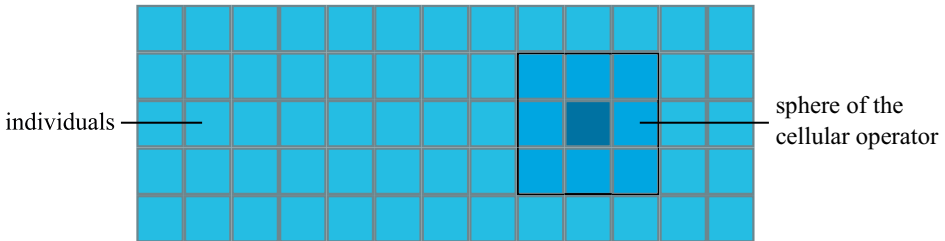


Fig. 2: Schematic presentation of the Moore neighbourhood for a cellular operator

An implementation can be found in [LL06] by Luo and Liu. They develop a genetic algorithm to solve the 3-SAT problem. The individuals are arranged in a 2-dimensional toroidal grid and the Moore neighbourhood is used. This grid structure avoids borders where individuals have fewer neighbours. They use an Intel Pentium 4 and a Nvidia GeForce 6200 for comparison. The results are comparable and the GPU reaches a speed-up of 5.

5.2.2 Island model

The island model tries as well to reduce the amount of global communication. This is also accomplished by the use of neighbourhood relationships. Separated subpopulations are used for this purpose. They only exchange few individuals in set intervals by the additional migration operator. This can lead to better solutions because the different populations are more likely to not get stuck at the same local optimum. Then again arbitrary models can be used within one island. Figure 3 shows a schematic representation.

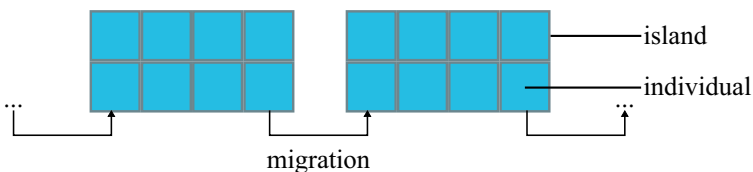


Fig. 3: Schematic presentation of the island model with migration

Pospíchal, Jaroš and Schwarz present in [PJ09], [PJS10] and [PSJ10] a genetic algorithm based on the island model. Therein each thread is assigned to one individual and a thread block corresponds to an island. Therefore the fast shared memory and synchronisation, which are only available within one block, can be used for each subpopulation. A migration is performed between blocks in intervals. The best individuals of a subpopulation replace the worst. This exchange is done asynchronously in a circle. Therefore a set interval isn't guaranteed, due to different thread blocks that run independently. Though this is no big issue as EAs are stochastic algorithms. The implementation is compared with a sequential

one of the same algorithm. The algorithms are compared on an Intel i7-920 and Nvidia GeForce 8800 GTX, GTX 260 and GTX 285 for different test functions and the knapsack problem. The results are speed-ups up to 1000 for a sufficient number of individuals and generations. As mentioned earlier such values are unrealistic and caused by less optimised CPU implementations. The island model converges faster than a standard algorithm, while the quality of results stays the same. Additionally the compiler option `-use_fast_math` is tested. This instructs the compiler to replace mathematical operations with faster but less precise ones. The runtime can be reduced further without deterioration of the solution quality.

5.2.3 Multi-GPU

Using multiple GPUs is the logical continuation if the processing power of one GPU isn't enough. Multi-GPU configurations are still feasible for consumer computers and become more common. This use case is also addressed by modern APIs. When multiple GPUs in one computer should be used, the CUDA function `cudaSetDevice` can be called. Afterwards all commands use the specified device.

The communication between GPUs in one computer can take place directly via the Peripheral Component Interconnect Express (PCIe) bus. The transfer bandwidth between GPUs is limited by PCIe bandwidth, which is significantly slower, than memory access to GPU main memory. If multiple computers are needed, the communication between them can be done via the Message Passing Interface (MPI). Thereby communication costs are further increased, because the data has to be copied from the GPU to the main memory of the computer, then transferred to another computer and copied into memory of the other GPU.

Jaroš presents in [Ja12] an implementation of a genetic algorithm for multiple GPUs. Different implementation options are discussed at the beginning. Mapping individuals to threads is restricted by the size of the individuals and therefore only applicable for small problems. A thread block per individual is only useful for large individuals. Using thread groups is proposed as a middle course. The individuals are divided into subpopulations and each GPU is assigned to one such island. The used operators are uniform crossover, tournament selection and mutation via one bit flip. The migration between islands is done in a ring topology. The communication between GPUs is done via MPI. Furthermore the memory organisation is described in detail. Two different memory layouts for the population are discussed. These are the gene- and chromosome-based layouts. The gene-based structure stores the same genes of the different individuals in succession. The chromosome based structure stores individuals continuously. This is favoured because it preserves data locality and leads to better cache utilisation. Figure 4 schematically represents these layouts. The same algorithm is also implemented for the CPU. The knapsack problem with 10,000 items is used to compare both implementations. The hardware used are two servers each with two Intel Xeon X5650 and seven Nvidia GTX 580. Both implementations scale well for sufficiently large islands. The graphic cards need bigger islands to be fully utilised. The fourteen GPUs reach a speed-up of 35 when compared to the four CPUs and 194 com-

pared to one CPU. Scaled down to one GPU this results in a speed-up of 9 and 14 times respectively.

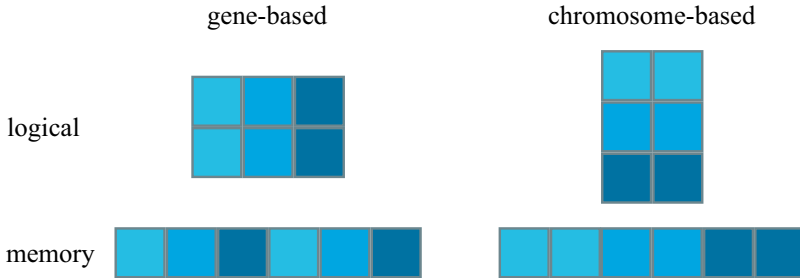


Fig. 4: Schematic presentation of the gene- and chromosome-based memory layout

6 Summary

This paper presents different approaches to performing evolutionary algorithms on graphic cards. It is shown that a complete GPU implementation is possible and sensible. The best implementation can differ depending on structure and complexity of the problem. In general the following recommendations can be given. The algorithm should be implemented completely on the GPU to avoid slow transfer between CPU and GPU memory. Also the architectural features should be considered. This includes the avoidance of branching and global communication, which can be achieved via cellular operators or an island model. These methods are also able to increase the quality of results. The chromosome based memory layout is recommended to increase locality and therefore makes better use of caches.

It's possible to reduce the runtime by an order of magnitude for sufficiently large problems, if a fair comparison between a GPU and a CPU implementation is considered. This corresponds to the ratio of theoretical peak performance of CPUs and GPUs. Therefore speed-ups of 100 or even more are unrealistic and result from comparisons with less optimised CPU implementations. Common mistakes are the use of only one core on multi-core processors or bad access patterns, which reduce the effectiveness of caches.

It's noteworthy that nearly all publications use CUDA for their implementation since it was released. No information could be found on the use of newer frameworks like OpenACC or OpenMP 4.0. This is because of the lack of freely available compilers supporting these standards. At present only commercial compilers offer useful support for these standards. However the GNU Compiler Collection aims to support these too¹⁰.

Overall it can be seen, that only a few publications release their source code. This compromises the traceability and impedes further use or improvement. It also makes more detailed analysis and comparison impossible.

¹⁰ <https://gcc.gnu.org/wiki/OpenACC>

References

- [Be08] Bergman, Keren; Borkar, Shekhar; Campbell, Dan; Carlson, William; Dally, William; Denneau, Monty; Franzon, Paul; Harrod, William; Hiller, Jon; Karp, Sherman; Keckler, Stephen; Klein, Dean; Lucas, Robert; Richards, Mark; Scarpelli, Al; Scott, Steven; Snively, Allan; Sterling, Thomas; Williams, R. Stanley; Yelick, Katherine: , ExaScale Computing Study: Technology Challenges in Achieving Exascale Systems. <http://www.cse.nd.edu/Reports/2008/TR-2008-13.pdf>, 2008.
- [Ca12] Cavuoti, Stefano; Garofalo, Mauro; Brescia, Massimo; Pescap e, Antonio; Longo, Giuseppe; Ventre, Giogio: Genetic Algorithm Modeling with GPU Parallel Computing Technology. Neural Nets and Surroundings, Proceedings of 22nd Italian Workshop on Neural Nets, WIRN 2012, 2012.
- [Ch07] Chitty, Darren M.: A data parallel approach to genetic programming using programmable graphics hardware. In: GECCO '07: Proceedings of the 9th annual conference on Genetic and evolutionary computation. volume 2, ACM Press, London, pp. 1566–1573, 2007.
- [De09] Debattisti, Stefano; Marlat, Nicola; Mussi, Luca; Cagnoni, Stefano: Implementation of a Simple Genetic Algorithm within the CUDA Architecture. GPUs for Genetic and Evolutionary Computation Competition at 2009 Genetic and Evolutionary Computation Conference, 2009.
- [FWW07] Fok, Ka-Ling; Wong, Man-Leung; Wong, Tien-Tsin: Evolutionary Computing on Consumer-Level Graphics Hardware. In: IEEE Intelligent Systems. volume 22, pp. 69–78, 2007.
- [Ja12] Jaroř, Jiřı: Multi-GPU island-based genetic algorithm for solving the knapsack problem. In: WCCI 2012 IEEE World Congress on Computational Intelligence. IEEE, pp. 1–8, 2012.
- [JP12] Jaroř, Jiřı; Pospıchal, Petr: A Fair Comparison of Modern CPUs and GPUs Running the Genetic Algorithm under the Knapsack Benchmark. Lecture Notes in Computer Science, 2012(7248):426–435, 2012.
- [Li07] Li, Jian-Ming; Wang, Xiao-Jing; He, Rong-Sheng; Chi, Zhong-Xian: An Efficient Fine-grained Parallel Genetic Algorithm Based on GPU-Accelerated. In: 2007 IFIP International Conference on Network and Parallel Computing Workshops. IEEE, Liaoning, pp. 855–862, 2007.
- [LL06] Luo, Zhongwen; Liu, Hongzhi: Cellular Genetic Algorithms and Local Search for 3-SAT problem on Graphic Hardware. In: 2006 IEEE Congress on Evolutionary Computation. IEEE, Vancouver, pp. 2988–2992, 2006.
- [Nv16] GPU applications. <http://www.nvidia.com/object/gpu-applications.html>.
- [Ow05] Owens, John D.; Luebke, David; Govindaraju, Naga; Harris, Mark; Krger, Jens; Lefohn, Aaron E.; Purcell, Timothy J.: A Survey of General-Purpose Computation on Graphics Hardware. In: Eurographics 2005, State of the Art Reports. pp. 21–51, 8 2005.
- [PJ09] Pospıchal, Petr; Jaroř, Jiřı: GPU-based Acceleration of the Genetic Algorithm. Genetic and evolutionary computation conference, 2009.
- [PJS10] Pospıchal, Petr; Jaroř, Jiřı; Schwarz, Josef: Parallel Genetic Algorithm on the CUDA Architecture. In: In Applications of Evolutionary Computation, LNCS 6024. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 442–451, 2010.

- [PSJ10] Pospíchal, Petr; Schwarz, Josef; Jaroš, Jiří: Parallel Genetic Algorithm Solving 0/1 Knapsack Problem Running on the GPU. In: 16th International Conference on Soft Computing MENDEL 2010. Brno University of Technology, pp. 64–70, 2010.
- [RMPF08] Robilliard, Denis; Marion-Poty, Virginie; Fonlupt, Cyril: Population Parallel GP on the G80 GPU. In: Genetic Programming, pp. 98–109. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.
- [Sa11] Salmon, John K.; Moraes, Mark A.; Dror, Ron O.; Shaw, David E.: Parallel Random Numbers: As Easy As 1, 2, 3. In: Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis. SC '11, ACM, New York, pp. 16:1–16:12, 2011.
- [SK10] Sanders, Jason; Kandrot, Edward: CUDA by Example: An Introduction to General-Purpose GPU Programming. Addison-Wesley, Boston, 2010.
- [TO15] Coprozessoren TOP500. <http://www.top500.org/statistics/list/>.
- [WW09] Wong, Man-Leung; Wong, Tien-Tsin: Implementation of Parallel Genetic Algorithms on Graphics Processing Units. In: Intelligent and Evolutionary Systems. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 197–216, 2009.

Recent Developments in Example-based Texture Synthesis for Graphics Rendering

Carsten Rudolph¹

Abstract: Textures are essential to allow real-time rendering of computer graphics. While they are easy to use and widely supported by current graphic frameworks, their creation still involves a large amount of manual work. Within the recent years algorithms have been developed that help artists automating common tasks in texture creation. Those tasks can involve texture creation, manipulation and fitting them to certain constraints like tiling, size and spatial uniformity. Example-based Texture Synthesis describes the creation of arbitrarily large textures from a typically smaller and limited exemplar whilst maintaining certain user-defined constraints. In [We09] the authors presented a tutorial-fashioned introduction towards Texture Synthesis combined with a state of the art overview on current algorithms. The goal of this paper is to give a basic introduction, an overview over the recent developments in the area and a comparison of current state-of-the-art algorithms, determining up- and downsides of each of the presented approaches.

Keywords: Computer Graphics, Rendering, Textures, Texture Synthesis, Example-based Texture Synthesis

1 Introduction

In the recent years the available performance and amount of memory on graphics cards has grown considerably, leading to a more realistic look of current graphics applications such as video games, movies or simulations [NV15]. Thus the demand for high quality and resolution textures has grown equally. This typically increases the required work to create great looking textures.

Texture Synthesis aims to automate reoccurring tasks involved in texture creation and give artists a tool to create textures fitting their requirements. Example-based Texture Synthesis is a form of Texture Synthesis where a usually low-resolution exemplar is used to create a completely new texture that should resemble the exemplars global appearance while increasing its original resolution. Also certain constraints can be applied to the exemplar to create textures that can preserve spatial uniformity or make the texture tileable. There are also other constraints but those two are typically common when creating textures for computer graphics.

In 2009 Wei *et al.* [We09] wrote a good introduction into various applications of Texture Synthesis in form of a tutorial for readers who are new to the field. Following that approach, this paper wants to:

¹ Technische Universität Chemnitz, Straße der Nationen 62, 09111 Chemnitz, carsten.rudolph@s2014.tu-chemnitz.de

- Give the reader a short introduction into the topic by describing common algorithms (sections 2 and 3).
- Compare different approaches and try to name applications where one algorithm can be preferred over another one (section 3).
- Give an overview over the recent developments within the field of Texture Synthesis (section 4).

This paper only describes different non-parametric synthesis methods, which include pixel- and patch-based approaches and Texture Optimization. Those are typically used for static, non-volumetric textures, which are most common in the field of real-time rendering. For a detailed insight into other applications, like synthesis of animated or solid textures or real-time Texture Synthesis, the reader is advised to start with the work of Wei *et al.* [We09].

2 Texture Synthesis

Textures are 2- or 3-dimensional sets of values, used to model geometric features that would be too expensive to calculate them each on their own. They can contain different surface information that get interpreted during the rendering process. Typical texture-encoded surface data are normals, depth and color information, but they are not limited to those examples.

Due to the infinite variety of surfaces in the real world, each texture is different. The main question for analytically dealing with textures is to find an appropriate representation for the content of the texture. Section 2.1 tries to implement a general-purpose definition of the appearance of single characteristics within a single texture.

However, different algorithms might choose different descriptions of textures. Section 2.2 shows the basic principles of different common algorithms and their approaches.

2.1 Texture Classes

The distribution of single characteristics of a certain texture highly influences its global and local visual appearance. In order to compare and test algorithms it is important to classify their input. The authors of [Ka15] have chosen to use the following notion to classify their textures:

- **structured** a set of differently scaled, but regularly shaped features (Figure 1a).
- **regular** a set of features which are roughly of same size and shape (Figure 1b).
- **cellular** a set of features with varying size and shape, but clearly delimited towards their neighbors (Figure 1c).

- **semi-structured/stochastic** a random background with randomly spread features of roughly same size and shape (Figure 1e).
- **large-/small-scaled features** a random background with some features that are not regularly spread over the surface and do not necessarily have the same size nor shape (Figure 1d).

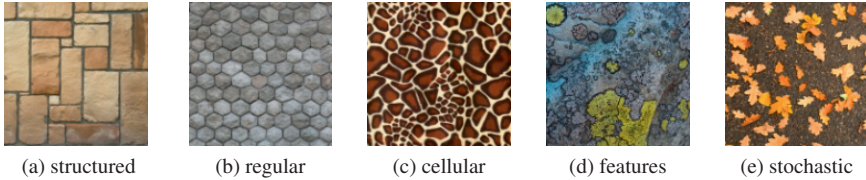


Fig. 1: Examples of texture classes. The textures are classified in the order they appear. [Ka15]

Classifying textures is a good way to make assertions about how good an algorithm works. To be general-purpose an algorithm should produce good quality outputs for a variety of exemplars from each of those classes.

2.2 Algorithm Classes

Texture Synthesis has a broad field of application, with each one having certain requirements for algorithms. In the field of non-parametric Texture Synthesis there are three major principles to classify algorithms [We09]:

- **Pixel-based** algorithms synthesize their output in a pixel-wise manner. This means, they interpret input textures as a set of pixels and synthesize the result pixel-by-pixel. Each algorithm step synthesizes a new pixel within the output.
- **Patch-based** algorithms increase the unit of interpretation to whole patches instead of pixels. Those are typically larger features that are copied into the output texture where they may be manipulated later on. The main difference between those algorithms is how they fill up the remaining space in a way that all features are well-distributed without any gaps, repetitions or artifacts.
- **Texture Optimization** is a technique proposed from [Kw05a] which aims to combine both previously mentioned approaches by synthesizing the texture per-pixel. Unlike pure pixel-based approaches it considers all of the pixels and interprets the mismatches between input-/output pixels as an energy function. This function can be minimized in order to achieve an output that looks similar to the input. The energy function and it's solver determines the quality of the output.

The first two principles mainly differ in the size of the unit of interpretation. Whilst pixel-based algorithms interpret textures as plain pixel-sets without any sense of what the texture actually represents, patch-based algorithms try to analyze the texture first to find out

logically connected areas (so called patches), thus increasing the complexity of such algorithms. The differences are explained in detail within the sections 3.1 and 3.2.

There are other principles for other applications, like solid or liquid Texture Synthesis ([Kw07], [Yu09]), and there are also approaches using existing principles in real-time applications ([Le08], [WL02]).

3 Previous Work

Many state-of-the-art Texture Synthesis algorithms rely on the ability to choose a synthesis unit (pixel or patch) from the exemplar that get's copied into the synthesis result during a synthesis step. A convenient method to do this are Markov Random Fields (MRF). Efros *et al.* [Ef99] were one of the first who utilized this mathematical tool for Texture Synthesis. Their algorithm is described in detail in section 3.1.

In patch-based approaches the synthesis process is typically split into two phases: An *analysis phase* that extracts characteristics of the sample image and the actual *synthesis phase* that combines those characteristics to create a new image. During the synthesis phase the algorithms are using local properties of the sample patches and global functions like uniqueness constraints to spread patches equally and prevent obvious artifacts like heavy repetition. As an example authors of *Kaspar et al.* [Ka15] are utilizing the analysis phase to create so called "guidance channels", which describe homogeneous areas of the sample and their relation towards their environment. In synthesis stage those channels are used along functions like a *global uniqueness constraint* to distribute patches evenly. Their method is described in detail in section 4.2.

3.1 Pixel-based Approaches

As mentioned before the algorithm described in [Ef99] was one of the first to use Markov Random Fields to describe the spatial neighborhood of pixels. The approach is simple and elegant and thus a good starting point for readers who are not familiar with other algorithms.

The basic idea of the algorithm is to initialize the synthesis result with a random field from the exemplar that gets copied to the output. All pixels neighboring the seed are synthesized in a greedy manner, selecting the next pixels to synthesize in a inside-out fashion. The neighborhood of the current pixel to synthesize is described by an Markov Random Field, which can be seen as a window around the pixel. The size of the MRF is the only user-defined parameter the algorithm takes and should resemble the size of the largest patch inside the exemplar to prevent sampling artifacts.

The neighborhood gets compared to the neighborhood of all pixels from the exemplar, resulting in a set of similar candidates for the synthesis. If multiple candidates are found, a random one will be chosen and copied to the output, resulting in a new neighborhood for

the next pixel. The algorithm continues with a pixel that has a neighborhood with as many already synthesized pixels as possible. Figure 2 shows a selection of three pixels that are candidates for copying into the output.

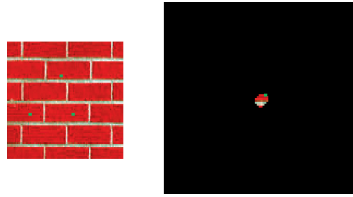


Fig. 2: Candidates for the synthesized pixel feature similar spatial neighborhoods. If multiple candidates exist in the exemplar, a random match is picked.

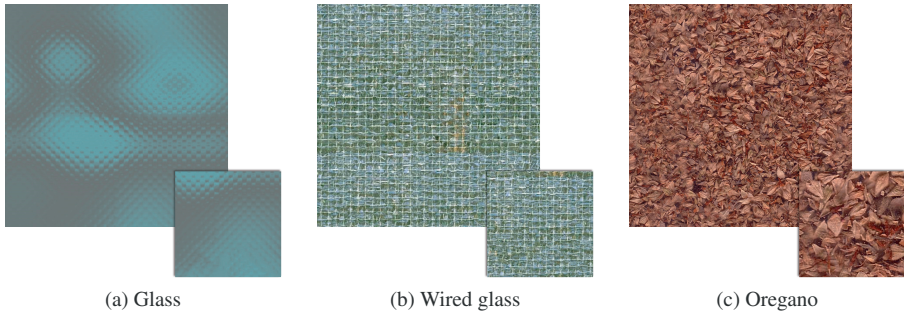


Fig. 3: Synthesis results based on the implementation of [Ef99]. The larger images are the input samples. The results were kept at low resolution for performance reasons.

Sample image	Input resolution	Output resolution	MRF size	approx. Time
Glass	512x512	256x256	20	8 days
Wired glass	512x512	256x256	30	10 days
Oregano	1024x1024	256x256	30	18 days

Tab. 1: The performance of the algorithm depends heavily on the resolution of the input and output image, just as the size chosen for the MRF window. The table shows that the algorithm is not suited for today's typical texture resolutions, even if the quality is very high if the MRF size is chosen cautiously.

Besides its simplicity, the algorithm has the disadvantage that it can be inefficiently slow, depending on the size of the output texture and the size of the MRF. Also the pixel-by-pixel approach prevents the algorithm from taking great advantage of parallelization, because the output of one synthesis step directly influences the neighborhood of pixels for available for the next synthesis steps. The overall performance of the algorithm depends on the following factors: sample size, MRF size and output size. Each algorithm step synthesizes a new pixel in the output image, so the number of steps equals the number of pixels inside the output image. During each step the whole sample image is scanned, so the number of processed pixels equals approximately the number of input pixels times the numbers of pixels within the MRF (even if not all pixels are considered during the MRF analysis). The overall performance can be described by $\mathcal{O}(p_{in} * p_{out} * s_{MRF})$.

For visualization purposes we benchmarked a set of sample images from different texture classes with different resolutions. The results were achieved on an Intel® Core™ i7-5820 @ 3.30GHz. Note that the process only was able to utilize one core. Figure 3 shows the results created from the test run. Table 1 shows the parameters used for the synthesis process and the time it took to create the results. The output resolution has been chosen smaller than the input resolution, because the time it takes to create an 512 square pixel output from an 256 square pixel input does not differ from a 512 square pixel input with a 256 square pixel output, as shown above. Whilst the synthesis of the images in figure 4 (with an sample size of 90 square pixels) only took a few hours to complete, the time consumption raises much higher when synthesizing textures from or for today’s typical resolutions of 512 square pixels or above.

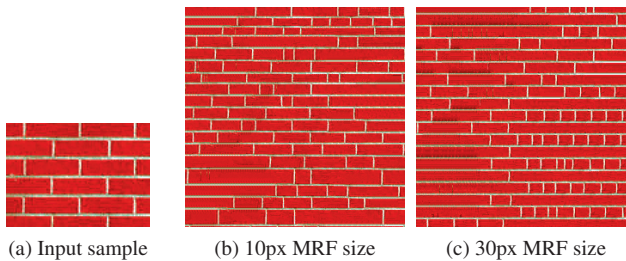


Fig. 4: An sample input is synthesized with two different MRF sizes: The lower the MRF size is, the greater the *randomness* gets in terms of patch distribution.

In terms of quality, the local neighborhood approach may result in increased noise or garbage regions, especially if the MRF size has been chosen to low. Figure 4 shows the impact of different MRF sizes. In both cases the window was not large enough to capture global neighborhood information (like the mortar stripes around a whole brick in this example). As a rule of thumb the MRF should be large enough to capture the largest structure inside the texture.

To address the performance issue *Wei et al.* [WL00] are adopting the previously described approach and replace the full MRF with a fixed neighborhood and the inside-out synthesis fashion with a scaline-based approach. The fixed neighborhood enables the algorithm to take advantage of tree-based algorithms, like *tree-based vector quantification (TSVQ)*, as the authors suggest. Later authors based their research on this method and combined it with other tree-structures, like *kd-trees* (*Kwatra et al.* [Kw05b]) or *k-coherence* (*Tong et al.* [To02]), which is the best in terms of quality and performance according to [We09].

3.2 Patch-based Approaches

Patch-based Texture Synthesis describes a technique that extends pixel-based approaches by using a different interpretation of synthesis units. Instead of single pixel, whole patches of different size get copied to the output. Naturally this cannot be done in an optimal manner, because it is not guaranteed that a patch that exactly fits a hole inside the output can be found in the exemplar. Another challenge of those approaches is to uniformly seed

patches from the exemplar in the synthesis result to prevent unnatural looking repetitions or artifacts. Therefore different algorithms exist, that ultimately differ in the way how they combine patches together to create a naturally-looking result.

Praun et al. [PFH00] described an algorithm that simply overwrites new patches over existing ones, generating surprisingly good results for stochastic textures, but does not work conveniently for structured textures. Therefore *Lian et al.* [Li01] described an algorithm that blends both patches together, which can cause blurry regions, while *Efros et al.* [EF01] use dynamic programming techniques to find an optimal path through both patches and cut them there.

3.3 Texture Optimization

Texture Optimization is an approach, proposed by *Kwatra et al.* [Kw05b], that combines the advantages of both, pixel- and patch-based approaches. Their implementation performs the actual synthesis step in units of pixels, but unlike traditional pixel-based approaches, Texture Optimization considers all of the pixels and generates an energy function from the mismatches between their values in the exemplar and the output. The interesting thing is that this (quadratic) function directly appears to resemble the quality of the output: optimizing (e.g. minimizing) the function leads to better quality results.

4 Recent Developments

In the past years many scientists researched problems with current Texture Synthesis algorithms, mainly trying to improve their performance and the quality of their results. This section goes a little bit more into detail on the *PatchMatch* ([Ba09], [Ba10]) algorithm, which was a significant milestone for Texture Optimization algorithms. *Image Melding* ([Da12]) further improved it and *Kaspar et al.* [Ka15] utilized both methods to present a *Self-Tuning Texture Optimization algorithm*.

4.1 PatchMatch and Image Melding

Previous algorithms mostly derived the nearest-neighbor search problem from [Ef99]: the more pixels the input and output image has, the more time it takes to search the whole input for fitting synthesis candidates in each synthesis iteration. *Ashikhmin* [As01] showed that significant performance gains can be achieved by limiting the search space in the exemplar to a patch around the source neighbors of the neighboring pixels of the target, thus exploiting *local coherence*. *Barnes et al.* [Ba09] based their *PatchMatch*-algorithm on that coherence assumption. Also they prototyped a GPU-based implementation of their nearest-neighbor search algorithm, generating a even better speedup.

In terms of quality, *PatchMatch* utilizes user-provided *guidance channels* to solve the problem of poor synthesis results where the boundaries of the synthesized target deliver only

weak or none constrains for natural looking completion results. However, *guidance channels* in general are user-provided thus only shifting the search for fitting patches towards user suggestions. This prevents a highly or fully automated process and may even increase the overall workload for the actual artist.

The introduction of *guidance channels* as a pre-calculated map of characteristics from the input sample also firstly split up the synthesis process into two distinct stages: the *analysis phase* and the actual *synthesis phase*, which is typical for all following Texture Optimization algorithms, which later aim on (partially) automating this task.

Darabi et al. [Da12] generalized the PatchMach algorithm to be able to apply patch-based approaches to the whole family of image melding problems, where Example-based Texture Synthesis represents a special case with one input source. They introduced three major changes towards patch-based algorithms:

- Extending the degrees of freedom of the search space to handle not only the basic transformations translation, rotation and scale, but also reflection and non-uniform scale.
- Using the image gradient as an additional property to represent a patch. The gradient is stored as a *guidance channel* during analysis phase.
- Changing the interpretation of *energy* in Texture Optimization thus increasing the sharpness of the result.

During synthesis phase, which can be seen as a generalized application of *hole-fitting* or *image completion* for PatchMatch and Image Melding implementations, the algorithm searches for fitting patches within the guidance channels from the analysis phase. In order to find fitting patches it performs a search and votes for the best choice. It does so by solving a screened Poisson equation (*Poisson fusion*), instead of using a Fourier-based solver. This proved to give good quality results while still reducing the complexity.

4.2 Self-Tuning Texture Optimization

Kaspar et al. [Ka15] base their implementation on *Image Melding* by removing some of its features which are not common for two-dimensional textures, which are typically rectified and do not exhibit perspective distortions. Therefore they disabled searching among rotation and scale, improving the performance of the search process. Also they do not use some analysis filters like gain, bias adjustments or gradient channels. To increase texture sharpness, they are not using the *Poisson fusion* described in 4.1. Instead they quantize the nearest-neighbor field to integer locations to improve performance of the search process.

Besides those changes to *Image Melding* the authors also introduce several improvements over existing algorithms which they derived from a survey of common problems. Those problems arise when applying an algorithm to textures of different texture classes:

- **Structured and large featured** textures often suffer from missing non-local information which is explicitly derived from large structures. Repetitions are a typical symptom of this problem. The algorithm solves the problem by using *guidance channels* and introducing a technology to calculate them automatically.
- **Regular and stochastic** structures may not be evenly spread due to random initialization of the synthesis result. The authors introduce a new smart initialization technique that is build from random blocks, rather than pixels, which improves output quality for the mentioned classes, whilst not affecting the quality for textures without any regularity.

Additionally a *spatial uniformity constraint* ensures that the output image looks globally similar to the exemplar by keeping track and constraining the number of occurrences of certain pixels, while improving performance towards other approaches, like bidirectional similarity ([Si08]).

5 Further Research

In this paper we've presented previous and current developments in the field of non-parametric Example-based Texture Synthesis. Since computation performance, especially graphics memory on current GPU's have significantly been increased, higher resolution textures can be used to create photo-realistic virtual environments. The demand for such textures increases with the supply of computation performance, increasing both, the demand for Texture Synthesis algorithms in general to automate common manual texture editing tasks and the stress for existing algorithms. Current algorithms are delivering good results for many, but not all textures. The lack of an general-purpose solution prevents Texture Synthesis from being broadly used within the graphics industry. Also the poor performance of many of those algorithms does not allow viable interactive editing.

Researchers significantly improved existing algorithm approaches in the last years by carefully evaluating their downsides and trying to apply their benefits to other algorithms. For example pure pixel-based algorithms have been replaced by Texture Optimization, which combines the advantages of patch- and pixel-based approaches. By critically evaluating current state-of-the-art implementations, chances for a fast and quality algorithm to be developed within the next years are high.

6 Acknowledgments

This paper has been written as a student research project. The author wants to thank Tom Uhlmann, staff member of the faculty *Graphical Data Processing and Visualization* at the Chemnitz University of Technology, for the support.

References

- [As01] Ashikhmin, Michael: Synthesizing natural textures. In: Proceedings of the 2001 symposium on Interactive 3D graphics. ACM, pp. 217–226, 2001.
- [Ba09] Barnes, Connelly; Shechtman, Eli; Finkelstein, Adam; Goldman, Dan: PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics-TOG*, 28(3):24, 2009.
- [Ba10] Barnes, Connelly; Shechtman, Eli; Goldman, Dan B; Finkelstein, Adam: The Generalized PatchMatch Correspondence Algorithm. In: European Conference on Computer Vision. September 2010.
- [Da12] Darabi, Soheil; Shechtman, Eli; Barnes, Connelly; Goldman, Dan B; Sen, Pradeep: Image melding: combining inconsistent images using patch-based synthesis. *ACM Trans. Graph.*, 31(4):82, 2012.
- [Ef99] Efros, Alexei; Leung, Thomas K et al.: Texture synthesis by non-parametric sampling. In: Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on. volume 2. IEEE, pp. 1033–1038, 1999.
- [EF01] Efros, Alexei A; Freeman, William T: Image quilting for texture synthesis and transfer. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques. ACM, pp. 341–346, 2001.
- [Ka15] Kaspar, Alexandre; Neubert, Boris; Lischinski, Dani; Pauly, Mark; Kopf, Johannes: Self Tuning Texture Optimization. *Computer Graphics Forum*, 2015.
- [Kw05a] Kwatra, Vivek; Essa, Irfan; Bobick, Aaron; Kwatra, Nipun: Texture Optimization for Example-based Synthesis. Technical report, Georgia Institute of Technology, 2005.
- [Kw05b] Kwatra, Vivek; Essa, Irfan; Bobick, Aaron; Kwatra, Nipun: Texture optimization for example-based synthesis. In: *ACM Transactions on Graphics (TOG)*. volume 24. ACM, pp. 795–802, 2005.
- [Kw07] Kwatra, Vivek; Adalsteinsson, David; Kim, Theodore; Kwatra, Nipun; Carlson, Mark; Lin, Ming: Texturing Fluids. Technical report, Department of Computer Science, University of North Carolina at Chapel Hill, 2007.
- [Le08] Lefebvre, Sylvain: Filtered Tilemaps (in Shader X6). *Shader X6: Advanced Rendering Techniques*, pp. 63–72, 2008.
- [Li01] Liang, Lin; Liu, Ce; Xu, Ying et al.: Real-time texture synthesis using patch based sampling. Microsoft Research, 2001.
- [NV15] nVidia CUDA C Programming Guide.
- [PFH00] Praun, Emil; Finkelstein, Adam; Hoppe, Hugues: Lapped textures. In: Proceedings of the 27th annual conference on Computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co., pp. 465–470, 2000.
- [Si08] Simakov, Denis; Caspi, Yaron; Shechtman, Eli; Irani, Michal: Summarizing visual data using bidirectional similarity. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, pp. 1–8, 2008.
- [To02] Tong, Xin; Zhang, Jingdan; Liu, Ligang; Wang, Xi; Guo, Baining; Shum, Heung-Yeung: Synthesis of bidirectional texture functions on arbitrary surfaces. *ACM Transactions on Graphics (TOG)*, 21(3):665–672, 2002.

- [We09] Wei, Li-Yi; Lefebvre, Sylvain; Kwatra, Vivek; Turk, Greg: State of the Art in Example-based Texture Synthesis. In: EUROGRAPHICS 2009. 2009.
- [WL00] Wei, Li-Yi; Levoy, Marc: Fast texture synthesis using tree-structured vector quantization. In: Proceedings of the 27th annual conference on Computer graphics and interactive techniques. ACM Press/Addison-Wesley Publishing Co., pp. 479–488, 2000.
- [WL02] Wei, Li-Yi; Levoy, Marc: Order-Independent Texture Synthesis. CoRR, abs/1406.7338, 2002.
- [Yu09] Yu, Qizhi; Neyret, Fabrice; Bruneton, Eric; Holzschuch, Nicolas: Scalable real-time animation of rivers. In: Computer Graphics Forum. volume 28. Wiley Online Library, pp. 239–248, 2009.

Counting Performance: Hardware Performance Counter and Compiler Instrumentation

Jan-Patrick Lehr¹

Abstract: Analyzing applications for their runtime behavior, especially in the light of efficient resource utilization, involves iterative measurements and the interpretation of the data gathered. For fine grained analysis often hardware performance counters are monitored. Since compiler instrumentation augments the program with calls to a measurement system, it interacts with compiler optimizations. Currently, it is unclear to which degree the instrumentation influences the characteristics of the resulting binary. To determine the behavioral change introduced through instrumentation we conduct a series of measurements on a subset of the SPEC CPU 2006 benchmark suite. We show that although runtime increases up to acceptable 10% of the original runtime some hardware performance counter are significantly perturbed. However, it is also possible that hardware performance counter deviate only slightly from the values measured in the original binary, even though the benchmark's runtime increases substantially. In particular, the program *444.namd* showed an increase in store instructions by $2x$ with only 3% runtime overhead, whereas *450.soplex* did not show significant change in mispredicted branches, when exhibiting an increase in runtime by $3x$. We investigate whether the validity of a hardware performance counter can be determined via static analysis. Therefore, we outline a new tool based on the MAQAO binary analysis framework to compute and compare the static instruction mix of binaries to identify the introduced change in the static instruction mix due to instrumentation. The static instruction mix is comparable to a histogram, denoting how many instructions of a certain category are found in each function of the binary. We conclude that static analysis of the binary's instruction mix may describe induced change for hardware performance counters. Finally, we outline directions of further research in the field of perturbation detection and predictive modelling.

Keywords: High Performance Computing, Performance Analysis, Compiler Instrumentation, Hardware Performance Counter, Binary Analysis

1 Introduction

Modern hardware offers increasing performance at the price of increasing complexity. Especially in the area of high performance computing (HPC), big advances have been made within the last decades. However, the increasing compute capability poses strong demands on software architects in order to make efficient use of the available hardware.

Since many users of HPC are experts in other domains, such as physics or mechanical engineering, a deep understanding of the available hardware cannot be assumed. Bischof et al. pointed out that, based on economic considerations, performance analysis and tuning is important [BaMI11]. Performance analysis and tuning is an iterative process consisting of measurement, analysis and tuning, with the goal to improve the application's performance

¹ TU Darmstadt, Scientific Computing, Mornewegstrasse 30, 64293 Darmstadt

and efficiency. In order to give guidelines and help analysts to carry out the task, the process has been modelled by Iwainsky et al. [Iw11].

To gain insight about the application's behavior a performance analyst (PA) performs a series of measurements, either manually or automatically. Those measurements are, in general, performed using *sampling* or *instrumentation* and are taken with respect to a certain metric. For a more detailed explanation of the two approaches the reader is referred to Morris et al. [Mo10]. The program that is measured is called a performance proxy, as it is used to approximate the original application's behavior in regard of the metrics monitored. As with any measurement, the system under observation is also influenced by the measurement carried out [Ma91]. This influence usually is referred to as runtime overhead.

Besides measuring runtime, often hardware performance counters are monitored. A commonly used library to abstract from the low-level and machine dependent interface of hardware performance counters is the PAPI library [Br00]. State of the art measurement systems like Score-P [Me11] or HPC Toolkit [Ad10] offer support for this interface. However, instrumentation, as used by Score-P, might actually have two types of influence on a program's behavior [MRW92]. Since it changes a program's source code, it might also change the compiler's decisions on optimizations it can apply. Consequently, not only the measurement system's runtime overhead is introduced, but a potentially differently optimized binary is observed.

If a PA uses measurements taken from such an instrumented binary as the basis for a tuning decision, the question arises whether the collected data is obtained from a valid performance proxy of the original application. Current best practice suggests that whenever the measured runtime was within 10% of the original runtime, the performance proxy is considered valid. However, this guidance only reflects runtime and leaves aside other important characteristics of an application.

We conducted a series of experiments regarding the influence of automatic compiler instrumentation on hardware performance counters. The study is carried out using a subset of the SPEC CPU 2006 benchmark suite consisting solely of C/C++ benchmarks as the toolchain that we use is currently only capable of dealing with C/C++ programs.

To help the PA in determining whether the collected values resulted from a valid performance proxy, we investigate whether hardware performance counter perturbation can be correlated to the change in the static instruction mix within the binary. The static instruction mix is comparable to a histogram, denoting how many instructions of a certain category are found in each function of the binary. To this end, we developed a tool, based on the MAQAO binary analysis framework by Djoudi et al. [Dj05], that computes the static instruction mix.

The paper is structured as follows: Section 2 provides a more detailed motivation and background. We present experimentally gathered results for PAPI events in Section 3. Our definition of the static instruction mix as well as a description of our tool is given in Section 4. We conclude our findings in Section 5 and give an outline of possible future work and application areas of instruction mix comparison in Section 6.

2 Background

To investigate whether an application is constrained by a specific machine limitation or does not use a given platform efficiently, an analyst needs to obtain an understanding of the application's runtime behavior. In order to construct a picture of the application's behavior, a series of measurements is carried out, gathering a variety of metrics. The metrics typically include runtime, performance counter and, in the case of parallel applications, communication waiting times.

Inspecting and relating all of the available information leads to a hypothesis why the target application suffers from a specific performance problem. Depending on the identified problem, the analyst decides which limiting factors are to be approached first. If the performance proxy, however, does not reflect the original application's behavior with respect to the metric it was used to monitor, the PA may draw a wrong conclusion. As a result, tuning effort is spent to resolve an artefact which resulted from the measurement and is not necessarily present in the original application. Typically recorded hardware performance counter differ from platform to platform, but may include:

- Level 1 and 2 instruction and data cache accesses / hits / misses
- Performed branch instructions / mispredicted branches
- Load and store instructions
- Stalled cycles / cycles without instruction issue

In this work, the effects of instrumentation will be inspected. The impact of sampling on the various hardware performance counters needs yet to be studied, but will not be part of this work. Instrumentation augments the application with calls to a measurement system. Thus, it guarantees the observation of events and is therefore useful for certain tasks, e.g. discovery of communication patterns in parallel applications. Instrumentation can be carried out manually or automatically; in this work we focus on automatic compiler instrumentation, as provided by major compilers including GCC², Clang³ or the Intel compiler⁴.

List. 1: Uninstrumented example code.

```
int factorial(int n){
    if(n == 0){
        return 1;
    }

    return factorial(n-1) * n;
}
```

List. 2: Instrumented example code.

```
int factorial(int n){
    __cyg_profile_func_enter(&factorial ,
        __builtin_return_address(1));
    if(n == 0){
        __cyg_profile_func_exit(&factorial ,
            __builtin_return_address(1));
        return 1;
    }

    int _tval_ = factorial(n-1) * n;
    __cyg_profile_func_exit(&factorial ,
        __builtin_return_address(1));
    return _tval_;
}
```

Automatic compiler instrumentation introduces calls to a measurement interface at the

² see <http://gcc.gnu.org/>

³ see <http://clang.llvm.org/>

⁴ see <http://software.intel.com/en-us/c-compilers>

start and exit of every function as can be seen exemplarily in Listing 2. Especially in C++ codes, this can lead to massive runtime overhead, if one considers a modern, object-oriented programming style with many small functions. However, the introduced runtime overhead is only one part, as the additional function calls may also change the compiler's decisions about optimizations, including inline expansion and tail call elimination. Thus, compiler instrumentation might influence the target application's behavior in more subtle ways.

2.1 Related Work

Mytkowicz et al. showed that capturing software metrics using instrumentation may significantly perturb hardware performance counters [My07]. In their work they use sampling to create program traces of vanilla⁵ and instrumented versions⁶ and capture a selection of hardware performance counters, including level 1 data and level 2 total cache misses. For both traces, the correlation of two performance counters values within the trace is computed. After aligning the traces, these correlations are compared and the results show that adding instrumentation significantly alters the correlation values. In comparison to their approach, our research targets static analysis of the binary instead of the comparison of multiple traces. The possibility to determine the influence of instrumentation statically, would decrease the necessity to perform time and resource consuming baseline measurements.

Malony et al. proposed a methodology to automatically detect measurement overhead in time stamps of program trace events [MRW92]. In their approach, they construct a propagation model of incurred runtime overhead and are able to recover trace event time within small error margins. In order to decrease the necessary analysis time, Malony et al. proposed an on-the-fly compensation method for overhead mitigation in [MS05]. They extend their former work to account for parallel applications and the special needs to capture communication correctly. However, both approaches focus solely on time and the relative order between different events in a potentially parallel application. While this is valuable information for the analysis of applications, their approach regards solely runtime.

Kashnikov et al. use the static loop analyzer provided with the MAQAO infrastructure to extract low-level assembly features and characterize loops to discover potential optimization opportunities [Ka13]. While their approach to extract important features, such as vectorization ratio, is similar, the actually extracted features are different and the overall goal differs. Kashnikov et al. focus on improving existing compiler optimization or apply further binary optimization, our work focuses on determining the validity of measured performance counter values.

Moseley et al. focus on the discovery of regions for potential improvements in compiler optimization techniques. While the process involves the comparison of assembler instructions, it focuses on the performance of short instruction sequences in terms of runtime.

⁵ Here: the application without any instrumentation applied.

⁶ Here: the application compiled with any kind of instrumentation applied.

To this end, they match event logs from program traces and correlate the respective assembly instructions [MGP09a]. Although their correlation approach is interesting, the primary goal is drawing the user’s attention to potentially poorly applied compiler optimizations [MGP09b].

Static binary analysis is frequently used for reverse engineering and cross instruction set architecture translation [CE99, CS00, TC02]. Especially today’s high-level language abstractions, such as virtual calls, and transformations applied by optimizing compilers pose challenges to the correct reverse engineering of the binary.

To the best of our knowledge static binary analysis has not been applied to determine whether an instrumented binary is a valid performance proxy with respect to a certain hardware performance counter.

3 Experiments

In order to study the impact of instrumentation on hardware performance counters we conduct a series of measurements on a C/C++ subset of SPEC⁷ CPU 2006 benchmark programs, using the Clang compiler in version 3.8. The benchmarks are written either in C (*403.gcc*, *429.mcf*, *433.milc*, *456.hmmmer*, *458.sjeng*, *462.libquantum*, *464.h264href*, *470.lbm* and *482.sphinx3*) or in C++ (*444.namd*, *447.dealII*, *450.soplex*, *453.povray* and *473.astar*). For a more detailed explanation of the benchmarks, see [He06].

To measure the PAPI counters, we developed a lightweight wrapper library, which allows us to capture PAPI events defined in its high-level API. The library is loaded using *libmonitor* [Kr13], a lightweight mechanism to inject measurement facilities into applications. All measurements are conducted on nodes from phase 2 of the Lichtenberg cluster of TU Darmstadt⁸. Each node is equipped with two Intel Xeon E5-2680 v3 processors, 64GB main memory and is used exclusively for the measurement. The processes are pinned to a specific core and hyper threading as well as frequency scaling are disabled in order to obtain comparable results across runs.

3.1 Measuring PAPI Counters for Vanilla and Instrumented Binaries

In our measurements we obtain values for exactly two PAPI events. The total number of instructions completed (PAPI_TOT_INS) is measured in every run. The second metric is one of nine commonly used metrics: Level 1 instruction (L1I) cache misses (PAPI_L1_ICM), level 2 instruction (L2I) cache total accesses (PAPI_L1_ICA), level 2 instruction and data cache misses (PAPI_L2_ICM, PAPI_L2_DCM), load and store instructions (PAPI_LD_INS, PAPI_SR_INS), conditional and mispredicted branches (PAPI_BR_CN, PAPI_BR_MSP) as well as the number of cycles stalled on any resource (PAPI_RES_STL). The counter values are monitored for a vanilla version as well as for an automatically instrumented version of

⁷ See <http://www.spec.org>

⁸ see www.hhlr.tu-darmstadt.de

the benchmark. We do not link the instrumented binary to an actual implementation of a measurement system, but use the empty default implementation in `glibc`⁹. The goal is to gather an understanding of hardware performance counter behavior in the sole presence of additional calls to a measurement system, without actually capturing data. Thus, the perturbation observed stems from solely the instrumentation instructions and its impact on the compiler's optimization decisions.

Benchmark	Runtime	Load	Store	L1I Misses	L2I Misses	L2D Misses	Cond. Branches	Br. Mispred.
403.gcc	1.405	1.608	1.579	1.091	1.121	0.985	1.000	1.200
429.mcf	1.186	1.422	1.561	1.506	1.508	1.028	1.075	1.611
433.milc	1.052	1.171	1.379	1.250	1.249	0.984	1.000	1.010
444.namd	1.038	1.344	1.959	1.135	1.073	1.002	1.000	1.000
447.dealII	13.136	7.282	14.736	17.479	7.788	0.764	1.428	6.247
450.soplex	3.081	3.271	11.122	2.892	1.932	0.650	1.093	1.033
453.povray	3.170	2.300	2.801	2.657	1.808	1.738	1.062	3.304
456.hammer	1.024	1.010	1.018	1.007	0.935	1.000	1.000	1.009
458.sjeng	1.488	1.360	1.586	5.557	0.786	1.001	1.005	1.263
462.libquantum	1.002	1.019	1.070	1.449	1.465	1.011	1.000	0.923
464.h264ref	1.374	1.289	2.085	1.140	1.181	0.993	1.000	1.370
470.lbm	1.053	1.000	1.000	0.923	1.015	1.027	1.000	0.997
473.astar	1.818	1.970	2.775	1.653	1.473	0.913	1.000	1.063
482.sphinx3	1.132	1.054	1.325	1.104	1.115	0.998	1.000	1.305

Tab. 1: Measured counts for PAPI load and store instructions as well as L1I, L2I, L2D cache misses and conditional branch instructions as well as conditional branch instructions mispredicted. The values are given as relative factors with the vanilla version's value being the baseline (1.0).

Table 1 lists the benchmarks and their runtime increase for an automatic compiler instrumented version. The factors listed show relative behavior of several PAPI counters when vanilla and instrumented binaries are compared. Taking overhead runtime as an indicator for the validity of the measurement suggests that *433.milc*, *444.namd*, *456.hammer*, *462.libquantum* and *470.lbm* are valid performance proxies. In contrast, *447.dealII*, *450.soplex* and *453.povray* show an increase of 13x and 3x in runtime and, thus, would not be considered valid. Both, *429.mcf* and *482.sphinx3* do not exhibit large overhead and are within 20% overhead, whereas *458.sjeng*, *464.h264ref* and *473.astar* show runtime increases between 37% and 82%.

As it can be seen in Table 1 and Figure 1, *444.namd* would be considered valid, although the increase in store instructions is close to 2x. Such a deviation can lead the performance analyst to drawing the wrong conclusions and trying to investigate and fix not an inherent problem of the application but an artefact of the measurement. The same can also be true for *433.milc*, although, the increase in store instructions is not as severe as it is for *444.namd*.

For *462.libquantum*, where nearly no runtime increase is seen, the L1I and L2I cache misses increase by nearly 50%. In the analysis one can conclude that it would be necessary to change code layout in order to improve code locality in the binary. However, as the increase is due to the instrumentation added, it is uncertain whether the vanilla version would benefit from such changes.

⁹ see <http://www.gnu.org/software/libc/>

On the other hand, for *403.gcc*, the counted number of L1I cache misses is nearly unchanged when compared to the vanilla version. Additionally, the L2I cache misses are close to a 10% deviation compared to the vanilla version. This suggests that the number of these events counted is usable for analysis purposes, even though the runtime increase is 1.4x.

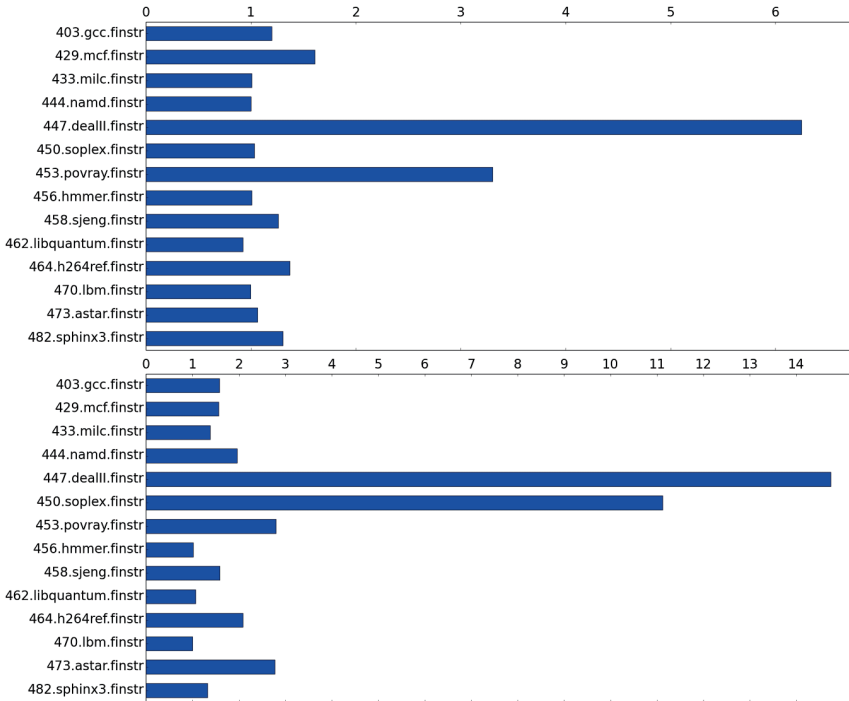


Fig. 1: Depicting the increase for mispredicted branches (top) and store instructions (bottom) for the instrumented versions of the benchmarks.

The results also show that a measurement, which would not be considered a valid performance proxy, as it lies outside of 10% of the vanilla runtime, can indicate the original's behavior with respect to a certain hardware performance counter. All programs, except *447.deallI*, show no or very little change with respect to the measured number of conditional branch instructions committed. Additionally, the counted number of mispredicted branches is nearly unchanged for some of the benchmarks for which the measurement would not be considered valid, see Figure 1. For example, *450.soplex* shows nearly no change with respect to committed, as well as mispredicted, conditional branches, but would not be considered valid, as its runtime increased by 3x.

In summary, we conclude that requiring only runtime to be within 10% of the vanilla version's runtime seems to be valid as a general guideline. However, especially the results for *444.namd* and *450.soplex* motivate a detailed evaluation of the subtle interplay between instrumentation and hardware performance counters.

4 Static Instruction Mix

In the last section we showed that instrumentation does influence some of the PAPI performance counters considerably and that runtime increase alone is not always a good measure for hardware performance counter perturbation. To be able to investigate the correlation between our findings and changes at the binary level, this section illustrates our concept of a *static instruction mix* for binaries. It also outlines challenges when inspecting binaries as one may need to reconstruct properties, such as transformed call sites. In addition, we present our approach for computing the static instruction mix based on the MAQAO infrastructure.

The static instruction mix is comparable to a histogram, denoting how many instructions of a certain category are found in each function of the binary. We chose seven categories: `Arith`, `Mem`, `Calls`, `Branches`, `Unconditional Branches`, `Stack Ops` and `Unclassified`, with the latter containing operations which are not clearly of any other category. The `Mem` category in general captures memory related operations like `MOV` operations, which correlate to PAPI load and store events. Consequently, we do not consider register to register move operations as `MEM` operations, as PAPI does not count these as load or store events. `Calls`, `Branches` and `Unconditional Branches` should influence the PAPI cycles stalled counter as well as branch predictor related metrics. The other categories are chosen to be able to inspect the impact on stack memory, compute arithmetic to memory ratio and possibly missed optimization opportunities to eliminate arithmetic operations.

We use MAQAO's call graph analysis to build the static call graph. Since compiler optimization may turn a function call to a jump, we fix up lost call sites doing a jump-to-function-label analysis. The analysis iterates over all jump instructions in the binary with a function name as jump target. If the label does not have an additional offset, the jump is considered to be an optimized call.

One challenge during the reconstruction of the call graph are virtual calls, as can be found in C++. Another challenge are indirect calls through function pointers, which could not be resolved at compile time but depend on runtime values. Since virtual calls are usually implemented using jump tables, to be able to perform the dispatch based on the runtime type of an object, it is not easy to reliably reconstruct them [TC02]. Our implementation currently ignores this fact and does reconstruction based only on calls and jumps to function labels. However, this is not a limitation of the general approach but of the current technology used.

After the construction of the call graph, we compute the *inclusive instruction mix*, which is defined as the sum of all instruction mixes for functions \hat{f} reachable from a given function f . Doing so for the *main* function results in the instruction mix for the binary, without static initialization. Having both, the instruction mix per function as well as the inclusive instruction mix for a function and all reachable descendents, it is possible to compare the change on a function to function level and on a subtree to subtree level. This allows to evaluate the effects of instrumentation on inline expansion and the resulting static instruction mix.

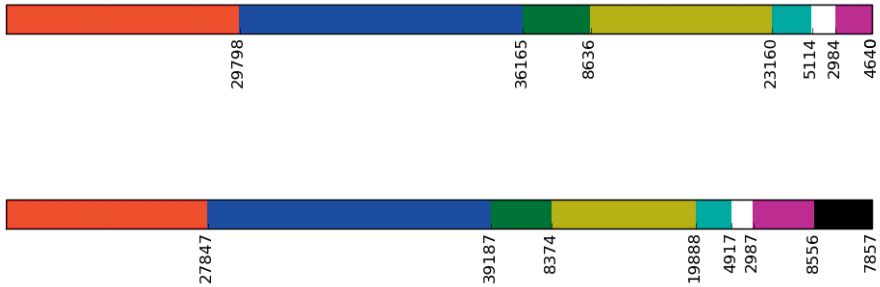


Fig. 2: The instruction mix change between vanilla (upper) and instrumented (lower) version of 453.povray benchmark. Arith is shown in red, Mem is shown in blue, Calls are shown in green, Branches are shown in beige, Unconditional Branches are given in turquoise, Unclassified are shown in white, Stack Ops are colored pink. For the instrumented version, calls to the measurement interface are given in black. The numbers are total counts of the respective instruction mix category.

In general, an increase towards the Mem category can be seen for all of the benchmark binaries. As an example, Figure 2 depicts the inclusive static instruction mix for the *453.povray* benchmark with the vanilla binary at the top and the instrumented binary at the bottom, respectively. It can be seen that the value for Branches as well as Arith within the binary decreases for the instrumented version. These two artefacts, in our opinion, do correlate as the computation of a condition for a Branch instruction itself counts as Arith operation. Thus, if fewer branches exist, then less conditions must be evaluated leading to a decreasing number of Arith operations.

In addition, the decreasing number of Arith operations can be explained by missed optimization opportunities. Usually, if the compiler’s vectorizer hits a computation it is able to vectorize, it emits a vectorized as well as a scalar version of the computation, thus increasing the number of total arithmetic operations. Seeing a decrease in this category is likely to show missed opportunities for such optimization. The lower number of conditional branches can be seen as another indicator of this change, as within an optimized binary, the compiler needs to emit runtime checks, whether the vectorized computation or the scalar version needs to be called.

At the same time the number of Stack Ops increases for the instrumented binary when compared to the vanilla version. This correlates with more functions requiring local stack space being present in the instrumented binary (2025) than in the vanilla version (1545). In the stack frame, the compiler has to allocate space for local variables. Further, it needs to save live registers¹⁰ before calling any function. If the register save was not necessary in the vanilla version, the additional calls to the measurement system, most likely, demand

¹⁰ A register is said to be live if its value is read within any subsequent instruction in the instruction sequence without an interpositional write operation.

the saving of at least one register. This leads to either more stack operations or to additional memory operations.

5 Conclusion

Performance analysis and tuning is important for efficient use of HPC resources. Since monitoring a system necessarily introduces perturbation into the system, the analyst needs to be aware of the change that was introduced by the instrumentation. For instrumentation, we showed that solely relying on the usually accepted guideline that a measurement is valid (in the sense that the instrumented program properly reflects the characteristics of the uninstrumented program) when its runtime does not deviate more than 10% of the original program's runtime, is not always true. In particular, hardware performance counters show significant differences. Consequently, additional information for determining the validity of hardware performance counter measurements is needed.

We propose to inspect the static instruction mix at the binary level to determine the perturbation of hardware counters due to instrumentation, as the binary more than the source code reflects the really running program. To this end, we developed a static binary analysis tool based on the MAQAO framework. The tool classifies every instruction into one of seven categories to reveal changes within the mix of assembly instructions. The classification is performed at the granularity of functions, thus, it draws a clearer picture of the application than simply counting assembly instructions from start to end. Additionally, the call graph is constructed and the inclusive instruction mix is computed, to give a high-level view on the changes in the instruction mix of the binary.

We showed that instrumentation, in general, gears the instruction mix more towards memory operations and increases the number of stack operations. The number of conditional branch instructions decreased for a subset of the benchmarks, as well as the number of arithmetic operations. Thus, the change in the static instruction mix may be helpful to determine the validity of gathered hardware performance counter values.

6 Future Work

Some of the artefacts that can be seen in the static instruction mix may serve as a descriptor of introduced change. However, further investigation is needed to quantify the relation between properties of the binary and perturbation of hardware performance counters and in this section, we outline avenues for future research.

An interesting property of the binary is the distance between function calls within the vanilla and instrumented versions as it indicates the compiler's ability to inline functions. In addition the number of functions and the distance between the code of the respective bodies is of interest, as it can serve as an indication for level one and two instruction cache misses.

Although in current architectures the floating point operation counters are disabled, these counters are of interest in older architectures. To determine the impact of instrumentation on other compiler optimizations, the kind of arithmetic operations which are found in the binary is of interest. To this end, the operand-type mix and differences in the vectorization ratio can be computed between the binary, between single functions, or between subtrees of the respective call graphs.

Since static analysis cannot determine where an application spends most of its time, using profile information to weight the instruction mixes can also benefit the analysis. Especially loop driven programs, like many scientific applications, spend most of their runtime in a limited amount of code. Thus, a tool used for performance data assessment can use already available profiling information to weight the respective static information.

To further investigate the change induced by instrumentation, experiments with more sophisticated instrumentation techniques are also of interest. For example, Score-P offers support to ignore `inline` marked functions from instrumentation. Since inlining is a powerful compiler optimization, it is worth investigating whether this reduces the induced change.

Finally, it is worth exploring whether the instruction mix can be used to predict the likelihood that the instrumented binary captures a certain metric to a sufficient degree.

In general, further research is necessary to investigate whether the static instruction mix can be used as a descriptor for other attributes, such as energy consumption. Being able to correlate the static instruction mix, maybe with added profile information, with the energy behavior of an application, would provide opportunities for energy-aware work scheduling systems.

Acknowledgements

I would like to thank Christian Iwainsky for his guidance and valuable discussions.

References

- [Ad10] Adhianto, Laksono; Banerjee, Sinchan; Fagan, Mike; Krentel, Mark; Marin, Gabriel; Mellor-Crummey, John; Tallent, Nathan R.: HPCToolkit: Tools for performance analysis of optimized parallel programs. *Concurrency and Computation: Practice and Experience.*, 22(6):685–701, 2010.
- [BaMI11] Bischof, Christian; an Mey, Dieter; Iwainsky, Christian: Brainware for green HPC. *Computer Science - Research and Development.*, 27(4):227–233, 2011.
- [Br00] Browne, S.: A Portable Programming Interface for Performance Evaluation on Modern Processors. *Intl. Journal of High Performance Computing Applications.*, 14(3):189–204, 2000.
- [CE99] Cifuentes, Cristina; Emmerik, Mike Van: Recovery of jump table case statements from binary code. In: *Proc. of the 7th Intl. Workshop on Program Comprehension.* pp. 192–199, 1999.
- [CS00] Cifuentes, Cristina; Simon, Doug: Procedure abstraction recovery from binary code. In: *Proc. of the 4th Europ. Software Maintenance and Reengineering.* pp. 55–64, 2000.

- [Dj05] Djoudi, Lamia; Barthou, Denis; Carribault, Patrick; Lemuet, Christophe; Acquaviva, Jean-Thomas; Jalby, William et al.: Maqao: Modular assembler quality analyzer and optimizer for itanium 2. In: The 4th Workshop on EPIC architectures and compiler technology. volume 200, 2005.
- [He06] Henning, John L.: SPEC CPU2006 benchmark descriptions. *ACM SIGARCH Computer Architecture News.*, 34(4):1–17, 2006.
- [Iw11] Iwainsky, Christian; Altenfeld, Ralph; an Mey, Dieter; Bischof, Christian: Enhancing brainware productivity through a performance tuning workflow. In: *Euro-Par 2011: Parallel Processing Workshops*. Springer, pp. 198–207, 2011.
- [Ka13] Kashnikov, Yuriy; de Oliveira Castro, Pablo; Oseret, Emmanuel; Jalby, William: Evaluating architecture and compiler design through static loop analysis. In: *Intl. Conf. on High Performance Computing and Simulation, HPCS*. pp. 535–544, 2013.
- [Kr13] Krentel, Mark W.: Libmonitor: A tool for first-party monitoring. *Parallel Computing*, 39(3):114–119, 2013.
- [Ma91] Malony, Allen D.: Event-based Performance Perturbation: A Case Study. In: *Proc. of the 3rd ACM SIGPLAN Symp. on Principles and Practice of Parallel Programming, PPOPP '91*, ACM, New York, NY, USA, pp. 201–212, 1991.
- [Me11] an Mey, Dieter; Biersdorf, Scott; Bischof, Christian; Diethelm, Kai; Eschweiler, Dominic; Gerndt, Michael; Knüpfer, Andreas; Lorenz, Daniel; Malony, Allen; Nagel, Wolfgang E.; Oleynik, Yury; Rössel, Christian; Saviankou, Pavel; Schmidl, Dirk; Shende, Sameer; Wagner, Michael; Wesarg, Bert; Wolf, Felix: Score-P: A Unified Performance Measurement System for Petascale Applications. In: *Competence in High Performance Computing 2010.*, pp. 85–97. Springer Science + Business Media, 2011.
- [MGP09a] Moseley, Tipp; Grunwald, Dirk; Peri, Ramesh: Chainsaw: Using Binary Matching for Relative Instruction Mix Comparison. In: *18th Intl. Conf. on Parallel Architectures and Compilation Techniques, PACT*. pp. 125–135, 2009.
- [MGP09b] Moseley, Tipp; Grunwald, Dirk; Peri, Ramesh: OptiScope: Performance Accountability for Optimizing Compilers. In: *Intl Symp. on Code Generation and Optimization. Institute of Electrical & Electronics Engineers (IEEE)*, 2009.
- [Mo10] Morris, Alan; Malony, Allen D.; Shende, Sameer; Huck, Kevin: Design and Implementation of a Hybrid Parallel Performance Measurement System. In: *39th Intl. Conf. on Parallel Processing. Institute of Electrical & Electronics Engineers (IEEE)*, 2010.
- [MRW92] Malony, Allen D.; Reed, Daniel A.; Wijshoff, Harry A. G.: Performance measurement intrusion and perturbation analysis. *IEEE Transactions on Parallel and Distributed Systems.*, 3(4):433–450, 1992.
- [MS05] Malony, Allen D.; Shende, Sameer S.: Models for On-the-Fly Compensation of Measurement Overhead in Parallel Performance Profiling. In: *Euro-Par 2005 Parallel Processing.*, pp. 72–82. Springer Science + Business Media, 2005.
- [My07] Mytkowicz, T.; Diwan, A.; Hauswirth, M.; Sweeney, P. F.: Understanding Measurement Perturbation in Trace-based Data. In: *IEEE Intl. Parallel and Distributed Processing Symposium, IPDPS*. pp. 1–6, 2007.
- [TC02] Troger, J.; Cifuentes, C.: Analysis of virtual method invocation for binary translation. In: *Proc. of the 9h Working Conference on Reverse Engineering*. pp. 65–74, 2002.

Non-Intrusive Load Monitoring: A Review and Outlook

Christoph Klemenjak¹ und Peter Goldsborough²

Abstract: With the roll-out of smart meters the importance of effective non-intrusive load monitoring (NILM) techniques has risen rapidly. NILM estimates the power consumption of individual devices given their aggregate consumption. In this way, the combined consumption must only be monitored at a single, central point in the household, providing various advantages such as reduced cost for metering equipment. In this paper we discuss the fundamental building-blocks of NILM, first giving a taxonomy of appliance models and device signatures and then explaining common supervised and unsupervised learning methods. Furthermore, we outline a fundamental algorithm that tackles the task of NILM. Subsequently, this paper reviews recent research that has brought novel insight to the field and more effective techniques. Finally, we formulate future challenges in the domain of NILM and smart meters.

Keywords: Non-Intrusive Load Monitoring, Appliance Modelling, Smart Metering

1 Introduction

Non-Intrusive Load Monitoring (NILM) techniques extract the power consumption of single appliances out of aggregated power data. Given that a measurement device employing NILM must only be installed at a single point, none of the individual appliances have to be equipped with metering devices. In Carinthia, Austria, field trials using such technology are currently being implemented³. These smart measurement devices, installed by energy suppliers, pave the way for sophisticated disaggregation algorithms and possibly also recommender-systems. Such a system would be able to detect devices that have a need for maintenance and give appliance-specific feedback to the consumer. Especially older household devices consume a lot more energy than new ones. In [BKB10] measurements of an aged household refrigerator are reported that consumed three times more energy than a new refrigerator. This is a very specific example of a problem that could be solved by load-disaggregation systems. Moreover, research has shown that appliance-specific feedback can save up to 12 % of annual power consumption [Ar13]. Data acquisition units such as smart meters operate at a central point in the household's power distribution network, generating measurement data of the total power consumption. Currently available commercial smart meters were shown to have a measurement deviation of 10-20 % [ZR11].

This shows clearly that actions have to be taken to improve future metering units. [KEE15] states that the recent trend of metering units to make use of more sophisticated energy monitoring ICs, providing more precise measurements, has improved the effectiveness of NILM.

¹ Alpen-Adria-Universität Klagenfurt, Lakeside Park B10a, 9020 Klagenfurt, christoph.klemenjak@aau.at

² Technische Universität München, Arcisstrasse 21, 80333 München, peter.goldsborough@tum.de

³ http://gewerbe.kelag.at/content/page_strom_smartmeter.jsp

This paper is further structured as follows: Before discussing NILM algorithms, the rudimentary concepts and fundamental vocabulary of the field will be introduced. To investigate appliances, we must first classify them according to their particular features and behaviour. This step is examined in Section 2. Once measurement data is available, it must be analysed to extract relevant appliance signatures, allowing tracking of said appliances. We examine this topic in Section 3. Naturally, a NILM system requires to be aware about the circuits it is monitoring and thus to detect them. For this, a wide range of learning approaches are applied, which we discuss in Section 4. After touching upon these introductory concepts, we describe a fundamental NILM algorithm published by Hart [Ha92]. Since the publication of this algorithm in 1992, the field of load-disaggregation has seen a tremendous amount of further research and novel approaches. We present a selection of these ideas in Section 6. Lastly, in Section 7, we discuss future challenges faced by the NILM community.

2 Appliance types

Appliances differ in the number of operational states and their power consumption behaviour. For the purpose of further discussion and analysis, we therefore present three abstract models that are commonly used in research to represent and characterise appliances. In detail, we will examine models for on/off appliances, multi-state appliances and infinite-state appliances.

2.1 On/Off appliances

The first type of device is the so-called *on/off* appliance. This class includes common household appliances such as a toaster or a light bulb. Such appliances consume only *one* specific amount of power when active. For the large part, on/off appliances are purely resistive. Appliances with a small reactive part are assumed to be linear. A well-known fact is that electric power is additive. This fact is exploited when describing a set of on/off appliances. The total power consumed at time-instant t is the sum of all power signals $P_i(t)$. To modulate the power signal a switch process $a_i(t) \in \{0, 1\}$ is introduced. The product of the switch processes and the power signals models the power consumption of a given appliance. The total power $P_{total}(t)$ can therefore be estimated by:

$$P_{total}(t) = \sum_{i=1}^N a_i(t)P_i(t) + e(t) \quad (1)$$

The additive term $e(t)$ describes the deviation between the actual sum of the modulated power signals and the measured total power. To estimate the state of the appliances, the deviation $e(t)$ has to be minimised. In general, the problem with this is that the complete set of power signals $P_1(t) \dots P_N(t)$ is not known. A second issue is that from a high measurement uncertainty in estimating the total power $P_{total}(t)$, a bad interpretation of the switching process may follow. Many appliances may be estimated to be turning on and off

at the same time. As a solution the *Switch Continuity Principle* was introduced in [Ha92]. It states that in a small time interval the number of appliances changing their state is also small. Consequently, we assume that in a small enough time window the number of state transitions is zero. The sampling frequency of the acquisition unit has thus to be high enough to detect such time windows. Between two such intervals, in which the total power consumption is steady, appliances which change their state can be identified.

2.2 Multi-state appliances

The second type of appliances are multi-state appliances, which have more than one state of operation. Each of these states has a specific power consumption. A common way to represent this class of devices is the finite state machine (FSM) model. The graphic rendition of such a FSM consists of several circles, each corresponding to a specific state of operation with a well-defined power consumption. At the transition from one state to the other, visualised by an edge, the power draw increases or decreases by the difference in consumption between the two states of operation. As an example, let there be a finite state machine model with two states of operation, as illustrated in Figure 1. State A represents a power consumption of 500 W and state B a consumption level of 750 W. At the transition from state A to state B the power consumption of the appliance rises with an amount of 250 W. In contrast to that the power consumption decreases by 250 W from state B to state A. This is analogous to Kirchhoffs law, as the sum of the power changes is zero.

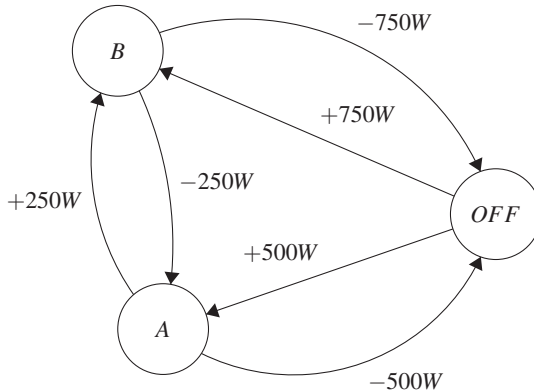


Fig. 1: Finite state machine for an electric heater

2.3 Infinite-state appliances

Lastly, there exist also appliances whose observable set of states is not finite. For example, the power consumption of light-dimmers changes continuously with no consistent step change. Such infinite-state appliances represent a challenge to model and identify. Figure 2 shows the power consumption of such a continuously-varying power consumption. While

on/off appliances as well as multi-state appliances change their power consumption in one clear and observable step, infinite-state appliance's power draw shows a smooth pattern. One subclass of infinite-state appliances are those continuously consuming energy, even when set in standby. Examples of such devices include fire detectors or TVs.

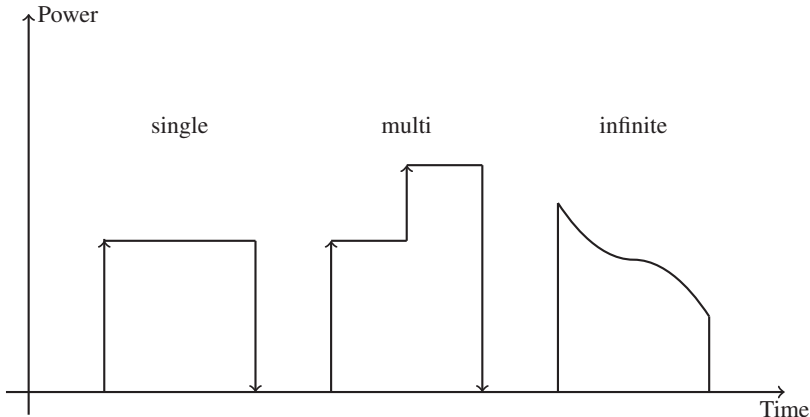


Fig. 2: Power consumption of different appliance types

3 Appliance signatures

Appliance *signatures* describe characteristics specific to certain devices, which can thus be used to identify and classify them. The importance of such signatures was already pointed out in [Ti05], where a taxonomy of voltage-current signatures was introduced to classify appliances. In general, appliance signatures can be seen as measurable parameters, which provide device-specific information extracted from physical quantities. Another taxonomy was introduced in [Ha92], where two classes of non-intrusive appliance signatures are described: *steady-state signatures* and *transient-state signatures*.

3.1 Steady-state signatures

Steady-state signatures comprise features extracted from appliances when they are not currently transitioning between two states but are operating at a steady level of power consumption. More specifically, a steady-state signature is the result of analysing the difference in certain characteristics between two steady states of operation. Such a characteristic may be, but is not limited to, the change in power consumption as was depicted in Figure 1. In general such features can be categorised into the following groups:

- **Power Change:** Real and reactive power are the physical quantities of greatest interest, since they provide very characteristic information about appliances. To detect such features, the power consumption is estimated and plotted as shown in Figure

3. One major difficulty associated with this is the fact that certain power signatures may *overlap*. This overlap results in a bad detection probability especially for appliances with low power consumption. Implementations such as [Hu11] implement rely on these signatures.

- **V-I Features:** The problem with overlaps can be solved by adding additional information about the appliances. By analysing the V-I characteristics, for instance the root-mean-squared (RMS) values of voltage and current, appliances with a similar power consumption may be further described and distinguished.
- **V-I Trajectory:** Another method, using current and voltage signals, is to classify devices by extracting features out of the V-I trajectory. The shape of this trajectory shows useful characteristic features such as asymmetry, looping direction, and enclosed area. A recent application of analysing these features can be found in [HFA14].
- **Harmonics:** In [Ha92], Hart states that analysing the harmonics of a device's current waveform by means of a Fourier Analysis can provide additional information about an appliance's characteristics. In particular, it was found that some non-linear appliances such as motors or light-dimmers produce current waveforms containing a specific set of harmonics, which can further aid in classification.

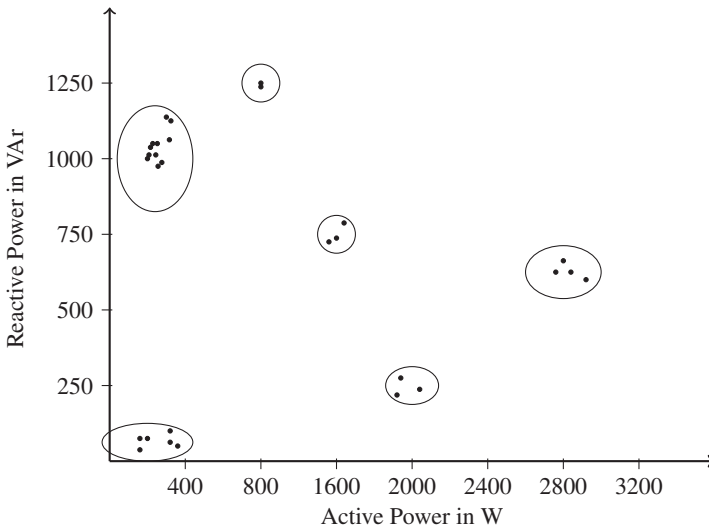


Fig. 3: Distribution of appliances in a traditional P-Q plane

Extraction of such steady-state signatures does not necessarily demand for high-end metering hardware. RMS values of current and voltage as well as frequent power readings provide a good basis to extract steady-state signatures. Already low-cost hardware such as introduced in [KEE15] can be used to identify steady-state signatures.

3.2 Transient-state signatures

Situations exist where two different appliances may have very similar power consumption profiles, reducing chances of correct identification of either device. In such cases, examining the transients, i.e. the consumption behaviour of an appliance when *transitioning* between two steady states of operation, can provide vital information. The transient signature of an appliance is strongly influenced by the physical task it performs [La03]. For instance, the turn-on current of a computer system differs massively from a lighting system due to charging capacitors. The shape, size, and duration of such a transient can thus aid in distinguishing between two appliances when their steady-state signatures alone do not provide a sufficient basis for identification. At the same time, as noted in [Ha92], it must be considered that such momentary transition events are less easily detectable than steady-state operation and may require a higher sampling frequency. Another approach takes emitted voltage noise into account. Each appliance in state operation transmits noise back to the power line. This noise can be measured and categorised into on/off transient noise, steady-state line voltage noise and steady-state continuous noise [Zo12].

3.3 Ambient Appliance Features

Both steady-state and transient characteristics are extracted from what are arguably the most obvious sources: the power, voltage or current draw of an appliance. However, in recent research, appliance-specific patterns have also been extracted from environmental, ambient or behavioural sources. Such techniques exploit the external impact of appliances, such as their heat-dissipation or light-emission. For this purpose, [BSM10] advocates the collection of data from environmental sensors. An implementation of this idea was proposed in [BR11], where the deployment of electromagnetic field detectors (EMF) to combine information about energy wastage and power-consumption profiles was examined. In the same spirit, [GTM13] discusses the fact that home appliances emit sound waves (noise). A system is suggested that correlates information about energy consumption with sound recordings of the respective appliance. In contrast to information provided by sensors about appliances directly, there exists also the paradigm of Context-Aware Power Management (CAPM) [DAS01]. CAPM techniques typically examine signatures not necessarily extracted from appliances themselves, but from their environment, users or usage behaviour. For instance, [KJ12] explores behavioural patterns including duration of use and time of day. In [KJ12] and [Ar13] it is stated that such contextual information may also include location or even weather patterns. Furthermore, [Pr08] studies appliance-user interaction to facilitate load-disaggregation. The behaviour and presence of human beings is traced by a set of motion sensors in the building and combined with other NILM techniques. To gather such ambient data, a wireless sensor network was proposed by [DAS01].

3.4 Optimal Sampling Frequency for Signature Detection

The number and kinds of steady-state and transient appliance features recognisable from an aggregated consumption sample is strongly connected to the frequency of measurement

in the earliest stages of disaggregation. As noted in [KJ11], a wide range of utilised sampling frequencies is reported in past literature. Many of the steady-state appliance features described above, such as current harmonics or V-I trajectories and especially transient features are more realistically attainable at higher sampling frequencies. With *high* frequencies we usually mean several kHz, although select approaches have even employed MHz readings [GRP10]. High-frequency sampling rates not only allow more fine-grained and detailed analysis of device-signatures, but are also more flexible. The obvious benefit of having more samples available than too few, is that when high-resolution data is not required or too bulky to store, it can always be down-sampled to lower frequencies. On the flip side, metering hardware for high sampling rates is practically non-existent in households today, making NILM techniques requiring sampling rates in the region of 1 Hz more practical and immediately applicable. 1 Hz readings as used by Hart's algorithm [Ha92] allow for reasonably effective examination of active and reactive power measurement. More recently, attempts have been made to better adjust NILM algorithms to the low-frequency sampling of conventional smart meters. For example, [KBN10] makes attempts to perform load disaggregation using discriminative sparse coding techniques on power samples provided only on an hourly basis.

4 Learning Approaches

Learning approaches for NILM can fundamentally be divided into *supervised* and *unsupervised techniques*. The distinction between a supervised and an unsupervised algorithm is whether or not ground-truth data about individual appliance features is available to train the algorithm. If such device-specific information is present, meaning that the algorithm knows a priori about the appliances it is monitoring, the learning approach is limited to disaggregation only. On the other hand, an unsupervised algorithm need not only perform load-disaggregation, but additionally detect which appliances exist in the circuit it is monitoring.

4.1 Supervised Learning Approaches

Supervised approaches feed the system with existing device-specific information, such as its power consumption profile. This data may either already exist, such as in the case of the REDD dataset [KJ11], or is the result of an initial training phase, in which a database of appliances and their signatures is collected [AL16]. The actual load-disaggregation is commonly performed by one of two techniques: *Optimisation* or *pattern recognition*. We will elaborate on either approach in the following paragraphs.

- **Optimisation:** A straightforward method to solve the load disaggregation task is to model it as optimisation problem. Obtaining the solution for such problems is well-researched and builds on a simple concept. The extracted appliance features are compared to an existing database consisting of appliance features. When the deviation between the database's entry and the extracted feature can be minimised, the best match is obtained [AL16]. For a small number of appliances, this approach may very well be feasible. However, as discussed in [EBE15], the performance of this method deteriorates with an increasing number of loads, while the complexity increases. Another weak point of this approach is that it may have significant difficulties in distinguishing between loads with overlapping signatures.
- **Pattern Recognition:** This approach detects appliances by means of clustering and mapping state-changes to a feature space [Zo12]. An example of such clustering is given in Figure 3. As outlined by Hart in [Ha92], the identified appliance features in the PQ plane are divided into clusters. Given this initial separation, the clusters are compared to those already known to the supervised system. In further detail, [Zo12] identifies two main approaches: Bayesian classifiers and heuristic methods. For the former, it is assumed that two operating states of an appliance are independent of each other. While research has shown promising results for the Bayesian approach, the independence of states is clearly an ideal but not practical model. For example, the power state of a computer monitor usually depends directly on the power state of the connected computer.

4.2 Unsupervised Learning Approaches

Supervised learning approaches require an initial training phase and input of external, labeled data. Practically speaking, for the average household, such data does not exist. Therefore, unsupervised learning approaches, which are able to operate without a priori information, are a promising alternative. Unsupervised disaggregation techniques are required not only to perform load-disaggregation, but must further train themselves *online*. This means that appliances need to be identified and extracted from the aggregate power signal and their models added to the database of existing devices. The quality of the load-disaggregation is thus additionally dependent on the ability of the system to correctly identify existing devices. Methods of probabilistic analysis such as Hidden Markov Models (HMMs) and extensions thereof are especially suited to this task [AL16]. An HMM is a probabilistic graphical model that differs from standard Markov models in that the states are not directly observable, but can only be estimated probabilistically given certain observations. For NILM purposes, an appliance can be described as an HMM with n hidden states $S = \{s_1, \dots, s_n\}$ representing the appliance's states of operation. Also, we define an observation or emission matrix describing the probability for the appliance to be in a certain state s at time slice t given the observation (emission) of an aggregate power consumption signal. Lastly, there exists a transition matrix $T = (a_{i,j}) \in \mathbb{R}^{n \times n}$ where $a_{i,j}$ represents the likelihood for a transition of the appliance from state s_i to state s_j between two time slices t and $t + 1$. More specifically, $a_{i,j} = P(x_{t+1} = s_j | x_t = s_i)$ with $a_{i,j} > 0$ and $\sum_{j=0}^n a_{i,j} = 1$. Factorial Hidden Markov Models (FHMM) are an extension of the basic

HMM. An FHMM models not only a single but many independent hidden state chains in parallel, with the emission (the aggregate power consumption) being thus a function of all states combined. In [EBE15] it is stated that this can help reduce the number of parameters maintained by the system.

5 Hart's NILM algorithm

The algorithm introduced by Hart in [Ha92] is considered fundamental in the NILM community and is the basis of many of today's load-disaggregation techniques. For this reason, we will outline its basic operation briefly. The general concept of Hart's algorithm is to meter a household's aggregate power consumption, identify appliances and then track their behaviour. The algorithm executes the following tasks:

1. **Measure Power and Voltage:** Measurements of the aggregate power and RMS voltage signal are recorded at a sampling frequency of 1 Hz.
2. **Calculate Normalised Power:** The estimated power signals are normalised (smoothed) depending on the power line voltage. This allows for immediate comparison of power levels.
3. **Edge Detection:** An edge-detection algorithm is applied to the normalized power signals. This algorithm extracts steps in power consumption and labels the time instants.
4. **Cluster Analysis:** The output of the edge detection algorithm is used to create points in the PQ-plane. Points nearby are clustered.
5. **Build Appliance Models:** From the clusters obtained finite state machine (FSM) models are created. The simplest state machine is an on/off appliance consisting of two symmetrical events at the PQ-plane.
6. **Track Behaviour:** The estimated appliance models are tracked. Whenever a modelled appliance performs a state transition, the algorithm recognises this behaviour.
7. **Tabulate Statistics:** Statistics and characteristics of the models obtained so far are calculated and tabulated. These statistics may also be used to predict the future behaviour of the monitored state machines.
8. **Appliance Naming:** In the final step, the algorithm attempts to assign each observed FSM to an actual appliance in the system. For this, Hart recommends Bayesian, maximum-likelihood-multiple-hypothesis or other methods from detection theory.

6 Recent approaches

A more recent approach suggests a rethinking of NILM itself and introduces a new way of implementing the algorithm. The authors of [Ba14] introduce a new modelling of NILM in an application-centric way. The approach demands for real-time processing right after metering, which is termed *online* NILM. Basically it is suggested to divide NILM into three steps: device detection, modelling, and device-tracking. The novelty of this approach is that device detection and modelling are usually said to be offline tasks as part of algorithm-training. For the method proposed, device recognition and monitoring is implemented online and in real-time. Smart meters would thus transmit the measurement data immediately to the cloud or server. For example, the online service could be hosted by the power utility itself, improving its immediate ability to forecast future power consumption. The crux of this idea lies in computation. Such a system would have to perform NILM across hundreds of households in real-time. We identify this in particular as a challenge.

The most common learning algorithms used for load-disaggregation today rely on optimisation or Factorial Hidden Markov Models (FHMM). In [KK15], Kelly et al. very recently employed a novel learning approach based on artificial neural networks (ANN). For this, the authors implemented three separate ANN architectures. The first is a *recurrent neural network*, which learns appliance features on a training dataset to then estimate the appliance consumption level given an aggregate sample. The second architecture utilises a *denoising autoencoder* (dAE), often used for signal reconstruction and denoising, such as for removing grain from an old image or reverberation from an audio track. For the dAE, the learning task is to extract a device's load from an aggregate sample, by viewing the consumption of other appliances as the signal's unwanted noise component. Lastly, a standard neural network was used to regress start and end time as well as power consumption for each activation of a device. We note that little research has been done on the application of these modern machine learning techniques to NILM. Yet, [KK15] shows that neural networks beat conventional load-disaggregation algorithms in almost every metric, inviting further investigation into these promising new learning approaches.

7 Conclusion

In this paper we discussed the concept of NILM, appliance models, appliance signatures and learning methods as well as recent improvements and trends in the field of load-disaggregation. We are certain further research is necessary. One fundamental question posed is *where* and on what platforms data processing and NILM algorithms are performed. The first and simplest option is the measurement device itself, meaning the smart meter or metering units installed in the household. This would require sophisticated hardware that is capable of performing NILM in real-time. In general, such hardware is more expensive and energy-consuming than conventional measurement devices. Therefore another approach is to perform data processing on a device in the home network. Single-board computing devices such as the Raspberry Pi⁴ or BeagleBoard⁵ could be well suited

⁴ <https://www.raspberrypi.org>

⁵ <https://beagleboard.org>

to such a task. With the surge of cloud-computing services in recent years, the employment of such an online service becomes a possibility as well. However, along with data-transmission across networks, away from the home and into the cloud, security concerns will and must be raised. We estimate that the majority of the population would feel unease in sending their private household data to external servers.

In conclusion, we would like to re-emphasise our belief in the very certain potential of load-disaggregation techniques to improve the consumption patterns of individuals and reduce energy wastage in the grid. At the same time, we acknowledge that non-intrusive load-monitoring is still a very open and ongoing field of research and that no current approach is perfect. We express our hope that this will change in the near future.

References

- [AL16] Aiad, M.; Lee, P.H.: Unsupervised approach for load disaggregation with devices interactions. *Energy and Buildings*, 116:96–103, 2016.
- [Ar13] Armel, K. Carrie; Gupta, Abhay; Shrimali, Gireesh; Albert, Adrian: Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy*, 52(0):213 – 234, 2013.
- [Ba14] Barker, Sean; Kalra, Sandeep; Irwin, David; Shenoy, Prashant: NILM redux: The case for emphasizing applications over accuracy. In: *NILM-2014 Workshop*. 2014.
- [BKB10] Benyoucef, Dirk; Klein, Philipp; Bier, Thomas: Smart Meter with Non-Intrusive Load Monitoring for Use in Smart Homes. In: *IEEE International Energy Conference*. pp. 96–101, 2010.
- [BR11] Bergés, Mario; Rowe, Anthony: Appliance classification and energy management using multi-modal sensing. In: *Proceedings of the Third ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*. ACM, pp. 51–52, 2011.
- [BSM10] Bergés, Mario; Soibelman, Lucio; Matthews, H Scott: Leveraging data from environmental sensors to enhance electrical load disaggregation algorithms. In: *Proceedings of the 13th International Conference on Computing in Civil and Building Engineering*, Nottingham, UK. volume 30, 2010.
- [DAS01] Dey, Anind K; Abowd, Gregory D; Salber, Daniel: A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-computer interaction*, 16(2):97–166, 2001.
- [EBE15] Egarter, D.; Bhuvana, V.P.; Elmenreich, W.: PALDi: Online load disaggregation via particle filtering. *IEEE Transactions on Instrumentation and Measurement*, 64(2):467–477, 2015.
- [GRP10] Gupta, Sidhant; Reynolds, Matthew S.; Patel, Shwetak N.: *ElectriSense: Single-point Sensing Using EMI for Electrical Event Detection and Classification in the Home*. In: *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*. *UbiComp '10*, ACM, New York, NY, USA, pp. 139–148, 2010.
- [GTM13] Guvensan, M Amac; Taysi, Z Cihan; Melodia, Tommaso: Energy monitoring in residential spaces with audio sensor nodes: TinyEARS. *Ad Hoc Networks*, 11(5):1539–1555, 2013.

- [Ha92] Hart, George W: Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891, 1992.
- [HFA14] Hassan, Taha; Fahad, Javed; Arshad, Naveed: An Empirical Investigation of V-I Trajectory Based Load Signatures for Non-Intrusive Load Monitoring. *IEEE Transactions on Smart Grid*, 5(2):4–31, Mar 2014.
- [Hu11] Huang, Shyh-Jier; Hsieh, Cheng-Tao; Kuo, Lun-Chia; Lin, Chun-Wei; Chang, Che-Wei; Fang, Shyang-An: Classification of home appliance electricity consumption using power signature and harmonic features. In: *Power Electronics and Drive Systems (PEDS), 2011 IEEE Ninth International Conference on*. IEEE, pp. 596–599, 2011.
- [KBN10] Kolter, J. Z.; Batra, Siddharth; Ng, Andrew Y.: Energy Disaggregation via Discriminative Sparse Coding. In (Lafferty, J.; Williams, C.; Shawe-taylor, J.; Zemel, R.s.; Culotta, A., eds): *Advances in Neural Information Processing Systems 23*, pp. 1153–1161. 2010.
- [KEE15] Klemenjak, Christoph; Egarter, Dominik; Elmenreich, Wilfried: YoMo: the Arduino-based smart metering board. *Computer Science-Research and Development*, pp. 1–7, 2015.
- [KJ11] Kolter, J Zico; Johnson, Matthew J: REDD: A public data set for energy disaggregation research. In: *Workshop on Data Mining Applications in Sustainability (SIGKDD)*, San Diego, CA. volume 25. Citeseer, pp. 59–62, 2011.
- [KJ12] Kolter, J. Z.; Jaakkola, Tommi: Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation. In (Lawrence, Neil D.; Girolami, Mark A., eds): *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics (AISTATS-12)*. volume 22, pp. 1472–1482, 2012.
- [KK15] Kelly, Jack; Knottenbelt, William J.: Neural NILM: Deep Neural Networks Applied to Energy Disaggregation. *CoRR*, abs/1507.06594, 2015.
- [La03] Laughman, Christopher; Lee, Kwangduk; Cox, Robert; Shaw, Steven; Leeb, Steven; Norford, Les; Armstrong, Peter: Power Signature Analysis. *IEEE power and energy magazine*, pp. 56–63, 2003.
- [Pr08] Priyantha, Nissanka B; Kansal, Aman; Goraczko, Michel; Zhao, Feng: Tiny web services: design and implementation of interoperable and evolvable sensor networks. In: *Proceedings of the 6th ACM conference on Embedded network sensor systems*. ACM, pp. 253–266, 2008.
- [Ti05] Ting, KH; Lucente, Mark; Fung, George SK; Lee, WK; Hui, SYR: A taxonomy of load signatures for single-phase electric appliances. In: *IEEE PESC (Power Electronics Specialist Conference)*. pp. 12–18, 2005.
- [Zo12] Zoha, Ahmed; Gluhak, Alexander; Imran, Muhammad Ali; Rajasegarar, Sutharshan: Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey. *Sensors*, 12(12):16838–16866, 2012.
- [ZR11] Zeifman, Michael; Roth, Kurt: Nonintrusive appliance load monitoring: Review and outlook. *IEEE Transactions on Consumer Electronics*, pp. 76–84, 2011.

Ein Ansatz zur Einführung von Complex Event Processing zum workfloworientierten Software-Monitoring

Sebastian Niehaus¹

Abstract: In vollautomatisierten Geschäftsworkflows sollte nicht nur die Ausführung, sondern auch die Überwachung der ausführenden Software automatisiert ablaufen. Bei dieser Überwachung bleiben meist diejenigen Ausfälle unberücksichtigt, bei denen die jeweilige Software zwar noch als aktiver Prozess läuft und Ressourcen beansprucht, aber nicht mehr dem eigentlichen Zweck nachgeht. Ein solcher Ausfall wird erst deutlich, wenn man den gesamten Workflow betrachtet. Im vorliegenden Beitrag wird anhand des Auftragseingangsprozesses eines Unternehmens aus der Fleischwarenindustrie eine Möglichkeit vorgestellt, diese Art von Ausfällen mittels Complex Event Processing (CEP) zu erkennen. Dabei wird die Zielsetzung verfolgt, ein ressourcensparendes Echtzeit-Verfügbarkeits-Monitoring für den Auftragseingangsworkflow zu schaffen.

Keywords: Software-Monitoring, Complex Event Processing, Geschäftsworkflow

1 Einleitung und Problemstellung

In automatisierten Geschäftsworkflows ist der fehlerfreie Workflowablauf abhängig von der Verfügbarkeit der genutzten Softwaremodule. Zur Gewährleistung dieser Verfügbarkeit ist es notwendig jeden Ausfall der Software sofort zu erkennen. Im vorliegenden Beitrag wird ein Unternehmen betrachtet, welches in der Fleischwarenindustrie tätig ist und daher besonders auf fehlerfreie Workflowabläufe angewiesen ist. Zu dem Unternehmen gehören über zehn Schlacht- und Verarbeitungsbetriebe, die in ganz Deutschland verteilt sind, um möglichst kurze Lieferzeiten zu ermöglichen. Von diesen Standorten werden die Produkte direkt zu den Kunden geliefert. Da besonders in der Produktion von frischer Ware die zu produzierenden Artikelzahlen täglich bedarfsgerecht bestimmt werden müssen, besteht für Kunden die Möglichkeit, die Bestellung EDI-basiert zu übertragen. Diese Bestellungen werden vollkommen automatisiert verarbeitet und an den entsprechenden Produktionsstandort übermittelt, damit die Produktion und die Lieferung schnellstmöglich erfolgen können. Dem Kunden ist es somit möglich, die Bestellung rund um die Uhr aufzugeben. Die eingegangenen Kundenaufträge werden anschließend sofort bearbeitet und als Produktionsaufträge an die jeweiligen Produktionsstandorte übertragen.

Fällt in diesem Prozess beispielsweise die Software zur Standortzuordnung der

¹ Hochschule Osnabrück, Fakultät Wirtschafts- und Sozialwissenschaften, Landwehrstraße 7a, 49393 Lohne, Sebastian.Niehaus@hs-osnabrueck.de

Kundenaufträge aus, hat das zur Folge, dass ein oder mehrere Aufträge nicht rechtzeitig in der Produktion ankommen und somit auch nicht bedient werden können. Um einem solchen Ausfall entgegenzuwirken bzw. ihn rechtzeitig erkennen zu können, läuft der Prozess auf hochverfügbarer Hardware ab, die über ein entsprechendes Monitoring-System permanent überwacht wird. Darüber hinaus sendet jede bearbeitende Software eine Fehlermeldung, sofern ein Prozessschritt nicht ordnungsgemäß ausgeführt werden kann. Es treten aber immer wieder Ausfälle auf, bei denen Softwaremodule nicht mehr ihrem ursprünglichen Zweck nachgehen und dementsprechend auch keine Meldung mehr über Erfolg oder Misserfolg der Aktivität senden. Dennoch nehmen sie als aktiver Prozess Hardwareressourcen in Anspruch. Ausfälle dieser Art werden meist erst mit Verzögerung, aufgrund von Hinweisen von Anwendern, erkannt und erfordern auch eine manuelle Kontrolle. Diese späte Erkennung ist darauf zurückzuführen, dass in dem betrachteten Beispielprozess gegenwärtig nur ein Softwareinternes Monitoring sowie ein Infrastruktur-Monitoring eingesetzt werden. In [Hei15] beschreibt Hein ebenfalls, dass durch eine reine Infrastruktur-Überwachung Softwareausfälle in den meisten Fällen erst spät erkannt werden. Darüber hinaus werden die Notwendigkeit und der Wert des Monitorings auf Softwareebene aufgezeigt.

2 Vorgehensweise und Zielsetzung

Die Zielsetzung der Arbeit besteht darin, eine Software zum Monitoring des eingangs beschriebenen Workflows zu konzipieren und deren Einführung zu beschreiben. Das Ziel des Monitorings ist aber nicht die Leistungsüberwachung, wie es beispielsweise beim Business Activity Monitoring (BAM) der Fall ist[Sch13]. Die Software soll lediglich inaktive Softwaremodule erkennen und demnach eine reine Verfügbarkeitsüberwachung leisten. Es soll aber nicht nur die Möglichkeit bestehen, eine Aussage über die Verfügbarkeit proaktiver Softwaremodule zu treffen, sondern auch über Softwaremodule, die nicht zur Datenübertragung dienen. Während der gesamten Einführung wird stets darauf geachtet, die Lösung möglichst leicht übertragbar auf andere Prozesse zu gestalten und somit ein Referenzmodell für ähnliche Projekte zu bieten. Daher soll das Monitoring-System so konzipiert sein, dass nach Möglichkeit keinerlei Modifikationen an den ausführenden Softwaremodulen vorgenommen werden und außerdem keine Eingriffe in den Prozessablauf nötig sind.

Daraus wird folgende Forschungsfrage abgeleitet:

Wie lässt sich Complex Event Processing (CEP) zum workfloworientierten Software-Monitoring einsetzen und welche Vorteile ergeben sich aus dieser Art des Software-Monitorings?

Im Folgenden wird zunächst das Complex Event Processing erklärt und anschließend das entwickelte Konzept der Begrenzung des CEP auf Workflows beschrieben. Darauf aufbauend wird die Einführung des Complex Event Processing erläutert und anhand von ausgewählten Szenarien die Herausforderungen bei der Einführung in die Praxis

beschrieben. Zuletzt werden ein Fazit sowie ein Ausblick auf mögliche aufbauende Projekte geboten. Als vergleichbare Arbeiten sind an dieser Stelle [HG09] und [Hal07] zu nennen, die zwar BAM in der Praxis umsetzen und somit leistungsorientiertes Monitoring als primäres Ziel verfolgen. Dennoch können beide Arbeiten ebenfalls als Referenzmodelle für die Einführung von CEP in Geschäftsprozessen betrachtet werden. In [Hei15] wird zur Softwareüberwachung eine Nutzersimulation vorgestellt. Bei dieser Art der Softwareüberwachung findet ebenfalls eine Prozessorientierung statt.

3 Complex Event Processing (CEP)

Complex Event Processing ist eine Softwaretechnologie, die in dem Buch „The Power of Events“ von David Luckham erstmals beschrieben wird [Luc02]. „Allerdings hat CEP viele unabhängige Wurzeln in der Forschung, wie beispielsweise die ereignisorientierte Simulation über aktive Datenbanken und Netzwerkmanagement bis zum temporalen Schließen in der KI“ [EB09].

Complex Event Processing hat zum Ziel, Ereignisse (Events), die miteinander in Verbindung stehen, in Echtzeit zu überwachen. Ein Ereignis kann dabei alles sein, was passiert oder was passieren kann [LS11]. Im Allgemeinen bezieht sich ein Ereignis auf die Veränderungen eines Zustands, also auf Änderungen des Wertes einer Eigenschaft eines realen oder virtuellen Objekts. Wird ein Wert über einen Zeitraum betrachtet, ergeben sich daher Ereignisströme. In diesen Ereignisströmen bestehen Zusammenhänge zwischen den Ereignissen der unterschiedlichen Ströme. In welcher Art sie im Zusammenhang stehen, wird in Ereignismustern abgebildet. Im Wesentlichen wird zwischen einfachen Ereignismustern und komplexen Ereignismustern unterschieden. Einfache Ereignismuster enthalten dabei Ereignisse und boolesche Operatoren, die die Verbindung zwischen den einzelnen Ereignissen aufzeigen. Oftmals reichen diese einfachen Ereignismuster zur Darstellung der Zusammenhänge nicht aus. In diesem Fall müssen weitere Operatoren zur Darstellung herangezogen werden [Alv10]. Mit Hilfe dieser Ereignismuster lassen sich Ereignisregeln formulieren. Eine Ereignisregel beschreibt eine spezifische Reaktion, die beim Erkennen eines Musters ausgeführt wird. Sie besteht aus einem Bedingungsteil und einem Aktionsteil. Der Bedingungsteil enthält ein oder mehrere Ereignismuster und der Aktionsteil enthält eine Aktion oder Reaktion, die beim Auftreten des Musters ausgeführt werden soll [BD15, S.9ff.]. In der Abb.1 wird ein Ereignismuster dargestellt, welches als

$$(A \wedge B \wedge C) \rightarrow D \rightarrow E \rightarrow (G \wedge F)$$

zu interpretieren ist. Dabei legen die Sequenzoperatoren (\rightarrow) die zeitliche Reihenfolge fest, ebenfalls werden boolesche Operatoren verwendet, wie in diesem Fall das logische Und (\wedge). Jede Variable (in der formalen Darstellung) und jeder Punkt (in der Grafik) stellen ein Ereignis dar. Für jedes dargestellte Ereignis wäre zu einem anderen Zeitpunkt

auch ein anderes Ereignis eines beliebigen Ereignistyps, also jede andere Änderung eines beliebigen Wertes, möglich.

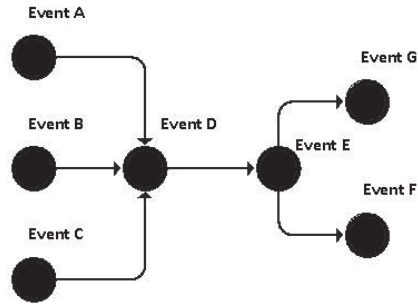


Abb. 1: Ereignisbeziehungen

In der Literatur finden sich zahlreiche Übersichten über kommerzielle und Opensource-CEP-Tools. Bruns und Dunkel nennen eine kleine Auswahl dieser Tools [BD15, S.46] und verweisen zudem für weitere Toolübersichten auf Vidačković et.al. und Vincent. Beide bieten eine sehr umfassende Übersicht und beschreiben zudem die Eigenschaften der Tools grob [VRR10], [Vin14]. In [Rob10] bietet Robins neben einer Übersicht über CEP-Tools, einige Hinweise zur Implementierung. Der Markt für CEP-Lösungen wächst sehr rasant und es besteht ein immer größerer Bedarf an CEP-Lösungen in der Praxis[Lea09].

4 CEP mit Workflowbegrenzung

Für das workfloworientierte Software-Monitoring soll beim CEP nicht mehr jedes Ereignis betrachtet werden, welches mit einem anderen Ereignis in Beziehung steht. Es sollen lediglich diejenigen Ereignisse der Ereignistypen betrachtet werden, die dem zu überwachenden Geschäftsworkflow zugeordnet werden können. Ziel dieser reduzierten Betrachtung soll die Komplexitätsreduzierung und Problemorientierung sein.

Wird für das Beispiel aus Abb.1 unterstellt, dass jedes Ereignis einen anderen Ereignistypen hat und die Ereignistypen der Ereignisse C und G für einen zu betrachtenden Workflow irrelevant sind, können sie ignoriert werden. Damit reduziert sich die Betrachtungsmenge um zwei Ereignistypen, womit ebenfalls die Reduzierung der jeweiligen Ereignisse dieses Ereignistyps einhergeht (Abb. 2). Formal reduziert sich das Ereignismuster damit auf:

$$(A \wedge B) \rightarrow D \rightarrow E \rightarrow F$$

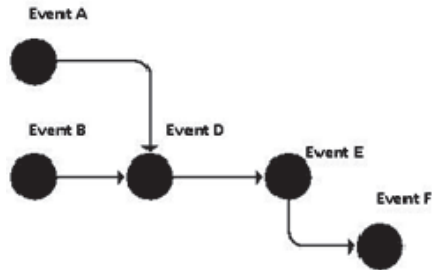


Abb. 2: Reduzierte Sicht auf Eventbeziehungen

Damit ist die Komplexität des Ereignismusters sichtlich verringert worden. Da die Software im Geschäftsworkflow auf Verfügbarkeit überwacht werden sollen, werden keine Geschäftsereignisse betrachtet, sondern rein technische Ereignisse. Für jedes technische Ereignis müssen Sensoren eingesetzt werden, was in diesem Fall Softwaremodule zur Ereigniserfassung sind. Das könnten beispielsweise Module zur Prüfung von Änderungen in Dateiverzeichnissen sein. Jeder Sensor erfasst dabei nur Ereignisse eines Ereignistyps. Die Sensoren werden so gesetzt, dass Ereignismuster entstehen, die eine Aussage über die Verfügbarkeit der Software erlauben. Dabei sollte für die Messwerte jedes Sensors bestimmt werden, inwieweit sie mit den anderen Werten korrelieren und somit relevant für das Ereignismuster sind. Es sollten allerdings nicht nur Korrelationen zwischen einzelnen Ereignissen verschiedener Ereignistypen untersucht werden, sondern auch die Korrelation zu Ereignismengen. Sensoren, die irrelevante Ereignisse erfassen, können entfernt werden.

5 Systemeinführung

Die Einführung von workfloworientiertem CEP zum Software-Monitoring setzt explizites Wissen über den Geschäftsprozess und den verwendeten Ressourcen voraus. Dieses Wissen ist essenziell für die Identifizierung der Ereignisse bzw. zunächst für die Eventtypen. Zur Generierung dieses Wissens wird vor der Systemeinführung zunächst die Durchführung der Prozessanalyse beschrieben und anschließend die Identifikation der Ereignisse.

5.1 Geschäftsprozessanalyse

Aufgrund der Zielsetzung Software zu überwachen, werden diese in der Geschäftsprozessmodellierung als Ausgangspunkt genutzt. Dabei wird der Ansatz der subjektorientierten Geschäftsprozessmodellierung verfolgt, die im Wesentlichen aus zwei Schritten besteht[SFG09]. Im ersten Schritt werden alle Subjekte und ihre

Interaktionsbeziehungen identifiziert. Das heißt, dass die beteiligten Softwaremodule und deren Nachrichtenaustausch identifiziert werden. Dabei sollten nicht nur die Softwaremodule, sondern genau der Client und der Server bestimmt werden. Im zweiten Schritt wird das Subjektverhalten spezifiziert. Das bedeutet, dass jedes Subjekt mittels der Abbildung seines Verhaltens verfeinert wird. Dabei werden streng sequenziell die Zustände und Zustandsübergänge der einzelnen ausgeführten Aktivität beschrieben. Damit werden Geschäftsereignisse identifiziert, die allerdings nicht für die Verfügbarkeitsüberwachung der Software genutzt werden. Zur Unterstützung der Einhaltung der subjektorientierten Modellierungsmethode wird das Modellierungstool Nautilus verwendet, das durch einen Formulierungsassistenten diese Methodik unterstützt. Für die Modellierung wird die eEPK-Notation genutzt, da diese eine umfassende Sicht auf den Geschäftsprozess bietet, welche für die Identifizierung der Messpunkte der technischen Ereignisse hilfreich ist. Allerdings muss zusätzlich für jede Inputentität und jede Outputentität im Model ergänzt werden, wo die Entität abgelegt ist bzw. wie die Entität weiter übertragen wird.

5.2 Identifikation der technischen Ereignisse und Ereignismuster

Für die Identifikation der technischen Ereignisse werden die Geschäftsaktivität und die angebotenen Modellelemente im Geschäftsprozessmodell betrachtet. Es wird für jede Geschäftsaktivität ein Messpunkt identifiziert, von dem aus eine Aussage über die Ausführung getroffen werden kann. Außerdem müssen Messpunkte identifiziert werden, deren Ereignisse eine Verbindung zu den Ereignissen des Aktivitätsmesspunkts aufweisen. Dafür werden auch vorhergehende Aktivitäten betrachtet. Diese Identifizierung wird anhand von zwei Szenarien beschrieben, die sich abstrahiert auf alle anderen Prozessszenarien des Beispielprozesses übertragen lassen:

- Szenario 1: Eine Software zur Konvertierung von Bestellungen liest Dateien verschiedener Datentypen beim Auslösen eines Zeittriggers aus einem Importverzeichnis und konvertiert diese in ein einheitliches Format. Anschließend werden die Dateien in das Exportverzeichnis der Software verschoben.
- Szenario 2: Eine Software zur Leistungsüberwachung prüft, ob alle Dateien ordnungsgemäß konvertiert werden. Bei der Software handelt es sich um keine proaktive Software. Jede Datei in dem Exportverzeichnis wird geöffnet und geprüft, aber nicht verändert oder weiter verarbeitet. Nach einem festdefinierten Zeitpunkt werden die Dateien aus dem Exportverzeichnis geladen und weiterverarbeitet.

Für Szenario 1 lassen sich zwei Messpunkte identifizieren, die gemeinsam betrachtet eine Aussage über die Verfügbarkeit der Software erlauben. Wird die Anzahl der Dateien im Import- und Exportverzeichnis überwacht, sollten die Änderungen in den Verzeichnissen immer ungefähr gleich sein. Damit bestehen zwei Ereignistypen, die miteinander in Verbindung stehen und deren Ereignisse somit als komplexe Ereignisse betrachtet werden können. Die Anzahl im Exportverzeichnis kann minimal abweichen,

da sich im Importverzeichnis Dateitypen befinden könnten, die nicht verarbeitet werden können. Diese Abweichung sollte allerdings für das Verfügbarkeits-Monitoring außer Acht gelassen werden. In der Tab.1 sind Messwerte enthalten, die genau diesen Sachverhalt abbilden. Zum Zeitpunkt T1 und T3 werden alle Dateien konvertiert und zum Zeitpunkt T2 wird eine Datei nicht konvertiert. Würde die Anzahl der Dateien im Exportverzeichnis stark von der Anzahl der Dateien im Importverzeichnis abweichen oder dem Wert 0 entsprechen, ist die Software oder eine Komponente der Software nicht mehr lauffähig. Demnach sind alle Ereignismuster zu identifizieren, bei denen die Änderung des Messwerts im Importverzeichnis stark von der Änderung des Messwerts im Exportverzeichnis abweicht. Für die Darstellung in den Tabellen bleibt unberücksichtigt, dass die Anzahl im Importverzeichnis und im Exportverzeichnis nicht zum selben Zeitpunkt gemessen wird, sondern die Messung der Dateianzahl im Exportverzeichnis erst nach der Konvertierung erfolgt.

Messzeitpunkt	Importverzeichnis	Exportverzeichnis
T1	71	71
T2	64	63
T3	73	73

Tab.1: Messungen für Szenario 1

Alternativ könnten auch Datenbanktabellen oder Dateigrößen überwacht werden. Dabei müssen ggf. Änderungen der Dateigröße bei Bearbeitungsprozessen einbezogen werden. Bei Softwaremodulen in denen Dateien oder andere Datenobjekte nicht abgelegt, sondern direkt weiter übertragen werden, können auch alternative Übertragungswege überwacht werden.

Für Szenario 2 lassen sich keine Daten- oder Dateibewegungen überwachen, allerdings ist die Hardwareressourcennutzung einer solchen Software abhängig von der Anzahl der Dateien. Durch diese Korrelation lässt sich auch die Veränderung der Ressourcennutzung als komplexes Ereignis betrachten. Allerdings müssen bei der Messung des Ressourcenverbrauchs die Ereignisse über einen definierten Zeitraum betrachtet werden. Daher erfolgt auf einem Ereignis bei den Messwerten im Exportverzeichnis eine Vielzahl von Ereignissen bei den Messwerten der Ressourcennutzung. In der Tab.2 ist dieser Sachverhalt abgebildet. Zum Zeitpunkt T1 ergibt sich ein Ereignis an beiden Messpunkten. An nachfolgenden Messzeitpunkten nur noch bei der RAM-Nutzung. Zur Ermittlung besserer Korrelationswerte sollten mehrere Ressourcen einbezogen werden. Sollte die Software nicht mehr das Exportverzeichnis analysieren, würde die Ressourcennutzung einen anderen Verlauf aufweisen. Eine Meldung müsste demnach dann erfolgen, wenn die Klasse von Ereignissen aus der RAM-Nutzung einen für die Anzahl an Dateien untypischen Verlauf hat.

Messzeitpunkt	Exportverzeichnis	RAM-Nutzung
T1	71	3814
T2	71	3867
T3	71	3854

Tab.2: Messungen für Szenario 2

5.3 Umsetzung in der Software

Für die Erfassung der Ereignisse müssen Softwaremodule eingerichtet werden, die als Sensoren agieren. Da im Beispielprozess alle Softwaremodulen auf Microsoft Windows-Betriebssystemen ablaufen, werden für die Softwaremodule PowerShell-Skripte eingesetzt. Um die PowerShell-Skripte nicht nur auf dem lokalen Rechner einzusetzen, werden WMI-Klassen eingesetzt, die in den PowerShell-Skripten verwendet werden. Außerdem werden Befehle aus den Cmdlets verwendet, die einen direkten Eingriff in bestehende Softwaremodule ermöglichen und somit auch Aussagen über einzelne Komponenten liefern. Jedes Skript agiert als ein Sensor, der an nur einem Ereignispunkt eingesetzt wird. Ein Sensor ist dabei eine Ausgabeschnittstelle oder Anwendungscodeerweiterung zur Datenerfassung. Er gibt beim Aufruf beispielsweise die Anzahl der aktuell im Verzeichnis vorhandenen Dateien zurück. Zur Analyse und zur anschließenden Meldung bei Ausfällen wird kein herkömmliches CEP-Tool eingesetzt, da es sich nicht ausreichend an die Prozesse und die Problemstellung anpassen lässt. Stattdessen wird der PRTG Network Monitor eingesetzt, da sich in diesem die Weiterverarbeitung der Rückgabewerte individuell gestalten lässt[Tim15]. Die Skripte werden mit dem Skriptensensor in das Tool eingebunden und anschließend wird im Formelsensor der Zusammenhang abgebildet. Der Skriptensensor und der Formelsensor sind Möglichkeiten zur individuellen Datendarstellung im PRTG Network Monitor. Dabei lassen sich mit dem Skriptensensor externe Datenquellen einbinden und mit dem Formelsensor mehrere Datenquellen zusammenfassen und Berechnungen durchführen. Jede Software ist somit in einem Formelsensor abgebildet und die unterschiedlichen Datenquellen, die somit auch eine Ereignisquelle sind, werden zu einer Aussage zusammengefasst. Zur besseren Übersicht sollten alle Sensoren in einem Prozesssensor vereint werden. Damit lassen sich alle zu überwachenden Applikationen, die dem gleichen Prozess zugeordnet werden können, schneller erfassen. Die Abb. 3 zeigt den Prozesssensor des Beispielprozesses. In dem Prozesssensor sind der Konverter (Szenario 1) und das Filetransfersystem (Szenario 2) eingebunden. Ein Ausfall einer Software würde zu einer Statusänderung im Prozess führen.

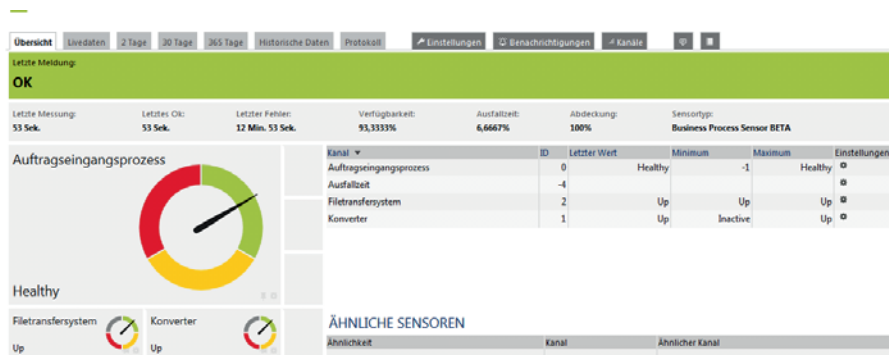


Abb. 3: Auftragseingangsprozess im PRTG Network Monitor

6 Kritische Reflektion und Ausblick

Wird CEP workfloworientiert zum Verfügbarkeits-Monitoring eingesetzt, lassen sich Softwareausfälle und sogar einzelne inaktive Softwarekomponenten identifizieren. Außerdem lässt sich, durch eine Beschränkung auf geschäftskritische Prozesse, ein Monitoring realisieren, was auf geschäftskritische Software beschränkt ist. Diese Konzentration ermöglicht die Einsparung von Ressourcen.

Allerdings setzt die Erkennung eines Ausfalls einer Software oder einer einzelnen Softwarekomponente einen nahezu perfekten Zusammenhang zwischen den Ereignissen der unterschiedlichen Messpunkte voraus. Daher stellen Verzeichnisse, die durch mehrere Softwaremodule unterschiedlicher Prozesse genutzt werden, eine Herausforderung dar. Ebenfalls ergeben sich Probleme bei der Zuordnung von Messwerten der Hardwareressourcennutzung, sofern Softwaremodule mehrmals ablaufen. Eine weitere Schwierigkeit besteht in der Fehlermeldung bei Ereignismustern, die außerhalb des erwarteten Wertebereichs liegen, aber dennoch keinen Störung der Software als Ursprung haben. Diesem Problem könnte in einer möglichen Erweiterung der vorgestellten Methodik durch Event stream processing (ESP) entgegen gewirkt werden. Zudem muss für die Inbetriebnahme des implementierten Konzepts die Fehlerbehandlung und Fehlerzuständigkeit genau geregelt werden. In dem Zusammenhang sollte auch untersucht werden, in welchem Maße sich die Fehlerbehandlung automatisieren lässt, damit der vollautomatisierte Workflowablauf gewährleistet werden kann. Darüber hinaus wäre eine Evaluation zur Aufdeckung möglicher Chancen und Risiken des beschriebenen Konzepts sinnvoll und ein Vergleich zu anderen Softwareüberwachungskonzepten aufzuzeigen. Ein weiteres Folgeprojekt wäre die Identifikation von Prozessen im Unternehmen, auf die dieses Konzept übertragen werden kann. Dabei könnte auch untersucht werden, in welchem Umfang eine Übertragbarkeit auf automatisierte Produktionsprozesse oder nur teilautomatisierte Prozesse möglich und nötig ist.

Danksagung

Ein besonderer Dank gilt Prof. Dr. Andreas Schmidt, der mich nicht nur bei der Umsetzung des vorgestellten Projekts, sondern auch bei dem Verfassen dieses Beitrages tatkräftig unterstützt hat. Dieser Beitrag entstand im Rahmen eines Praxisprojekts im Bachelorstudiengang Betriebliches Informationsmanagement an der Hochschule Osnabrück.

Literaturverzeichnis

[BD15] Bruns, R.; Dunkel, J.: Complex Event Processing. Komplexe Analyse von massiven Datenströmen mit CEP, Wiesbaden: Springer Vieweg, 2015.

- [EB09] Eckert, M.; Bry, F.: Complex Event Processing (CEP), *Informatik_Spektrum_32_2*, S.163-167, 2009.
- [Hal07] Hoontae, K.; Yong-Han, L.; Hongsoon, Y., Nam Wook C.: Design and Implementation of a Personalized Business Activity Monitoring System. In (Julie A. J. Hrsg.): *Human-Computer Interaction. HCI Applications and Services*, Berlin, Heidelberg: Springer Verlag, 2007.
- [Hei15] Hein, M.: Ende-zu-Ende-Monitoring. *Durchblick verschaffen*, IT Administrator Sonderheft II/2015, S.68-70,2015.
- [HG09] Heinz, G.; Greiner, T.: Business Activity Monitoring mit Stream Mining am Fallbeispiel TeamBank AG, *HMD – Praxis der Wirtschaftsinformatik Volume 268*, S.82-89, 2009.
- [Lea09] Leavitt, N.: Complex-Event Processing Poised for Growth, *IEEE Computer Society: Computer*, S. 17-20, 2009.
- [LS11] Luckham, D.; Schulte, R. et al: Event Processing Glossary – Version 2.0, Event Processing Technical Society, http://www.complexevents.com/wp-content/uploads/2011/08/EPTS_Event_Processing_Glossary_v2.pdf, Stand: 18.04.2016.
- [Luc02] Luckham, D.: *The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise*, Addison-Wesley, 2002.
- [Rob10] Robins, D. B.: *Complex Event Processing*, *Advanced Topics in Software Systems (CSEP 504)*, 2010.
- [Sch13] Schmidt, W.: *Business Activity Monitoring (BAM)*, In (Rausch, P et al.): *Business Intelligence and Performance Management*, London: Springer-Verlag, 2013.
- [SFG09] Schmidt, W.; Fleischmann, A.; Gilber, O.: *Subjektorientiertes Geschäftsprozessmanagement*, *HMD – Praxis der Wirtschaftsinformatik Volume 266*, S.52-62, 2009.
- [Tim15] Timmerman, T.: *Marktübersicht Monitoring. Von Äpfel und Birnen*, IT Administrator Sonderheft II/2015, S.29-35,2015.
- [Vin14] Vincent, P.: *CEP tooling market survey 2014*. <http://www.complexevents.com/2014/12/03/cep-tooling-market-survey-2014/> (2014). Stand: 18.04.2016.
- [VRR10] Vidackovic, K., Renner, T., Rex, S.: *Marktübersicht Real-Time Monitoring Software. Event Processing Tools im Überblick*, Stuttgart: Fraunhofer Verlag, 2010

GI-Edition Lecture Notes in Informatics

- P-1 Gregor Engels, Andreas Oberweis, Albert Zündorf (Hrsg.): Modellierung 2001.
- P-2 Mikhail Godlevsky, Heinrich C. Mayr (Hrsg.): Information Systems Technology and its Applications, ISTA'2001.
- P-3 Ana M. Moreno, Reind P. van de Riet (Hrsg.): Applications of Natural Language to Information Systems, NLDB'2001.
- P-4 H. Wörn, J. Mühlung, C. Vahl, H.-P. Meinzer (Hrsg.): Rechner- und sensor-gestützte Chirurgie; Workshop des SFB 414.
- P-5 Andy Schürr (Hg.): OMER – Object-Oriented Modeling of Embedded Real-Time Systems.
- P-6 Hans-Jürgen Appelpath, Rolf Beyer, Uwe Marquardt, Heinrich C. Mayr, Claudia Steinberger (Hrsg.): Unternehmen Hochschule, UH'2001.
- P-7 Andy Evans, Robert France, Ana Moreira, Bernhard Rumpe (Hrsg.): Practical UML-Based Rigorous Development Methods – Countering or Integrating the extremists, pUML'2001.
- P-8 Reinhard Keil-Slawik, Johannes Magenheim (Hrsg.): Informatikunterricht und Medienbildung, INFOS'2001.
- P-9 Jan von Knop, Wilhelm Haverkamp (Hrsg.): Innovative Anwendungen in Kommunikationsnetzen, 15. DFN Arbeitstagung.
- P-10 Mirjam Minor, Steffen Staab (Hrsg.): 1st German Workshop on Experience Management: Sharing Experiences about the Sharing Experience.
- P-11 Michael Weber, Frank Kargl (Hrsg.): Mobile Ad-Hoc Netzwerke, WMAN 2002.
- P-12 Martin Glinz, Günther Müller-Luschnat (Hrsg.): Modellierung 2002.
- P-13 Jan von Knop, Peter Schirmbacher and Viljan Mahni_ (Hrsg.): The Changing Universities – The Role of Technology.
- P-14 Robert Tolksdorf, Rainer Eckstein (Hrsg.): XML-Technologien für das Semantic Web – XSW 2002.
- P-15 Hans-Bernd Bludau, Andreas Koop (Hrsg.): Mobile Computing in Medicine.
- P-16 J. Felix Hampe, Gerhard Schwabe (Hrsg.): Mobile and Collaborative Business 2002.
- P-17 Jan von Knop, Wilhelm Haverkamp (Hrsg.): Zukunft der Netze –Die Verletzbarkeit meistern, 16. DFN Arbeitstagung.
- P-18 Elmar J. Sinz, Markus Plaha (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2002.
- P-19 Sigrid Schubert, Bernd Reusch, Norbert Jesse (Hrsg.): Informatik bewegt – Informatik 2002 – 32. Jahrestagung der Gesellschaft für Informatik e.V. (GI) 30.Sept.-3. Okt. 2002 in Dortmund.
- P-20 Sigrid Schubert, Bernd Reusch, Norbert Jesse (Hrsg.): Informatik bewegt – Informatik 2002 – 32. Jahrestagung der Gesellschaft für Informatik e.V. (GI) 30.Sept.-3. Okt. 2002 in Dortmund (Ergänzungsband).
- P-21 Jörg Desel, Mathias Weske (Hrsg.): Promise 2002: Prozessorientierte Methoden und Werkzeuge für die Entwicklung von Informationssystemen.
- P-22 Sigrid Schubert, Johannes Magenheim, Peter Hubwieser, Torsten Brinda (Hrsg.): Forschungsbeiträge zur "Didaktik der Informatik" – Theorie, Praxis, Evaluation.
- P-23 Thorsten Spitta, Jens Borchers, Harry M. Sneed (Hrsg.): Software Management 2002 – Fortschritt durch Beständigkeit
- P-24 Rainer Eckstein, Robert Tolksdorf (Hrsg.): XMIDX 2003 – XML-Technologien für Middleware – Middleware für XML-Anwendungen
- P-25 Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Commerce – Anwendungen und Perspektiven – 3. Workshop Mobile Commerce, Universität Augsburg, 04.02.2003
- P-26 Gerhard Weikum, Harald Schöning, Erhard Rahm (Hrsg.): BTW 2003: Datenbanksysteme für Business, Technologie und Web
- P-27 Michael Kroll, Hans-Gerd Lipinski, Kay Melzer (Hrsg.): Mobiles Computing in der Medizin
- P-28 Ulrich Reimer, Andreas Abecker, Steffen Staab, Gerd Stumme (Hrsg.): WM 2003: Professionelles Wissensmanagement – Erfahrungen und Visionen
- P-29 Antje Düsterhöft, Bernhard Thalheim (Eds.): NLDB'2003: Natural Language Processing and Information Systems
- P-30 Mikhail Godlevsky, Stephen Liddle, Heinrich C. Mayr (Eds.): Information Systems Technology and its Applications
- P-31 Arslan Brömme, Christoph Busch (Eds.): BIOSIG 2003: Biometrics and Electronic Signatures

- P-32 Peter Hubwieser (Hrsg.): Informatische Fachkonzepte im Unterricht – INFOS 2003
- P-33 Andreas Geyer-Schulz, Alfred Taudes (Hrsg.): Informationswirtschaft: Ein Sektor mit Zukunft
- P-34 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenber, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 1)
- P-35 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenber, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 2)
- P-36 Rüdiger Grimm, Hubert B. Keller, Kai Rannenber (Hrsg.): Informatik 2003 – Mit Sicherheit Informatik
- P-37 Arndt Bode, Jörg Desel, Sabine Rathmayer, Martin Wessner (Hrsg.): DeLFI 2003: e-Learning Fachtagung Informatik
- P-38 E.J. Sinz, M. Plaha, P. Neckel (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2003
- P-39 Jens Nedon, Sandra Frings, Oliver Göbel (Hrsg.): IT-Incident Management & IT-Forensics – IMF 2003
- P-40 Michael Rebstock (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2004
- P-41 Uwe Brinkschulte, Jürgen Becker, Dietmar Fey, Karl-Erwin Großpietsch, Christian Hochberger, Erik Maehle, Thomas Runkler (Edts.): ARCS 2004 – Organic and Pervasive Computing
- P-42 Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Economy – Transaktionen und Prozesse, Anwendungen und Dienste
- P-43 Birgitta König-Ries, Michael Klein, Philipp Obreiter (Hrsg.): Persistence, Scalability, Transactions – Database Mechanisms for Mobile Applications
- P-44 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): Security, E-Learning, E-Services
- P-45 Bernhard Rumpe, Wolfgang Hesse (Hrsg.): Modellierung 2004
- P-46 Ulrich Flegel, Michael Meier (Hrsg.): Detection of Intrusions of Malware & Vulnerability Assessment
- P-47 Alexander Prosser, Robert Krimmer (Hrsg.): Electronic Voting in Europe – Technology, Law, Politics and Society
- P-48 Anatoly Doroshenko, Terry Halpin, Stephen W. Liddle, Heinrich C. Mayr (Hrsg.): Information Systems Technology and its Applications
- P-49 G. Schiefer, P. Wagner, M. Morgenstern, U. Rickert (Hrsg.): Integration und Datensicherheit – Anforderungen, Konflikte und Perspektiven
- P-50 Peter Dadam, Manfred Reichert (Hrsg.): INFORMATIK 2004 – Informatik verbindet (Band 1) Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 20.-24. September 2004 in Ulm
- P-51 Peter Dadam, Manfred Reichert (Hrsg.): INFORMATIK 2004 – Informatik verbindet (Band 2) Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 20.-24. September 2004 in Ulm
- P-52 Gregor Engels, Silke Seehusen (Hrsg.): DELFI 2004 – Tagungsband der 2. e-Learning Fachtagung Informatik
- P-53 Robert Giegerich, Jens Stoye (Hrsg.): German Conference on Bioinformatics – GCB 2004
- P-54 Jens Borchers, Ralf Kneuper (Hrsg.): Softwaremanagement 2004 – Outsourcing und Integration
- P-55 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): E-Science und Grid Ad-hoc-Netze Medienintegration
- P-56 Fernand Feltz, Andreas Oberweis, Benoit Otjacques (Hrsg.): EMISA 2004 – Informationssysteme im E-Business und E-Government
- P-57 Klaus Turowski (Hrsg.): Architekturen, Komponenten, Anwendungen
- P-58 Sami Beydeda, Volker Gruhn, Johannes Mayer, Ralf Reussner, Franz Schweiggert (Hrsg.): Testing of Component-Based Systems and Software Quality
- P-59 J. Felix Hampe, Franz Lehner, Key Pousttchi, Kai Rannenber, Klaus Turowski (Hrsg.): Mobile Business – Processes, Platforms, Payments
- P-60 Steffen Friedrich (Hrsg.): Unterrichtskonzepte für informatische Bildung
- P-61 Paul Müller, Reinhard Gotzhein, Jens B. Schmitt (Hrsg.): Kommunikation in verteilten Systemen
- P-62 Federrath, Hannes (Hrsg.): „Sicherheit 2005“ – Sicherheit – Schutz und Zuverlässigkeit
- P-63 Roland Kaschek, Heinrich C. Mayr, Stephen Liddle (Hrsg.): Information Systems – Technology and its Applications

- P-64 Peter Liggesmeyer, Klaus Pohl, Michael Goedicke (Hrsg.): Software Engineering 2005
- P-65 Gottfried Vossen, Frank Leymann, Peter Lockemann, Wolfrid Stucky (Hrsg.): Datenbanksysteme in Business, Technologie und Web
- P-66 Jörg M. Haake, Ulrike Lucke, Djamshid Tavangarian (Hrsg.): DeLFI 2005: 3. deutsche e-Learning Fachtagung Informatik
- P-67 Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinhage (Hrsg.): INFORMATIK 2005 – Informatik LIVE (Band 1)
- P-68 Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinhage (Hrsg.): INFORMATIK 2005 – Informatik LIVE (Band 2)
- P-69 Robert Hirschfeld, Ryszard Kowalczyk, Andreas Polze, Matthias Weske (Hrsg.): NODe 2005, GSEM 2005
- P-70 Klaus Turowski, Johannes-Maria Zaha (Hrsg.): Component-oriented Enterprise Application (COAE 2005)
- P-71 Andrew Torda, Stefan Kurz, Matthias Rarey (Hrsg.): German Conference on Bioinformatics 2005
- P-72 Klaus P. Jantke, Klaus-Peter Fähnrich, Wolfgang S. Wittig (Hrsg.): Marktplatz Internet: Von e-Learning bis e-Payment
- P-73 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): "Heute schon das Morgen sehen"
- P-74 Christopher Wolf, Stefan Lucks, Po-Wah Yau (Hrsg.): WEWoRC 2005 – Western European Workshop on Research in Cryptology
- P-75 Jörg Desel, Ulrich Frank (Hrsg.): Enterprise Modelling and Information Systems Architecture
- P-76 Thomas Kirste, Birgitta König-Riess, Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Informationssysteme – Potentiale, Hindernisse, Einsatz
- P-77 Jana Dittmann (Hrsg.): SICHERHEIT 2006
- P-78 K.-O. Wenkel, P. Wagner, M. Morgens-tern, K. Luzi, P. Eisermann (Hrsg.): Land- und Ernährungswirtschaft im Wandel
- P-79 Bettina Biel, Matthias Book, Volker Gruhn (Hrsg.): Softwareengineering 2006
- P-80 Mareike Schoop, Christian Huemer, Michael Rebstock, Martin Bichler (Hrsg.): Service-Oriented Electronic Commerce
- P-81 Wolfgang Karl, Jürgen Becker, Karl-Erwin Großpietsch, Christian Hochberger, Erik Maehle (Hrsg.): ARCS'06
- P-82 Heinrich C. Mayr, Ruth Breu (Hrsg.): Modellierung 2006
- P-83 Daniel Huson, Oliver Kohlbacher, Andrei Lupas, Kay Nieselt and Andreas Zell (eds.): German Conference on Bioinformatics
- P-84 Dimitris Karagiannis, Heinrich C. Mayr, (Hrsg.): Information Systems Technology and its Applications
- P-85 Witold Abramowicz, Heinrich C. Mayr, (Hrsg.): Business Information Systems
- P-86 Robert Krimmer (Ed.): Electronic Voting 2006
- P-87 Max Mühlhäuser, Guido Rößling, Ralf Steinmetz (Hrsg.): DELFI 2006: 4. e-Learning Fachtagung Informatik
- P-88 Robert Hirschfeld, Andreas Polze, Ryszard Kowalczyk (Hrsg.): NODe 2006, GSEM 2006
- P-90 Joachim Schelp, Robert Winter, Ulrich Frank, Bodo Rieger, Klaus Turowski (Hrsg.): Integration, Informationslogistik und Architektur
- P-91 Henrik Stormer, Andreas Meier, Michael Schumacher (Eds.): European Conference on eHealth 2006
- P-92 Fernand Feltz, Benoît Otjacques, Andreas Oberweis, Nicolas Poussing (Eds.): AIM 2006
- P-93 Christian Hochberger, Rüdiger Liskowsky (Eds.): INFORMATIK 2006 – Informatik für Menschen, Band 1
- P-94 Christian Hochberger, Rüdiger Liskowsky (Eds.): INFORMATIK 2006 – Informatik für Menschen, Band 2
- P-95 Matthias Weske, Markus Nüttgens (Eds.): EMISA 2005: Methoden, Konzepte und Technologien für die Entwicklung von dienstbasierten Informationssystemen
- P-96 Saartje Brockmans, Jürgen Jung, York Sure (Eds.): Meta-Modelling and Ontologies
- P-97 Oliver Göbel, Dirk Schadt, Sandra Frings, Hardo Hase, Detlef Günther, Jens Nedon (Eds.): IT-Incident Mangament & IT-Forensics – IMF 2006

- P-98 Hans Brandt-Pook, Werner Simonsmeier und Thorsten Spitta (Hrsg.): Beratung in der Softwareentwicklung – Modelle, Methoden, Best Practices
- P-99 Andreas Schwill, Carsten Schulte, Marco Thomas (Hrsg.): Didaktik der Informatik
- P-100 Peter Forbrig, Günter Siegel, Markus Schneider (Hrsg.): HDI 2006: Hochschuldidaktik der Informatik
- P-101 Stefan Böttinger, Ludwig Theuvsen, Susanne Rank, Marlies Morgenstern (Hrsg.): Agrarinformatik im Spannungsfeld zwischen Regionalisierung und globalen Wertschöpfungsketten
- P-102 Otto Spaniol (Eds.): Mobile Services and Personalized Environments
- P-103 Alfons Kemper, Harald Schöning, Thomas Rose, Matthias Jarke, Thomas Seidl, Christoph Quix, Christoph Brochhaus (Hrsg.): Datenbanksysteme in Business, Technologie und Web (BTW 2007)
- P-104 Birgitta König-Ries, Franz Lehner, Rainer Malaka, Can Türker (Hrsg.) MMS 2007: Mobilität und mobile Informationssysteme
- P-105 Wolf-Gideon Bleek, Jörg Raasch, Heinz Züllighoven (Hrsg.) Software Engineering 2007
- P-106 Wolf-Gideon Bleek, Henning Schwentner, Heinz Züllighoven (Hrsg.) Software Engineering 2007 – Beiträge zu den Workshops
- P-107 Heinrich C. Mayr, Dimitris Karagiannis (eds.) Information Systems Technology and its Applications
- P-108 Arslan Brömme, Christoph Busch, Detlef Hühnlein (eds.) BIOSIG 2007: Biometrics and Electronic Signatures
- P-109 Rainer Koschke, Otthein Herzog, Karl-Heinz Rödiger, Marc Ronthaler (Hrsg.) INFORMATIK 2007 Informatik trifft Logistik Band 1
- P-110 Rainer Koschke, Otthein Herzog, Karl-Heinz Rödiger, Marc Ronthaler (Hrsg.) INFORMATIK 2007 Informatik trifft Logistik Band 2
- P-111 Christian Eibl, Johannes Magenheimer, Sigrid Schubert, Martin Wessner (Hrsg.) DeLFI 2007: 5. e-Learning Fachtagung Informatik
- P-112 Sigrid Schubert (Hrsg.) Didaktik der Informatik in Theorie und Praxis
- P-113 Sören Auer, Christian Bizer, Claudia Müller, Anna V. Zhdanova (Eds.) The Social Semantic Web 2007 Proceedings of the 1st Conference on Social Semantic Web (CSSW)
- P-114 Sandra Frings, Oliver Göbel, Detlef Günther, Hardo G. Hase, Jens Nedon, Dirk Schadt, Arslan Brömme (Eds.) IMF2007 IT-incident management & IT-forensics Proceedings of the 3rd International Conference on IT-Incident Management & IT-Forensics
- P-115 Claudia Falter, Alexander Schliep, Joachim Selbig, Martin Vingron and Dirk Walthert (Eds.) German conference on bioinformatics GCB 2007
- P-116 Witold Abramowicz, Leszek Maciszek (Eds.) Business Process and Services Computing 1st International Working Conference on Business Process and Services Computing BPSC 2007
- P-117 Ryszard Kowalczyk (Ed.) Grid service engineering and management The 4th International Conference on Grid Service Engineering and Management GSEM 2007
- P-118 Andreas Hein, Wilfried Thoben, Hans-Jürgen Appelrath, Peter Jensch (Eds.) European Conference on ehealth 2007
- P-119 Manfred Reichert, Stefan Strecker, Klaus Turowski (Eds.) Enterprise Modelling and Information Systems Architectures Concepts and Applications
- P-120 Adam Pawlak, Kurt Sandkuhl, Wojciech Cholewa, Leandro Soares Indrusiak (Eds.) Coordination of Collaborative Engineering - State of the Art and Future Challenges
- P-121 Korbinian Herrmann, Bernd Bruegge (Hrsg.) Software Engineering 2008 Fachtagung des GI-Fachbereichs Softwaretechnik
- P-122 Walid Maalej, Bernd Bruegge (Hrsg.) Software Engineering 2008 - Workshopband Fachtagung des GI-Fachbereichs Softwaretechnik

- P-123 Michael H. Breitner, Martin Breunig, Elgar Fleisch, Ley Pousttchi, Klaus Turowski (Hrsg.)
Mobile und Ubiquitäre Informationssysteme – Technologien, Prozesse, Marktfähigkeit
Proceedings zur 3. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2008)
- P-124 Wolfgang E. Nagel, Rolf Hoffmann, Andreas Koch (Eds.)
9th Workshop on Parallel Systems and Algorithms (PASA)
Workshop of the GI/ITG Special Interest Groups PARS and PARVA
- P-125 Rolf A.E. Müller, Hans-H. Sundermeier, Ludwig Theuvsen, Stephanie Schütze, Marlies Morgenstern (Hrsg.)
Unternehmens-IT: Führungsinstrument oder Verwaltungsbürde
Referate der 28. GIL Jahrestagung
- P-126 Rainer Gimnich, Uwe Kaiser, Jochen Quante, Andreas Winter (Hrsg.)
10th Workshop Software Reengineering (WSR 2008)
- P-127 Thomas Kühne, Wolfgang Reisig, Friedrich Steimann (Hrsg.)
Modellierung 2008
- P-128 Ammar Alkassar, Jörg Siekmann (Hrsg.)
Sicherheit 2008
Sicherheit, Schutz und Zuverlässigkeit
Beiträge der 4. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)
2.-4. April 2008
Saarbrücken, Germany
- P-129 Wolfgang Hesse, Andreas Oberweis (Eds.)
Sigsand-Europe 2008
Proceedings of the Third AIS SIGSAND European Symposium on Analysis, Design, Use and Societal Impact of Information Systems
- P-130 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
1. DFN-Forum Kommunikationstechnologien Beiträge der Fachtagung
- P-131 Robert Krimmer, Rüdiger Grimm (Eds.)
3rd International Conference on Electronic Voting 2008
Co-organized by Council of Europe, Gesellschaft für Informatik und E-Voting, CC
- P-132 Silke Seehusen, Ulrike Lucke, Stefan Fischer (Hrsg.)
DeLFI 2008:
Die 6. e-Learning Fachtagung Informatik
- P-133 Heinz-Gerd Hegering, Axel Lehmann, Hans Jürgen Ohlbach, Christian Scheideler (Hrsg.)
INFORMATIK 2008
Beherrschbare Systeme – dank Informatik Band 1
- P-134 Heinz-Gerd Hegering, Axel Lehmann, Hans Jürgen Ohlbach, Christian Scheideler (Hrsg.)
INFORMATIK 2008
Beherrschbare Systeme – dank Informatik Band 2
- P-135 Torsten Brinda, Michael Fothe, Peter Hubwieser, Kirsten Schlüter (Hrsg.)
Didaktik der Informatik – Aktuelle Forschungsergebnisse
- P-136 Andreas Beyer, Michael Schroeder (Eds.)
German Conference on Bioinformatics GCB 2008
- P-137 Arslan Brömme, Christoph Busch, Detlef Hühlein (Eds.)
BIOSIG 2008: Biometrics and Electronic Signatures
- P-138 Barbara Dinter, Robert Winter, Peter Chamoni, Norbert Gronau, Klaus Turowski (Hrsg.)
Synergien durch Integration und Informationslogistik
Proceedings zur DW2008
- P-139 Georg Herzwurm, Martin Mikusz (Hrsg.)
Industrialisierung des Software-Managements
Fachtagung des GI-Fachausschusses Management der Anwendungsentwicklung und -wartung im Fachbereich Wirtschaftsinformatik
- P-140 Oliver Göbel, Sandra Frings, Detlef Günther, Jens Nedon, Dirk Schadt (Eds.)
IMF 2008 - IT Incident Management & IT Forensics
- P-141 Peter Loos, Markus Nüttgens, Klaus Turowski, Dirk Werth (Hrsg.)
Modellierung betrieblicher Informationssysteme (MobIS 2008)
Modellierung zwischen SOA und Compliance Management
- P-142 R. Bill, P. Korduan, L. Theuvsen, M. Morgenstern (Hrsg.)
Anforderungen an die Agrarinformatik durch Globalisierung und Klimaveränderung
- P-143 Peter Liggesmeyer, Gregor Engels, Jürgen Münch, Jörg Dörr, Norman Riegel (Hrsg.)
Software Engineering 2009
Fachtagung des GI-Fachbereichs Softwaretechnik

- P-144 Johann-Christoph Freytag, Thomas Ruf, Wolfgang Lehner, Gottfried Vossen (Hrsg.)
Datenbanksysteme in Business, Technologie und Web (BTW)
- P-145 Knut Hinkelmann, Holger Wache (Eds.)
WM2009: 5th Conference on Professional Knowledge Management
- P-146 Markus Bick, Martin Breunig, Hagen Höpfner (Hrsg.)
Mobile und Ubiquitäre Informationssysteme – Entwicklung, Implementierung und Anwendung
4. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2009)
- P-147 Witold Abramowicz, Leszek Maciaszek, Ryszard Kowalczyk, Andreas Speck (Eds.)
Business Process, Services Computing and Intelligent Service Management
BPSC 2009 · ISM 2009 · YRW-MBP 2009
- P-148 Christian Erfurth, Gerald Eichler, Volkmar Schau (Eds.)
9th International Conference on Innovative Internet Community Systems
I²CS 2009
- P-149 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
2. DFN-Forum
Kommunikationstechnologien
Beiträge der Fachtagung
- P-150 Jürgen Münch, Peter Liggesmeyer (Hrsg.)
Software Engineering
2009 - Workshopband
- P-151 Armin Heinzl, Peter Dadam, Stefan Kirm, Peter Lockemann (Eds.)
PRIMIUM
Process Innovation for Enterprise Software
- P-152 Jan Mendling, Stefanie Rinderle-Ma, Werner Esswein (Eds.)
Enterprise Modelling and Information Systems Architectures
Proceedings of the 3rd Int'l Workshop EMISA 2009
- P-153 Andreas Schwill, Nicolas Apostolopoulos (Hrsg.)
Lernen im Digitalen Zeitalter
DeLFI 2009 – Die 7. E-Learning Fachtagung Informatik
- P-154 Stefan Fischer, Erik Maehle, Rüdiger Reischuk (Hrsg.)
INFORMATIK 2009
Im Focus das Leben
- P-155 Arslan Brömme, Christoph Busch, Detlef Hühnlein (Eds.)
BIOSIG 2009:
Biometrics and Electronic Signatures
Proceedings of the Special Interest Group on Biometrics and Electronic Signatures
- P-156 Bernhard Koerber (Hrsg.)
Zukunft braucht Herkunft
25 Jahre »INFOS – Informatik und Schule«
- P-157 Ivo Grosse, Steffen Neumann, Stefan Posch, Falk Schreiber, Peter Stadler (Eds.)
German Conference on Bioinformatics 2009
- P-158 W. Claudepein, L. Theuvsen, A. Kämpf, M. Morgenstern (Hrsg.)
Precision Agriculture
Reloaded – Informationsgestützte Landwirtschaft
- P-159 Gregor Engels, Markus Luckey, Wilhelm Schäfer (Hrsg.)
Software Engineering 2010
- P-160 Gregor Engels, Markus Luckey, Alexander Pretschner, Ralf Reussner (Hrsg.)
Software Engineering 2010 –
Workshopband
(inkl. Doktorandensymposium)
- P-161 Gregor Engels, Dimitris Karagiannis, Heinrich C. Mayr (Hrsg.)
Modellierung 2010
- P-162 Maria A. Wimmer, Uwe Brinkhoff, Siegfried Kaiser, Dagmar Lück-Schneider, Erich Schweighofer, Andreas Wiebe (Hrsg.)
Vernetzte IT für einen effektiven Staat
Gemeinsame Fachtagung
Verwaltungsinformatik (FTVI) und
Fachtagung Rechtsinformatik (FTRI) 2010
- P-163 Markus Bick, Stefan Eulgem, Elgar Fleisch, J. Felix Hampe, Birgitta König-Ries, Franz Lehner, Key Pousttchi, Kai Rannenberg (Hrsg.)
Mobile und Ubiquitäre Informationssysteme
Technologien, Anwendungen und Dienste zur Unterstützung von mobiler
Kollaboration
- P-164 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2010: Biometrics and Electronic Signatures
Proceedings of the Special Interest Group on Biometrics and Electronic Signatures

- P-165 Gerald Eichler, Peter Kropf, Ulrike Lechner, Phayung Meesad, Herwig Unger (Eds.)
10th International Conference on Innovative Internet Community Systems (I²CS) – Jubilee Edition 2010 –
- P-166 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
3. DFN-Forum Kommunikationstechnologien Beiträge der Fachtagung
- P-167 Robert Krimmer, Rüdiger Grimm (Eds.)
4th International Conference on Electronic Voting 2010
co-organized by the Council of Europe, Gesellschaft für Informatik and E-Voting.CC
- P-168 Ira Diethelm, Christina Dörge, Claudia Hildebrandt, Carsten Schulte (Hrsg.)
Didaktik der Informatik
Möglichkeiten empirischer Forschungsmethoden und Perspektiven der Fachdidaktik
- P-169 Michael Kerres, Nadine Ojstersek, Ulrik Schroeder, Ulrich Hoppe (Hrsg.)
DeLFI 2010 - 8. Tagung der Fachgruppe E-Learning der Gesellschaft für Informatik e.V.
- P-170 Felix C. Freiling (Hrsg.)
Sicherheit 2010
Sicherheit, Schutz und Zuverlässigkeit
- P-171 Werner Esswein, Klaus Turowski, Martin Juhrisch (Hrsg.)
Modellierung betrieblicher Informationssysteme (MobIS 2010)
Modellgestütztes Management
- P-172 Stefan Klink, Agnes Koschmider, Marco Mevius, Andreas Oberweis (Hrsg.)
EMISA 2010
Einflussfaktoren auf die Entwicklung flexibler, integrierter Informationssysteme
Beiträge des Workshops der GI-Fachgruppe EMISA (Entwicklungsmethoden für Informationssysteme und deren Anwendung)
- P-173 Dietmar Schomburg, Andreas Grote (Eds.)
German Conference on Bioinformatics 2010
- P-174 Arslan Brömme, Torsten Eymann, Detlef Hühnlein, Heiko Roßnagel, Paul Schmücker (Hrsg.)
perspeGktive 2010
Workshop „Innovative und sichere Informationstechnologie für das Gesundheitswesen von morgen“
- P-175 Klaus-Peter Fähnrich, Bogdan Franczyk (Hrsg.)
INFORMATIK 2010
Service Science – Neue Perspektiven für die Informatik
Band 1
- P-176 Klaus-Peter Fähnrich, Bogdan Franczyk (Hrsg.)
INFORMATIK 2010
Service Science – Neue Perspektiven für die Informatik
Band 2
- P-177 Witold Abramowicz, Rainer Alt, Klaus-Peter Fähnrich, Bogdan Franczyk, Leszek A. Maciaszek (Eds.)
INFORMATIK 2010
Business Process and Service Science – Proceedings of ISSS and BPSC
- P-178 Wolfram Pietsch, Benedikt Krams (Hrsg.)
Vom Projekt zum Produkt
Fachtagung des GI-Fachausschusses Management der Anwendungsentwicklung und -wartung im Fachbereich Wirtschafts-informatik (WI-MAW), Aachen, 2010
- P-179 Stefan Gruner, Bernhard Rumpe (Eds.)
FM+AM'2010
Second International Workshop on Formal Methods and Agile Methods
- P-180 Theo Härder, Wolfgang Lehner, Bernhard Mitschang, Harald Schöning, Holger Schwarz (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW) 14. Fachtagung des GI-Fachbereichs „Datenbanken und Informationssysteme“ (DBIS)
- P-181 Michael Clasen, Otto Schätzel, Brigitte Theuvsen (Hrsg.)
Qualität und Effizienz durch informationsgestützte Landwirtschaft, Fokus: Moderne Weinwirtschaft
- P-182 Ronald Maier (Hrsg.)
6th Conference on Professional Knowledge Management
From Knowledge to Action
- P-183 Ralf Reussner, Matthias Grund, Andreas Oberweis, Walter Tichy (Hrsg.)
Software Engineering 2011
Fachtagung des GI-Fachbereichs Softwaretechnik
- P-184 Ralf Reussner, Alexander Pretschner, Stefan Jähnichen (Hrsg.)
Software Engineering 2011
Workshopband
(inkl. Doktorandensymposium)

- P-185 Hagen Höpfner, Günther Specht, Thomas Ritz, Christian Bunse (Hrsg.)
MMS 2011: Mobile und ubiquitäre Informationssysteme Proceedings zur 6. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2011)
- P-186 Gerald Eichler, Axel Küpper, Volkmar Schau, Hacène Fouchal, Herwig Unger (Eds.)
11th International Conference on Innovative Internet Community Systems (I²CS)
- P-187 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
4. DFN-Forum Kommunikationstechnologien, Beiträge der Fachtagung 20. Juni bis 21. Juni 2011 Bonn
- P-188 Holger Rohland, Andrea Kienle, Steffen Friedrich (Hrsg.)
DeLFI 2011 – Die 9. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. 5.–8. September 2011, Dresden
- P-189 Thomas, Marco (Hrsg.)
Informatik in Bildung und Beruf INFOS 2011
14. GI-Fachtagung Informatik und Schule
- P-190 Markus Nüttgens, Oliver Thomas, Barbara Weber (Eds.)
Enterprise Modelling and Information Systems Architectures (EMISA 2011)
- P-191 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2011
International Conference of the Biometrics Special Interest Group
- P-192 Hans-Ulrich Heiß, Peter Pepper, Holger Schlingloff, Jörg Schneider (Hrsg.)
INFORMATIK 2011
Informatik schafft Communities
- P-193 Wolfgang Lehner, Gunther Piller (Hrsg.)
IMDM 2011
- P-194 M. Clasen, G. Fröhlich, H. Bernhardt, K. Hildebrand, B. Theuvsen (Hrsg.)
Informationstechnologie für eine nachhaltige Landwirtschaft Fokus Forstwirtschaft
- P-195 Neeraj Suri, Michael Waidner (Hrsg.)
Sicherheit 2012
Sicherheit, Schutz und Zuverlässigkeit Beiträge der 6. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)
- P-196 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2012
Proceedings of the 11th International Conference of the Biometrics Special Interest Group
- P-197 Jörn von Lucke, Christian P. Geiger, Siegfried Kaiser, Erich Schweighofer, Maria A. Wimmer (Hrsg.)
Auf dem Weg zu einer offenen, smarten und vernetzten Verwaltungskultur Gemeinsame Fachtagung Verwaltungsinformatik (FTVI) und Fachtagung Rechtsinformatik (FTRI) 2012
- P-198 Stefan Jähnichen, Axel Küpper, Sahin Albayrak (Hrsg.)
Software Engineering 2012
Fachtagung des GI-Fachbereichs Softwaretechnik
- P-199 Stefan Jähnichen, Bernhard Rumpe, Holger Schlingloff (Hrsg.)
Software Engineering 2012
Workshopband
- P-200 Gero Mühl, Jan Richling, Andreas Herkersdorf (Hrsg.)
ARCS 2012 Workshops
- P-201 Elmar J. Sinz Andy Schürr (Hrsg.)
Modellierung 2012
- P-202 Andrea Back, Markus Bick, Martin Breunig, Key Pousttchi, Frédéric Thiesse (Hrsg.)
MMS 2012: Mobile und Ubiquitäre Informationssysteme
- P-203 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
5. DFN-Forum Kommunikationstechnologien
Beiträge der Fachtagung
- P-204 Gerald Eichler, Leendert W. M. Wienhofen, Anders Kofod-Petersen, Herwig Unger (Eds.)
12th International Conference on Innovative Internet Community Systems (I²CS 2012)
- P-205 Manuel J. Kripp, Melanie Volkamer, Rüdiger Grimm (Eds.)
5th International Conference on Electronic Voting 2012 (EVOTE2012)
Co-organized by the Council of Europe, Gesellschaft für Informatik und E-Voting.CC
- P-206 Stefanie Rinderle-Ma, Mathias Weske (Hrsg.)
EMISA 2012
Der Mensch im Zentrum der Modellierung
- P-207 Jörg Desel, Jörg M. Haake, Christian Spannagel (Hrsg.)
DeLFI 2012: Die 10. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V.
24.–26. September 2012

- P-208 Ursula Goltz, Marcus Magnor, Hans-Jürgen Appelrath, Herbert Matthies, Wolf-Tilo Balke, Lars Wolf (Hrsg.)
INFORMATIK 2012
- P-209 Hans Brandt-Pook, André Fleer, Thorsten Spitta, Malte Wattenberg (Hrsg.)
Nachhaltiges Software Management
- P-210 Erhard Plödereder, Peter Dencker, Herbert Klenk, Hubert B. Keller, Silke Spitzer (Hrsg.)
Automotive – Safety & Security 2012
Sicherheit und Zuverlässigkeit für automobile Informationstechnik
- P-211 M. Clasen, K. C. Kersebaum, A. Meyer-Aurich, B. Theuvsen (Hrsg.)
Massendatenmanagement in der Agrar- und Ernährungswirtschaft
Erhebung - Verarbeitung - Nutzung
Referate der 33. GIL-Jahrestagung
20. – 21. Februar 2013, Potsdam
- P-212 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2013
Proceedings of the 12th International Conference of the Biometrics Special Interest Group
04.–06. September 2013
Darmstadt, Germany
- P-213 Stefan Kowalewski, Bernhard Rumpe (Hrsg.)
Software Engineering 2013
Fachtagung des GI-Fachbereichs Softwaretechnik
- P-214 Volker Markl, Gunter Saake, Kai-Uwe Sattler, Gregor Hackenbroich, Bernhard Mitschang, Theo Härder, Veit Köppen (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW) 2013
13. – 15. März 2013, Magdeburg
- P-215 Stefan Wagner, Horst Lichter (Hrsg.)
Software Engineering 2013
Workshopband
(inkl. Doktorandensymposium)
26. Februar – 1. März 2013, Aachen
- P-216 Gunter Saake, Andreas Henrich, Wolfgang Lehner, Thomas Neumann, Veit Köppen (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW) 2013 – Workshopband
11. – 12. März 2013, Magdeburg
- P-217 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
6. DFN-Forum Kommunikationstechnologien
Beiträge der Fachtagung
03.–04. Juni 2013, Erlangen
- P-218 Andreas Breiter, Christoph Rensing (Hrsg.)
DeLFI 2013: Die 11 e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. (GI)
8. – 11. September 2013, Bremen
- P-219 Norbert Breier, Peer Stechert, Thomas Wilke (Hrsg.)
Informatik erweitert Horizonte
INFOS 2013
15. GI-Fachtagung Informatik und Schule
26. – 28. September 2013
- P-220 Matthias Horbach (Hrsg.)
INFORMATIK 2013
Informatik angepasst an Mensch, Organisation und Umwelt
16. – 20. September 2013, Koblenz
- P-221 Maria A. Wimmer, Marijn Janssen, Ann Macintosh, Hans Jochen Scholl, Efthimos Tambouris (Eds.)
Electronic Government and Electronic Participation
Joint Proceedings of Ongoing Research of IFIP EGOV and IFIP ePart 2013
16. – 19. September 2013, Koblenz
- P-222 Reinhard Jung, Manfred Reichert (Eds.)
Enterprise Modelling and Information Systems Architectures (EMISA 2013)
St. Gallen, Switzerland
September 5. – 6. 2013
- P-223 Detlef Hühnlein, Heiko Roßnagel (Hrsg.)
Open Identity Summit 2013
10. – 11. September 2013
Kloster Banz, Germany
- P-224 Eckhart Hanser, Martin Mikusz, Masud Fazal-Baqaie (Hrsg.)
Vorgehensmodelle 2013
Vorgehensmodelle – Anspruch und Wirklichkeit
20. Tagung der Fachgruppe Vorgehensmodelle im Fachgebiet Wirtschaftsinformatik (WI-VM) der Gesellschaft für Informatik e.V.
Lörrach, 2013
- P-225 Hans-Georg Fill, Dimitris Karagiannis, Ulrich Reimer (Hrsg.)
Modellierung 2014
19. – 21. März 2014, Wien
- P-226 M. Clasen, M. Hamer, S. Lehnert, B. Petersen, B. Theuvsen (Hrsg.)
IT-Standards in der Agrar- und Ernährungswirtschaft Fokus: Risiko- und Krisenmanagement
Referate der 34. GIL-Jahrestagung
24. – 25. Februar 2014, Bonn

- P-227 Wilhelm Hasselbring,
Nils Christian Ehmke (Hrsg.)
Software Engineering 2014
Fachtagung des GI-Fachbereichs
Softwaretechnik
25. – 28. Februar 2014
Kiel, Deutschland
- P-228 Stefan Katzenbeisser, Volkmar Lotz,
Edgar Weippl (Hrsg.)
Sicherheit 2014
Sicherheit, Schutz und Zuverlässigkeit
Beiträge der 7. Jahrestagung des
Fachbereichs Sicherheit der
Gesellschaft für Informatik e.V. (GI)
19. – 21. März 2014, Wien
- P-230 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2014
Proceedings of the 13th International
Conference of the Biometrics Special
Interest Group
10. – 12. September 2014 in
Darmstadt, Germany
- P-231 Paul Müller, Bernhard Neumair,
Helmut Reiser, Gabi Dreo Rodosek
(Hrsg.)
7. DFN-Forum
Kommunikationstechnologien
16. – 17. Juni 2014
Fulda
- P-232 E. Plödereder, L. Grunske, E. Schneider,
D. Ull (Hrsg.)
INFORMATIK 2014
Big Data – Komplexität meistern
22. – 26. September 2014
Stuttgart
- P-233 Stephan Trahasch, Rolf Plötzner, Gerhard
Schneider, Claudia Gayer, Daniel Sassiati,
Nicole Wöhrle (Hrsg.)
DeLFI 2014 – Die 12. e-Learning
Fachtagung Informatik
der Gesellschaft für Informatik e.V.
15. – 17. September 2014
Freiburg
- P-234 Fernand Feltz, Bela Mutschler, Benoît
Otjacques (Eds.)
Enterprise Modelling and Information
Systems Architectures
(EMISA 2014)
Luxembourg, September 25-26, 2014
- P-235 Robert Giegerich,
Ralf Hofestädt,
Tim W. Nattkemper (Eds.)
German Conference on
Bioinformatics 2014
September 28 – October 1
Bielefeld, Germany
- P-236 Martin Engstler, Eckhart Hanser,
Martin Mikusz, Georg Herzwurm (Hrsg.)
Projektmanagement und
Vorgehensmodelle 2014
Soziale Aspekte und Standardisierung
Gemeinsame Tagung der Fachgruppen
Projektmanagement (WI-PM) und
Vorgehensmodelle (WI-VM) im
Fachgebiet Wirtschaftsinformatik der
Gesellschaft für Informatik e.V., Stuttgart
2014
- P-237 Detlef Hühnlein, Heiko Roßnagel (Hrsg.)
Open Identity Summit 2014
4.–6. November 2014
Stuttgart, Germany
- P-238 Arno Ruckelshausen, Hans-Peter
Schwarz, Brigitte Theuvsen (Hrsg.)
Informatik in der Land-, Forst- und
Ernährungswirtschaft
Referate der 35. GIL-Jahrestagung
23. – 24. Februar 2015, Geisenheim
- P-239 Uwe Aßmann, Birgit Demuth, Thorsten
Spitta, Georg Püschel, Ronny Kaiser
(Hrsg.)
Software Engineering & Management
2015
17.-20. März 2015, Dresden
- P-240 Herbert Klenk, Hubert B. Keller, Erhard
Plödereder, Peter Dencker (Hrsg.)
Automotive – Safety & Security 2015
Sicherheit und Zuverlässigkeit für
automobile Informationstechnik
21.–22. April 2015, Stuttgart
- P-241 Thomas Seidl, Norbert Ritter,
Harald Schöning, Kai-Uwe Sattler,
Theo Härder, Steffen Friedrich,
Wolfram Wingerath (Hrsg.)
Datenbanksysteme für Business,
Technologie und Web (BTW 2015)
04. – 06. März 2015, Hamburg
- P-242 Norbert Ritter, Andreas Henrich,
Wolfgang Lehner, Andreas Thor,
Steffen Friedrich, Wolfram Wingerath
(Hrsg.)
Datenbanksysteme für Business,
Technologie und Web (BTW 2015) –
Workshopband
02. – 03. März 2015, Hamburg
- P-243 Paul Müller, Bernhard Neumair, Helmut
Reiser, Gabi Dreo Rodosek (Hrsg.)
8. DFN-Forum
Kommunikationstechnologien
06.–09. Juni 2015, Lübeck

- P-244 Alfred Zimmermann, Alexander Rossmann (Eds.)
Digital Enterprise Computing (DEC 2015)
Böblingen, Germany June 25-26, 2015
- P-245 Arslan Brömme, Christoph Busch, Christian Rathgeb, Andreas Uhl (Eds.)
BIOSIG 2015
Proceedings of the 14th International Conference of the Biometrics Special Interest Group
09.–11. September 2015
Darmstadt, Germany
- P-246 Douglas W. Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.)
INFORMATIK 2015
28.9.-2.10. 2015, Cottbus
- P-247 Hans Pongratz, Reinhard Keil (Hrsg.)
DeLFI 2015 – Die 13. E-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. (GI)
1.–4. September 2015
München
- P-248 Jens Kolb, Henrik Leopold, Jan Mendling (Eds.)
Enterprise Modelling and Information Systems Architectures
Proceedings of the 6th Int. Workshop on Enterprise Modelling and Information Systems Architectures, Innsbruck, Austria
September 3-4, 2015
- P-249 Jens Gallenbacher (Hrsg.)
Informatik
allgemeinbildend begreifen
INFOS 2015 16. GI-Fachtagung Informatik und Schule
20.–23. September 2015
- P-250 Martin Engstler, Masud Fazal-Baqaie, Eckhart Hanser, Martin Mikusz, Alexander Volland (Hrsg.)
Projektmanagement und Vorgehensmodelle 2015
Hybride Projektstrukturen erfolgreich umsetzen
Gemeinsame Tagung der Fachgruppen Projektmanagement (WI-PM) und Vorgehensmodelle (WI-VM) im Fachgebiet Wirtschaftsinformatik der Gesellschaft für Informatik e.V., Elmshorn 2015
- P-251 Detlef Hühnlein, Heiko Roßnagel, Raik Kuhlisch, Jan Ziesing (Eds.)
Open Identity Summit 2015
10.–11. November 2015
Berlin, Germany
- P-252 Jens Knoop, Uwe Zdun (Hrsg.)
Software Engineering 2016
Fachtagung des GI-Fachbereichs Softwaretechnik
23.–26. Februar 2016, Wien
- P-253 A. Ruckelshausen, A. Meyer-Aurich, T. Rath, G. Recke, B. Theuvsen (Hrsg.)
Informatik in der Land-, Forst- und Ernährungswirtschaft
Fokus: Intelligente Systeme – Stand der Technik und neue Möglichkeiten
Referate der 36. GIL-Jahrestagung
22.-23. Februar 2016, Osnabrück
- P-254 Andreas Oberweis, Ralf Reussner (Hrsg.)
Modellierung 2016
2.–4. März 2016, Karlsruhe
- P-255 Stefanie Betz, Ulrich Reimer (Hrsg.)
Modellierung 2016 Workshopband
2.–4. März 2016, Karlsruhe
- P-256 Michael Meier, Delphine Reinhardt, Steffen Wendzel (Hrsg.)
Sicherheit 2016
Sicherheit, Schutz und Zuverlässigkeit
Beiträge der 8. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)
5.–7. April 2016, Bonn
- P-257 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
9. DFN-Forum
Kommunikationstechnologien
31. Mai – 01. Juni 2016, Rostock
- P-258 Dieter Hertweck, Christian Decker (Eds.)
Digital Enterprise Computing (DEC 2016)
14.–15. Juni 2016, Böblingen
- P-259 Heinrich C. Mayr, Martin Pinzger (Hrsg.)
INFORMATIK 2016
26.–30. September 2016, Klagenfurt

The titles can be purchased at:

Köllen Druck + Verlag GmbH

Ernst-Robert-Curtius-Str. 14 · D-53117 Bonn

Fax: +49 (0)228/9898222

E-Mail: druckverlag@koellen.de

Gesellschaft für Informatik e.V. (GI)

publishes this series in order to make available to a broad public recent findings in informatics (i.e. computer science and information systems), to document conferences that are organized in co-operation with GI and to publish the annual GI Award dissertation.

Broken down into

- seminars
- proceedings
- dissertations
- thematics

current topics are dealt with from the vantage point of research and development, teaching and further training in theory and practice. The Editorial Committee uses an intensive review process in order to ensure high quality contributions.

The volumes are published in German or English.

Information: <http://www.gi.de/service/publikationen/lni/>

ISSN 1617-5468

ISBN 978-3-88579-653-4

“INFORMATIK 2016” was held from Sept. 26-29 2016 at the Alpen-Adria-Universität Klagenfurt (Austria). Under the tagline “Informatics: from humans for humans” aspects of developing and running human centered IT systems have been discussed from a variety of different perspectives. This volume contains all contributions that have been presented at the conference: key note speeches, workshop papers, papers accepted for the students conference and the doctoral symposium, as well as summaries of the selected tutorials.