



CERC2017

Collaborative European Research Conference

Karlsruhe University of Applied Sciences – Karlsruhe, Germany

22 - 23 September 2017

www.cerc-conference.eu

Proceedings

**40 Extended abstracts and
and 23 selected full papers**

Editors (in alphabetical order) :

Udo Bleimann

Bernhard Humm

Robert Loew

Stefanie Regier

Ingo Stengel

Paul Walsh

ISSN : 2220 - 4164

Table of Content

Extended Abstracts and selected Full Papers

Chapter 1

Computer Science

Social Network-Based Knowledge, Content, and Software Asset Management Supporting Collaborative and Co-Creative Innovation	16
Salman, Mertens, Vu, Fuchs M., Heutelbeck, Hemmje	
Context-Aware Documentation in the Smart Factory	26
Beez, Bock, Deuschel, Humm, Kaupp	
User Centric Web Usage Mining as a Tool for Social Sciences	38
Bettermann, Schlegel	
Cloud Interoperability: Approaches and Challenges	42
Moynihan, Abdi	
From Monolithic to Modern Software Development – An Approach to Migrate and Scale Legacy Applications Using Software Containers, Microservices and Cloud Computing	47
Krämer, Wiener, Abecker, Nimis	
Linked Open Data for Organizational Knowledge Bases – Towards a Linked Data Wiki	50
Frank, Zander	
Multi-Access Edge Computing: Managing the Resources of the Edge of the Internet	58
Frank, Fuhrmann, Ghita	
Interactive Guidance and Real-Time Fault Recognition for Manual Operations Through a Stationary AR System and Visual Object Detection	61
Lutze, Hansert, Lunt, Schlegel	

Evaluation of Machine Learning Algorithms for Smurf Detection Hippe, Dornheim, Zeitvogel, Laubenheimer	65
3D Human Surface Registration Using Subdivision Surfaces Zeitvogel, Laubenheimer	69
Cost-Effective Semi-Automatic Ontology Development from Large Domain Terminology Humm, Ossanloo	73
From Raw Data to Smart Documentation: Introducing a Semantic Fusion Process for Cyber-Physical Systems Kaupp, Beez, Humm, Hülsmann	83
A Weighted Hybrid Recommender System Approach for Product Configuration Benz, Ullmann, Wetzel, Felic, Alexakis	95
New Approach in 3D Mapping and Localization for Search and Rescue Missions Brenner, Gelfert, Rust	105
The System Design and Implementation of V-Book System Based on AR Technology Wu, Mao, Wang, Cong, Li	112
The Application of Big Data Analytics Technologies for the Predictive Maintenance of Industrial Facilities in Internet of Things (IoT) Environments Rieger, Regier, Stengel, Clarke	118
Reinforcement Learning Based Approach for Optimal Deep Drawing Process Control Dornheim, Link	129

Chapter 2

Bioinformatics / Health informatics

Quantified Self in the Healthcare: State of the Art and Impact on the Health Insurance Widmann, Atkinson, Stengel, Clarke, Regier	134
---	-----

Comparative Analysis of Mycobacterium Tuberculosis Genome	138
Manning, Palu, Lawlor, MacAogáin, Kelly, Walsh P.	
A Dynamic Product Line for an Electronic Health Record Management System for Cancer	142
Coym, Humm, Spitzer, Walsh P.	
A Process Mining Framework to Simulate and Predict Patient Pathways and Health Related Costs Based on Medical Claim Data	150
Sahin, Accorsi, Frey-Luxemburger, Knahl, Atkinson	
Evidence-Based Medical Recommendations for Personalized Medicine	155
Humm, Lamba, Landmann, Steffens, Walsh P.	
Preforming Real-Time Emotion Classification Using an Intel RealSense Camera, Multiple Facial Expression Databases and a Support Vector Machine	164
Healy, Keary, Walsh P.	
Predicting Premature Termination of Treatment in Psychotherapy for Borderline Personality Disorder	176
Bohus, Deuschel, Gimbel, Görg, Humm, Schüller, Turan	
Utilizing High Performance Computing Techniques for Efficiently Anonymizing Sensitive Patient Data	186
Ntalaperas, Bouras, Di Fuccio	
Challenges and Limitations of Large-Scale Molecular Biology Analysis and Visualization Tools from the User Experience Perspective	190
Palu, Lawlor, Kelly, Walsh P.	
An Interactive Interface for Statistical Modelling of Gene Expression Data	197
Briand, Palu, Kelly, Walsh P.	
Case Study: Management of Affective Data and Data Fusion in eHealth Appliances under Consideration of Legal Frameworks	200
Kowohl, Engel, Dassler, Hemmje	

Chapter 3

Security

Elicitation of Security Requirements for Migration of OCR Software to the Cloud	210
Ringmann, Langweg	
Understanding Opportunities and Risks in Online Gaming Environments: An Ethnographic Approach	213
Sanders	
IT Security Policy Enforcement in IoT Gateways	218
Koca	
Insecurity Refactoring as a Novel Method to Improve Manual Code Inspection Skills	222
Schuckert	
Expert System for Risk Analysis of Technical Systems	228
Single, Schmidt, Denecke	
On Measuring the Behaviour of Users in the Area of Security-Sensitive Platforms	232
Schalles	
Gamifying Cyber Security Awareness via Mobile Training Apps	236
Alotaibi, Furnell, Stengel, Papadaki	

Chapter 4

Business

Electricity Cost Reduction Potential of Industrial Processes Using Real Time Pricing in a Production Planning Problem	240
Fleschutz, Bohlayer, Bürger, Braun	
Multi-Objective Optimization of a Waste Heat Driven Heating Supply System Using Mixed-Integer-Programming	249
Bohlayer, Fleschutz, Bürger, Braun	

An Exploration of Experiential Marketing in the Fast Moving Consumer Goods Industry	253
Barry, Leahy, Fenton	
Judging Competencies, Abilities and Expertise Based on Enterprise Social Network Profile Pages	256
Schemmann, Behrendt, Kinkel	
Practice-Oriented Training as a Sales Engineer: Case Study Analysis of a Role Play	260
Murzin, Ben Amor	
Measuring and Visualizing the Value Added through the Industrial Internet of Things	263
Gebhard, Ersari, Wirth	
The Effect of Cognitive and Individual-Level Factors in Wearable Computing Adoption – Empirical Evidence from Germany	268
Gribel, Regier, Stengel, Furnell	
A Review of Challenges in Blockchain Adoption for Supply Chains	276
Buchner, Stengel	

Chapter 5

Learning

Thoughts on Presenting Learning Recommendations in the Context of the Project SensoMot	284
Schneider	
Mobile Learning in Nature: A Digital Field Guide to Plants	287
Henning P., Henning J., Manegold, Fuchs K.	
Machine Learning for User Learning: Supporting Users in Different Stages of Expertise	290
Schuhmacher, Kaiser	
Supporting 21st Century Learning through Digitally-Enabled Collaborative Learning Spaces	294
Holton	

Operationalizing Learning Behavior as Spatio-Temporal Trajectories	297
Fuchs K., Henning P.	

Chapter 6

Media

Acoustic Reading Experience: Aligning Sound Events to Text Using Gaze Tracking to Improve Immersion in Reading	308
Hill, Wölfel	
Sensory Modalities Influence on Musicians' Latency Measurements	311
Victoria	
Virtual Reality – Quo vadis? How to Address the Complete Audience of an Emerging Technology	315
Sieß, Beuck, Wölfel	
Fostering Human-to-Human Communication and Collaboration in Physical Space by Playful Augmentation and Information Inequality	325
Conzelmann, Lange, Wölfel	

Chapter 7

Interdisciplinary

Experimental Raman Spectra of Hydrocarbons up to Flame Relevant Temperatures	330
Dieter, Geyer	
Bioactives from Saccharina Latissima and Gut Health	335
O'Connell, O'Shea, McGlynn	
Protection of Chemical Reactors with Partial-Least-Square-modeling (PLS) for Safety Devices	340
Biernath, Goerlich, Denecke, Schmidt	

Chapter 8

Traffic and Transportation

A vehicular Trajectory-Driven Method for Signalized Corridor Control under Connected Vehicles Environment	346
Gutesa, Lee, Besenski, Spasovic	
Towards Automated Driving: Development of an Unmanned Driving Vehicle for Road Maintenance	350
Reuber, Holldorb, Wachsmann	
Development of a Concept of Need-Based De-Icing Salt Storage Capacities and the Implementation in a Dimensioning Guideline	354
Wachsmann, Holldorb, Schappacher	
Assessing the Value of GPS Travel Data for Travel Forecasting Models	358
Casello, Xu, Fard, Mikkila	
Dealing with the Rise of Transportation Network Vehicle Service: Rethinking the Regulatory Regime for TNVS towards Sustainable Urban Mobility	362
Gamboa	
Innovation Coaching in Foresight Processes for Distributed Product Development	366
Hahn, Albers, Stöckner, Niever, Walter, Kerres, Bursac	
Real-World Laboratory: Promoting Pedestrians	376
Haeussler	

Chapter 9

Civil Engineering

The Successful Integration of Health and Safety in the Engineering Curriculum of the Graduate Engineer	380
Moloney, Walsh D.	

List of Authors

Abdi	42	Abecker	47
Accorsi	150	Albers	366
Alexakis	95	Alotaibi	236
Atkinson	134, 150	Barry	253
Beez	26, 83	Behrendt	256
Ben Amor	260	Benz	95
Besenski	346	Bettermann	38
Beuck	315	Biernath	340
Bock	26	Bohlayer	240, 249
Bohus	176	Bouras	186
Braun	240, 249	Brenner	105
Briand	197	Buchner	276
Bursac	366	Bürger	240, 249
Casello	358	Clarke	118, 134
Cong	112	Conzelmann	325
Coym	142	Dassler	200
Denecke	228, 340	Deuschel	26, 176
Di Fuccio	186	Dieter	330
Dornheim	65, 129	Engel	200
Ersari	263	Fard	358
Felic	95	Fenton	253
Fleschutz	240, 249	Frank	50, 58
Frey-Luxemburger	150	Fuchs K.	287, 297
Fuchs M.	16	Fuhrmann	58
Furnell	236, 268	Gamboa	362
Gebhard	263	Gelfert	105
Geyer	330	Ghita	58
Gimbel	176	Goerlich	340
Gribel	268	Gutesa	346
Görg	176	Haeussler	376
Hahn	366	Hansert	61
Healy	164	Hemmje	16, 200
Henning J.	287	Henning P.	287, 297
Heutelbeck	16	Hill	308
Hippe	65	Holldorb	350, 354
Holton	294	Humm	26, 73, 83, 142, 155, 176
Hülsmann	83	Kaiser	290
Kaupp	26, 83	Keary	164
Kelly	138, 190, 197	Kerres	366
Kinkel	256	Knahl	150

Koca	218
Krämer	47
Landmann	155
Langweg	210
Lawlor	138, 190
Lee	346
Link	129
Lutze	61
Manegold	287
Mao	112
Mertens	16
Moloney	380
Murzin	260
Nimis	47
Ossanloo	73
O'Shea	335
Papadaki	236
Reuber	350
Ringmann	210
Sahin	150
Sanders	213
Schappacher	354
Schlegel	38, 61
Schneider	284
Schuhmacher	290
Sieß	315
Spasovic	346
Steffens	155
Stöckner	366
Ullmann	95
Vu	16
Walsh D.	380
Walter	366
Wetzel	95
Wiener	47
Wu	112
Xu	358
Zeitvogel	65, 69

Kowohl	200
Lamba	155
Lange	325
Laubenheimer	65, 69
Leahy	253
Li	112
Lunt	61
MacAogáin	138
Manning	138
McGlynn	335
Mikkila	358
Moynihan	42
Niever	366
Ntalaperas	186
O'Connell	335
Palu	138, 190, 197
Regier	118, 134, 268
Rieger	118
Rust	105
Salman	16
Schalles	232
Schemmann	256
Schmidt	228, 340
Schuckert	222
Schüller	176
Single	228
Spitzer	142
Stengel	118, 134, 236, 268, 276
Turan	176
Victoria	311
Wachsmann	350, 354
Walsh P.	138, 142, 155, 164, 190, 197
Wang	112
Widmann	134
Wirth	263
Wölfel	308, 315, 325
Zander	50

Preface

Research in our days is moving more and more from individual to collaborative. In this way different research partners from different countries are linked together, research is speeded up and made more sustainable. In the age of fast research and developments visions together with solid expertise lead to ground breaking results in interdisciplinary areas like e.g. Bioinformatics, Self-driving cars, etc.

CERC is aligning with this tendency, bringing together experienced researchers from science, engineering, business, humanities and arts, thus enabling young and experienced researchers to present and discuss the results of their research.

The organisation of an interdisciplinary and collaborative conference requires contributions of a number of persons. We would like to thank researchers for submitting their papers to CERC and many reviewers that helped to guaranty the quality of papers. The conference would not have been possible without the significant contributions of the local organising committee and without the support and the facilities provided by the University of Applied Science Karlsruhe.

Last but not least, I would like to thank my fellow program chairs for the regular discussions that helped to solve problems and to keep the conference on track.

Prof. Dr. Ingo Stengel

General Conference & Programme Co-Chair, CERC2017
Karlsruhe, September 2017

Chapter 1

Computer Science

Social Network-Based Knowledge, Content, and Software Asset Management Supporting Collaborative and Co-Creative Innovation

Munir Salman^{1, 4}, Jana Mertens², Binh Vu², Michael Fuchs³, Dominic Heutelbeck², Matthias Hemmje¹

¹FernUniversität in Hagen, Faculty for Multimedia and Computer Science, Hagen, Germany

²Forschungsinstitut für Telekommunikation und Kooperation e.V. (FTK), Dortmund, Germany

³Wilhelm Büchner University of Applied Science, Pfungstadt, Germany

⁴Munich Center for the Economics of Aging (MEA) at the Max Planck Institute for Social Law and Social Policy
e-mail: Munir.Salman@fernuni-hagen.de

Keywords: Social Networking Platforms, Knowledge Management and Transfer, Applied Gaming, Digital Ecosystem, Collaborative and Co-creative innovation processes

Abstract

This paper addresses a number of scenarios for the collaborative and co-creative aggregation of innovation knowledge, content, and corresponding software assets from social network-based software repositories supporting, e.g., Source Code Management (SCM), such as e.g., GitHub (*Build software better, together*, 2017). In the resulting collaborative, co-creative, and social-network enabled innovation approach, such innovation knowledge resources, supplementary content, and corresponding software assets are aggregated into a service-based, social network-enabled, knowledge-driven ecosystem portal supporting innovation for an exemplary use case of applied games communities (i.e. AG researchers, software developers, and players). In addition, we propose methods for the harvesting of relevant knowledge contained in know-how oriented co-creative user-interactions surrounding these assets such as Stack Overflow (*Stack Overflow*, 2017), as well as the long-term archival of these knowledge resources to capture implicit innovation and co-creation knowledge associated with the software-assets. The paper will explain the implementation of the software harvesting as well as innovation and co-creation knowledge acquisition scenario in detail.

1 Introduction and motivation

The market for non-leisure games (Applied Games, AGs) is part of an exemplar emerging innovative, technology- and knowhow-driven ecosystem. However, as such it is still fragmented and needs to achieve critical mass. Nevertheless, its growth potential is widely recognized and even suggested to exceed the growth potential of, e.g., the leisure games market. Therefore, to become more competitive the relevant applied gaming independent software developers and small and medium sized enterprises as well as their customer and user communities require innovation support.

The launch of innovative products and services by Small and Medium-sized Enterprises (SMEs) of the AG industry constitutes an enormous challenge considering fierce global competition combined on the one hand and limited budgets on the other hand. Therefore, such SMEs need external, i.e., collaborative and co-creative innovation support, e.g., from research groups and independent innovative software developers, enabling them to collaboratively and co-creatively achieve a crucial competitive advantage of being faster than others (Paukert et al., 2011). Accelerating the discovery of new (scientific and technical, S&T) findings, their take-up into innovative software developments including the technical realization of the innovation as well as the market launch (Grupp, 1997; Haß, 1983) has to be fostered. The success of this process is increasingly dependent on the use of advanced information and knowledge technologies for building web-based and service-oriented innovation ecosystem support environments that support the collaborative and co-creative knowledge sharing

and corresponding collaborative and co-creative innovation processes systematically and efficiently (Specht et al., 2002). Such environments depend on the availability of a number of advanced Information, Content, and Knowledge Management (ICKM) technologies and processes and have to adapt to a wide variety of innovative practices, cultures, organizational context, business models, and application areas (Becker et al., 2015). Independent of the domain, innovation is a knowledge-intensive process (Paukert et al., 2011). Therefore, the RAGE project (netvision, 2017) is an exemplary reference use case aiming at supporting this challenge. RAGE will help to seize these opportunities by making available an interoperable set of innovative Applied Game (AG) technology assets, as well as knowledge and supplementary content about proven innovative practices of co-creatively using software-asset based AG innovations in various real-world contexts. As described in (Salman et al., 2015), RAGE will finally provide a centralized access to a wide range of innovative applied gaming software-assets, relevant innovation and co-creation information, knowledge, learning content, and related social-network enabled community services, as well as related scientific and technical documents and media (e.g., publications, presentations) including knowledge resources like, e.g., taxonomies, educational resources, quality assurance dialogs etc. within an online innovation co-creation community portal called the RAGE Knowledge Management Ecosystem Portal (RAGE KM-EP).

2 Problem statement, goals, and objectives

The main objectives of the RAGE KM-EP are to allow its participants to get hold of advanced, innovative but at the same time co-creatively usable best practice knowledge, supplementary content resources, and technology assets (supporting technology push), to enable them effectively (Manlio Del Giudice et al., 2014) and efficiently to get access to the associated innovative business cases and commercial opportunities. Therefore, the main driver of the RAGE KM-EP is to connect, equip, and enable SMEs and individual industry players (e.g., independent innovative game developers) with a set of advanced technology resources (so-called assets) and strategies (i.e., know-how) from academic Research and Technical Development (RTD) communities to strengthen their capacities to penetrate an emerging market (non-leisure) and to develop an innovative, i.e., unique sales position. The initial concepts, ideas, and awareness supporting the ideation phase of the innovation process come through the content aggregation of Scientific and Technical (S&T) publications, presentations and other supplementary Information-, Content- and Knowledge- (ICK) resources as described in (Salman et al., 2016) along, e.g., the integration with the Mendeley (Mendeley, 2017) and SlideShare (SlideShare.net, 2017) platforms.

In consequence, the integration with the GitHub software repository and its Source Code Management (SCM) facilitates the integration and aggregation of software-assets from social-network based systems into the development phase of the innovation and co-creation processes. In this way, AG researchers as well as independent innovative software developers are seamlessly enabled to import innovative AG software-assets into the RAGE KM-EP. The imported assets will be, on the one hand, semantically annotated to support searching, browsing, and access and on the other hand, will be linked to other objects (i.e., taxonomies, supplementary documents, media, and educational resources, as well as best practice knowhow as, e.g., contained in Question and Answer (Q&A) dialogs etc.) within the RAGE KM-EP. In this way, the RAGE KM-EP will realize centralized access to a wide range of AG innovation ICK resources referring to innovative AG software-assets, services, and resources by the arrangement of a well-managed and structured asset repository, digital library, and media archive system. The Stack Overflow system, e.g., supports AG researchers and innovative independent software developers to post questions to the AG Stack Overflow community and to co-creatively get answers, comments, notifications, etc. without switching from the RAGE KM-EP to Stack Overflow and vice versa. Moreover, relevant AG best practice knowledge can be imported into the KM-EP as, e.g., a Q&A dialog which on the other hand, can be semantically

annotated and analyzed by means of applying technologies for, e.g., Natural Language Analysis (NLA) for supporting discourse analysis. That will support to arrange, e.g., co-creation workshops and offer training courses on an online training portal, covering training for developers and educators, self-sustainable co-creative production of innovation-oriented assets and documentation, training material, workshops, and collaboration activities.

In the remainder of this paper, section 3 describes the relevant State of the Art and Related Work with respect to relevant portal technologies supporting KM within innovation and co-creation processes considering the model of knowledge management success as described in (Jennex and Olfman, 2010; Jennex et al., 2016). Section 4 provides an introduction to the use case scenarios of the GitHub and Stack Overflow KM-EP integration based on knowledge- and service-driven perspectives for the digital innovation and co-creation ecosystem environment. Furthermore, section 5, more specifically and technically reviews the integration possibilities of GitHub (Build software better, together, 2017) and Stack Overflow (Stack Overflow, 2017) into the RAGE KM-EP using their Application Programming Interfaces (APIs). Besides, this section will describe the implementation of the RAGE KM-EP GitHub and Stack Overflow integration use cases that are described in section 4. Finally, the paper will present an initial cognitive walkthrough evaluation as well as conclusions and future works.

3 State of the art and related work

The work presented in this paper is related to a number of topics in RTD. The conceptual approach of the KM-EP is based on the reflections of KM (Bullinger et al., 1997) and Service Management (Fließ, 2008). Building, amongst others, on the SEKI model of Nonaka und Takeuchi (Nonaka and Takeuchi, 1995), the Innovation Knowledge Lifecycle Meta Model (IKLC) extends the basic model with respect to supporting innovation (Paukert et al., 2011). Defined by people, especially practitioners, in a shared domain engaging in a process of collective learning, innovation, and co-creation (Wenger, 2011), the KM-EP is aiming at creating and supporting the digital environment for a virtual Communities of Practice (VCoP) in the innovation domain AG to enable successful asset-based innovation ideation, conceptualization, proof-of-concept, and finally co-creative innovation, i.e., application solution development and will later also support corresponding commercialization.

A well-known similar example of related work is GitHub (Build software better, together, 2017). GitHub is a software repository and SCM portal for developers, which currently has over 10 million users. Users can upload or start a software project and collaborate co-creatively on its development with other users in the community (Dabbish et al., 2012). Currently active and past projects are retained and can be used as examples and sources of knowledge. GitHub provides an API (GitHub API v3 Developer Guide, 2017) based on the REST service-oriented architecture (Fielding, 2000) which can be used for integration with the RAGE KM-EP. The current version (v3) of the GitHub RESTful API supports full read- and write-access on the GitHub resources (i.e., repositories, organizations, users, issues, statistics, activities, commits and tags).

Stack Exchange (Hot Questions - Stack Exchange, 2017) is another example of a collaborative innovation and co-creation support environment. Stack Exchange is a popular Q&A online community platform. Users can post questions on a topic and also associate the question with a maximum number of five tags, other users provide answers. Besides, participants are not compensated, e.g., financially for their services, but users do gain reputation points when other users “up-vote” their questions and/or answers. This approach as a co-creatively learning community with a strong reputation-seeking element creates valuable public good in the sense of a knowledge resource. Therefore, (Posnett et al., 2012) named Stack Exchange as a gift economy. One of the most important communities of Stack Exchange is Stack Overflow, the preeminent site for programmers to find, ask, and answer questions about software development. Similarly, Stack Exchange provides a RESTful API which can be used to post and pull questions and answers, as well as to get access to

the resources of Stack Exchange (i.e., questions, answers, comments, badges, events, user information and tags) from other systems and platforms.

The KM-EP was built based on Typo3 (Binh Vu, 2015a) and, therefore, can be extended with the help of Typo3 extensions. Our work extends the RAGE KM-EP modules Content & Knowledge Management and User & Community Management.

4 Modelling the KM-EP platform and its use case scenarios

The concept of a supporting collaborative and co-creative innovation processes by creating an adequate innovation and co-creation ecosystem support environment considering incoming innovation opportunities as well as innovation needs and demands, harvesting and working with outputs of RTD, i.e., innovation knowledge creation processes and the satisfaction of innovation demands based on innovative RTD outcomes has been extended to the concept of the KM-EP. Additionally, our BMC of the collaborative and co-creative KM-EP environment is based on the business model framework as described by Osterwalder et al. in (Osterwalder and Pigneur, 2011) using nine basic building blocks, covering four areas of business: Customers, offer, infrastructure and financial aspects. The framework has been used to analyze the Leisure and Applied Games industry business models.

Whilst the BMC is a tool by which businesses can effectively articulate their value proposition, it does have its limitations critically in the very linear way it presents transactions and relationships. The RAGE KM-EP RTD requires a more sophisticated and richer approach provided by Platform Design (Choudary, 2015). The Platform Design provides us with attributes of modern product driven services, and these attributes should remain at the core of the products and services provided through the RAGE KM-EP, namely fast, personalized, relevant and human. Besides, as the AG market and its innovation and co-creation ecosystem is fragmented the Platform Design approach provides RAGE with the opportunity to “bring order to a disordered market” and to facilitate the development of “business models that allow multiple sides (producers and consumers) to interact by providing an infrastructure that connects them” (Choudary, 2015). Platform Design thinking incorporates the elements of service design.

To ensure the goal of co-creatively stimulating both the demand and supply side for applied games, a fundamental disruptive (Christensen et al., 2010) action is required, with the aim of achieving a state of systemic innovation within the sector. In this way, the asset-based approach of the RAGE KM-EP is transformative: it requires the development of new, i.e., innovative and co-creative business models and takes the sector beyond the “products and customer service” narrative in to the collaborative and co-creative platform design narrative.

Taken the SEKI and IKLC considerations into account (Becker et al., 2015), the KM process underlying the KM-EP can be described as follows. Driven by the need for competitive products and services, and at the same time acting responsive, i.e., co-creative to the customer’s fast changing requirements the AG industry indicates a high demand for collaborative and co-creative innovation. To establish an appropriate innovation and co-creation support environment and at the same time a kind of a body of state-of-the-art and best-practice knowledge to support these the KM-EP supports collecting conceptual AG RTD knowledge, ideas, methods, and corresponding document and media resources (e.g., scientific publications and presentations), software-assets, documentation, Q&A dialogs, training material, and best practices from the heterogeneous and dispersed AG landscape to provide relevant communities the opportunity to co-creatively participate, to share and to benefit from these RTD resources to create new and innovative outcomes. In this way the content aggregation of, e.g., S&T publications and presentations as described in (Salman et al., 2016) along the integration with the Mendeley (Mendeley Homepage, 2017) and SlideShare (SlideShare.net, 2017) platforms has already been implemented. The following subsections describe and model the integration with the GitHub and Stack Overflow based innovation and co-creation support features using the Unified

Modeling Language (UML) (Welcome To UML Web Site!, 2017). The detailed modeling of the RAGE KM-EP BMC is an ongoing agile and co-creative innovation process itself and can (because of limitations in space) not be presented at this time in this paper.

5 Service-driven approach and the GitHub use case scenarios

Taken the modelled features of the RAGE KM-EP into account, the KM-EP is integrating streams of innovation and co-creation knowledge, content and asset resources of the innovation “supply side”, i.e. from RTD actors and is making it collaboratively available to the “demand side” i.e. innovative software developers and SMEs in the AG innovation ecosystem.

Thus, the offered features received a service character. In this case S&T innovation and co-creation knowledge represents an intangible commodity and the result of the service process has an intangible character as well. The process could be highly integrative, if the customer, frequently users/consumers themselves, will later be integrated constantly into a co-creative phase of an innovation preparation process (Corsten and Gössinger, 2007). An advantage of the system is the automated execution of the described service processes. In consequence, the integration with the GitHub software repository permits RAGE KM-EP users to import software-assets including their metadata into the RAGE KM-EP. Furthermore, software-assets can be annotated with semantic representations and classified with taxonomies. In this way, AG software-assets (e.g. software-frameworks) can be found as well as effectively be accessed and re-used within the RAGE KM-EP. AG Asset-developers can, e.g., import their own repositories into the RAGE KM-EP while administrators are allowed to import additional other repositories using a Uniform Resource Locator (URL) of repositories. RAGE KM-EP users can then look for AG assets using the full-text as well as the semantic search engine which is integrated into the RAGE KM-EP or by means of faceted browsing through the taxonomies within the RAGE KM-EP. Besides, RAGE users can also download the entire repository as a ZIP-file that can be archived using an archiving or even long term archiving system based on OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting, 2017). Another use case of the integration with GitHub which is considered to be implemented in the future is to keep the imported repository along with its associated issues, comments, changes, etc., synchronized with the original one from GitHub. This helps to support the user collaboration, workflows, and co-creative communication. Likewise, the Travis CI (Travis CI - Test and Deploy Your Code with Confidence, 2017) platform uses synchronization technology based on WebHook (Web Hooks / FrontPage, 2017). Hence, RAGE KM-EP users stay well-informed about the issues, changes, and events from the community uninterruptedly. The described use case scenarios show that expert knowledge, information, and expertise from outside the user community, as well as community collaboration support were taken into account.

6 Knowledge-driven approach and the Stack Overflow use case scenarios

To handle the available domain-specific tacit knowledge within a technology based infrastructure it must be transformed into machine-readable representation. Therefore, the applied gaming users or user groups have to be supported to transform their co-creative communication, information, Q&A dialogs, and experiences into digital objects (e.g., text, dialog, image, video, recorded speech, etc.) and provide them to the related AG RTD communities. The integration with the Stack Overflow as a collaborative co-creation support feature surrounding the AG-assets permit RAGE KM-EP users to post questions to the Stack Overflow community without switching from the RAGE KM-EP to Stack Overflow and to get answers about the posted questions directly back into the RAGE KM-EP. Thus, Q&A dialogs are transformed into digital co-creation support objects which can be archived, shared, accessed, disseminated, and re-used (again using an archiving system based on OAI-PMH (Open

Archives Initiative Protocol for Metadata Harvesting, 2017)), annotated with semantic representations, and classified with taxonomies.

After a successful authentication and authorization RAGE KM-EP users can post new questions or edit a given one (i.e. changes the question, change its associated tags, add a comment, etc.).

Furthermore, users can browse the answers within the RAGE KM-EP for a posted question. Another use case similar to the synchronization of the GitHub repository is to track the questions, answers, and their related changes such as issues and comments in order to support users' co-creative collaboration, interaction, and communication. Besides, the KM-EP will use the collective data of Stack Overflow (i.e., questions, answers, associated tags, and related user data) in order to fulfill the automated expert recommendation use case. Thus, the KM-EP supports its users to find content that arose from harvesting tacit knowledge and curating it with explicit SA representations and to adapt, extend, and link/enrich it in this way semantically with additional supplementary background knowledge to support later re-use and especially internalization of this knowledge (i.e., Internalization as from the SEKI point of view). This step is called cognitive value creation or the generation of knowledge (Lave and Wenger, 1991).

Connecting to this process the KM-EP offers Internalization (as understood in SEKI) resp. knowledge transfer support in the sense of a learning management process. It serves as a huge knowledge library or database. E.g., Stack Overflow assigns reputations to user profiles to measure how much the community trust the user (i.e., measurement of the users' competence). The more reputation users earn, the more privileges they can gain.

7 Proof of concept prototype to support co-creative software asset integration

The following section presents the main technical integration possibilities of the GitHub and the Stack Overflow in the backend and frontend of the RAGE KM-EP, as well as the implementation

for the use cases that were described in section 4. Figure 1 displays our software architecture extension of the two components Ingest and Content Manager within the RAGE KM-EP.

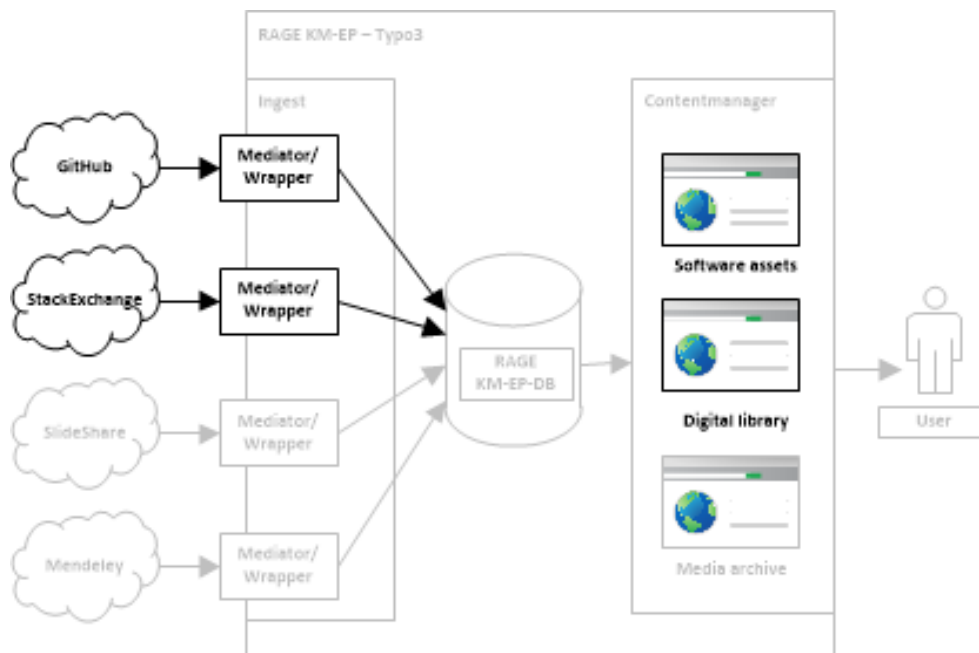


Figure 1. Ingest and content manager: extending the KM-EP software architecture

The data integration is based on the Mediator/Wrapper architecture as described in (Wiederhold, 1992) using the RESTful APIs of the GitHub and Stack Exchange (GitHub API v3 Developer Guide, 2017) (“Stack Exchange API,” 2017). The RAGE KM-EP extensions have to be registered on the GitHub and Stack Exchange developer portal as an application and a valid “client ID and secret key” has to be available in order to be allowed to use their APIs. Besides, security issues are taken into account to get access to the resources and software-assets using the OpenID (OpenID Foundation website, 2017), the OAuth Protocol (“OAuth 2.0 — OAuth,” 2017), and HTTPs for authentication and authorization. We use the MythCRM-System (Binh Vu, 2015b) to manage applications’ and users’ access based on role management. Both extensions GitHub as well as Stack Exchange are registered as an OpenID-Client in the MythCRM-System. Furthermore, users have to be registered too, and associated to roles in order to get access to the resources and to be allowed to perform functions (e.g., import private or public repositories from GitHub, post questions and import Q&A dialogs from Stack Exchange, etc.) of the GitHub-, and Stack Exchange-extensions. Furthermore, the integration of the user interface is based on the Model View Controller (MVC) paradigm as described in Extbase for Typo3. We use the owner name of the repository and the repository name to identify the repository that would be read. All metadata of the repository including the RAGE-Asset Metadata (if available) are displayed in separated tabs and categories in order to enhance the usability and the user-friendliness of the implemented interface (cf. section 6). After that, the user can either select repositories or search for a public repository using a repository’s URL and import them. Once a repository was imported, the metadata-array will be automatically mapped to the attributes of the RAGE KM-EP software-object. Moreover, user can also adjust the mapping-table manually. On the other hand, the Stack Exchange API returns every response in a common “wrapper” object using JSON or JSONP with the callback query parameter. The API support three object types i.e., question-, answer-, and user-object. Each object has an ID which we use to identify contents of a posted object. Besides, each question-object is assigned to a user. The answers’ references of the question are saved in the field “answers” as an array of answer-objects. We use the associated tags to identify questions and answers that are interesting or relevant to the AG developers. A tag is a word or phrase that describes the topic of the question or the answer. The answer-, and the user-object have similar structure to the question-object which makes the integration and the mapping of the data to the RAGE KM-EP attributes easy.

8 Preliminary evaluation

In order to drive decision making and further RTD of our prototype and designs we conducted a cognitive walkthrough (Polson et al., 1992) with AG RTD experts from the user communities of the RAGE project for each use case described in session 4. This preliminary, i.e., initial qualitative evaluation is necessary in order to ensure that functions adequately meet the requirements and needs of the targeted AG user communities. This section presents a preliminary evaluation based on cognitive walkthrough for the integration use case with GitHub. **Fehler! Verweisquelle konnte nicht gefunden werden.** displays the user interface after successfully connection to the GitHub repository of the logged in user. The user can then import his own repositories (Figure 5, tab “your repositories”) or search for a public repository using an URL (**Fehler! Verweisquelle konnte nicht gefunden werden.**, tab “URL search”). The assets of the selected repositories are displayed in the tab “Assets to import” (**Fehler! Verweisquelle konnte nicht gefunden werden.**). There the user can select assets to be imported. **Fehler! Verweisquelle konnte nicht gefunden werden.** shows the data mapping interface of the imported asset. User can accept the proposed data mapping or edit it.

9 Conclusion and outlook

In summary, it is a big advantage to aim at supporting the integration of software repositories and SCM such as GitHub, as well as collaboration- and co-creation Q&A support environments such as Stack Overflow including content, information, and knowledge capturing, management, sharing, and dissemination support through their RESTful API into the KM-EP. This will on the one hand facilitate to provide a wide range of supporting services in the field of collaborative innovation and co-creation knowledge transfer and -creation to overcome low market access and small market share of SMEs and independent innovative software developers in the innovation domain as, e.g., in our reference use case the AG market, to create new effective technology based assets in order to build new innovative products and services, e.g., learning-oriented games. On the other hand, it focuses on identifying collaboration and co-creation opportunities between individuals and among groups, to support matchmaking, collaboration, ideation, conceptualization, proof-of-concept, as well co-creation and collaborative innovation of application solutions between stakeholders, and to identify and provide support for innovation opportunities and creativity efforts. That allows communities (such as technology providers, game developers and educators, game industries and researchers) to engage themselves in a co-creative VCoP, create their own assets and post them to the ecosystem's repository without major effort and to benefit from achieving (business) results.

The innovation potential of the KM-EP based collaborative and co-creative ICKM and innovation process support is based on the following innovation and co-creation ICK resource harvesting, management, and access features: a huge, mostly entire collection of community specific innovation and co-creation knowledge (e.g., content like media objects, software-assets, Q&A dialogs, and best practices), a structured approach of innovation and co-creation knowledge access, search, browse, collaboration and co-creation tools, as well as social-network enabled NLA discourse analysis tools to foster efficient innovation and co-creation knowledge creation and transformation processes into marketable technology assets. Consequently, these new assets will be used to produce new applied games or game-based learning application solutions. Hence, the economic value of the developed software assets including their innovation and co-creation ICK resources will be economized and could be monetarily evaluated at the market. Therefore, innovation and co-creation ICK resources become a supplementary economic asset and the process of innovation and co-creation ICK resource creation, harvesting, archival, and reuse becomes a value adding service. With the design and development of a comprehensive approach as pursued with the KM-EP, ethical issues are taken into account. The integration of users' profiles from different resources, as well as the use of

Import from GitHub

Your repositories [URL search](#) [Assets to import](#)

Connected to your GitHub account [rageassetdeveloper](#) (disconnect). Select the repositories you want to import.

10 records per page Search:

<input type="checkbox"/>	Name	Description	RAGE metadata file exists
<input type="checkbox"/>	rageassetdeveloper/AnotherRageAsset	This is another RAGE asset. It does not contain a RAGE metadata XML file.	No
<input type="checkbox"/>	rageassetdeveloper/MyRageAsset	This is an Applied Game asset.	Yes

Showing 1 to 2 of 2 entries

[Import Selections](#) [Previous](#) [1](#) [Next](#)

Figure 2. KM-EP User Interface after successfully connecting to the GitHub repository

Listing for Import Wrappers

[BibTex](#) [OAI-PMH](#) [Mendeley](#) [SlideShare](#) [GitHub](#)

[Tools +](#)

Source	Destination		
full_name	title	Edit	Delete
html_url	repositoryUrl	Edit	Delete
description	description	Edit	Delete
updated_at	swDate	Edit	Delete
rage_keyword	keywords	Edit	Delete

Figure 3. Data mapping interface of the software assets

features carrying out analyses on top of KM-EP user data have ethical implications in terms of privacy and data protection and require appropriate information and consent in the terms and conditions of use, as well as compliance to national and international data protection regulations. The consideration of such ethical and legal requirements is incorporated in the system design and development process in terms of an ethics-by-design approach. This means that data protection and privacy is already taken into account when the system is being designed. Design principles, such as purpose binding, ensure that personal information is only accessible, if there is a need for it when performing a certain action. The system can also control data access by respecting personal settings which data should be available to others or the public. Other ethics-enabled features include the modification or deletion of personal data. Therefore, the KM-EP supports the interconnectedness, the knowledge exchange and the harmonization of standards of the applied gaming sector. Coupled with suitable business models the KM-EP is aiming at helping the AG industry to collaboratively and co-creatively drive strategy, support problem solving, build capabilities and knowledge competencies, cross fertilize ideas and increase opportunities for co-creative innovation to assert themselves against big games companies.

Acknowledgements and disclaimer



This publication has been produced in the context of the RAGE project. The project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 644187. However, this paper reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

References

- Becker, J., Lankveld, G. van, Steiner, C., Hemmje, M. (2015). "Realizing an Applied Gaming Ecosystem: Towards Supporting Service-based Innovation Knowledge Management and Transfer", in: *Proceedings of the 4th International Conference on Games and Learning Alliance, GALA 2015, Rome, Italy*.
- Binh Vu, D. (2015a). *Realizing an Applied Gaming Ecosystem - Extending an Education Portal Suite towards an Ecosystem Portal*, Master Thesis. Technische Universität Darmstadt, Germany.
- Binh Vu, D. (2015b). *Customer Relationship Management based on Identity Management for Scientific Associations*, Bachelor Thesis. Technische Universität Darmstadt, Germany.
- GitHub (2017). "Build software better, together". <https://github.com> (accessed 5.19.17).
- Bullinger, H.-J., Wörner, K., Prieto, J. (1997). *Wissensmanagement heute*. Fraunhofer Institut für Arbeitswirtschaft und Organisation, Stuttgart.
- Choudary, S.P. (2015). *Platform Scale: How an emerging business model helps startups build large empires with minimum investment*, 1st edition, Platform Thinking Labs.
- Christensen, C., Johnson, C.W., Horn, M.B. (2010). *Disrupting Class: How Disruptive Innovation Will Change the Way the World Learns*, Updated, Expanded ed., McGraw-Hill Professional, New York.
- Corsten, H., Gössinger, R. (2007). *Dienstleistungsmanagement*, 5th edition. ed. Oldenbourg Wissenschaftsverlag.
- Dabbish, L., Stuart, C., Tsay, J., Herbsleb, J. (2012). "Social Coding in GitHub: Transparency and Collaboration in an Open Software Repository", in: *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, CSCW '12*. ACM, New York, NY, USA, pp. 1277–1286.
- Fielding, R.T. (2000). *Architectural styles and the design of network-based software architectures*. University of California, Irvine.
- Fließ, S. (2008). *Dienstleistungsmanagement: Kundenintegration gestalten und steuern*. Gabler Verlag.
- GitHub API v3 Developer Guide (2017). <https://developer.github.com/v3/> (accessed 5.20.17).
- Grupp, H. (1997). *Messung und Erklärung des technischen Wandels – Grundzüge einer empirischen Innovationsökonomik*. Springer Verlag, Berlin Heidelberg.

- Haß, H.-J. (1983). *Die Messung des technischen Fortschritts*. München.
- Mendeley Homepage (2017). <https://www.mendeley.com/> (accessed 6.12.17).
- Hot Questions - Stack Exchange (2017). <https://stackoverflow.com/> (accessed 5.21.17).
- Jennex, M.E., Olfman, L. (2010). *A model of knowledge management success*. ME Jennex Smolnik Eds Strateg. Knowl. Manag. Success Explor. Organ. Effic. IGI Glob., pp 14–31.
- Jennex, M.E., Smolnik, S., Croasdell, D. (2016). "The Search for Knowledge Management Success", in: *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pp. 4202–4211.
- Lave, J., Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.
- Manlio Del Giudice, P., Vincenzo Maggioni, P., Ma, Z., Huang, Y., Wu, J., Dong, W., Qi, L. (2014). "What matters for knowledge sharing in collectivistic cultures? Empirical evidence from China", in *J. Knowl. Manag.*, vol. 18, pp. 1004–1019.
- netvision (2017). RAGE. <http://rageproject.eu/> (accessed 6.11.17).
- Nonaka, I., Takeuchi, H. (1995). *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press.
- OAuth 2.0 — OAuth (2017). <https://oauth.net/2/> (accessed 5.30.17).
- Open Archives Initiative Protocol for Metadata Harvesting (2017). <http://www.openarchives.org/pmh/> (accessed 6.10.17).
- OpenID Foundation website (2017).
- Osterwalder, A., Pigneur, Y. (2011). *Business Model Generation: Ein Handbuch für Visionäre, Spielveränderer und Herausforderer*, 1st edition, Campus Verlag, Frankfurt New York.
- Paukert, M., Niederée, C., Hemmje, M. (2011). "Knowledge in Innovation Processes", in: *Encyclopedia of Knowledge Management*.
- Polson, P.G., Lewis, C., Rieman, J., Wharton, C. (1992). "Cognitive walkthroughs: a method for theory-based evaluation of user interfaces", in: *Int. J. Man-Mach. Stud.* 36, pp. 741–773.
- Posnett, D., Warburg, E., Devanbu, P., Filkov, V. (2012). "Mining Stack Exchange: Expertise Is Evident from Initial Contributions", in: *2012 International Conference on Social Informatics*. pp. 199–204.
- Salman, M., Fuchs, M., Vu, B., Heutelbeck, D., Hemmje, M., Becker, J., Brocks, H. (2016). "Integrating Scientific Publication into an Applied Gaming Ecosystem", in: *Annual International Conference on Computer Games Multimedia & Allied Technologies, CGAT 2016*. Singapore.
- Salman, M., Heutelbeck, D., Hemmje, M. (2015). "Towards Social Network Support for an Applied Gaming Ecosystem", in: *Proceedings of the 9th European Conference on Games Based Learning ECGBL 2015*. Steinkjer, Norway, pp. 721–728.
- SlideShare.net (2017) <https://www.slideshare.net> (accessed 6.12.17).
- Specht, G., Beckmann, C., Amelingmeyer, J. (2002). *F&E-Management*. Schäffer-Poeschel-Verlag, Stuttgart.
- Stack Exchange API (2017). <http://api.stackexchange.com/> (accessed 6.8.17).
- Stack Overflow (2017). <https://stackoverflow.com/> (accessed 5.19.17).
- Travis CI (2017), "Test and Deploy Your Code with Confidence". <https://travis-ci.org/> (accessed 6.10.17).
- Web Hooks / FrontPage (2017). <https://webhooks.pbworks.com/w/page/13385124/FrontPage> (accessed 6.11.17).
- Welcome To UML Web Site! (2017). <http://www.uml.org/> (accessed 6.13.17).
- Wenger, E. (2011). *Communities of practice: A brief introduction*.
- Wiederhold, G. (1992). "Mediators in the architecture of future information systems", in: *Computer*, vol. 25, pp. 38–49.

Context-Aware Documentation in the Smart Factory¹

Ulrich Beez[†], Jürgen Bock^{††}, Tilman Deuschel[†], Bernhard G. Humm[†], Lukas Kaupp[†]

{ulrich.beez,tilman.deuschel,bernhard.humm,lukas.kaupp}@h-da.de

[†]Hochschule Darmstadt -

University of Applied Sciences

Haardtring 100

64295 Darmstadt, Germany

juergen.bock@kuka.com

^{††}KUKA Roboter GmbH

Zugspitzstraße 140

86165 Augsburg, Germany

Keywords: Smart factory, technical documentation, ontology, semantic matching, robotics

Abstract: When machine errors occur in factories, it is important to act quickly in an appropriate way. Depending on the complexity of the error situation and the skill level of the personnel, it is important to identify the appropriate technical documentation quickly. This paper presents a methodology for semantically matching symptoms and causes in error situations, and automatically presenting solutions to end-users in an intuitive way. For demonstration purposes, the use case of robotics application development is chosen.

1. Introduction

Whenever personnel of smart factories are confronted with machine errors, a critical aspect is the time required to put the machine back to operation. A machine error may reduce the overall productivity and hence impact the factories' competitiveness (Zhang and Ordóñez, 2012, p IX; Saalman and Hellingrath, 2016, p 601).

We illustrate this by means of the example use case of robot application development for the smart factory. Developers program robots to support dedicated factory processes. One feature of modern robots like e.g. KUKA LBR iiwa 14 R820 is its impedance control mode, i.e., motions can be programmed based on measured forces and torques. (KUKA, 2017). A common use case for impedance-controlled motion is the search of the physical workspace, e.g., a desk. The robot moves in one direction until it senses a specific force, i.e., until it reaches contact. Similar to a human elbow, the joints of the LBR iiwa have angle limits (An et al., 1989, p 1251; KUKA, 2017, p 30). In case of an impedance-controlled motion, a joint can reach its maximum angle if the expected force is not detected during the motion, and no default behaviour is programmed, resulting in a sudden stop during movement. To continue the desired movement, the application developer has to first determine which joint has hit its angle limit. Then, he has to rotate the joint out of its maximum angle and use a combination of other joints for moving to the desired position.

In general, error isolation until its root-cause is found can be a complex and time-consuming task (Lettin and Winterholer, 2017, p 2). The time required depends on a multitude of factors, such as the machine's complexity, the personnel's skill level and the availability of other information sources. Recent government and industry initiatives have focused on optimizing the manufacturing processes by studying on smart factories that use highly interconnected machines. For Germany: Industrie 4.0

¹ The project ProDok 4.0 is funded by the German Federal Ministry of Education and Research (BMBF) within the framework of the Services 2010 action plan under funding no. 02K14A110.

(Henninger et al., 2013), For the USA: The Industrial Internet (Industrial Internet Consortium 2014), For France: Industrie du Futur (2015). More precisely, according to Henninger et al. (2013), those initiatives have the potential to compensate short time deficiencies, allow flexible reaction to errors and even improve the work-life balance. The charts of Bauernhansl et al. (2014, p 31) indicate that the so-called next industrial revolution has the potential to introduce high savings in different industry cost areas. The German industry association for digital business Bitkom (2017) offers a closer insight: One of their recent press releases contains an excerpt of a study conducted in 314 companies with 3 or more employees stating that for Germany there is a possibility for an increase in productivity of up to 78.5 billion (10^9) euros until the year 2025. With regard to the number of ongoing research projects and their topics, a not sufficiently answered question might be how smart factories with increased connectivity can be utilized to achieve such a growth in productivity. One indicator is the web page of the German Federal Ministry of Education and Research (2017) which states in Germany, there are over 300 research project partner locations for ongoing projects within the smart factory domain. Some projects examine holistic approaches, e.g. the recently completed CyberSystemConnector (CSC) project interlinks distributed technical documentation (CyberSystemConnector, 2017). Other projects focus on specific scenarios. E.g. Lin et al. (2012) is introducing a combination of statistics and fuzzy logic for machine failure detection. A study of real time control systems for sensor networks in the same domain is conducted by Nguyen (2017).

In this paper, we propose a novel methodology for reducing the time span for fixing machine errors in smart factories by automatically providing appropriate technical documentation taking machine context into account. In the example use case presented above the robot application developer will automatically be alerted when a robot joint has reached its maximum angle. The information of how to move the robot back into the desired position will be provided in an intuitive way, integrated in the developer's workflow.

This paper is structured with five sections in total. Section 2 describes the problem. In Section 3, the proposed solution is described, which is evaluated in Section 4. A review of related literature can be found in Section 5. The papers' last section comprises conclusions and future work.

2. Problem Description

Based on expert interviews conducted with personnel of manufacturers for smart factory equipment (four robot application developers, one knowledge engineer, three support engineers) the following requirements have been collected:

- R1) In case a machine error occurs in a smart factory, personnel shall be enabled to *resolve the error quickly and with little effort*.
- R2) *Appropriate technical documentation* shall be provided, indicating causes and solutions of machine errors.
- R3) The provided technical documentation shall *match the machine context* in which the error occurred.

R4) The technical documentation shall be provided *automatically*, alerting the user as soon as the machine error occurs.

R5) *Devices* shall be supported which support the smart factory workflow, e.g., desktop computer, tablet PC, smartphone, or smartwatch.

R6) User *interaction* shall be *fast* to not disturb the personnel's workflow.

3. A Method for Context-Aware Documentation in the Smart Factory

3.1. User Interaction Concept

We explain the interaction concept of the methodology by means of the example use case introduced above. As soon as the robot stops during hand guiding, the user is alerted. This alert may be pushed to a suitable device, e.g., the development workstation, a tablet PC or even a smartwatch. The alert indicates the symptom of the machine error, i.e. 'Joint: the robot has stopped'. See Figure 1 for a screenshot of a dashboard view on the development workstation.

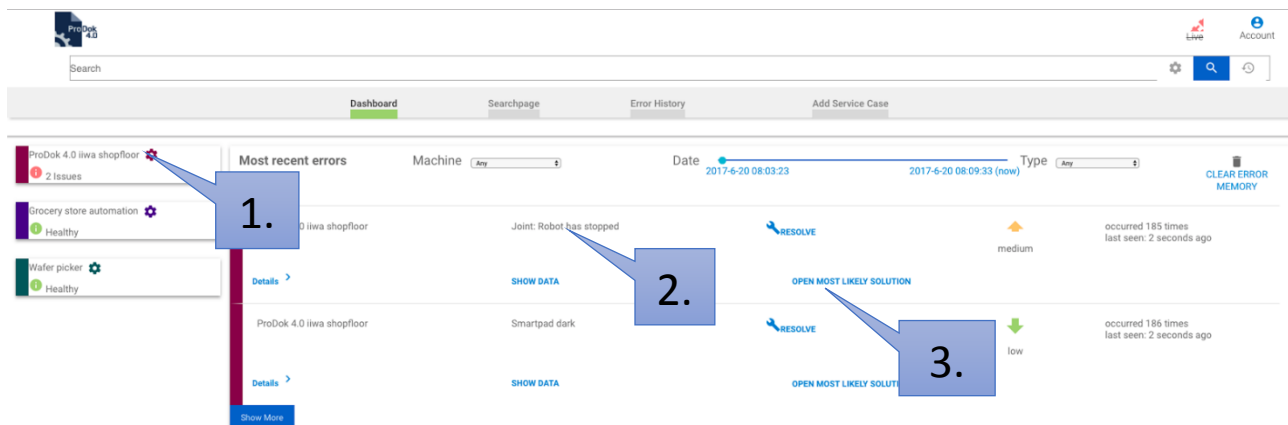


Figure 1: Dashboard View

An important aspect of the screen design is the clarity of information presentation, following the ISO Standard 9241-110 (ISO, 2006), respectively 9241-210 (ISO, 2010). The user centred design approach is applied, which includes close contact to the end user to define requirements collaboratively and test intermediate prototypes iteratively (Garret, 2011).

The dashboard component implements the interaction design pattern of sequence-of-use, which follows the mental model of spatial alignment of semantically related objects (Koffka, 2014). Elements that share a semantic relationship are grouped together and the interaction design provides subsequent interaction-steps for the main tasks.

The dashboard's most important components are:

- The overview functionality of all connected devices as a list (Figure 1, Mark 1).
- A table containing the most recent errors for all connected devices. Each column displays the error symptom alongside a shortcut towards the solution. E.g. 'Joint: the robot has stopped'

(Figure 1, Mark 2) and in addition a navigation component reducing the effort a user has to invest for finding a solution, e.g. ‘open most likely solution’, (Figure 1, Mark 3)

With a single interaction, e.g., a click on ‘open most likely solution’ (Figure 1, Mark 3), the user is provided with a solution to the most likely cause of this machine error. See Figure 2.

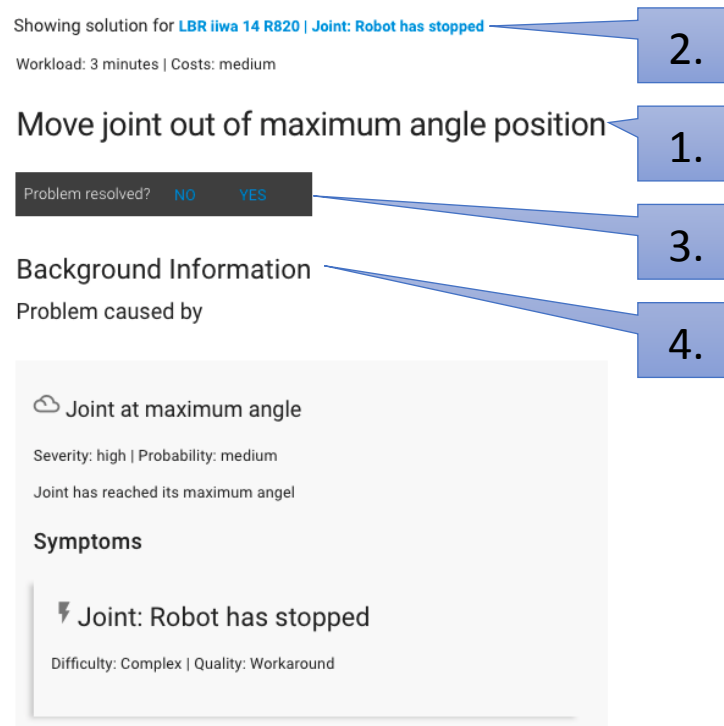


Figure 2: Solution View

The conceptual ideas of the Solution View (Figure 2) are as follows:

- Present the solution to the machine error, e.g. the solution text ‘Move joint out of maximum angle position’ (Figure 2, Mark 1).
- Provide a quick view to the user regarding error context and symptom, e.g. the blue headline ‘LBR iiwa 14 R820 | Joint: the robot has stopped’ (Figure 1, Mark 2).
- Collect user feedback for a solution, e.g. ‘Problem resolved?’ (Figure 2, Mark 3).
- Display solution background information to the user, matching both error context and error, e.g. the cause and symptom (Figure 2, Mark 4).
- In case that several causes for the machine error symptom (Joint: the robot has stopped) exist, they are sorted by likelihood in descending order. On the solution view following the most likely solution, the less likely ones are displayed.

3.2. Software Architecture

The software architecture is shown below in Figure 3 and consists of the 3 *layers* (Starke, 2015): *Presentation*, *Logic* and *Data*. Each layer encompasses different modules. Following Figure 3 the purpose of each module is described.

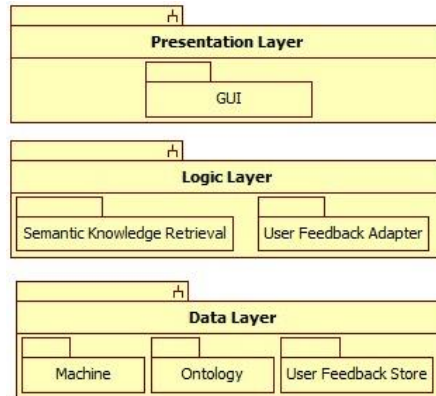


Figure 3: UML Class Diagram – Layer View

Presentation Layer: It contains the Graphical User Interface (*GUI*) which serves as entry point for the user.

Logic Layer: This layer encompasses two modules (a) the *Semantic Knowledge Retrieval*, with the purpose of providing accurately fitting documentation and (b) the *User Feedback Adapter* which has the purpose of handling user feedback.

Data Layer: Three modules are located here, (a) the *Machine*, which sends out event and context information, (b) the *Ontology*, containing hierarchies of products and errors, both interlinked, and modularized technical documentation and (C) the *User Feedback Store*, containing collected user feedback.

For providing an inside view of the components on the different layers, their interaction is shown in Figure 4 which is explained afterwards applying the introductions use case.

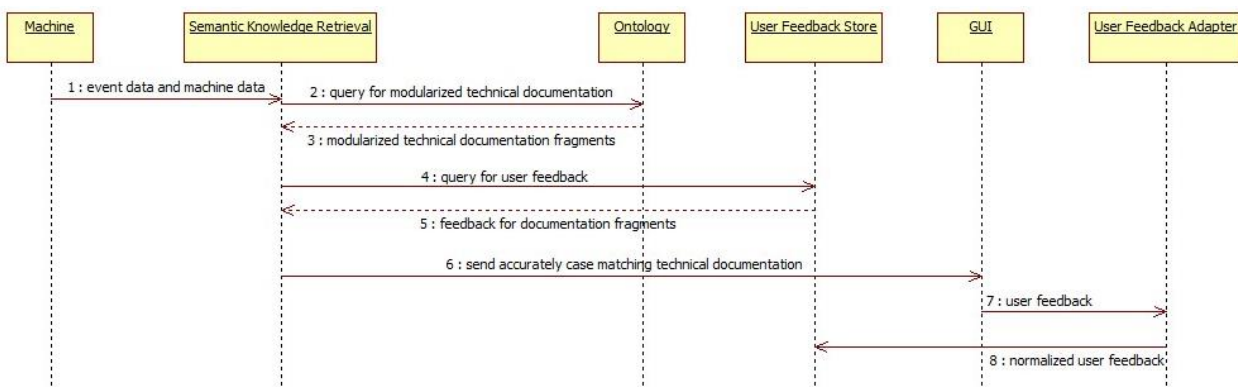


Figure 4: UML Sequence Diagram – Smart Documentation Retrieval

- Step 1: As soon as an error occurs, the machine sends *event data* and *machine data* to the ‘*Semantic Knowledge Retrieval*’ component. *Event data* describes the machine error in detail, e.g. ‘Joint 3 maximum angle reached, emergency stop’ on 13:45:17. *Machine data* contains context information, e.g. the robots’ digital identification plate with manufacturer and type, e.g., ‘KUKA’ and ‘LBR iiwa 14 R820’.
- Step 2,3: Modularized technical documentation is retrieved from the ontology by querying for the event data in combination with the machine data. The modularized technical documentation consists of the documentation fragments *symptom*, *cause*, and *solution* (SCS).
- Step 4,5: User feedback is queried for each of the previously retrieved SCSs.
- Step 6: *Technical documentation* accurately matching both the machine event and the personnel’s preference is forwarded to the ‘*Graphical User Interface*’ (GUI) component.
- Step 7,8: As soon as the GUI sends *user feedback* it is saved in the feedback store after normalization.

3.3. Ontology

An *ontology* specifies concepts and their relationships. One purpose of an ontology is to bridge terminology across different domains (Busse et al., 2015). Here, the event data and machine data, both originating from the machine, are bridged to the technical documentation. The ontology can be queried to retrieve *symptoms*, *causes* and *solutions* (SCS), matching both product and error data. See Figure 4 with an example following the W3C recommendation for concepts and abstract syntax (W3C, 2014).

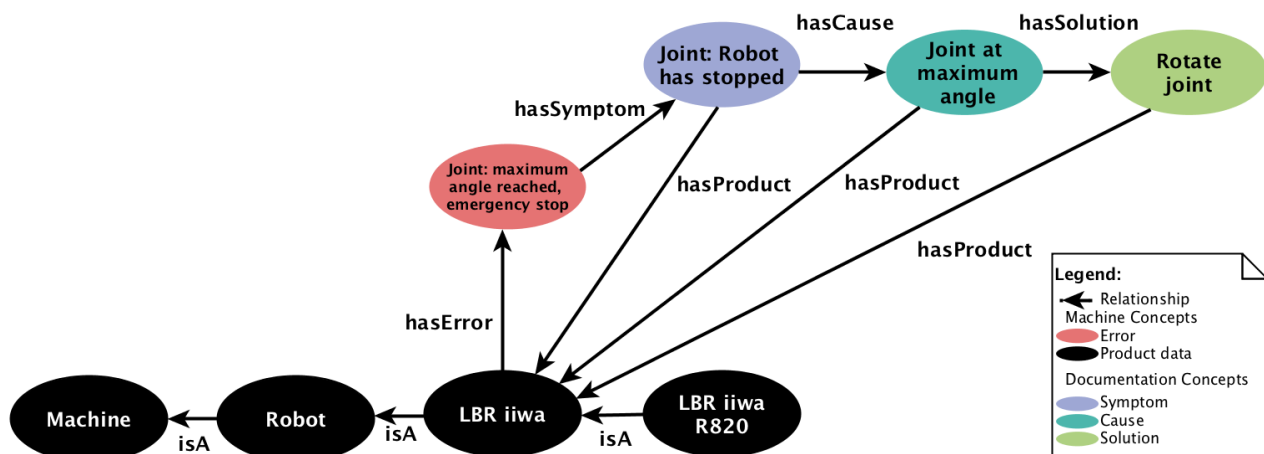


Figure 4: Ontology example

Within the ontology, different concepts are modelled:

- (a) Hierarchies of products and errors, both interlinked, e.g.:
 - ‘LBR iiwa R820’ ‘isA’ ‘LBR iiwa’
 - ‘LBR iiwa’ ‘has Error’ Joint: maximum angle reached, emergency stop’

- (b) Technical documentation splitted in symptoms, their causes and solutions (SCS) with linkages to (a), e.g.:
- ‘Joint: maximum angle reached, emergency stop’ ‘hasSymptom’ ‘Joint: Robot has stopped’
 - ‘Joint: Robot has stopped’ ‘hasCause’ ‘Joint at maximum angle’
 - ‘Joint at maximum angle’ ‘hasSolution’ ‘Rotate joint’
 - all linked to ‘LBR iiwa’ via ‘hasProduct’

The ontology enables modelling transitive relationships like ‘isA’. ‘LBR iiwa R820’ has no direct relationship with any error, symptom, cause, or solution. However, due to the ‘isA’ relationship with ‘LBR iiwa’, the relationships to the respective error, symptom, cause, and solution can be inferred.

In addition to the label attribute shown in Figure 4, concepts and relationships may have additional attributes. E.g., a solution may have an additional attribute containing a detailed description on how to apply the solution. An additional attribute for SCS vertices may contain information about the target user group.

3.4. Semantic Knowledge Retrieval

The aim of the ‘*Semantic Knowledge Retrieval*’ component is providing appropriate technical documentation. The component is triggered by the machine as soon as an error (event) occurs on the machine. The trigger mechanism hands over event data and machine data. Those data are used to query the ontology for technical documentation (symptoms, causes and solutions - SCS). Subsequently, the component queries the user feedback store to retrieve the feedback for the SCSs in order to rank the SCSs according to user preferences. Finally, the ranked SCSs are forwarded to the GUI. Semantic knowledge retrieval is a most complex process including pre-processing, semantic enrichment, ontology querying, and ranking. A detailed description is beyond the scope of this paper and can be found in Kaupp et al. (2017).

3.5. User Feedback Adapter

The ‘*User Feedback Adapter*’ is tracking interactions and feedback originating from the user while operating the GUI. Each user interaction such as a mouse click is recorded and stored. Recorded data are: user name, timestamp, GUI component (e.g. solution view), and concrete interaction (e.g. mouse click on ‘open most likely solution’ for the symptom ‘Robot has stopped’ at ‘KUKA LBR iiwa 14 R820’).

3.6. Prototype Implementation

The components have been implemented prototypically, serving as a demonstrator for the evaluation. The front end is programmed in Typescript using Angular2+ and Material-Design-Light. The back end uses Java as programming language in combination with the Spring library. To enable fast querying, Apache Solr has been selected as technology for implementing the ontology. The ontology has been filled with demo data. User feedback is stored in a PostgreSQL relational database. Communication between the individual components is realized using the technologies Web socket, RESTful, and Spring Message Routing. The robot is connected via OPC UA.

4. Evaluation

We evaluate the method and prototypical implementation presented in Section 3 by comparing it with the requirements specified in Section 2.

R1 - *Resolve error quickly and with little effort* - As shown in Section 3.1, a user can navigate from an error alert to an appropriate solution with a single interaction. In contrast to manual lookups without connection to the machine or even paper-based approaches, the proposed methodology can be assumed to enable the personnel to resolve errors more quickly and with reduced effort. Another aspect indicating the fulfilment of this requirement is the reduction of possible user typing errors.

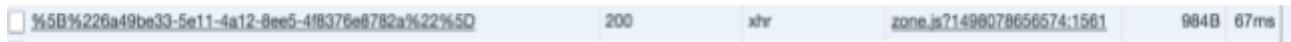
R2 - *Appropriate technical documentation* - The appropriateness of the solution presented is dependent on the users' preference, including his skill level, and the availability of appropriate documentation. The solution proposed in Section 3 offers functionalities for collecting user feedback and for storing it in the user feedback store (Section 3.5). After querying the ontology for technical documentation fragments (SCS), the user feedback store is consulted prior to re-ordering the SCSs (Section 3.4). In addition, the ontology can include information regarding the target user group (Section 3.3). Hence, the appropriateness of the documentation relies on both, the ontology and the user feedback store. Using the demo use case from the introduction alongside with demo data, the solution shown in Figure 2 is appropriate for a junior application developers skill level for programming a LBR iiwa R820, hinting towards the fulfilment of R2. For fully evaluating requirements R1 and R2, a comprehensive survey with factory personnel and real technical documentation needs to be conducted. This is planned as future work.

R3 - *Match the machine context* - The ontology is consulted as soon as a connected machine has communicated its error event and context. As stated in Section 3.3, the ontology acts as bridging mechanism between the robot and the technical documentation. Hence, a properly filled ontology is a prerequisite for mapping the machine context to technical documentation. The ontology contains relationships between products, errors and modularized technical documentation (SCS) and is used for querying SCSs matching both the error and machine context. Thus, it can be assumed that R3 is met.

R4 - *Automatically* - As stated in Section 3.1 and shown in Figure 1, the proposed methodology includes a navigation component that provides an automatic shortcut towards technical documentation with the most likely solutions. In favour of providing an overview of all connected machines and errors occurring on them, the concept includes a dashboard view where the navigation component is provided per machine error. Hence, the proposed solution includes an automatic path towards error solutions and fulfils R4.

R5 - *Devices* - The layered software architecture provides application programmers interface (API) between presentation and logic layer. GUIs for different device classes can be implemented, thus providing the basis for fulfilling R5. As explained in Section 3.6, the demonstrator is programmed using the Material-Design-Light library. Its grid mechanism is used while implementing the prototype, resulting in a responsive web page. It can be used in a web browser, on a tablet PC, and on a smartphone.

R6 - *Fast interaction* - The methodology presented in Section 3 can be implemented efficiently, allowing response times below 1 s for user interactions. Employing the prototype implementation (Section 3.6), an initial performance measurement has been conducted. A screenshot of the web browsers' debug console is shown in Figure 5, displaying an end to end response time of 67 ms for a symptom lookup. The test was conducted using the prototype locally, running on a MacBook Pro (Late 2016, i7-6567U @ 3.30GHz, 16 GB RAM) on battery, with a limited amount of test data.



%5B%226a49be33-5e11-4a12-8ee5-4f8376e8782a%22%5D	200	xhr	zone.js?1498078656574:1561	984B	67ms
--	-----	-----	----------------------------	------	------

Figure 5: Symptom lookup in 67 ms; using the prototype locally.

5. Related Work

Bunte et al. (2016) introduced an interface powered by natural language with the purpose of interacting with a smart device (system). Their solution is suitable for answering questions like “Are there any anomalies in the system?” by combining different techniques including natural language processing, ontologies, analytic and synthetic algorithms organized hierarchically. However, they do not provide means for providing an overview of machine status as the dashboard component introduced in Section 3 does. In addition, their solution relies on the user typing in their question, opening the possibility for misspelling. Our solution offers more automation thanks to the navigation component towards a technical documentation which includes solutions.

Hornung et al. (2014) introduced a new kind of semi-supervised robot anomaly detection, using a data-driven approach. By utilizing cascading machine learning techniques, their solution is able to detect anomalies while little information is present. The employed machine learning techniques are maps of valid and invalid data combined with a distance measurement and a Support Vector Machine. The technique of detecting anomalies from raw data in situations where little information is present could be an addition for the pure event-based approach described in Section 3. In contrast to the authors' proposed method in Section 3, their approach does not include a mapping of technical documentation to the detected errors/anomalies.

The CSC project (CyberSystemConnector, 2017) introduces an approach for storing technical documentation on individual machines or even machine parts in a smart factory. In contrast, we propose a central storage for technical documentation. We argue that the centralized approach eases the editing and updating process of technical documentation to a great extent. In case the information is stored in a distributed manner, i.e. each machine is carrying its own documentation, the need for an individual updating mechanism arises. Wang et al. (2017) compare a client-server paradigm with a mobile agent paradigm in the context of predictive maintenance, outlining pros and cons. Our approach follows the client-server paradigm whereas the CSC project follows the agent paradigm. Wang et al. confirm our view that resource management is eased following the client-server paradigm.

A concept regarding knowledge acquisition and process mapping in context of a smart factory is proposed by Panfilenko et al. (2016). By combining a semantic media wiki with process knowledge in BPMN and techniques of object character recognition (OCR) on hand-written incident reports, their solution is capable of mapping detected anomalies onto existing knowledge. Their demonstrator

focuses on manually collected incident reports and does not implement an automatic connection to the machine as the authors' solution does.

In accordance with the proposed software architecture, numerous publications in the context of smart factory propose a layered architecture. Examples are the cyber-physical systems architecture for Industry 4.0-based manufacturing systems (Lee et al., 2015), the reference architecture model for Industry 4.0 (DIN, 2016), and the event-driven manufacturing information system architecture for Industry 4.0 (Theorin et al., 2016).

6. Conclusions and Future Work

When machine errors occur in factories, it is important to act quickly in an appropriate way. In this paper, we have presented a methodology for semantically matching symptoms and causes in error situations, and automatically presenting solutions to end-users in an intuitive way. The methodology has been implemented prototypically. As stated in the evaluation section 4, the first prototype shows promising results.

To implement the proposed concept for a more production like environment, a number of next steps are required. The ontology needs to be filled with comprehensive technical documentation. An interface to a state-of-the-art knowledge management system is currently being implemented, providing the ontology. A thoroughly conducted study is planned which will compare the stand-alone knowledge management system with our integrated, context-aware solution. Learnings from the study will be used to improve the solution before it can replace an existing stand-alone system in production.

A future addition could be the incorporation of automatic fixes for common machine errors. Users could then be offered an additional option: 'automatically apply solution'. May our contribution help improve the productivity and efficiency of smart factories.

References

- An, K. N., Kaufman, K. R. and Chao, E. (1989). "Physiological considerations of muscle force through the elbow joint", In: *Journal of Biomechanics*, vol. 22, 11-12, pp. 1249-1256. doi: 10.1016/0021-9290(89)90227-3
- Bauernhansl, T., Hompel, M. ten and Vogel-Heuser, B., eds. (2014). *Industrie 4.0 in Produktion, Automatisierung und Logistik: Anwendung, Technologien, Migration*, Wiesbaden, Springer Vieweg. doi: 10.1007/978-3-658-04682-8
- Bitkom (2017). *IT-Unternehmen bauen Angebote für die Industrie 4.0 aus* [Online]. Available at <https://www.bitkom.org/Presse/Presseinformation/IT-Unternehmen-bauen-Angebote-fuer-die-Industrie-40-aus.html> (Accessed 21 March 2017).
- Bunte, A., Diedrich, A. and Niggemann, O. (2016). "Integrating semantics for diagnosis of manufacturing systems", In *IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*. Berlin, Germany, 06/09/2016 - 09/09/2016, IEEE, pp. 1-8. doi: 10.1109/ETFA.2016.7733721
- Busse, J., Humm, B., Lubbert, C., Moelter, F., Reibold, A., Rewald, M., Schluter, V., Seiler, B., Tegtmeier, E. and Zeh, T. (2015). "Actually, What Does "Ontology" Mean?: A Term Coined by Philosophy in the Light of Different Scientific Disciplines", In: *Journal of Computing and Information Technology*, vol. 23, no. 1, p. 29. doi: 10.2498/cit.1002508
- CyberSystemConnector (2017). *Maschinendokumentation intelligent erstellen und nutzen* [Online]. Available at <https://www.cyber-sc.de/site/> (Accessed 8 June 2017).
- DIN (2016) *SPEC 91345:2016-04: Referenzarchitekturmodell Industrie 4.0 (RAMI4.0)*, Berlin: Beuth.
- Garrett, J. J. (2011). *The elements of user experience: User-centered design for the Web*, 2nd ed, Berkeley, Calif., New Riders.

- German Federal Ministry of Education and Research (2017). Industrie 4.0 Standortverteilung Deutschland [Online]. With filter active projects. Available at <https://www.bmbf.de/de/zukunftsprojekt-industrie-4-0-848.html> (Accessed 5 May 2017)
- Henninger, K., Wahlster, W. and Helbig, J. (2013). *Recommendations for Implementing the Strategic Initiative Industrie 4.0: Final Report of the Industrie 4.0 Working Group* [Online]. Available at http://www.acatech.de/fileadmin/user_upload/Baumstruktur_nach_Website/Acatech/root/de/Material_fuer_Sonderseiten/Industrie_4.0/Final_report__Industrie_4.0_accessible.pdf (Accessed 13 January 2017).
- Hornung, R., Urbanek, H., Klodmann, J., Osendorfer, C. and van der Smagt, P. (2014). "Model-free robot anomaly detection", *IEEE/RSJ International Conference on Intelligent Robots and Systems 2014*, pp. 3676-3683. doi: 10.1109/IROS.2014.6943078
- Industrial Internet Consortium (2014). *About Us* [Online]. Available at <http://www.iiconsortium.org/about-us.htm> (Accessed 8 June 2017).
- French Embassy in Berlin (2015). *Slidset: Industrie du Futur* [Online]. Available at <http://www.ambafrance-de.org/Vorstellung-des-neuen-franzosischen-Plans-Industrie-du-Futur-in-der-Botschaft>.
- ISO (2006). *DIN ISO 9241-110:2006: DIN EN ISO 9241-110 Ergonomie der Mensch-System-Interaktion - Teil 110: Grundsätze der Dialoggestaltung (ISO 9241-110:2006); Deutsche Fassung EN ISO 9241-110:2006: Perinorm* [Online]. Available at <http://perinorm-s.redi-bw.de/volltexte/CD21DE04/1464024/1464024.pdf> (Accessed 28 June 2013).
- ISO (2010). *DIN ISO 9241-210:2010: Ergonomie der Mensch-System-Interaktion - Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme* [Online]. Available at <http://perinorm-s.redi-bw.de/volltexte/CD21DE05/1728173/1728173.pdf> (Accessed 29 October 2014).
- Kaup, L., Beez, U., Humm, B. G. and Hülsmann, J. (2017). "From Raw Data to Smart Documentation: Introducing a Semantic Fusion Process for Cyber-Physical Systems", In: *Proceedings of the 2017 Collaborative European Research Conference (CERC 2017)*, Karlsruhe, Germany.
- Koffka, K. (2014). *Principles of Gestalt psychology*, Mimesis Edizioni.
- KUKA (2017). *Product Brochure: Sensitive robotics_LBR iiwa* [Online]. Available at https://www.kuka.com/-/media/kuka-downloads/imported/9cb8e311bfd744b4b0eab25ca883f6d3/kuka_lbr_iiwa_broschuere_de.pdf (Accessed 23 May 2017).
- Lee, J., Bagheri, B. and Kao, H.-A. (2015). "A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems", In: *Manufacturing Letters*, vol. 3, pp. 18-23. DOI: 10.1016/j.mfglet.2014.12.001 (Accessed 19 May 2017).
- Lettner, D. and Winterholer, M. (2017). *Embedded software verification and debugging*, New York, NY, Springer. doi: 10.1007/978-1-4614-2266-2
- Lin, L. and Lee, H.-M. (2012). "Machine Failure Diagnosis Model Applied with a Fuzzy Inference Approach", in *Watada, J. (ed) Intelligent decision technologies: Proceedings of the 3rd International Conference on Intelligent Decision Technologies (IDT2011)*, Berlin, New York, Springer, pp. 185-190. doi: 10.1007/978-3-642-22194-1_19
- Nguyen, V.-A.-Q. (2017). "Study on realtime control system in IoT based smart factory: Interference awareness, architectural elements, and its application", In: *2017 Seventh International Conference on Information Science and Technology (ICIST)*. Da Nang, Vietnam, IEEE, pp. 25-28. doi: 10.1109/ICIST.2017.7926774
- Panfilenko, D., Poller, P., Sonntag, D., Zillner, S. and Schneider, M. (2016). "BPMN for knowledge acquisition and anomaly handling in CPS for smart factories", In: *2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*. Berlin, Germany, 06/09/2016 - 09/09/2016, IEEE, pp. 1-4. doi: 10.1109/ETFA.2016.7733686
- Saalmann, P. and Hellingrath, B. (2016). "Coordination in Spare Parts Supply Chains", In: *2016 International Conference on Collaboration Technologies and Systems (CTS)*. Orlando, FL, USA, IEEE, pp. 601-604. doi: 10.1109/CTS.2016.0111
- Starke, G. (2015). *Effektive Software-Architekturen: Ein praktischer Leitfaden*, 7th edn, München, Hanser. doi: 10.3139/9783446444065
- Theorin, A., Bengtsson, K., Provost, J., Lieder, M., Johnsson, C., Lundholm, T. and Lennartson, B. (2016). "An event-driven manufacturing information system architecture for Industry 4.0", In: *International Journal of Production Research*, vol. 55, no. 5, pp. 1297-1311. DOI: 10.1080/00207543.2016.1201604 (Accessed 22 May 2017).
- W3C (2014). *RDF 1.1 Concepts and Abstract Syntax* [Online]. Available at <https://www.w3.org/TR/rdf11-concepts/> (Accessed 16 June 2017).
- Wang, J., Zhang, L., Duan, L. and Gao, R. X. (2017). "A new paradigm of cloud-based predictive maintenance for intelligent manufacturing", In: *Journal of Intelligent Manufacturing*, vol. 28, no. 5, pp. 1125-1137. DOI: 10.1007/s10845-015-1066-0.

Zhang, C. and Ordóñez, R. (2012). *Extremum-seeking control and applications: A numerical optimization-based approach*, New York, London, Springer. doi: 10.1007/978-1-4471-2224-1

User centric web usage mining as a tool for Social Sciences

Thomas Bettermann, Thomas Schlegel
Hochschule Karlsruhe – Technik und Wirtschaft
Institut of Ubiquitous Mobility Systems (IUMS)
e-mail: thomas.bettermann@hs-karlsruhe.de

Keywords: Web Usage Mining, Behavior Tracking, Algorithmic Media, Mobile-enabled

1. Motivation and Background

Through the Internet as one of the most important information and communication media and the ever-increasing offer of digital media, the media landscape is changing fundamentally. Especially younger generations of people receive their information from social networks, in particular Facebook (PEW 2015) as well as other media that are personalized, filtered and individually created. Research also shows that information gathering on the web is often done through online offers of established, traditional media (Bernhardt et al. 2015). Beside the targeted information search, the random stumbling over information, called 'Incidental Exposure', on the Internet and especially in the social media area is of central importance (Valerina & Vaccari). This information behavior takes place as a result of technical suggestion mechanisms, which take over the selection of content relevant to the individual users. This also means that research on the use of media cannot be covered by the usual applied methods of social science research anymore, since the content of the media is partly tailored to each consumer and additionally changing over time, e.g. by new comments read below an article, which makes analysis based on articles in media unreliable. De Vreese and Neijens (2016, p.70) analyzed 204 articles on the measurement of media usage, and concluded that 94% of the studies still rely on self-reported data. The drawback is that the results are retrospectively collected and can be distorted by memory performance, social desirability and other impacts of human information reproduction. With the increasing shift in media consumption to online media, there is the chance that media usage behavior can be recorded by means of technical measurements to overcome these shortcomings. The precise design of so-called tracking studies has so far been established neither from a social-science methodological point of view nor with regard to the technical implementation. Therefore, it seems to be a great contribution and challenge to design and implement a technical tool that captures media usage behavior to generate social-scientifically valid study data and interpretable variables. Within the scope of a joint research project in close collaboration between communication and computer scientists, we have designed and implemented a tracking tool which we will describe in this paper.

2. Related work

In computer science, Web Mining is a major topic that is differentiated in Web Content Mining, Web Usage Mining and Web Structure Mining (Kosala and Blockeel 2000). Web Usage Mining examines the navigation behavior on the Web and is thereby related to user tracking. Another important factor is the distinction between the source of the data that could be collected on the server side, the client side or a proxy server. For business purposes (e-commerce), server log files of a single web page are often processed and analyzed in order to improve their usability (Grace et al. 2011). Proxy servers act as intermediary between several clients and the Internet and, like the server-side approach, provide log files that contain all the web page requests and responses (Pierrakos et al. 2003). The client-side data collection is generally implemented by modified browsers or browser extensions. This makes it possible to capture and analyze interactions with the browser for multiple domains. There are several approaches for client level web usage mining. An early approaches characterize browsing strategies

(Catledge and Pitkow 1995) with the help of a modified web browser. An instrumented version of a Firefox browser and an intermediary to inject java script in visited web sites was used in a long term study to investigating the users browsing behavior (Weinreich et al. 2008). With the usage of a Firefox browser add-on, the DERI Online Behavior Study had the goal to create a dataset in non-intrusive, completely anonymous and privacy-preserving way (v. d. Weth and Hauswirth 2013).

3. Technical concept and implementation

Based on the considerations of how to follow the online user behavior on the server, proxy or client level, we chose the client-level approach, because this allows to directly capture user interactions as well as the site content for multiple domains. The tracking solution therefore consists of a client-side component for capturing the web browsing activity as well as a server-side component which collects, stores and processes the data for all participants. The implementation of the tracking tool is done on the client side as a browser add-on for Mozilla Firefox. At the time of development, Firefox was the only widespread browser to support add-ons on desktop and mobile devices. In spite of the passive observations of the users' behavior, no statements can be made about the motive or the reasons of the web usage behavior. For this purpose, Experience Sampling Method survey (in short: ESM) was integrated, which allows to ask the user context related questions. An overview of the overall system and its components is shown in Figure 1. The individual components are described below.

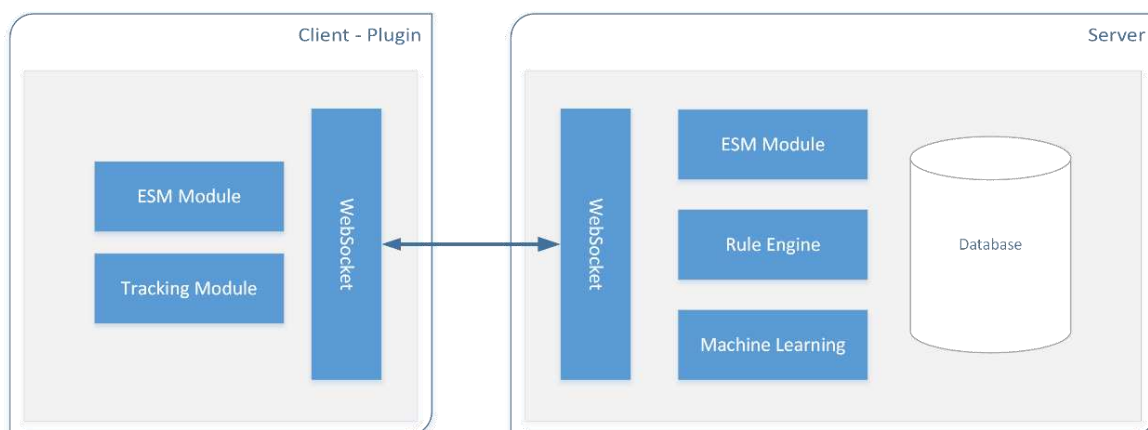


Figure 1 Technical Overview

Browser-Plugin

The browser add-on basically records the events that are triggered by the user in the browser and sends them to the server component. A distinction is made between two event sources. Events, such as opening a tab or entering a URL into the address bar, are triggered directly by the browser interface. Other events are generated by interacting with a website, such as clicking a link or scrolling on a website. An event contains information about the type of interaction, the time of the occurrence, as well as additional information depending on the event, including context.

Server

The browser events of all users are received and stored on the server side in a time series memory. This is a large quantity of raw data, which can be difficult to read or to be examined in this form. In order to simplify the further analysis, in the next step the data are preprocessed in a suitable form. For this purpose, a rule engine is used, which uses a rule set to derive activities from the raw events, which are more meaningful and also contain new, derived information. For example, a "SiteView" activity is derived from an event to load a page in a tab and the close-event of the same tab. This

activity describes the viewing of a website and contains additional information to the references to the source events, such as the residence time on the website. Closely linked to this a rough categorization of the website by their URL is performed in almost real-time. The category is then attached as additional information to the "SiteView" activity. This categorization is also used to pop up ESM questions on specific categories. If certain pre-configurable conditions are met, ESM question are presented to the user by the browser add-on. These conditions are defined on the server side as rules in the rule engine. For example, when a user visits a website of a particular category and stays for a certain time, a rule is raised and a question is sent to the browser add-on, which displays the question in the browser and sends the responses of the user back to the server. For this scenario, an in real-time categorization of the pages visited is indispensable.

4. Results & Discussion

The tracking tool has been used as part of the research project in two pre-tests and one full case study, which were carried out by communication scientists. The case study showed the functionality of the tracking tool as well as the viability of the approach. The web usage behavior of 83 subjects was tracked over a period of two weeks. Around 3 million events were captured and 125.000 web sites were visited. The full record has a size of 16 gigabytes. Problems with the tool could be recognized and improved already in the preliminary studies. We could see, for example, that the recording of certain events (e.g. scrolling events) is not practical, as they increased data volume significantly and client-side performance problems could occur. Therefore, we did not include these events in the case study.

5. Outlook

In the next step, the collected observation data must be transformed into socio-scientific analysis variables. Due to the large amount of data, preparation and analysis can no longer be carried out manually and must therefore be computer-assisted. Since the raw data, i.e. events and full texts of the web pages, have been stored and are available, the analysis can be carried out using various methods, for example machine learning for the automatic categorization of the visited pages using the full texts. Furthermore, rules can be added to the rule engine in order to aggregate the individual actions into larger activities, respectively socio-scientific variables, in order to answer the specific research questions. Further improvement of the tool itself is also carried out to allow for extended and new applications in studies.

Acknowledgement

This publication has been produced in the context of the project "Algorithmic Transformation of the Public Sphere. Chances and risks related to the development of information and communication technologies for citizens' political information literacy" and is funded by the German Federal Ministry of Education and Research with a project duration of two years (10/2015 – 09/2017) under the guidance of Prof. Dr. Lutz M. Hagen and Prof. Dr.-Ing. Thomas Schlegel.

References

- Bernhardt, U., Dohle, M., & Vowe, G. (2015). Wer nutzt wie das „Web 2.0“ für Politik? Der Stellenwert von Social Media in politischen Kontexten. In: K. Imhof, R. Blum, H. Bonfadelli, O. Jarren, & V. Wyss (Hrsg.), *Demokratisierung durch Social Media? Mediensymposium 2012*. (S. 41–54). Wiesbaden: Springer VS.
- Catledge, L. and Pitkow, J. (1995). Characterizing Browsing Strategies in the World-Wide Web. *Computer Networks and ISDN Systems*, 27(6), pp. 1065 - 1073.

- Grace L. K. J., Maheswari V. and Nagamalai D. (2011). Web Log Data Analysis and Mining. In: N. Meghanathan, B. K. Kaushik and D. Nagamalai, eds., *Advanced Computing. CCSIT 2011. Communications in Computer and Information Science*, vol 133. Berlin, Springer VS, pp. 459–469.
- Kosala, R. and Blockeel, H. (2000). Web Mining Research: A Survey, *ACM SIGKDD Explorations Newsletter*, 2(1), pp. 1-15.
- Pierrakos, D., Paliouras, G., Papatheodorou, C. and Spyropoulos, C. (2003). Web Usage Mining as a Tool for Personalization: A Survey, In: *User Modeling and User-Adapted Interaction*, 13(4), pp. 311-372.
- Pew Research Center (2015). The Evolving Role of News on Twitter and Facebook. Pew Research Center, pp. 1-10. Available at: <http://www.journalism.org/files/2015/07/Twitter-and-News-Survey-Report-FINAL2.pdf> [Accessed: 2017-06-29].
- Vreese, C. de and Neijens, P. (2016). Measuring Media Exposure in a Changing Communications Environment. In: *Communication Methods and Measures*, 10(2-3), pp. 69–80.
- Weinreich, H., Obendorf, H., Herder, E. and Mayer, M. (2008). Not Quite the Average: An Empirical Study of Web Use, In: *ACM Trans. Web*, 2(1), pp. 5:1–5:31.
- Weth, C. v. d., and Hauswirth, M. (2013). DOBBS: Towards a Comprehensive Dataset to Study the Browsing Behavior of Online Users. In: *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, IEEE, pp. 51-56.

Cloud Interoperability: Approaches and Challenges

Donagh Moynihan, and Samane Abdi

Department of Computer Science, Cork Institute of Technology, Bishopstown, Cork, Ireland.

Email: Donagh.Moynihan@mycit.ie; Samane.Abdi@cit.ie

Abstract: Cloud computing is the primary platform for provisioning infrastructure. Cloud computing environments have evolved into vendor specific infrastructure islands. The deployment of applications over different infrastructures or different geographic locations is often not a possibility. These issues have motivated the initiative of interconnected multiple clouds or Interclouds in order to attain a better quality of service, reliability and flexibility within cloud computing. There are multiple projects and organisations working on achieving cloud interoperability. We demonstrate a comprehensive up to date analysis on the cloud interoperability efforts, including the challenges and approaches with the direction of achieving an Inter-cloud.

Keywords: Cloud Computing, Interoperability, Challenges

1. Introduction

Cloud computing is a paradigm that allows users with different computing demands to access a shared pool of resources (e.g., network, storage, database, servers, services and applications). Many commercial cloud providers have emerged in recent years, and each typically provides its own cloud infrastructure, APIs and application description formats to access the cloud resources. This limits the flexibility that cloud end users would like to process, when it comes to deploy applications over different infrastructures in different geographic locations, or to migrate a service from one cloud provider's to another. This limitation is called vendor lock-in. To enable seamless sharing of resources from a pool of cloud providers, efforts have emerged recently to facilitate cloud interoperability, i.e., the ability for multiple cloud providers to work together (Toosi et al., 2014). Cloud interoperability can be achieved by two general approaches: 1) the common standards; and 2) the brokers. The first approach requires cloud providers to adopt and implement standard interfaces, protocols and architectural components that will assist in cloud interoperability. The second approach, brokers, involves a third party that translates the messages between the different cloud provider interfaces. This allows cloud consumers to switch and negotiate between cloud providers and effectively allow the clouds become interoperable. In this paper we conduct a comprehensive up to date analysis on the different approaches on cloud interoperability and their challenges.

2. Cloud Interoperability on Different Approaches

In this section we analyse the cloud interoperability through different approaches including: cloud ontology, semantic models, cloud pattern, and model-based (which are classified as the common standards); multi-agent systems (which is classified as the brokers and middle wares).

2.1 Cloud Ontology

Ontologies offer the means of explicit representation of the meaning of different terms or concepts, together with their relationships (Moscato et al., 2011). Cloud ontology deals with the cloud issues by focusing on the application layer (SaaS) rather than the IaaS layer. The main goal in the use of ontology for cloud computing is that of obtaining transparent and simple access to heterogeneous

cloud computing resources and to avoid locked-in proprietary solutions. The mOSAIC project (Moscato et al., 2011) as a sample of this approach uses ontology techniques and enables interoperability among existing cloud solutions, platforms and services, both from end-user and developer side. This solution has produced an architectural framework that utilises a standard protocol and open source API for cloud federation.

Interoperability through Cloud Ontology

Ontologies are believed to be the next trend for solving heterogeneity problem in the integration of services. However, ontologies alone cannot resolve heterogeneity problems for all domains because there is not a single ontology that can address specifications for all domains of interest (Abdi, 2014). Many of the existing works assume semantic mapping between ontologies is performed manually and that an automated dynamic integration solution is required. This is also significant in achieving cloud interoperability globally, as scalability, automation and standardisation are essential in allowing all domains of the cloud interoperate and communicate. Cloud ontologies alone cannot achieve these requirements, therefore, additional cloud interoperability approaches and measures are required.

2.2 Semantic Models

The second approach is quite similar to the first, with the implementation of data modelling. The challenge in this approach is moving the data from a schema-less data store (such as Google Bigtable) to a schema-driven data such as a relational database. Semantic-enabled services provide an abstraction layer over the cloud APIs (PaaS) and cloud resources (IaaS) so that the application layer (SaaS) benefits from the intelligent selection of services, with automation of different tasks.

Interoperability through Semantic Models

In cloud computing platforms although the interfaces are different, their operations' semantics are similar. This similarity means metadata added through annotations pointing to generic operations models would play a key role in consolidating the APIs and enable interoperability among the heterogeneous cloud environments. This enables automatic advanced discovery and composition of cloud services, and can be seen, enriched with proper inference rules and description of a certain number of provider's cloud services to support inter-cloud interoperability. Incorporating semantic technologies will greatly assist in overcoming the severe diversity in cloud computing platforms.

2.3 Cloud Pattern

On the subject of cloud interoperability, the Cloud Standards Customer Council (C. Baudoin, 2013) suggest that one approach that may help with the migration of applications to cloud computing is the use of patterns such as Amazon Web Services (AWS, 2013) and Microsoft Azure (Microsoft Azure cloud design patterns, 2014) and CloudPatterns.org (CloudPatterns.org, 2016). These catalogues describe general functionalities and behaviors and they propose architectural models that are much less bound to specific cloud platforms, thus resulting in less detail and better flexibility.

Interoperability through Cloud Pattern

Having explored some of the cloud vendor's catalogues and determined their inflexibility and platform-dependent objectives, patterns can be used to describe and model existing cloud applications in a very easily understandable manner, tracing back the different cloud implementations to a set of well-known and stable solutions. Cloud Patterns therefore offer an easier understanding of the exact functionalities and responsibilities of a specific cloud application component. The idea that these

characteristics could later be substituted with similar characteristics enhancing cloud interoperability. Cloud patterns have the potential to be utilised to enhance portability and interoperability between cloud platforms allowing integration even at the IaaS and PaaS service model layers. Further advantages of using cloud patterns, and particularly an agnostic pattern, when used as a canvas on which to develop a new application. It would be possible to implement each of the pattern's participants with services and components exposed by different cloud vendors.

2.4 Model-Based

A model based approach is based on the theory of “model once, generate anywhere”, which is particularly relevant when it comes to designing and managing applications across multiple services, as well as migrating them from one service to another.

Interoperability through Model-Based Approach

In the context of cloud computing, model-driven development allows developers to design software systems in a cloud agnostic way, and to be supported by model transformation techniques into the process of instantiating the system into specific multiple clouds. MODAClouds (MODAClouds, 2016), ARTIST (ARTIST, 2016) and PaaSage (PaaSage, 2016) are some of the research groups and projects focused on model-based approaches that will interoperate the PaaS and SaaS service models.

2.5 Multi Agent Systems

A multi-agent system removes the control from the cloud provider by acting as a mediator or broker for the cloud user in order to consume cloud resources from diverse cloud providers. The brokers are essentially a third party that translates the messages between the different cloud provider interfaces. A multi-agent system represents a distributed computing paradigm based on multiple interacting agents that are capable of intelligent behavior (D. Talia, 2011). The interacting agents are self-contained software programs embodying domain knowledge and having the ability to behave with a specific degree of independence to carry out actions needed to achieve specified goals. This allows cloud consumers to switch and negotiate between cloud providers and effectively allow the clouds become interoperable. Multi-agent systems have the ability to address interoperability across all the service models, IaaS, PaaS and SaaS. However, the main challenge for a multi-agent system such as a cloud agency and cloud broker is to attain global cloud interoperability or the Inter-cloud is scalability.

Interoperability through Multi-Agent Systems

Multi-agent systems seem to offer one of the most effective approaches to solve a number of interoperability issues and automate a number of activities, in particular brokering, negotiation, management, monitoring and reconfiguration in multiple clouds. Talia (2011) highlights that despite the differences between cloud computing and multi-agent systems, they are two distributed computing models. Therefore several common problems can be identified and several benefits can be obtained by the integrated use of cloud computing systems and multi-agents. On the other side, software agents can be used as basic components for implementing intelligence in cloud computing systems making them more adaptive, flexible and autonomic in resource management, service provisioning, and in running large-scale applications. From the two approaches discussed, scalability is achieved through the introduction of multiple agents to communicate with the cloud providers. As discovered in the mOSAIC project Discovery and Mapping Service, there is still an element of manual intervention present. To remove manual intervention, it is evident there has to be a degree of

compromise and cooperation from the cloud provider on what the components such as languages, APIs, and communication protocols are being applied to deliver their cloud platform. Nevertheless, a multi-agent system approach certainly offers substantial progress and contribution towards achieving cloud interoperability among heterogeneous cloud providers.

3. Conclusions

This paper conducted an up to date analysis on the cloud interoperability efforts, including the ongoing research projects and studies in the area. An overview of the cloud interoperability approaches including semantic models, multi-agent systems, cloud patterns, and model-based is discussed. Each of these approaches have their own challenges and advantages towards achieving interoperability: The model-based approach has limitations on deployment and adaptability with different clouds; Multi-agent systems are useful for large-scale interoperability in clouds federation; The cloud pattern provides flexible and platform-independent objectives that can be used by different cloud vendors; and finally, the semantics play a significant role in resolving cloud computing interoperability which can be achieved through ontology techniques. Future direction of the research will investigate the progression of the interoperability to achieve an Inter-cloud (Toosi et al., 2014).

References

- Abdi, S. (2014). Integration of Heterogeneous Policies for Trust Management, in: *Computer Software and Applications Conference Workshops (COMPSACW)*, 2014 IEEE 38th International, pp. 413–419. doi:10.1109/COMPSACW.2014.71
- ARTIST (2016). Artist Project - Advanced Software-based Service provisioning and Migration of legacy Software [online]. URL <http://www.artist-project.eu/> [accessed 3.7.16].
- AWS (2013). AWS-CloudDesignPattern [online]. URL http://en.clouddesignpattern.org/index.php/Main_Page [accessed 2.27.16].
- Baudoin, C. (2013). Migrating Applications to Public Cloud Services: Roadmap for Success [online]. URL <http://www.cloud-council.org/deliverables/CSCC-Migrating-Applications-to-Public-Cloud-Services-Roadmap-for-Success.pdf> [accessed 2.21.16].
- CloudPatterns.org (2016). Cloud Computing Patterns [online]. URL <http://cloudpatterns.org/> [accessed 2.28.16].
- Cretella, G., Di Martino, B. (2012). Towards a Semantic Engine for Cloud Applications Development, in: *2012 Sixth International Conference on Complex, Intelligent and Software Intensive Systems (CISIS)*, pp. 198–203. doi:10.1109/CISIS.2012.159
- Martino, B.D., Cretella, G., Esposito, A., Sperandio, R.G. (2014). Semantic Representation of Cloud Services: A Case Study for Microsoft Windows Azure, in: *Proceedings of the 2014 International Conference on Intelligent Networking and Collaborative Systems. IEEE Computer Society*, Washington, DC, USA, pp. 647–652. doi:10.1109/INCoS.2014.76
- Martino, B.D., Cretella, G., Esposito, A., Willner, A., Alloush, A., Bernstein, D., Vij, D., Weinman, J. (2015). Towards an Ontology-Based Intercloud Resource Catalogue – The IEEE P2302 Intercloud Approach for a Semantic Resource Exchange, in: *2015 IEEE International Conference on Cloud Engineering (IC2E)*, pp. 458–464. doi:10.1109/IC2E.2015.76
- Microsoft Azure cloud design patterns (2014). Cloud Design Patterns: Prescriptive Architecture Guidance for Cloud Applications [online]. URL <https://msdn.microsoft.com/en-us/library/dn568099.aspx> [accessed 2.21.16].
- MODAClouds (2016). MODAClouds Project [online]. URL <http://www.modacLOUDS.eu/> [accessed 2.28.16].
- Moscato, F., Aversa, R., Di Martino, B., Fortis, T., Munteanu, V. (2011). An analysis of mOSAIC ontology for Cloud resources annotation, in: *2011 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp. 973–980.
- PaaSage (2016). PaaSage - Model-based Cloud Platform Upperware [online]. URL <http://www.paasage.eu/> [accessed 3.5.16].
- Sheth, A., Ranabahu, A. (2010). Semantic Modeling for Cloud Computing, Part 2. *IEEE Internet Computer*. 14, pp. 81–84. doi:10.1109/MIC.2010.98
- Talia, D. (2011). Cloud Computing and Software Agents: Towards Cloud Intelligent Services [online]. URL http://ceur-ws.org/Vol-741/INV02_Talia.pdf [accessed 3.17.16].

Toosi, A.N., Calheiros, R.N., Buyya, R. (2014). Interconnected Cloud Computing Environments: Challenges, Taxonomy, and Survey. *ACM Comput Surv* 47, 7: pp. 1–7. doi:10.1145/2593512

From Monolithic To Modern Software Development

An approach to migrate and scale legacy applications using software containers, microservices and Cloud Computing

Marius Krämer¹, Patrick Wiener¹, Andreas Abecker² and Jens Nimis¹

Faculty of Management Science and Engineering

¹Karlsruhe University of Applied Sciences

²Disy Informationssysteme GmbH
me@mariuskraemer.com

Keywords: Software Container, Microservices, Cloud-based Scaling, Software-as-a-Service

Digitalization is one major trend in the industry at which big data, cloud computing or artificial intelligence define the cornerstones of new business models. Today, more often than not, a myriad of software companies lack corresponding strategies to modernize their monolithic application stack due to predominant scepticism about new technologies as well as new development approaches. However, to foster a competitive and innovative market position, it is inevitable to migrate from monolithic to modern software development. The related challenges when dealing with monolithic applications are threefold, including the (1) deployment, (2) update process and (3) scaling of individual components, which make improvements in organizing, operating and scaling necessary. Some of those challenges include local installations, cross checking existing references and libraries when updating, or not knowing which piece of code is located in which part of the application. Monolithic applications tend to be huge collections of code thus requiring a lot of effort when scaling. While many software companies successfully apply modern technologies and strategies to move towards distributed systems and cloud-native applications there are still a lot of competitors following a traditional monolithic approach for software development.

Our goal is to develop a generic concept that addresses the presented challenges by leveraging three key technologies, namely (1) software containers, (2) microservices and (3) cloud-based scaling to modernize legacy applications eventually operating as a Software-as-a-Service (SaaS). While it is not necessarily required to apply these technologies and architectural design, they greatly intertwine and are beneficial for the overall transformation process. The three stages of the concept that evolve around the aforementioned cornerstones are described in the following as depicted in Figure 1.

The first stage uses software containers as a fundamental building block. A containerised monolith allows for the use of Container-as-a-Service, which is defined as a forth cloud computing service level that unites developing, deploying, updating and running applications (Burns, 2015). The containerisation process of the application is exemplified using Docker¹, demonstrating that a monolith can be dockerised without further changes to its architecture, as shown in Figure 1.

In the second stage, we introduce microservices to the concept. With microservices there is the need to introduce organizational changes to a development process, by using concepts like agile and lean development (Anderson, 2003) such as an iterative SCRUM (Schwaber, 2004) process instead of the commonly known waterfall approach (Bell & Thayer, 1976).

¹ <https://www.docker.com/>

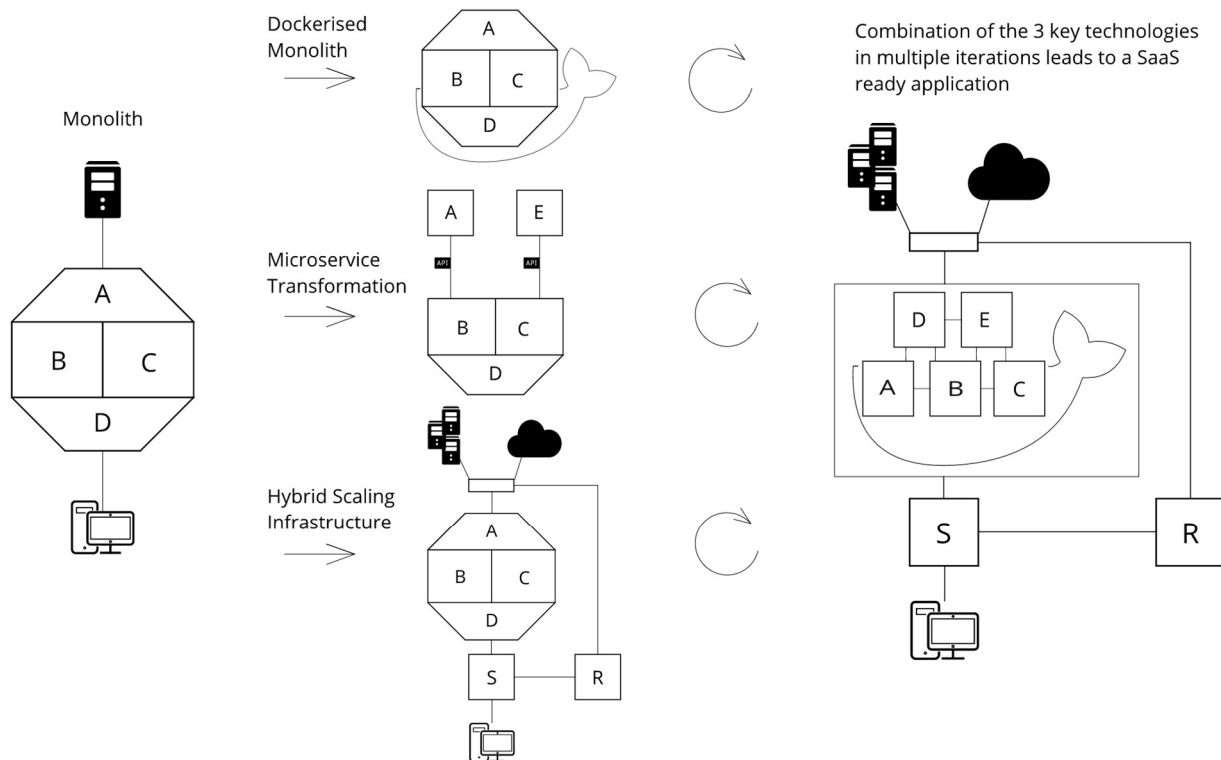


Figure 1: Approach to migrate and scale legacy applications using software containers, microservices and cloud computing.

Three different approaches can be used to transform an application's architecture into multiple microservices. The first approach is based on the law of holes meaning that if in an untenable position, it is best to stop carrying on and exacerbating the situation (Healy, 2013). Thus, new functionalities and features are attached to the monolith as additional microservices by using an API, as shown in Figure 1 by microservice “E”. The second approach separates front- and backend, but is only applicable to some special use cases, which is why this approach is not further discussed. The most common approach can be called “extract don't expand”. It combines the first approach by gradually extracting existing services from the monolith and attaching them as microservices, as shown in Figure 1 by extracting functionality “A” from the monolith and re-attaching as a microservice. To decide on the order of extraction, we developed a prioritisation matrix.

The third stage centres around scaling an application in a hybrid or entirely cloud-based environment. We evaluate existing scaling mechanisms, which are used in practice and are currently under development in research (Xiao et al., 2014; AWS, 2017; Scalr, 2017). One example is the prediction-based scaling mechanism Scryer, developed by Netflix (Jacobson et al., 2013). We leverage different scaling mechanisms to design a holistic approach to scale applications including (1) decision-based scaling, (2) reactive scaling and (3) predictive scaling element, as represented by the scaling engine “S” in Figure 1. The decision-based scaling enables the user to scale an application based on guidelines across a hybrid environment. The reactive scaling element scales an application based on metrics of the current workload. The predictive scaling element uses past metrics to scale an application in advance of the arising workload. An automatic scaling mechanism further requires a routing service, which is represented by the element “R” in Figure 1.

Eventually, the three key technologies can be combined to easily migrate into a SaaS-model. Any combination showed as valid and brings significant benefits to deploying, updating and scaling an application. For example, it is more efficient to scale the service, on which the actual workload appears, instead of scaling the entire monolithic application. Containerised microservices make this scaling even faster and allow for a clear separation between services and the monolith. The final decision is about deploying the newly formed application with in a single or multi-tenant architecture. With multiple iterations of all three key technologies it is possible to successfully transform a monolithic application into a SaaS-ready application, as shown in the final architecture in Figure 1.

We conducted a proof of concept in cooperation with a local software company to modernize their monolithic application design. The monolith includes a desktop and a web application, while the company's long-term goal is to develop an independent web application. At first the monolith was containerised and could be deployed as a Docker container. In a first step, the web monolith was extracted, which was consequently missing some functionalities from the desktop application. We identified the needed modules in the desktop application and prioritised them for extraction according to the proposed prioritisation matrix. We evaluated a common use case, in which the traffic will randomly exceed the applications capacities. The usual scaling approach was to manually start new instances. We showed how to displace this approach with an automatic scaling mechanism. A second use case features the seasonal customer base, which uses the application mainly for 3 months of the year. In addition to that the main traffic is generated at two fixed times of the day. A predictive scaling mechanism can greatly improve this use case. As a forth step we stated how to combine all technologies and how to deploy the application in a SaaS-model, using a single-tenant architecture.

This work is not only a generic conceptual guidance for companies to modernize their legacy software but further proves that the new technologies and development approaches can be used to cope with the presented challenges. Despite the initial success, the concept still needs some further testing and adjustment. Hence, supplementary research will focus on deploying the concept to a wide range of use cases and refining the individual stages.

References

- Anderson, David J. (2003). *Agile Management for Software Engineering: Applying the Theory of Constraints for Business Results*. Pearson Education.
- AWS (2017). *Auto Scaling Documentation*. Available at <https://aws.amazon.com/de/documentation/autoscaling/> [24.08.2017].
- Bell, T. E. & Thayer, T. A. (1976). Software requirements: Are they really a problem?, *Proc. 2nd Int. Conf. on Software engineering*. IEEE Computer Society Press.
- Burns, Brendan. (2015). Brendan Burns, *Google: An Overview of Kubernetes*. Available at <https://soundcloud.com/thenewstackmakers/brent-butler-of-google-gives-an-overview-of-kubernetes> [24.08.2017].
- Healy, Dennis (1996). *Healy's First Law of Holes is to stop digging*. Available at <https://business.highbeam.com/794/article-1G1-19124747/healey-first-law-holes-stop-digging-so-he-believes> [24.08.2017].
- Jacobson, D., Yuan, D. & Joshi, N. (2013). *Scryer: Netflix's Predictive Auto Scaling Engine*. The Netflix Tech Blog. Available at <https://medium.com/netflix-techblog/scryer-netflixs-predictive-auto-scaling-engine-a3f8fc922270> [24.08.2017].
- Scalr (2017). *Documentation*. Available at <https://scalr-wiki.atlassian.net/wiki/> [24.08.2017].
- Schwaber, K. (2004). *Agile Project Management with Scrum*. Microsoft Press. Redmond, Washington.
- Xiao, Z., Chen, Q. & Luo, H. (2014). Automatic Scaling of Internet Applications for Cloud Computing Services, in: *IEEE Transactions on Computers*, vol. 63, no. 5, pp. 1111-1123.

Linked Open Data for Organizational Knowledge Bases Towards a Linked Data Wiki

Matthias Frank
FZI Research Center for Information Technology
Haid-und-Neu-Str. 10-14
76131 Karlsruhe, Germany
e-mail: frank@fzi.de

Stefan Zander
Hochschule Darmstadt - Fachbereich Informatik
Schoefferstrasse 8B
64295 Darmstadt, Germany
e-mail: stefan.zander@h-da.de

Abstract. Although approaches for semantic Wikis are used in organizations for more than one decade, the integration and usage of Linked Open Data in organizational knowledge bases is still not widely adopted. This leads to unnecessary redundant work when establishing new organizational knowledge bases, as even the facts about well-known and well described entities have to be collected and maintained in each new context. In order to improve the process of building up new organizational knowledge bases, we provide an overview of current Linked Open Data approaches with respect to the challenges in organizational context and discuss their advances and limitations. In addition, we discuss our idea of an approach for a Linked Data wiki for organizational knowledge bases.

Keywords: Linked Open Data, Semantic Web, Wiki systems, Knowledge Engineering.

1. Motivation

One decade after the first semantic wiki approaches have been introduced within semantic Web-based research projects, more and more organizations start to use semantic wikis in a corporate environment. Especially Semantic MediaWiki (SMW), an extension for the popular MediaWiki engine, is deployed in many corporate settings. SMW enables semantic query construction with respect to organization-specific vocabularies and contexts and allows to incorporate query results as first-class page content. Organizations like enterprises, NGOs or civil services can benefit from such approaches. However, although existing semantic wiki approaches like Semantic MediaWiki, Ontowiki or even Wikibase make use of established semantic Web technologies on a syntactical and lightweight semantic level¹, they hardly support a common schema knowledge on a semantic level. Rather than supporting the user in reusing established vocabularies for a common terminology, the Tbox of ontologies, and linking to existing entities in the Abox, current semantic wiki approaches focus on building organization specific ontologies, which do not benefit from the growing amount of Linked Open Data (Bizer et al, 2009; Heath and Bizer, 2011). Exploiting knowledge from external sources like other organizations or Linked Open Data as well as sharing knowledge in a meaningful way is hard due to the lack of a common vocabulary and linkage in these approaches.

The present approach aims to overcome the limitation of a common schema knowledge or aligned ontology in organizational knowledge bases by supporting the annotation of organization-specific schema knowledge with the common and well-established terminology of Linked Open Vocabularies, hence extending the schema knowledge by interlinking modelled entities with Linked Open Data entities. Based on the resulting extended and interlinked schema knowledge, additional information can be provided for given entities that helps users in maintaining attributes of individuals described in the wiki system. As a consequence, our work elaborates around the following research questions:

¹ Most of these systems use a knowledge representation system, which is a subset of the decidable and well-understood *SROIQ(D)* description logic upon which the Web Ontology Language (OWL) is built.

- **RQ1:** How can we assist users of organizational wikis in establishing meaningful (i.e., semantic) relations² to Linked Open Data entities?
- **RQ2:** How can we keep track of the provenance of statements in an organizational wiki, especially if these statements are inferred or gathered from Linked Open Data?
- **RQ3:** How can we evaluate and interpret potential uncertain, incomplete, inconsistent or redundant Linked Open Data correctly in order to increase the informative value of an organizational knowledge base?

We hypothesise that the information value of organizational knowledge bases will increase with the integration of Linked Open Data. For the evaluation of this hypothesis, we test our approach with the existing organizational wiki of our research group and compare the information derived from Linked Open Data with already verified information. Based on that evaluation, we calculate the rate of correctly derived information in relation to the false derivations. If this relation is better than the relation of our existing wiki, we regard our hypothesis as confirmed for this specific use case. The hypothesis testing is subject of future work

2. Related Work

This chapter covers related work on semantic wiki software and Linked Data management. Both parts are essential for our approach. A short discussion at the end of this chapter lays the foundation for our approach towards an integrated Linked Data wiki.

2.1. Semantic Wiki Software

Software applications for creating semantic wikis already exist; one of the best-known application is Semantic MediaWiki (Krötsch et al., 2006). As many other wiki approaches, SMW is based on the MediaWiki engine, the technical base for Wikipedia. The latest release³ of Semantic MediaWiki supports the development of organization-specific knowledge bases and enables the querying of contained data (so-called *facts*) within the wiki in structured and well-defined ways. It is also possible to export semantically described facts to an external RDF store, which allows to use the W3C standardized query language SPARQL for extended query capabilities. More extensions for MediaWiki exist that provide better syntactical linkage of data modelled in SMW and RDF data like the Triple Store Connector (ontoprise GmbH, discontinued), the SparqlExtension⁴, or the RDFIO⁵ extension. All these approaches have in common that they provide semantic Web technology merely on a syntactical layer rather than a full-fledged integration on a semantic layer. Only the LinkedWiki⁶ extension focuses on exploiting LOD for organizational knowledge bases, which are built on top of the MediaWiki engine.

Koren (2015) presents the MediaWiki extension *Cargo*, since (according to his words) most semantic wiki applications are just used to structure and query data within an organizational wiki rather than integrating data on a semantic level or providing facilities of addressing semantic heterogeneity. The Cargo extension also provides functionality for structuring and querying data, however without employing any semantic Web technology. Cargo instead provides a wrapper around relational databases and exploits the well-established functionality of SQL including the limitation that

² Those links must be established on the basis of the model-theoretic semantics of the formal language and the logical theory upon which a schema is defined.

³ <https://github.com/SemanticMediaWiki/SemanticMediaWiki/releases>

⁴ <https://www.mediawiki.org/wiki/Extension:SparqlExtension>

⁵ <https://www.mediawiki.org/wiki/Extension:RDFIO>

⁶ <https://www.mediawiki.org/wiki/Extension:LinkedWiki>

entailment regime computations are hardly possible. So this means Cargo is not capable of exploiting the full feature set of Linked Open Data, right? If so, we should add that here.

One example for a non-MediaWiki based semantic wiki applications is OntoWiki (Auer et al., 2007). OntoWiki focuses on modelling a machine readable knowledge base without providing a knowledge presentation for human readers like free text and natural language.

Although the introduced semantic wiki software applications support semantic Web technology like the Resource Description Framework (RDF) or even the SPARQL Protocol and RDF Query Language (SPARQL) on a syntactic level, the data integration across multiple data sources still requires a lot of manual effort due to the establishment of a common data scheme on a semantic level.

2.2. Linked Data Management

Vrandečić and Krötzsch (2014) describe the collaborative data scheme in WikiData⁷ as one possible solution for a common data scheme in order to extend schema knowledge in other wikis, especially Wikipedia⁸. However, this approach also defines a data schema that is independent from Linked Open Vocabularies (LOV)⁹. In contrast to the WikiData approach, the OpenAnno approach (Frank and Zander, 2016) focuses on mapping individually created ontologies to LOV in order to support the interlinkage of local knowledge bases with existing Linked Open Data sources¹⁰ in a semi-automated fashion. In recent years many other approaches for Linked Data management, some of them are introduced in the following, especially with respect to our research questions.

The Versioning and Evolution Framework for RDF Knowledge Bases (Auer and Herre 2006) aims on RQ1 with a compatibility concept between ontologies and provides an assistant for changes which involves the user in the decision whether or not to accept a change. Regarding RQ2, the authors use ontology versioning to keep track of different versions of an ontology and provide the possibility to allow branching and merging operations. For RQ3, the approach is based on atomic changes like additions or deletions of statements to or from an RDF graph which are aggregated to a hierarchy of changes and facilitate the human reviewing process on various levels of detail. The changes can be annotated with meta-information and classified as ontology evolution patterns. The advantage of this approach is that it is similar to well-known versioning approaches as they are widely used in software development like the popular GIT-system. However, the work on this framework has been discontinued in favour of OntoWiki.

The SoftWiki approach (Auer et al 2007) provides semantic wiki representations for building an enterprise knowledge base. SoftWiki enables users to create, enrich, and manage defined requirements as demanded in RQ1. It provides web-based accessibility for ease of use. No installation is required on the user side and collaborators can be invited through a weblink. Provenance information as needed for RQ2 are not implemented in SoftWiki. RQ3 is supported by the traceability of changes and optional comments and discussions for every single part of the requirements engineering knowledge base. The advantage of the SoftWiki approach is that it has already been applied to a real business context. However, the approach is still on an early stage and further evaluation is needed. Especially the cloud based approach may not fit the security policies of organizations.

⁷ <https://www.wikidata.org/>

⁸ <https://www.wikipedia.org/>

⁹ <http://lov.okfn.org/dataset/lov>

¹⁰ <https://www.w3.org/standards/semanticweb/data>

The Linked Data Washing Machine approach (Auer 2011) aims on creating knowledge out of interlinked data. Our RQ1 is addressed by adaptive user interfaces and interaction paradigms to empower users to formulate expressive queries for exploiting the rich structure of linked data. Users are able to give feedback on the automatically obtained suggestions in order to improve them. User interaction has to preserve privacy, ensure provenance, and be regulated using access control. Authoring tools should hide technicalities of the RDF, RDFS, or OWL data models and assist the user through what-you-see-is-what-you-get (WYSIWYG). Regarding RQ2, different information structures need to be seamlessly combinable in a provenance preserving way in a single visualization or authoring environment even if the information to be visualised or authored is obtained or stored in various linked data sources. Addressing RQ3, the authors investigate unsupervised and supervised machine learning techniques to enable knowledge base maintainers to produce high quality mappings. They also use a semi-automatic repair method to increase the quality of Linked Data. Users have to be enabled to effortlessly give feedback to improve quality of Linked Data. Tools and services should be deployed to classify and interlink datasets automatically, to assess their information quality, and suggest enrichments and repairs to the published datasets. The advantage of the Linked Data Washing Machine is the integrative approach which combines the individual challenges rather than regarding them isolated. However, this approach still lacks on practicality of the discussed solution and remains on a theoretical stage.

The what-you-see-is-what-you-mean (WYSIWYM) approach aims on authoring of structured content based on Schema.org (Khalili and Auer, 2013). The authors describe the manual composition process aiming at the creation of documents which use semantic knowledge representation formalism as needed for RQ1. The manual composition is supported by a graphical user interface. Regarding RQ2, the work does not focus on provenance, origin or source of LOD. For RQ3, the approach provides a set of quality attributes for semantic content authoring (SCA) systems with corresponding user interfaces for their realization. Those include usability, automation, generalizability, collaboration, customizability, and evaluability. The paper provides a consolidated literature review of existing approaches including in-depth review of four SCA-systems.

The crowdsourcing quality assessment methodology (Acosta et. Al., 2013) aims on improving the quality of Linked Data in general. Although this approach does not contribute to RQ1 or RQ2 since it focuses on quality improvement of existing LOD in DBpedia only, rather than supporting the interlinkage of new organizational specific entities, the approach provides a valuable contribution regarding RQ3. Using a list of data quality dimensions (criteria). The authors discuss four quality dimensions: Accuracy, relevancy, representational consistency, and interlinking.

OntoWiki (Frischmuth et. al., 2015) follows a different approach by providing an authoring, publication and visualization interface for the Data Web. For RQ1, OntoWiki supports navigation through RDF knowledge bases using SPARQL-generated lists, tables and trees. However, the authors do not mention the support for creating new links as they focus on accessing existing links only. Regarding RQ2, their RDFauthor approach builds on RDFa by preserving provenance information in RDFa representations following the named-graph paradigm and by establishing a mapping from RDFa view representations to authoring widgets. Regarding RQ3, a number of tools in addition to OntoWiki are discussed that focus on data linking, quality improvement, enrichment, evolution and visualization. The advantage of the OntoWiki approach is the comprehensive user interface for arbitrary RDF knowledge graphs. However, there is a risk of overloading the user interface with more features which may decrease the usability.

Another approach makes use of the statements of Wikidata by using DBpedia concepts in order to exploit the benefits of both approaches (Ismayilov et. al., 2015). Regarding RQ1, the approach used

the human-readable Wikipedia article identifiers to create IRIs for concepts in each Wikipedia language edition, use RDF and Named Graphs as its original data model, and provide <http://wikidata.dbpedia.org/> as a Linked Data interface and SPARQL endpoint. For RQ2, Wikidata uses language-independent numeric identifiers and developed its own data model, which provides better means for capturing provenance information. Wikidata has a smaller dataset than DBpedia but higher quality and provenance information due to manual curation. Provenance extractors can be used to export as much knowledge as possible. Extractors can get labels, aliases, descriptions, different types of sitelinks, references, statements, and qualifiers.

Rautenberg et al (Rautenberg, 2016) discuss a workflow management system for linked data processing called LODFlow. Regarding our RQ1, they use LODFlow to create and manage the execution of workflows which interact with workflow participants. In addition, they provide visual programming frontends to enable users to construct their applications as a visual graph by connecting nodes together. RQ 2 is also addressed by LODFlow as it can help to preserve provenance by adding comprehensive metadata such as the version, invocation, and configuration of the tool execution in a concrete workflow instantiation. The authors plan workflows for Linked Data datasets maintenance to enable provenance extraction and reproducibility over time. RQ3 is partly addressed by LODFlow engine to interpret the resources from the Linked Data Workflow Project Ontology and invoking other tools. They employ Luzzu for quality analysis. The advantage of LODFlow is that it is tested and applied to a large-scale real-world use case. However, the complexity of a full workflow management system for linked data aims to data scientists and cannot be used without special training. Therefore it does not fit to our intended use as an organizational wiki system which can easy be used by any employee.

2.3. Discussion of Related Work

We have shown that current semantic wiki applications provide technical integration of semantic Web technology on a syntactic level. We have also discussed initial approaches for enriching organizational knowledge bases with addition information from LOD. However, the introduced semantic wiki approaches do not support the annotation and interlinkage of organizational knowledge bases with LOD on a semantic level. Such a recommendation system is provided by OpenAnno, but it is not integrated in any of the introduced semantic wiki applications. The statements maintained by one of these semantic wiki applications cannot be updated by external services as the statements contained within a wiki are always considered as master data. When importing statements from external sources into an organizational wiki, none of the introduced semantic wiki applications consider the context or the linkage of the data. Both is important in order to evaluate given statements, especially when they are inconsistent, redundant or ambiguous.

3. The Approach

By separating the semantic statements from the non-semantic part of the wiki as illustrated in Figure 1, we avoid the issue of syncing statements between the wiki and the knowledge base. Additionally, we are able to maintain and curate the semantic statements outside of the wiki without causing inconsistent data. This separation of semantic and non-semantic data is therefore a prerequisite for the transparent integration of statements from the wiki itself and external statements from LOD.

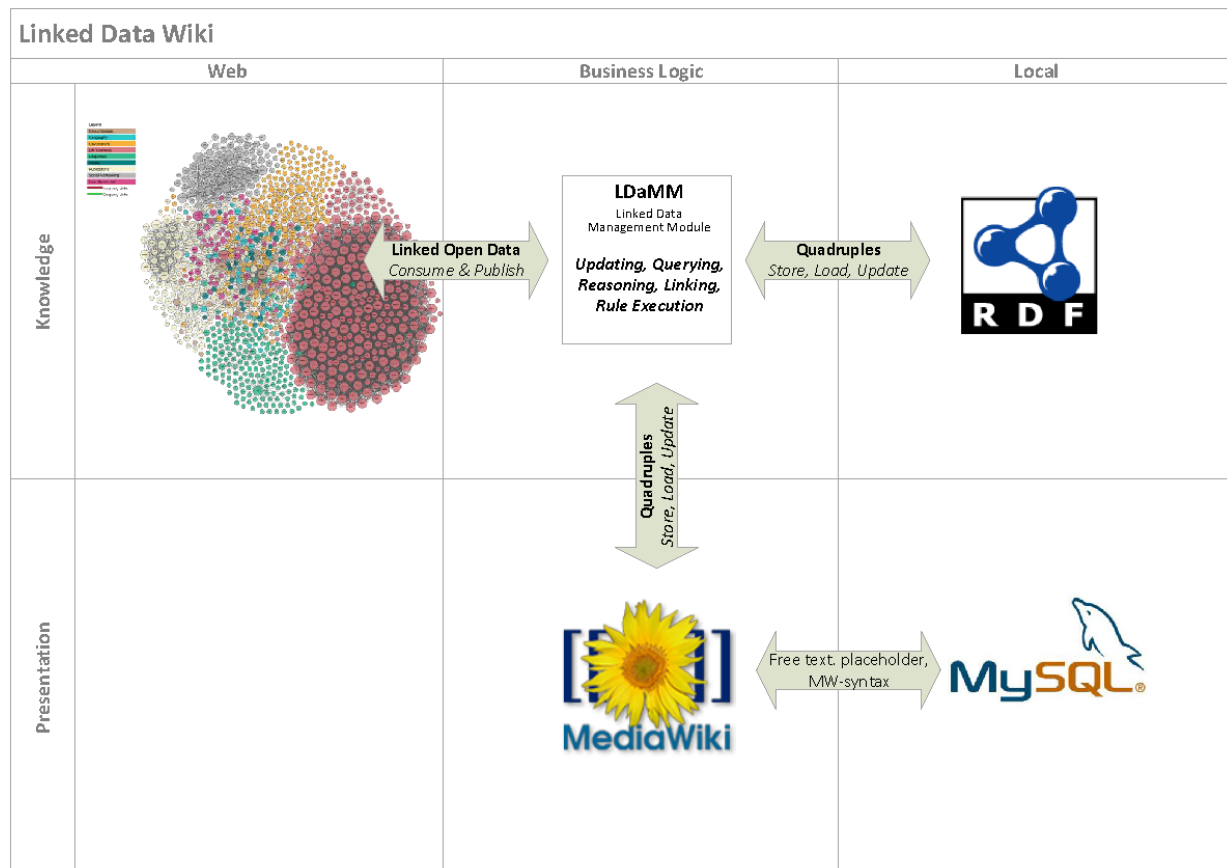


Figure 1: Architecture for out intended Linked Data Wiki system

In contrast to providing only wiki-based statements within an organizational wiki, our approach does also include external statements from multiple LOD sources. The inclusion of external statements causes issues when the same entity is described in multiple sources which increases the likelihood of incorporating potentially redundant or inconsistent data. We address this issue in our approach by exploiting gathered provenance information and evaluating the statements based on a ranking derived from contained provenance statements. The ranking is influenced by the interlinkage of the source as an indicator of reference and by the evaluation of statements by wiki users.

Another issue is the potential amount of provenance information. Although this provenance information is necessary in order to evaluate the trustworthiness of statements, it would be confusing for users to show all available provenance information for each statement. We address this challenge by evaluating the provenance information in the background and just showing the resulting statement to the user with an option to expand the underlying provenance-based derivation of the statement.

Our approach is technically based on MediaWiki in combination with an RDF-store. The contribution of our approach is therefore not on a technical layer, but aims to support the schema integration on a semantic layer. We provide a set of established Linked Open Vocabulary (LOV) to encourage the reuse of these vocabularies in organizational wikis. The resulting organizational knowledge base using LOV is the foundation for suggestions of annotations from Linked Open Data (LOD). These annotations allow to enrich the organizational knowledge base with additional information from LOD. In order to distinguish organization specific statements from statements gathered from LOD, we track the provenance information of each statement. The provenance information is stored with-

in *named graphs*¹¹ in the RDF store, which extends the statements' triple-model consisting of subject, predicate and object to quadruples, containing an ID for each statement. This ID, which is a graph IRI, allows us to attach provenance information to each statement. Using the provenance information, we can also handle uncertain or inconsistent data and provide the data consumers with the latest and most probable information.

One characteristic of our approach is the strict separation of statements in the triple store maintained by our extension and the non-semantic part of the wiki, like free text, MediaWiki syntax and placeholders for the semantic statements, which are still maintained by the MediaWiki engine.

4. Conclusion

At the stage of our work discussed in this paper, we have developed the concept to assist users of organizational wikis in establishing and curating meaningful relations to Linked Open Data entities by building adequate SPARQL queries based on the user's input and the given context (RQ1). By executing queries on general entities like instances of cities, we can show that adequate LOD records exist to support the construction of organizational knowledge bases. Further research efforts have to prove that this is also the case for other knowledge domains. We keep track of the provenance of each statement retrieved from LOD by assigning unique identifiers to each statement and attach the provenance information to that ID (RQ2), which enables software agents to consider the provenance of statements for reasoning. Exploiting the gathered provenance information for semantic reasoning is a precondition to increase the informative value of an organizational knowledge base (RQ3). However, further research has to be carried out for a proper evaluation and will be provided in future articles.

References

- Acosta, A., Zaveri, A., Simperl, E., Kontokostas, D., Auer, S., Lehmann, J. (2013). "Crowdsourcing Linked Data Quality Assessment". *International Semantic Web Conference* (2), pp. 260-276.
- Auer, S., Jungmann, B. and Schönefeld, F. (2007). „Semantic Wiki Representations for Building an Enterprise Knowledge Base". *Reasoning Web 2007*. Dresden, Germany.
- Auer, S. (2011). "Creating knowledge out of interlinked data: making the web a data washing machine". *WIMS 2011*: 4
- Auer, S., Herre, H. (2006). "A Versioning and Evolution Framework for RDF Knowledge Bases". *Ershov Memorial Conference 2006*, pp. 55-69.
- Auer, S., Jungmann, B., Schönefeld, F. (2007.) "Semantic Wiki Representations for Building an Enterprise Knowledge Base". *Reasoning Web 2007*, pp. 330-333.
- Bizer, C., Heath, T. and Berners-Lee, T. (2009). "Linked Data - the story so far", *International Journal on Semantic Web and Information Systems*, 5, (3), pp. 1-22. (doi:10.4018/jswis.2009081901).
- Frischmuth, P., Martin, M., Tramp, S., Riechert, T., Auer, S. (2015). "OntoWiki - An authoring, publication and visualization interface for the Data Web". *Semantic Web* 6(3), pp. 215-240.
- Heath, T., Bizer, C. (2011). "Linked Data: Evolving the Web into a Global Data Space", in *Linked Data: Evolving the Web into a Global Data Space*, 1, Morgan & Claypool.
- Ismayilov, A., Kontokostas, D., Auer, S., Lehmann, S., Hellmann, S. (2015). "Wikidata through the Eyes of DBpedia". *CoRR abs/1507.04180*.
- Khalili, A., Auer, S. (2013). "WYSIWYM Authoring of Structured Content Based on Schema.org". In: *WISE (2) 2013*, pp. 425-438.
- Koren, Y. (2015). "Cargo and the future of Semantic MediaWiki". In: *SMWCon Spring 2015*. St. Louis, MO, USA.

¹¹ <https://www.w3.org/2004/03/trix/>

Krötzsch M., Vrandečić D. and Völkel, M (2006). “Semantic MediaWiki”. *International Semantic Web Conference 2006*. Athens, GA, USA.

Rautenberg, S., Ermilov, I., Marx, E., Auer, S., Ngonga Ngomo, A. (2015). “LODFlow: a workflow management system for linked data processing”. At: *SEMANTICS 2015*, pp. 137-144.

Vrandečić, D. and Krötzsch, M. (2014). “Wikidata: a free collaborative knowledgebase”. In: *Communications of the ACM*, Vol. 57 No. 10, pp. 78-85.

Multi-Access Edge Computing: Managing the Resources of the Edge of the Internet

H. Frank¹, W. Fuhrmann², B. Ghita³

¹ IQDoQ
Document Management Company of
MATERNA,
Germany
e-mail: harald.frank@iqdoq.de

² Department of Computer Science
Hochschule Darmstadt
University of Applied Sciences,
Germany
e-mail: woldemar.fuhrmann@h-da.de

³ Centre for Security, Communications
and Network Research
Plymouth University,
United Kingdom
e-mail: bogdan.ghita@plymouth.ac.uk

Keywords: Multi-Access Edge Computing, Black Rider, Relocation Policy

1. Introduction

A lot of new techniques for the edge of the mobile internet was developed in the last few years. These reduces the distance between mobile devices and the applications in the cloud-infrastructure. In order to optimize the controlling of the distributed components at the different locations a centralised logical unit is needed. This paper proposes a new controlling unit between the mobile transport network and the cloud-infrastructure, called Mobile Edge Computing (MEC) Manager, which reuse the concept of the Black Rider (BR), which is context- and policy-based itself (Frei *et al.*, 2013). This centralised processing unit bundled information from the mobile transport network as well as from the cloud-infrastructure to optimize the distribution of the applications at the edge of the mobile internet. The paper provides an architectural overview.

2. Related Work

In order to define and understand the concept of application mobility, this section will provide an overview of the state of the art in the area of Multi-access edge computing (MAEC) and the relevant functional concepts used by it.

The European Telecommunications Standards Institute (ETSI), as part of its work on, mobile radio, converged, broadcast and internet technologies, produced a series of six specifications in relation to MAEC (ETSI Consortium, 2016). Other specifications are under study in the consortium.

The ETSI approach will be used as a reference architecture and the basis for the novel extensions proposed in this research. There are also proposals in other research papers, as for example Cloud RAN (Pompili *et al.*, 2015). However, the ETSI architecture seems to be the most consolidated architecture for the execution of applications in the close proximity to end users (Tran *et al.*, 2017).

The ETSI standard leaves the determination mechanism to select the concrete mobile edge host unspecified. The ETSI standard recommendations vary depending on the implementation, configuration and operator deployment. Therefore, a general mechanism is missing, which control the optimal selection of the mobile edge host and support the relocation of the applications on the mobile edge hosts from the user equipment point of view.

In addition, relocation and coordination between application and transport are not covered by the ETSI architecture. Especially in a small cell network the range of each individual cell is limited. Mobility support becomes more important, and a solution for fast process migration may become necessary. Ensuring fair resource sharing and load balancing is also an essential problem. Both aspects are described as open research issues in the literature (Tran *et al.*, 2017).

Therefore, an analysis about the generic virtualisation and mobility context, the limitations of the ETSI standard and novel enhancements resulting in more efficient mobility is to be explored by the research project.

3. Research Design

The research design consist of the following steps. After finalising of a first literature review, categorising of necessary technologies and defining the requirements for the new mechanism, the outline of the architecture with the necessary components, the refining of the architecture, the creating of a simulation-system and finally the evaluation will be conducted. This order was chosen to ensure the quality and reliability of the research progress. The definition and refining of the components of the novel architecture is necessary to create a simulation-system. This will be reached with appropriate use cases. The simulation-system is useful for the evaluation of the findings.

Currently the research works are in the face of refining the architecture. After that step the creating of the simulation-system is intended to be start.

4. Solution Approach

The idea is to create an additional mechanism to collect the information and provide the results to other network components. BR provides a similar concept – collecting information about the mobile device and delivering it to other network components (Frei *et al.*, 2013). However, the BR architecture does not provide any mechanism for collecting information on the actual infrastructure, including cloudlets (Frank *et al.*, 2016).

A first coarse architecture concept containing the entity BR and the new entity MEC Manager for the mobility management of the applications in the MAEC environment is shown in Figure 1.

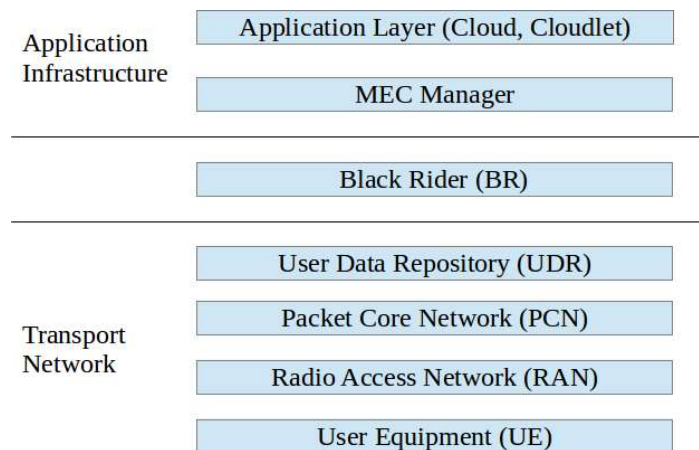


Figure 1: Reference architecture of the MEC Manager

On top there is the application layer including the Cloud and Cloudlets infrastructure. The MEC Manager gets information on the mobility management context and bearer management context of the mobile users from the BR, for example by subscribing appropriate user events, and can make its own decisions on application mobility management within the application infrastructure. The User Data Repository (UDR) stores all user data, except user content and user context data, which are required for transport network control. The transport network consists of user equipment, radio base

stations of the radio access network and the packet core network. The BR monitors the mobility behaviour of the mobile user in the transport network and makes optimized decisions on the traffic distribution in the transport network. Based on these decisions in the transport network, the MEC Manager uses appropriate mobility management and bearer management information from the BR and his own collected information from the cloudlet infrastructure to optimize the behaviour of the Application Layer, like movement of virtual machines (VM) between cloudlets in order to improve the quality of experience (QoE) for the mobile user.

Thereby the MEC Manager can execute context- and policy-based decisions to control the components of the cloud infrastructure. By using the MEC Manager important information about the network-infrastructure and their conditions are available at a central point in the network. As a result, new opportunities for the optimization for other network components are available. In the final architecture, parts of the BR architecture can be reused. Therefore, a common use of the database of the BR is conceivable.

5. Conclusion and Outlook

In this paper it's shown how to extend the ETSI standard for enabling MAEC by employing cloudlets and the concept of black rider. Through the use of the concept called Black Rider it is principally possible to establish a central management instance here called MEC Manager that enables the distribution and management of mobile applications in the cloud infrastructure.

As a next step a first control architecture is developed based on existing blocks as cloudlets and BR. Furthermore, interfaces between different providers have be defined so that VMs can be transported in cloudlets over provider boundaries.

References

- ETSI Consortium (2016), "Mobile Edge Computing", available at: <http://www.etsi.org/technologies-clusters/technologies/mobile-edge-computing>.
- Frank, H., Fuhrmann, W. and Ghita, B. (2016), "Mobile Edge Computing: Requirements for Powerful Mobile Near Real-Time Applications", in Paul Dowland et al. (Ed.), *Proceedings of the Eleventh International Network Conference (INC 2016)*, Plymouth University, Plymouth, pp. 63–66.
- Frei, S., Fuhrmann, W. and Ghita, B. (2013), "Generic Real-Time Traffic Distribution Framework: Black Rider", Nassau, Bahamas.
- Pompili, D., Hajisami, A. and Viswanathan, H. (2015), "Dynamic provisioning and allocation in Cloud Radio Access Networks (C-RANs)", *Ad Hoc Networks*, No. 30, pp. 128–143.
- Tran, T.X., Hajisami, A., Pandey, P. and Pompili, D. (2017), "Collaborative Mobile Edge Computing in 5G Networks: New Paradigms, Scenarios, and Challenges", *IEEE Communications Magazine*, pp. 54–61.

Interactive guidance and real-time fault recognition for manual operations through a stationary AR system and visual object detection

Dipl. Medieninf. Tom Lutze, Jonas Hansert B.Sc., Dipl.-Inform. Björn Lunt, Prof. Dr.-Ing Thomas Schlegel
Institute of Ubiquitous Mobility Systems
University of Applied Sciences Karlsruhe, Germany
e-mail: tom.lutze@hs-karlsruhe.de | jonas.hansert@hs-karlsruhe.de

Keywords: hand detection, static augmented-reality, guided manual workflows

Introduction and Background

In the field of manual assembly and operations in industrial manufacturing processes, occasional mistakes are unavoidable due to human error which can lead to high costs for companies, since they can also have an impact on further process steps. By monitoring and analyzing manual process steps in real-time for possible errors, interactive augmented-reality systems can immediately inform the worker of any mistakes and propose actions for correction directly in the context. Since many activities in assembly are tied to a fixed workplace and must be carried out there, it is in this cases much more comfortable for the employee to obtain AR-based information over a stationary system without the need to carry additional interfering equipment.

State of the art

Several approaches have been proposed for manual assembly guidance using augmented reality with head-mounted-displays (HMD) (Yuan et al. 2008) or in situ projection which is well suited for manual task assistance and compared to HMD, tablets or paper it is faster and less error-prone. In addition, there is no need for wearing a device, preserving the workers comfort and freedom of movement (Funk, Kosch and Schmidt, 2016). Projection-based AR provides a bigger field of view than currently available AR-glasses and is much more robust with respect to changing light conditions (Sand et al. 2016). Multimodal natural user interfaces like speech and gesture input are used in combination with a HMD by Siltanen et al. (2007) to guide users solving a 3D puzzle. Henderson and Feiner (2011) developed a research prototype AR application in the field of maintenance of military vehicles and found out, that AR can reduce the required time for task location as well as the need for head and neck movements. AR can also improve the task efficiency by guiding technicians via virtual models, animations and graphical symbols (De Crescenzo et al. 2011). The mentioned works focused specific aspects of our target system and confirmed the possible benefits for the employees in the corresponding application environments. Our implementation is intended to offer a considerably higher added value for the user by the additional fault detection and real-time feedback through AR-based instruction steps, closing the loop.

Approach and Methodology

Our approach consists of four components (Figure 1): the detection system with hand and object recognition, the analysis and fault detection, the feedback system giving a visual and acoustic feedback about the next steps and on top of them the knowledge base containing and delivering all static information needed by the other components. The stationary AR-system should be based on consumer technology. For this purpose, we use conventional webcams, the Microsoft Kinect technology and projectors. On the software side, we use OpenCV for image processing and adapt it accordingly to our system.

One essential part of our system is to know the position of the hands and other objects on the workbench. To detect them we use a Microsoft Kinect 2.0 system. For hand and object detection we use only the RGB image in a first step. We read the RGB image from the camera and convert it to

HSV color model. Then we use a filter to get only the pixels with a hue and saturation in space of human skin color. We smoothen the image with a combination of dilatation and erosion filters. Based on this image we search the contours of every connected area. Contours smaller than a threshold are discarded because the probability that it is an error and not a hand is very high. For remaining contours, we compute the convex hull and based on this a smallest rectangle covering the hand. Additionally we calculate the center of mass from the contours. This method has a very high accuracy and a high robustness against changes in lighting and background.

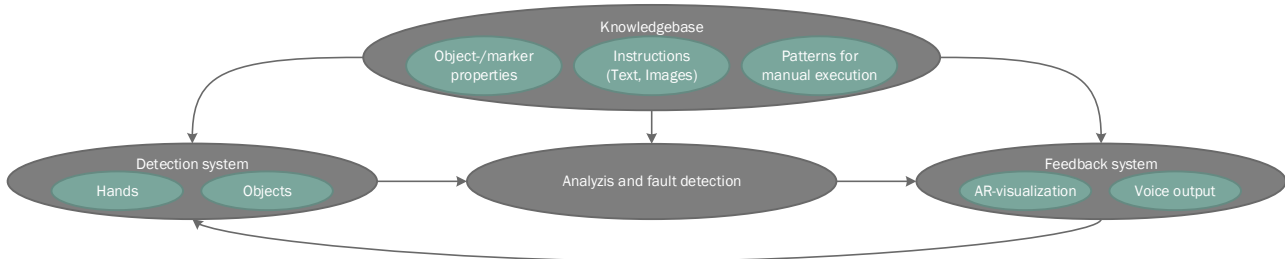


Figure 1: Our system architecture, consisting of the three basic components. The hand- and object detection system, analysis and fault detection and the feedback system. The knowledgebase provides all necessary information to the three basic components.

To compute the position of the hands and objects on the table, it is necessary to have a transformation between the coordinate systems of camera, projector and table. The registration procedure has only to be performed after changing the relative positions between camera, projector and table. We use ArUco markers (Garrido-Jurado et al. 2014) and detect them in the camera image. At first, we project four markers at four predefined positions in the projectors coordinate system on the worktop and detect them in the camera image (Figure 2). We assume a planar worktop and can calculate the perspective transformation between the two coordinate systems, using the positions of the four markers in both coordinate systems. We do the same with printed markers placed on the four edges

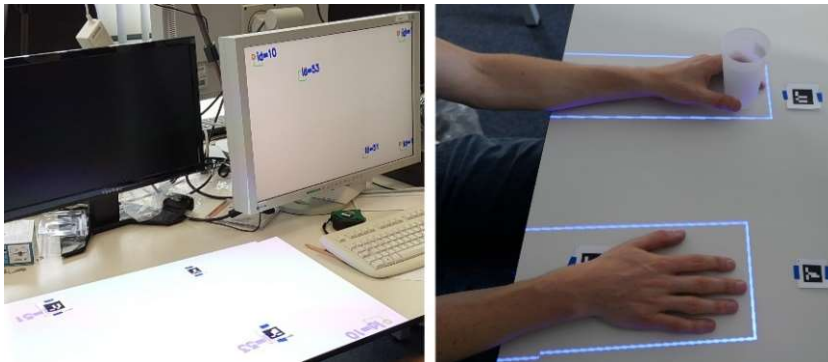


Figure 2: The left picture shows the process of referencing the real and virtual coordinate systems via markers. The picture on the right shows the detected hands which is visualized by a blue frame.

of the worktop and define a coordinate system with root in one corner of the worktop, axes along the edges and the distance to the other corners in millimeter. As result we have the transformation between the projector and the camera as well as between the table and the camera. Then we can calculate the transformation between worktop and projector by a multiplication of the inverse of the first matrix with the second.

As result we can transform the position of the hands and other objects detected on the worktop to the worktop coordinate system, combine this with the information about static objects like boxes for the components and create an image that shows the next step. We transform this image to the projector coordinate system and show it with the projector on the worktop.

Test scenario and setup

We have developed a scenario in which an employee performs manual activities at a demonstration plant for automatic filling of beverages. In the scenario of our demonstrator a customer selects a cup

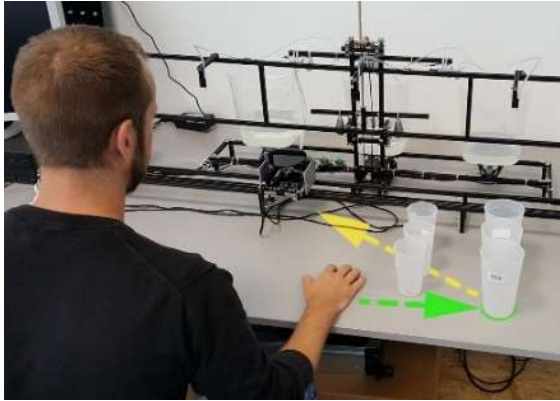


Figure 3: The conceptual implementation of AR-based work instructions with our filling machine demonstrator.

size and if he wants a cap on it. The worker has to take the correct sized cup and place it on the object holder. The filling line fills the chosen drink in it and, if wished, the worker has to put the correct cap on top. To assist the worker, at first our detector system has to detect the position of all the cups in different sizes and the position of the hands. The caps are placed in boxes with a fixed position on the worktop, known by the knowledgebase. Then the analysis and fault detection part of our system uses the information from the detector system and knowledgebase and determines the next instruction together with the selection of the customer. The feedback system visualizes this instruction. In this case, it generates an arrow from the right hand to the cup and from the cup to the object

holder. Then it transforms this image to the projector coordinate system and shows it on the worktop. Now we complete the loop and detect the positions of hands and objects a second time. The instructions will be the same until the cup is placed to the object holder and the augmented reality visualization is adapted to the new position of the hands and cups.

When the cup is placed in the object holder the filling line starts to mix the drink. While this happens, there can be displayed instructions to prepare everything for the next step. In our example, an arrow from the right hand to the correct sized cap is projected. When the filling line has finished and the cap is grasped, the next instruction is an arrow from both hands to the cup: From the left hand with a symbol visualizing to grasp the cup and from the right hand a symbol visualizing to put the cap on the cup. The last step is visualized by an arrow with a symbol that shows where the filled cup must be placed. With this scenario and setup, we can check how reliably the detection and projection works in the practice of manual assembly. It is particularly interesting how reliable the system works with fast hand movements, different hand and finger positions as well as with objects held in the hands. Furthermore we want to evaluate how the information has to be presented and how well different methods are suitable for the employees to issue recommendations for action.

Conclusion and outlook

In the same way of our simplified scenario we can give instructions for more complex tasks. Up to now, the employees need to look in a printed manual or use a computer, maybe with dirty hands. Our system detects the situation on the worktop automatically so there is no explicit interaction from the user needed. This fact makes our system very easy to use, reduces faults and allows to work as efficient as possible because it adapts to the speed of the employee and always gives the next instruction exactly when it is needed.

Our system for hand and object detection works very well. In some cases, it might be advantageous to have additional information like position of fingertips or wrist. We could use this positions for a better fault detection, more detailed guidance and more intuitive gestures. These positions can be extracted with a convolution neural network described by Tompson et al. 2014. We currently working on an implementation of this convolutional neural network and further topics like workflow integration.

References

- Yuan, M. L., Ong, S. K., & Nee, A. Y. C. (2008). Augmented reality for assembly guidance using a virtual interactive tool. *International Journal of Production Research*, 46(7), 1745–1767.
- Funk, M., Kosch, T., & Schmidt, A. (2016). Interactive Worker Assistance: Comparing the Effects of In-Situ Projection, Head-Mounted Displays, Tablet, and Paper Instructions. *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 0–5.
- Henderson, S., & Feiner, S. (2011). Exploring the benefits of augmented reality documentation for maintenance and repair. *IEEE Transactions on Visualization and Computer Graphics*, 17(10), 1355–1368.
- De Crescenzo, F., Fantini, M., Persiani, F., Di Stefano, L., Azzari, P., & Salti, S. (2011). Augmented reality for aircraft maintenance training and operations support. *IEEE Computer Graphics and Applications*, 31(1), 96–101.
- Sand, O., Sebastian, B., Paelke, V., & Carsten, R. (2016). *Virtual, Augmented and Mixed Reality* (Vol. 9740).
- Siltanen, S., Hakkarainen, M., Korkalo, O., Salonen, T., Saaski, J., Woodward, C., Kannetis, T., Perakakis, M. and Potamianos, A. (2007). Multimodal User Interface for Augmented Assembly. *2007 IEEE 9th Workshop on Multimedia Signal Processing*.
- Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F.J., Marín-Jiménez, M.J. (2014). Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition, Volume 47, Issue 6, Pages 2280-2292, ISSN 0031-3203*
- Tompson, J., Stein, M., LeCun, Y., Perlin, K. (2014). Real-Time Continuous Pose Recovery of Human Hands Using Convolutional Networks. *ACM Trans. Graph.* 33, 5, Article 169 (September 2014).

Evaluation of machine learning algorithms for Smurf Detection

Raphael Hippe, Johannes Dornheim, Samuel Zeitvogel, Astrid Laubenheimer
Intelligent Systems Research Group (ISRG)
Karlsruhe University of Applied Sciences, Germany
e-mail: info@raphaelhippe.de

Keywords: game analytics, data preparation, machine learning

Introduction

In amateur eSports—competitive gaming—detecting *smurfs* is a real issue for tournament organizers, like CompeteLeague. A player is *smurfing* when a secondary account is used to hide his or her true identity. In this paper, results of a two-layer model with the purpose of detecting smurf accounts for the game League of Legends (LoL), based on various machine learning algorithms, are presented.

Related Work

In *Player Skill Modeling in Starcraft II* (Avontturr et al. 2013) an approach for player skill level prediction is presented. The classification model is based on information tracked by the game such as *actions per minute* or *effective actions per minute* and divides the players into 7 leagues according to the Starcraft II ranking. The presented model reaches an average accuracy of 44.9% and an average misclassification distance of 1.55 Leagues.

Yan and Randell (2005) created a definition of *cheating*: “Any behavior that a player uses to gain an advantage over his peer players or achieve a target in an online game is cheating if, according to the game rules or at the discretion of the game operator (i.e. the game service provider, who is not necessarily the developer of the game), the advantage or the target is one that he is not supposed to have achieved”. By this definition *smurfing* would categorize as *cheating* because players who try to play against a weaker opposition violate the given ruleset. *Cheating* techniques and defense techniques are further discussed by Laurens et al (2007). The focus of both papers lies primarily on exploits of the game clients and the different variations of those exploits. To the knowledge of the authors no directly related research has been conducted to detect smurf accounts or predict a player's skill level. Therefore, the related work in a broader sense of game analytics was factored into research phase. The book *Game Analytics* (El-Nasr et al. 2016) offers good insights of background knowledge in the given field.

Further the paper *Player Behavior and Optimal Team Composition for Online Multiplayer Games* (HY Ong et al. 2015) presents research on the examined game, LoL. Here the prediction of the winning teams based on team compositions using unsupervised learning as well as classification algorithms is focused.

Scenario

Currently game-data has to be checked for all game accounts by online tournament organizers manually in order to keep their tournaments free from *smurfs*. Since this is a time-consuming and recurring task, a method which automatically provides a pre-selection of game accounts which likely are *smurfs*, is desired.

Classification Model

Two models were created to solve the *Smurf Detection* problem. Both consist of two layers, with the inner layer being the classification and the outer one a majority vote. The classification layer holds either one - or in the case of the second model five—classifiers. The second model features five classifiers, one for each position in the game, to learn the differences of the positions. The classification layer has been trained on match data, which means they classify a player based on one match. Afterwards the classification-results—20 per player—get passed onto the second layer, where a majority vote takes place in order to determine whether a player gets classified as a *smurf* account or not. Vote-thresholds can be manually set to influence the precision-recall balance based on the given application scenario.

Data

The data produced by LoL can be roughly split into two groups: Account metadata and gameplay data. While the former covers information such as time-based activity data of an account or information about accounts it often plays together with, the latter focuses on the data actually produced from game participation. A first naïve approach tested by the main author was creating account networks trying to find clusters of accounts which play together regularly with two accounts within that group not sharing a game and therefore are assumed to be used by the same person. Since this only worked in 28 out of 100 testcases, in which the main as well as the smurf account were known, the focus was led onto gameplay data. Furthermore, knowledge about the game from the main author as well as the research presented by Avonturr et al. (2013), suggests, focusing on the data produced in the early minutes of the game. This assumption was later supported by Scikit-learn's feature importance measures obtained from algorithms such as the *Random Forest Classifier*. Using that knowledge, new features were aggregated such as the *perMinuteFeature*, normalizing cumulative features, such as *goldEarned*—the amount of gold a player earned during the game—by the duration of the game. Additionally, the *performance* feature was introduced with great success. Due to the fact, that contribution towards a win in LoL differs from position to position, three variations of the *performance* feature have been created based on domain knowledge. Further, the difference of the *performance* is supposed to help with the *role lane* problem: LoL codes the information what *position* a player took in a match with two variables—*role* and *lane*. The *one-hot* method was applied in order to translate these categorical data, since the used machine learning algorithms have to use purely numeric data. Due to combinatorics, a lot of features, which together only hold one information, would be produced using this approach. A different possible angle to deal with this problem will be discussed in the *modeling* section. Building on the feature creation step, three data sets were produced: the training, validation and the test set. Since the smurf rate in CompeteLeague matches has been estimated to 10% the validation and test set have a distribution of 90% non-smurf (negative class) and 10% smurf accounts (positive class). In contrast, the training set has a balanced distribution to prevent the model from predicting only the negative class. Each set has 20 matches per player with the training set consisting of 340, the validation set of 146 and the test set of 354 players. The constant 20 matches per player has been chosen, because it was the highest number of matches per player for known smurf accounts available without going too far in the past or dropping rapidly in player numbers for the different data sets. Besides that, it is also important to notice, that the non-smurf partition is randomly picked from the big player base of LoL. Therefore, it is likely that single smurf accounts are part of the non-smurf ground-truth, since the author has no knowledge about the accounts being smurf accounts or not for this partition.

Evaluation

In Figure 1 evaluation results of the two models using different algorithms as well as different thresholds for the majority vote are presented. The constant at 0.1 precision is the established baseline, which represents the expected value for random guessing. Evaluating the model showcased that most classification algorithms such as *XGBoost*, *Random Forest* as well as the *Support Vector Machine* are very close together in terms of *precision* and *recall* for the given task. Furthermore, the model using only one classifier instead of five proofed to be slightly more precise for the given task and data. Finally, the chosen model instance achieved a precision value of 20% with a recall of 70% respectively, since the built model is planned to be used as a recommendation system. With the new model only the recommended accounts will get checked manually, instead of all accounts. Therefore, *recall* has been valued higher than *precision* and aggregated evaluation metrics such as the F1-score have not been used.

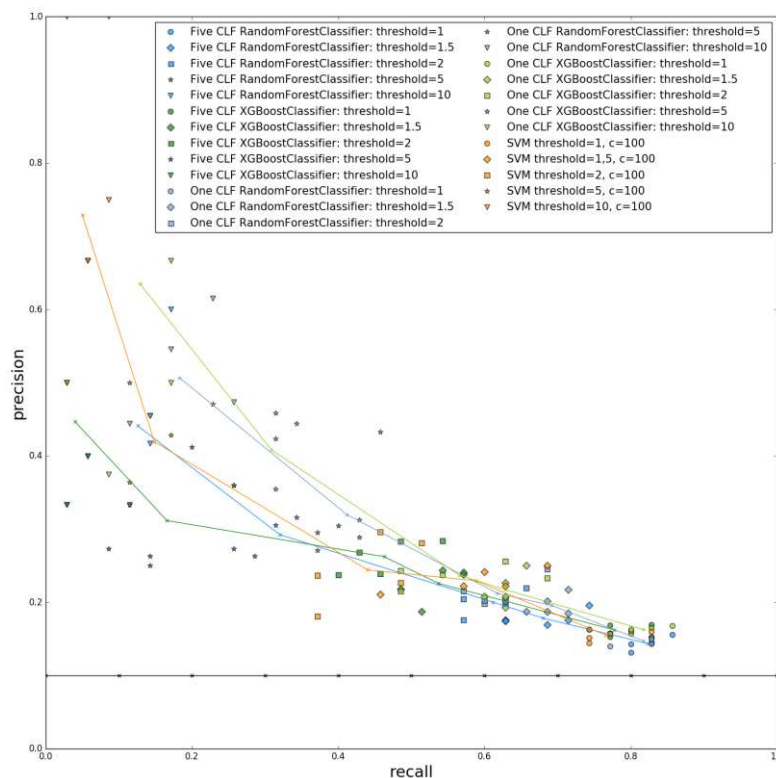


Figure 1: Result comparison of the two models with different classification algorithms and thresholds

Future work

Future work contains steps such as keeping the data sets up to date, because the game gets constantly updated. Further, more and primarily cleaner data has to be gathered in order to be able to train more advanced and complex models and thereby improve the results.

Conclusion

A first solution of the *Smurf Detection* Problem—which often occurs in amateur eSports tournaments—has been found using classification algorithms such as *Random Forest* or *XGBoost*. A model surrounding these algorithms has been build predicting 20 matches per player followed by a majority vote to determine the class of set account. The introduced solution achieved a precision of 20% with a recall of 70% respectively.

This work was conducted in the framework of a bachelor thesis at Karlsruhe University of Applied Sciences—Intelligent Systems Research Group in cooperation with CompeteLeague.

References

- Avontuur, T., Spronck, P. and Van Zaanen, M. (2013). *Player Skill Modeling in Starcraft II*. In *AIIDE*, October.
- El-Nasr, M.S., Drachen, A. and Canossa, A. (2016). *Game analytics*. Springer London Limited.
- Laurens, P., Paige, R.F., Brooke, P.J. and Chivers, H. (2007). A novel approach to the detection of cheating in multiplayer online games. In: *12th IEEE International Conference on Engineering Complex Computer Systems*. pp. 97-106.
- Ong, H.Y., Deolalikar, S. and Peng, M. (2015). *Player Behavior and Optimal Team Composition for Online Multiplayer Games*. arXiv preprint arXiv:1503.02230.
- Yan, J. and Randell, B. (2005). A systematic classification of cheating in online games. In: *Proceedings of 4th ACM SIGCOMM workshop on Network and system support for games*, pp. 1-9.

3D Human Surface Registration using Subdivision Surfaces

Samuel Zeitvogel, Astrid Laubenheimer
Intelligent Systems Research Group (ISRG)
Karlsruhe University of Applied Sciences, Germany
e-mail: samuel.zeitvogel@hs-karlsruhe.de astrid.laubenheimer@hs-karlsruhe.de

Keywords: Registration, 3D, Subdivision Surface

We propose a surface registration scheme for 3D measurement data using a skinned and subdivided human surface model as a basis. We make use of existing continuous optimization methods used for fitting subdivision surfaces to noisy 3D data and apply them to 3D human surface registration. We present first promising results.

Introduction and Related Work

Template-based 3D surface registration is the process of aligning a template to 3D data, e.g. to point clouds or triangulated surfaces. Thereby, the template is a parametric representation of 3D objects with a fixed topology but variable geometry. The most common templates are given by parametric boundary representations, e.g. 3D meshes consisting of triangles with movable edges. Template-based 3D surface registration is a key component in today's pipeline for generation of deformable surface models. In our case we want to align a 3D human body template to corpora of 3D-scanned humans given as boundary representations. The goal is to deform the template in such a way that the spatial distance to given 3D data is small. A second requirement is that the deformation of the template should preserve anatomically meaningful landmarks, e.g. the templates thumb tip should be registered to data representing a thumb tip and not another fingertip. Once the template is registered to the corpora it is possible to correlate the vertices and faces across the scans to formulate and train a generative human body model i.e. by using principal component analysis to model subject specific body shape and an underlying skeleton to model body pose.

The seminal work of Blanz & Vetter(1999) registered a template of a human face to 3d scans of human faces using optical flow to construct a deformable face model. Angelov et al. (2005) cast the registration problem as a correspondence problem using graphical models. Hirshberg, et al. (2012) use a crude body model to steer the registration process in a bootstrapping fashion. Khamis, et al., (2015) jointly train a deformable human hand model. They use a low-resolution subdivision surface model (Catmull & Clark, 1978) as a template for the model. Cheng, et al. (2004) fits a closed subdivision surface to unordered point clouds.

Contribution

Our registration process takes an unordered point cloud and estimates the vertex positions of a predefined template (see Figure 1) which best describes the input. This template consists of 1140 vertices with 2276 quadrilateral faces and is used as the control cage of an underlying Catmull-Clark subdivision surface.

We propose a 3D human surface registration framework using a simple hand-crafted human model to regularize the process and make use of a subdivision surfaces template instead of the more commonly used piecewise linear mesh templates. One drawback of using non-differentiable surfaces during registration is the need for alternating optimization schemes whereas subdivision surface templates allow off-the-shelf optimization routines. While neither the use of subdivision surfaces for registration and model training in a joint optimization framework (Khamis, et al., 2015) nor a

regularization with the help of a human body model (Hirshberg, et al., 2012) are new, we are not aware of any work fusing these approaches for 3D human body registration.

We define a cost function which consists of two major terms: The first term is the data term with penalizes deviations between the current estimation of the registration and the data on which the registration is performed on. The registered mesh has the same topology as the human model used for regularization. The term is defined as the sum of the squared distances between each data point to its closest point on the subdivision surface of the registration mesh. During optimization, we vary the vertex positions of the mesh and the surface correspondences for each data point to minimize this error and therefore force the registration mesh to align with the data points. The subdivision surface is differentiable everywhere (Peters & Reif, 1998) which enables us to use continuous optimization of all parameters including the surface correspondences which slide along the registration surface during the energy minimization process. As a fall-back strategy, we sample points on the registration surface and assign each data point to the closest surface point. This is similar in spirit to the classical non-rigid iterative closest point algorithm for surface registration. If we use this term alone the vertex positions of the mesh could move freely. This could result in a mesh which describes the surface without making sure that the anatomical landmarks of the mesh land on the corresponding data points. A regularization of some kind is needed to restrict the movement of the vertex positions of the mesh.

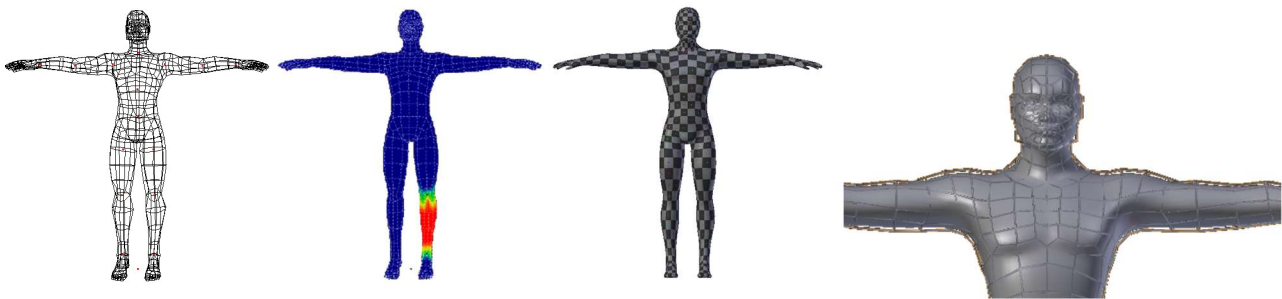


Figure 1: Regularizing Model (enumerated from left to right): (a) the model control mesh in rest pose with the corresponding joint positions. (b) blend weights corresponding to the left knee joint.

Warmer colours indicate stronger influence. (c) Textured view of the underlying subdivision surface. (d) Control cage and corresponding subdivision surface. The topology of the model mesh is also used as the registration template.

Our second term is a regularization term which penalizes the deviation of the registration mesh to a hand-crafted blend skinned human surface model. This model uses an underlying skeleton which consists of 15 joints. The skeleton is coupled to the model surface with blend weights. The blend weights specify the extend of the effect of the joint rotations to the model surface. The model surface is allowed to deform but with restrictions. Allowed deformations are uniform scaling. Other deformations are associated with some cost encouraging the optimization process to deform the model surface in rest pose as rigid as possible. The joint positions are bound to the model mesh in rest pose.

During optimization, we optimize the registration mesh vertex positions, the registration surface points which correspond to each data point, the model rest pose parameters (uniform scaling factor, model rest pose vertex positions, joint positions) as well as the global model pose and the joint rotations. The presented cost function is nonlinear and non-convex. To reduce the chance of converging to local minima we augment our data term with a set of landmarks which were picked on the registration template and the point cloud manually. The landmarks are only used during initialization. Overall, the cost function is formulated as a nonlinear least squares objective which is optimized using the Levenberg-Marquardt algorithm.

Evaluation, Results and Conclusion

With the described approach, we have conducted preliminary experiments on the SCAPE dataset (Anguelov, et al., 2005), a small subset of the CAESAR dataset (Robinette, et al., 1999) and the Faust training dataset (Bogo, et al., 2014). Since our template has spread fingers but the SCAPE and the CAESAR dataset consist of human scans which form their hands to fists. We removed the hands from the scans of these datasets in order to avoid the influence of misleading information. Qualitative results are shown in Figure 2. We do not have access to the CAESAR dataset. Instead of using the original scan data we resort to a pre-registered version of the dataset (Yang, et al., 2014). We also used cleaned-up meshes for the other datasets for our initial evaluation. We treat the pre-registered scans as unordered point clouds.

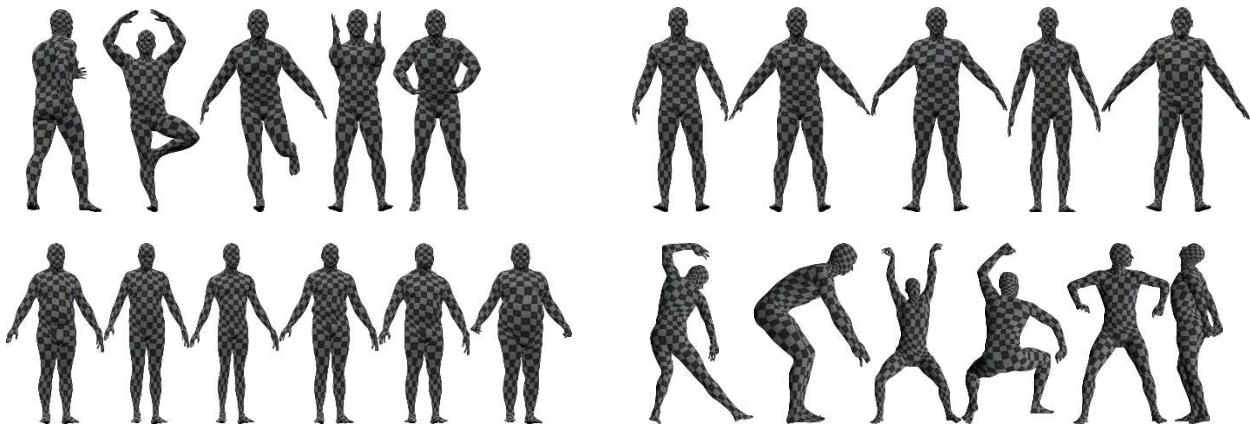


Figure 2: Qualitative Results: Cherry-picked registrations acquired with our proposed approach. The checkerboard texture highlights local deformation and corresponding surface points across the registrations. (top left) five registrations of one person in five different postures (FAUST). (top right) five registrations of five persons in a similar A-posture (FAUST). (bottom left) six registrations of six persons in a similar A-posture (CAESAR). (bottom right) six registrations of one person in six different postures (SCAPE). Best viewed on a computer screen with zoom.

We conducted preliminary quantitative evaluation on the FAUST dataset. We use a publicly available evaluation framework provided by Chen & Koltun (2015). The benchmark consists of 100 pairs of 3D human scans. Each pair consists of a source scan and a target scan. The goal is to match multiple points on the surface of the source scan to the corresponding location on the target scan. The benchmark is split into an intra-subject challenge and an inter-subject challenge. The results are shown in Figure 3. Our approach is not directly comparable to the competing approaches as we make use of manually placed landmarks and our approach can only be applied to the object class of the template. Our current implementation requires approximately half an hour to process a single point cloud.

Nevertheless, our qualitative and quantitative results indicate that our proposed approach can be used to reparametrize existing cleaned up surface meshes and noisy real-world 3D data from heterogeneous corpora to a 3D subdivision surface template. Interpolating between registrations and incorporating our approach in a pipeline for 3D deformable model generation is subject to future work.

Acknowledgements

The authors would like to thank the German Federal Ministry of Education and Research (BMBF) for funding the presented research under grant #03FH061PX5.

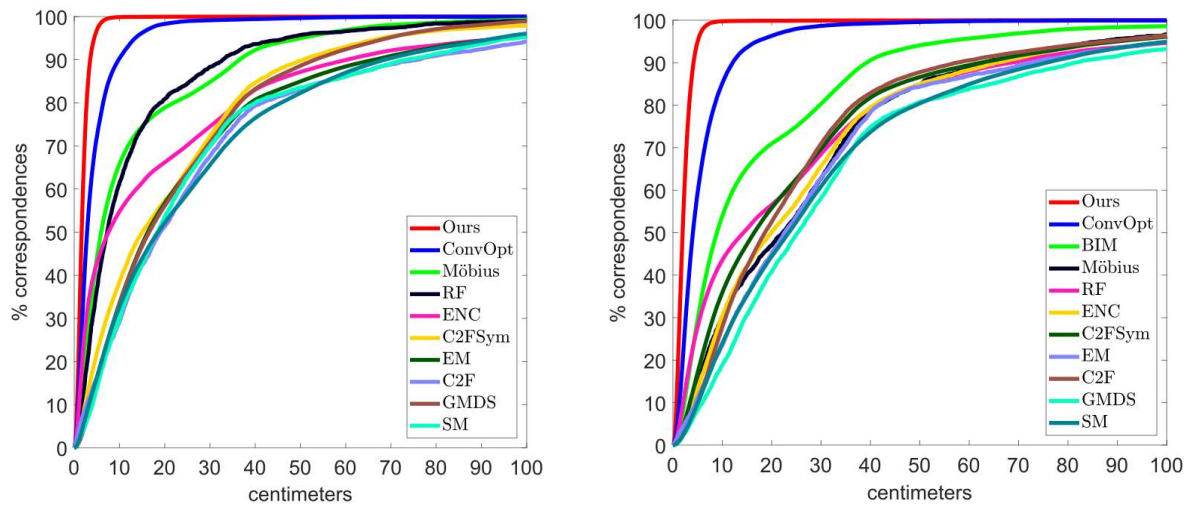


Figure 3: Quantitative Results: This figure shows the error distribution of different methods for the intra-subject challenge (left) and the inter-subject challenge (right). The curves show the percentage of correspondences which deviate from the ground truth by less than a varying threshold (higher is better). The results of Chen & Koltun (2015) are labelled ConvOpt. A detailed review of the evaluation methodology and the competing methods is described in their work.

References

- Anguelov, D. et al., (2005). SCAPE: Shape Completion and Animation of People. *ACM Trans. Graph.*, #jul#, Band 24, pp. 408-416.
- Blanz, V. & Vetter, T., (1999). *A morphable model for the synthesis of 3D faces*. s.l., s.n., pp. 187-194.
- Bogo, F., Romero, J., Loper, M. & Black, M. J., (2014). *FAUST: Dataset and evaluation for 3D mesh registration*. Piscataway, IEEE.
- Catmull, E. & Clark, J., (1978). Recursively generated B-spline surfaces on arbitrary topological meshes. *Computer-aided design*, Band 10, pp. 350-355.
- Cheng, K.-S.D. et al., (2004). *Fitting subdivision surfaces to unorganized point data using SDM*. s.l., s.n., pp. 16-24.
- Chen, Q. & Koltun, V., (2015). *Robust nonrigid registration by convex optimization*. s.l., s.n., pp. 2039-2047.
- Hirshberg, D., Loper, M., Rachlin, E. & Black, M. J., (2012). *Coregistration: Simultaneous alignment and modeling of articulated 3D shape*. s.l., Springer-Verlag, pp. 242-255.
- Khamis, S. et al., (2015). *Learning an efficient model of hand shape variation from depth images*. s.l., s.n., pp. 2540-2548.
- Peters, J. & Reif, U., (1998). Analysis of algorithms generalizing B-spline subdivision. *SIAM Journal on Numerical Analysis*, Band 35, pp. 728-748.
- Robinette, K. M., Daanen, H. & Paquet, E., (1999). *The CAESAR project: a 3-D surface anthropometry survey*. s.l., s.n., pp. 380-386.
- Yang, Y. et al., (2014). *Semantic parametric reshaping of human body models*. s.l., s.n., pp. 41-48.

Cost-Effective Semi-Automatic Ontology Development from Large Domain Terminology

Bernhard G. Humm
Hochschule Darmstadt - University of Applied Sciences
Haardtring 100, 64295 Darmstadt, Germany
bernhard.humm@h-da.de

Hesam Ossanloo
OBJECT ECM GmbH,
Grimm 3, 20457 Hamburg, Germany
hesam.ossanloo@gmail.com

Abstract: When developing a semantic application, the best-case scenario is that there is already a suitable off-the-shelf ontology available. If no suitable ontology is available, a new ontology needs to be developed, which may be costly. This paper describes a methodology for cost-effectively and semi-automatically developing an ontology, given a large domain terminology. The methodology is outlined by means of an example from the domain of software component search.

Keywords: Ontology, ontology learning, semantic web, machine learning, ETL

1. Introduction

Having a suitable ontology plays a most important role in implementing semantic applications. (Maedche and Staab, 2001) state: “The Semantic Web relies heavily on formal ontologies to structure for comprehensive and transportable machine understanding. Thus, the proliferation of ontologies factors largely in the Semantic Web’s success”.

When developing a semantic application, the best-case scenario is that there is already a suitable off-the-shelf ontology, developed by domain-experts, available for the application domain. Examples are “Medical Subject Headings” (MESH) (Rogers, 2017) in the medical domain or GND (Haffner, 2017) in the library domain.

If no suitable off-the-shelf ontology is available for an application domain, a new ontology needs to be developed. This may be costly. A good starting point for developing an ontology is a set of terms which are relevant in the application domain. We call such a set of terms a *domain terminology*. However, if such a domain terminology is large, i.e., contains several thousand terms, then manually developing an ontology including all relevant terms is costly.

This paper describes a methodology for cost-effectively and semi-automatically developing an ontology, given a large domain terminology. We demonstrate the methodology in detail by means of an example from the domain of software component search.

Software development today means, to a large extend, integrating existing software components. An important task of the architect of a software solution is to identify suitable software components. (Humm and Ossanloo, 2016) present a semantic search engine for software components called “SoftwareFinder”. The functionality of software components is usually described by *terms*, also called tags, e.g., “database management system”, “word processor”, “machine learning”, etc. “SoftwareFinder” uses a simple ontology for normalizing the term sets of various software hosting sites like apache.org, for a semantic faceted search, a semantic AutoSuggest service, and for recommending similar software components.

Despite intensive research, the authors did not find any off-the-shelf ontology suitable for software component search. However, the combined term set of the software hosting sites “apache.org”,

“sourceforge.net”, and “alternativeto.net” includes more than 20,000 terms. The methodology presented in this paper has successfully been used for developing a suitable ontology for software component search based on this term set.

2. Problem Statement

In this section, we detail the problem by means of pre-conditions:

1. In a specific application domain, an ontology is needed and requirements for the ontology are specified.
2. A large terminology for the application domain, including several thousand terms, exists. These terms need to be included in the ontology.
3. There is no ontology available which meets the specified requirements.

If these conditions are met, the methodology for developing a suitable ontology in a cost-effective, semi-automatic way can be applied. This methodology is described in the following section.

3. Methodology

In this section, we explain the proposed methodology by means of the example of software component search. See Fig. 1 for a methodology overview in BPMN Notation.

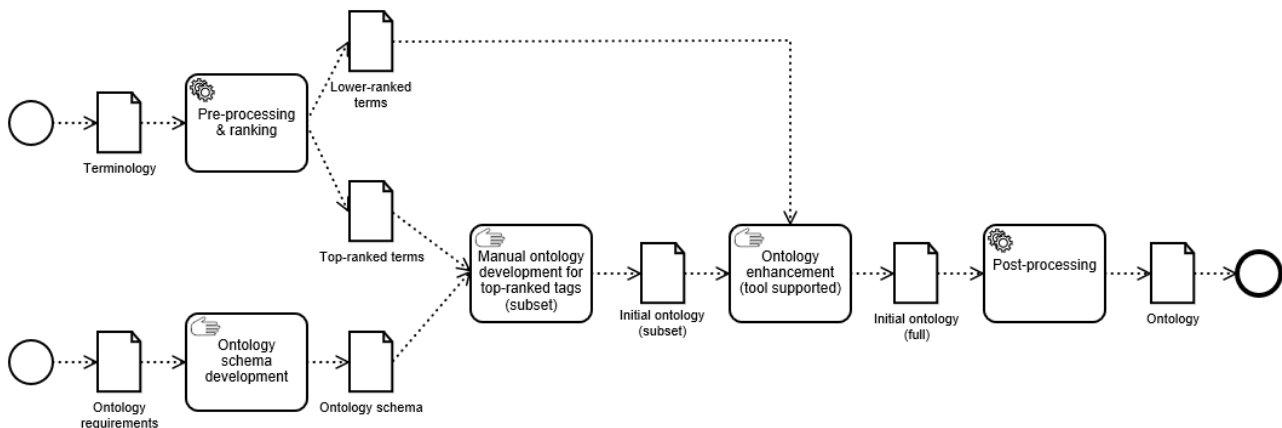


Figure 1: Methodology overview

We explain the process steps in the following sections.

3.1. Ontology Schema Development

Based on the ontology requirements, the domain-specific ontology schema needs to be developed.

In the example of software component search, the ontology requirements are as follows. All terms for describing software components shall be assigned a semantic category, e.g., the category “business” for the term “enterprise resource planning” or the category “development” for the term “database management system”. For SoftwareFinder, 12 semantic categories have been identified. Acronyms shall be associated with terms, e.g., “DBMS” to “database management system”. Also,

synonyms shall be associated, e.g., “database” with “database management system”. Please note: technically speaking, “database” and “database management system” are no synonyms. However, we assume that users who manually tagged software components in hosting sites like sourceforge.net used both terms interchangeably.

Based on those requirements, the ontology schema is simple, consisting of a single entity “Concept” only. See Fig. 2 for a UML class diagram.

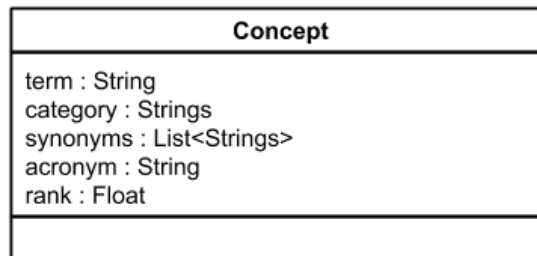


Figure 2: Ontology schema: entity Concept

A concept consists of the term as identifier, a semantic category, and optionally an acronym and a list of synonyms. A rank denotes the importance of the term in relation to other terms.

3.2. Pre-Processing and Ranking

In this step, the terminology is normalized and anomalies are handled in a domain-specific way. Furthermore, all terms are ranked according to relevance.

In the example of software component search, the following configuration data is used for normalization:

1. Blacklist (Ignore list): Some terms have no relevance for the software component search. E.g., the term “Other/Nonlisted” is used in sourceforge.net for describing software components. The terminology for SoftwareFinder is created by crawling software hosting sites like sourceforge.net and extracting all terms being used for describing software components. While the term “Other/Nonlisted” may be useful in the context of the hosting site, it provides no useful information for the software component search. Therefore, such terms are specified in a blacklist which is used for removing them from the terminology as a pre-processing step.
2. Composite terms: Some terms are composites including several concepts. E.g., the term “audio/video” indicates functionality for audio processing as well as for video processing. To improve searching, the term will be split into the two terms “audio” and “video”.

Ranking terms according to relevance is domain-specific. In the example of software component search, the number of software components that a term is assigned to is used as a heuristic for its relevance: the more often a term is used the more relevant it is considered and, hence, the higher its rank.

The output of the step “pre-processing & ranking” is list of normalized terms, sorted according to a relevance ranking with the top-ranked terms first.

3.3. Manual Ontology Development

In this step, the domain expert takes a manageable subset of the terms, e.g., the first 500 top-ranked terms, and creates the corresponding concepts and relationships manually.

In the software component search example, this includes manually assigning a suitable semantic category to each term. The output of this step is the first version of the ontology filled with a subset of the concepts containing the terms and their semantic categories.

3.4. Ontology enhancement

In this step, the initial full ontology is developed, also including the lower-ranked terms. For this, domain-specific software tools may be used interactively.

In the example of software component search, simple software tools can be developed and used for identifying candidates for synonyms and acronyms. E.g., containment of one term in another indicates a potential synonym, e.g., “word processor” is contained in “word processors”. Please note: linguistically speaking, “word processor” and “word processors” are no synonyms. However, we assume that users who manually tagged software components in hosting sites like sourceforge.net used both terms interchangeably.

Also, pattern matching techniques like Jaro-Winkler (Winkler, 1999) are used to find similarities which may indicate potential synonyms, e.g., “word processor” and “word processing”. Using such tools, the domain expert can interactively enhance the ontology.

To find the potential acronyms among the terms within the ontology, the algorithm to extract abbreviations by (Hearst, 2003) is used. The output of the algorithm is a list of all terms in the ontology with the suggested potential acronyms. The domain expert can go through the list manually and extend the ontology accordingly.

Techniques for pattern matching and finding similarities are commonly used in approximate string matching (Hall and Dowling, 1980) (Jokinen et al., 1996) (Navarro, 2001). These techniques have widespread applications, in data linkage (Christen et al., 2004) (Christen and Goiser, 2007) (Porter et al., 1997) (Winkler, 2006) (Yancey, 2004) (Yancey, 2005), duplicate detection (Bilenko and Mooney, 2003) (Borgman and Siegfried, 1992) (Cohen et al., 2003), information retrieval (Gong and Chan, 2006) (Keskustalo et al., 2003), spelling errors correction (Damerau, 1964) (Kukich, 1992) (Pollock and Zamora, 1984), approximate database joins (Gravano et al., 2001), bioinformatics, and health informatics (Friedman and Sideli, 1992).

Table 1 shows the similarity scores of different techniques for potential synonyms of the term “word processor”. A score of 0 denotes no similarity at all and 1 denotes identity. Since lowercase and uppercase letters are distinguished, all terms are lowercased.

	“word processing”	“word processors”	“many word processors”	“processor word”
Jaro-Winkler similarity score	0.977	0.998	0.852	0.793
Jaccard metric similarity score	0.666	0.923	0.666	0.6
Normalized Levenshtein similarity score	0.8	0.933	0.7	0.285
Cosine similarity score	0.8	0.960	0.816	0.750
Sorensen-Dice coefficient similarity score	0.8	0.96	0.8	0.75

Table 1: Comparison of similarity scores for the term “word processor”

The comparison shows that the Jaro-Winkler technique delivers the best results for this use case. This is in line with the comprehensive evaluation of different pattern matching techniques by Christen (2006). After various optimization steps, the Jaro-Winkler threshold of 0.9 appears to be optimal for this use case.

3.5. Post-Processing

In this step, the ontology is finalized. For this, domain-specific heuristic processes or machine learning techniques may be used.

In the example of software component search, the semantic categories for all lower-ranked terms (in total more than 20,000) need to be predicted. As input for prediction, the co-occurrence of terms in software components may be used. This is based on the assumption that terms, which are often used together for describing software components, belong to the same semantic category.

To verify this assumption, we have experimented with various supervised machine learning techniques (Russell and Norvig, 2009). The training set is based on the manually classified semantic categories of the top-ranked terms (about 700 terms). Figure 3 shows the feature vector used for training as UML class diagram.

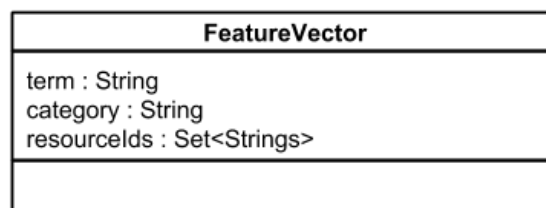


Figure 3: Feature vector entity class diagram

The feature vector contains the term (the identifier), the semantic category (the category to be predicted) and a set of unique identifiers of the software products, in which the term is used (resource Ids). This set represents the co-occurrence of the terms in concrete software products and is used as a bag-of-words-model (Nigam et al., 2000). We used the data science platform RapidMiner (<https://rapidminer.com/>) and applied more than 10 machine learning techniques, including deep neural networks, Bayesian classifiers and decision trees. Using cross-validation (Larson, 1931), the overall accuracy was evaluated. However, the accuracy of all approaches ranged between 27% and 37% which is not considered sufficient.

To improve prediction accuracy, a domain-specific heuristic approach was implemented. For each term, all terms are collected which co-occur in any software product. The semantic category most often used in those co-occurring terms is used as a prediction of the semantic category. Consider the following examples:

1. “word processor” (expected category: “General”): General:259 (correct!), Development:38, Infrastructure:37, Business:12, Multimedia:8, Communication:7, Standards:6, Science:5, Humanities:5, Entertainment:3, Engineering:2, Security:1
2. “database management system” (expected category: “Infrastructure”): Infrastructure:81 (correct!), Development:43, General:31, Business:15, Standards:3, Science:1
3. “hypertext markup language” (HTML, expected category: “Standards”): Development:353 (incorrect!), General:147, Standards:63, Infrastructure:38, Communication:11, Business:4, Multimedia:2, Humanities:2

The prediction accuracy compared to machine learning approach could be improved. Out of 778 terms, 391 terms were correctly classified, 287 were incorrectly classified, and 109 could not be classified. The accuracy of predictions is 58%. A prediction is not possible for terms without classified co-occurring terms.

For SoftwareFinder, the prediction accuracy of the heuristic approach is sufficient since an incorrect semantic category is not mission critical. For example, the only effect of misclassifying HTML is that in the semantic autosuggest and topic pie features, HTML will be displayed under the category “Development” instead of “Standards”. The user of SoftwareFinder may be surprised by this but it will certainly not impede the semantic search.

4. Related Work

Methodologies for developing ontologies were first brought to attention by the early works of Leśniewski (Lesniewski, 1930) in which he uses a symbolic language rather than a colloquial language to develop logical views. He introduced a mathematical approach to defining the concepts in an ontology and the logical relation among those concepts in the ontology.

Based on his early works, different methodologies have been introduced to on developing ontologies. Jones and his colleagues (Jones et al., 1998) have addressed the problem of significant variety in ontologies even if they have been constructed for very similar purposes. They postulated the need for a good methodology for developing ontologies. They performed a comprehensive survey on different methodologies for developing an ontology. In the following, we compare those methodologies with our approach.

KBSI IDEF5 (Mayer et al., 1995) has some similarities to the methodology presented in this paper. Their approach involves:

1. Organizing and scoping: this step is similar to define ontology specification step of our methodology
2. Data collection: this step is the pre-requisite of our methodology
3. Data analysis and initial ontology development: those steps are similar to our steps of “Pre-processing & ranking” and “Manual ontology development”
4. Ontology refinement and validation: this step is similar to our “Ontology enhancement” and “Post-processing” step

However, in KBSI IDEF5, there are no automatic or semi-automatic approaches in order to reduce the cost of ontology development.

The TOVE (Toronto Virtual Enterprise) approach (Gruninger and Fox, 1994) is most interesting for the emphasis on ontology evaluation, providing in completeness theorems. These theorems are useful in a number of ontology maintenance tasks, e.g. assessing the extendibility of an ontology – any extension must be able to preserve the validity of the completeness theorems – or to provide a benchmark for ontologies.

METHONTOLOGY: Initially described in (Gómez-Pérez et al., 1996) and updated in (Fernández-López et al., 1997), they have a similar approach to KBSI IDEF5 and TOVE. Like TOVE, the most distinctive aspect of METHONTOLOGY is the focus on maintenance. The main difference between the two is that in METHONTOLOGY the focus is on comprehensively addressing the maintenance stage of the life cycle of an ontology whereas TOVE utilises more formal techniques to address a more limited number of maintenance issues. Again, they also do not mention ways of automatic or semi-automatic approaches.

Ontolingua: The guides for use of the Ontolingua server (Farquhar et al., 1995) (Farquhar et al., 1997a) (Farquhar et al., 1997b) contain advice on browsing, developing, maintaining and sharing ontologies stored at the server. One of the main benefits in using the Ontolingua server is the access it provides to a library of previously defined ontologies, which is based on having already some ontologies on the server.

CommonKADS and KACTUS: they introduce a methodology (Schreiber et al., 1995) (Wielinga et al., 1994) to develop an ontology which is constructed from a library of small-scale ontologies, which requires the mapping between the various ontologies included in the development of the new ontology. If there are no domain-specific ontologies available, this methodology is not applicable.

The Plinius project (Mars et al., 1994) attempts the semi-automatic extraction of knowledge from natural-language texts, namely the title and abstracts of bibliographic document descriptions in the online version of Engineered Materials Abstracts. Although the authors claim that these principles apply generally in the development of ontologies, they also suggest that they are not always sensible or even feasible.

ONIONS (ONtologic Integration Of Naive Sources) (Gangemi et al., 1996) (Steve and Gangemi, 1996) is motivated by the knowledge integration problem, i.e., how to integrate heterogeneous sources of information in knowledge acquisition. This problem is addressed through the creation of a formal domain ontology by the integration of existing repositories of knowledge. If there are no domain-specific ontologies available, this methodology is not applicable.

The MENELAS ontology was designed as part of a natural language understanding system (Bouaud et al., 1994) (Bouaud et al., 1995). Four principles in the development of taxonomic knowledge in ontologies are described which helps mainly at a fine level of detail. It does, however, take a rather idealized view of taxonomies which may not be applicable to many domains.

The PHYSSYS method (Borst et al., 1996) aims to facilitate the selection of pre-existing ontologies based on dynamic knowledge construction rather than the simple selection of knowledge components. If there are no domain-specific ontologies available, this methodology is not applicable.

The SENSUS approach (Swartout et al., 1996) is based on the assumption that if two knowledge bases are built on a common ontology, knowledge can be shared between them more readily since

they share a common underlying structure. If there are no domain-specific ontologies available, this methodology is not applicable.

To summarize, we are not aware of an existing methodology for semi-automatic ontology development from large domain terminology in a cost-effective way.

5. Conclusions and Future Work

The methodology presented in this paper has been successfully used for developing an ontology for software component search. It has considerably reduced the time for developing and maintaining the ontology for the semantic search application SoftwareFinder. The ontology meets all requirements specified and is successfully being used in SoftwareFinder.

The methodology is general and we expect it to be useful in many application domains where an ontology is to be developed from a large domain terminology. However, the use of the methodology in other application domains and the evaluation thereof is subject to future work.

References

- Bilenko, M., Mooney, R.J. (2003). "Adaptive duplicate detection using learnable string similarity measures", in: *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, pp. 39–48.
- Borgman, C.L., Siegfried, S.L. (1992). "Getty's synonameTM and its cousins: A survey of applications of personal name-matching algorithms". *J. Am. Soc. Inf. Sci.* 43, 459.
- Borst, P., Benjamin, J., Wielinga, B., Akkermans, H., et.al. (1996). "An application of ontology construction", in: *Proc. of ECAI-96 Workshop on Ontological Engineering*. Budapest. p. 24.
- Bouaud, J., Bachimont, B., Charlet, J., Zweigenbaum, P. (1995). "Methodological principles for structuring an "ontology,"" in: *Proceedings of the IJCAI'95 Workshop on "Basic Ontological Issues in Knowledge Sharing"*. pp. 19–25.
- Bouaud, J., Bachimont, B., Charlet, J., Zweigenbaum, P. (1994). "Acquisition and structuring of an ontology within conceptual graphs", in: *Proceedings of ICCS*. pp. 1–25.
- Christen, P. (2006). "A comparison of personal name matching: Techniques and practical issues", in: *Data Mining Workshops, 2006. ICDM Workshops 2006*. Sixth IEEE International Conference On. IEEE, pp. 290–294.
- Christen, P., Churches, T., Hegland, M., et.al. (2004). "Febri-a parallel open source data linkage system". In: *Lect. Notes Comput. Sci.* pp. 638–647.
- Christen, P., Goiser, K. (2007). "Quality and Complexity Measures for Data Linkage and Deduplication". In: *Qual. Meas. Data Min.* 43, pp. 127–151.
- Cohen, W., Ravikumar, P., Fienberg, S. (2003). "A comparison of string metrics for matching names and records", in: *Kdd Workshop on Data Cleaning and Object Consolidation*. pp. 73–78.
- Damerau, F.J. (1964). "A technique for computer detection and correction of spelling errors". In: *Commun. ACM* 7, pp. 171–176.
- Farquhar, A., Fikes, R., Pratt, W., Rice, J. (1995). *Collaborative ontology construction for information integration. Technical Report KSL-95-63*, Stanford University Knowledge Systems Laboratory.
- Farquhar, A., Fikes, R., Rice, J. (1997a). "The ontolingua server: A tool for collaborative ontology construction". In: *Int. J. Hum.-Comput. Stud.* 46, pp. 707–727.
- Farquhar, A., Fikes, R., Rice, J. (1997b). "Tools for assembling modular ontologies in Ontolingua". In: *Proc AAAI* 97.
- Fernández-López, M., Gómez-Pérez, A., Juristo, N. (1997). *Methontology: from ontological art towards ontological engineering*.

- Friedman, C., Sideli, R. (1992). "Tolerating spelling errors during patient validation". In: *Comput. Biomed. Res.* 25, pp. 486–509.
- Gangemi, A., Steve, G., Giacomelli, F. (1996). "ONIONS: An ontological methodology for taxonomic knowledge integration", in: *Proceedings of the Workshop on Ontological Engineering, ECAI96*.
- Gómez-Pérez, A., Fernández, M., De Vicente, A. (1996). "Towards a method to conceptualize domain ontologies", in: *ECAI96 Workshop on Ontological Engineering*, Budapest. pp. 41–51.
- Gong, R., Chan, T.K. (2006). "Syllable alignment: A novel model for phonetic string search". In: *IEICE Trans. Inf. Syst.* 89, 332–339.
- Gravano, L., Ipeirotis, P.G., Jagadish, H.V., Koudas, N., Muthukrishnan, S., Srivastava, D., et. al. (2001). "Approximate string joins in a database (almost) for free", in: *VLDB*. pp. 491–500.
- Gruninger, M., Fox, M.S. (1994). "The design and evaluation of ontologies for enterprise engineering", in: *Workshop on Implemented Ontologies, European Workshop on Artificial Intelligence*, Amsterdam, The Netherlands.
- Haffner, A. (2017). GND ONTOLOGY.
- Hall, P.A., Dowling, G.R. (1980). "Approximate string matching". In: *ACM Comput. Surv. CSUR 12*, pp. 381–402.
- Hearst, M.S. (2003). A simple algorithm for identifying abbreviation definitions in biomedical text.
- Humm, B.G., Ossanloo, H. (2016). "A Semantic Search Engine for Software Components". In: *Isaias, P. (Ed.), Proceedings of the International Conference WWW/INTERNET 2016*. IADIS Press, Mannheim, pp. 127–135.
- Jokinen, P., Tarhio, J., Ukkonen, E. (1996). "A comparison of approximate string matching algorithms". In: *Softw. Pract. Exp.* 26, pp. 1439–1458.
- Jones, D., Bench-Capon, T., Visser, P. (1998). Methodologies for ontology development.
- Keskustalo, H., Pirkola, A., Visala, K., Leppänen, E., Järvelin, K. (2003). "Non-adjacent digrams improve matching of cross-lingual spelling variants", in: *String Processing and Information Retrieval. Springer*, pp. 252–265.
- Kukich, K. (1992). "Techniques for automatically correcting words in text". in: *ACM Comput. Surv. CSUR 24*, pp. 377–439.
- Larson, S.C. (1931). "The shrinkage of the coefficient of multiple correlation". In: *J. Educ. Psychol.* 22, p. 45.
- Lesniewski, S. (1930). 'Über die Grundlagen der Ontologie'. *Comptes Rendus*, pp. 111–132.
- Maedche, A., Staab, S. (2001). "Ontology learning for the Semantic Web". In: *IEEE Intell. Syst.* 16, 72–79. doi:10.1109/5254.920602
- Mars, N., de Jong, H., Speel, P., ter Stal, W., van der Vet, P. (1994). "Semi-automatic knowledge acquisition in Plinius: an engineering approach. Knowledge-based systems workshop", in: Unknown.
- Mayer, R.J., Menzel, C.P., Painter, M.K., Dewitte, P.S., Blinn, T., Perakath, B. (1995). "Information integration for concurrent engineering (IICE) IDEF3 process description capture method report". KNOWLEDGE BASED SYSTEMS INC COLLEGE STATION TX.
- Navarro, G. (2001). "A guided tour to approximate string matching". In: *ACM Comput. Surv. CSUR 33*, pp. 31–88.
- Nigam, K., McCallum, A.K., Thrun, S., Mitchell, T. (2000). "Text Classification from Labeled and Unlabeled Documents Using EM. Mach Learn 39", pp. 103–134. doi:10.1023/A:1007692713085
- Pollock, J.J., Zamora, A. (1984). "Automatic spelling correction in scientific and scholarly text". In: *Commun. ACM* 27, pp. 358–368.
- Porter, E.H., Winkler, W.E., others (1997). *Approximate string comparison and its effect on an advanced record linkage system*, in: *Advanced Record Linkage System*. US Bureau of the Census, Research Report. Citeseer.
- Rogers, F. B. (2017). Medical Subject Headings (MESH).
- Russell, S.J., Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*, 3rd ed. Pearson Education, Upper Saddle River, New Jersey.
- Schreiber, G., Wielinga, B., Jansweijer, W., others (1995). "The KACTUS view on the 'O'word", in: *IJCAI Workshop on Basic Ontological Issues in Knowledge Sharing*. pp. 159–168.

- Steve, G., Gangemi, A. (1996). "ONIONS methodology and the ontological commitment of medical ontology ON8. 5", in: *Proceedings of Knowledge Acquisition Workshop*.
- Swartout, B., Patil, R., Knight, K., Russ, T. (1996). "Toward distributed use of large-scale ontologies", in: *Proc. of the Tenth Workshop on Knowledge Acquisition for Knowledge-Based Systems*. pp. 138–148.
- Wielinga, B., Schreiber, A.T., Jansweijer, W., Anjewierden, A., Van Harmelen, F. (1994). *Framework and formalism for expressing ontologies*. ESPRIT Proj. 8145.
- Winkler, W.E. (2006). "Overview of record linkage and current research directions", in: Bureau of the Census. Citeseer.
- Yancey, W.E. (2005). "Evaluating string comparator performance for record linkage". In: *Statistics 5*.
- Yancey, W.E. (2004). "An adaptive string comparator for record linkage". In: *Statistics 02*.

From Raw Data to Smart Documentation: Introducing a Semantic Fusion Process for Cyber-Physical Systems¹

Lukas Kaupp[†]Ulrich Beez[†]Bernhard G. Humm[†]Jens Hülsmann^{††}

{lukas.kaupp,ulrich.beez,bernhard.humm}@h-da.de

jhuelsmann@isravision.com

[†] Hochschule Darmstadt -^{††} ISRA Vision AG

University of

Albert-Einstein-Allee 36-40

Applied Sciences

45699 Herten, Germany

Haardtring 100

64295 Darmstadt, Germany

Abstract: Machine outage is a considerable problem in smart factories. This paper introduces a novel *Semantic Fusion Process for Cyber-Physical Systems (SFP-CPS)*. It helps reducing machine outage in factories by automatically detecting problems in CPSs and providing suitable documentation to assist factory personnel in resolving the problem. The *SFP-CPS* operates on raw data collected from the CPS in real-time. Raw data gets normalized, semantically enriched and mapped onto a knowledge base, resulting in a semantic description of the error and its course together with a technical instruction of how to solve the error.

Keywords: Smart factory, smart manufacturing, cyber-physical system, technical documentation, semantic matching, semantic enrichment, inspection machine, stream analysis

1. Introduction

Machine outage in a factory is considerably a worst case scenario for businesses (Saalman & Hellingrath, 2016). Minimizing outage time is key to optimize production flows. In order to assist businesses, manufacturers introduced support contracts with multiple service level agreements (SLA). A maintenance or error situation can include multiple manufacturing support levels within the highly integrated, interconnected, automated machinery environment (Mousheimish et al., 2016). Simplifying the problem at the production site, enabling professionals to solve encountered problems, could help reduce the life cycle costs of production machinery (Gorecky et al., 2014; Saalman & Hellingrath, 2016).

Cyber-physical systems (CPS) (Petnga & Austin, 2016) are defined as physical systems like machines in a factory with sophisticated computation and networking.

This paper introduces a novel *Semantic Fusion Process for CPS (SFP-CPS)*. It helps reducing machine outage in factories by automatically detecting problems in CPSs and providing suitable documentation to assist factory personnel in resolving the problem. Raw data gets normalized, semantically enriched and mapped onto a knowledge base, resulting in a semantic description of the error and its course together with a technical instruction of how to solve the error.

¹This study takes place within the project ProDok 4.0, funded by the German Ministry of Education and Research (BMBF) within the framework of the Services 2010 action plan under funding no. 02K14A110. Executive steering committee is the Karlsruher Institut für Technologie - Karlsruhe Institute of Technology (KIT). Project partners are KUKA AG, ISRA VISION AG, dictaJet Ingenieurgesellschaft mbH and Hochschule Darmstadt - University of Applied Sciences. Glass inspection machine (Type FS5D) logs provided by ISRA VISION AG. All rights on example log data remains solely to ISRA VISION AG. Usage must be requested. Content of the knowledge base will not be released to public.

Technical instructions are provided in form of so-called smart documents. A smart document is modularized, containing the building blocks symptom, cause, and solution.

The SFP-CPS acts passively collecting already generated runtime information on the CPS and can thus be rapidly adapted to additional detectors without touching the functional part of the implementation. In this paper, we illustrate the SFP-CPS by means of an example use case in the domain of quality inspection in glass manufacturing processes.

The remainder of this paper is structured as follows. Section 2 specifies the problem by means of requirements. Section 3 describes the SFP-CPS in detail. Section 4 evaluates the approach and Section 5 compares it with related work. Section 6 concludes the paper and outlines future work.

2. Problem Statement

Machine outage time in factories is considered a worst case. Fast assistance for factory personnel can reduce the outage. The following requirements for SFP-CPS have been developed in cooperation with specialists and factory personnel, to take a highly integrated, interconnected, automated machinery environment into account. They conform to the standard template of the International Requirements Engineering Board (IREB) (Pohl, 2016).

- R1.*Suitable documentation*: The SFP-CPS shall provide suitable smart documents in order to assist personnel in a smart factory in solving problems with CPSs.
- R2.*Right-on-time*: The SFP-CPS shall provide documentation right-on-time, depending on the problem criticality, reducing the time span for fixing the encountered problem.
- R3.*Applicable*: It shall be possible to apply the SFP-CPS to various application domains within the smart factory, e.g., quality inspection in glass manufacturing processes.
- R4.*Event-triggered*: The SFP-CPS shall be triggered automatically whenever a problem occurs in a CPS.
- R5.*Based on raw data*: The SFP-CPS shall collect raw data from CPS. Various raw data formats may be handled.
- R6.*Non-interfering*: The raw data collection of the SFP-CPS shall not interfere with the operation of the CPS.
- R7.*Extensible*: The application of the SFP-CPS in an application domain shall be extensible with new modules (e.g. parsers, algorithms, etc.), in order to cover new application requirements.

3. Semantic Fusion Process

The main purpose of the SFP-CPS is to assist personnel in problem situations of CPS. The SFP-CPS is described in detail in this section.

3.1. Example Use Case

The authors present the SFP-CPS by means of an example from the domain of quality inspection in manufacturing processes, here glass production (Wagner 2008; Jebesen-Marwedel & Brückner, 2010). A conveyor delivers glass towards a glass inspection machine, which detects glass defects and rates material quality (Beyerer et al., 2016). The glass inspection machine consists of cameras and lights and server machines with inspection software (ISRA VISION AG, 2015). An error can occur in every component of the setup, including the interconnections and on both the software level or the hardware level. Errors in the

inspection system may lead to uninspected glass or glass of unknown quality and thus impact the yield of the plant. With today's quality requirements for glass products, only quality inspected glass can be sold to customers.

Common machinery errors are camera failures, network errors, lighting issues or a wrong configuration of external system parameters from the plant. Known events are detected by the CPS and get reported. Typically, these issues cannot be solved by the CPS on its own, i.e. a lost communication signal with another node due to a physical damage of the cable.

Less obvious or even previously unknown problems can cause an altered behavior in the defects detection (over or under detection) and cause unreliable quality information.

Often, hints for these kinds of problems can be found if the distributed information the system generates at runtime is considered as a whole. The SFP-CPS aims at detecting errors within the inspection machine based on different indicators from this runtime information and give a statement on cause and, if possible, point to a solution to fix the problem.

3.2. Process Overview

Figure 1 shows the SFP-CPS as BPMN diagram. Any internal state changes, exceptions, sensor data, and communication between components (software or hardware components) in a CPS is saved into log files. Each line in a log file corresponds to a log event (raw event). Raw events are the inbound data of the SFP-CPS. Multiple log events get collected within a *data stream* and are delivered in real-time into the process.

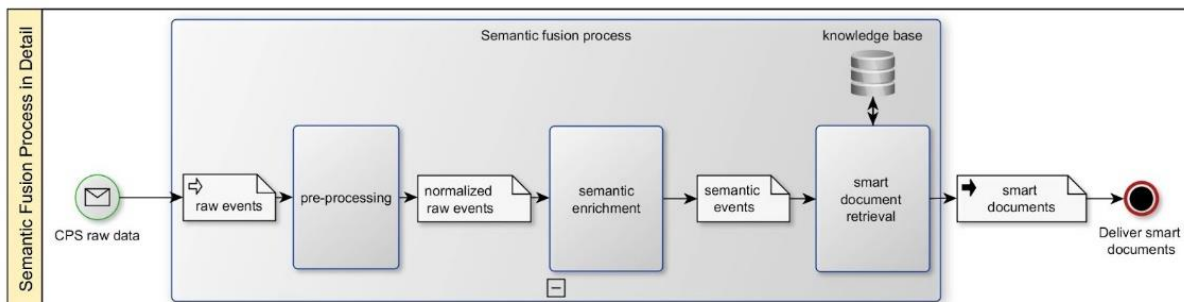


Figure 1: Semantic fusion process

The semantic fusion process consists of three components. In the *pre-processing* step, differently formatted log events are normalized to a defined schema. In the *semantic enrichment* step, normalized raw events are refined using analysis methods, generating semantic events. In the *smart document retrieval* step, these semantic events enable a fine-grained query to SFP-CPSs' knowledge base and lead to a composition of smart documents.

3.3. Pre-Processing

In order to support multiple log events with different encodings and formats from different software modules and hardware modules, a pre-processing step is needed. Implementing the idea of (Nuñez & Borsato, 2017) on the input data, log events are normalized using a defined schema, enabling SFP-CPS to be independent on the format of inbound data. Figure 2 shows the pre-processing step by means of the example event 'GlassBreakBegin'.

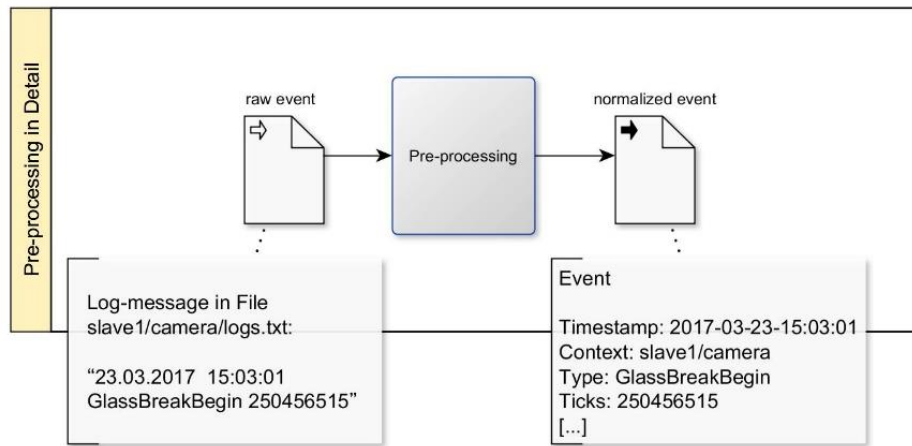


Figure 2: Pre-processing of the event GlassBreakBegin

‘GlassBreakBegin’ indicates the detection of a break within the glass layer. It is reported by the CPS as a formatted log message as shown in Figure 2. Each event has generalized attributes such as timestamp, context and type. The ‘timestamp’ attribute indicates the time when the error occurred. The ‘context’ attribute reflects the origin within the CPS, e.g. for a camera event on slave 1 ‘/slave1/camera’. The ‘type’ attribute classifies the event, here ‘GlassBreakBegin’. In addition, an event may have specific event attributes e.g. ticks (conveyor belt position). The formatted log message of the raw event gets parsed and the extracted data are stored in an event.

3.4. Semantic Enrichment

Each normalized raw event alone may not be sufficient to identify the CPS problem. Consequently, in a second step, the normalized log events get lifted semantically in the semantic enrichment process. Figure 3 shows the semantic enrichment process with four sub-processes. Normalized raw events are input data for the semantic enrichment process. The process can apply *filters*, *pattern matching*, *value progression analysis*, and *time progression analysis* to generate semantic events.

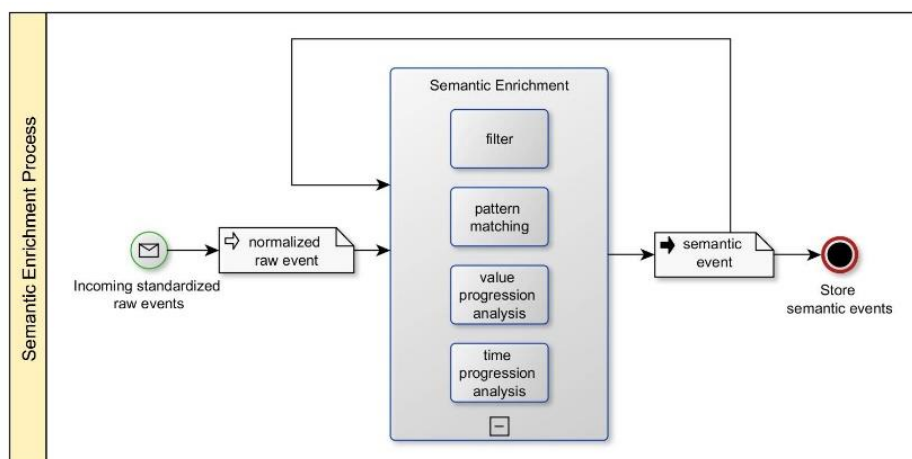


Figure 3: Semantic enrichment process

In addition, semantic events can be used as input data for semantic enrichment. In case of an inbound semantic event, higher semantic events can be generated.

Figure 4 displays a filtering operation on a normalized event stream, here filtering ‘GlassBreak’ events.

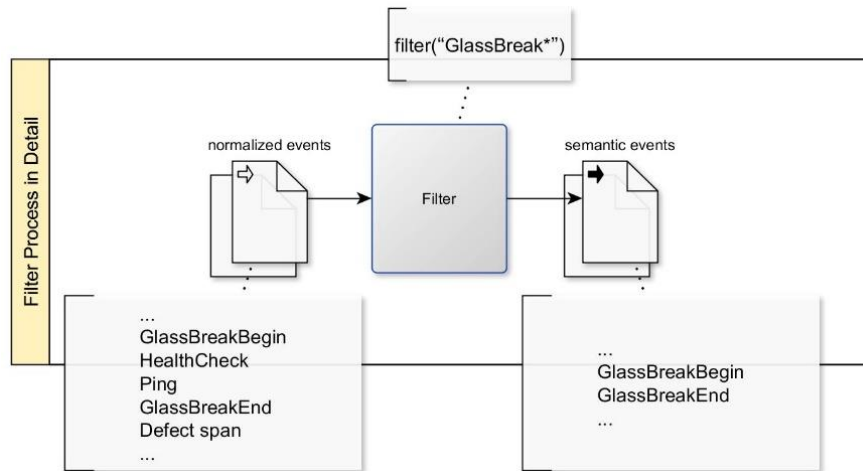


Figure 4: Example Filter

On the left side, the unfiltered data stream is shown containing multiple normalized events. On the right side, only filtered events are shown. For the filter operation, the authors use pseudo code oriented at common complex event processing languages like Apache Flink.

Figure 5 shows an example for pattern matching, identifying corresponding ‘GlassBreakBegin’ and ‘GlassBreakEnd’ events.

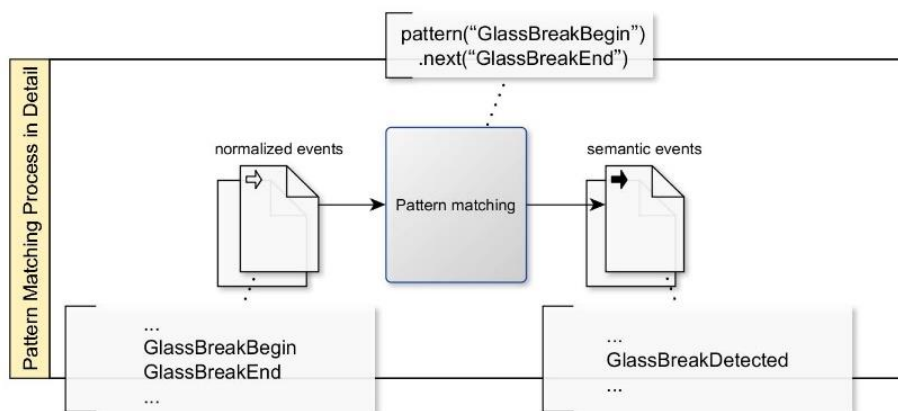


Figure 5: Example Pattern Matching

The pseudo code specifies a pattern in which a ‘GlassBreakBegin’ event immediately followed by a ‘GlassBreakEnd’ event in the data stream is detected. Each recognized pattern can be used to generate a semantic ‘GlassBreakDetected’ event.

Figure 6 shows an example of a value progression analysis for generating the semantic ‘SpeedChanged’ event. The speed of the inspection process may vary, due to the versatility of the glass production process. A speed change may have an effect on defect detection and, therefore, is an important semantic information.

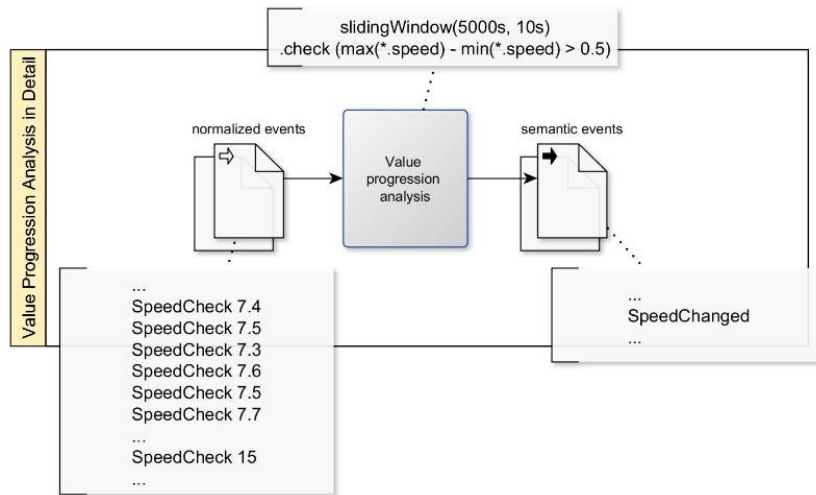


Figure 6: Example value progression analysis

Each ‘SpeedCheck’ event delivers a snapshot of the speed of the conveyor belt. The pseudo code for implementing the value progression analysis uses sliding time windows of size 5000s with an overlap of 10s. For each time window, it is checked whether the difference in speed exceeds a threshold (here: 0.5). In this case, a new semantic event ‘SpeedChanged’ is generated.

Figure 7 illustrates the time progression analysis by means of the example of the ‘SignalLost’ event. The ‘SignalLost’ event indicates a connectivity failure, e.g., between a camera and a server.

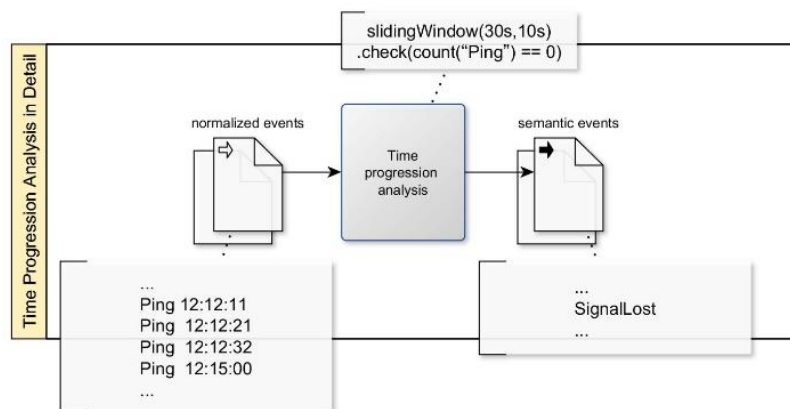


Figure 7: Example time progression analysis

Each component within the CPS regularly sends ping events. If no ping event occurs within half a minute, it is considered as a loss of signal. The pseudo code shown in Figure 9 uses a sliding time window of 30 s with an overlap of 10 s. If no ping event occurred within the time window then a ‘SignalLost’ event is generated.

As shown in Figure 3, semantic events generated by semantic enrichment processes can be used as input for other semantic enrichment processes. So, a chain of successive enrichment processes can be established. For example, high conveyor belt speed typically implies a thinner glass layer. With the glass thickness the properties of the defects change and in the transition between different thicknesses a lot of defects may occur. The semantic event ‘SpeedChanged’ introduced above may then be used for generating a semantic ‘ThicknessChanged’ event.

3.5. Smart Document Retrieval

Technical documentation is stored in a *knowledge base*. Technical instructions are provided in form of *smart documents*. A smart document is modularized, containing the building blocks symptom, cause, and solution. This structure forms the schema of the knowledge base. The authors call it the *symptom / cause /solution model (SCS model)*.

A *symptom* is a misbehavior of any form as visual, physical or nonphysical (software-related) aspect (Hornung et al., 2014). A *cause* can be the origin of a symptom; however one cause can be linked to multiple symptoms and symptoms can have multiple causes. In addition, a *solution* covers one or more causes and a cause can be fixed by multiple solutions. In addition, each solution can have a semantic context defining the scope of the solution (e.g. *server*). See Figure 8.

Through this modularized structure it is possible to assemble a smart document consisting of one symptom, one cause and one solution according to the semantic context for a given semantic event. The smart document is the final output of the SFP-CPS semantically interlinking semantic events and documentation for a specific CPS problem.

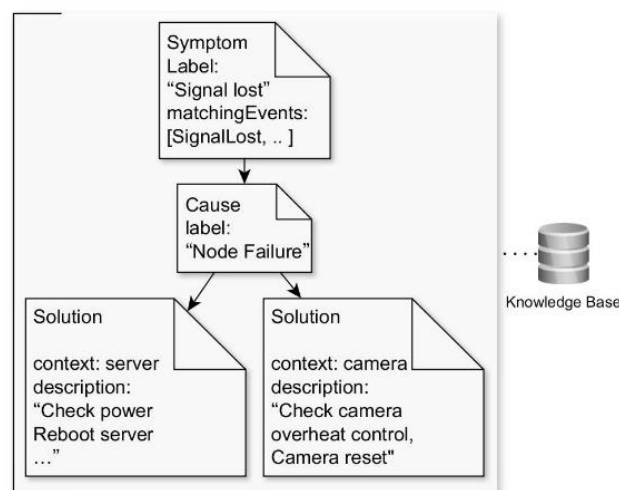


Figure 8: SCS model with example data

For instance, the previously generated semantic event ‘SignalLost’ can be mapped onto the symptom ‘signal lost’ using the ‘matchingEvents’ attribute. For example, the ‘SignalLost’ symptom within the Knowledge Base has a ‘node failure’ cause with two different solutions.

In Figure 9, the smart document retrieval process is shown, consisting of a ‘semantic symptom matching’ process and a ‘ranking and filtering’ process.

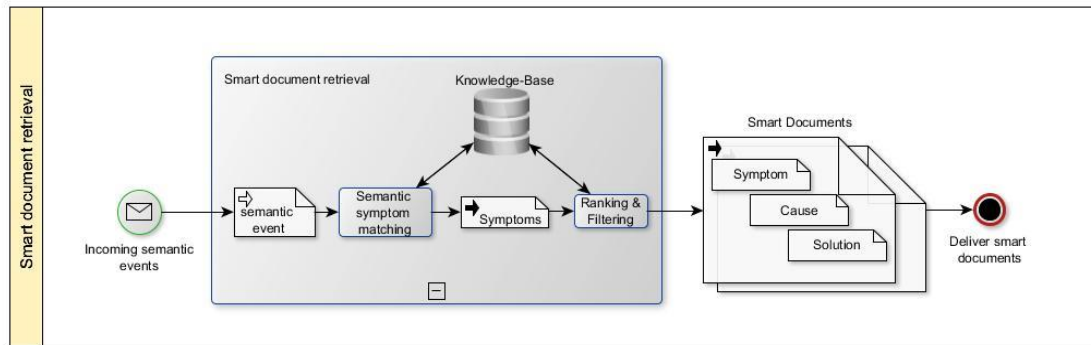


Figure 9: Smart document retrieval process

The semantic symptom matching process queries the SFP-CPS knowledge base. A query on the ‘matchingEvents’ attribute to determine the right symptom is such an example. Multiple symptoms can have the same ‘matchingEvent’ like ‘SignalLost’. In order to provide suitable smart documents, the symptoms get filtered by semantic context. A ranking can be applied according to filtered causes and solutions. For example, the likelihood of a problem / cause / solution in past usages may be used as ranking criterion.

3.6. Prototypical Implementation

The SFP-CPS has been implemented prototypically for the application use case of glass quality inspection. See Figure 10.

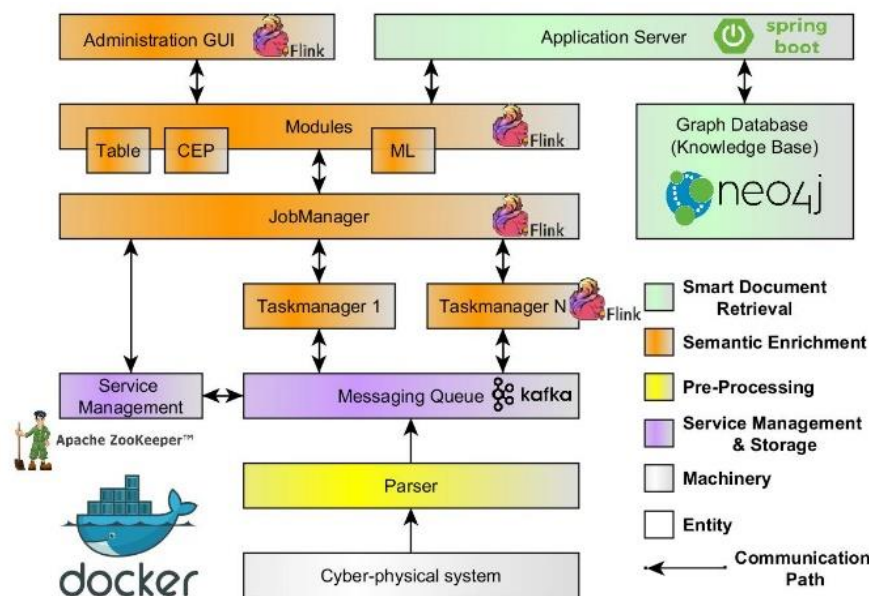


Figure 10: Prototypical implementation of the semantic fusion process

The parser pre-processes all incoming log events from CPS and forwards the processed and normalized log events into the Apache Kafka queue. Multiple Apache Flink tasks consume the queue and perform semantic enrichment, with filtering, pattern matching as well as value and time progression analysis. Job manager and task managers are parts of Apache Flink and enable for distributed parallel processing of the normalized events. The normalized events are

deposited in Apache Kafka queue. Through Flinks' job manager administration GUI, new semantic enrichment tasks can be added. The Spring Boot application server consumes the enriched events and queries the Neo4j graph database (knowledge base) for matching documentation. Parser, Apache Flink Jobmanager, Apache Flink Taskmanager, Apache Kafka with Apache Zookeeper, Neo4j and Spring Boot application server are shipped in individually configured docker containers. Docker enables upscaling the implementation, from a local development to multi server environment used for production. For instance, through Docker the prototype is cloud-ready and can be distributed to a private or hybrid Platform as a Service (PaaS) (Mell & Grance, 2011) cloud provider such as Amazon Web Services or Google Cloud Platform. An example user interface supporting the SFP-CPS and closer insides regarding the knowledge base can be found in Beez et al. (2017).

4. Evaluation

The SFP-CPS should assist machinery personnel in problem situations by providing highly relevant documentation containing solutions. For evaluation purposes, the requirements are compared to the SFP-CPS and its prototypical implementation (Section 3).

- R1.*Suitable documentation*: The SFP-CPS provides smart documents via the component 'smart document retrieval' (Section 3.5). Whether those documents are, indeed suitable and assist personnel in a smart factory in solving problems with CPSs, needs to be evaluated in a field study. This is subject to future work.
- R2.*Right-on-time*: The SFP-CPS as described in Section 3.2, continuously listens to raw data events generated by the CPS, processing them and retrieving suitable smart documents. The processing time from raw data to smart documentation has been measures for the prototype implementation, described in Section 3.6. In a Lenovo X1 Carbon (v. 2014, i5-4210U @ 2,4 GHz, 8 GB RAM), the total processing time was below 200ms for sample raw data events. This indicates that R2 can be met for production scenarios. As the SFP-CPS is an assistance process for factory personnel, 200 ms are an acceptable time span to provide guidance in a problem situation. Because the SFP-CPS does not interact with other machines, provide life safety functions or trigger emergency shutdowns, a shorter reaction time is not necessary. However, different problems of different problem criticalities in production scenarios need to be evaluated further.
- R3.*Applicable*: The SFP-CPS provides a process in which application domain specific business logic can be plugged in. In pre-processing, custom parsers can be implemented. In semantic enrichment, custom filters, pattern matching rules, value and time progression analysis rules can be provided. The knowledge base must be filled with concrete symptoms, causes, and solutions. To verify R3, the SFP-CPS has been implemented prototypically for two application domains: glass quality inspection as described in Section 3.1, and robot application development. Both implementations have been successful and indicate high potential for productive use.
- R4.*Event-triggered*: As shown in Figure 2, the SFP-CPS is triggered by CPS raw data events. Therefore, R4 is met.
- R5.*Based on raw data*: As shown in Figure 2, the SFP-CPS is filled with raw data from the CPS. Furthermore, the technology used in the prototype does not rely on any data format or data type. The pre-processing step (Section 3.3) standardizes all incoming log events that can also be updated by new log event formats.

- R6. *Non-interfering*: The SFP-CPS listens passively to log events without any impact on functional part of the CPS, in the example use case the glass inspection machinery. The SFP-CPS does not affect certified software modules and hardware modules. Therefore, current modules do not require a cost and time intensive recertification.
- R7. *Extensible*: Extensibility is enabled by a modular process and sub process architecture. The technologies used in the prototypical implementation also support this modular system architecture. There are four components: pre-processing, semantic enrichment, smart document retrieval and knowledge representation. Each of the four components is extensible. In the pre-processing component, new parsers can be applied for e.g. new input formats. In the semantic enrichment component, new rules can be integrated. In the smart document retrieval component, the ‘semantic symptom matching’ component as well as the ‘ranking and filtering’ component can be modified to improve the output quality and runtime performance. Finally, the knowledge base can be enhanced by new solutions and new domain information, such as symptoms and causes.

5. Related Work

The authors identified two main streams in the research domain of decision support systems for CPSs. On the one hand, ontologies are often employed as the main methodology to map error events to a knowledge base and thus to reason solutions, as presented for example by Bunte et al. as well as Petnga & Austin. On the other hand, complex event processing is used to enrich events to find solutions, see e.g. (Wang et al., 2016).

Both, Bunte et al. as well as Petnga & Austin introduce a framework (Bunte et al., 2016; Petnga & Austin, 2016), which can be interpreted as following the closed world assumption (Elçi et al. 2008). Both start with a pre-filled machine ontology. Both frameworks employ the web ontology language (OWL) for formalising machine events and infer helpful information. However, Bunte et al. split their ontology into a semantic instance model (SIM) with machine-specific knowledge (e.g. errors) and a semantic meta model (SMM) for common knowledge (e.g. solutions) (Bunte et al., 2016). Petnga and Austin perform a similar approach and split their ontology into domain specific ontologies (DSO) with machinery information (e.g. errors) and a decision support knowledge bases (DSKB), covering common knowledge (e.g. solutions) (Petnga & Austin, 2016). Since all mapping information is modeled into the SIM/DSO (Bunte et al., 2016; Petnga & Austin, 2016), both approaches lack the pre-processing step. With extensibility in mind, the SFP-CPS is designed to be enhanceable, especially pre-processing, semantic enrichment and smart document retrieval with knowledge base. The component-based approach has the advantage of more extension points within and in between process steps. An informal interview with domain experts with anecdotic evidence shows that the component-based approach can use synergy effects, because the experts can stick to their domain. For example, machine professionals can edit knowledge base entries and application developers can program pattern matching rules.

(Wang et al., 2016) introduce a context-aware semantic complex event processing framework (CSCEP). Their framework interconnects an OWL-based ontology with complex event processing. We adapt the idea of context-aware complex event processing, accommodate it in pre-processing (context attribute), semantic enrichment (context-aware algorithms) and smart document retrieval (knowledge base attribute) of SFP-CPS. In contrast to their work, we apply the context to a realistic application use case, here glass quality inspection. We provide a concrete concept for structuring the knowledge base (SCS model). Furthermore, we evaluate the feasibility of the approach via measurements of a realistic prototype implementation.

6. Conclusions and Future Work

The authors presented the SFP-CPS to assist machinery personnel in problem situations at a CPS in the glass quality inspection domain. As stated in Section 3, the SFP-CPS converts raw data into semantic events and assembles smart documents on top of the SCS model automatically. Due to its automation, the SFP-CPS can be seen as an enabler for on-site professionals to shorten CPS outage times. Another key aspect of the SFP-CPS is its extensibility which is caused by its modularity, see Sections 3.2 and 4. The evaluation shows that the SFP-CPS does not interfere with existing workflows, because it listens passively to log events of already well-tested and certified software modules and hardware modules within a CPS without affecting them. This hints towards the broad applicability of SFP-CPS in the smart factory domain.

Future work includes a field study to verify if the documentation is, in fact, ‘suitable’ in productive application scenarios. Next, the machine metadata (configuration files, the current machine state and multiple sensor data) will be integrated into the semantic enrichment. In addition, machine learning is planned to be integrated into semantic enrichment and smart document retrieval. Therefore, it is necessary to obtain new raw event data, especially for unsupervised machine learning. A study on probabilistic reasoning in smart document retrieval is also planned for the near future. Håkansson et al. present multiple reasoning strategies for a CPS (Håkansson et al., 2015). A selection of these strategies will also be applied in the SFP-CPS.

It is intended to use the SFP-CPS productively in real smart application settings. May it help optimizing production flows in smart factories.

References

- Beez, U., Bock, J., Deuschel, T., Humm, B. G., Kaupp, L. (2017). “Context-Aware Documentation in the Smart Factory”. In: *Proceedings of the 2017 Collaborative European Research Conference (CERC 2017)*: Karlsruhe, Germany
- Beyerer, J., León, F.P. & Frese, C. (2016). *Automatische Sichtprüfung: Grundlagen, Methoden und Praxis der Bildgewinnung und Bildauswertung*, Springer-Verlag.
- Bunte, A., Diedrich, A. & Niggemann, O. (2016). “Integrating semantics for diagnosis of manufacturing systems”. In: *2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*. Available at: <http://dx.doi.org/10.1109/etfa.2016.7733721>.
- Elçi, A., Rahnama, B. & Kamran, S. (2008). “Defining a Strategy to Select Either of Closed/Open World Assumptions on Semantic Robots”. In: *2008 32nd Annual IEEE International Computer Software and Applications Conference*. Available at: <http://dx.doi.org/10.1109/compsac.2008.182>.
- Gorecky, D. et al. (2014). “Human-machine-interaction in the industry 4.0 era”. In: *2014 12th IEEE International Conference on Industrial Informatics (INDIN)*. Available at: <http://dx.doi.org/10.1109/indin.2014.6945523>.
- Håkansson, A., Hartung, R. & Moradian, E. (2015). “Reasoning Strategies in Smart Cyber-Physical Systems”. In: *Procedia computer science*, 60, pp.1575–1584.
- Hornung, R. et al. (2014). “Model-free robot anomaly detection”. In: *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Available at: <http://dx.doi.org/10.1109/iros.2014.6943078>.
- ISRA VISION AG (2015). *The NEW Standard In Float Glass Inspection FLOATSCAN-5D Product Line*, ISRA VISION AG. Available at: http://www.isravision.com/media/public/prospekte2013/Brochure_Floatscan_5D_Product_Line_2013-05_EN_low.pdf [Accessed June 13, 2017].
- Jebsen-Marwedel, H. & Brückner, R. (2010). *Glastechnische Fabrikationsfehler: “Pathologische” Ausnahmestände des Werkstoffes Glas und ihre Behebung; Eine Brücke zwischen Wissenschaft, Technologie und Praxis*, Springer-Verlag.

Mell, P.M. & Grance, T. (2011). *The NIST definition of cloud computing*, Available at: <http://dx.doi.org/10.6028/nist.sp.800-145>.

Mousheimish, R., Taher, Y. & Zeitouni, K. (2016). “Toward the Support of Challenging Service Level Agreements (SLAs) in Manual and Context-Dependent Activities”. In: *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)*. Available at: <http://dx.doi.org/10.1109/compsac.2016.209>.

Nuñez, D.L. & Borsato, M. (2017). “An Ontology-based Model for Prognostics and Health Management of Machines”. In: *Journal of Industrial Information Integration*. Available at: <http://dx.doi.org/10.1016/j.jii.2017.02.006>.

Petnga, L. & Austin, M. (2016). “An ontological framework for knowledge modeling and decision support in cyber-physical systems”. In: *Advanced Engineering Informatics*, 30(1), pp.77–94.

Pohl, K. (2016). *Requirements Engineering Fundamentals, 2nd Edition: A Study Guide for the Certified Professional for Requirements Engineering Exam - Foundation Level - IREB compliant*, Rocky Nook, Inc.

Saalmann, P. & Hellingrath, B. (2016). “Coordination in Spare Parts Supply Chains”. In: *2016 International Conference on Collaboration Technologies and Systems (CTS)*. Available at: <http://dx.doi.org/10.1109/cts.2016.0111>.

Wagner, E. (2008). *Glasschäden: Oberflächenbeschädigungen, Glasbrüche in Theorie und Praxis: Ursachen, Entstehung, Beurteilung*,

Wang, W. et al. (2016). “A framework for context-aware semantic complex event processing”. In: *2016 12th World Congress on Intelligent Control and Automation (WCICA)*. Available at: <http://dx.doi.org/10.1109/wcica.2016.7578689>.

A Weighted Hybrid Recommender System Approach for Product Configuration

Daniel Benz[†], Nils M. Ullmann^{††}, Gilles Wetzel[†], Artur Felic³, Spiros Alexakis^{†††}

[†] Karlsruhe University of Applied Sciences, Germany

^{††} Flensburg University of Applied Sciences, Germany

^{†††} CAS Software AG, Germany

e-mail: daniel.benz@cas.de, nils.ullmann@hs-flensburg.de, gilles@wetzel.lu, artur.felic@cas.de, spiros.alexakis@cas.de

Keywords: Recommender Systems, Product Configuration, Mass Customization

Abstract

Mass Confusion is known as an information overload in mass customization environments that is caused by raising product variety. Recommender Systems support users to find individual solutions, but need to consider the challenges of product configuration systems. In this paper, a weighted hybrid recommender system is presented that addresses the challenges of product configuration systems and mass customization requirements. Based on user behaviour aware collaborative filtering and constraint solving techniques, it enables product managers to influence given recommendations while dynamic weights influence the behaviour based on the progress, efficiency and confidence about data. Preliminary results show promising results, but uncover the need for further research in this area.

1. Introduction

Mass Customization aims at individualization and personalization of products and services to meet customer preferences while maintaining costs equivalent to mass production (Tiihonen et al., 2017). At the same time, individualization leads to huge product variety and an information overload for the user known as *Mass Confusion* (Huffman et al., 1998). Recommender systems (RS) can counteract the latter aspect by guiding the user towards an optimally personalized configuration.

Especially in the field of product configuration, classical recommender systems are rather unsuitable for recommendations due to domain complexity and the variety of options. In addition, consumer behaviour differs compared to online shops. Complex products are bought less frequently than simple, standardized products (Tiihonen et al., 2010). Ratings or shopping histories are usually unavailable. This leads to a Cold-Start problem in the domain of RS and makes it difficult to infer recommendations.

Collaborative Filtering (CF) is used in RS to recommend items based on the similarity of users, i.e. the users purchase history and ratings. The absence of a purchases history leads to a Cold-Start Problem and thus no recommendations can be generated. The advantage is the recommendations' quality due to complete knowledge of past transactions. The approach can be improved by including user behaviour, e.g. clickstreams (Cho et al., 2002).

Constraint-Based (CB) RS generates recommendations based on previously defined rules, which requires manual effort by domain experts. The cold-start problem can be countered, since

no historical data is needed. Additionally, the quality of the recommendations in case of sparse data is improved.

The hybridization of CB and CF aims to mitigate disadvantages of classical recommender systems by combining the advantages of multiple approaches. Scenarios for RS in product configuration are categorized in (Tiihonen et al., 2010) as follows:

1. Recommending product characteristics and features
2. Recommending base line products
3. Recommending complete configurations
4. Complement unfinished configurations
5. Recommending sub-configurations

2. Related Work

A surge of interest in recommender systems overall was caused by the Netflix prize, resulting in increased research endeavours in this area (Koren et al., 2009). The Netflix competition brought considerable advances, especially the introduction of complex machine learning models. Recent work in the field of recommender systems is focused on hybrid models that combine collaborative filtering and content based recommender systems. These approaches try to utilize the strengths of both approaches to limit the cold-start problem, data sparsity issues and increase the accuracy of predictions.

According to (Burke, 2002), hybrid systems combine several approaches in various ways. Burke discusses six hybridization techniques e.g. weighting, switching, and mixing to improve the recommendations' quality. The constraint-based approach is widely used for configuration systems (Felfernig et al., 2013) and extended by a weighted majority voter (in case of features which are not specified with user requirements) in (Felfernig and Burke, 2008). The work in (Fargier et al., 2016) takes a different approach, which can be used if there is no prior knowledge about the user and his requirements. They introduce a user-based CF recommender for product configuration where no manual effort is needed. (Zanker and Jessenitschnig, 2009) also tries to overcome the issue to personalize interaction with anonymous first-time visitors with a switching hybridization approach consisting knowledge-based and collaborative strategies. Our approach uses a dynamic weighting mechanism for the hybridization to combine CB and CF which adapts itself according to the available data and user requirements.

3. A Weighted Hybrid Recommender System Approach

Figure 1 illustrates the architecture of our approach. The highlighted modules constitute the hybrid weighted RS, combining a CB and a CF RS. The database stores user profile and sales history together with the product knowledge defined by domain experts.

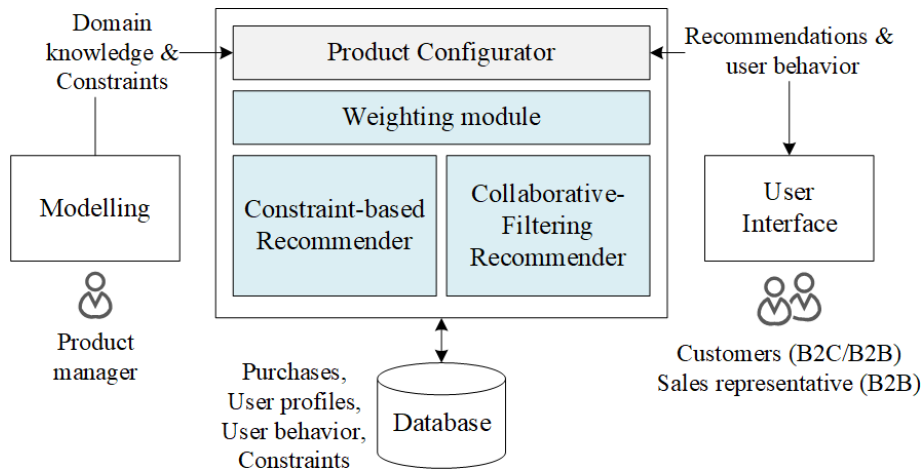


Figure 1: Architectural Overview of the Recommender System

The dynamically weighted linear combination of different recommender techniques takes the configuration progress, the efficiency of the constraints and the evaluation of a collaborative filtering into account. In contrast to static approaches like (Blecker et al., 2004), the weights are dynamically adapted during runtime.

The score R of an item i for a user u can be determined by the following equation:

$$R(u, i) = \alpha * CB(u, i) + \beta * CF(u, i) \quad \text{Equation 1}$$

$$\alpha + \beta = 1 \quad \text{Equation 2}$$

The Factors α and β are dynamic weights for the CB and CF approaches and are determined by:

- *Constraint Efficiency:*
Describes the fraction of all constraints that were satisfied based on historical configurations. Reduces the influence of poorly or incorrectly modelled constraints.
- *Interest Factor:*
Discriminates the CB approach if no interests are defined for the given user.
- *Completeness Factor:*
The further advanced the configuration, the more the CB approach is discriminated as the CF approach has more information to work with.

3.1 Constraint-Types

In order to compare users to identify the similarity, the presented approach is based on comparable user model features called interests. The set of interests is defined by the product manager (see Figure 1). Interests can be defined hierarchically to reflect the relationship between parent and child nodes. Weights, which represents their importance, can be attached to every node. Additionally, the relationships between these interests can also have weights which define the similarity between them.

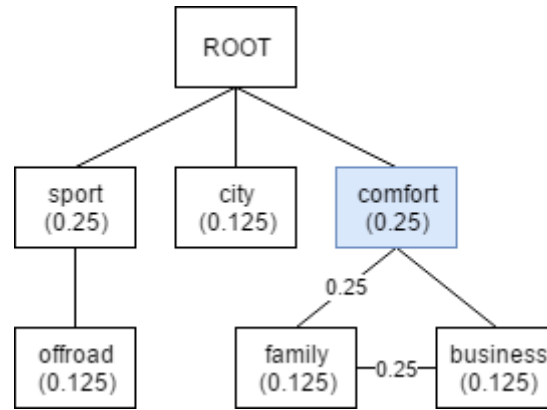


Figure 2: Example of an interest hierarchy

The recommender uses two different types of constraints:

1. *Mapping Constraints* map user information, usually collected from the user during usage of the software, to the interests to create a user model. For example, the *age* or *hobbies* of a user could give an insight on the interests *sport* or *comfort* (see Figure 2). Hence, the product manager is enabled to reflect the preferred mapping with this type of constraint.
2. *Recommendation Constraints* are used to calculate the recommendations and map interests to specific product features. They consist of a condition and an action part. The condition consists of a logical grouping of the interests and the action part specifies the characteristic recommendations or numerical restrictions for variables or numerical features of the configuration. For example, the identified interest *sport* in the condition part could be mapped to higher performance engines for car configurators in the action part.

3.2 Calculation of recommendations

In order to compute the recommendations from the CB, applicability and satisfiability scores are calculated from the condition and action apart of the Recommendation-Constraints. For that, aspects of Fuzzy Logic are used. The applicability and satisfiability scores represent the truth-values of the condition and action parts. The calculation of the applicability score also takes into account the weights of the interests. The satisfiability score indicates how well the Recommendation-Constraints are fulfilled. To get truth-values between $[0, 1]$, suitable t- and s-norms are used to analyse the logical groupings.

The maximum function is used for the s-norm:

$$T_v(p \vee q) = \max[T_v(p), T_v(q)] \quad \text{Equation 3}$$

The following equation is used for the t-norm:

$$T_v(p \wedge q) = \begin{cases} \min[T_v(p), T_v(q)] ; \text{if } T_v(p) > 0 \wedge T_v(q) > 0 \\ \frac{1}{\frac{1}{\max[T_v(p), T_v(q)]} + 1} ; \text{if } T_v(p) > 0 \vee T_v(q) > 0 \\ 0 ; \text{else} \end{cases} \quad \text{Equation 4}$$

In constraint-based recommender systems, the lack of recommendations in situations where item constraints are missing, it is a common situation that no items can be recommended. But it is possible that there are items, which at least fulfil some of the constraints or they satisfy them partially. With the help of these s- and t-norms, the partial applicability and satisfiability is possible to minimize missing recommendations.

3.3 Recommending product characteristics and features

By modelling *recommendation constraints*, interests of a user profile are mapped to suitable product characteristics or variable values (see *Figure 3*). The interests are defined hierarchically. A fuzzy-logic relaxation approach allows partially satisfied constraints during constraint solving to avoid an "over-constrained" rule base. The user-based CF approach exploits a configuration matrix instead of a user-item matrix. The matrix consists of built-in features augmented with additional information, i.e. user profiles and clickstreams. The normalized data, the interests and the configurations are weighted whereby an optimization is possible. A similarity matrix is compiled from the cosine distances, which are used as similarity measure. The recommendations of characteristics and features can be used to complement unfinished configurations.

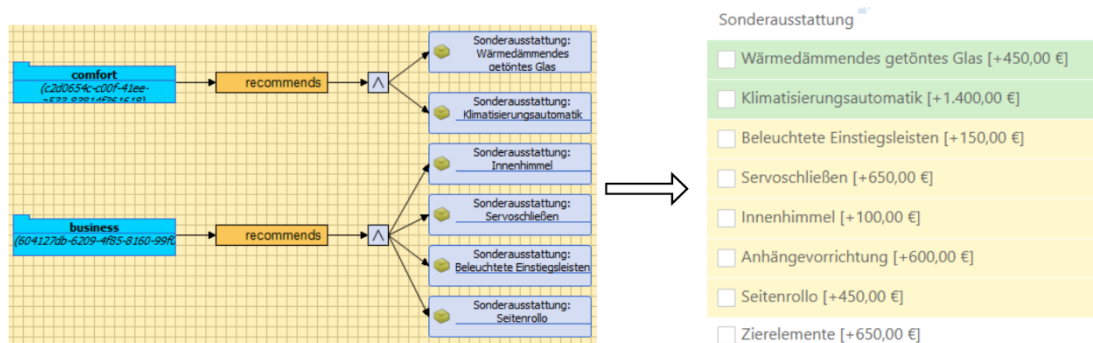


Figure 3: Example for Recommender Constraints

3.4 Recommending base line products

Recommender constraints between user interests and sales products can be modelled. During configuration, a score indicates the fit of a base product to a user. A user-item matrix, reflecting all users and purchased items, provides the best fitting base product (see *Figure 4*). When user data is unavailable at start, the modelled recommender constraints will be dominant, whereas the influence of the user-item matrix will rise with the amount of generated information. Depending on the domain, several algorithms can be used on the user-item matrix. Commonly leveraged are nearest-neighbour algorithms or matrix factorization. The former calculates the similarity among users by exploiting their purchase behaviour as well as demographic

information, in case they are available. Therefore, a similarity measure like the cosine similarity is used. The cosine similarity fits the problem domain, because only non-zero dimensions of the vectors have to be considered. In many cases, the user-item matrix is quite sparse, since most users only purchase a few products of the entire product catalogue. Let \bar{X} and \bar{Y} be two user-profiles with n features, then the cosine similarity is given by the following equation:

$$Sim_{cos}(\bar{X}, \bar{Y}) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad \text{Equation 5}$$

Matrix factorization, alternatively, decomposes large user-item matrices into a product of smaller matrices that have smaller dimensions and thus reduces the calculation cost at run-time. The product of matrices can be used to predict a specific rating for a user-item combination.



Figure 4: Recommendations of Base Products

3.5 Recommending complete configurations and sub-configurations

Complete configurations in the database are analysed for (partial) recommender constraint fulfilment. Scores that indicate the fit between users and complete configurations (see Figure 5) as well as sub-configurations are calculated. Complete configurations in the database are matched with current configuration steps. Configurations must match the current sub-configuration with all configured characteristics. Additionally, further features (i.e. user profile and interests) must also be as similar as possible.

Auto16			
10. August 2017			
Preisart: 39.860,00 €			
Ratings:			
	Rating	%	Info
comfort	1,08	70	5+ (0)/6+ (1)
sport	0,7	46,67	4+ (2)/10+ (0)
family	0,91	62,5	4+ (0)/5+ (1)
Score: 58,48			
Utility: 1,78			

Figure 5: Example Scores for Complete Configurations

3.6 Rating of configurations

In order to recommend configurations by the CB, they are rated to show how much they satisfy each of the interest dimensions. For that, the applicability and satisfiability scores are used. With the following formula, which is similar to the formula used in (Peischl et al., 2010), such scores are calculated. C is the set of Recommendation-Constraints, int is an interest and $Config$ is a configuration. The function $applscore$ calculates the applicability score and the function $satscore$ calculates the satisfiability score.

$$score(C, int, Config) = \frac{1}{\sum_{\forall c \in C_{app}} applscore(c, int)} * \sum_{\forall c \in C_{app \wedge sat}} satscore(c, Config) \quad \text{Equation 6}$$

The next step to recommend configurations is the personalization of the scores by using the user model in order to get a user specific utility-score. e is the user weight of the interest i and s is the score of a specific interest which was calculated with the formula above.

$$utility(x) = \sum_{i=1}^n e_i s_i(x) \quad \text{Equation 7}$$

3.7 Characteristic recommendation while configuring a product

To get characteristic recommendations while configuring, the applicability scores of the Recommendation-Constraints are calculated with the given user model. If this value is positive, recommendations can be derived. With the applicability score also a prioritization is possible, if there are multiple recommendations for a feature or if there are oppositional recommendations.

3.8 Calculation of characteristic recommendations using user behaviour

In order to be able to analyse user behaviour, user-specific data must first be collected. For this purpose, you can look at the features that have been built-in (BI) and build-out (BO). Built-in describes a characteristic selection, a build-out a deselection of a specific characteristic.

Table 1: Clickstreams for given configuration.

	Event1	Event2	Event3	Event4	Event5	Event6	Event7
Conf1	BI _A	BI _B	BI _C	BI _D	BI _E	BI _F	BI _V
Conf2	BI _E	BI _C	BO _c	BI _A	BO _E	BI _B	
Conf3	BI _C	BI _V	BO _C	BI _D	BI _A	BI _G	
Conf _U	BI _A	BI _C	BI _D				

The following equation shows a distance calculation between two sequences S_1 and S_2 from clickstreams, adapted from (Wang et al., 2013).

$$D_n(S_1, S_2) = 1 - \frac{|T_n(S_1) \cap T_n(S_2)|}{|T_n(S_1) \cup T_n(S_2)|} \quad \text{Equation 8}$$

A first approach to the order in the sequences is to present the clickstream as a collection of continuous partial sequences of length n . These are called n -grams. The following table shows how the most similar clickstreams can be determined for configuration u . The set $[1, 3] \in \mathbb{N}$ for instance encompasses all permutations, like $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$.

Table 2: Similarity matrix between conf_u and conf_1 , conf_2 and conf_3 .

conf_u	Conf1	Conf2	Conf3	kNN = 2
n [1,3]	0.20	0.10	0.16	$\text{conf}_1, \text{conf}_3$
n [1,4]	0.16	0.09	0.14	$\text{conf}_1, \text{conf}_3$
n [2,4]	0.05	0	0	conf_1

3.9 Relaxation-approach based on the Interest-Hierarchy

As described above, a situation can occur that no recommendations can be calculated. The mechanisms to prevent these situations are called relaxation approaches. Normally, the set of constraints that have to be fulfilled is shortened or the constraints themselves are altered. This system uses an oppositional approach: The set of constraints which can be satisfied will be enlarged. For that, the Interest-Hierarchy is used. The set of interests which are activated in the user model is enlarged with similar interests to the already available interests. Also the Fuzzy Set Theory is used for that in a form similar to (Borràs et al., 2012).

Each interest of the Interest-Hierarchy represents a Fuzzy Set I_d , where the membership-function μ_{I_d} is abgeleitet from the Interest-Hierarchy and the relationship-weights. The instantiated Interests int_i from the user model are the elements, which are zugewiesen to the Fuzzy Sets according to the membership-functions. For each int_i there is an equivalent set I_d to which the interest can be zugewiesen with the membership-value 1. This set is called S_{int_i} . For the assignment to other sets, the membership-value is calculated with $\mu_{I_d}(\text{int}_i)$.

$$\mu_{I_d}(\text{int}_i) = \begin{cases} 1; & \text{if } S_{\text{int}_i} = I_d \\ \text{weight}_{\text{sibling}}; & \text{if } \text{siblings}(I_d, S_{\text{int}_i}) = \text{true} \\ \text{weight}_{\text{parent}}; & \text{if } \text{parent}(I_d, S_{\text{int}_i}) = \text{true} \\ \text{weight}_{\text{child}}; & \text{if } \text{child}(I_d, S_{\text{int}_i}) = \text{true} \\ 0; & \text{else} \end{cases} \quad \text{Equation 9}$$

The weights and membership-values for each user interest, which were assigned to a set, are added, so that each set I_d gets a value. A Fuzzy Set I_d of an interest is “activated”, if the value is > 0 and the equivalent interest is considered by the calculation of recommendations by including the membership-values while calculating the applicability and satisfiability scores. With the help of this mechanism there are more Recommendation-Constraints to get characteristic recommendations and Recommendation-Constraints which contribute to the user specific score for a configuration for the calculation of configuration recommendations.

4. Research Methodology

In the first iteration, we designed and implemented standalone prototypes for constraint-based and collaborative filtering RS for product configuration. For each RS, we reviewed the

literature, formulated research questions with respect to product configuration and did qualitative and quantitative evaluations. The results have shown that none of the standalone RS fully covers a suitable RS for product configuration. Hence, in the second iteration, we reviewed the literature for hybrid recommender systems. Although there are many approaches for E-Commerce, open questions arose from a product configuration perspective. How can recommender systems be dynamically combined to leverage the advantages of classic RS for product configuration? Hence, we designed and prototyped a hybrid RS with dynamic weights. At present, we are at prototyping stage and heading towards a qualitative evaluation via interviews with focus groups and a quantitative evaluation utilizing known measures for recommender systems, i.e. precision, recall and the F-Measure, for reproducibility.

5. Preliminary Results

We have achieved promising results using internal data sets, which need further validation by larger data sets. Furthermore, the results have shown that our approach performs at least as good as CB or CF in product configuration, because in edge cases (*no data* - only CB or *massive data* - only CF) it falls back to one of these approaches. Overall, our results were always better than random. The use of constraint-based recommender stays dependent on the product manager's knowledge who defines the constraints.

6. Conclusion and Future Work

We introduced a weighted and hybrid recommender system approach based on collaborative filtering and a constraint based recommender technique. The cold start problem is solved by constraint-based approach. With the rising amount of data during runtime, collaborative filtering improves the recommendation quality. Dynamic weighting is applied to realize dynamic adaption during runtime.

Further research is required for solving the issue regarding changing product knowledge base. Old configurations cannot be used in collaborative filtering since they are incoherent with current constraints. Additionally, a content-based recommender could increase the prediction quality.

7. References

- Blecker, T. and Di Kreutler, G. (2004). "An Advisory System for Customers' Objective Needs Elicitation in Mass Customization", in: *Proceedings of the 4 th Workshop on Information Systems for Mass Customization (ISMC 2004) at the fourth International ICSC Symposium on Engineering of Intelligent Systems (EIS 2004)*, University of Madeira Funchal, Portugal, February 29th - March 2, 2004. ICSC Interdisciplinary research, 2004.
- Borràs, J., Valls, A., Moreno, A. and Isern, D. (2012). "Ontology-Based Management of Uncertain Preferences in User Profiles", in: *Advances in Computational Intelligence. 14th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU 2012, Catania, Italy, July 9-13, 2012. Proceedings, Part II*, vol. 298. Springer, Berlin/Heidelberg; pp. 127–136.
- Burke, R. (2002). "Hybrid Recommender Systems: Survey and Experiments", in: *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331–370.
- Cho, Y. H., Kim, J. K. and Kim, S. H. (2002). "A personalized recommender system based on web usage mining and decision tree induction", in: *Expert Systems with Applications*, vol. 23, no. 3, pp. 329–342.

Fargier, H., Gimenez, P. and Mengin, J. (2016). "Recommendation for product configuration. An experimental evaluation", in: *Elise Vareilles, Lars Hvam and Cipriano Forza (Hg.): 18th International configuration workshop. Proceedings of the 18th International Configuration workshop within CP 2016 conference, September 5-6, 2016, Toulouse, France*. École des Mines d'Albi-Carmaux; pp. 9–16.

Felfernig, A. and Burke, R. (2008). "Constraint-based Recommender Systems. Technologies and Research Issues", in: *Proceedings of the 10th international conference on Electronic commerce*. New York, NY, USA: ACM (ICEC '08), 3:1-3:10.

Felfernig, A., Jeran, M., Ninaus, G., Reinfrank, F. and Reiterer, S. (2013). "Toward the Next Generation of Recommender Systems. Applications and Research Challenges", in: *Multimedia Services in Intelligent Environments: Advances in Recommender Systems*. Springer International Publishing, Heidelberg; pp. 81–98.

Huffman, C. and Kahn, B. E. (1998). "Variety for sale. Mass customization or mass confusion?", in: *Journal of retailing*, vol. 74, no. 4, pp. 491–513.

Koren, Y., Bell, R. and Volinsky, C. (2009). "Matrix factorization techniques for recommender systems", in: *Computer*, vol. 42, no. 8, pp. 30-37, Aug. 2009.

Peischl, B., Zanker, M., Nica, M. and Schmid, W. (2010). "Constraint-Based Recommendation for Software Project Effort Estimation", in: *JETWI* vol. 2, no. 4, pp. 282–290.

Tiihonen, J. and Felfernig, A. (2010). "Towards recommending configurable offerings", in: *International Journal of Mass Customisation*, vol. 3, no. 4, pp. 389–406.

Tiihonen, J. and Felfernig, A. (2017). "An introduction to personalization and mass customization", in: *Journal of Intelligent Information Systems*, vol. 49, no. 1, pp. 1–7.

Wang, G., Konolige, T., Wilson, C., Wang, X., Zheng, H. and Zhao, B. Y. (2013). "You Are How You Click. Clickstream Analysis for Sybil Detection", in: *USENIX Security Symposium*, vol. 9, pp. 1–8.

Zanker, M. and Jessenitschnig, M. (2009). "Case-studies on exploiting explicit customer requirements in recommender systems", in: *User Model User-Adap Inter*, vol. 19, no. 1-2, pp. 133–166.

New Approach in 3D Mapping and Localization for Search and Rescue Missions

Sven Brenner, Sebastian Gelfert, Hendrik Rust
Faculty of Management Science and Engineering
Karlsruhe University of Applied Sciences, Germany
e-mail: sebastian.gelfert@hs-karlsruhe.de

Keywords: search and rescue robots, multi-sensor system, 3D mapping

1 INTRODUCTION

Recent catastrophes such as the Fukushima nuclear disaster in 2011 or Nepal earthquake in 2015 showed that Search And Rescue (SAR) robots can be used to efficiently support rescue teams in finding persons in danger or gathering information more effectively. In Fukushima emergency responders deployed a SAR robot, to check on conditions in the surroundings and allowed workers a safe distance from hazardous radiation (Roman, 2015).

Due to the fact that SAR robots are nowadays mostly tele-operated, they can reach dangerous places that rescue teams cannot. In the future more and more semi-autonomous or autonomous SAR robots will be develop.

Figure 1 shows a layout of such a scenario. On communication level, commands by the rescue team or the operator will be send via a man-machine interface (MMI) to a SAR robot, located in a hazardous environment. In this scenario, the mission of SAR robots is finding persons in danger. Position data and ambient data (e.g. temperature) will return simultaneously to the rescue team or the operator (Rust, 2001).

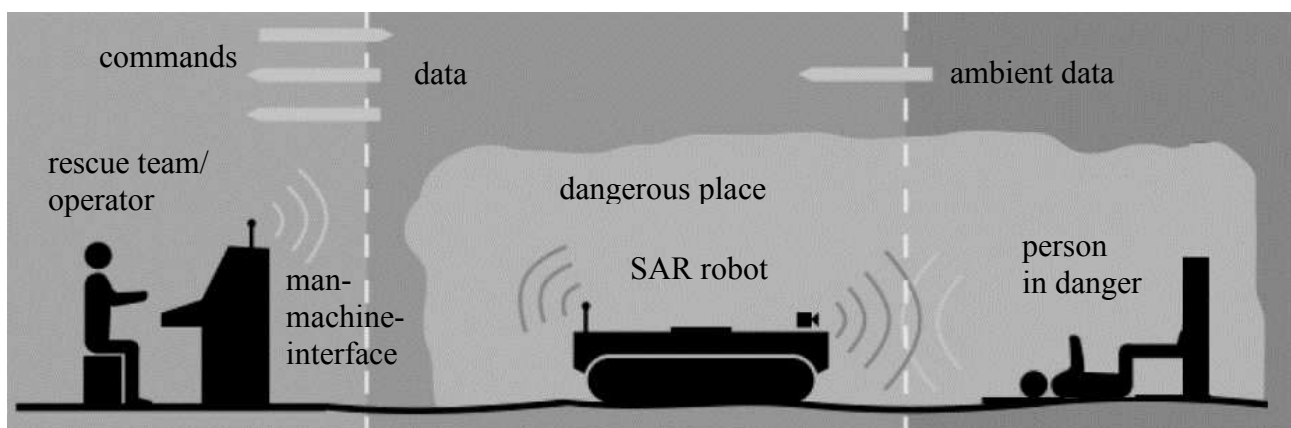


Figure 1: Layout of scenario (Rust, 2001).

To accomplish a successful SAR mission, a perfect communication as well as a good indoor mapping and localization approach is required.

Since more than 20 years, the main challenge is acquiring information about the disaster in harsh environmental conditions. These conditions are designated by smoke, dust or chemical steam.

Until now, several SAR robots are under development and some have been used in emergency situations. Figure 2 shows the first serial robot type ZELOS R&D developed by ACTIVEROBOTICX. The SAR robot has a chain drive, uses a 4 Ah lithium ion battery as drive unit and has a weight of merely 25 kilogram. Beside the SAR robot an industrial radio control unit constitutes the MMI. This platform is a very robust, tested and cost effective basis for the described scenarios.



Figure 2: SAR robot.

The aim of this paper is to provide an insight into a new comprehensive low-cost approach in accurate real-time 3D indoor mapping and localization in SAR missions with high/medium visibility. It uses a multi-sensor system in combination with local and/or cloud-based data processing. Furthermore, the focus is on quantitative results at different transmittances.

2 RELATED WORK

Area mapping and localization for SAR missions is a research topic since many years, made good progress in terms of performance and has been extensively discussed in the literature.

A low bandwidth radar-based scanning-technology for mapping is presented by Fritsche and Wagner (2015a) and Fritsche and Wagner (2015b). The radar technology is capable of showing a substantial mapping quality, but there is a need for improvement regarding resolution of the radar sensors and in addition it is only available in 2D. Another disadvantage is, that this approach is very cost-intensive compared to other sensor technologies.

Another work of accurate indoor localization is given by Winterhalter et. al. (2015). They also show how to 2D outline the environment using imaging data. This cost-efficient approach is not suitable in harsh environmental conditions.

Langerwisch et al. (2014) shows how to construct environmental 3D models generated by fusing ground-based and aerial light detection and ranging cloud data. The results are very promising, but further investigations, on the accuracy of the sensor data fusion, have to be done. Indeed, this approach is cost-efficient, not suitable in harsh environmental conditions but usable in environments with high/medium visibility.

Within this paper, a new comprehensive low-cost approach of accurate real-time 3D indoor mapping and localization in environments with high/medium visibility is discussed. It uses a multi-sensor system in combination with local and/or cloud-based data (on an external server) processing and merges imaging and laser technology.

With this approach no satellite navigation such as the Global Positioning System (GPS) is necessary.

3 NEW APPROACH AND CONTRIBUTION

In this work, the challenge of accurate real-time 3D mapping and localization in environments with high/ medium visibility will be discussed. The main contribution of this new comprehensive low-cost research approach is the combination of a multi-sensor system with local and/or cloud-based data processing at these conditions.

The multi-sensor system consists of a Lenovo Phab 2 Pro with Tango technology:

- High resolution 16 mega pixel camera
- Depth sensing, near infrared element
- Wide angle motion tracking camera (135 degree vertical, 115 degree horizontal opening angle)
- Accelerometer
- Gyroscope data (Lenovo, 2017)

All elements are combined to a small and integrated low-cost system compared to other sensor system configurations.

Figure 3 shows the SAR robot equipped with the multi-sensor system.

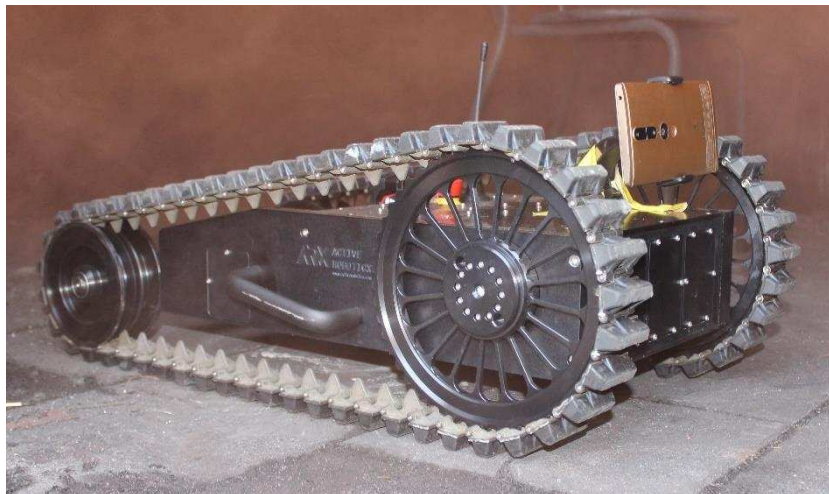


Figure 3: SAR robot with multi-sensor system.

The approach consists of three steps:

In the first step “3D indoor mapping” the multi-sensor system will be used to accurately 3D map the environment until a person in danger is localized or the environment is mapped completely (to find hazard sources). The motion tracking technology in the multi-sensor system hereby provides the orientation and position in the space. Generally the orientation and position of the system is called pose. On the basis of prominent spots in the environment, the system tracks its position relatively to its origin. In addition, a depth sensing near infrared element measures the emissive and inclined beams for distance determination.

In the second step “local and/or cloud-based data processing” the data of the 3D map will be processed, memorized and provided with distinctive characteristic attributes. The data of the pose consists of a vector for movement as well as a quaternion for determination of rotation.

In the third step “localization” a second multi-sensor system is used to match the processed 3D map by comparing real and memorized information. The multi-sensor system recognizes familiar environmental points by constellation of different area points. The position can be evaluated without re-mapping the environment. A fast area recognition can be accomplished by archiving the data via random access memory (RAM).

The multi-sensor system can be fixed on SAR robots or handheld equipment for rescue forces. The path, to a localized person in danger, can be notified by augmented way-points or way-arrows. With this comprehensive approach a SAR mission scenario could be: SAR robots are searching for persons in dangerous places and rescue teams will be guided to the localized person via augmented information. In summary this visual positioning system runs with sensors instead of trilateration methods like GPS.

The research method in the presented new approach consists of a series of experiments with the multi-sensor system and comparable measurements at different transmittances. No simulation approach was chosen, because usually the environmental conditions are too complex and simulations are far away from realistic behavior. The main objective of this approach is, that the results of the experiments getting comparable at realistic environmental settings.

4 EXPERIMENTAL SET-UP

Figure 4 shows an overview of the experimental set-up. The test room, 5 meter long and 4 meter wide, is used as an exemplary scenario for a SAR mission. It is equipped with predefined home furnishing objects (wall unit, dining table with four chairs, couch) in different dimensions (black). The room has a total volume of 60 cubic meter. The multi-sensor system is moving with a speed of 0.1 meter per second along the dotted path at a height of approximately 1 meter. To reach the desired transmittance, a fog machine (type Mc CRYPT A-900) with a fog output of 110 cubic meter per minute is installed in one corner. The following transmittances are set: 100 %, 70 %, 50 %, 40 %.

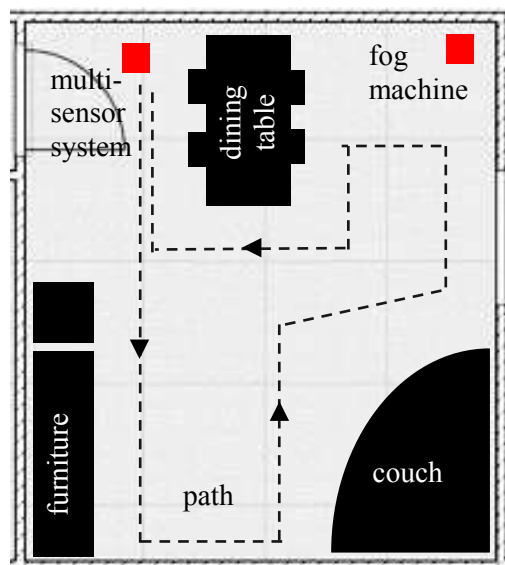


Figure 4: Overview of experimental set-up.

Figure 5 shows an image of the experimental set-up in bird's eye view. Furthermore the image shows, that the test room is very tight and there is no much room to maneuver.



Figure 5: Image of experimental set-up.

5 RESULTS

Figure 6 and 7 shows the results of the final 3D indoor map after data processing at different transmittances. At a transmittance of 100 %, the range of visibility proceeds towards unlimited. Rust (2001) defines the range of visibility with object detection, but he doesn't give any definition about the object identification. The conducted experiment at foggy environment shows that at a transmittance of 70 % the range of visibility is greater than or equal to 4 meter. In this fog contaminated environment, object detection is reasonably applicable with infrared sensing. The results are getting fuzzy as well as not quantifiable at transmittances less than 50 %. The range of visibility proceeds between 0 to 4 meters.

In summary the map shows proper object information in comparison with other 3D mapping research approaches like radar technology with point cloud. At transmittances greater than 70 %, tele-operated navigation and object detection is possible in environments with high/medium visibility. In return, navigation and object detection is not possible at transmittances less than 70 % with the presented multi-sensor system. For this purpose other sensor technologies are necessary, e.g. radar technology. Furthermore an intuitive navigation and simple object detection is feasible by the 3D image with photo-realistic texture. Improvements can be done by reducing the period of time for the 3D indoor mapping process (moving speed less than 0.1 meter per second) as well as a systematic and methodic area motion of the multi-sensor system. Reflection, brightness and darkness are further identified challenges during the mapping process.



100 % transmittance
Range of visibility: ∞

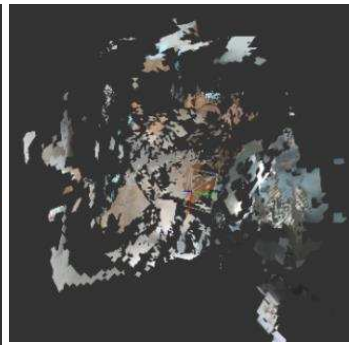
Figure 6: Experimental results with different transmittances (3 meter above ground).



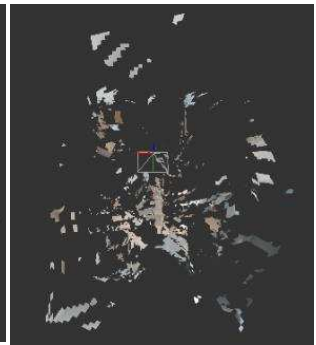
100 % transmittance
Range of visibility:
 ∞



70 % transmittance
Range of visibility:
 ≥ 4 meter



50 % transmittance
Range of visibility:
2 - 4 meter



40 % transmittance
Range of visibility:
0 - 2 meter

Figure 7: Experimental results with different transmittances (top view).

6 CONCLUSION AND FUTURE WORK

In this research approach, the challenge of accurate real-time 3D indoor mapping and localization in SAR missions with smoke contaminated environments is faced. A new low-cost approach is introduced, that uses a multi-sensor system based on a Lenovo Phab 2 Pro with Tango technology in combination with local and/or cloud-based data processing to guide rescue teams. The first results show that this approach is reasonably applicable in environments with high/medium visibility, especially where chemical substances need to be detected.

Nevertheless, the approach leads to problems with regards of no mapping results (grey zones). In addition, the uses to additional sensors has to be examined. A suitable sensor could be radar sensor, because this technology, due to the frequency, is insensitive against larger particles in environments with high/medium visibility. By fusing the radar technology with infrared thermography, a new approach is generated. Moreover, the robustness of the multi-sensor system at realistic environmental conditions has to be assured as well as an intuitive man-machine-interface.

Another topic to work on is the integration of the system in SAR robots (e.g. mechanical and electronical integration, data interface). Additional effort has to be spent on minimizing the unit in terms of size and usability for rescue teams. Based on the data of the multi-sensor system, the evolution of SAR robots in regard to autonomous functions from tele-operated to semi-autonomous/autonomous systems need to be examined.

References

- Fritsche, P. and Wagner, B. (2015a), "Scanning Techniques with Low Bandwidth Radar for Robotic Mapping and Localization," in *Informatics in Control, Automation and Robotics 12th International Conference (INCINCO 2015)*. IEEE, 2015; pp. 321-335.
- Fritsche, P. and Wagner, B. (2015b), "Comparison of two radar-based scanning-techniques for the use in robotic mapping," in *Informatics in Control, Automation and Robotics 12th International Conference (INCINCO 2015)*. IEEE, 2015; pp. 365-372.
- Lenovo (2017), Operating Manual of Lenovo Phab 2 Pro, Lenovo.
- Langerwisch et al. (2014), "Construction of 3D Environment Models by Fusing Ground and Aerial Lidar Point Cloud Data," in *International Conference on Intelligent Autonomous Systems (IAS 2014)*. IEEE, 2014; pp. 473-485.
- Roman, J. (2015), RISE OF THE MACHINES, NFPA Journal, July/August 2015; pp. 48-55.
- Rust, H. (2001), Objekterfassung in verrauchter Umgebungen für teleoperierte kettenbasierte Manipulatorfahrzeuge, University of Stuttgart.
- Winterhalter et al. (2015), "Accurate Indoor Localization for RGB-D Smartphones and Tablets given 2D Floor Plans," in *International Conference on Intelligent Robots and Systems (IROS)*. IEEE/RSJ, 2015; pp. 3138-3143.

The System Design and Implementation of V-Book System Based on AR Technology

Wu Jingqian[†], Mao Yufan^{††}, Wang Tianxin^{†††}, Cong Fuze^{††††}, Li Jiazhen^{††††}

[†] Beijing National Day School, China

^{††} School attached to Central Academy of Art and Design, China

^{†††} The Second High School Attached to Beijing Normal University, China

^{††††} Beijing University of Posts and Telecommunications, China

e-mail: 964391132@qq.com

Keywords: AR, iOS, Windows Azure

Abstract: The quality of the learning materials has a great impact on children and adolescents' learning effect and efficiency. The V-Book system based on augmented reality technology (AR) can combine text information, auxiliary video and other presentation information effectively. The teaching & learning community and online V-Book page editing are also included in the system. The expectation investigation from potential consumers and system testing have been completed for system design and verification.

1. Introduction

The quality of the learning materials has great impact on children and adolescents' learning effect and efficiency. At present, most of the children and adolescents learn from paper books (Ni, 2012). Some students may feel reluctant to learn because they think textbooks are boring and difficult to understand. When this happens, teachers and parents often help to find some video information, do experiments or give a demonstration to explain new knowledge to them. Our challenge is to find a way to rationally combine text information and auxiliary video or any other interesting materials to form a new type of book which is easy for students to understand.

Augmented Reality (AR) is a live view of a physical, real-world environment whose elements are augmented by computer-generated sensory input such as sound, video, graphics or Global Positioning System (GPS) data (Azuma et al., 2001; Chen et al., 2016; Wu et al., 2013; En.wikipedia.org, 2017). AR has been used in several areas, including advertising, games, and encyclopedias to make the contents more vivid. Therefore, there has been an increasing interest in applying AR technology into education field (Chen et al., 2016). Ibáñez, Castro and Delgado-Kloos. (2017) use AR technology in physics learning which suggested that AR technology has a positive impact on students' motivation. Dinis et al. (2017) develop AR game-based applications in civil engineering education which can improve learning effectiveness. However, the exiting AR systems are built for students' study, fewer projects focus on both teaching and learning by combining the vivid video with static text, and providing customizable AR service with information exchange system.

After the AR technology investigation and the expectations investigation from potential consumers, a new system is developed by applying AR technology to the teaching and learning field. The name of the system is V-Book. The evaluation of the system shows that it can improve the teaching and learning effect.

2. Expectations from Potential Consumers

Before developing, it is important to know the market demand. A questionnaire was developed to evaluate the requirements based on high school students' point of view. All the questions are designed to find out the potential consumer's attitude towards the technologies will be used in the system, which including the combination of video and paper materials, AR, online community. All the investigation participants are students, mainly high school students, ranging from age of 6 to 22. The number of participant students is 216, which is sufficient to conduct a valid result. The results of our investigation are shown in Fig. 1 and Fig. 2.

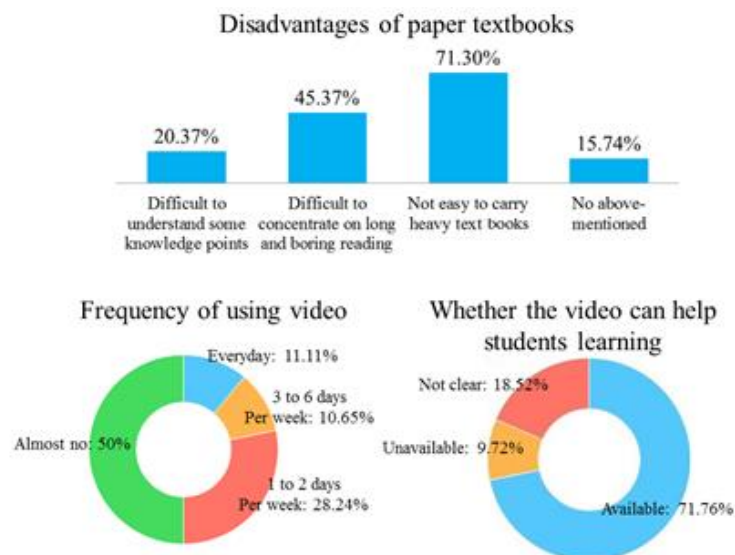


Fig.1: Results of the questionnaire 1 (participants number=216)

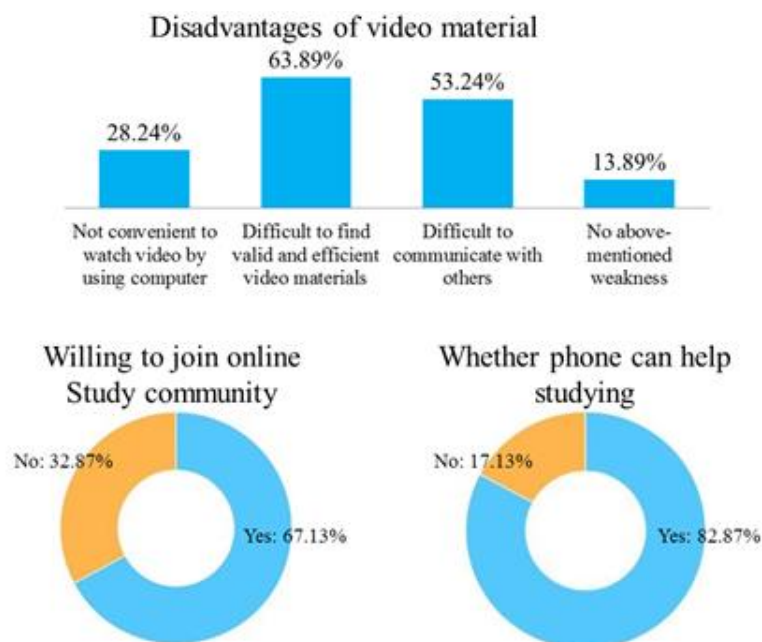


Fig. 2: Results of the questionnaire 2 (participants number=216)

The figures show that there are drawbacks for both paper and video material separately, and students are willing to use cell phone to study. Under such a background, we start our project, aiming to provide a better environment and suitable method of teaching and studying.

3. System Design and Implementation

V-Book system consists of four parts: V-Book page, iOS APP, web interface and server. In the paper book, some of the text will be replaced by the representative pictures, which leads to the augmented reality video. The system structure diagram is shown in Figure 3.

When students want to learn more details about knowledge, they can use their mobile phone APP to take photos of the picture on V-Book Page. The App on the iOS device recognizes the picture and sends requests to the system back end. Then the video related to the picture will be found in the database and sent back to the APP. After being received by the cellphone, the video will be played for detailed knowledge.

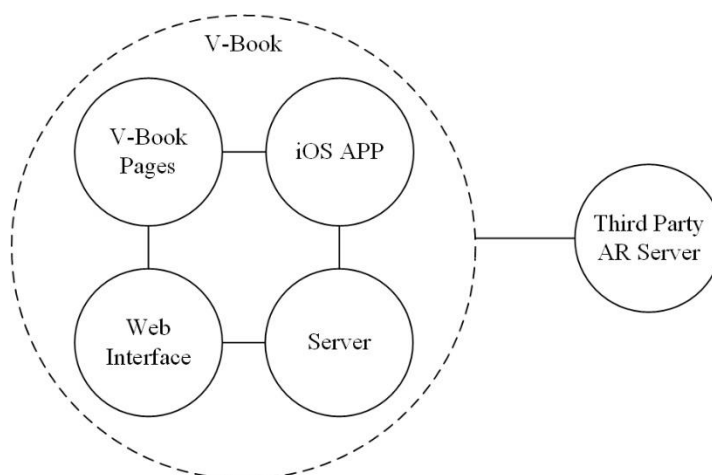


Fig. 3: The system structure of V-Book

Teachers can upload the text, the picture and video through the V-Book website. After being edited and confirmed by our team members, the information will be stored in the database. Then the V-Book pages will be automatically created and displayed on the website. Users can print the V-Book pages at any time.

The Web server structure is divided into three parts: front end, back end and database. In the front end, HTML and Cocoa are used. Flask framework and a third party library called QJARlib form the back end. SQLite is used for database design and implementation. The system data flow is shown in Figure 4.

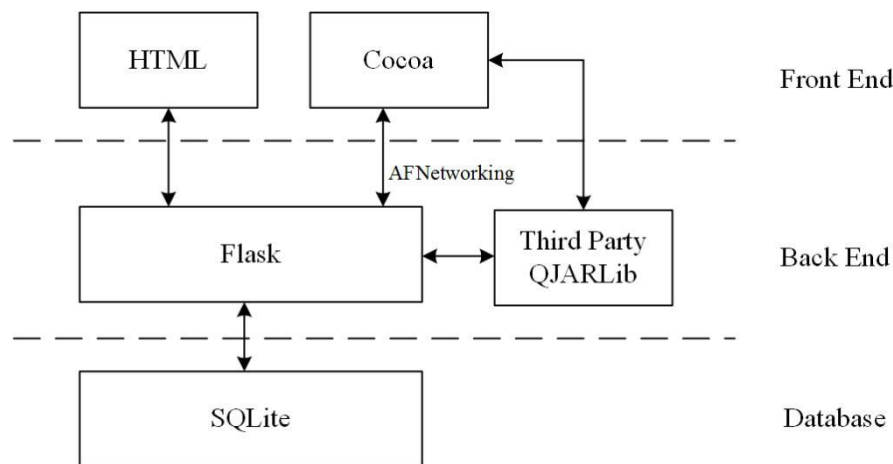


Fig. 4: The system data flow

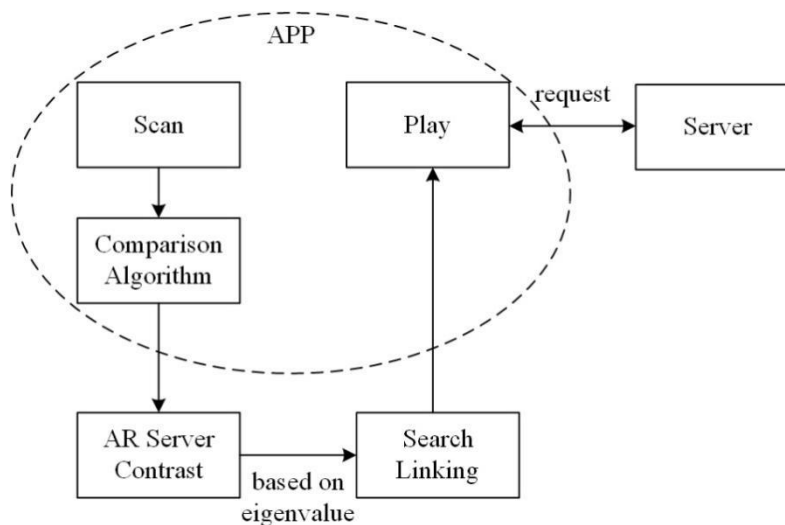


Fig. 5: The implementation of V-Book AR function

The implementation of V-Book AR function is shown in Fig. 5. When a user uses V-Book to scan a V-Book page picture, the App will call a comparison algorithm. After the algorithm has calculated an Eigenvalue, the calculated Eigenvalue is sent to AR server for comparison. This is a really vital process because only by doing that can the system know whether we are scanning the right picture and which picture it is. Then, the V-Book app requests the related video resource from web server, and the video will be sent back to the phone side finally.

Microsoft Azure is a cloud computing service created by Microsoft for building, deploying, and managing applications and services (Cusumano, 2010; Wall et al., 2012). Based on Azure, the service of online study community, video and uploading is provided. Concretely, the virtual machine service of Azure is used to deploy V-Book back-end application.

4. System Testing

The interface of the V-Book app is shown in Fig. 6. When users open the app for the first time, they need register and create an account. Users can directly log in after the registration. The second picture in Fig. 6 shows the manipulation interface. Click on the Camera button, users can enter the scan interface to scan V-Book pages and watch video afterwards. Clicking on the book store button in the

third picture, users can enter the interface of the V-Book shop. In the shop, there are many kinds of V-Book pages, including details about the book, such as the author, the main content, and the comments from other readers.



*: Book; **: Book Store

Fig. 6 The interface of the V-Book App

Our project has a clear target market and audience. V-Book is intended for teachers and students. In the future, we would collaborate with publishing houses for the publication of V-Book pages. V-Book accumulation mechanism can encourage teachers and students to upload their materials and share with others. A free public welfare online-study community and a complete commercial closed-loop between content provider and content user will be our future task.

5. Conclusion and Outlook

In this paper, the V-Book system has been developed and implemented based on AR technology. The system achieves a combination of the text, auxiliary video and other presentation information effectively and achieves online community and customizable AR service. The system testing was completed. However, the AR application relies on the high performance of mobile device, which, in turn, increases the cost of the system and makes it difficult to promotion. Besides, the online website should be designed as consumer friendly as possible so that it can maintain users using our product while attracting more potential consumers. So future works may include optimizing application on cheaper devices, the development of more AR content and the comment drop-down loading function. With the improvement mentioned above, V-Book can not only improve students' efficiency in study, but also help strengthen the linkage between the students and the teachers.

Acknowledgments

This work is supported by National Special Fund on Basic Research from Ministry of Science and Technology of China (2014FY110100 & 2014FY210200)

Bibliography:

- Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S. and MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, 21(6), pp. 34-47.
- Chen, P., Liu, X., Cheng, W., and Huang, R. (2017). A review of using Augmented Reality in Education from 2011 to 2016. In: *Innovations in Smart Learning*, pp. 13-18.
- Cusumano, M. (2010). Cloud computing and SaaS as new computing platforms. *Communications of the ACM*, 53(4), p. 27.
- Dinis, F., Guimarães, S. and Carvalho, R. (2017). Virtual and Augmented Reality game-based applications to Civil Engineering Education. In: *Global Engineering Education Conference (EDUCON), 2017 IEEE*. pp. 1683-1688.
- En.wikipedia.org. (2017). *Augmented reality*. [online] Available at: https://en.wikipedia.org/wiki/Augmented_reality [Accessed 2 Sep. 2017].
- Ibáñez, M., Castro, A. and Delgado-Kloos, C. (2017). An Empirical Study of the Use of an Augmented Reality Simulator in a Face-to-Face Physics Course. In: *In Advanced Learning Technologies (ICALT), 2017 IEEE 17th International Conference on Advanced Learning Technologies*. pp. 469-471.
- Ni, C. (2012). *An Investigation on the Learning Mode and Transformation of Middle School Students in the New Curriculum Reform*. Master. Huazhong Normal University.
- Wall, D., Kim, I., Jung, J., DeLuca, T., Nelson, T. and Wall, D. (2012). Cloud Computing for Comparative Genomics with Windows Azure Platform. *Evolutionary Bioinformatics*, p. 527.
- Wu, H., Lee, S., Chang, H. and Liang, J. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, 62, pp. 41-49.

The Application of Big Data Analytics Technologies for the Predictive Maintenance of Industrial Facilities in Internet of Things (IoT) Environments

Thomas Rieger [†], Stefanie Regier ^{††}, Ingo Stengel ^{††}, Nathan Clarke [†]

[†] School of Computing and Mathematics, Plymouth University, United Kingdom

^{††} Karlsruhe University of Applied Sciences, Germany

e-mail: thomas.riege@plymouth.ac.uk

Keywords: Big Data Analytics, Predictive Maintenance, Industrial Internet of Things (IIoT), Real-time Processing, Data Streams

Abstract

Software systems for the analysis of large amounts of data (Big Data Analytics) are applied in more and more areas. The subject of Predictive Maintenance (PdM) using Big Data Analytics systems is present in industrial manufacturing processes, especially in combination with Industrial Internet of Things (IIoT) environments. Typically, such applications are based on established Big Data computing using distributed storage and processing frameworks like Hadoop and distributed computing paradigm like Map Reduce (Marz, et al., 2015). Analytics applications running purely on frameworks like Hadoop can suffer from hours of data lag. Applying real-time processing on data streams could close this gap. In addition, with real-time processing, further problematic characteristics of Big Data can be addressed (e.g. high velocity combined with high volume). Streaming Analytics technologies can open up entirely new technological opportunities (Gupta, et al., 2016). On the other hand, real-time and stream processing places special demands on how data is processed and predictions are made. Some of the common algorithms to build up predictive models might become inefficient or even impracticable. This also applies to the manifold forms of combinations of such algorithms.

This paper investigates into the issues of real-time and stream processing in predictive maintenance applications in industrial environments. A deeper look on the characteristics of Predictive Maintenance applications in Industrial Internet of Things (IIoT) environments and the relevance of real-time processing in this field will be provided as a result of a literature research. Based on supplementary primary market research, it is determined which algorithm is the most important for Predictive Maintenance and real-time processing. It is then discussed if there is a need for further research to find a novel approach by adapting or innovative combining the selected algorithm in terms of its internal structure and functioning to significantly improve real-time and stream processing abilities.

1. Introduction

The motivation behind this paper is to investigate into the issues of real-time and streaming demands of predictive maintenance application in industrial environments. In addition to literature research, the most relevant algorithm used for predictive analytics in this field is determined by conducting primary market research, based on expert interviews as a qualitative procedure.

The following research questions have been formulated and will be addressed in this report:

- a) What are the characteristics and requirements, real-time and stream processing applications place on predictive algorithms?
- b) What are the methods and algorithms most relevant for Predictive Maintenance (PdM), taking Big Data, real-time requirements and data streams into account?

The first section of this paper describes the literature research undertaken regarding the basic role of Predictive Maintenance in Industrial Internet of Things (IIoT) environments and the relevance of real-time processing in this field. In addition, the algorithmic approaches for Predictive Maintenance mentioned in the literature will be summarized. The subsequent section describes the applied research methodology which comprises the intensive literature research accompanied by a qualitative study in the form of exploratory expert interviews. This section ends with a summary of the results of the market research. The paper then closes with a conclusion section containing a proposal for further research to define a novel approach which improves the real-time performance of a selected algorithm for a specific application case using a defined data set.

2. Literature Review

2.1. Predictive Maintenance in (I)IoT

From IT perspective, industrial automation systems are about managing systems and processes. Communication is focused on the process-, production-, and product data. Until now, storing raw data for later analysis has not been a priority (Marz, et al., 2015). In traditional industrial environments, following the automation pyramid (see Figure 1), sensors usually send their data over a fieldbus to a controller. A hierarchical structure with monolithic systems and non-standardized communication protocols is one of the usual characteristics of such environments. The full and unprocessed sensor data, hereinafter referred to as raw data, occurs only on the field level and the process level. From the perspective of the data this traditional approach usually does not provide any “transparency” through the layers of the automation pyramid (Bauer, et al., 2017). In an IIoT environment, intelligent devices are connected directly to the internet. The collected (raw) data is sent to a central service, which is often cloud-based. In that sense, the sensor data bypasses the traditional automation pyramid. Continuous streams of data with high volume and high velocity are available to IT-Systems.

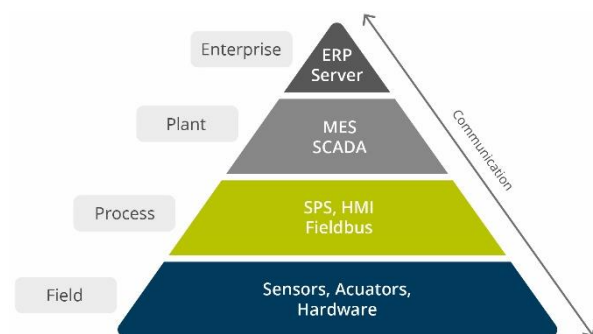


Figure 1: The traditional Automation Pyramid following (Bauer, et al., 2017)

Due to intelligent devices, more and more data is generated in industrial systems. Connecting these devices in IIoT environments makes it possible to store and analyse this data. As a result, the use of Big Data Analytics is useful for storage and analysis. For maintenance purposes, it is recommended to take all the data into consideration contributing to the issue of maintenance. This comprises condition-monitoring data, feedback from routine inspections, failure data, maintenance resources, work orders, overhaul and refurbishment plans, and so forth (Zhang, et al., 2014). Beside sensors, maintenance data comes from systems like Enterprise Asset Management (EAM) system, Enterprise Resource Planning (ERP) system, Condition monitoring System (CMS), Supervisory Control and Data Acquisition (SCADA) system or Safety Instrumented System (SIS) (Zhang, 2016).

Beside diagnostic approaches, like Condition Based Monitoring (CBM), a relevant field of application for the analysis of such data are prognostics like Predictive Maintenance (PdM). The main purpose of PdM is to avoid unscheduled outages of machines and plants by anticipating a failure before its possible occurrence. Time is a critical variable in prognostics, which is less important in diagnostics. In principle, maintenance should be carried out at the latest possible point, in order to avoid breakdowns and maximizing the service life at the same time. Scheduled service at normal working hours should be the result. The maintenance costs are also taken into account as they were expected to grow rapidly while degradation (Peng, et al., 2010). With PdM, maintenance costs have the potential to be reduced. Scheduled and preventive maintenance is much cheaper than unscheduled and reactive repair and maintenance. The availability of production capacities will be more predictable (eoda, 2014). In literature, the different PdM strategies are often summarised under the term “Reliability Centered Maintenance (RCM)”. According to Figure 2, the different maintenance strategies are reactive maintenance, preventive maintenance, predictive maintenance and proactive maintenance (NASA, 2008). Proactive maintenance is also often called descriptive maintenance and defined as a combination of descriptive analytics (describe why certain things have happened) and predictive analytics. The goal of descriptive maintenance is not only to correct errors, but to avoid them.

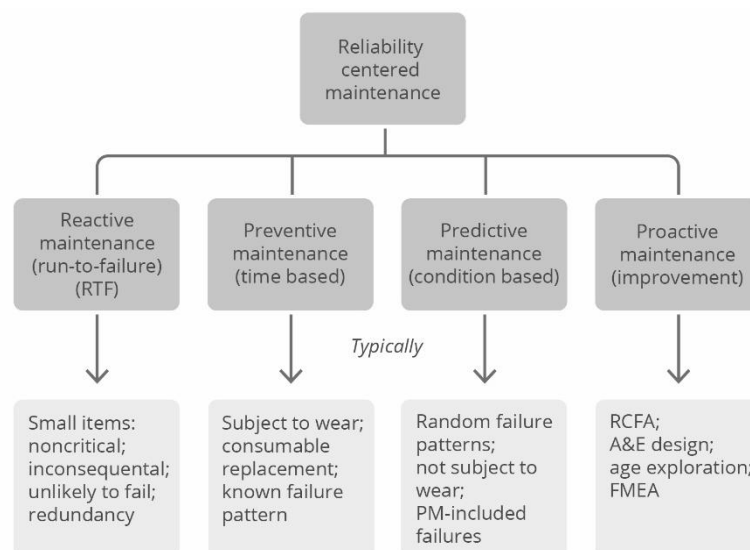


Figure 2: Maintenance strategies for Reliability Centered Maintenance (RCM) (NASA, 2008)

Because an expensive and risky reactive maintenance strategy is avoidable by applying preventive and predictive strategies, more and more companies improve their maintenance organisation by implementing PdM systems. In addition to the usual aims like Remaining Useful Life (RUL) and Mean Time to Failure (MTTF), also strategic and economic indicators have been introduced like Total Productive Maintenance (TPM) (Fredriksson, et al., 2012).

PdM assumes that the failure behaviour or characteristic is predictable. In other words, it presumes that equipment deteriorates deterministically following a well-defined sequence. Unfortunately, the assumption is not reflected in reality, where failure behaviour is normally a function of equipment aging, environmental effects, process drifting, complex interactions between components and systems, and many other factors (Kothamasu, et al., 2006). Predictive systems therefore have to make abstractions of their reality (models) and perform specific data preparation and data handling to deal with effects like high dimensionality, nonlinearity and context-drifts.

2.2. The Relevance of Real-time Processing

In IIoT environments, systems are more and more equipped with sensors delivering an ever-growing number of measurements of different types. Measurements are positions, temperatures, pressures, vibrations, flow rates, and many others. Therefore, the state measurements of a machine or other equipment tend to be high-dimensional. Sensors typically generate a continuous stream of data. The term of data streams refers to data continuously generated at a high rate. Normally, data streams are also temporally ordered, fast changing, and potentially infinite (Krawczyk, et al., 2015). The high volume and the high complexity of the data put massive demands on existing data processing techniques (Zhang, et al., 2017).

Established Big Data computing typically uses distributed storage and processing frameworks, like Hadoop, and distributed computing paradigm like Map Reduce. Such systems typically consist of two parts: First, storing high volumes of data fast and safely and second, performing calculations on the stored data and provide insights by batch or offline processing. Traditional incremental systems become very complex when processing large amounts of data. Complex systems are often faulty, difficult to maintain and to develop further, as well as difficult to scale. As a result, Big Data systems also have to be simple and easily comprehensible besides the efficient processing of large volumes of data. Therefore, the desired properties for a Big Data system are (Marz, et al., 2015):

Robustness and fault tolerance	Extensibility
Low latency reads and updates	Ad hoc queries
Scalability	Minimal maintenance
Generalization	Debug ability

Table 1: Properties for a Big Data System (Marz, et. al., 2015)

The process of collecting data from the production to interpret and process it in terms of failure detection and prediction of possible breakdowns is very complex and places high demands on the underlying IT-Infrastructure. If this process is not fast enough, this leads to short-term planning and immediate actions in maintenance and repair. In IIoT systems the representation of the data used by systems like Condition Based Maintenance (CMB) or PdM is often defined in the form of a virtual representation or virtual clone of the physical components of the factory. This virtual representation is running in parallel to the physical factory. The drivers of the virtual representation are continuous streams of raw data of the factory systems and sensors. If the data is not transferred and processed in real-time, the virtual representation runs after the physical process in terms of time. With respect to a PdM system this could lead to unintentional reactive maintenance because of insufficient lead time to plan the maintenance tasks (Bauer, et al., 2017). However, how fast real-time processing needs to be strongly depends on the application case. According to Rippel et al. (2015) in micro manufacturing systems, where vast volumes of micro parts are manufactured with high speed, the term real-time means microseconds. With systems for fault detection and PdM the rejection rate of the manufactured micro parts decrease by increasing processing speed (Rippel, et al., 2015). In other scenarios, the term of real-time can mean seconds, minutes or hours. For example in offshore wind turbines the frequency with which the data is available is mostly minutes and hours (Freitag, et al., 2015).

Processing continuous streams of data in real-time demands much more sophisticated programming techniques than working on a stored and immutable set of data having no time concerns. Especially the discipline of Massively Parallel Processing (MPP) comes into account (Chen, et al., 2014). There are lots of emerging technologies especially in the area of stream processing, like in-memory processing technologies, micro-batch stream processing or strongly ordered processing. Beside new frameworks and tools for stream processing, the way how data is stored usually differs strongly from batch-oriented approaches. Key-value based data stores, maps, sorted maps, or column-oriented databases like Apache Cassandra are in fashion. There are a multitude of different approaches for storing data, all having their advantages and disadvantages. In classical IT-Systems, Relational Database Management Systems (RDBMS) are more or less the standard. Normally, batch-oriented Big Data Systems rely on distributed file systems like Hadoop Distributed File System (HDFS), often supplemented by a relational database. In real-time and streaming environments the selection of the proper data store depends very strongly on the individual use case and individual data. In such systems, it is very common to use several complementary data stores at the same time (Marz, et al., 2015).

Frequently, dimension reduction techniques are used in a preliminary step to reduce the complexity of the data. This makes processing easier and faster and supports real-time capabilities (Lee, et al., 2014). Various strategies are used to reduce complexity. Thus, certain approaches, such as clustering or sliding windows, aim at reducing the data to be processed by different forms of partitioning. The following prediction algorithms can thereby be substantially relieved.

Another approach tries to evaluate if the failure of a certain component of a machine has a critical impact. Complex predictions are then only performed for those critical components. Figure 3 shows such an approach. The downtime caused by the failure of a component is shown as a function of the failure frequency. The aim is to define the most critical components in terms of impact and cost. According to Lee et al. (2014) the use of PdM is only recommended for components with low failure frequency and high downtime. The components that have a high failure frequency and cause high downtime must be eliminated "by design". All components that cause low downtime should be treated with standard approaches of Reactive and Preventive Maintenance. The respective definition of the limits as well as the classification of each component must be done by an expert, according to the

circumstances and the desired objectives (Lee, et al., 2014). This has to be done for each application case as every physical machine is different. No generally accepted definition is possible. With this approach the amount of data to be processed and the effort required to implement real-time Predictive Analytics applications can be significantly reduced (Lee, et al., 2014).

In addition to concentrating only on critical components, there are different approaches in which the prediction analytics are only performed in case of certain events (Ait-Alla, et al., 2015).

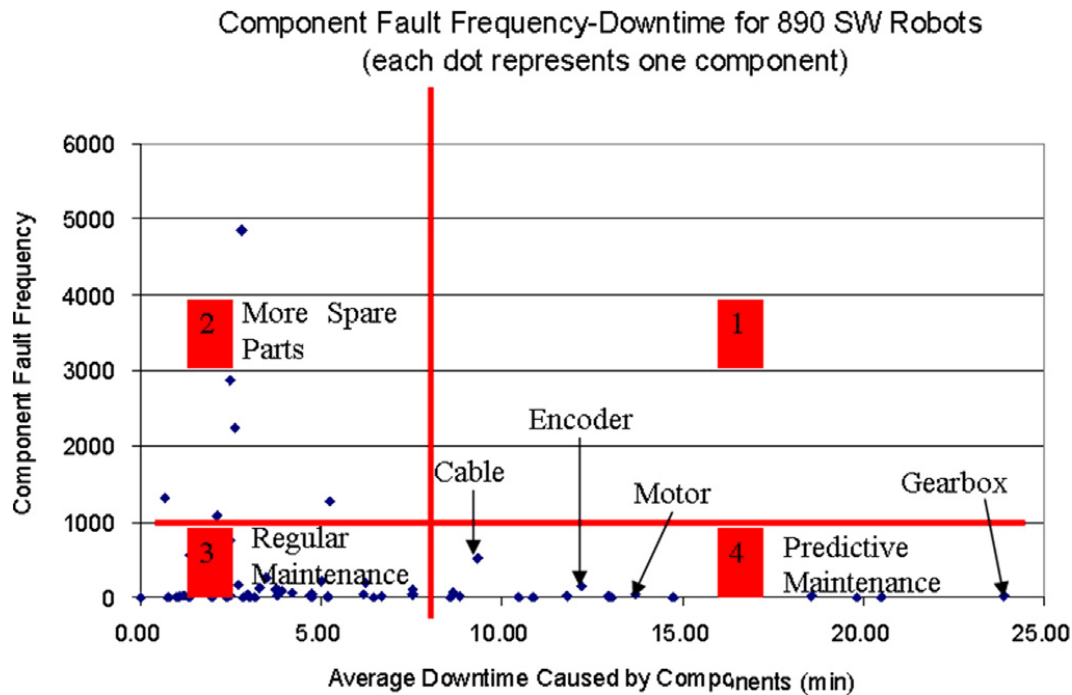


Figure 3: Four quadrant chart for identifying critical components, taken from (Lee, et al., 2014)

In order to be able to process large amounts of data with complex and demanding algorithms in real-time, it is often attempted to divide an elaborate processing step into many small processing steps. This segmentation is intended to ensure that the individual steps can be processed in parallel on several computers / nodes (scale out). One approach is the vertical division of the monolithic processing sequence into individual steps in the form of a pipeline. These individual processing steps must be very simple and independent. The connection between the individual processing steps is performed exclusively via the data (intermediate data), whereby the input data of a processing step corresponding to the output data of the previous processing step. When using pipelines, the data must be parallelisable and distributable. The processing logic (algorithms) must be divisible into independent individual steps (Orenstein, et al., 2015). As proposed by Marz et al. (2015), the intermediate data may consist exclusively of key-value pairs. This ensures that the individual processing steps are actually independent of each other. Another important aspect is that files of key-value pairs are very easy to parallelize and distribute (Marz, et al., 2015).

2.3. Algorithmic Approaches for Predictive Maintenance

In summary, the literature points out that the complexity when dealing with Big Data and real-time demands is enormous and appropriate solutions always depend on the concrete application case. There is no general rule for applying specific algorithms. As shown in this section, the applied algorithms are very diverse. The use of reduction techniques can be regarded as one of the common approach in the viewed literature. It is also very common that systems not only focus on predictive functions. Systems are often a combination of CBM, Complex Event Processing (CEP), fault detection and predictive functions.

Artificial Neural Networks (ANN) are mentioned by all literature when high complexity and uncertainty is present. For modern data-driven systems, coping with data streams and real-time demands, there is an unanimous opinion in the viewed literature that ANN's will become the most relevant approach in future systems. As a disadvantage of ANN's the high complexity is usually mentioned. The literature mostly agreed that there is a lack of knowledge and practical experience in handling ANN's. This is mostly argued by the hidden inner structure of an ANN's which makes its inner functioning completely opaque. Because of its distributed structure ANN's are meant to be ideal for distributed computing.

With the growing demand for real-time analytics ANN's seem to be an appropriate approach. However there is little research and experience in this field as most current projects are not really faced with hard real-time demands and therefore fall back to less complex approaches. Research has to be undertaken regarding many different aspects of ANN's. Beside research regarding the complexity, there is also an essential need for research regarding real-time operation and optimisation with respect to modern real-time and stream processing IT technologies.

3. Research Methodology

In addition to the literature research the method intended to be the most important in the field of streaming analytics for predictive maintenance was determined by primary market research. Primary research refers to the systematic gathering of information directly in the field (field research). The fundamental research design for the primary market research was based on expert interviews as a qualitative procedure.

According to Bogner et al. (2014), the expert interview is very popular as a qualitative survey tool within social research, even though, among the followers of qualitative research, it is not universally accepted as an independent acquisition method (Bogner, et al., 2014). One reason for this is the fact that the expert interviews are comparatively less theoretically and methodologically examined and supported. There are different definitions for the concept according to experts. Furthermore, there is criticism that the interviewer often fails to meet the criteria of openness and not influencing the expert (Bogner, et al., 2014).

In the present case, the implementation of expert interviews was obvious. On one hand, the aim was to obtain a comprehensive picture of the opinions of experts. On the other, it was also important to gather different opinions and experiences from experts on Predictive Analytics and the specific requirements placed on methods in the application area of Predictive Maintenance (PdM) in IIoT environments and Streaming Analytics. Furthermore, expert knowledge of the evaluation and validation of models was important. In contrast to focus groups, the expert interview allows different experts to be addressed in different ways and to develop the guidelines from interview to interview.

Apart from concerns regarding to the incompatibility of the various experts, the comparatively reduced organisational effort demonstrated the benefit of conducting expert interviews.

Despite all efforts, distortions and influences during the interview cannot be completely ruled out. Therefore, critical reflection, systematic control, and transparent discussion of such effects are required. Interference effects are known, for example, in surveys of specific age groups or target groups with similar characteristics and usually occur independently of the type of questioning. In addition, problems caused by expert and technical vocabulary, the frequent lack of imagination, diminishing capacity to concentrate, the feeling of being overwhelmed, incipient time pressure or strong digressions are often amplified by the open nature of general questioning (De Vaus, 2013). In the present study, the target group consisted exclusively of experts who are constantly dealing intensively with the subject area in their professional activities and are also accustomed to situations of intensive conversation via web conference. It is therefore assumed, that no relevant distortions and influencing factors needed to be considered during the preparation, execution or evaluation of the interviews.

As already mentioned in this paper, there are few recommendations in the literature with regard to the systematic evaluation and generation of information from expert interviews. Finkbeiner (2016) offers a recommendation for a model for the analytical evaluation of content. The order of content within individual interviews in this model is irrelevant for the evaluation. Instead, the emphasis is on the search for similar thematic units (Finkbeiner, 2016). The approach in this study was based on this. First, a text document with relevant thematic blocks was prepared on the basis of the questionnaire, before the interviews were conducted. If new issues arose during the interviews, these were supplemented accordingly in the document.

It was possible to make audio recordings of all interviews. The audio recordings were evaluated promptly after each interview. Important passages from the interviews could be paraphrased or quoted verbatim and assigned to the previously defined thematic blocks, and transcribed. The respective author was colour coded. For time reasons a complete transcription was waived. After this step had been completed for all interviews, the text passages were reorganized and reduced to their essential components through the abstraction of common motifs. The originator of the respective statements remained identifiable. Based on this, superordinate statements or findings on the respective thematic blocks were collected in note form. These superordinate statements formed the outline of the following evaluation section, in which the respective findings were formulated in detail.

4. Results of Market Research

It was stated by nearly all experts, that current PdM projects are more likely to be the first step in the form of simple systems. Contrary to the external impression which is obtained from the topics Industry 4.0 (I4.0) und IIoT, industrial projects are often only at the beginning. In most projects online data is not available yet. The main reason for this is the not adequately developed IT infrastructure in many areas. In addition to this, industrial customers are often missing an exact idea of what knowledge is contained in their data and how to explore and utilize it.

Therefore, in current systems, the goal is usually the detection of errors and the rapid initiation of measures. This is often based on CBM and CEP systems. In addition, data analytics is also carried out with the aim of determining optimization potentials in the plants.

Predictive and highly adaptive online systems that can process data streams in real-time will be standard in the opinion of experts in the future. Predictive Maintenance (PdM) will become an important application in the area of Predictive Analytics. Autonomous transport and conveyor systems as well as self-organizing production plants require intensive communication with people (Human to Machine Communication H2M, Machine to Human Communication M2H) and machines (Machine to Machine Communication M2M).

Consensus opinion of the experts was that in future the ability of highly adaptive online systems to process data streams in real-time will be a key factor. Because of the not yet been sufficiently developed communication infrastructure in industrial environments and the lack of experience in the field of Predictive Analytics, this is not yet the focus. Due to the rapidly advancing development in the area of I4.0 and IIOT, real-time systems will be the normal case in the future.

According to the experts, in current projects with real-time requirements mostly reduction techniques are used to reduce the amount of data to be processed. The actual algorithms are not real-time capable and not suitable for distributed and parallel processing. Here, there is still a lack of extensive knowledge and experience to make algorithms real-time capable. It was also mentioned that the adaptive learning process must also happen in real time.

Overall, the experts confirmed that PdM is one of the important areas in IIoT and I4.0 environments and is still under development in practise. Beside the fact, that current implementations are mostly simple, because of a lack of knowledge and experience in many areas, they also confirmed, that in future scenarios methods and systems are essential that are able to cope with high complexity as well as real-time demands. To handle highly complex data and ensure adaptive and flexible behaviour, accompanied by the ability to deal with uncertainty, the surveyed experts agree that deep learning approaches and specifically Artificial Neural Networks (ANN) represent the key future procedures. From the results of the expert interviews it can be derived, that there is a need for deeper research in the area of the application of ANN in Big Data and real-time scenarios. Besides many other questions, an important open question is how to apply ANN for PdM applications in modern stream processing architectures and systems. A novel approach which is flexible enough to cope with the very different requirements of PdM application cases and which is easy to use is necessary.

5. Conclusions

From the results of the literature research and the conducted qualitative study, it can be derived that PdM systems in industrial environments are in a far less developed state than would be expected based on the topics I4.0 and IoT. The objective of PdM systems is ultimately always the reduction of costs. Failures and unscheduled downtime cause the highest costs. Predicting failures and scheduling maintenance and repairs based on predictions is intended to avoid such unscheduled downtime.

Even though the degree of automation in industrial environments is very high, complete communication infrastructures from the field level to the higher-level IT and cloud systems is rare. The direct streaming of raw data from the field into IT or cloud systems is far from being standard. The unlimited networking and autonomous action and communication of things on the principle of an IoT landscape are still rare in industrial environments. In addition to technical reasons, such as a lack of standardization as well as security concerns of the companies certainly play an important role in this situation.

The question which methods and algorithms are used most frequently by the experts has largely been answered in a consistent manner by the fact that this strongly depends on the application case.

Therefore, no general statements can be made. Certain methods and algorithms are, in principle, better or worse suited for a particular application. Methods from the field of Machine Learning (ML) and Artificial Intelligence (AI) are increasingly gaining acceptance, according to the opinion of the surveyed experts. Additionally, the opinion of the majority of the surveyed experts is that deep learning and Artificial Neural Networks (ANN) represent the key future procedures. An important feature of ANNs is the ability to create models with incomplete and partial non-existent expert knowledge. ANNs are therefore also suitable for analysing unknown data in an evaluation phase and for determining suitable methods. By permanent learning, adaptive systems are achieved, whose quality can be enriched with additional expert knowledge. Due to their structure, neural networks are also very good for parallel and distributed processing. However, ANNs are rapidly become very complex and processor intensive. The combination with reduction techniques, filters or, alternatively, the application of novel and more optimized approaches is advisable.

Related literature and surveyed experts agree that in order to be able to use neural networks in future PdM applications, deeper knowledge and experience is necessary. This applies in general to the construction and the use of neural networks, as well as to the processing of Big Data streams in compliance with real-time requirements.

Based on the conclusions mentioned and especially the perception that ANN's will become the most relevant approach in future real-time oriented PdM applications, additional research should be undertaken to find a novel approach by adapting or innovative combining existing algorithms in terms of its internal structure and functioning to significantly improve real-time and stream processing abilities.

To investigate on such a novel approach a specific ANN algorithm should be selected and adopted for a clear demarcated application scenario with a fixed and well balanced dataset. Also, deeper investigation on neural networks for the application in PdM systems needs to be carried out. Besides neural networks, technologies in the field of stream processing should be investigated in more detail with regard to PdM systems.

An additional but no less import question is how to measure and evaluate results in a representative way. So far, few approaches for measuring, evaluating and benchmarking of results have been developed. Moreover, those approaches are usually not verifiable as generally valid.

References

- Ait-Alla, Abderrahim, et al. (2015). "Real-time fault detection for advanced maintenance of sustainable technical systems". *Proceedings of the CIRP*. 2015, 41.
- Bauer, Dennis, Stock, Daniel and Bauernhansl, Thomas (2017). "Movement towards service-orientation and app-orientation in manufacturing IT". *10th CIRP Conference on Intelligent Computation in Manufacturing Engineering - CIRP ICME '16*. 2017.
- Bogner, Alexander, Littig, Beate and Menz , Wolfgang (2014). *Interviews mit Experten*. Wiesbaden : VS Verlag für Sozialwissenschaften, 2014.
- Chen, C.L Philip and Zhang, Chun-Yang (2014). "Data-intensive applications, challenges, techniques". *Information Sciences*. 2014, 275.
- De Vaus, David (2013). *Surveys in Social Research* . London, New York : Routledge, 2013.
- eoda (2014). *Predictive Maintenance (mit R). Potenziale und Möglichkeiten der freien Statistiksprache R für neue Geschäftsmodelle im Industrie 4.0 Zeitalter*. Kassel, eoda gmbH, 2014.

Finkbeiner, Patrick (2016). "Qualitative Research: Semi-structured Expert Interview". In: *Social Media for Knowledge Sharing in Automotive Repair*. s.l. : Springer International Publishing, 2016.

Fredriksson, Gustav and Larsson, Hanna (2012). *An analysis of maintenance strategies and development of a model for strategy formulation – A case study*. Master of Science Thesis. Schweden, Chalmers University of Technology, 2012.

Freitag, Michael, et al. (2015). "A Concept for the Dynamic Adjustment of Maintenance Intervals by Analysing Hereogeneous Data". In: *Applied Mechanics and Materials*.p. 794, 2015.

Gupta, Sumit and Shilpi (2016). *Real-Time Big Data Analytics*. s.l. Packt Publishing, 2016.

Kothamasu, Ranganath, Huang, Samuel H and VerDuin, William H. (2006). "System health monitoring and prognostics — a review of current paradigms and practices". In: *The International Journal of Advanced Manufacturing Technology*. 28, 2006, Vol. 9.

Krawczyk, Bartosz and Wozniak, Michal (2015). "Data stream classification and big data analytics". In: *Neurocomputing*. 2015, p. 150.

Lee, Jay, et al. (2014). "Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications". In: *Mechanical Systems and Signal Processing*. 2014, p. 42.

Marz, Nathan and Warren, James (2015). *Big Data. Principles and best practices of scalable real-time data systems*. New York : Manning Publications Co., 2015.

NASA (2008). *RCM Guide Reliability- Centered Maintenance Guide For Facilities and Collateral Equipment*. 2008.

Orenstein, Gary, et al. (2015). *Building Real-Time Data Pipelines - Unifying Applications and Analytics with In-Memory Architectures*. Sebastopol, CA : O'Reilly Media, Inc., 2015. 978-1-491-93547-7.

Peng, Ying, Dong, Ming and Zuo, Ming Jian (2010). "Current status of machine prognostics in condition-based maintenance: a review". In: *The International Journal of Advanced Manufacturing Technology*. 1-4, 2010, p. 50.

Rippel, Daniel, Lütjen, Michael and Freitag, Michael (2015). Simulation of Maintenance Activities for Micro-Manufacturing Systems by use of Predictive Quality Control Charts.

Zhang, Liangwei and Karim, Ramin (2014). *Big Data Mining in eMaintenance: An Overview*. 2014.

Zhang, Liangwei (2016). *Big Data Analytics for Fault Detection and its Application in Maintenance*. Doctoral Thesis. s.l. : Luleå University of Technology, 2016.

Zhang, Liangwei, et al. (2017). "Sliding Window-Based Fault Detection From High-Dimensional Data Streams". In: *IEEE Transactions on Systems, Man, and Cybernetics*. 2017, Vol. 47, 2.

Reinforcement Learning Based Approach for Optimal Deep Drawing Process Control

Johannes Dornheim, Norbert Link
 Intelligent Systems Research Group (ISRG)
 Karlsruhe University of Applied Sciences, Germany
 e-mail: johannes.dornheim@hs-karlsruhe.de

Keywords: Online Control, Adaptive Optimal Control, Reinforcement Learning, Machine Learning

Introduction

In this paper, a *Reinforcement Learning* (RL) based approach for production process parameter optimization tasks is described and results for a sample application are presented. Under the assumption of time-discrete process behaviour, production processes can be modelled as finite or fixed horizon *Markov Decision Processes* (MDP). The goal of optimization thereby is the maximization of a reward signal, which is defined over the final state of the processed product. The MDPs action-space is defined by either discrete or continuous process parameters. Based on the MDP framework, the proposed algorithms can solve (1) offline, simulation-based problems like sample-efficient search for optimal process-parameters, and (2) online optimization problems like adaptive optimal control of stochastic, partially-observable or non-stationary production environments. In this paper a class of production process problems is specified and a RL based approach is described. Furthermore, first results of the application on a metal sheet deep drawing process are given.

Related Work

The outcome of deep drawing is influenced by different process parameters, as described by (Singh & Agnihotri, 2015). Besides fixed value parameters, like blank shapes and thickness variation, time-varying blank holder forces (BHF) are important for the resulting process quality: “It is evident that even small changes in BHF can lead to failure during the process. These failures can be avoided if a variable BHF is applied, but the correct trajectories need to be chosen” (Singh & Agnihotri, 2015). Therefore, many publications are treating time-variable BHF optimization problems in analytic and application specific manners. In (Volk, et al., 2011) a fuzzy logic controller is described for obtaining dedicated optimal holding force trajectories for segmented blank holders. Other recent publications in this context are investigating different pressure schemes for a simulated deep drawing process (Wifi & Mosallam, 2007) and for a segmented deep drawing process (Tommerup & Endelt, 2012). These methods have in common, that proposed solutions are based on problem specific analytical methods. (Senn & Link, 2010) propose research on BHF optimization tasks with a more general approach. More precisely, they use finite element models (FEM) and define explicit state descriptions, which are reconstructed from simulated observables by regression analysis and dimension reduction. In (Senn, 2012) *Approximate Dynamic Control* is used to calculate a simulation-optimal BHF control policy for online application.

Problem Class – Production Processes

Production processes can be modelled as special cases of finite-horizon MDP, commonly denoted by $(S, A, P(s, a, s'), R(s), \gamma)$, where the discount factor γ is one. The reward function R is non-zero for terminal states, where the main reward is given according to the measured quality of the final process result. In addition, negative intermediate reward can be given depending on actions or per time step to model energy and time costs. In online applications of optimization algorithms, based on the MDP

framework, the actual state description $s(t)$ has to be derived from the sequence of sensor and actor values observed so far until time t . The derivation has to make sure that the estimated state $s(t)$ can be treated as a Markovian-state in the context of the given optimization problem. This implies that actions can be found for every s which are optimal in terms of the given reward-function without taking historical state-estimations into account.

Approach

The optimization goal in the MDP framework is to find the optimal policy $\pi^*(S) \rightarrow A$, which maximizes the expected accumulative future reward for all states $s \in S$. Within RL (Sutton & Barto, 2011), in contrast to *Dynamic Programming* and *Approximate Dynamic Programming*, this is reached by raw experience in interactive settings. The aim of using RL for production process control is (a) to find optimal policies prior to production (ppo) by using simulation environments and (b) later refining the identified policy in online process control (opc) for adaption to non-stationary, unknown or non-simulatable process behaviour aspects. Both steps are currently investigated offline based on an interactive simulation environment for the given sample application of deep drawing. The methods developed and investigated so far are variations of the q-learning (Watkins & Dayan, 1992) algorithm adopted to the described problem class in combination with *Artificial Neural Networks* (ANN) for non-linear function approximation. The conceptual core of q-learning is the q-function $q_\pi(s_t, a_t) = \mathbb{E}_\pi(\sum_{k=t}^T \gamma^k R_{k+1} | s_t, a_t)$, which is updated incrementally online, based on experience-tuples (s_t, a_t, s_{t+1}, R) for a given learning-rate α by the update rule

$$q(s_t, a_t) \leftarrow q(s_t, a_t) + \alpha [R + \gamma \max_{a_{t+1}} q(s_{t+1}, a_{t+1}) - q(s_t, a_t)]. \quad (1)$$

After convergence of (1), the optimal policy $\pi^*(s)$ is found by determining $\operatorname{argmax}_a q^*(s, a)$ for all states $s \in S$.

The BHF-controlled deep-drawing example process can be represented as fixed-horizon MDP with time-step-dependent disjoint state-spaces. This is due to constant progress of the process (forced by constant stamp-speed), and

a defined terminal state (given by cup height). The q-function is approximated by ANN, one for each time-step. The methods were developed and tested for the ppo scenario while the opc scenario is still subject to ongoing work. In the ppo case, due to the use of deterministic FEM simulations, the current state is fully described by the trajectory of control-actions leading to the state, as depicted in Figure 1. In

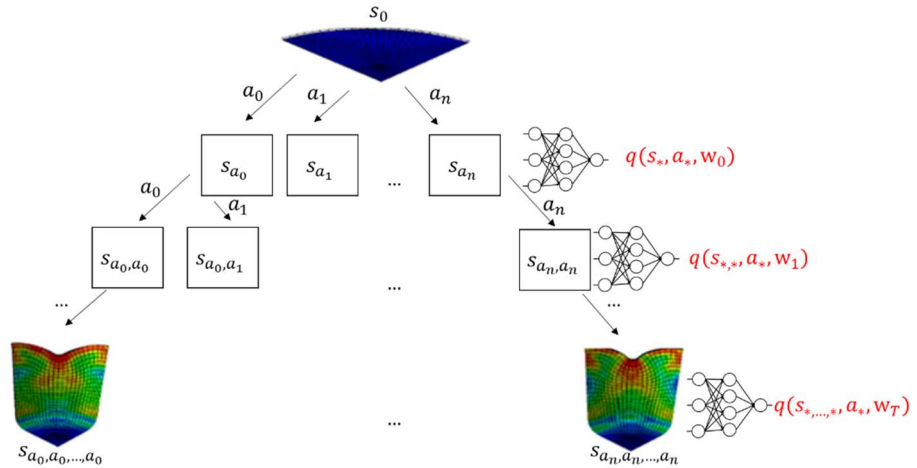


Figure 1: Time-layered Artificial Neural Networks for Q-Function approximation

the opc case external influence and stochastic process behaviour are imposed, to which the optimal control policy has to be adapted. The true state information is not available anymore and sensor observable trajectories are used in addition to action trajectories as state descriptors, instead.

Evaluation and Results

An FEM-simulated cup deep drawing process was used for evaluation of the two developed ppo approach variants in comparison to a classic hill-climbing optimization as a baseline. The simulation was using a rotationally symmetric 2D cup model and a simple material model to yield the elongation and *von Mises* stress $\sigma_{v,M}$ fields. While the stamp force was constant, the blank holder force (BHF) was controlled by selecting one of seven discrete force values at five discrete times (the BHF path) to finally get a cup with low internal stresses and good material efficiency. The performance of the optimization methods is measured by the number of sample processes required to reach the given quality value (*Empirical Sample Efficiency* ESE). The reward function, plotted in Figure 2, is defined as $R = 10 - n_1(\|\sigma_{v,M}\|_2) + n_2(m_r)$, where $\|\sigma_{v,M}\|_2$ is the L2-Norm of element-wise *von Mises* stresses, m_r is the residual material and n_i are normalization-functions, enabling equal balancing of multiple reward function terms by forcing the value-ranges into $[0, 10]$ based on empiric values.

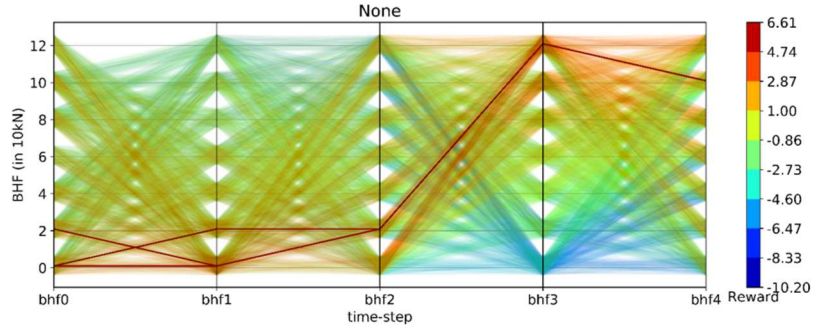


Figure 2: Full investigation of all possible trajectories: Reward by Trajectory (Color), Actions as parallel dimensions. Stacked ascending by reward (jitter added to y-axis for visualization purposes). Three highlighted trajectories: $R > 6.3$

The optimization methods are executed, repeatedly 100 times each, until a trajectory with $R > 6.3$ (highlighted trajectories in figure 2) is reached. The resulting ESE is depicted in figure 3 in form of box-plots over a log-scale for sample-counts. Evaluated algorithms are: (a) discrete space *Hill-Climbing* optimization with a mean of 3412.01 samples used (b) q-learning with a tabular q-function (mean: 686.13 samples) and (c) q-learning with approximated q-functions (approximative approach as described in the previous chapter, mean: 71.43 samples). Both q-learning agents are acting ϵ -greedy with $\epsilon = 0.1$ and due to the deterministic environment a learning rate α of 1.0. ANN for (c) are retrained every 10 iterations.

Discussion and Future Work

The published results are indicating a particularly high ESE and a good generalization ability of the approximative approach (c). Evaluations of the approach in comparison to other common *meta-heuristics*, like genetic algorithms, are part of ongoing work.

Future work focuses on two methodological improvements: (1) approaches for sample efficient adaption of stochastic, unknown and non-stationary process behaviour and (2) reduction of the computational complexity by using a *Recurrent Neural Network* for Q-Function approximation and simultaneous state estimation, instead of multiple ANN. In the current architecture the state-space-dimension d_s for final states is dependent on the time-step-count T and the number of observables $d_s = T * |o|$, further the count of needed ANN equals T .

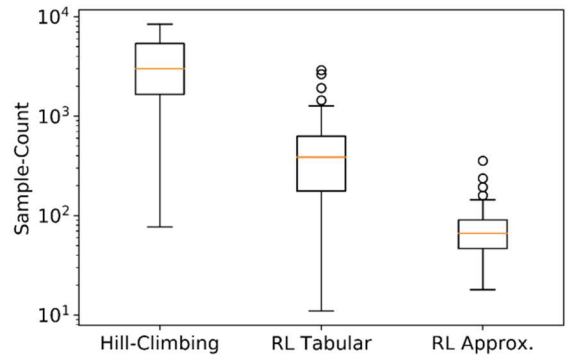


Figure 3 empirical sample-efficiency distribution statistics on logarithmic scale by algorithm for naïve hill-climbing and RL approaches

Acknowledgement

This work has been supported by the DFG Research Training Group 1483 "Process chains in manufacturing". The authors would like to thank the ITM at the KIT for providing the finite element deep drawing models.

References

- Senn, M., (2012). *Optimale Prozessführung mit merkmalsbasierter Zustandsverfolgung*. s.l.:KIT Scientific Publishing.
- Senn, M. & Link, N., (2010). Hidden state observation for adaptive process controls. s.l., *Proceedings of the Second International Conference on Adaptive and Selfadaptive Systems and Applications*.
- Singh, C. P. & Agnihotri, G., (2015). Study of deep drawing process parameters: a review. *International Journal of Scientific and Research Publications*, Volume 5, pp. 1-15.
- Sutton, R. S. & Barto, A. G., (2011). Reinforcement learning: An introduction.
- Tommerup, S. & Endelt, B., (2012). Experimental verification of a deep drawing tool system for adaptive blank holder pressure distribution. *Journal of Materials Processing Technology*, Volume 212, pp. 2529-2540.
- Volk, M., Nardin, B. & Dolšak, B., (2011). Application of numerical simulations in deep-drawing process and holding system with segments' inserts. *Strojniški vestnik-Journal of Mechanical Engineering*, Volume 57, pp. 697-703.
- Watkins, C. J. C. H. & Dayan, P., (1992). Q-learning. *Machine learning*, Volume 8, pp. 279-292.
- Wifi, A. & Mosallam, A., (2007). Some aspects of blank-holder force schemes in deep drawing process. *journal of achievements in materials and manufacturing engineering*, Volume 24, pp. 315-323.

Chapter 2

Bioinformatics / Health informatics

Quantified Self in the healthcare: State of the art and impact on the health insurance

Sadrick Widmann, Shirley Atkinson, Ingo Stengel, Nathan Clarke, Stefanie Regier
School of Computing, Electronics and Mathematics University of Plymouth, United Kingdom
Faculty of Computer Science and Business Information Systems, University of Applied Sciences Karlsruhe, Germany
{ingo.stengel|stefanie.regier}@hs-karlsruhe.de; {sadrick.widmann|shirley.atkinson|N.Clarke}@plymouth.ac.uk

Keywords: Quantified Self, healthcare, health insurance

Abstract

The simple integration of tracking an individual's health into daily life by automated recording enables to receive valuable insights. Eventually, the gathered personal data in combination with Big Data and Internet of Things technology is expected to offer preventive and personalized medicine. The Quantified Self enjoyed great popularity in the last years. Especially the healthcare sector pursues these recent developments with great interest. This paper focuses on the Quantified Self in the healthcare sector, aiming at developing a state of the art and highlighting the potential use in the health insurance.

Introduction

Users collect personal data across and within different platforms and devices. Thus creating context-specific profiles of themselves. The concept behind this is known as Quantified Self (Sharon, 2017). In recent years, the Quantified Self has enjoyed great popularity. Especially in the healthcare sector, new applications and devices are emerging, aiming at certain areas of or an overall view on health (Sharon, 2017; Shin and Biocca, 2017). Thereby, a better well-being, a healthy lifestyle and a behavioral change are in the foreground (Nake, Alissandrakis and Zbick, 2016). These health application are based on internet-related or mobile computing technologies and designated as eHealth and mHealth (Eysenbach, 2001; Free *et al.*, 2013). Obviously, the increased use of health surveillance applications has a positive impact on the users' health, but beyond this, it indicates the concern of users in monitoring their health, gaining insights and attaining a good health. Following up on this mindset, with the Quantified Self the healthcare sector has taken a major step forward in providing individual treatments.

State of the Art: Quantified Self

The focus of health insurances is on the one hand to explore the field of Quantified Self regarding a further support of disease prevention and on the other hand to improve customer loyalty by yet unique incentives like grants, discounts or other benefits (Shemkus, 2015).

Wearables and smart devices enable the effortless participation of people. This innovation facilitates self-tracking of people across a variety of sensors, e.g. pulse and blood pressure or motion sensors (Sharon, 2017). There are even special devices, like Fit Bit, for monitoring and recording of the necessary health data (Swan, 2009). Moreover, wearables shifted an active participation, in which users had to manually track and enter the personal data, towards an automated recording of data. In addition to the wearables, much more will be achieved in the Internet of Things area, improving the quality of the data, increasing the volume of collected data, and simplifying the integration into daily life (Swan, 2012b). Furthermore, the application of Big Data technologies will enable and improve the Quantified Self and the health services in the future. Accurate and real- or at least near-time

analysis allows direct and individual feedback for users. Therefore, the attractiveness on the user side increases and the possible medical added value enhances (Swan, 2012a). Additionally, user experience and gamification motivate to the user interactions (Swan, 2012b). To conclude the combination of the Internet of Things and Big Data, offering smart devices and enabling the Quantified Self to a wide range of users, will initiate the next important step towards a personalized and preventive health environment.

Some scientific research on the Quantified Self, its possible application for behavioral change and healthy lifestyle has already been conducted, e.g. by Swan and Sharon. Whereas, the research about quantified self and personal, preventive medicine along with the accompanying application in health insurance is still in its infancy. In particular, the linkage of information technology and smart devices into a comprehensive approach for health insurances should be further deepened.

In this light, it is important to consider the concerns of users regarding social and privacy aspects (Sharon, 2017). Users are willing to share their data with companies, such as Google, Netflix or Spotify, for a corresponding service. Companies that gain the trust of the users and offer a value-added service, can expect a great potential within the context of Quantified Self (Weitzman *et al.*, 2012). Health insurance companies own this kind of leap of faith which they can use to introduce “Connected Services” based on the Quantified Self (Park Associates, 2015). Besides a healthy lifestyle or the medical prevention, health insurances can also open up new business areas and redesign the relationship to customers completely. In this way, not only the medical component is redefined, but the business relationship can also change.

Methodology for “a framework for connected services in the health insurance” and outlook

As mentioned in the previous chapter, the current research focused on the Quantified Self from the user perspective and in particular with the aim to improve health and promote behavioral change. Up to now, there is only limited research on a comprehensive view on the quantified self in the health sector and no research on the integrated view on the quantified self and related information systems in health insurances. To harness the potential of the quantified self for health insurances, a framework to establish connected services for insurance holders must be developed. Thereby, currently existing research as well as concepts must be evaluated and qualitative interviews with stakeholders from the health insurance sector must be conducted to collect necessary requirements for the framework.

To make the Quantified Self available on a large scale and to pave the way for health insurances, the interplay of Big Data and Internet of Things must be explored much more. It is also necessary to examine the interplay between the different stakeholders, such as physicians and research institutes, whose expertise is necessary in the development of the described applications, much more. The aim of this work is to develop a framework for information systems of health insurances to provide connected services in the context of personalized and preventive medicine. The framework shall consist of a reference software architecture as well as a reference procedure model to introduce and integrate the information systems of a health insurance. Further research will now focus on the state of the art health care systems to develop and evaluate a suitable framework for the above mentioned “Connected Services” for health insurances.

To achieve this research goals and to acquire the necessary insights, different research methods must be evaluated and their advantages and disadvantages to solve the research question should be considered.

The following table highlights methods, like experiments, surveys, and prototyping, applied in comparable research setups and published in several papers:

Table: Overview of research projects with comparable setup

Title	Author	Release year	Period	Participants	Methodology
Software architecture for integration of institutional and social learning environments	Miroslava Raspopović, Svetlana Cvetanović, Dušan Stanojević, Mateja Opačić	2016	4 courses with 15 lessons within 2 weeks	72	Prototype with survey.
Validation of the Fitbit One® for physical activity measurement at an upper torso attachment site	Keith M. Diaz, et al.	2016	hours	13	Experiment
Visualizing Quantified Self Data Using Avatars	Isabella Nake, Aris Alissandrakis, and Janosch Zbick	2016	1 month	10	Prototype with survey.
Tracking Health Data Is Not Enough: A Qualitative Exploration of the Role of Healthcare Partnerships and mHealth Technology to Promote Physical Activity and to Sustain Behavior Change	Miyamoto SW, Henderson S, Young HM, Pande A, Han JJ	2016	hours	30 (group size of 8-12)	Focus group interviews with prior survey
Mobile Health Technology Evaluation	Kumar, Santosh et al.	2013	hours	26	Workshop
Modular Architecture of Added-Value Applications for German Healthcare Telematics	Sebastian Dünnebeil	2013	-	-	Evaluation against criteria collected during requirements engineering

As the table clearly illustrates, a multi-method design, with a combination of conceptual and empirical, quantitative research is often used and proved in a comparable setup. Especially the combination of prototyping, to evaluate the developed concepts, models or methods, and surveys, as quantitative, empirical research component to provide evidence for the research output, delivers promising results.

Based on the assessment of the existing research methodologies, the evaluation of *the upcoming framework for connected services in the health insurance* will be achieved with the following setup. Firstly, qualitative interviews will be conducted with experts in the field of information technology and health insurance. This information will be used as requirements and conditions to develop the framework and as evaluation criteria to assess its integrity. Secondly, a prototype based on the framework shall be implemented. A group of participants will test the prototype. Afterwards, the prototype is verified against the evaluation criteria and regarding its applicability by a survey, conducted among participants testing the prototype. Testing and evaluation based on a prototype helps

to collect views and opinions of users as well as identify faults and problems in the framework. Moreover, suggestions can be used to improve the quality and usability of the framework. Especially experimental prototypes are useful as proof of concept and will help to assess the feasibility, the validation and the completion of the framework. Finally, the framework characteristics are analyzed against the evaluation criteria derived during the qualitative interviews. Testing against the design specification, helps to ensure that a full and relevant evaluation of a prototype is carried out and allows to evaluate the developed framework fully, containing all necessary features.

References

- Eysenbach, G. (2001). "What is e-health?" in: *Journal of Medical Internet Research*, 3(2), p. e20. doi: 10.2196/jmir.3.2.e20.
- Free, C. *et al.* (2013). "The Effectiveness of Mobile-Health Technologies to Improve Health Care Service Delivery Processes: A Systematic Review and Meta-Analysis" in: *PLOS Medicine*, 10(1), p. e1001363. doi: 10.1371/journal.pmed.1001363.
- Kim, H. and Xie, B. (2017). "Health literacy in the eHealth era: A systematic review of the literature" in: *Patient Education and Counseling*, 100(6), pp. 1073–1082. doi: 10.1016/j.pec.2017.01.015.
- Nake, I., Alissandrakis, A. and Zbick, J. (2016). "Visualizing Quantified Self Data Using Avatars", in: Alma Leora Culén, L. M., Irini Giannopulu, and Birgit Gersbeck-Schierholz (ed.) *ACHI 2016 : The Ninth International Conference on Advances in Computer-Human Interactions. ACHI 2016: The Ninth International Conference on Advances in Computer-Human Interactions*, International Academy, Research and Industry Association (IARIA), pp. 57–66. Available at <http://urn.kb.se/resolve?urn=urn:nbn:se:lnu:diva-52381> (Accessed: 4 May 2016).
- Park Associates (2015). "Consumer comfort with sharing health data in exchange for insurance discounts varies from 26-42% depending on connected fitness device owned", 12 August. Available at <https://www.parksassociates.com/blog/article/chs-2015-pr6>.
- Sharon, T. (2017). "Self-Tracking for Health and the Quantified Self: Re-Articulating Autonomy, Solidarity, and Authenticity in an Age of Personalized Healthcare" in: *Philosophy and Technology*, 30(1), pp. 93–121.
- Shemkus, S. (2015). "Fitness trackers are popular among insurers and employers – but is your data safe?" in: *The Guardian*, 17 May. Available at <https://www.theguardian.com/lifeandstyle/2015/apr/17/fitness-trackers-wearables-insurance-employees-jobs-health-data> (Accessed: 27 June 2017).
- Shin, D.-H. and Biocca, F. (2017). "Health experience model of personal informatics: The case of a quantified self" in: *Computers in Human Behavior*, 69, pp. 62–74. doi: 10.1016/j.chb.2016.12.019.
- Swan, M. (2009). "Emerging Patient-Driven Health Care Models: An Examination of Health Social Networks, Consumer Personalized Medicine and Quantified Self-Tracking" in: *International Journal of Environmental Research and Public Health*, 6(2). doi: 10.3390/ijerph6020492.
- Swan, M. (2012a). "Health 2050: The Realization of Personalized Medicine through Crowdsourcing, the Quantified Self, and the Participatory Biocitizen" in: *Journal of Personalized Medicine*, 2(3). doi: 10.3390/jpm2030093.
- Swan, M. (2012b). "Sensor Mania! The Internet of Things, Wearable Computing, Objective Metrics, and the Quantified Self 2.0" in: *Journal of Sensor and Actuator Networks*, 1(3). doi: 10.3390/jsan1030217.
- Weitzman, E. R. *et al.* (2012). "Willingness to share personal health record data for care improvement and public health: a survey of experienced personal health record users" in: *BMC Medical Informatics and Decision Making*, 12(1), p. 39. doi: 10.1186/1472-6947-12-39.
- Wu, L. (2016). "A Digital Fingerprint For Your Tastebuds" in: *Forbes Magazine*, 30 September. Available at <https://www.forbes.com/sites/lesliewu/2016/08/31/a-digital-fingerprint-for-your-tastebuds/#7f3fca04922a> (Accessed: 27 June 2017).

Comparative Analysis of *Mycobacterium tuberculosis* Genome

J. Manning¹, C. Palu^{1,2}, B. Lawlor¹, M. MacAogáin¹, B. Kelly¹, P. Walsh¹

¹ NSilico Life Science Ltd

² University College Cork

Cork, Ireland

e-mail: cintia.palu@nsilico.com

Keywords: antibiotic resistance, translational bioinformatics, personalized medicine.

1. Introduction

The aim of this project is to explore the potential applied use of bioinformatics resources into infectious diseases healthcare practice. Assorted genome samples of the microorganism *Mycobacterium tuberculosis* were analysed using the Simplicity™ bioinformatics platform and the Comprehensive Antibiotic Resistance Database (CARD). *M. tuberculosis* is the most common cause of tuberculosis (TB) infections in humans and it's estimated to have caused the deaths of up to 1.8 million people in 2016 (WHO, 2016). Like all bacterial infections, TB is treated with antibiotics. The type and amount of antibiotics prescribed is dependent on the severity of the infection. However, due to the improper use of antibiotics and the ability of prokaryotes to quickly adapt to their environment, *M. tuberculosis* resistance to common antibiotics has become a major health issue (Mahmoudi & Iseman, 1993). This has further increased the severity of tuberculosis infections as they are becoming increasingly difficult to treat.

The most commonly used antibiotic to treat tuberculosis infections is isoniazid, which inhibits lipid synthesis (Timmins & Deretic, 2006). The other common treatments use streptomycin, rifampicin, ethambutol or pyrazinamide (Haas, et al., 1997) (Bodmer, et al., 1995) (Campbell, et al., 2011) (Andreis, et al., 2014) (Takiff, et al., 1994) (Feuerriegel, et al., 2012). Cases of multidrug-resistant strains are usually treated with bedaquiline, and it is, therefore, the last line of treatment for TB (Matteelli, et al., 2010).

2. Methodology

The data used to compile this report was obtained from the National Centre for Biotechnology Information's (NCBI) Sequence Read Archive (SRA). For consistency, only *M. tuberculosis* genomes sequenced using unpaired reads and the Illumina Genome Analyser® were selected. The first step of the analysis was to run Simplicity™ *De Novo* Assembly pipeline for each file. Secondly, the Simplicity™ Genome Comparison Pipeline was launched for all of the files, resulting in a phylogenetic tree that allowed us to compare the relationships of each strain of *M. tuberculosis*. Finally, each file was analysed for antibiotic resistance using the CARD's Resistance Gene Identifier (RGI) tool. This online tool provided a basis for comparing the results from Simplicity's Resistance Profiler pipeline.

3. Results and Discussion

We analysed 26 samples from experiments published in 2016 and 2017. The majority of the samples came from Africa: 5 from a *M. tuberculosis* genomic study in Ethiopia (ERP010284); 6 from HIV co-infected and uninfected patients in Kampala, Uganda (SRP095230) and 6 from unspecified African countries (SRP077478). We also analysed 2 samples from India, that actually are common H37Rv laboratory strains (SRP081552) and 7 from unknown location (NA, ERP010057). Both African countries and India are greatly affected by tuberculosis infections as they have an

underdeveloped healthcare system and they have the highest occurrences of tuberculosis worldwide (WHO, 2016). Each genome assembly was executed in less than 24h by Simplicity and the genome comparisons also were executed in less than 8h period.

The phylogenetic tree in Figure 1 presents the genetic relationship between the studied samples and the antibiotic resistance profile. Samples from Ethiopia (except for 1), India and the N/A location are all contained within separate groups based on location, suggesting that the samples within each of these groups display a relationship between genetics and geographic location. In contrast, it can be seen that both the samples from Kampala and the unspecified African locations do not have exclusive common ancestors. However, by analysing the phylogenetic tree it can be seen that the samples from Kampala are genetically similar whereas the samples from the unspecified African locations are more diverse. The diversity of the unspecified African samples is to be expected as they were obtained from an experiment that used samples from multiple African countries (NCBI, 2015).

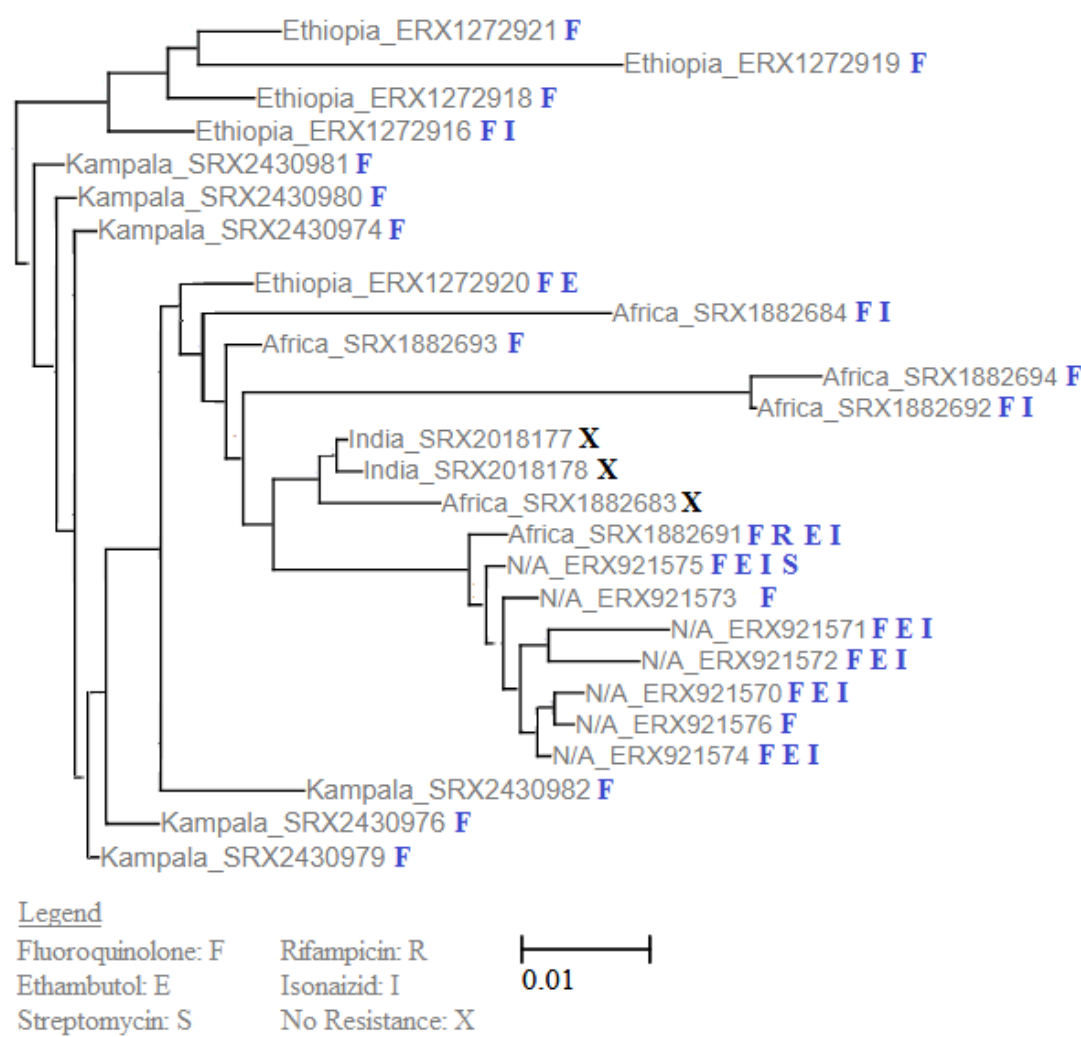


Figure 1: Phylogenetic Tree. A phylogenetic tree showing the genetic similarity and geographical location of each sample. The samples are named after geographic location followed by their SRA database ID. The resistance profile is presented on the right side of each sample, according to the legend.

Among our results, the samples from the N/A location are the most genetically similar, which indicates that they are likely to originate from the same geographic location. The N/A samples belong

to a large experiment analysing widely distributed clinical strains. Additionally, there is 1 file from the unspecified African locations that is closely related to the N/A samples and together they display the widest spectrum of resistance. Interestingly, these samples also share a common ancestor with 3 samples (1 African and 2 Indian), that do not display resistance to any of the 5 common antibiotics. The lack of resistance in the two samples from Indian is most likely due to the fact that they originated from the laboratory strain H37Rv. The close relationship between the 2 Indian samples and the sample SRX1882683 implies that sample SRX1882683 likely comes from the same strain. This is not surprising as the strain H37Rv is one of the most common and widely studied strains of *M. tuberculosis* (Camus, et al., 2002).

Regarding the other 23 samples, they all are resistant to fluoroquinolone, according to CARD analysis. This is unexpected as fluoroquinolone is not 1 of the 5 primary antibiotics that are commonly used to treat tuberculosis infections. This resistance may be due to increased usage of fluoroquinolones and the high frequency of mutations in the *GYRA* gene that is responsible for fluoroquinolone resistance (Thee, et al., 2015). According to our results, 9 samples are resistant to isoniazid, the most common TB treatment, and 7 are resistant to ethambutol. No geographic relationship between resistance to these 2 medications was found, being present in samples from Ethiopia, Africa, and N/A. Only 1 sample presented resistance to rifampicin (from Africa) and only 1 is resistant to streptomycin (N/A location), these 2 samples are the ones with the broadest resistance profile. No strain resistant to pyrazinamide or bedaquiline was identified.

Factors responsible for antibiotic resistance can be transmitted through horizontal gene transfer between bacteria. The results obtained from this project encourage further investigation into this process. Considering that multiple genes and mutations can confer resistance to the same antibiotic, performing sequence comparisons across the genes known to cause antibiotic resistance would provide relevant evidence on how these organisms acquired resistance.

4. Conclusion

The use of whole genome sequencing data enables the creation of an effective treatment plan and avoids unnecessary patient stress and costs of redundant treatments, by supporting the determination of patient-specific antibiotic regimes according to the infection strains. Furthermore, the data obtained from individual patients can be combined to monitor disease outbreaks across large areas, providing insight into how a disease evolves as it spreads and can also be used to prepare strain specific treatment plans. Bioinformatics pipelines such as the ones offered by Simplicity™ are key elements to bring this personalised medicine approaches to general practice, by speeding the analysis process and making it available to the non-bioinformatician public.

Acknowledgement

Paul Walsh and Brian Kelly are supported through the MetaPlat, grant agreement number H2020-MSCA-RISE 690998, funded under European Commission Horizon 2020 Programme. Cintia Palu is supported through Irish Research Council Enterprise Partnership Scheme (Postgraduate).

References

- Andreis, K. et al. (2014). "Acquired Resistance of *Mycobacterium tuberculosis* to Bedaquiline". *PLOS One*, 9(7), pp. 1-11.
- Bodmer, T., Zurcher, G., Imboden, P. & Telenti, A. (1995). „Mutation position and type of substitution in the beta-subunit of the RNA polymerase influence in-vitro activity of rifamycins in rifampicin-resistant *Mycobacterium tuberculosis*". In: *Antimicrobial Chemotherapy*, 35(2), pp. 345-348.

- Campbell, P. et al. (2011). "Molecular Detection of Mutations Associated with First- and Second-Line Drug Resistance Compared with Conventional Drug Susceptibility Testing of *Mycobacterium tuberculosis*". In: *Antimicrobial Agents and Chemotherapy*, 55(5), pp. 2043-2041.
- Camus, J.-C., Pryor, M. J., Médigue, C. & Cole, S. T. (2002). "Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv". In: *Microbiology*, Volume 148, pp. 2967-2973.
- Feuerriegel, S. et al. (2012). "Sequence analysis for detection of first-line drug resistance in *Mycobacterium tuberculosis* strains from a high-incidence setting". In: *Microbiology*, 12(90), pp. 1-10.
- Haas, W. et al. (1997). "Molecular analysis of katG gene mutations in strains of *Mycobacterium tuberculosis* complex from Africa". In: *Antimicrobial agents and chemotherapy*, 41(7), pp. 1601-1603.
- Mahmoudi, A. & Iseman, M. D. (1993). "Pitfalls in the Care of Patients With Tuberculosis Common Errors and Their Association With the Acquisition of Drug Resistance". *JAMA*, 270(1), pp. 65-68.
- Matteelli, A., Carvalho, A. C., Dooley, K. E. & A, K. (2010). "TMC207: the first compound of a new class of potent anti-tuberculosis drugs". In: *Future Microbiology*, 5(6), pp. 849-858.
- NCBI, 2015. *NCBI BioProject*. Available at: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA300846>, (Accessed 07 July 2017).
- NCBI, 2016. *NCBI BioProject*. Available at: <https://www.ncbi.nlm.nih.gov/bioproject/PRJEB9003>, (Accessed 07 July 2017).
- Takiff, H. et al. (1994). "Cloning and nucleotide sequence of *Mycobacterium tuberculosis* gyrA and gyrB genes and detection of quinolone resistance mutations". In: *Antimicrobial Agents and Chemotherapy*, 38(4), pp. 773-780.
- Thee, S. et al. (2015). "Fluoroquinolones for the treatment of tuberculosis in children". In: *Tuberculosis*, 95(3), pp. 229-245.
- Timmins, G. S. & Deretic, V. (2006). "Mechanisms of action of isoniazid". In: *Molecular Microbiology*, 62(5), pp. 1220-1227.
- WHO, 2016. *Global tuberculosis Report 2016*. Available at: http://www.who.int/tb/publications/global_report/en/, (Accessed 18 April 2017).
- WHO, 2016. *World Health Organisation*. Available at: <http://www.who.int/features/factfiles/tuberculosis/en/>, (Accessed 21 February 2017).

A Dynamic Product Line for an Electronic Health Record Management System for Cancer Care ¹

Matthias Coym, Bernhard G. Humm, Patrick Spitzer
Department of Computer Science
Hochschule Darmstadt – University of Applied Sciences
Haardtring 100, 64295 Darmstadt, Germany
bernhard.humm@h-da.de

Paul Walsh
NSilico Lifescience Ltd.
Bishopstown, Co. Cork, Ireland
paul.walsh@nsilico.com

Keywords: Product Line Engineering, Medical Information System, Electronic Health Record, HL7

1. Introduction

The field of medical information technology is advancing, standards are maturing, and systems such as electronic health records (EHRs) are increasingly being adopted (Yamamoto et al., 2012). However, clinical IT requirements are also continually evolving and expanding rapidly (Anhøj, 2003; Bates et al., 2014). Therefore, more efficient approaches are needed to deal with large amounts of EHR data while flexibility is needed to cope with the rapid evolution and expansion of diverse data models from different disease domains which is in turn driven by evolution in clinical care. Moreover the dynamic nature of medical informatics requirements can cause a prohibitive cost when it comes to integrating new features, such as customising treatment pathways, supporting multi-tenancy deployments and integrating and interoperating securely with third party eHealth systems. Each new eHealth software feature may requires both the extension of the data model and the potential need to integrate data with external IT systems, while maintaining trust and patient privacy, so as to provide a similar set of features across a range of deployments.

Therefore a medical information system for managing *electronic health records (EHR)* must be highly dynamic due to ever changing medical methods and practices. One way of dealing with dynamic requirements is to use a software product line engineering approach, which entails the use of software engineering methods, tools and techniques for creating a suite of software that share a common, managed set of features, from a shared set of software assets using a common means of production, to efficiently meet requirements of a particular sector².

The concept has emerged from the automotive industry where a product line of different cars models can be efficiently manufactured, from a common set of component parts. Application of this *product line* approach for EHR management must be even more dynamic in order to configure different clinical practices in various hospitals. This paper presents a dynamic

¹ This work was funded by the European Commission, Horizon 2020 Marie Skłodowska-Curie Research and Innovation Staff Exchange, under grant no 644186 as part of the project SAGE-CARE (SemAntically integrating Genomics with Electronic health records for Cancer CARE).

² Software Product Lines Carnegie Mellon Software Engineering Institute Web Site

product line for an EHR Management System for Cancer Care. EHR evolution is facilitated by the HL7 Reference Information Model³. A sophisticated rights management is being used for configuring tenant-specific services.

2. Problem Statement

The described concept shall outline the architecture of a product line for an EHR Management System in cancer care. The product line is not designed for the mass market.

The following fundamental requirements were identified:

1. Dynamic derivation of concrete products:
Concrete products shall be derivable from a software platform via configuration, covering all possible variants. Meaning that various medical specialties, medical information services and technical services have to be considered. Furthermore, configuration may concern every layer of the architecture.
2. Configurability on different levels:
Products meet the requirements of a customer on different levels, e.g. a default configuration for a hospital and a differing specific configuration for a single physician. Consequently, where appropriate, different configuration levels must be considered.
3. Multiple products on one instance:
The product line must be able to derive multiple products, each with its own configuration and data, on one running product line instance. It is mandatory, that concrete products must not know if they share their product line instance with other products.
4. Fundamental security mechanisms:
Only valid users shall have access to their application. They should have their own account with its own corresponding roles and rights. For example a melanoma physician must not be able to view patients, which are not assigned to his area of responsibility.
5. Change Tracking:
Changes to the persistent data in the database shall be traceable. Additionally, data must not be deletable to ensure compliance standards.
6. Flexible data model:
To meet the needs of all products, the data model must be flexible and extensible. Negative effects of the flexibility on efficiency and convenience shall be counteracted.
7. Flexible user interfaces:

³ <http://www.hl7.org/implement/standards/rim.cfm>

A flexible user interface for EHR management with reusable elements shall be designed, enabling to build up patient data dynamically.

3. Concepts

3.1. Architecture

The product line architecture is structured in three layers, which are loosely coupled. An overview of the architecture is shown in Figure 1.

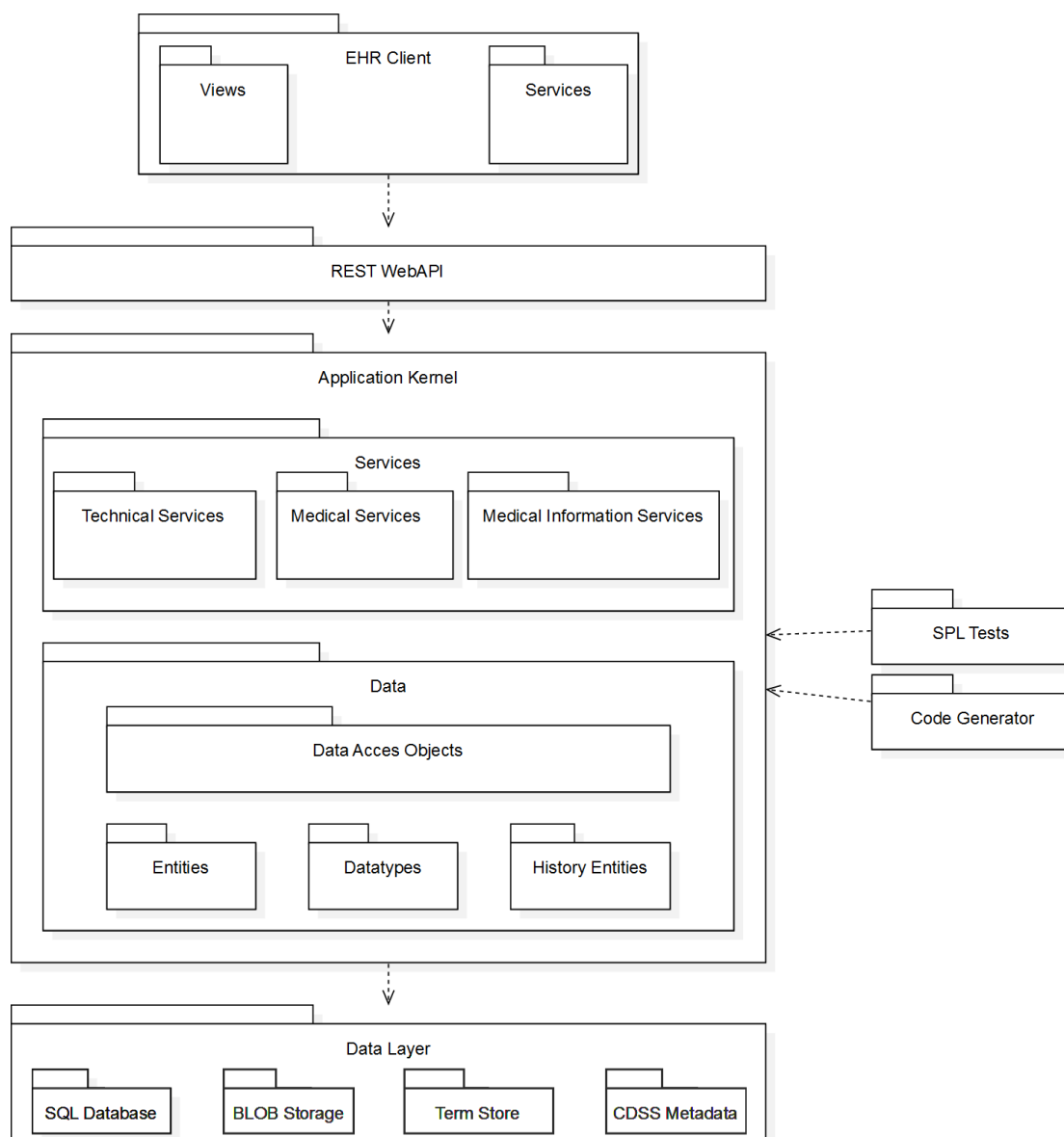


Figure 1: Product Line Architecture Overview

The top layer is the *EHR client*, which is the configurable client application for all products of the product line. It is designed as a web client which is loosely coupled via a *REST WebAPI*. The view package contains view components, being made up of the graphical user interface

(GUI) and corresponding presentation logic. The service package contains reusable client logic, shared by multiple view components. The next layer consists of the *REST WebAPI* and the *Application Kernel*. The *REST WebAPI* provides an interface for the client and is designed according to the Representational State Transfer (REST) principle (Fielding, 2000). The *Application Kernel* contains the business logic of the application. It is made up of different types of services and a data access layer. The data access layer uses data access objects (DAOs) and an object-relational mapper (ORM) for accessing a relational database. The packages *Entities*, *DataTypes* and *HistoryEntities* contain the data items. The bottom layer contains the persistent data of the Software Product Line. It is made up of different storage types. The relational *SQL Database* stores the EHR data, the *BLOB Storage* is used for storing files, e.g. images, while the *Term Store* and *CDSS Metadata* are used for the *Medical Information Services*. See (Beez et al., 2015) and (Idelhauser et al., 2016) for further details.

3.2. Configuration Hierarchy

The product line can be configured on three different levels, as shown in Figure 2.

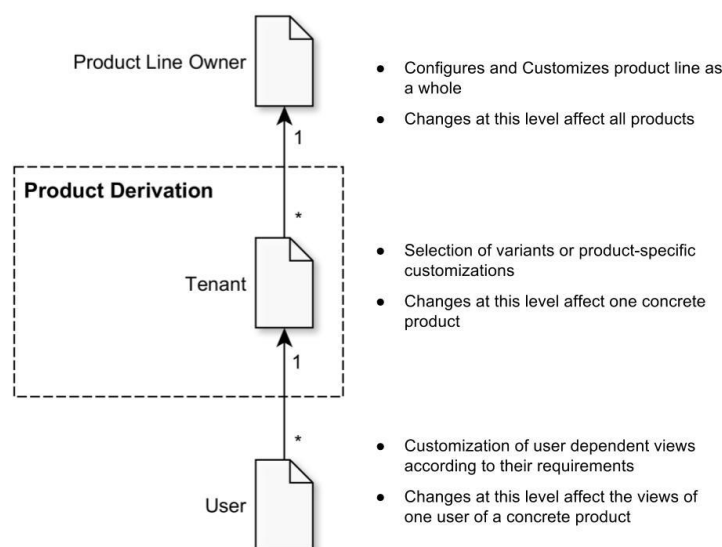


Figure 2: Configuration Hierarchy

At the highest level the *Product Line Owner* decides which features are supported by the Product Line in the first place. At the second level the product derivation takes place. The *Tenant* selects specific features from the product line called variants to create a specific product according to its needs. A tenant may be a hospital group, an individual hospital, or a hospital department. At the lowest level, each user of the product has the option to customize the views according to his or her needs.

In a classic product line, the product derivation takes place at compilation time. In a dynamic product line this must be possible at runtime. To make this possible, the configuration is tightly integrated in the access management. Tenant rights are given according to the chosen variants. The individual user rights are a subset of the tenant rights. When a user logs in with his or her

credentials, the configuration is loaded and the correct views are shown. If a user has access to more than one medical service, he is able to switch between these services.

3.3. Access Management

For the access management of the dynamic product line, we use the approach of *role-based access control* (Ferraiolo and Kuhn, 1992). The implementation is additionally based on the security, authentication, and authorization documentation for ASP.NET Web API (Anderson, Pasic and Dykstra, 2012).

The attributes of a user are bundled in an object called the user account, which contains a userID as the unique identifier. For the authentication process, a user has to provide his or her userID and a credential to the system. The credential provides evidence for the user's claimed identity, e.g. a password which is only known to the user and the system. If userID and credential match, the user is authenticated. A security context or token can be bound to the user's connection as evidence that the user has already been authenticated. After the authentication, the requests of the user can be authorized. Authorization can be defined per operation or resource that an interface offers and is realized via a role based access control model.

Access permission to the EHR system depend upon two factors. The first are the permissions granted to individual users are, as usually, according to their role in the hospital as shown in Figure 3. An administrator for example might have the right to manage users but should not be able to access the patient data.

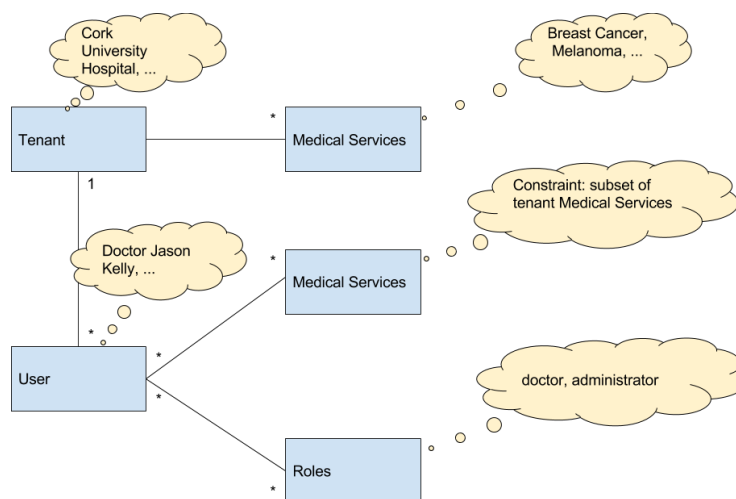


Figure 3: Access Management

Additionally, there are permissions granted to individual users according to their tenant's access permissions like the access to medical services. As described in Section 3.1, a tenant's product is derived from his configuration of variability points. For example, the Cork University Hospital might have a product with the breast cancer medical service, but not the melanoma medical service. Therefore a consultant from Cork University Hospital is not able to access the melanoma medical service.

Apart from the restrictions at the application layer and client layer, access is also restricted at the database. It must be ensured that no tenant can access data of another tenant. To ensure this, each database entry is enhanced by a tenantID. Every database access is performed with this tenantID as an extra restriction. This approach is called *row level security* or *shared table* (Jacobs and Aulbach, 2007).

3.4. Flexible Data Model

The flexible data model is based on HL7 Reference Information Model (RIM) (ISO, 2006). This offers the following advantages:

1. It is a common, maintained, and well documented ISO-Standard for health care information systems.
2. The model adds the flexibility to extend the data model without changing the relations between entity classes.
3. The HL7 RIM appends the ability to create new links between concrete objects at run-time of the application through the usage of association classes.
4. Compatibility to other systems is ensured, which implement the HL7 RIM and therefore between all products of the software product line.

See Figure 4.

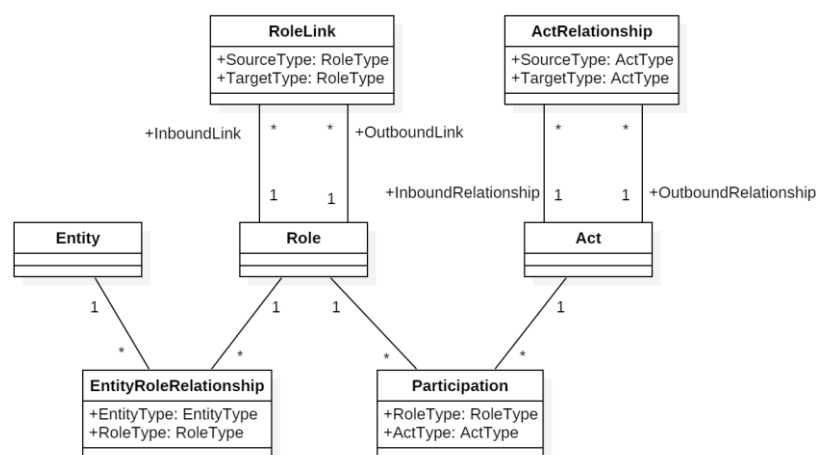


Figure 4: Flexible Data Model oriented at HL7 RIM

In HL7 RIM according to (ISO, 2006), the association between entity and role is not realized with an association class. In order to be consistent with the other association classes *Participation*, *RoleLink* and *ActRelationship*, we introduced a new association class called *EntityRoleRelationship*. This allows dynamically changing relationships between entities, roles and acts, while still being compatible with HL7 RIM.

With the flexible data model it is possible to support a rich variant of products, while still allowing tenant-specific associations. For example, one hospital might prefer to associate a

medication directly with the patient while another hospital prefers, instead, to associate it with the cancer issue attached to the patient. Both preferences can be realized without modifying data model or any database table.

Also refer to Humm and Walsh (2015).

4. Related Work

Neither the combination of *Dynamic Software Product Line* with *reconfigurable multi tenant aware Software as a Service* applications, nor *EHR Management Systems* (EHRMS) are new research fields. The combination of an *EHR Management System* with a *Dynamic Software Product Line*, however, is still an unexplored field.

Bahga and Madisetti outlined a cloud-based approach for the design of an interoperable EHRMS (Bahga and Madisetti, 2013). They describe the architecture for an EHRMS, but with the explicit focus on semantic interoperability, data integration, and security. Albeit, the paper focuses on single-system engineering instead of creating a product line.

Kuo (2011) summarizes the general opportunities and challenges of cloud computing to improve health care services, to benefit healthcare research and to change the face of health information technology.

The SPLiCE (Software Product Line for healthCarE) project proposes a model-driven engineering method for healthcare information systems. A Software Product Line shall thereby be created, which integrates clinical data models, described according to the “openEHR” specifications, and architecture models, specified in the “Acme” architecture description language (Gomes et al., 2012).

5. Conclusion and Future Work

In this paper, we have presented a dynamic product line for an EHR Management System for Cancer Care. EHR evolution is facilitated by the HL7 RIM. A sophisticated rights management is being used for configuring tenant-specific services.

We have implemented the product line concept prototypically and applied it to three medical services: melanoma treatment, breast cancer treatment, and a receptor database. The initial results are most promising and the reception by clinical consultants is positive. Therefore, it is planned to further develop the prototype to a productive system.

Towards this end, future work is needed. After completing development, extensive testing is required including performance testing. Furthermore, we plan an extensive usability study with end users.

References

- Anderson, R., Pasic, A. and Dykstra, T. (2012). *Security, Authentication, and Authorization in ASP.NET Web API*. [online] Docs.microsoft.com. Available at: <https://docs.microsoft.com/en-us/aspnet/web-api/overview/security/> [Accessed 1 Sep. 2017].
- Anhøj, J. (2003). "Generic Design of Web-Based Clinical Databases". In: *Journal of Medical Internet Research*, 5(4), p.e27.
- Bahga, A. and Madiseti, V. (2013). "A Cloud-based Approach for Interoperable Electronic Health Records (EHRs)". In: *IEEE Journal of Biomedical and Health Informatics*, 17(5), pp.894-906.
- Bates, D., Saria, S., Ohno-Machado, L., Shah, A. and Escobar, G. (2014). "Big Data In Health Care: Using Analytics To Identify And Manage High-Risk And High-Cost Patients". In: *Health Affairs*, 33(7), pp.1123-1131.
- Beez, U., Humm, B.G. and Walsh, P. (2015), "Semantic AutoSuggest for Electronic Health Records", In: *2015 International Conference on Computational Science and Computational Intelligence (CSCI)*, 760-765.
- Fielding, R. (2000). *Architectural Styles and the Design of Network-based Software Architectures*. University of California. Available at: https://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation.pdf [Accessed 27 Jun. 2017].
- Ferraiolo, D.F. and Kuhn, D.R. (1992). "Role Based Access Controls". In: *15th National Computer Security Conference*. pp.554 - 563.
- Gomes, A., Ziviani, A., Correa, B., Teixeira, I. and Moreira, V. (2012). "SPLiCE". In: *Proceedings of the 2nd ACM SIGHIT symposium on International health informatics - IHI '12*. pp. 721-725.
- Humm, B.G. and Walsh, P. (2015). "Flexible yet Efficient Management of Electronic Health Records". In: *2015 International Conference on Computational Science and Computational Intelligence (CSCI)*, 771-775.
- Idelhauser, J., Beez, U., Humm, B.G., Walsh, P. (2016). "A Clinical Decision Support System for Personalized Medicine". In: *U. Bleimann, B. Humm, R. Loew, I. Stengel, P. Walsh (Eds): Proceedings of the 2016 European Collaborative Research Conference (CERC 2016)*, pp 132 - 145. Cork, Ireland, 2016. ISSN 2220 - 4164.
- Jacobs, D. and Aulbach, S. (2007). "Ruminations on Multi-Tenant Databases". In: *BTW Proceedings, volume 103 of LNI*. Aachen Germany: GI, pp.514--521.
- Kuo, A. (2011). "Opportunities and Challenges of Cloud Computing to Improve Health Care Services". In: *Journal of Medical Internet Research*, 13(3), p.e67.
- ISO (2006). *International Organization for Standardization (ISO): Health informatics - HL7 version 3 - Reference information model - Release 1*. Available at: http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=40399.
- Yamamoto, K., Sumi, E., Yamazaki, T., Asai, K., Yamori, M., Teramukai, S., Bessho, K., Yokode, M. and Fukushima, M. (2012). "A pragmatic method for electronic medical record-based observational studies: developing an electronic medical records retrieval system for clinical research". In: *BMJ Open*, 2(6), p.e001622.

A Process mining framework to simulate and predict patient pathways and health related costs based on medical claim data

Fatih Sahin, Rafael Accorsi, Monika Frey-Luxemburger, Martin Knahl, Shirley Atkinson
 University of Plymouth, Faculty of Science and Engineering
 University of Furtwangen, Faculty of Business Information Systems
 e-mail: fatih.sahin@plymouth.ac.uk

Keywords: Process mining, health insurance, care flow, complex event data, predictive simulation

1 Introduction

Health insurances around the world follow various legal frameworks, fee models and insurance plans but they face all one common issue - tremendously high health care cost: In 2014 16.6 % of the Gross domestic product from the United States was spend for health costs followed by Switzerland with 11.4 % which correlates to 9024 US Dollar and 6787 US Dollar per capita (OECD,2016).

There are several reasons influencing the high costs like the longer life expectation or rise of chronic disease and obesity. According to the WHO (World Health Organization, 2016) chronic illness occurrence will increase by 57% in the year 2020.

Further health insurances operate usually within a framework where they pay for health care services based on the service provided rather than the treatment efficiency or outcome. This leads to a chain of cost pressure and cuttings: Governments or policies decrease the reimbursement rate for a provided services, which decreases also the revenue of a health care providers but do not improve the quality of patient care in any matter. Hereby it is worth to mention that overuse or doctor shopping led also to this climate within the health care domain (Brenner and Hricak 2010; Wang et al. 2008; Baillargeon, et al. 2000). According to a report from the Institute of Medicine 750 billion US Dollar out of \$2.6 trillion have been spent for actions, which did not have any positive impact on the health status of patients in the United States (Newhouse and Garber 2013).

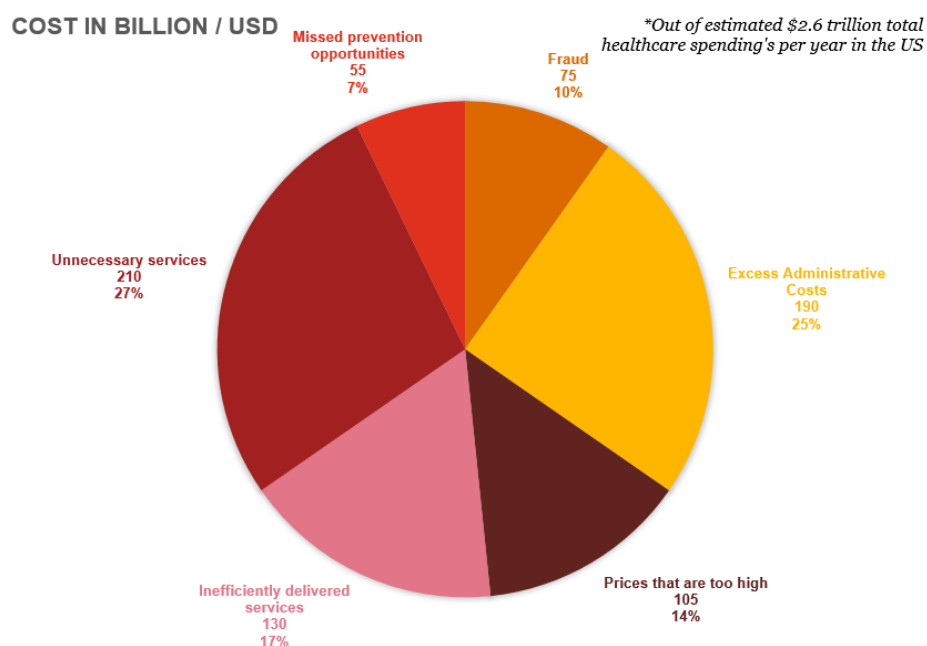


Figure 1: Ineffective health care spending distribution (Kliff 2016)

The illustration Fig.1 shows the various spending categories. Hereby, it is worth to highlight especially the spending categories ‘unnecessary services’, ‘inefficiently delivered services’ and ‘missed prevention opportunities’ – which we see as potential improvement areas of the relevant processes based on the usage of process mining techniques which will be introduced later.

In the majority of western countries an enormous amount of data is generated and processed within the health care domain (Wullianallur and Raghupathi 2014). The health sector, in particular the health insurance, is the one that achieves one of the highest level of process automation and digitization. Paradoxically, although it is also the sector, which could gigantically profit from the insights inherent to the collected data in various ways, it currently largely fails to do so. In fact, health insurances are quite ineffective when it comes to proactive actions (Goetzel 2009). They are particularly weak and rudimentary when it comes to understanding the patient pathway or treatment pathway, predicting the likelihood that someone with a specific disease could also suffer in near future or long term from another disease respectively simulating and predicting cost.

To address the issue of lack in transparency and limited view on health care processes, the potential impact of process mining techniques are presented in this paper. We introduce a novel approach to monitor and predict patient care flow in order of performance measurement and improved cash flow for health insurances.

2 Methodology

Process mining is a relatively young research discipline which can be classified between computational intelligence and data mining on the one hand and process modeling and analysis on the other hand. ‘The idea of process mining is to discover, monitor and improve real processes (i.e., not assumed processes) by extracting knowledge from event logs readily available in today’s (information) systems. Process mining includes (automated) process discovery (i.e., extracting process models from an event log), conformance checking (i.e., monitoring deviations by comparing model and log), social network/ organizational mining, automated construction of simulation models, model extension, model repair, case prediction, and history-based recommendations’ (Van der Aalst et. al. 2010).

In contrast to typical mining techniques like e.g. classification, regression or clustering which do not concentrate on the full end to end process but more in singular aspects, process mining concentrates on the complete life cycle of processes based on event data (Van der Aalst 2012).

Through its capability of a whole process chain coverage and experiments we have carried out with real data coming from health-insurances, we evidenced that process mining produces a data layer which allows us to mine patient pathways. Patients do not exist as a unique data point much more considering a typical patient history as multiple records about multiple diseases (event) at various times (timestamp). Further, in order to identify a relation between e.g. a current disease and previous disease it is mandatory to use an approach like process mining which is treating multiple events as a set of traces. Our experiments with other techniques, such as complex event processing, did not produce the expected results, as they only put events into a context, but not focusing on a particular patient’s profile.

3 Health Insurance Environment and Data Structure

Health insurances receive a large amount of medical claims respectively bills sent from health care service provider to health insurances. These claims usually contain patient master data (personal information) and information about the type of service provided by the e.g. physician and a diagnosis

code like the ‘International Classification of Diseases’ (ICD Code) maintained by the World Health Organization (WHO):

‘The ICD is designed as a health care classification system, providing a system of diagnostic codes for classifying diseases, including nuanced classifications of a wide variety of signs, symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or disease. This system is designed to map health conditions to corresponding generic categories together with specific variations, assigning for these a designated code, up to six characters long’ (World Health Organization 2014) .

Based on the available ICD codes it is also possible for health insurances to see diagnosis and allocate relevant costs based on a diagnosis. This ICD codes can be used as well to track the “patient flow” from a health insurance perspective and also to predict expected cost for a patient once an initial diagnosis is given.

Put another way, health insurances have a comprehensive data set about following key elements:

- Patient master information (Name, Address, Gender, Age)
- Type of service provided (e.g. ambulant therapy, stationary)
- Cost of service
- Disease code linked to each service provided (e.g. ICD Code)
- Timestamp when a specific service was provided
- Information about service provider (e.g. Gastroenterologist, Dentist)
- Health insurance plan of payer (e.g. premium package health insurance)
- Optional data (e.g. profession of insured person, marital status)

4 Contribution and Outlook

Based on the capabilities of process mining techniques, process nature, data obtained and our previous experiments we will define a target process mining framework considering the special conditions of the health insurance landscape and identify most appropriate class of process mining algorithms to obtain following key outcomes:

- A process mining algorithm and configuration tailored for mining patient pathways taking into consideration its particular characteristics and weighting different paths, e.g. depending on the relevance of a treatment.
- A continuous monitoring technique that combines complex event processing and the results of process mining to assess the state of a particular patient pathways on demand and on premise.
- A simulation method that allows us to predict pathways computed with different variables. This will allow us to forecast the impact of various patient behaviour to the health costs and cash flow of a health insurance.

Although challenging, the establishment of above described points will contribute to research as there is no to very limited research on applying process mining techniques to improve transparency in health care related services from a health insurance perspective. The nature of process mining

algorithms and patient data available in health insurance legacy systems appear like a promising match which could lead among others to following benefits:

Target 1: The ability to have a full understanding how patients are moving during a disease and between diseases.

Potential benefits: Health insurance understands better root cause of cost and have improved transparency about cost drivers. Identification of high cost patients respectively high cost disease. Once a health insurance has the ability to see process chain between diseases or relations between diseases it can offer preventive actions to its members to avoid potential other disease.

Target 2: Compliance - Identification of process anomalies

Potential benefits: Health insurances could identify e.g. health service providers, which are not following official or recommended treatment guidelines for a specific disease. Further, e.g. knowledge gaps of physicians can be detected – which can be used to measure the performance of the physicians and recommend patients proactively health service provider which operate in line with medical guidelines.

Target 3: Cost monitoring and prediction

Potential benefits: Based on process patterns in the patient pathways or disease pathways health insurance can add a cost dimension to each specific process. This gives the health insurance a new ability of monitor, control and predict costs. Depending where in the process a patient is located it is possible to give a likelihood about every next activity in the process chain as well about the potential cost based on the path a patient is following. Process mining allows through model simulation to create predictive “what if analysis”. In other words, health insurances could play with process mining techniques various e.g. given patient scenarios through and identify the corresponding cost.

Target 4: Monitor treatment efficiency by followed path

Potential benefit: Capability to monitor and benchmark treatment efficiency from a health insurance perspective. Identification of ‘best practice’ treatment pathway for specific diseases based on treatment guidelines.

The proposed work will deliver a foundational frame for process mining operations based on health insurance data and evaluate existing process mining algorithms in regard of suitability to achieve above described outcomes which may lead to a significant patient care reduction and improved process transparency.

5 References

- Baillargeon, J., et al. (2000). The disease profile of Texas prison inmates. In: *Annals of epidemiology* 10.2, pp.74-80.
- Brenner, D. J., and Hricak, H. (2010). Radiation exposure from medical imaging: time to regulate?. In: *Jama* 304.2: pp. 208-209.
- Goetzel, R. Z., (2009). Do prevention or treatment services save money? The wrong debate. *Health Aff Millwood* 28(1), pp.37–41.
- Kliff, S. (2012). We Spend \$750 Billion on Unnecessary Health Care. Two Charts Explain Why. In: *Washington Post*. The Washington Post, 7 Sept. 2012. [Accessed 30 Sept. 2016].

Newhouse, J. P., and Garber, A. M., (2013). Geographic variation in health care spending in the United States: insights from an Institute of Medicine report. In: *JAMA* 310.12: pp.1227-1228.

OECD, 2016. Health spending indicator. doi: 10.1787/8643de7e-en [Accessed 30 September.2016].

Raghupathi, W., and Viju R. (2014). "Big data analytics in healthcare: promise and potential." In: *Health Information Science and Systems* 2.1: p.1.

Van Der Aalst, W., et al., (2011). Process mining manifesto. In: *International Conference on Business Process Management*. Springer Berlin Heidelberg.

Van Der Aalst, W., (2012). Process mining: Overview and opportunities. In: *ACM Transactions on Management Information Systems (TMIS)* 3.2: p. 7.

Wang, Y., et al., (2008). Will all Americans become overweight or obese? Estimating the progression and cost of the US obesity epidemic. In: *Obesity*16.10: pp. 2323-2330.

World Health Organization (2016). "The Global Burden of Chronic", n.d. [online] [Accessed 30 Sept. 2016].

World Health Organization (2014). "International Classification of Diseases (ICD)". [Accessed 10 July 2016].

Evidence-Based Medical Recommendations for Personalized Medicine ¹

Bernhard G. Humm, Fabian Lamba, Tino Landmann, Marc Steffens
Department of Computer Science
Hochschule Darmstadt – University of Applied Sciences
Haardtring 100, 64295 Darmstadt, Germany
bernhard.humm@h-da.de

Paul Walsh
NSilico Lifescience Ltd.
Nova Center, Belfield Innovation
Park, Dublin 4, Ireland
paul.walsh@nsilico.com

Keywords: Clinical Decision Support System, Evidence-Based Medicine, Personalized Medicine, Electronic Health Record

Abstract: Medical consultants face increasing challenges in keeping up-to-date with the rapid development of new treatments and medications. Evidence-based medical guidelines (EBM) are provided by healthcare institutions to support consultants in clinical decision making. However, due to time constraints in their daily routine, consultants rarely find time to research and cross-check medical decisions with such guidelines. This paper presents an approach for personalized EBM recommendations, tailored to each particular patient, without any research effort by the consultant. It is available at the point-of-care and requires electronic health records (EHR) to be semantically linked with EBM guidelines.

1. Introduction

Due to the rapid development of new treatments and medications, along with heavy workloads and limiting time constraints in their daily schedule, medical consultants face the challenge of continuously keeping up-to-date with the latest advances in clinical practice. In addition, treatment methods quickly become obsolete, change or are replaced by newly developed protocols (Marchant and Lindor, 2013).

Personalized medicine aims to tailor medical decisions, practices, interventions or products to the individual patient based on their predicted response or risk of disease (Academy of Medical Sciences, 2015). Evidence-Based Medicine (EBM) is a way of evaluating treatment in clinical practice. It is used to get the best available research for diagnosis and treatment to improve decision making (Masic et al., 2008). This is done by various types of research methods like systematic reviews, randomized controlled trials, meta-analysis, case report or practice guideline analysis (Hung et al., 2015). Most existing clinical decision support systems focus on providing evidence-based medical information, guidelines and summaries, written by networks of medical experts. By constantly updating information, these guidelines and summaries present the current state of medical best practice (Antes et al., 1998). But

¹ This work was funded by the European Commission, Horizon 2020 Marie Skłodowska-Curie Research and Innovation Staff Exchange, under grant no 644186 as part of the project SAGE-CARE (SemAntically integrating Genomics with Electronic health records for Cancer CARE).

consultants rarely have time to read and navigate through guidelines and such services. Also summaries and guidelines are sometimes written in a confusing way, which cost additional time and makes it more difficult to understand the required information (Obst et al., 2015).

This paper presents an approach for personalized EBM recommendations, tailored to each particular patient, without any research effort by the consultant. It is available at the point-of-care. This combines the advantages of personalized medicine and EBM, while alleviating the problems with current EBM solutions. We present the concept and prototypical implementation of a personalized EBM service which semantically links electronic health records (EHR) with published EBM guidelines. This service can be integrated into an EHR application as module of a Clinical Decision Support System (CDSS) (Kawamoto et al., 2005).

2. Problem Statement

Based on intense consultation with clinicians involved in the treatment of melanoma, and in accordance with Idelhauser et al. (2016), we have identified the following requirements for providing an EBM solution for CDSS:

1. Functional Requirements

- 1.1. *EBM recommendations*: The consultant shall get evidenced-based medical recommendations for treating cancer patients.
- 1.2. *Personalized*: The information provided shall be tailored to the medical condition of a particular patient, based on their EHR.
- 1.3. *Pro-active*: The EBM recommendations shall offer information pro-actively without additional data entry by the consultant.
- 1.4. *Easily comprehensible*: The EBM recommendation service shall provide information at the point of care and it should be possible to quickly retrieve more detailed information.
- 1.5. *Non-interfering*: The EBM recommendations shall not interfere with the consultant's EHR workflow.

2. Non-Functional Requirements

- 2.1. *Usability*: The EBM recommendations shall be intuitive to use and self-explanatory.
- 2.2. *Performance*: The application shall allow users to work in their own pace in a pleasant way; In particular, response times for all interactions with the EBM recommendations shall be less than 1s.
- 2.3. *Maintenance*: Maintaining changes in EBM recommendations data shall induce only a small effort.
- 2.4. *Extensibility*: The ongoing extension of the EBM recommendations with new information sources shall be facilitated with moderate implementation effort.

3. A Personalized EBM Recommendation Service

3.1. User Interaction Concept

We illustrate the user interaction concept of the personalized EBM recommendation service by means of an example in melanoma treatment. Fig. 1 shows anonymised summary data of a melanoma patient from an EHR application.

i Issue Data			
Type:	Melanoma in situ	Lesion Site:	Lower limb
Clinical Stage:	IB	Breslow Thickness:	0.8 mm
TStage:	PT4A	Mitosis:	1
Drugs:	Ipilimumab Interferon	BRAF Status:	

Figure 1: Patient-related issue data

In this example, the patient suffers from melanoma in situ at stage IB, with a Breslow thickness of 0.8 mm. Based on the EHR data, without interaction by the consultant, the relevant page of the NCCN EBM guideline (National Comprehensive Cancer Network, 2017) for melanoma treatment is retrieved (Fig. 2).

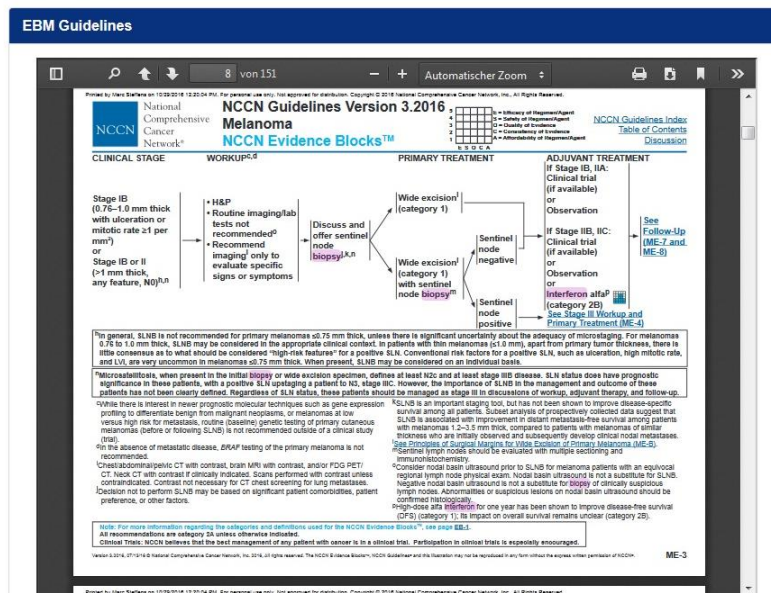


Figure 2: EBM recommendation

The guideline is structured as a decision tree with the relevant path (Stage IB, Breslow thickness 0.67 - 1.0 mm) displayed. Appropriate diagnosis and treatment procedures are recommended. Terms matching the EHR, e.g., interferon, are highlighted. The personalized EBM Guideline service is part of a CDSS which the consultant may or may not open when

working with the EHR. If interested, the consultant may read footnotes and follow hyperlinks for more details.

3.2. Information Sources

EBM recommendations are edited and published regularly by various organisations and companies. Tab. 1 gives an overview of prominent EBM information providers as a result of a survey we conducted.

Name	Description	API	Access	Volume
BMJ Best Practice	Evidence-based information to offer step-by-step guidance on diagnosis, prognosis, treatment and prevention.	yes	subscription	undisclosed
DynaMedPlus	Evidence-based clinical overviews and recommendations. Content updated daily. Also offers calculators, decision trees and unit and dose converters.	yes	subscription	> 3,200 topics and > 500 journals
EBMeDS	Platform-Independent web service CDSS with EBM module.	yes	commercial	undisclosed
Medscape / eMedicine	Largest clinical knowledge base available freely. Articles updated yearly. Also available as mobile application.	no	free, registration required	~6,800 articles
NCCN	Guidelines for treatment of cancer by site offering decision trees. Compiled by panels of experienced medicians.	no	free, registration required	~60 documents
Physician Data Query	Cancer database from the U. S. <i>National Cancer Institute</i> . Contains peer-reviewed information on cancer treatment in the form of summaries for patients and professionals.	no	public	Only cancer domain
UpToDate	Popular evidence-based POC tool for a wide range of disciplines but targeted on internal medicine. Extensive peer-review process to ensure accurate and precise recommendations	yes	subscription, some articles free	~8,500 topics

Table 1: EBM information providers

Some organizations like the US National Cancer Institute or the National Comprehensive Cancer Network provide guidelines for free. Commercial information providers like Wolters Kluwer (UpToDate) or EBSCO (DynaMedPlus) provide paid access to a large number of EBM guidelines published world-wide.

All information providers provide web access with full-text search. Some providers offer an application programmer's interface (API) for performing full-text search. However, to the best of our knowledge, no information provider offers an API for matching conditions, e.g., Breslow thickness 0.67-1.0 mm. Such conditional matching logic is required for the personalized EBM recommendation service.

3.3. Software Architecture

The personalized EBM recommendation service is part of a Clinical Decision Support System (CDSS), which is integrated into an EHR application. The application is organized as a three layer architecture where each layer consists of components that encapsulate logically separable units. See Fig. 3.

3.4. EHR Data Extraction

First, relevant the EHR data to be mapped to EBM guidelines are extracted from the EHR. In the melanoma example, this includes EHR attributes such as issue type, clinical stage, tumor stage, Breslow thickness, medication, ulceration, etc.

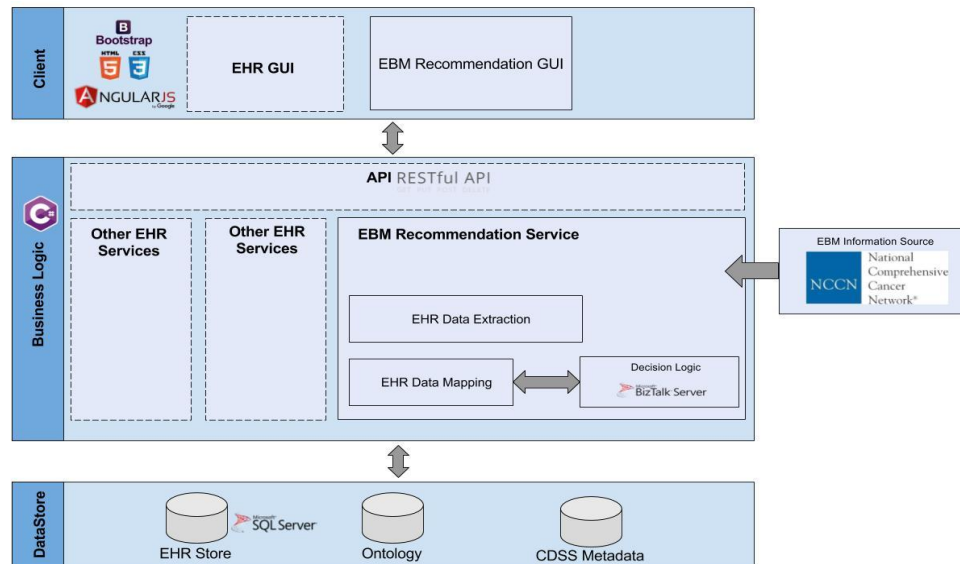


Figure 3: System architecture

3.5. Decision Logic

Identifying sections of EBM guidelines which are relevant for a particular patient under treatment requires more than full text search. Consider the example of Breslow thickness 0.8. Searching for the string “0.8” in the text of the NCCN guidelines will not match the relevant page ME-3, since on this page, the condition is formulated as “0.67-1.0 mm thick”. Therefore, some explicit decision logic is required for matching extracted EHR data to sections of the EBM guidelines. See Fig. 4 for an example rule.

Here, the following rule is shown: “If clinical stage is IB and Breslow thickness is between 0.76 - 1.0 mm, then section ME-3 on page 8 is relevant”. This rule is edited using a business rule composer (BRC), here MS BizTalk (Microsoft BizTalk, 2017).

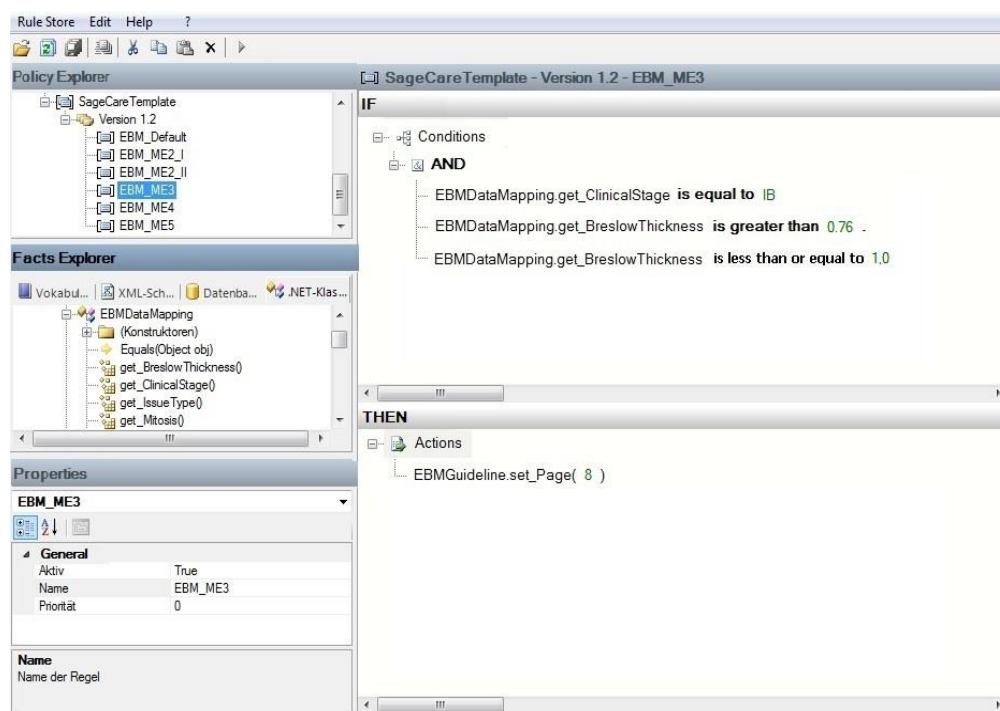


Figure 4: Example rule in business rule engine (MS Biztalk Server)

Applying the rules with the extracted EHR data as input in a business rule engine (BRE) will match the relevant section of the EBM guidelines which can then be displayed to the consultant in the CDSS GUI.

Using a BRC in combination with a BRE has advantages over coding the decision logic in a conventional programming language. It allows adding or modifying business rules by trained medical administrators whenever new or modified EBM guidelines are published.

3.6. Prototype Implementation

The EBM recommendation service has been implemented prototypically as part of an EHR application for melanoma treatment (Humm and Walsh, 2015; Beez et al., 2015; Idelhauser et al., 2016). The application is implemented in C# using .NET and MS SQL Server on the server side, and in HTML5 / CSS / JavaScript on the client side, using Bootstrap and AngularJS.

As a data source for the personalized EBM recommendation service, the NCCN guidelines have been selected. Microsoft BizTalk is as BRC / BRE. The HTML Inline Frame (Iframe) functionality for embedding documents is used to display the NCCN PDF files. Medical terms extracted from the EHR are highlighted in order to attract the consultant's attention.

4. Evaluation

In this section, we compare the concept and prototypical implementation of the personalized EBM service with the requirements from Section 2. As defined in Requirement 1.1, the physician receives evidence-based medical recommendations for the treatment of his patient. These recommendations are based on the clinical condition of the patient, as the

recommendation is selected by using the particular patient's specific EHR data. This makes the concept *personalized* (1.2).

When the consultant opens the patient's EHR, the service automatically displays the appropriate EBM recommendation in the CDSS, without requiring the consultant to enter additional data. This makes the service *pro-active* (1.3).

Requirement 1.4 (*easily comprehensible*) depends on the EBM guidelines selected. The NCCN guidelines provide step-by-step diagnosis and treatment instructions. Detail information may be obtained following hyperlinks. The consultant may or may not decide to open the CDSS GUI. So, he or she is not interrupted in his clinical workflow (1.5 *non-interfering*).

Freeing the consultant from the need to actively search for EBM recommendations also improves *usability* (2.1).

When opening a patient's EHR, the relevant EBM guideline is matched in the background. In the current implementation, this may take up to 5s. As soon as the consultant opens the CDSS GUI, all interactions such as scrolling or following hyperlinks have a response time of less than 300ms (tested with Mozilla Firefox performance analysis), clearly meeting Requirement 2.2 (*performance*).

The NCCN guidelines are updated regularly to always reflect state of the art. However, we expect modifications which actually affect whole sections of the EBM recommendations and, hence, decision logic rules, to be rare. If such a change is necessary, the *maintenance* effort is low (2.3) due to the use of a BRC.

Replacing the EBM information provider, respectively adding a new information provider like UpToDate (2.4 *extensibility*), requires programming. Due to the modular structure of the system architecture, only adaptors need to be implemented. Depending on the API to be integrated, the effort is estimated to be up to a few hundred lines of code. This could be considered moderate.

5. Related Work

(Kwag, 2016) evaluated 26 web-based point-of-care information sources in a study. Some of the most valued systems in terms of scope, evidence and quality were, among others, *DynamedPlus*, *UpToDate* and *BMJ Best Practice*. The sources were characterized by the fact, that they offer extensive evidence-based information to assist clinicians in decision-making. They all work in a similar way: in order to retrieve information, consultants can use full-text search (Ebsco Health, 2017; BMJ Best Practise, 2017; Wolters Kluwer, 2017). A condition-based search is not offered. In particular, semantic matching of EHR data with EBM guidelines as described in this paper is not possible.

IBM Watson is a system capable of answering questions posed in natural language. It has been adapted for various application domains, including health care and in particular oncology (IBM Watson Health, 2017). To keep up-to-date, Watson is constantly updated and trained with new information. Consultants can use the corpus of knowledge by asking relevant questions, using

a detailed search mask. However, automatic matching of relevant EBM guidelines based on the EHR is also not provided.

6. Conclusions and Future Work

EBM is a good method to assist physicians in decision-making, which results in an improved treatment of patients (Borgerson et al., 2005). So far, there are a number of systems and approaches to support consultants with EBM, and development is fast-paced. However, it is a great challenge to integrate these systems sensibly into the workflow of a clinician whose daily workload is high (Fowler et al., 2014). So far, adoption of EBM in clinical practice is low.

In this paper, we have presented a concept and prototypical implementation of a personalized EBM service to alleviate those issues. The focus was therefore on the usability of the service, in order to allow consultants to retrieve information they need without having to actively search in the system. For this purpose, the data extracted from the EHR was used to semantically match EBM guidelines and present the relevant sections to the consultant.

The EBM recommendation service has been integrated into an EHR application for melanoma treatment as part of a CDSS. Towards this end, future work is required. A comprehensive analysis with consultants in the field needs to take place. It is planned to put the EHR application including CDSS into productive use at hospitals as a commercial pilot system. This work may eventually help consultants improve patient care.

References

- Academy of Medical Sciences (2015). *Stratified, personalised or P4 medicine: a new direction for placing the patient at the centre of healthcare and health education (Technical report)*. Academy of Medical Sciences. May 2015. Retrieved 24/8/2016.
- Beez, U., Humm, B.G. and Walsh, P. (2015). "Semantic AutoSuggest for Electronic Health Records", *2015 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 760-765.
- Humm, B.G. and Walsh, P. (2015). "Flexible yet Efficient Management of Electronic Health Records". *2015 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 771-775.
- Hung, B. T., Long, N. P., Hung, L. P., Luan, N. T., Anh, N. H., Nghi, T. D., Hirayama, K. (2015). "Research Trends in Evidence-Based Medicine: A Joinpoint Regression Analysis of More than 50 Years of Publication Data", In: *PLoS ONE*, 10(4).
- Idelhauser, J., Beez, U., Humm, B.G., Walsh, P (2016). "A Clinical Decision Support System for Personalized Medicine". In: *U. Bleimann, B. Humm, R. Loew, I. Stengel, P. Walsh (Eds): Proceedings of the 2016 European Collaborative Research Conference (CERC 2016)*, pp 132 - 145. Cork, Ireland, 2016. ISSN 2220 - 4164.
- Marchant, G. E. and Lindor, R. A. (2013). "Personalized medicine and genetic malpractice". In: *Genetics in Medicine (Genet Med)*, 15(12), 921-2.
- Obst, O., Hofmann, C., Knüttel, H. and Zöller, P. (2013). "'Ask a question, get an answer, continue your work!' – Survey on the use of UpToDate at the universities of Freiburg, Leipzig, Münster and Regensburg", In: *GMS Medizin—Bibliothek—Information*, 13(3), 26.
- Kawamoto, K, Houlihan, C.A., Balas, E.A. and Lobach, D.F. (2005). "Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success", In: *BMJ: British Medical Journal*, 330(7494), 765.

Kwag, K. H., Gonzalez-Lorenzo, M., Banzi, R., Bonovas, S., Moja, L. (2016). "Providing Doctors With High-Quality Information: An Updated Evaluation of Web-Based Point-of-Care Information Summaries". In: *Journal of Medical Internet Res.* 2016 Jan; 18(1): e15. doi:10.2196/jmir.5234

National Comprehensive Cancer Network (2017). *Melanoma Skin Cancer Treatment* URL: <https://www.nccn.org/professionals/default.aspx> (visited /04/2017)

Microsoft Biztalk Business Rule Engine (2017). URL: <https://msdn.microsoft.com/en-us/library/aa561216.aspx> and <https://social.technet.microsoft.com/wiki/contents/articles/6480.biztalk-server-business-rule-engine-survival-guide.aspx> (visited 04/2017)

Ebsco Health (2017). *DynaMed Plus* URL: <http://www.dynamed.com/home/> (visited 05/2017)

IBM Watson Health (2017). *Watson for Oncology* URL: <https://www.ibm.com/watson/health/oncology-and-genomics/oncology/> (visited 05/2017)

UpToDate - Evidence-Based Clinical Decision Support at the Point of Care (2017). Wolters Kluwer URL: <http://www.uptodate.com/de/home> (visited 05/2017)

BMJ Best Practice - Clinical Decision Support Tool for Health Professionals (2017) URL: <http://bestpractice.bmj.com/info/> (visited 05/2017)

Antes, G. (1998). "Evidence-Based Medicine" In: *Der Internist ISSN: 0020-9554 Volume 39, Issue 9*, p. 900. doi: 10.1007/s001080050259

Borgerson, K. (2005). "Evidence-based alternative medicine?" Perspectives In Biology And Medicine [Perspect Biol Med] 2005 Autumn; Vol. 48 (4), pp. 504-513.

Fowler, S. A., Yaeger, L. H., Yu, F., Doerhoff, D., Schoening, P., Kelly B., (2014). "Electronic health record: integrating evidence-based information at the point of clinical decision making" In: *Journal of medical Library Association* 2014 Jan; 102(1): pp. 52–55. doi: 10.3163/1536-5050.102.1.010

Masic, I., Miokovic, M., Muhamedagic, B. (2008). "Evidence Based Medicine – New Approaches and Challenges" In. *Acta Informatica Medica* 2008 Dec.; 16(4): doi: 10.5455/aim.2008.16. pp. 219-225.

Preforming real-time emotion classification using an Intel RealSense camera, multiple facial expression databases and a Support Vector Machine

Michael Healy, Alphonsus Keary, Paul Walsh
Department of Computer Science
Cork Institute of Technology, Ireland
e-mail: michael.healy2@mycit.ie

Keywords : Affective Computing, Emotion Vision, Intel RealSense

This paper describes a new prototype for using the Intel RealSense™ (Intel, 2017) commercial camera system, a support vector machine (SVM) and multiple facial expression databases to predict emotional states with high accuracy. The system, called the Mobile Agitation Tracker, aims to be used for the detection of agitation and discomfort for patients suffering from cognitive decline disorders such as dementia. By mapping the Intel RealSense's 78 landmark points to an existing 68-point format we were successful in using existing facial expression databases as training data into machine learning algorithms. An experiment conducted by our research team also proved the effectiveness of the application.

1. Introduction

Dementia is a chronic or persistent disorder of the mental processes caused by brain disease or injury and marked by memory disorders, personality changes, and impaired reasoning. In 2011, over 47,000 people were diagnosed with dementia in Ireland alone (Pierce, et al., 2014). There is currently no known cure for dementia which means that full time care is the only approach to helping dementia patients.

A study conducted on 408 nursing home residents found that there is a direct link between agitated behaviours and cognitive decline. (Cohen-Mansfield, et al., 1990). The study used the Cohen-Mansfield Agitation Inventory (CMAI) (Cohen-Mansfield, 1991) which is used to score 29 different agitation states on a 7-point scale. These 29 states fall under three categories: aggressive behaviour, physically nonaggressive behaviour, verbally agitated behaviour.

Work on the Mobile Agitation Tracker (MAT) application has previously attempted to detect aggressive behaviour using pose data and other statistical analysis (Healy, et al., 2016). This paper explores whether the same technology can be used to enhance the ability of MAT to classify emotional states which a dementia patient may express such as disgust or anger. These are part of Paul Ekman's basic emotions (Ekman, 1992).

2. Related Work

Research conducted by Jayashree V. Patil & Prof. Preeti Bailke (Jayashree V, 2017), uses an Intel RealSense SR300 camera to produce a database of facial expressions and used this data to feed into a neural network to predict emotions. The authors compared the use of LIBSVM versus neural networks for classifying and found that the neural network was better at predicting happy and neutral facial expressions whereas the LIBSVM library was better at predicting surprised facial expressions. To perform this analysis, the authors captured 90 images of 30 different subjects performing happy, neutral, and surprised emotions. Using the data set the authors created, they trained a multi-layer

perceptron neural network that is capable of classifying three of Ekman's basic emotions. The authors also swapped and changed the features used in the algorithms to find the optimal feature set.

Another paper by Silva, V., Soares, F., Esteves, J.S., Figueiredo, J., Leão, C.P., Santos, C. and Pereira, A.P (Silva, et al., 2016) uses an Intel RealSense camera and a SVM to classify emotions using the Facial Action Coding System and Action Units as the input parameters. The Facial Action Coding System (FACS), associates the action of the muscles to the changes in facial appearance which are measured using Action Units (AUs). AUs are actions performed by a muscle or a group of muscles. The authors developed their own database of emotions to use as training data.

Research that uses the Intel RealSense camera to classify emotions has commonly developed a new database for their training data. The MAT application aims to use a different approach by enabling the use of existing high quality facial expression databases by mapping their 68-point landmark notations to the Intel RealSense's 78-point landmark notations. Input parameters to machine learning algorithms in the above research also varied from point deviation, AUs performed and landmark distances in a neural network. The MAT application will explore how the Euclidean distances between landmark points can be used as input into a SVM to perform classifications. It will also explore different SVM settings to try and find optimal classification accuracy.

3. Implementation

The MAT application uses an Intel RealSense camera which is capable of visual and emotional processing. The system is made up of three integrated optical devices, a full 1080P (1920 X 1080) camera, a VGA depth resolution camera, and an IR projector. IR projectors use infrared light points that are projected onto the users face or body. The VGA camera can then detect these points and use the differentiation of these dots to determine depth and distance between them.

The RealSense camera has a range of 0.2 – 1.2 meters and it also has stereo microphones for capturing audio (Intel, n.d.). A single camera can track up to five people at one time. This could be suitable in a care home scenario when multiple patients are in shared rooms or locations. The camera comes with an SDK for developing applications on the platform. The foundation of the SDK stack is made up of the SDK core, I/O module, and the algorithm modules. The SDK core manages the application pipeline execution and the I/O modules of the camera. The algorithm modules act as middleware for the main functions of the camera such as hand tracking, face tracking, and gesture recognition. They also allow development of applications through development frameworks (C++, C#, Java, Unity). A diagram of the SDK can be seen in Figure 1.

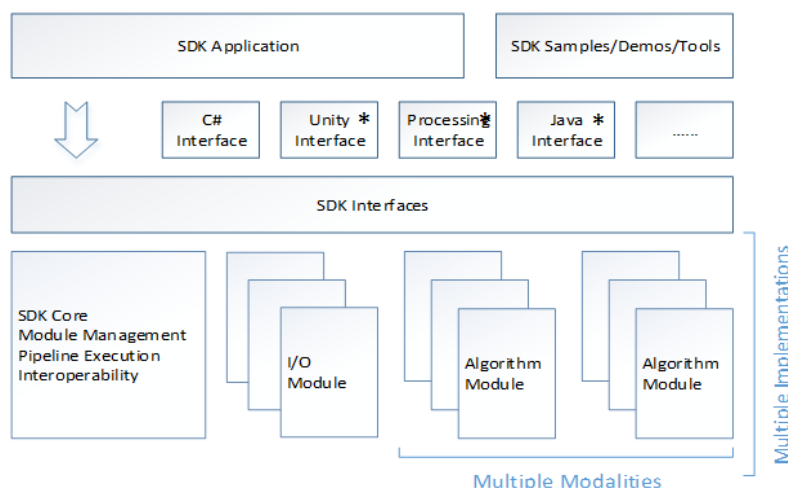


Figure 1: Intel® RealSense™ SDK Architecture

3.1. Machine Learning

Machine learning is a branch of artificial intelligence that gives computers the ability to learn without being explicitly programmed. A succinct summarisation would be “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience E ” (Mitchell, 1997). Machine learning algorithms are used in a wide range of applications and problems. Some modern uses of machine learning include spam filtering, speech recognition, facial recognition, document classification, and natural language processing. The three most commonly used machine learning applications are regression, clustering, and classification.

- Regression problems are a form of supervised learning. Supervised learning involves feeding “training” data into machine learning algorithms. Regression problems aim to predict continuous values for example, how much a house would cost given the size and location of the house rather than using discrete values which would list parameters as true or false.
- Clustering is a type of unsupervised learning. Clustering works by attempting to group data by mathematical similarities such as if their parameters were to be graphed, it would analyse how closely related the parameters are to each other.
- Classification problems are a form of supervised learning. Classification problems aim to assign predetermined labels to unseen data based on previous training examples. Classification was the machine learning method used in this research.

The MAT application uses training data from the Cohn-Kanade dataset (CK+) (Lucey, et al., 2010) and Multimedia Understanding Group (MUG) (Aifanti & Delopoulos., 2010) databases to create machine learning models used in the MAT system. The Cohn-Kanade database contains 593 images of six recognised facial expressions, that are FACS and AU coded, from 123 different subjects. Classification algorithms are useful with this type of dataset. Using a rule based system would require the detection of a series of AUs. If one of these AUs is missed the rule might not be met and the classification will be incorrect. In order to deal with all possibilities, within a rule based system, there will be a combinatorial explosion of rules to deal with such variations. This isn’t a problem when you use machine learning as the data is generalised to best fit a certain model and thus it does not have to fit the model perfectly.

3.2. Support Vector Machines

To assist with the complex mathematics of creating a classification model we used a Support Vector Machine (SVM). Support Vector Machines are used in supervised learning to assist with the generation of a model by using built-in algorithms to find the optimal hyperplane. This hyperplane is the largest minimum distance to the training examples which aims to clearly separate the different classes. New unseen examples are added to the same space and their class is predicted based on which side of the gap they fall. The SVM used is called LIBSVM (Lin & Chih-Jen, 2011). LIBSVM is an integrated software library for support vector classification, regression, and distribution estimation. It also has support for multi-class classification which enables the algorithm to compare the given data to multiple classes which would be useful when attempting to classify multiple emotional states. LIBSVM was originally written in C but now has support for a wide range of programming languages such as Java and Python. To generate the models using the SVM we first had to decide on the parameters which is discussed in the next section.

3.3. Landmark Points & Mapping

Facial landmarks are defined as the detection and localisation of certain key points on a human face. They are also known as vertices or anchor points. Current RealSense cameras can accurately distinguish 78 landmark points on a human face. The points are grouped around the following areas: the eyebrows, eyes, nose, mouth, and across the jaw line. Figure 2 below shows a collection of typical landmark points on a human face. Intel's RealSense camera allows the location of these points in a 2D or 3D space to be retrieved (x , y and z coordinates).



Figure 2: Intel® Real Sense™ SDK Landmark Data

This data can be problematic when most of the existing facial expression databases use 68-point annotations which can be seen in figure 3 below (Gross, et al., 2010).

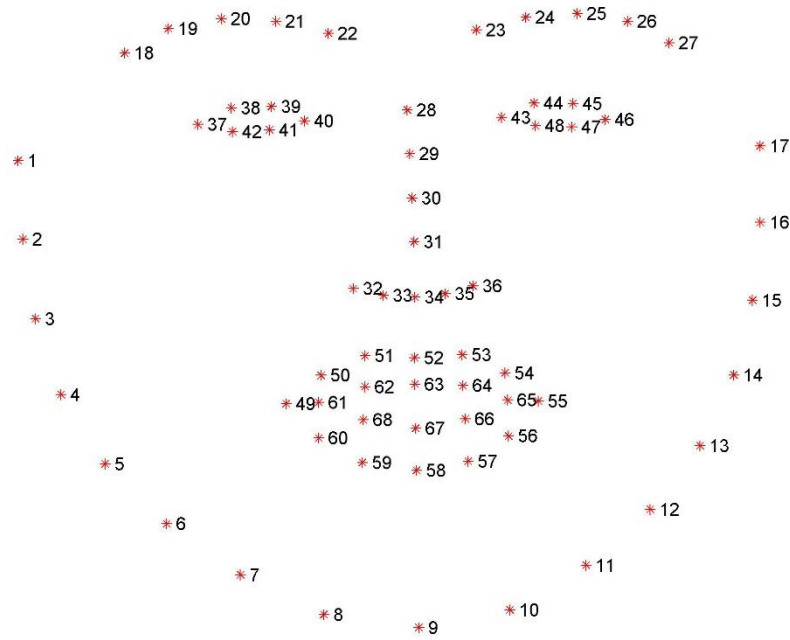


Figure 3: Multi-PIE 68-point mark-up

The differences to note is that the 68-point annotation style has 5 points at the end of the nose whereas the RealSense annotation only has 3. The RealSense camera detects landmarks under the eyebrow unlike the 68-point annotation which only tracks above the eyebrows. The RealSense camera can also track the eyes as landmarks which is not included in the 68-point annotations.

To create point 33 and point 35 seen in Figure 3, the midpoint formula was used to get the estimated landmark location between points 30,31 and points 31,32 in Figure 2. The formula used can be seen in Equation 1. For the rest of the points, they were mapped on a one to one basis to the nearest landmark. All extra landmarks the RealSense camera can track are ignored.

$$(x, y) = \left(\frac{x_1 - x_2}{2}, \frac{y_1 - y_2}{2} \right)$$

Equation 1: Midpoint Formula

3.4. Input Parameters

The distances from each landmark point to every other landmark point was chosen as input parameters to the SVM. This is because using the location of each landmark point would cause the data to be varied depending on the subject's alignment to the RealSense camera. This would make classification unfeasible in a live stream environment. This approach taken is ideally suited for the off-line processing of a video stream for affective computing analytic purposes.

Getting the landmark points was achieved using the RealSense SDK to query the positions of each landmark. Using the (x, y) coordinates provided by the SDK, the distance between the points can be calculated using the Euclidean distance formula seen in equation 2 below.

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Equation 2: Euclidean Distance Formula

This resulted in the computation of over 4,500 parameters which can lead to overfitting. Overfitting occurs when excess parameters can negatively affect the accuracy of predictions. This is because the parameters might not change throughout the different classes but their presence could mathematically skew the data to be wrongly classified. To overcome this problem, we removed the points along the jaw line as they were thought to have little significance in facial expressions. Finally, we needed to normalize the data. The chosen solution was to scale the data to real numbers between [-1, +1] using the formula below.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Equation 3: Feature Normalisation

4. Optimization & Experimentation

In machine learning, a “kernel” is a mathematical function for measuring similarities. The kernel function $K(x, y)$ describes the similarities between points ‘x’ and point ‘y’ (Bishop, 2006). There are many different kernels available to researchers and the most frequently used in machine learning are listed below:

- Polynomial kernel
- Radial basis function kernel (RBF)
- Sigmoid kernel
- Linear kernel

The MAT application uses the RBF kernel as there is a large amount of input parameters as features into the classifier. The RBF kernel can accommodate this as it can contain infinite dimensions in its calculations. The formula for the RBF kernel can be seen in Equation 4 below. In RBF SVM classification there are two values which can be optimised to suit the training data. These values are the ‘C’ value and the ‘γ’(gamma) value. The ‘C’ value describes the trade-off between training error and the flatness of the solution i.e. to decide how many outliers are considered when the SVM is doing its calculations. An optimal ‘C’ value will keep training errors small but will also generalise well. The ‘γ’(gamma) value controls the radius or reach of the C value. The following graphs were created with gnuplot (Thomas Williams, n.d.) and show a variety of different C & gamma values to find the optimal accuracy of classification.

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right)$$

Equation 4: Radial Basis Function Kernel

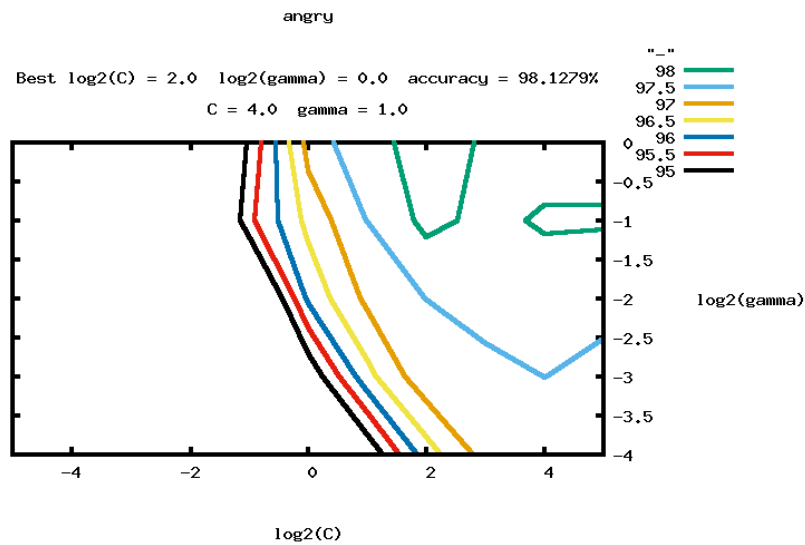


Figure 4: Optimal 'C' and Gamma value for angry model

Figure 4 shows the best 'C' and gamma values for the “angry” machine learning model to be trained with is 4 and 1.0, which gives an accuracy of 98% in cross-fold validation

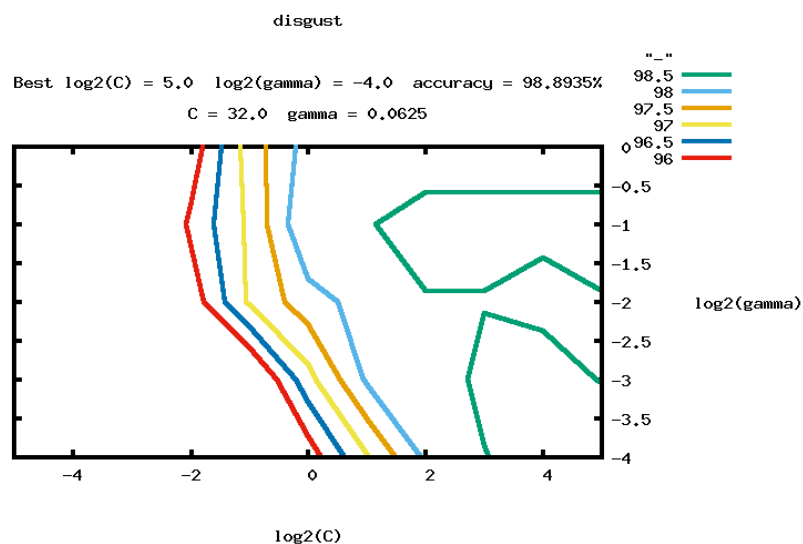


Figure 5: Optimal 'C' and Gamma value for disgust model

Figure 5 shows the “disgust” model should be trained with a 'C' value of 32 and gamma value of 0.0625, this shows that to classify a “disgust” expression more parameters need to be used in the calculations than the “anger” model seen in figure 4. This gave an accuracy of 98% in cross-fold validation

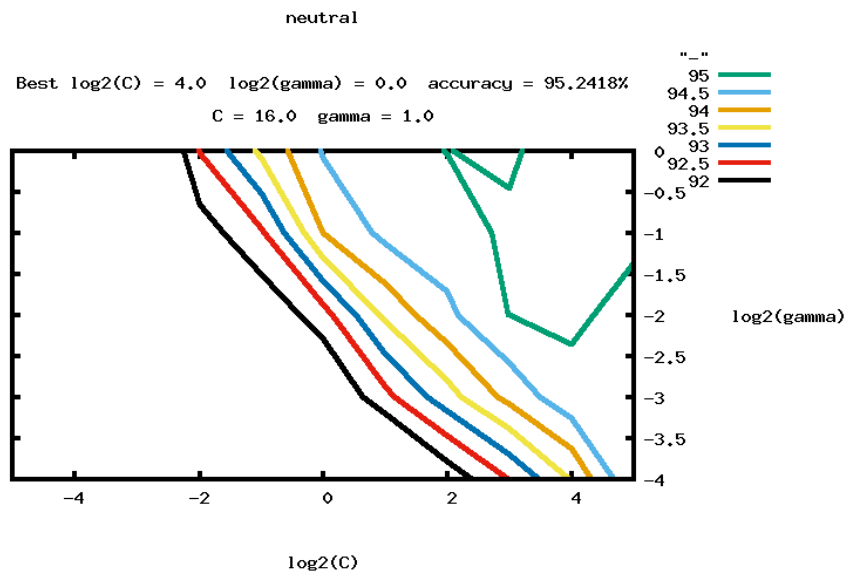


Figure 6: Optimal 'C' and Gamma value for neutral model

Figure 6 shows the optimal 'C' value to be 16 and gamma value to be 1.0 when training the “neutral” model. This gave a 95% accuracy in cross-fold validation. The accuracy from cross fold validation in the training examples was very high, however, the MAT application would need to classify subjects in real-time and using its own camera. Therefore, it was important to test its feasibility in the real world.

5. Results

An experiment conducted by our research team-involved a series of interactions with a selection of imagery expected to generate an emotional response in a participant. The MAT application was used to capture landmark data from each participant as they underwent this experiment. This data was then used for off-line analysis and is presented in this section. Naturally this data would be compared to a video stream of the experiments to prove expected results versus actual results. However, due to ethical reasons this was not possible. Instead, throughout the test the participant was asked to complete a self-evaluation that explained the types of emotions that the participant felt during the various parts of the experiment. For each series of interactions with the imagery, the participant was allowed select three emotions from a total of twenty and to score them on a scale of 1–5, with 5 being the strongest emotion felt.

To see how aggression detection was improved in MAT we decided to look at participants that gave a high score (4/5) for anger and disgust emotions during the experiment and see if MAT detected these expressions of emotion. Figure 7 below shows a timeline of emotions detected by MAT when the participant was interacting with the “disgust” imagery. The participant scored anger with a rating of 5 for this experiment. As seen in the graph, at 12:14:56 the MAT application classified the participant as expressing anger for a period of 18 seconds before briefly returning to a neutral expression followed by a second spout of anger. This is good evidence that the MAT application

successfully detected the aggression which was self-reported by the user. The following graphs display results from other participants.

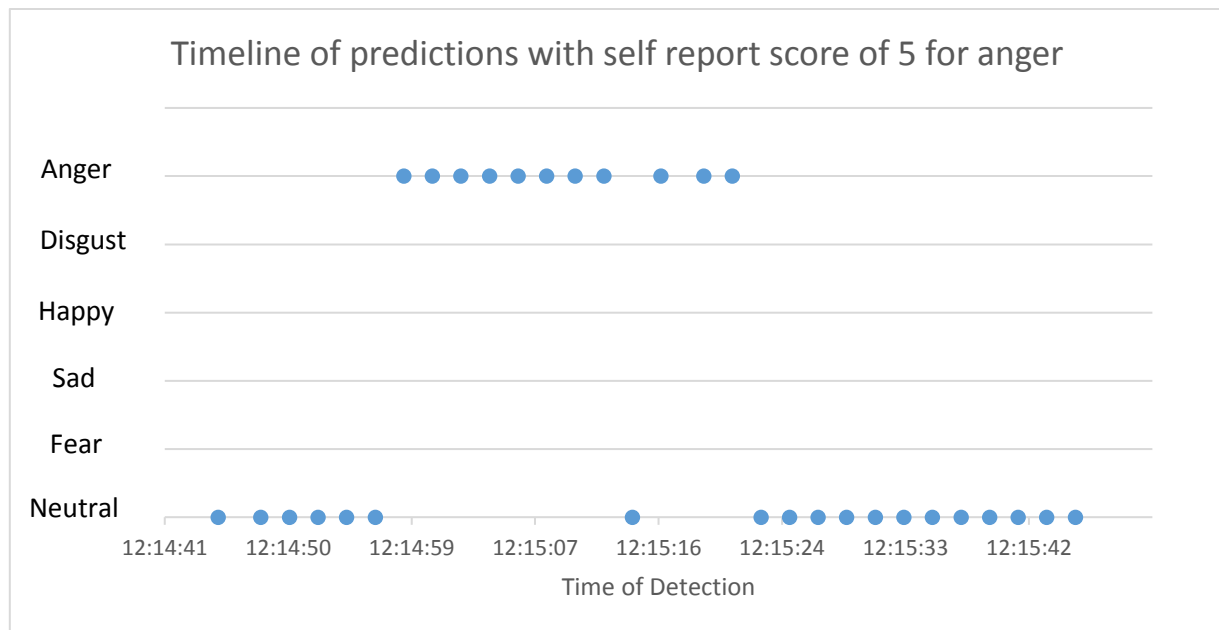


Figure 7: Participant 10, Experiment 3 - "Disgust" Imagery

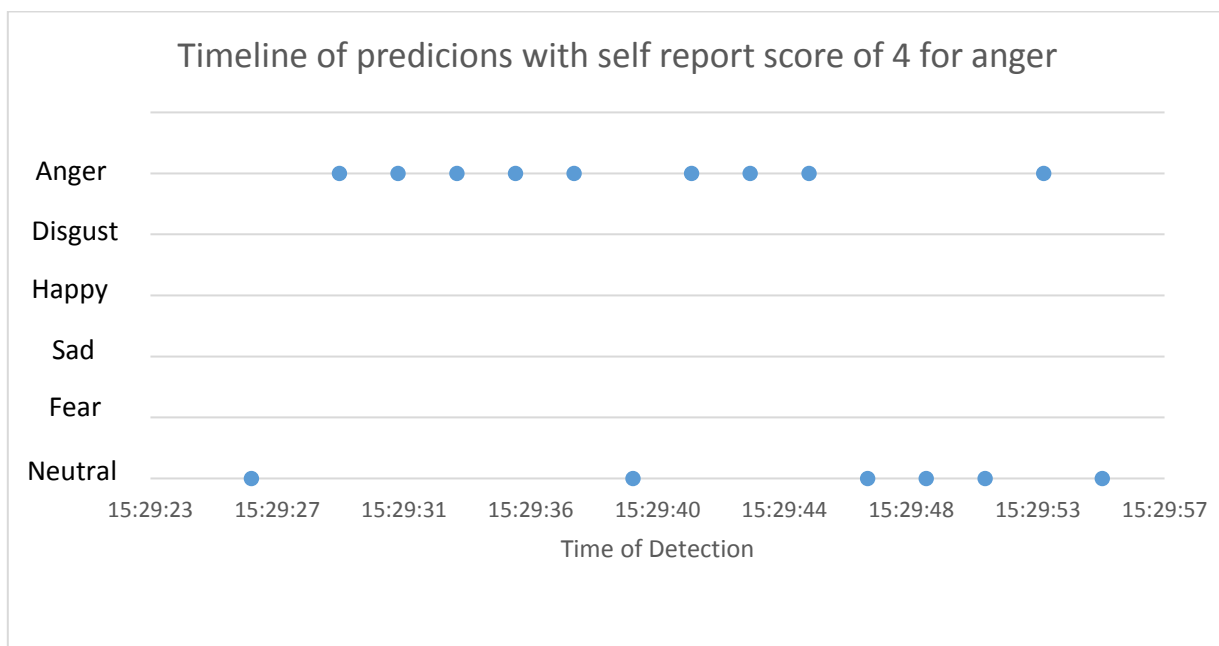


Figure 8: Participant 9, Experiment 3 - "Sadness" Imagery

Figure 8 shows an initial burst of anger with another two occurrences before the end of the experiment. This high-level detection of anger was seen to correlate with the participant's self-report score of 4.

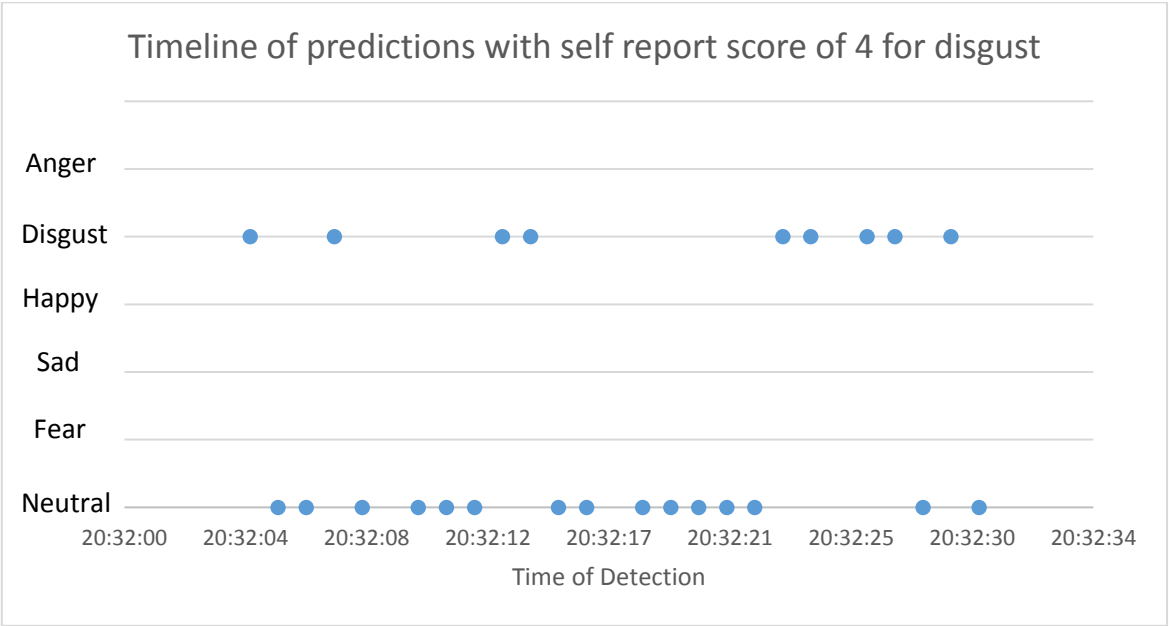


Figure 9: Participant 2, Experiment 3 - "Disgust" Imagery

Figure 9 shows several occurrences of a disgusted expression throughout the experiment with the best example beginning at 20:32:22.

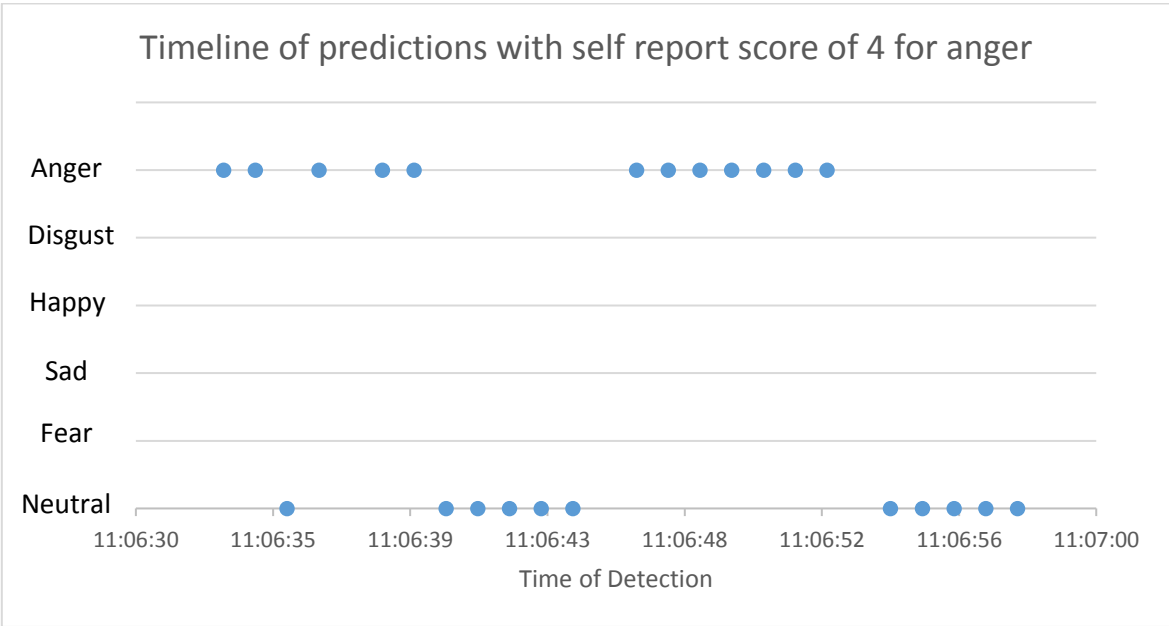


Figure 10: Participant 6, Experiment 3 - "Fear" Imagery

Figure 10 shows that the participant had an angry expression for the majority of the interaction with the images which would agree with the self-report score of 4 for anger.

The final chart below in Figure 11 is particularly interesting as the participant scored both disgust and hate at 3. It was interesting to see the MAT application also classified the hate emotion as anger.

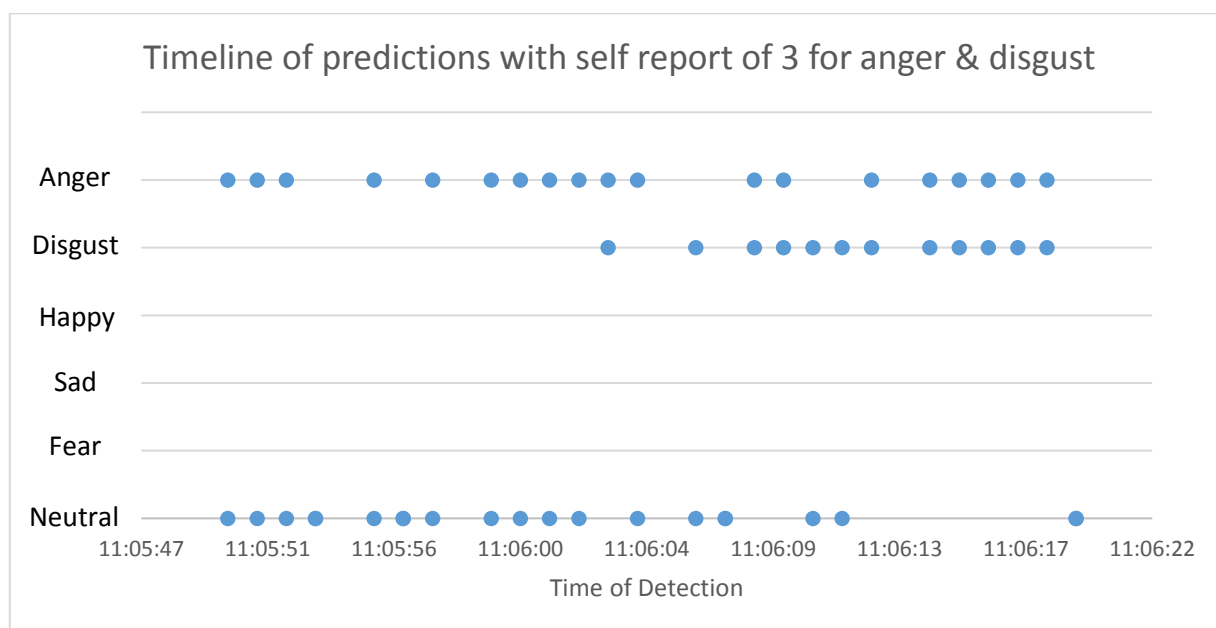


Figure 11: Participant 6, Experiment 3 - "Disgust" Imagery

6. Conclusion & Future Work

Using the participants self-reporting, we have proved that the Intel RealSense camera can be used to detect aggression using machine learning, a SVM, and training data from multiple existing sources. In the future, we would like to expand MATs emotional expression functionality to detect more emotional states which may be expressed by dementia patients (e.g. happiness). Our real-world related affective computing experiments are ongoing and the data collected will act as a valuable corpus for building on the preliminary results examined in this paper. It will also be possible to include additional affective computing related databases to the MAT training set to further improve its recognition accuracy.

References

- Aifanti, N. & Delopoulos., C. P. a. A. (2010). *The MUG Facial Expression Database*. Desenzano, Italy, Multimedia Interactive Services.
- Bishop, C., (2006). *Pattern recognition and machine learning*. s.l.:s.n.
- Cohen-Mansfield, (1991). "Cohen-Mansfield Agitation Inventory". *American Psychological Association*.
- Cohen-Mansfield, Marx & Rosenthal (1990). "Dementia and Agitation in Nursing Home Residents". *American Psychological Association*.
- Ekman, P. (1992). "An argument for basic emotions". *Cognition & Emotion*, pp. 3-4, 169-200.
- Gross, R., Matthews, I., Cohn, J. & Baker, T. K. a. S. (2010). "Multi-pie". *Image and Vision Computing*, 28(5), pp. 807-813.
- Healy, M., Keary, A. & Walsh, P. (2016). "Prototype proof of concept for a mobile agitation tracking system for use in elderly and dementia care use cases". *Proceedings of the CERC, Cork*.

Intel (2017). *Develop Immersive Experiences*. [Online] Available at: <https://software.intel.com/en-us/realsense/home> [Accessed 11 07 2017].

Intel (n.d.) *Specifications for the Intel® RealSense™ Camera F200*. [Online] Available at: <https://communities.intel.com/docs/DOC-24012> [Accessed 11 07 2017].

Jayashree V, P. a. P. B. (2017). "Real time facial expression recognition using RealSense camera and ANN". s.l., IEEE.

Lin, C.-C. C. & Chih-Jen (2011). "LIBSVM: a library for support vector machines". In: *ACM Transactions on Intelligent Systems and Technology*.

Lucey, P. et al. (2010). "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression", In : IEEE Computer Society Conference.

Mitchell, T. M. H. (1997). *Machine Learning*. s.l.:s.n.

Pierce, M., Cahill, S. & O'Shea, E. (2014). *Prevalence and Projects of Dementia in Ireland 2011 - 2046*, s.l.: NUIG.

Silva, V. et al. (2016). *Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*. s.l., IEEE.

Thomas Williams, C. K. (n.d.) *gnuplot*. [Online] Available at: <http://www.gnuplot.info> [Accessed 1 September 2017].

Predicting Premature Termination of Treatment in Psychotherapy for Borderline Personality Disorder ^{[1],[2]}

Martin Bohus[†], Tilman Deuschel^{†,††}, Stephan Gimbel[†], Nora Görg[†], Bernhard G. Humm[†], Martin Schüller^{†,††}, Erdem Turan[†]

[†] Hochschule Darmstadt –
University of Applied Sciences
Haardtring 100
64295 Darmstadt, Germany
bernhard.humm@h-da.de

^{††} Zentralinstitut für Seelische Gesundheit
J5
68159 Mannheim, Germany
martin.bohus@zi-mannheim.de

^{†††} Deuschel & Schüller GbR,
Sudetenstrasse 6
64853 Oetzberg, Germany
martin@jekyll-hyde.net

Keywords: Applied computing, health informatics, human-centered computing, empirical studies in interaction design, machine learning

Abstract: Premature termination is a central challenge in psychotherapeutic treatments. General moderators and mediators of premature termination are unclear. A mechanism that alerts therapists when a patient is likely to terminate the treatment would allow therapists to intervene and potentially ensure treatment success. The project on hand is applying a machine learning approach to analyse short daily verbal reports of individuals in ongoing psychotherapy in order to identify general and individual predictors of premature termination of psychotherapy. This paper presents the concept for a smartphone application to collect data from psychotherapy patients. Particular focus lies on usability and user experience targeted to the above-mentioned user group. Data collected from the application shall be used to predict the likelihood of premature termination of treatment using Machine Learning techniques. An initial evaluation of a prototypical implementation indicates high potential of this work in progress.

1. Introduction

Premature termination of treatment is a central challenge in psychotherapy which affects up to 25 % of all outpatients. Early treatment terminators receive only parts of evidence-based treatments and thus cannot benefit as much as treatment completers (Altmann et al., 2014; Saxon et al., 2016). Financial consequences include the waste of working hours of therapists as well as limited access to mental health services for others in need (Barrett et al., 2008) which adds up to the problem of waiting times of up to six months for psychotherapy in Germany (Bundespsychotherapeutenkammer, 2011).

Premature termination of treatment is also pronounced in patients with Borderline Personality Disorder (BPD) (Cinkaya et al., 2011), with dropout rates between 10 and 30 % (Bohus et al., 2004; Rüscher et al., 2008; Kröger et al., 2014). BPD is not only characterized by high health care usage and costs (Bode et al., 2016; Comtois and Carmel, 2016; Laurensen et al., 2016), but also by severe emotion regulation difficulties, identity disturbances, interactional difficulties, self-harm, suicide attempts, impulsivity, and outbursts of anger. Core symptoms of BPD, such as anger and self-harm, are effectively reduced by specialized treatments such as Dialectical Behavior Therapy (DBT) (Linehan, 2015) according to a meta-analysis (Stoffers et al., 2012). BPD patients require an average of 45 days less inpatient treatment in the year after

^[1] This project (HA project no. 517/16-29) is funded in the framework of Hessen ModellProjekte, financed with funds of LOEWE – Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz, Förderlinie 3: KMU-Verbundvorhaben (State Offensive for the Development of Scientific and Economic Excellence).

^[2] Trial registration: World Health Organization International Clinical Trials Registry Platform DRKS00011555.
<http://apps.who.int/clinicaltrialssearch/default.aspx>

completing DBT when compared to the year before DBT (Wagner et al., 2014). However, this data relates to treatment completers.

To date, reasons for premature termination of treatment are often analyzed post-hoc. This paper will present a novel concept and a prototype implementation for predicting premature termination *during* a patient's treatment. The goal is to alert therapists when a patient is likely to terminate. This will allow therapists to intervene and potentially ensure treatment success. For this, a computer application is designed that enables patients to regularly provide data relevant for the prediction of premature termination. To increase feasibility, the user experience of the application must be particularly targeted to patients. Machine Learning (ML) algorithms (Alpaydin, 2010) will be used to predict premature termination.

2. Related Work

Different clinical variables correlate with premature termination of BPD therapy, including a higher number of co-occurring psychiatric disorders, antisocial personality disorder, a poor therapeutic working alliance, avoidance of distressing feelings and experiences, increased anger levels, prior long-term psychiatric treatments and a lower education level (Kröger et al., 2014; Landes et al., 2016; Perroud et al., 2010; Rüsche et al., 2008; Spinhoven et al., 2007; Wnuk et al., 2013). Inconsistent findings exist for general psychopathology (Barnicot et al., 2011; Landes et al., 2016), and the number of previous suicide attempts (Kröger et al., 2014; Rüsche et al., 2008; Wnuk et al., 2013; Panfilis et al., 2012). Predictors of premature termination are usually identified by single baseline measures which excludes dynamic changes (e.g., personal and logistic problems, distress resulting from the therapy, arguments with other patients, critical life events) (Kröger et al., 2014; ter Huurne et al., 2017; Hernandez-Tejada et al., 2014). Furthermore, these predictors are based on mean-based group calculations and cannot be used to identify the individual risk of a patient. In contrast, repeated real-time assessments of potential predictors for premature termination would allow testing of dynamic changes in patients' everyday life (Ebner-Priemer and Trull, 2009). Mental health applications can be used for the assessment as there is a high demand for such tools (Bakker et al., 2016).

Most models that predict change in psychotherapy are based on a predefined set of predictors. For example, (Lutz et al., 1999) predicted an expected change in symptomatology by patient characteristics (e.g., chronicity of the problem, severity of intake symptomatology) in a multilevel model. This strategy allows comparing the actual change in a given patient with the expected change of this patient according to the model. Thus, clinicians can identify whether a patient is not on track and needs additional support. Another study computed an individual risk index for non-improvement of symptoms, also referred to as nonresponse (Delgadillo et al., 2016). The risk index included age, unemployment, disability, and baseline symptomatology. (Rubel et al., 2016) predicted session-to-session change in symptomatology by patient's self-reported coping experience and emotional involvement in the prior therapy session. However, when using a predefined set of predictors, other reasons for premature termination are excluded. A new way in the prediction of treatment termination could be to use ML algorithms. These algorithms can find relevant predictors on the basis of training data. For example, (Lutz et al., 2005) predicted symptom change of individual patients by the average change in statistical "nearest neighbors". These neighbors were similar to the individual patient with regards to baseline assessments (gender, age, baseline psychopathological symptomatology), and early symptom change. Other studies used ML to predict suicide risk or response to antidepressant medication (Chekroud et al., 2016; Passos et al., 2016).

To date research on the prediction of premature termination of BPD therapy mainly relied on predefined sets of baseline assessments. Statistical methods were limited to mean-based methods or simple ML analysis. To the best of our knowledge, no computer application for BPD patients predicted premature termination of treatment with repeated real-time data. Such

repeated assessments would allow the detection of dynamic processes which play a crucial role in premature termination. Furthermore, ML algorithms might detect relevant predictors for which no pre-assumptions have been made.

3. Requirements

The following requirements have been identified via extensive discussions with therapists and patients, applying the user-centered design method.

1. The application shall enable patients to regularly provide data relevant for the prediction of premature termination of treatment.
2. The usability and user experience of the application shall be targeted to BPD patients; in particular, regularly providing data shall not have a negative effect on the patient's wellbeing and on the therapeutic process. The patients may choose the device type according to their preference: smartphone, tablet PC or personal computer.
3. The application shall regularly compute a sufficiently accurate prediction of the likelihood of each registered patient of terminating the DBT treatment.
4. Therapists will be alerted instantly with a traffic light function when the prediction indicates the danger of premature termination of treatment.
5. Researchers shall be able to use anonymized data gathered by the application to analyze the effects of treatment procedures on premature termination of treatment.

4. User Interaction Concept

The user interaction concept is divided in three phases. *Pre-usage*: The patient starts DBT treatment and patients and therapists answer the initial questionnaires (see Table 1 for details). *Use*: In addition to standardized questionnaires, the patient provides a daily feedback about their therapy and everyday life. This part of the application is implemented as a personal diary. The diary will be provided verbally; the application then transfers the speech to text. This methodology of processing informal patient input has been chosen to not limit the models to a set of predefined predictors. In doing so we prevent the exclusion of other relevant factors. Descriptive analysis of the questionnaire data, e.g., average values of each patient, are presented visually to the therapist and to the patient. Moreover, an alert is presented to the therapist if patients are likely to terminate the treatment prematurely. *Post-usage* takes place after the therapy, when the therapist provides information about whether the patient has terminated the therapy prematurely or not. In case of premature termination reasons are provided by therapists and patients.

Table 1: Overview of data, collected in the different usage phases

	Pre-Use	Use	Post-Use
Patient	Date of birth, gender, highest educational degree, Acceptance and Action Questionnaire-II (AAQ-II) (Bond et al., 2011), previous treatment within the last ten years (including outpatient psychotherapy or psychiatric treatment, residential psychiatric, psychosomatic and psychotherapeutic treatment, inpatient psychiatric wards and finally previous drop-outs), level of impulsivity (Barratt Impulsiveness Scale, BIS-11; Patton et al., 1995), motivation to change (University of Rhode Island Change Assessment Scale, URICA, McConaughy et al., 1983), therapy expectancies (Therapy expectancies and therapy evaluation of patients, PATHEV, Schulte, 2005), quality of life (World Health Organization Quality of Life Assessment, WHOQOL-BREF, Skevington et al., 2004), severity of patients' psychopathological symptoms (Brief Symptom Inventory, BSI, Derogatis, 1992), Borderline Symptom List (BSL-23) (Bohus et al., 2004), if Posttraumatic Stress Disorder diagnosed additionally Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5; Weathers et al., 2013). In addition, patients are asked whether seeking treatment was mainly motivated by the patients themselves or by others.	Daily: Personal diary in which patients give feedback about the therapy and their everyday life Weekly: Borderline Symptom List (BSL-23). If patients are also diagnosed with Posttraumatic Stress Disorder, the Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5) is applied as well. Monthly: Therapeutic alliance (Working Alliance Inventory, short form, WAI-SR, Wilmers et al., 2008), as well as cohesiveness of groups (Therapeutic Factors Inventory Cohesiveness Scale, Lese and MacNair-Semands, 2000).	In case of premature termination of treatment: Reasons for termination
Therapist	Patient's diagnosis (ICD-10) (Graubner, 2014)		If premature termination of therapy: Ask for assumed reasons for termination. Patient's diagnosis (ICD-10) (Graubner, 2014) post therapy.

5. Software Architecture

We propose to implement the application as a three-layer-architecture (Ramirez, 2000), consisting of a presentation layer, business logic layer, and a data store, see Figure 1. Data visualisation and acquisition will be provided through a web application with dedicated views for therapist, patient, and researcher, according to their needs. This approach is flexible enough to allow the use of desktop as well as mobile clients such as iOS and Android. The application components provide functionality for managing all incoming and outgoing data (Data Management), compute benchmarks on patients' questionnaire data (Benchmarking), and predict premature termination of patients' treatments (Prediction). In order to do so, the system has several utility components: Access Management ensures data protection policies; Configuration allows configuring the application, e.g., specifying the interval in which patients have to complete specific questionnaires; Statistics / Machine Learning is a library providing suitable algorithms for benchmarking and prediction. The data store allows storing static data, such as questionnaire forms, uses, etc., as well as dynamic data, in particular questionnaire data provided by patients.

The following technology stack may be used for implementing the proposed software architecture: HTML5, JavaScript/TypeScript, CSS, and AngularJS for the presentation layer; a RESTful API as the interface between presentation layer and business logic layer; Python as the programming language for the business logic layer; MongoDB as database management system.

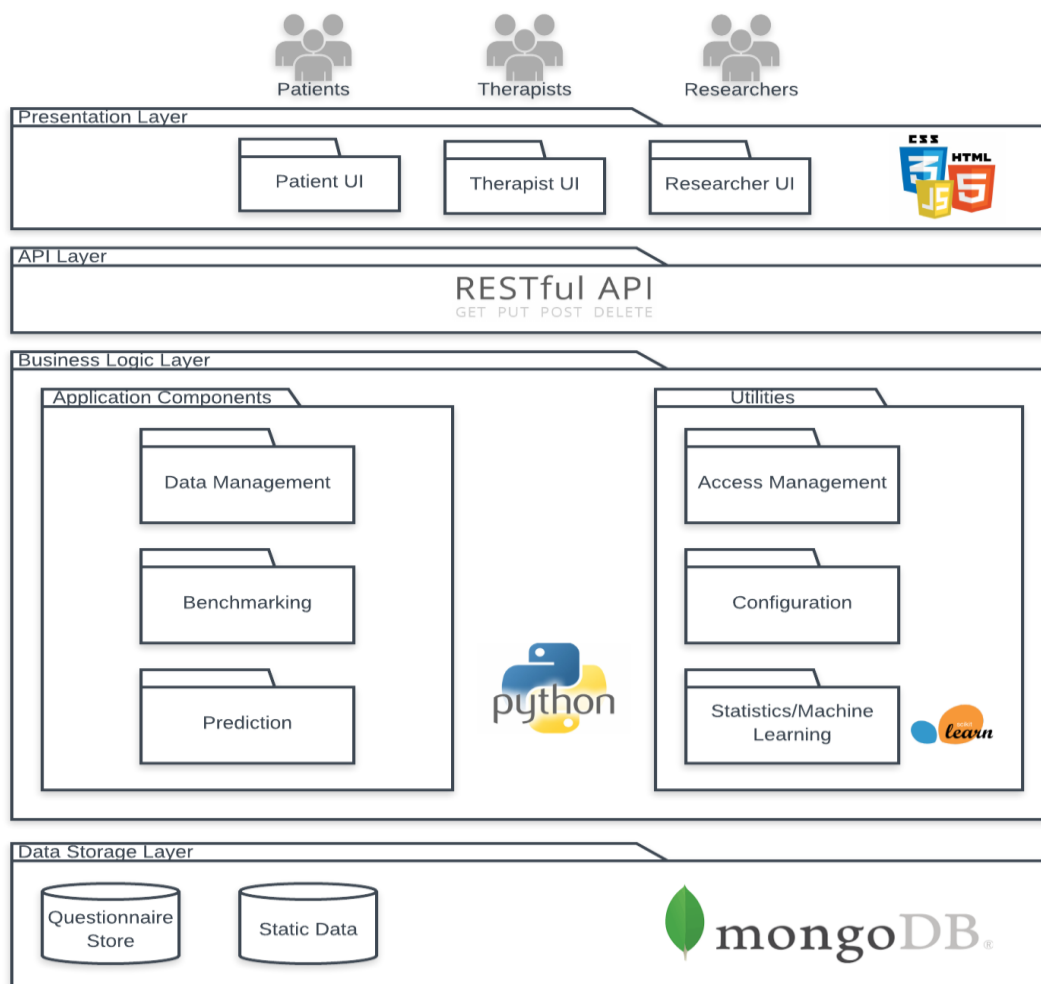


Figure 1: Proposed software architecture

6. User Evaluation of the Questionnaire Client

The data gathering and data storage components of the termination prediction application were prototypically implemented, using the technology stack presented in the previous section. Figure 2 shows the user interface of the questionnaire form on a mobile device. The application has been evaluated based on the requirements stated in Section 2. The termination prediction application allows patients to provide relevant data regularly for predicting premature termination of treatment (Requirement 1). To assess the user experience (Requirement 2) we performed an initial evaluation. The prototype has been qualitatively tested with four patients with either PTSD or BPD. The method of co-discovery has been applied (Jordan, 2010). This method engages participants in thinking aloud naturally without explicit instructions by the researcher which might have resulted in an artificial situation. Furthermore, the co-discovery is suited to investigate initial responses to prototypes. The patients were asked to test the application in pairs of two while thinking aloud. Two patients stated that they appreciate the

“minimalistic design” and the “clarity” of the application. The design would support focussing on the questions even under extreme distress, which is typical for patients with BPD. Further, the group stated a preference for the application over paper-pencil-questionnaires, because therapy already included a lot of paperwork (therapy homework, evaluation forms). Patients stated that the application made the data assessment more convenient and efficient. Patients expressed interest in seeing a weekly overview of their diary. This way, they could see how their overall attitude towards therapy and life might have changed. This would also serve as a response from the application to patients’ daily efforts. An initial evaluation of Requirement 3 (Prediction and Statistics / Machine Learning) can be found in the next section. Requirements 4-5 are covered in the system architecture: Therapist UI (Requirement 4), and Researcher UI / Questionnaire Store (Requirement 5). Their implementation and evaluation are subject to future work.

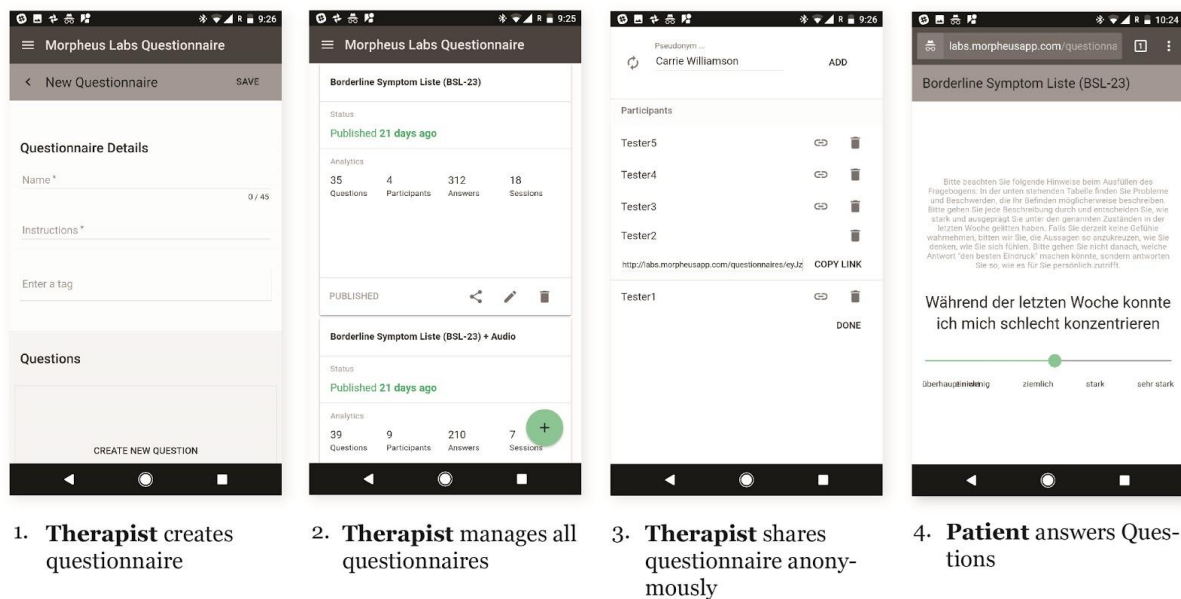


Figure 2: User interface of questionnaire handling

7. Initial Evaluation of the Prediction Component

A first prediction model has been composed and evaluated. In sum, 137 patients treated in the DBT unit at the Central Institute of Mental Health between 2014 and 2016 were available for a primary prediction of premature termination in DBT. In addition, 11 items on behavioural symptoms of BPD (e.g., self-harm, alcohol abuse) were assessed. These weekly paper and pencil measures of borderline symptomatology were collected as a standard procedure in the DBT clinic. One of the authors (N.G.) extracted information on whether the patients terminated the treatment prematurely from the clinical records of the patients. Based on this data, a prediction of premature termination was performed in Rapid Miner (<https://rapidminer.com/>). Several algorithms, among them, k-Nearest Neighbour, Logistic Regression, Deep Learning H2O, and Naive Bayes) were evaluated and optimized with respect to the F1-score for the class premature termination, see Table 2. Cross validation with folds set to 10 was used, splitting the data into a training and validation model to generate the prediction model. For all ML algorithms compared, Logistic Regression yields the best results (accuracy 72%, F-score for class termination: 0.53). However, the accuracy is not yet considered high enough. Future work is needed for improving the prediction accuracy.

Table 2: Performance of ML algorithms for classes termination / non-termination

	k-Nearest Neighbour			Logistic Regression			Deep Learning H2O			Naive Bayes (Kernel)		
Accuracy	54.01% +/- 11.67% (micro: 54.01%)			71.59% +/- 13.15% (micro: 71.53%)			69.40% +/- 15.35% (micro: 69.34%)			62.14% +/- 12.54% (micro: 62.04%)		
	True term.	True non-term.	Precision	True term.	True non-term.	Precision	True term.	True non-term.	Precision	True term.	True non-term.	Precision
Predict term.	27	47	36.49%	22	18	55.0%	23	22	51.11%	16	25	39.02%
Predict non-term.	16	47	74.60%	21	76	78.35%	20	72	78.26%	27	69	71.88%
Recall	62.79%	50.00%		51.16%	80.85%		53.49%	76.6%		37.21%	73.4%	
F1-Score	0.4616			0.5301			0.5227			0.3809		

8. Conclusions and Future Work

Rather than predicting premature termination by single baseline assessments, the application presented in this paper will predict potential treatment terminations before they occur. The prediction is based on repeated real-time assessments which are analyzed by ML algorithms. This approach will allow adjustments during therapy and could contribute to a decrease of premature treatment termination.

The current prototype allows the web-based creation of questionnaires and collection of data. Next, we will implement the Benchmarking component which provides descriptive feedback about the data to patients and therapists. Furthermore, the Prediction component will predict premature treatment termination using ML technology with sufficient accuracy. Data of patients with BPD who receive DBT in a residential DBT program will be collected. These datasets will allow the identification of further predictors of premature termination of treatment in addition to the existing data corpus. We expect this to increase the prediction accuracy.

In addition, we plan to include speech-to-text data from the personal diary in the prediction. Variables extracted from the personal diary that could be tested might include the number of pronouns (Campbell and Pennebaker, 2003), affect (Poria et al., 2017), or grouping of words into different topics (Xiao et al., 2015). We expect this to furthermore increase the prediction accuracy.

In addition, standardized user testings (see e.g. Krug, 2013; MacKenzie, 2013, or Knapp et al., 2016) will be employed to assess objective measures of efficiency and effectivity as well as subjective measures of satisfaction (ISO 1999), perceived usability (ISO 2006) and emotional values affecting the user experience (Hassenzahl, 2010). A triangulation of these measures is an important indicator to assess if the application is likely to be used regularly by patients and therapists. This method will be applied to test Requirement 2 with patients (high usability and user experience) and Requirement 4 with therapists (alert of possible premature termination). In order not to influence the ongoing data collection, Requirement 4 will be evaluated with selected therapists which do not treat patients in the data collection programme.

May this work support reducing premature termination of treatment in psychotherapy and, thus, help patients and therapists alike.

8. References

- Alpaydin, E. (2010). *Introduction to machine learning*, 2nd edition. Adaptive computation and machine learning series, MIT Press, Cambridge, Mass.
- Altmann, U., Wittmann, W., Steffanowski, A., Heymann, Friedrich von, Bruckmayer, E., Pfaffinger, I. et al. (2014). "Abbrüche antragspflichtiger ambulanter Psychotherapien. Häufigkeit, Risikofaktoren, Outcome", in: *Psychiat Prax*, vol. 41, no. 06, pp. 305–312.
- Bakker, D., Kazantzis, N., Rickwood, D., Rickard, N. (2016). "Mental Health Smartphone Apps. Review and Evidence-Based Recommendations for Future Developments", in: *JMIR Mental Health*, vol. 3, no. 1, p. 7.
- Barnicot, K., Katsakou, C., Marougka, S., Priebe, S. (2011). "Treatment completion in psychotherapy for borderline personality disorder – a systematic review and meta-analysis", in: *Acta Psychiatrica Scandinavica*, vol. 123, no. 5, pp. 327–338.
- Barrett, M. S., Chua, W., Crits-Christoph, P., Gibbons, M. B., Casiano, D., Thompson, D. O. N. (2008). "Early withdrawal from mental health treatment. Implications for psychotherapy practice", in: *Psychotherapy (Chicago, Ill.)*, vol. 45, no. 2, pp. 247–267.
- Bode, K., Vogel, R., Walker, J., Kröger, C. (2016). "Health care costs of borderline personality disorder and matched controls with major depressive disorder. A comparative study based on anonymized claims data", in: *The European Journal of Health Economics*, pp. 1–11.
- Bohus, M., Haaf, B., Simms, T., Limberger, M., Schmahl, C., Unckel, Christine, et al. (2004). "Effectiveness of inpatient dialectical behavioral therapy for borderline personality disorder. A controlled trial", in: *Behaviour Research and Therapy*, vol. 42, no. 5, pp. 487–499.
- Bond, F., Hayes, S., Baer, R., Carpenter, K., Guenole, N., Orcutt, H. et al. (2011). "Preliminary Psychometric Properties of the Acceptance and Action Questionnaire–II. A Revised Measure of Psychological Inflexibility and Experiential Avoidance", in: *Behavior Therapy*, vol. 42, no. 4, pp. 676–688.
- Bundespsychotherapeutenkammer (2011). "BPtK-Studie zu Wartezeiten in der ambulanten psychotherapeutischen Versorgung Umfrage der Landespsychotherapeutenkammern und der BPtK: Bundespsychotherapeutenkammer".
- Campbell, R. S.; Pennebaker, J. W. (2003). "The secret life of pronouns. Flexibility in writing style and physical health", in: *Psychological Science*, vol 14, no. 1, pp. 60–65.
- Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H. et al. (2016). "Cross-trial prediction of treatment outcome in depression. A machine learning approach", in: *The Lancet Psychiatry*, vol. 3, no. 3, pp. 243–250.
- Cinkaya, F., Schindler, A., Hiller, W. (2011). "Wenn Therapien vorzeitig scheitern", in: *Zeitschrift für Klinische Psychologie und Psychotherapie*, vol. 40, no. 4, pp. 224–234.
- Comtois, K. A., Carmel, A. (2016). "Borderline Personality Disorder and High Utilization of Inpatient Psychiatric Hospitalization. Concordance Between Research and Clinical Diagnosis", in: *The Journal of Behavioral Health Services & Research*, vol. 43, no. 2, pp. 272–280.
- Delgadillo, J., Moreea, O., Lutz, W. (2016). "Different people respond differently to therapy. A demonstration using patient profiling and risk stratification", in: *Behaviour Research and Therapy*, vol. 79, pp. 15–22.
- Derogatis, L. (1992). *SCL-90-R: Administration, Scoring and Procedures Manual-II for the Revised Version and Other Instruments of the Psychopathology Rating Scale Series*. Clinical Psychometric Research, Towson, MD.
- Ebner-Priemer, U. W., Trull, T. J. (2009). "Ambulatory Assessment", in: *European Psychologist*, vol. 14, no. 2, pp. 109–119.
- Graubner, B. (2014). *Systematisches Verzeichnis. Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme; 10th Revision - German Modification. Version 2014, Stand 20. September 2013*, Dt. Ärzte-Verl., Köln.
- Hassenzahl, M. (2010). "Experience design. Technology for all the right reasons", in: *Synthesis lectures on human-centered informatics*, vol., 8. Morgan & Claypool Publishers, San Rafael, Calif.
- Hernandez-Tejada, M. A., Zoller, J. S., Ruggiero, K. J., Kazley, A. S., Acierno, R. (2014). "Early Treatment Withdrawal from Evidence-Based Psychotherapy for PTSD. Telemedicine and in-Person Parameters", in: *The International Journal of Psychiatry in Medicine*, vol. 48, no. 1, pp. 33–55.
- ISO (1999). DIN EN ISO 9241 Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten - Teil 11. Anforderungen an die Gebrauchstauglichkeit (ISO 9241-11:1998).
- ISO (2006). DIN EN ISO 9241-110 Ergonomie der Mensch-System-Interaktion - Teil 110. Grundsätze der Dialoggestaltung (ISO 9241-110:2006); Deutsche Fassung EN ISO 9241-110:2006: Perinorm (DIN ISO 9241-110:2006).
- Jordan, P. W. (2010). *Designing pleasurable products. An introduction to the new human factors*, Taylor & Francis, London.

- Knapp, J., Zeratsky, J., Kowitz, B. (2016). *Sprint. How to Solve Big Problems and Test New Ideas in Just Five Day*, Bantam Press (Bantam Press / Jake Knapp), London.
- Kröger, C., Röepke, St., Kliem, S. (2014). “Reasons for premature termination of dialectical behavior therapy for inpatients with borderline personality disorder”, in: *Behaviour Research and Therapy*, vol. 60, pp. 46–52.
- Krug, S. (2013). *Don't make me think! A common sense approach to Web usability*, 3rd Edition, New Riders.
- Landes, S. J., Chalker, S. A., Comtois, K. A. (2016). “Predicting dropout in outpatient dialectical behavior therapy with patients with borderline personality disorder receiving psychiatric disability”, in: *Borderline Personality Disorder and Emotion Dysregulation*, vol. 3, no. 1, p. 9.
- Laurensen, E. M. P., Eeren, H. V., Kikkert, M. J., Peen, J., Westra, D., Dekker, J. M., Busschbach, J. V. (2016). “The burden of disease in patients eligible for mentalization-based treatment (MBT). Quality of life and costs”, in: *Health and Quality of Life Outcomes*, vol. 14, no. 1, p. 145.
- Lese, K. P., MacNair-Semands, R. (2000). “The Therapeutic Factors Inventory. Development of a Scale”, in: *Group*, vol. 24, no. 4, pp. 303–317.
- Linehan, M. M. (2015). *DBT Skills Training Manual*, 2nd Edition, Guilford Publications, New York, NY.
- Lutz, W., Leach, C., Barkham, M., Lucock, M., Stiles, W. B., Evans, C. et al. (2005). “Predicting Change for Individual Psychotherapy Clients on the Basis of Their Nearest Neighbors”, in: *Journal of Consulting and Clinical Psychology*, vol. 73, no. 5, pp. 904–913.
- Lutz, W., Martinovich, Z., Howard, K. I. (1999). “Patient profiling. An application of random coefficient regression models to depicting the response of a patient to outpatient psychotherapy”, in: *Journal of Consulting and Clinical Psychology*, vol. 67, no. 4, pp. 571–577.
- MacKenzie, I. S. (2013). *Human-computer interaction. An empirical research perspective*, Morgan Kaufmann, Waltham (MA).
- McConaughy, E. A., Prochaska, J. O., Velicer, W. F. (1983). “Stages of change in psychotherapy. Measurement and sample profiles”, in: *Psychotherapy: Theory, Research, and Practice*, vol. 20, pp. 368–375.
- Panfilis, C., Marchesi, C., Cabrino, C., Monici, A., Politi, V., Rossi, M., Maggini, C. (2012). “Patient factors predicting early dropout from psychiatric outpatient care for borderline personality disorder”, in: *Psychiatry Research*, vol. 200, no. 2–3, pp. 422–429.
- Passos, I. C., Mwangi, B., Cao, B., Hamilton, J. E., Wu, M., Zhang, X. Y. et al. (2016). “Identifying a clinical signature of suicidality among patients with mood disorders. A pilot study using a machine learning approach”, in: *Journal of Affective Disorders*, vol. 193, pp. 109–116.
- Patton, J. H., Stanford, M. S., Barratt, E. S. (1995). “Factor structure of the Barratt Impulsiveness Scale”, in: *Journal of Clinical Psychology*, vol. 51, no. 6, pp. 768–774.
- Perroud, N., Uher, R., Dieben, K., Nicastrò, R., Huguelet, P. (2010). “Predictors of Response and Drop-Out During Intensive Dialectical Behavior Therapy”, in: *Journal of Personality Disorders*, vol. 24, no. 5, pp. 634–650.
- Poria, S., Cambria, E., Bajpai, R., Hussain, A. (2017). “A review of affective computing. From unimodal analysis to multimodal fusion”, in: *Information Fusion*, vol. 37, pp. 98–125.
- Ramirez, A. O. (2000). “Three-Tier Architecture”, in: *Linux Journal*, vol. 75, Article No. 7.
- Rubel, J., Rosenbaum, D., Lutz, W. (2016). “Patients’ in-session experiences and symptom change. Session-to-session effects on a within- and between-patient level”, in: *Behavior Research and Therapy*, vol. 90, pp. 58–66.
- Rüsch, N., Schiel, S., Corrigan, P. W., Leihener, F., Jacob, G. A., Olschewski, M. et al. (2008). “Predictors of dropout from inpatient dialectical behavior therapy among women with borderline personality disorder”, in: *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 39, no. 4, pp. 497–503.
- Saxon, D., Firth, N., Barkham, M. (2016). “The Relationship Between Therapist Effects and Therapy Delivery Factors. Therapy Modality, Dosage, and Non-completion”, in: *Administration and Policy in Mental Health and Mental Health Services Research*, pp. 1–11.
- Schulte, D. (2005). “Messung der Therapieerwartung und Therapieevaluation von Patienten (PATHEV)“, in: *Zeitschrift für Klinische Psychologie und Psychotherapie*, vol. 34, pp. 176–188.
- Skevington, S. M., Lotfy, M., O’Connell, K. A. (2004). “The World Health Organization’s WHOQOL-BREF quality of life assessment. Psychometric properties and results of the international field trial. A Report from the WHOQOL Group”, in: *Quality of Life Research*, vol. 13, no. 2, pp. 299–310.
- Spinhoven, P., Giesen-Bloo, J., van Dyck, R., Kooiman, K., Arntz, A. (2007). “The therapeutic alliance in schema-focused therapy and transference-focused psychotherapy for borderline personality disorder”, in: *Journal of Consulting and Clinical Psychology*, vol. 75, no. 1, pp. 104–115.
- Stoffers, J. M., Völm, B. A., Rücker, G., Timmer, A., Huband, N. Lieb, K. (2012). “Psychological therapies for people with borderline personality disorder”, in: *Cochrane Database of Systematic Reviews*, vol. 8.

ter Huurne, E. D., Postel, M. G., Haan, H. A. de; van der Palen, J., DeJong, Cor A. J. (2017). "Treatment dropout in web-based cognitive behavioral therapy for patients with eating disorders", in: *Psychiatry Research*, vol. 247, pp. 182–193.

Wagner, T., Fydrich, T., Stiglmayr, C., Marschall, P., Salize, H., Renneberg, B. et al. (2014). "Societal cost-of-illness in patients with borderline personality disorder one year before, during and after dialectical behavior therapy in routine outpatient care", in: *Behaviour Research and Therapy*, vol. 61, 2014.

Weathers, F. W., Litz, B. T., Keane, T. M., Palmieri, P. A., Marx, B. P., Schnurr, P. P. (2013). *The PTSD Checklist for DSM-5 (PCL-5)*, National Center for PTSD.

Wilmers, F., Munder, T., Leonhart, R., Herzog, T., Plassmann, R., Barth, J., Linster, H. W. (2008). "Die deutschsprachige Version des Working Alliance Inventory - short revised (WAI-SR) - Ein schulenübergreifendes, ökonomisches und empirisch validiertes Instrument zur Erfassung der therapeutischen Allianz", in: *Klinische Diagnostik und Evaluation*, vol. 1, pp. 343–358.

Wnuk, S., McMain, S., Links, P. S., Habinski, L., Murray, J., Guimond, T. (2013). "Factors Related to Dropout From Treatment in Two Outpatient Treatments for Borderline Personality Disorder", in: *Journal of Personality Disorders*, vol. 27, no. 6, pp. 716–726.

Xiao, B., Imel, Z. E., Georgiou, P. G., Atkins, D. C., Narayanan, S. S. (2015). "Rate My Therapist. Automated Detection of Empathy in Drug and Alcohol Counseling via Speech and Language Processing", in: *PLoS One*, vol. 10, no. 12, e0143055.

Utilizing High Performance Computing Techniques for efficiently anonymizing sensitive patient data

D. Ntalaperas, A. Bouras
UBITECH Ltd
Di Fuccio Raffaele
University of Naples
e-mail: dntalaperas@ubitech.eu

Keywords: Data Anonymization, High Performance Computing

1. Introduction

Using Data Cubes for anonymizing large sets of patient data is an approach that has been well researched and has been demonstrated to provide a faithful representation of the original content, while also ensuring that the transformed data do not contain any information that can be reversed engineered to identify patients (Antoniades et al. 2012). The Data Cubes approach has been successfully used previously in the context of the Linked2Safety¹ project (Perakis et al. 2013) and is now being used in the context of the SAGE-CARE project² to transform clinical and genetic data of patients being treated for melanoma, in a representation that is safe to distribute among institutions.

The methodology that achieves this result has already been implemented and documented (Ntalaperas et al. 2016); the main purpose of the present work is to describe how the methodology is enhanced by using High Performance Computing (HPC) in order to be able to cope with very large data sets. Indeed, in the case of genetic data it may be possible that the size of original data may become very big, since the gene expression information or Single Nucleotide Polymorphisms (SNPs) that need to be recorded per patient may be in the order of hundreds of thousands.

2. Theory

Figure 1 depicts the methodology pipeline for creating a Data Cube from source clinical and genetic data. Data come from two files, one for clinical and one for genetic data. A module aligns the data so that rows and columns of the source data are correctly aligned based on patient and, if a mapping file is given by the user, a categorization schema is applied to variables selected. The categorization schema defines the ranges that are going to be used for aggregation; a category for variable BMI of 0-25 for example, denotes that all patients with a BMI lower than 25 will be treated the same and will be aggregated together. After categorization, the values for each category are aggregated and a Data Cube is created. Data are then perturbed by adding a small, statistically insignificant, random noise and cells with small values are removed. These last two actions, ensure that the resulting data set is fully anonymized (Antoniades et al. 2012), (Forgó et al. 2012). A complete description of the methodology can be found in (Ntalaperas et al. 2016).

¹ <http://www.linked2safety-project.eu/>

² <http://www.sage-care.eu/>

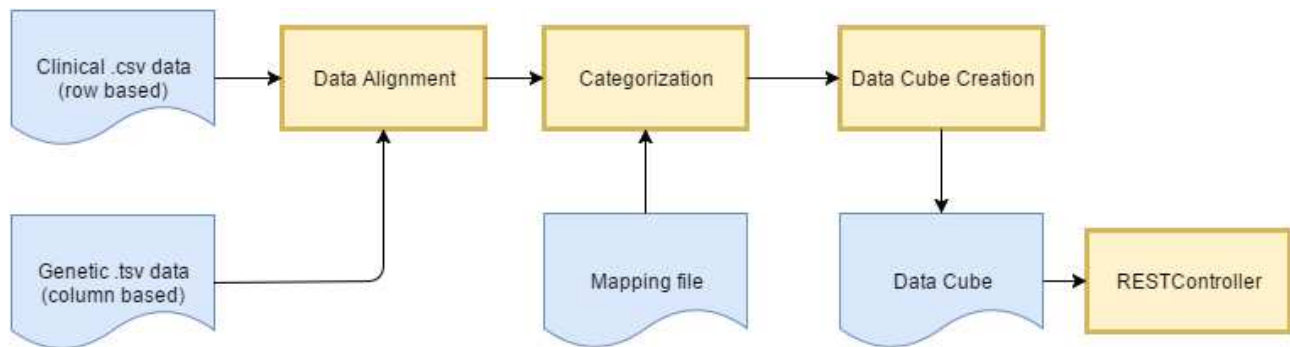


Figure 1: Data Cube Creation Methodology Pipeline

The most time consuming operations in the above pipeline are: a) the Data Alignment where each entry of the original two files need to be matched, b) the categorization where all data need to be traversed and be transformed to the categorical values defined in the mapping file and c) the aggregation of the data. Indeed, it is these three actions that involve operations that need to be performed on every entry of the source data.

3. Implementation and Results

The Data Cube creation algorithm was enhanced by parallelizing the most intensive, in terms of computation, steps. For the Data Alignment step, since each patient entry is independent from the other, a high level of parallelization was achieved. In case of N processors, each processor can be used to combine the two entries of the source files and construct the internal data representation that is to be used by the next steps. As is the case with the non-parallelized version, the main cost of this operation is the time required for I/O access.

In the case of categorization, there are n patient records and m variables to be categorized with each categorical variable having a list of m_i possible values. Parallelization can again be achieved to a high degree due to the fact that the categorization schema is applied to each one of the patient records independently. The only shared data is the mapping file and this is a read only, usually very small file, that can be loaded once in the memory. Data Cube creation finally, is a simple aggregation process where the values of each one of the cells of a multidimensional array is computed by aggregating the number of patient records that share a combination of categorically equal values that is represented by each cell.

Figure 2 depicts the typical work that can be performed by each processor during one iteration of the pipeline in the ideal case where the number of processors equals the number of the patients. Firstly, the processor will align the clinical and genetic data into a single entry. In the second step, it will iterate through each entry of the combined data and will perform a categorization operation according to the schematics defined in the Mapping file. Then, for the combined categorized entry, the corresponding cell in the Data Cube will be incremented. Consider for example that there three variables being monitored; namely BMI, smoking and diabetes. Suppose further that the categorized values of these variables for a specific patient are BMI=1 (corresponding to a normal BMI), smoking=1 (corresponding to a casual smoker) and diabetes=0 (corresponding to the patient not having diabetes). In the above described case, cell $D[1][1][0]$ will be incremented by the processor, so that when all processors end, cell $D[1][1][0]$ will have a value equal to the total number of patients with normal BMI who are casual smokers and do not have diabetes.

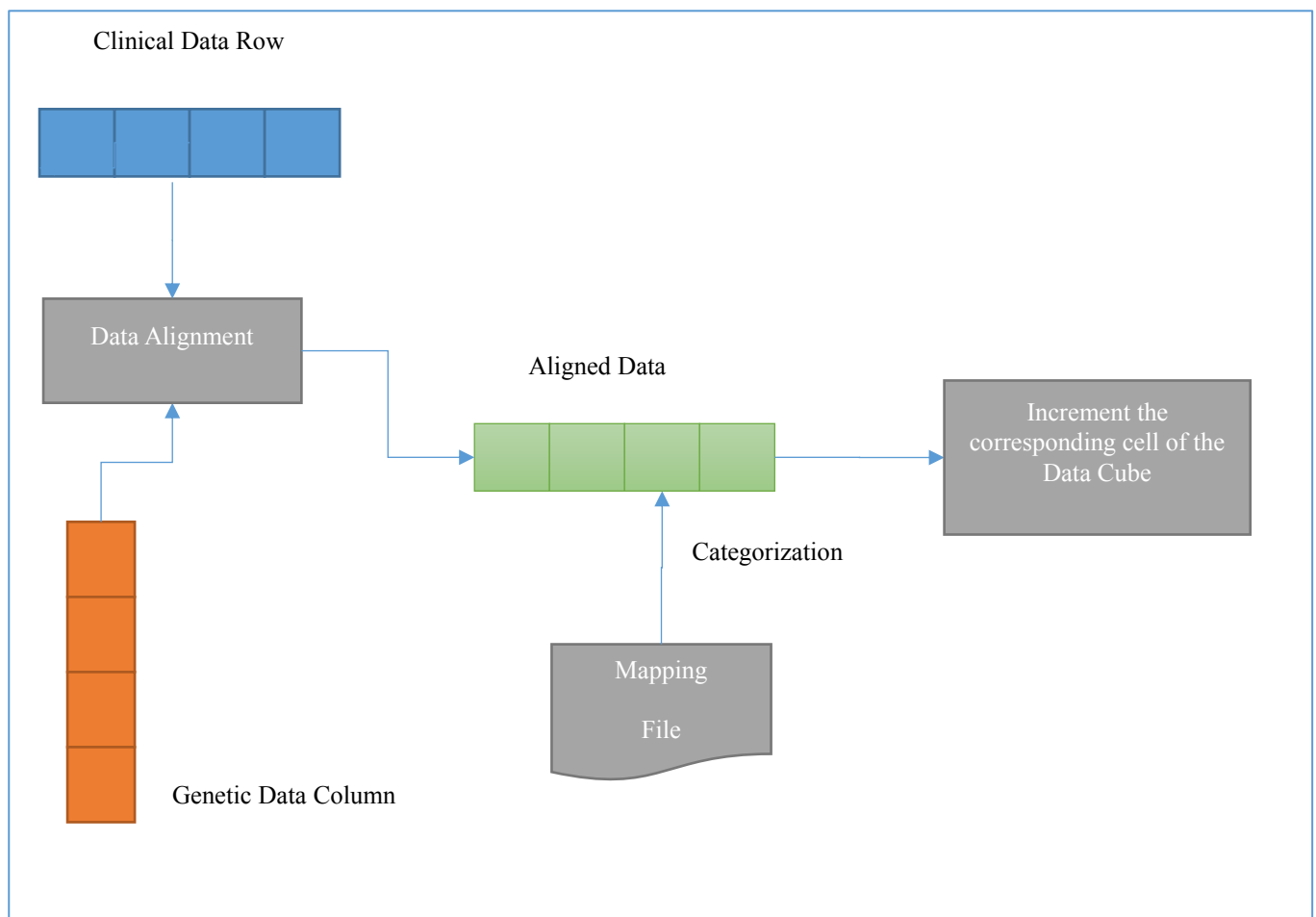


Figure 2: Workload for each processor

The above parallelization methodology has been implemented in the context of SAGE-CARE using the CUDA framework and was tested in a GeForce GTX 960M card. The control machine was the same as the test, the difference being that in the control the non parallelized version of the algorithm was run. First results in the case of a sample of 100.000 mock patients being monitored for three clinical and one genetic variable are depicted in Table 1.

Table 1: Measured response times for the parallelized and non parallelized algorithm. Times are averages of twenty runs

Non parallelized algorithm (ms)	Parallelized algorithm (ms)
4.817	12.8

It can be seen that the speedup achieved by the implementation is of the order of 10^2 . This can somewhat be understood empirically, since the number of cores present in the GTX 960M card are equal to 1024. So, in theory, the parallelized algorithm should run 1000 times faster; however idle

core times introduced during the transfer of data between the RAM and the GPU memory seem to have a limiting effect.

4. Conclusions

Using HPC has been demonstrated to offer a significant speedup during the generation of anonymized patient data. This facilitates the real time request and acquisition of data instead of having to wait or having to pre-generating data that could become outdated.

Future work consists of evaluating the response times of the algorithm to various combinations of sample test data as well as monitoring and measuring processor idle times. The latter will help to determine more accurately what are the bottlenecks that introduce the biggest deviations from the theoretical minimal times and it will provide hints to further enhance the efficiency of the parallelization schema.

References

- Antoniades, A. et. al. (2012). "The effects of applying cell-suppression and perturbation to aggregated genetic data", in: *12th Int. IEEE Conf. on Bioinformatics & Bioengineering (BIBE)*, IEEE.
- Forgó, N., Góralczyk, M. and Graf von Rex, C. (2012) "Security issues in research projects with patient's medical data", in: *12th Int. IEEE Conf. on Bioinformatics & Bioengineering (BIBE)*, IEEE.
- Ntalaperas, D., Mpouras, A. (2016). "An approach for anonymization of sensitive clinical and genetic data based on Data Cube Structures", in: *Collaborative European Research Conference, 2016*.
- Perakis, K. et. al. (2013). "Advancing Patient Record Safety and EHR Semantic Interoperability", in: *IEEE International Conference on Systems, Man, and Cybernetics*, IEE.

Challenges and limitations of large-scale molecular biology analysis and visualization tools from the user experience perspective

C. Palu^{1,2}, B. Lawlor¹, B. Kelly¹, P. Walsh¹

¹NSilico Life Science Ltd

²University College Cork
Cork, Ireland

e-mail: cintia.palu@nsilico.com

Keywords: data visualization, bioinformatics, large-scale biological data.

1. Introduction

Exploratory analysis is a key step in any data driven study and such exploration often relies on visualization to represent the information being studied. For molecular biologists and other researchers relying on large-scale datasets, visualization and statistical analysis can have an even higher value because the data dimension can limit manual exploration. The development of Next Generation Sequencing (NGS), over the last 2 decades and its present popularisation, combined with the discovery of the importance of biological features, such as small RNAs and epigenetics factors, has created a new level of data complexity, not only due to the large data volumes but also due to the complexity of interaction possibilities. Consequently, statisticians, computer scientists and bioinformaticians were pushed to create new solutions to store and analyse this data. Here we discuss the challenges regarding data analysis and visual representation, from the user point of view, based on feedback from an online survey.

Currently, the workflow to use large scale biological data relies on many professionals with complementary information, but gathering all the information and translating it into applied knowledge remains a big challenge (Gehlenborg et al., 2010). Frequently, the reason to not achieve the full discovery potential within this type of data is that the bench scientist does not master the knowledge to do the data analysis and the bioinformatician or statistician does not understand the biology of the dataset. Interpreting the results relies a lot on the data exploratory analysis and is also an important issue in this field, where visual representation is very beneficial to synthesise information and/or show tendencies.

In order to further understand what is currently limiting researchers involved in molecular biology research, we launched an online survey addressing the user experience regarding the visualization options currently available.

2. Methodology

The survey was presented in three parts: 1) Background – aiming to identify the respondent profile regarding familiarity with bioinformatics, education level and current job; 2) Biological data analysis, which had the objective of categorising the users study and analysis routine; 3) Visual representation of molecular biology data, which assessed the visual resources frequently used and their practice, the user experience and relevant features when evaluating software.

The questionnaire was created on Google Forms (Google, 2017) and a login was required to avoid duplicated entries (Figure 1). It had blended types of question: multiple-choice, check-box, open and Likert-scale and the data-analysis was carried according to the question type. An initial evaluation was done using Google Analytics, whereas the open questions were read and later summarized using

word-clouds (Zygomat, 2017). The background questions were used to separate the users into categories and evaluate variations according with the public of interest, using R (The R Foundation for Statistical Computing, 2017). The survey was available in July 2017 and was publicised in mailing lists from research institutions and social media groups related to the topic.

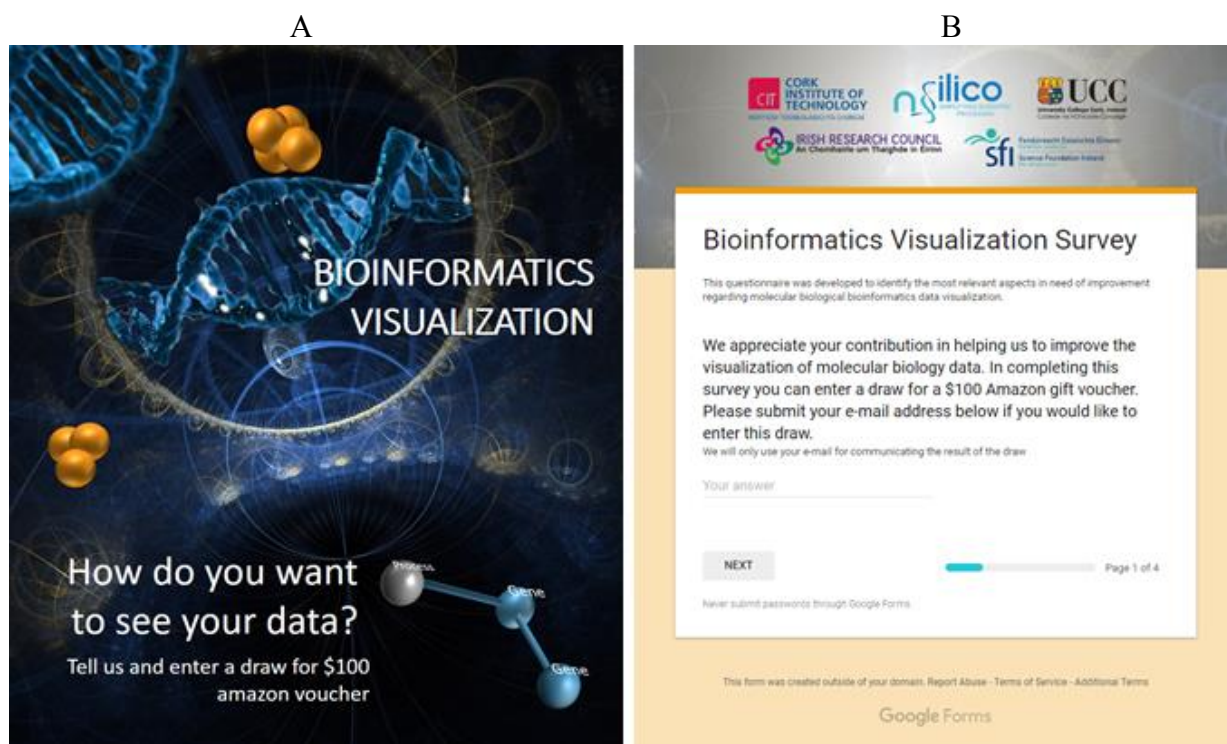


Figure 1: Survey marketing and interface

3. Results and discussion

We received 60 answers, whereas 46.7% of the respondents were bioinformaticians and 31.7% were able to use some bioinformatics resources (Figure 2). Nearly 75% are currently developing work in the academic field and we also received responses from industry professionals (13.3%). Regarding core interest, the majority of our users are involved with healthcare (45%), metagenomics (28.3%) and animal research (30%). A relevant number of respondents work with natural sciences (21.7%), prokaryote (15%) and fungi (8.3%).

Most of the feedback from users in the survey was gathered using Likert-scale responses. In general, all analysis had a score between seldom and average use, because analysis strategies were heterogeneous between the users (Figure 3). Genome/transcriptome comparison was the most popular type of analysis across the respondents, followed by assembly and differential expression. Further analysis was done by comparing responses from those who identified as bioinformaticians versus non-bioinformaticians. This showed that answers depended a lot on the person's background and as expected bioinformaticians have a higher use of all types of analysis, see Figure 4. In order to investigate if we were able to reach our intended target audience, we also compared the users total and maximum scores. Interestingly, at least 75% of our respondents use at least one of the analyses frequently, suggesting that we were successful in reaching a relevant target audience. The users also rely on network visual representation, heatmaps and genome/transcriptome browsers and bioinformaticians use these resources more frequently, on average (data not shown).

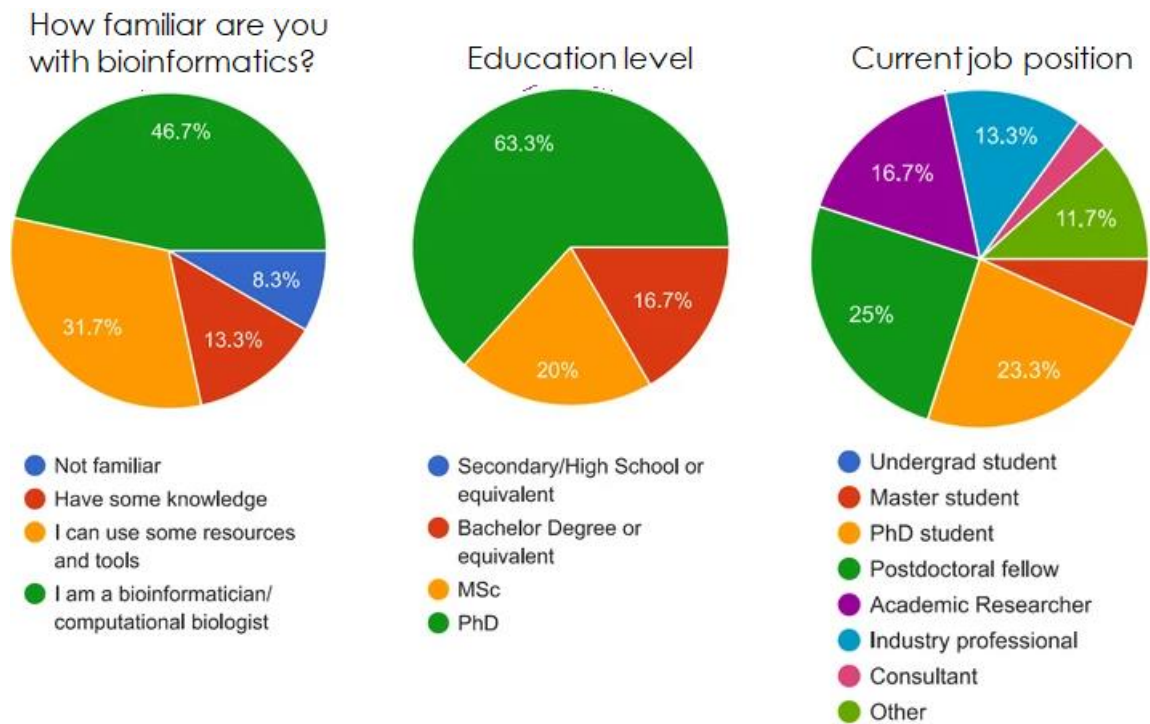


Figure 2: Respondents background

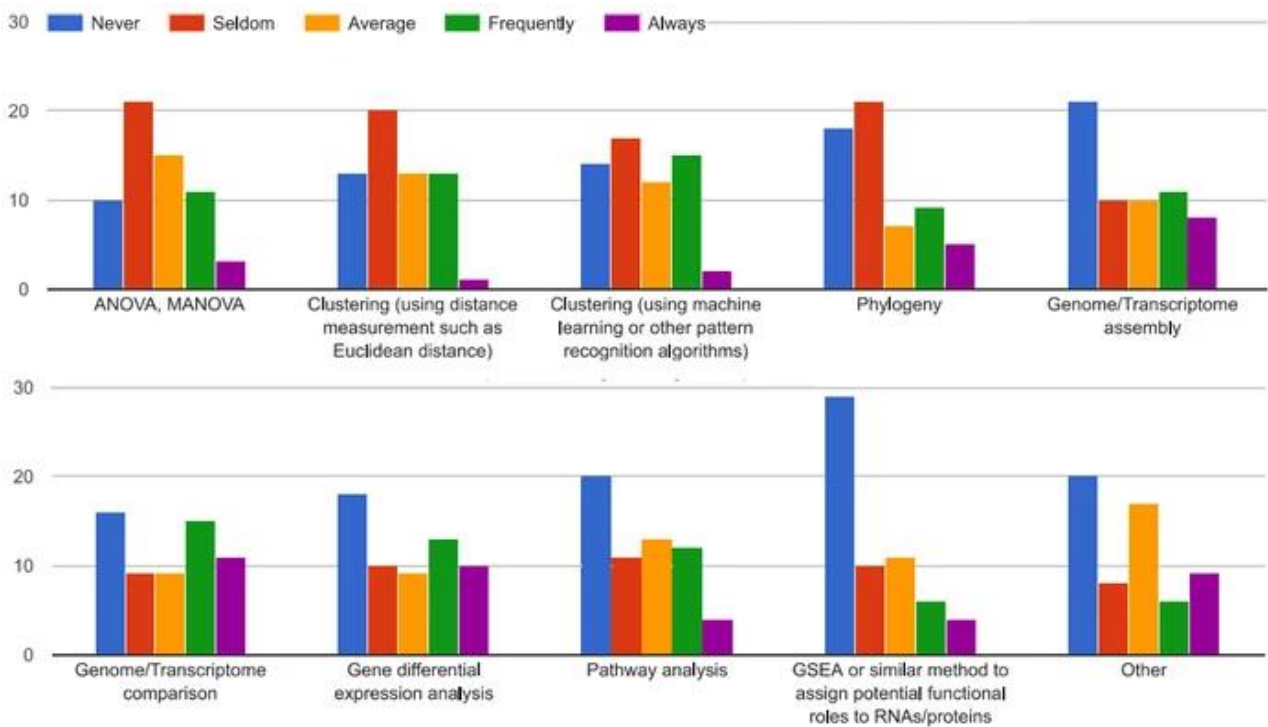


Figure 3: Analysis approach

Average points by analysis type, separated between bioinformatician and other professionals
(never = 0; seldom = 1; average = 2; frequently = 3; always = 4)

Analysis	bioinfo	other	Total
Genome/Transcriptome comparison	2.39	1.53	1.93
Gene differential expression analysis	2.32	1.31	1.78
ANOVA, MANOVA	1.54	1.66	1.60
Other	1.86	1.38	1.60
Genome/Transcriptome assembly	2.14	1.09	1.58
Clustering (using machine learning or other pattern recognition algorithms)	1.75	1.41	1.57
Clustering (using distance measurement such as Euclidean distance)	1.89	1.13	1.48
Pathway analysis	1.89	1.13	1.48
Phylogeny	1.57	1.19	1.37
GSEA or similar method to assign potential functional roles to RNAs/proteins	1.36	0.88	1.10

User individual profile analysis

Points by analyse type	Min	1st Qu	Median	Mean	3rd Qu	Max
Sum	0.00	9.75	17.00	15.50	21.25	31.00
Highest	0	3	3	3.167	4	4

Figure 4: Comparison of survey results between bioinformaticians and non-bioinformaticians.

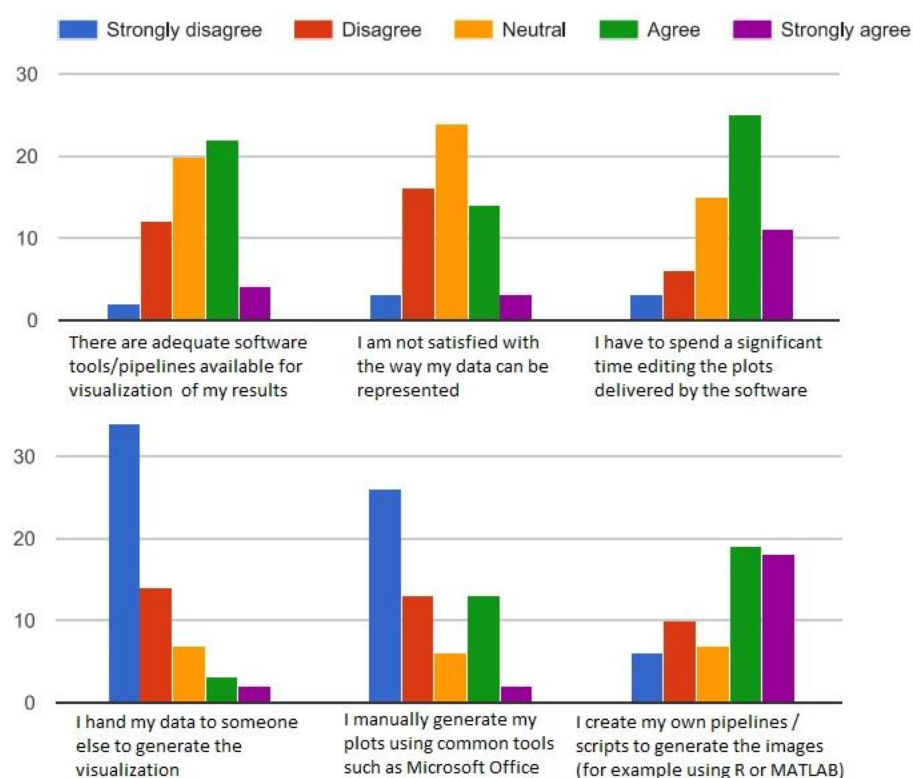


Figure 5: How do you produce visual representation of your data?

The majority of users generate visualization by themselves (Figure 5), whereas further analysis showed that 35 respondents create their own pipelines and scripts to generate images, whereas 23 are bioinformaticians. Twenty five percent of users rely on Microsoft Office resources and although they have a range of bioinformatics knowledge, they are less likely to create their own pipelines. Most of the respondents remained neutral regarding their satisfaction with the current options for data representation and the remainder are split between satisfaction and dissatisfaction. There was no relevant difference between bioinformaticians and non-bioinformaticians, however this topic can be positively related to the user ability and independence in creating pipelines and visualizations. Nevertheless, 60% agree that the time spent to generate visualizations is relevant, and 25% remained neutral in this regard. Those users also have a higher score in the later questions regarding input formatting.

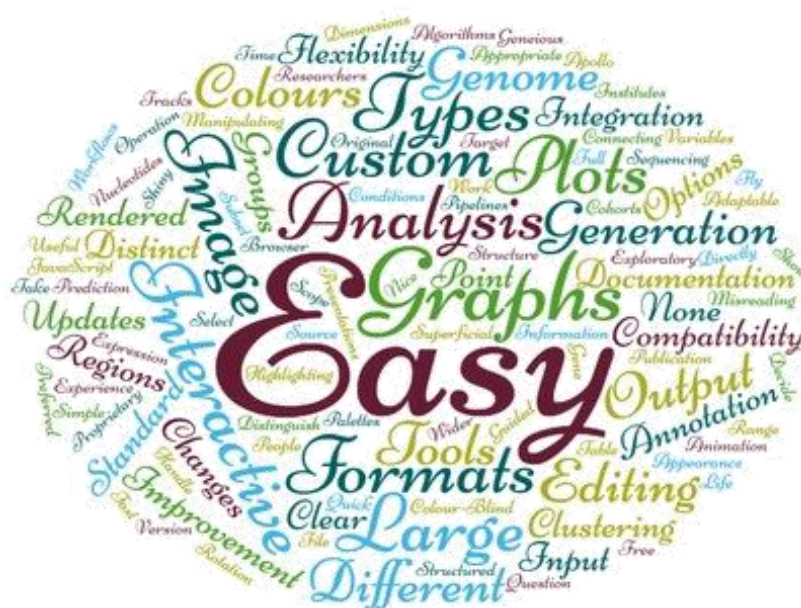


Figure 6: 'In your experience, what features would most improve molecular data visualization software?' answers word cloud

Independent of their knowledge background, clear documentation was the most frequent request and our respondents believes that "user-friendliness" is directly related to good quality documentation. Input type flexibility and auto-detection of format are also very popular expectations. They also want to have more control over the output, with adaptable, customisable visualizations (Figure 6).

Finally, to assess the user experience, we described 11 situations and asked if the users have ever encountered them and how serious the issues were. We also enquire how heavily the respondents would weight each of those issues when evaluating software (Figure 7). The poor documentation is the issue with highest score in general and bioinformaticians find it more difficult than others. The time spent to learn is also a relevant and is more of an issue for non-bioinformaticians. In fact, more than half of our audience consider documentation, interactivity and learning time relevant or very relevant aspects when evaluating analysis and visualisation software. The issue of payment charges for software split our respondents because it can be a blockage for many users, however 33.3% don't mind paying for such software. In general, the respondents opinions regarding the other relevant issues when evaluating a new software were variable and could not be explained by familiarity with bioinformatics.

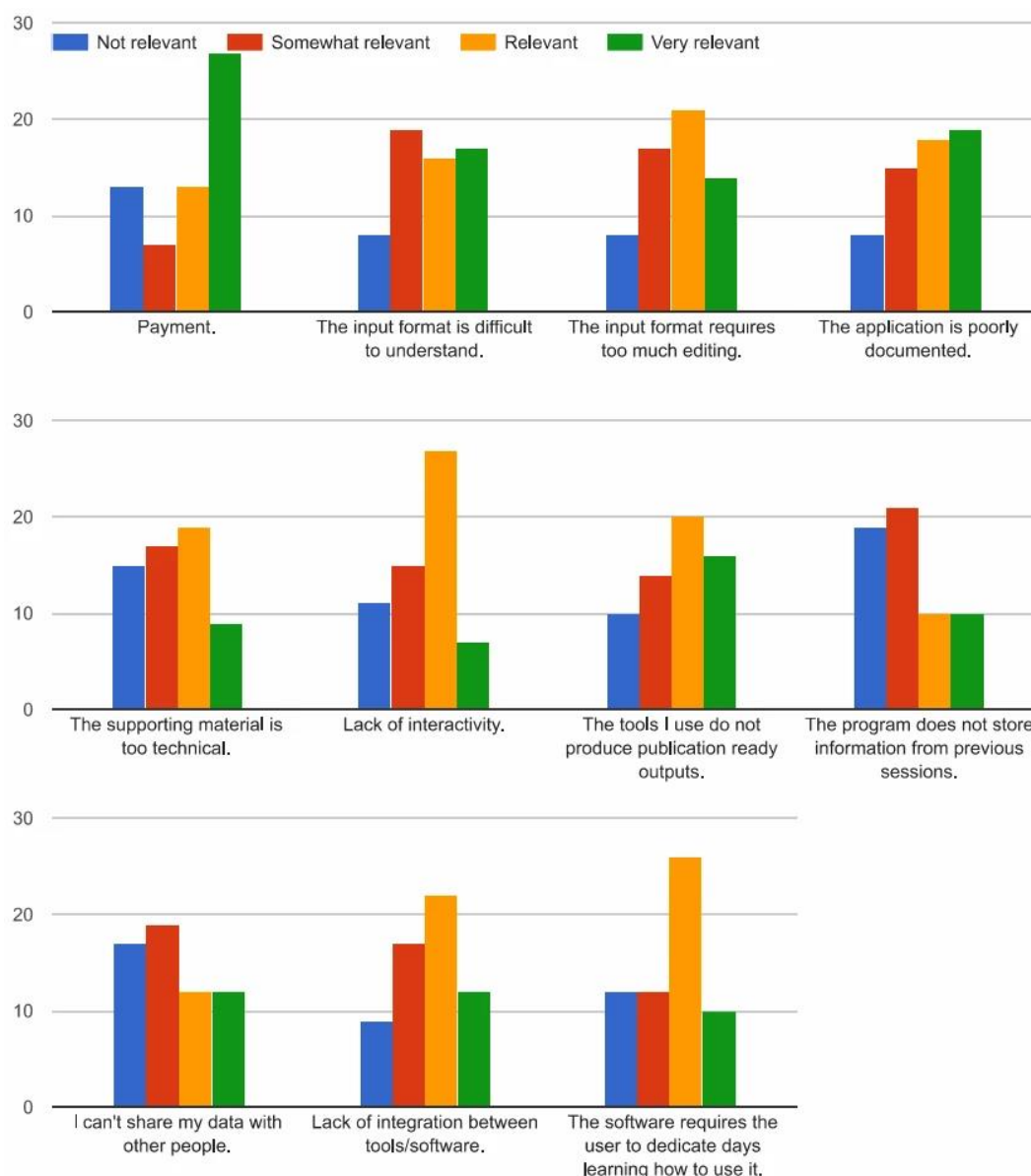


Figure 7: When evaluating a new software for visualizing data, how heavily would you weight each of those issues, if encountered?

4. Conclusions

The aim of this work was to assess user requirements regarding visualization resources for large-scale molecular biology data analysis. Despite the existence of assorted pipelines and tools, the analysis process and interpretation of such data remains a blockage. Technological development such as Next Generation Sequencing supports the advance of life sciences but it must be complemented by adequate strategies to handle, analyse and interpret the captured data. When developing solutions for biological large-scale data it is crucial to keep in mind the diversity of platforms and consequent variation in outputs. Additionally, the developer should dedicate some effort to deliver clear documentation and supporting material to guarantee the correct and efficient use of the product

5. Acknowledgment

Paul Walsh is funded under Science Foundation Ireland (SFI) Industry Fellowship Scheme Grant number 16/IFA/4342. Brian Kelly is supported through the RISE fellowship scheme, grant agreement

number H2020-MSCA-RISE 690998, funded under European Commission Horizon 2020 Programme. Cintia Palu is supported through Irish Research Council Enterprise Partnership Scheme (Postgraduate).

References

Gehlenborg, N. *et al.* (2010). "Visualization of omics data for systems biology", In: *Nature Methods*, 7(3s), pp. S56–S68. doi: 10.1038/nmeth.1436.

Google (2017). *Google Forms*. Available at: <https://docs.google.com/forms>.

The R Foundation for Statistical Computing (2017). 'R'.

Zygomatic (2017.) *Word clouds*. Available at: <https://www.wordclouds.com/>.

An Interactive Interface for Statistical Modelling of Gene Expression Data

K Briand¹, C. Palu^{1,2}, B. Kelly¹, P. Walsh¹

¹NSilico Life Science Ltd

²University College Cork
Cork, Ireland

e-mail: cintia.palu@nsilico.com

Keywords: differential expression analysis, RNAseq, translational bioinformatics

Introduction

Transcriptomics is a powerful tool for molecular biologists and healthcare scientists. It can reveal which genes are being actively translated in a sample. This is valuable information as cells with the same DNA sequence can produce very different patterns of transcription. Transcriptomics has many practical applications, such as comparing which genes have an increased or suppressed expression in different conditions, along timelines, between healthy and unhealthy tissues, or treatment and control samples.

A transcriptome analysis involves many steps, carried out by different experts, from sample isolation and sequencing to the alignment, data analysis and interpretation. As in most Next Generation Sequencing (NGS) studies, there is a bottleneck in bioinformatics analysis, which is a time consuming and crucial step that relies on the complementarity of knowledge of the wet-lab scientists and the bioinformatician. Many tools attempt to address this issue by automatizing the differential expression analysis (comparison of gene expression between sample groups). This potentially enables the bench scientists to explore their own data but these tools are still not effective at incorporating the complexity of some experimental designs thus limiting the information obtained. One of the most robust libraries for gene expression analysis is EdgeR (Robinson, et al., 2010), an R package uses a statistical model design based on contrasts and can apply General Linear Models (GLM) (Lun, et al., 2016) to identify differences in the expression patterns between groups. However, stabilising data models used with such packages can be challenging if the experiment relies on many variables and if there are frequent questions regarding the correct approach on online forums.

We present an implementation of a web-based interface to support the analysis of gene expression data for a bioinformatics platform known as *Simplicity* (Walsh, 2013). This project involves the development of an ASP.NET Model-View-Controller (MVC) (Leff, 2001) based application for managing and processing both clinical and genomic read count input files. We designed a single page application (SPA) model (Mikowski, 2013), in order to prevent the page from reloading frequently and improve the user experience. This also means that all calls to the controllers are made through AJAX calls. Each time an AJAX call is made to a controller, a new instance of this controller is created, and as each controller has its own model as an attribute, a new instance of the model class is created, so that controllers are not concerned with storing data (Hürsch, 1995). We have designed these classes to model the statistical analysis. These models include "Simple" and "Interaction" statistical analysis, where the user can quickly add variables to the statistical analysis model. In statistical analysis, two independent variables interact if the effect of one of the variables differs depending on the level of the other variable. Prior to running computationally expensive models, we conduct statistical validation. The validation consists of testing if the model is correct before doing the analysis. Its role is to tell the user if something is wrong before going further and potentially wasting valuable CPU cycles. The overview of the validation process is depicted in the flowchart in Figure 1 below.

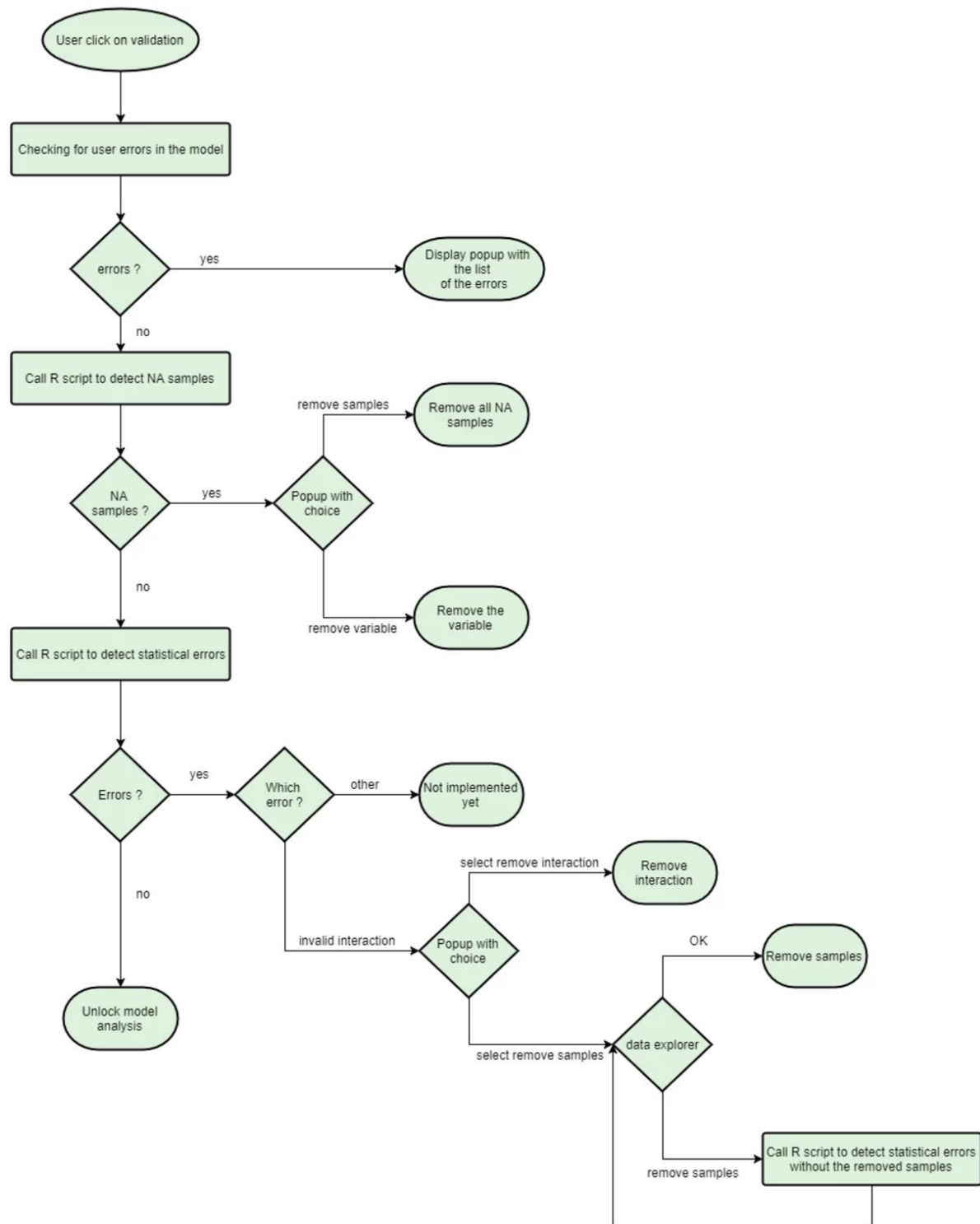


Figure 1: Statistical model validation flow chart.

Our implementation model covers variation on input file format and allows the definition of phenotypic variables' additive and multiplicative effects, which are used to test variation within samples, using edgeR (Robinson, et al., 2010).

Using the phenotypic data, we are able to test if there are enough samples in each category of the statistical model to provide meaningful analysis. The user can control sample filtering, dispersion method, threshold for logFC and FDR corrected p-value. The interface is intuitive and practical while is capable of handling complex data models in the necessary cases.

Acknowledgements

Paul Walsh is supported through the Science Foundation Ireland Industry Fellowship programme. Brian Kelly is supported through the RISE fellowship scheme, grant agreement number H2020-MSCA-RISE 690998, funded under European Commission Horizon 2020 Programme. Cintia Palu is supported through Irish Research Council Enterprise Partnership Scheme (Postgraduate).

References

- Hürsch, W. L. a. C. V. L. (1995). *Separation of concerns*.
- Leff, A. & R. J. T. (2001). "Web-application development using the model/view/controller design pattern", In: *Enterprise Distributed Object Computing Conference, EDOC*. s.l., IEEE.
- Lun, A., Chen, Y. & Smyth, G. (2016). "It's DE-licious: a recipe for differential expression". In: *Methods in Molecular Biology*, Issue 1418, p. 391–416.
- Mikowski, M. S. & P. J. C. (2013). *Single page web applications*. B&W.
- Robinson, M., McCarthy, D. & Smyth, G. (2010). "edgeR: a Bioconductor package for. In: *Bioinformatics*, Issue 26, pp. 139-140.
- Walsh, P. C. J. & S. R. D. (2013). "Accelerating in silico research with workflows: a lesson in simplicity". In: *Computers in biology and medicine*, 43(12).

Case Study: Management of affective data and data fusion in eHealth appliances under consideration of legal frameworks

Ulrich Kowohl[†], Felix Engel^{††}, Patrick Dassler[†] and Matthias Hemmje^{††}

[†] FernUniversität in Hagen, Universitätsstraße 1, Hagen, Germany,

^{††} Research Institute for Telecommunication and Cooperation, Dortmund, Germany,

Email: Ulrich.Kowohl@Fernuni-Hagen.de
Patrick.Dassler@Studium.Fernuni-Hagen.de
{Fengel,Mhemmje}@FTK.de

Abstract. With passing the law for General Data Protection Regulation (GDPR, EU 2016/679) the European Union strengthens and unifies personal data regulations throughout European member states. This case study introduces an approach on sensory and sentiment data fusion within the European funded Sensor Enabled Affective Computing for Enhanced Medical Care (SenseCare) project and outlines the impact of GDPR on Affective Computing (AC) data management. It discusses and evaluates the data management component of AC data fusion within SenseCare regarding the requirements introduced GDPR and outlines areas of further research to follow up.

Keywords: SenseCare, Information Lifecycle Management, Person Data Management, Medical Data Management, Sensory Data Fusion, Security by Design, General Data Protection Regulation

1. Introduction

The care sector faces particular challenges in consideration of the demographical change. People in need of care worldwide will increase from the level of around 350 million to 488 million in 2030 and 614 million in 2050 (Wimo et.al., 2009), (Wimo et.al, 2010a). This will result in escalating health care costs. Sensor Enabled Affective Computing for Enhancing Medical Care (SenseCare) is one of a set of research projects co-funded from the European Commission (EC) that aims to contribute research in eHealth appliances in home care settings. SenseCare aims on providing software services applied to the care of people with dementia. Therein, the SenseCare platform will integrate data streams from multiple sensors and will fuse these streams to provide a global assessment that includes objective levels of wellbeing and emotional state. SenseCare is thematically aligned with current EU Horizon 2020 (Wimo et.al. 2010b). Sensor data streams in SenseCare are applied to infer a person emotional state from various extracted features. The fused data will be consumed by other tools or stakeholders in order to use those for further data processing or to gather additional information about the wellbeing of the data subject via further fusion levels such as "situation assessment" or "impact assessment". This makes storage of AC data and AC affected personal data an important key aspect within the SenseCare AC software ecosystem. The comprehensive representation- and effective access to resources involved is a challenging task, because of the complex interrelations of involved resources and legal frameworks that have an impact on their later use. SenseCare introduces the need of data storage and management. Therefor SenseCare introduces a data management dimension which needs to be legal compliant. This case study will discuss two involved aspects that we faced in the SenseCare project: The integrated representation of data involved in AC analysis and considerations related to its reuse by means of influences from legal frameworks. Overall this paper introduces a possible solution to represent the relation between affective and its source data where sentiment data

has been mined from as well as it outlines legal requirements introduced by processing and storing private data. It outlines the effect of legal requirements to the current solution design and evaluates the current solution against legal requirements. Therefore, we will first introduce our approach in SenseCare towards realisation of semantic data fusion from sensory data streams and AC emotion analysis results to enable integrated storage as well as access and will then follow with an introduction of legal frameworks and their influence in this personal data processing- and management setting. Subsequent, we will document a comprehensive discussion about the impact legal frameworks on data representation, storage and access.

Semantic Data Fusion in SenseCare

Systematically teaming up computer and data science findings with psychological knowledge is the path to render optimal results for use in supporting people living with dementia to remain at home and their related care givers. Therefore, SenseCare Platform (SCP) is defined as a computer based medical support system which stores and manipulates with affective data by using medical data management technologies.

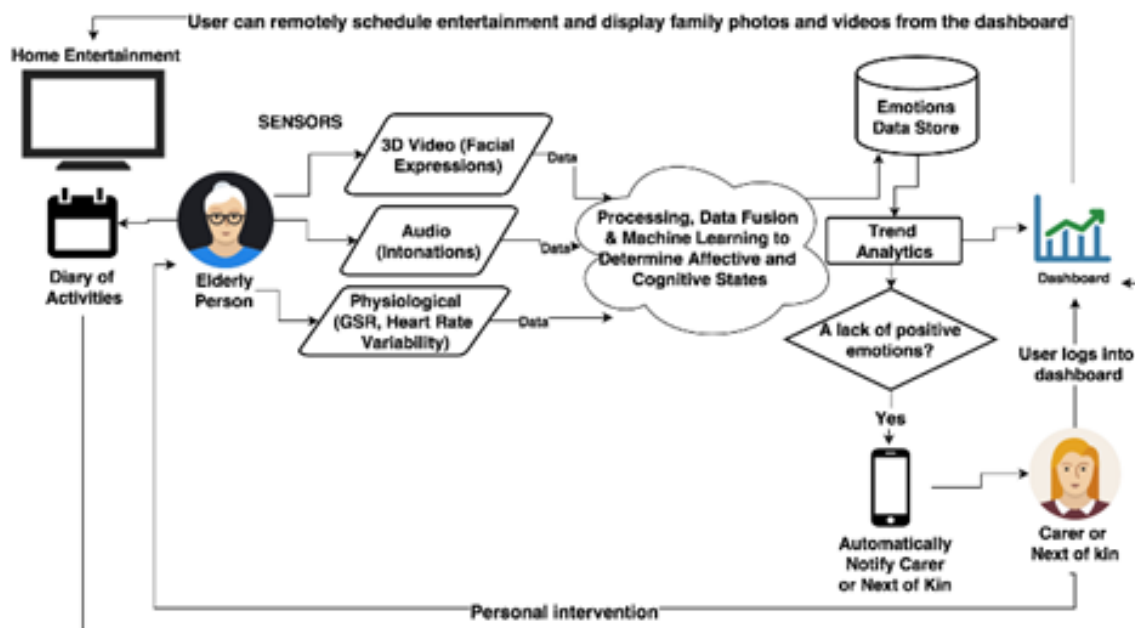


Figure 1: SenseCare Overview

SenseCare defines three different usage scenarios. Overall SCP uses recorded sensory data, in order to detect additional information about the monitored person with dementias (PwD) wellbeing. Emotional Monitoring defines the possibility of monitoring the emotional state of a PwD and serves as a live monitoring and feedback system in closed sessions e.g. therapy sessions. Assisted Living takes place at the PwD's home. SenseCare monitors the emotional state and gives feedback to other caregiving stakeholders. Shared Care Giving defines the care giving process by itself. Care givers are using the data gathered by SCP for care giving process optimization and documentation. Here additional data can be generated through the care giving process as well and linked to already existing data. To enable comprehensive access and use of streamed and analysed sensor data, the so called SenseCare semantic data fusion layer will create and integrate semantic representations of all

available resource in a global schema to avoid heterogeneity conflicts during access. This data has to be managed by an additional data management layer (Kowohl et.al., 2016), (Engel et.al., 2016).

Ontologies in the semantic data fusion layer will be specified to build a global structure that allows semantically enrichment and fusion of all resources available in SCP and their interrelations. Therefore the SenseCare Data Fusion Ontology (SDFO) has been developed consisting of three layers. First Layer (core) addresses all high-level health care related concepts as upper ontology. The Middle layer addresses concepts and interrelations of dementia care and connected health and serve as domain specific ontologies. Those ontologies do not serve all requirements from SenseCare, so that so called domain adapters will refine the domain ontologies with respect to further concepts and interrelations required. This architecture is displayed in Figure 2.



Figure 2: SenseCare Architecture

Starting from an ontology specification document based on competency questions, the relevant domains for the SDFO were identified. SDFO was found to be in need for concepts from biomedicine, emotions, time, locations, sensors and metadata for media files. A non-functional requirement for the development of the SenseCare data fusion layer imposes the reuse and integration of existing ontologies instead of the creation of SenseCare specific concepts as widely as possible. Two main reasons cause this non-functional requirement: First there are no domain experts for the mentioned fields involved in SCP development. Reuse of existing ontologies allows for the SDFO being scientifically correct. Second the integration of existing ontologies enhances SenseCare interoperability with other semantic-web-based health care applications.

During a thoroughly analysis of existing ontologies factors like domain, availability, form, obligations and represented concepts were analysed to assess the usability for integration in SDFO. For the biomedical module ontologies listed in OBO-Foundry (Smith et.al., 2007), (Noy et.al., 2009) were assessed to be suitable. OBO-Foundry lists a variety of ontologies of the medical and health care domain, which are optimized to be easily connectable as they use the same Upper Ontology (Basic Formal Ontology, BFO). For SDFO the domain ontologies DOID, OBI and OGMS are integrated. Further OBO-Foundry-listed ontologies were selected to represent emotional concepts (MFOEM) and locations (ENVO). Concepts to represent time instants and time periods are derived from the ontology OWL-Time. To represent the origin of measurement values the "Semantic Sensor Network Ontology" (SSN) has been selected for integration. Last, to represent MPEG-7 Metadata "Core Ontology for Multimedia" (COMM) (Arndt et.al., 2007) is used. All mentioned ontologies can be seen as part of the "Domain Ontology Layer" as depicted in Figure 2. A Core (or "Upper")

Ontology layer emerges from the fact that all mentioned ontologies are based on some Upper Ontologies by themselves. For the OBO-Foundry ontologies this Upper Ontology is BFO and the "Relation Ontology" (RO) which provides all RDF predicates relevant for BFO. SSN is based on "DOLCE Ultra Lite" (DUL) and COMM uses "DOLCE Very Lite". The Layer of "Domain Adapters" of SDFO includes all concepts that have been generated under a SenseCare-own namespace. The task of this adapter-layer-concepts is mainly to provide means of ontology integration such as predicates connecting classes from one ontology to classes from another. Furthermore, some very SenseCare-specific concepts that cannot be represented using any existing ontology are created under that layer such as references to an archive ID. The SDFO therefore follows the three-layer design as depicted in Figure 3.

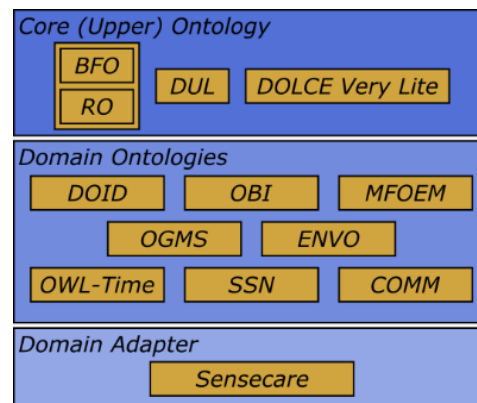


Figure 3: SDFO Architecture

All selected ontologies are available as RDF data and have been integrated in a single OWL file. Also, all concepts of the domain adapter layer have been included in that ontology file. Besides the creation of a single ontology file providing all necessary concepts to represent SenseCare's fused database, the use of this concepts has been specified. As an example, the connection of a PwD to an emotion measurement value shall be described. The PwD itself will be generated as a node of type "mf:human being" (a class which is part of the MFOEM ontology). Whilst "mf:human being" is part of the domain layer of SDFO it is derived as a subclass of "bfo:continuant" from the core layer. The recognized emotion of the PwD will be generated as a node of a specific emotion class from MFOEM, e.g. "mfoem:anger". All emotion classes as well as their super-class "mfoem:emotion process" are part of the domain layer. They are derived, once again, from a class of BFO from the core layer (bfo:occurrent). The necessary connection between the PwD and its emotion could be stored using the predicate "participates in" from RO.

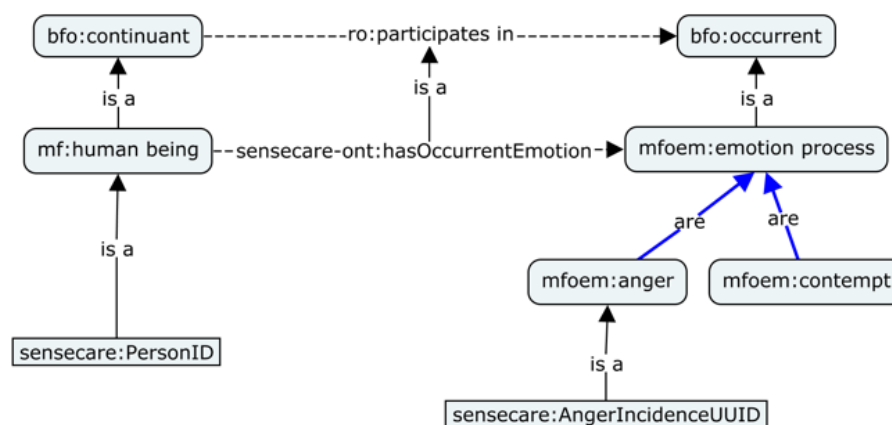


Figure 4: Representation of PwD in SDDFO

Whilst this predicate formally fulfils the requirements of "bfo:continuant" as domain and "bfo:occurrent" as range, it is very unspecific and would make the database unnecessary difficult to understand. It has been decided to generate a predicate "hasOccurrentEmotion" as a sub-predicate of "participates in". This predicate is part of the domain adapter of SDFO.

Figure 4 shows the usage of concepts of the core-, domain- and adapter-layer of SDFO to generate a particular information in the semantic data store.

GDPR and Privacy by Design

Research in the field of AC data management is inherently interconnected with personal data management and its processing. Therefore we will introduce within the next chapters the General Data Protection Regulation (GDPR) of the European Union (EU), the security by design principle which is demanded by GDPR as well as the regulatory needs of the EU member state Germany which have to be fulfilled by SenseCare as data management platform for affective data. With passing the law for General Data Protection Regulation (GDPR, EU 2016/679) the EU strengthens and unifies personal data regulations for all members. GDPR aims on easing international business by universal rules on handling data protection issues. The regulation follows strictly the privacy by design principles (Art. 25) which can be considered as legal framework on solution design related to data protection and security. (De Hert et.al., 2012) Privacy by Design (PbD) is an approach of systems and software engineering which takes privacy concerns already at system design phase into account and introduces security and privacy as one additional stakeholder into the process. PbD consists of 7 principles: Proactivity defines the way of thinking during solution concept phase to prevent data breaches by design concept. Security will be considered as one of the main Stakeholders. Security should be independent of human risk factors – e.g. due to misconfiguration of the system – for data breaches. Security audits should be established, the use of auditable standard technologies is encouraged. Tradeoffs between functionality and security should be avoided while privacy is the default within the solution. Purpose of data collection, disclosure, retention and usage has to be defined upfront, collection of data beyond purpose has to be limited so that collected data is kept to a minimum. Collected data has to be minimized in terms of identifiability, observability and link ability of personal information. Usage, retention and disclosure is limited to purpose, data has to be erased as soon as purpose vanishes. Security has to be applied on infrastructure level e.g. logging, transport layer as well (End to end security), entities have to be responsible for security of their dedicated contained information by themselves. The system has to be user centric, user produced metadata should be kept to a minimum, the user should be informed about the data which is stored about him. A consent of data collected has to be existing between the user and the system. Data within the system has to be complete and correct related to purpose (Accuracy) as well as compliant to regulatory. There should be no data access beyond purpose. Auditing Processes and security policies have to be visible within the entire system as well as accountability for collected data is laying at the system. Privacy of design principles introduce demand for continuous data protection and compliance. This applies throughout the entire lifecycle of data involved. (Schaar, 2010) German laws apply special requirements like specific retention periods for different types of medical data (§35 par. 2 BDSG, §630 par 3 BGB and §199 BGB) setting retention periods from 10 to 30 years for medical data.

Consideration of legal frameworks in an access component

GDPR has direct implication on SenseCare's data fusion layer, because processed sensory data is inherently personal and sensitive, its security is absolute required. Security can be defined by functional requirements. The SDFO allows connection of measurement data to the person it was acquired from which makes it necessary to manage them by Security by Design. Deletion after

vanishing of purpose, respecting retention periods and data correctness after deletion of remaining interconnected data has to be ensured. SenseCare infrastructure contains two data stores: Media Archive contains raw data and the semantic database contains fused RDF data. Both are considered through deletion and have to be consistent together. Retention period is currently configured in global scope. Deletion function takes a patient id and a timeframe as input. Given arguments are checked against retention period. Deletion in semantic database is depended on the database structure. Figure 5 shows a simplified version of a data set related to a patient. The Archive ID for each sensory data links from semantic store to raw data storage.

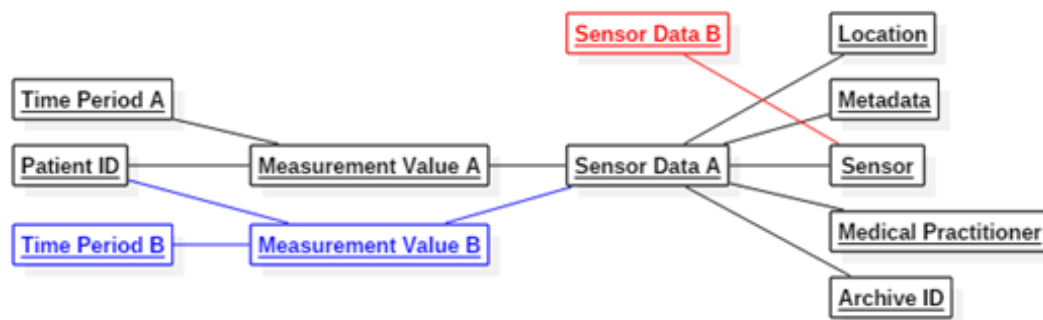


Figure 5: SenseCare semantic database structure

A difficulty in data deletion arises from the various connections linking data to be deleted to other data to be retained. For example, various measurement values are derived from the same sensor data and various sensor data is derived from the same sensor (Figure 5, marked in blue and red). Deletion in the SenseCare semantic data storage therefore follows a hierarchical approach: First, all measurement values fulfilling the selection criteria of `patient_id` and `time_period` is deleted. All affected sensor data are flagged as “dirty”. Second, all dirty sensor data is checked for further connection to measurement values. If no connections existing, data is deleted. Additional data of deleted sensor data is flagged (e.g. locations). Third, all flagged additional sensor data is checked for further connections to existing data. If nothing is found, data will be deleted. A deletion command for the media archive can be generated from the data fusion application if step three of deletion processing finds that an “Archive ID” can be deleted. The deletion process is transactional.

Discussion and evaluation against PbD

Previously we outlined requirements for managing fused data and data deletion in SenseCare. Deletion process has to be executed as soon as purpose (retention period) vanishes. Data deletion has to result in accurate data set – having purpose. We think that in the light of the coming sensitive data regulations, the management of retention period is a critical issue in SenseCare. Currently there is one global retention period configured within the system. As shown, the legal situation comprises different retention periods for types of medical data. To meet this, the ontology of SenseCare would need to store retention times connected to types. With changing legal requirements, update of retention periods is a valid use case and introduces the demand of retention period calculation. Figure 6 outlines the aspect that retention is determined as triple of data type, provenance and applicable time range and their connections.

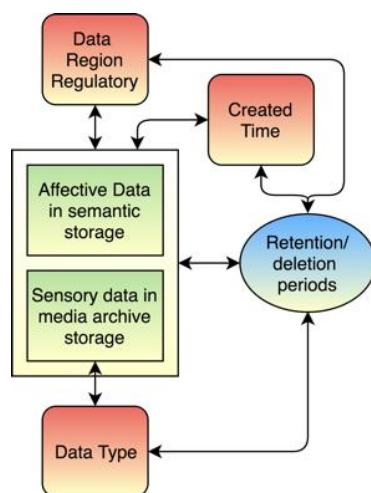


Figure 6: Time Range Triple Relational Model

Data accuracy has two cases: After deletion, remaining data has to be correct. This introduces the demand of transaction safety: Deletion process has to be persisted only in success cases. There are two storage systems involved, transactions have to be shared among both systems. Semantic storage using RDF4J (Eclipse, n.D.), as one RDF storage framework, is able to create shared transactions. The underlying raw archival system is a flat file system at the moment and offers no rollback functionality. Introducing this is matter of future research in SenseCare. SenseCare has a data management dimension on AC data storage. Additional correlations between data can occur by the Shared Care Giving Scenario spanning a dependency graph among correlations of data. Prerequisite of deletion is, that no transitive dependency of data within the deleted graph is still in retention period, otherwise links between data objects will be broken and data will remain inaccurate within the SenseCare System. Current implementation does not respect those data objects. Introducing this aspect is topic for further research. Deletion in raw file system however underlies the retention period process.

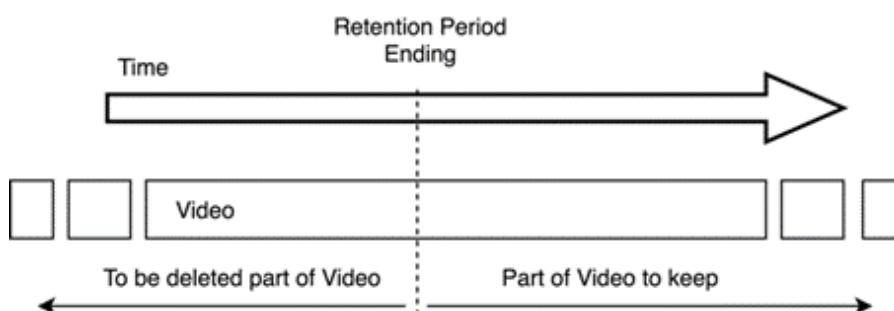


Figure 7; Time context through large sensory streams

This storage archives raw sensory data stream files. Those have an underlying time context as well. With passing time, retention time vanishes continuously throughout those video files (see Fig. 7) if no other dependencies exist on this entity. According to PbD principles, entities should manage security and privacy by themselves as far as possible. Implementing access control respecting continuous retention times is a tradeoff between effort and security. A special context for entities capable of handling continuous deletion by retention period configuration is subject for further research within the SenseCare project since SenseCare deals with long sensory data spanning among

possible large time periods. The Current implementation of deletion process is triggered on demand. Implementing a service triggering deletion is matter of future research as well.

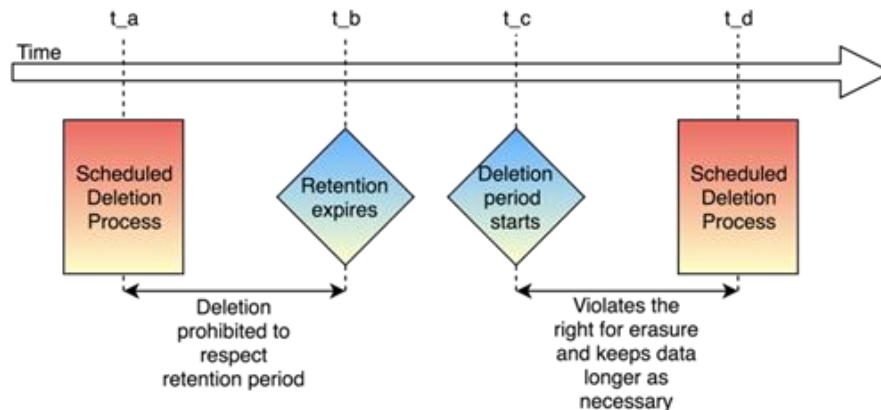


Figure 8: Macro view on central deletion trigger time context

There are several possibilities existing for implementing triggers. First option, a central scheduled component, adds an additional phase on top of the presented algorithm. As first step, scanning for deletion candidates is necessary. After candidate gathering, first phase starts the above presented algorithm for each candidate. The second solution is a deletion trigger within the entity itself. Entities check their deletion demand by themselves based on dependency preconditions as well as on retention time. Comparing both solutions bring pros and cons for both options. While first option is easily controllable and manageable, scanning through large amounts of data might be a complex and long running task. The scanning algorithm complexity is directly depending on the size of the underlying data set. Additionally, data can expire between two schedule cycles. This violates the purpose and disclosure limitations on PbD. Having entities managing their own retention and deletion, dependency checks are more complicated since there need to be a communication between entities. Additionally, it is hard to control deletion runs on entities as well as auditing these processes is hard since they happen in a distributed way. Distributed deletion triggers introduce possibly the need of transaction synchronization between two or more deletion triggers of related data. Conceptual evaluation of both solutions is matter of future research.

Conclusion

Within this paper we outlined the current approach on semantic data fusion and the underlying data management concept. We analysed this concept with regards to legal frameworks and showed that there is room for future research within the accuracy principle of PbD including shared transaction safety for raw file storage, the recognition process for to be deleted data and the deletion processes on sensory data stored on raw file storage. The current existing configuration of retention periods does not comply to legal requirements as well as the trigger for executing deletion process is matter of future research as well.

Acknowledgement

This publication has been produced in the context of the SenseCare project. This project has received funding from the European Union's H2020 Programme under grant agreement No 690862. This paper reflects the author's view, the EC is not responsible for any use that might made of the information it contains.

References

- Wimo, A. and Jönsson, L. (2009). “*Alzheimer Europe - Research - European Collaboration on Dementia - Cost of dementia - Prognosis to 2030*”. <http://www.alzheimer-europe.org/Research/European-Collaboration-on-Dementia/Cost-of-dementia/Prognosis-to-2030> (Accessed 23 Apr. 2015).
- Wimo, A. and Prince, M. (2010a). “*World Alzheimer Report 2010 - the global Economic Impact of Dementia*”. World Alzheimer Report. https://www.alz.org/documents/national/World_Alzheimer_Report_2010.pdf, (Accessed 17 Apr. 2015).
- Wimo, A. and Prince, M. (2010b). “*World Alzheimer Report 2010 - the global Economic Impact of Dementia*”. World Alzheimer Report. https://www.alz.org/documents/national/World_Alzheimer_Report_2010.pdf, (Accessed 17 Apr. 2015).
- Kowohl, U., Engel, F., Hemmje, M., Walsh, M. and Keary, A. (2016).” Towards lifecycle management of clinical records in health care environments”. In: *Proceedings of the CERC Conference* Cork.
- Engel F., Bond R., Keary A., Mulvenna M., Walsh P., Zheng H., Wang H., Hemmje M., Kowohl U., 2016, SenseCare: Towards an Experimental Platform for Home-Based, Visualisation of Emotional States of People with Dementia, Advanced Visual Interfaces Supporting Big Data Application
- Schaar, P. (2010). “Privacy by Design”. In: *Identity in the Information Society*, 3(2), pp.267-274.
- De Hert, P. and Papakonstantinou, V. (2012). “The proposed data protection Regulation replacing Directive 95/46/EC: A sound system for the protection of individuals”. In: *Computer Law & Security Review*, 28(2), pp.130-142.
- Smith, B., Ashburner, M. et al. (2007). “The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration”. In: *Nature Biotechnology*, pp.1251-1255.
- Noy N., Shah N., Whetzel P., Dai B., Dorf M., Griffith N., Jonquet C., Rubin D., Storey M., Chute C., (2009). “BioPortal: Ontologies and integrated data resources at the click of a mouse”, In: *Nucleic Acids Research, Volume 37*, pp. 170 – 173.
- Arndt R., Troncy R., Staab S., Hardmann L., Vacura M. (2007). Designing Well Founded Multimedia Ontology for the Web, Heidelberg, COMM, ISBN: 978-3-540-76298-0.
- Eclipse (n.D.). Eclipse Incubation, RDF4J, <http://rdf4j.org/>

Chapter 3

Security

Elicitation of security requirements for migration of OCR software to the cloud

S. Ringmann, H. Langweg
HTWG Konstanz
University of Applied Sciences, Germany
e-mail: sandra.d.ringmann@outlook.de

Keywords: security requirements, stakeholder interests, cloud

1. Introduction

In logistics, it is essential to identify the correct destination of a posted item. Routing discrete physical goods (as opposed to routing a continuous flow, e.g., through pipelines) in most cases requires manual processing upon sending and delivery. Hence, items are usually identified by the human-readable addresses of receiver and sender. This information is affixed to an item in handwriting, typewriting, as a bar code, a 2D-bar code, on a form, an address label, in a specific location on the packaging, or wherever and however the sender decides to put it.

At a routing node in the distribution network, the address is scanned, parsed, and used to group items with the same destination for further transport, e.g., by country, state, region, ZIP code, neighborhood, street. The sorting equipment consists of a set of cameras to scan an item. Highly specialized optical character recognition (OCR) software determines the location of an address on an item, considers the position and orientation of an item, and extracts a bitmap of the address information. That bitmap is then parsed for characters and bar codes. The result of the process is a structured dataset containing the address information, e.g., formatted as CSV or XML.

Scanning and parsing an address can in most cases be completed within 300-800 ms. The use of the equipment follows an irregular pattern. Items arrive at a node several times a day and must be ready for further transportation at fixed times. There are peaks like a higher amount of items at the beginning of an invoicing period, Christmas shopping season, or a lower number of shipments when people are on vacation. Long-term trends overlay yearly delivery patterns, e.g., a declining amount of personal correspondence by mail, a replacement of paper invoices by digital invoicing, a variation in direct mail advertising, improvements in digital-to-print services, an increasing market share of goods sent home instead of being picked up at a store. For logistics companies, it becomes harder to plan for the expected volume that needs to be handled.

Processing speed is determined by the speed of conveyor belts, speed of cameras, and speed of OCR. The bottleneck is the parsing of the address data captured by cameras. Automated parsing of addresses requires significant computing power to be completed fast and reliably. Companies cannot usually reduce address recognition quality to achieve higher throughput in times of high volume. Hence, computing capacity needs to be available for peak periods. Installing this capacity on site leads to underused expensive equipment. At the same time, deployment of software updates is tedious when each machine at each site for each customer potentially runs a different version of the OCR software. Therefore, there is an incentive to move from fixed peak capacity processing to outsourcing computation to shared computing resources in the cloud. While the economic case is straightforward, the technical challenges and its implications for the security of reliably processing sensitive personal data in the cloud are treated in this paper.

We investigate three different deployment models for OCR software: cloud (public or private hosting based on cloud technology), on-premise installation for legacy sites, and hybrid (combination of public/private cloud and on-premise). The goal is to be able to execute OCR software operating on the bitmaps captured at a network node. The deployment model has implications for costs, flexibility, scalability, recognition quality, and speed.

2. Related work

The framework for security requirements engineering developed by (Riaz et al., 2016) enables a security analyst to assess the security of an asset from multiple dimensions. The framework provides security goal patterns which are a combination of a security action (i.e., prevent, detect, respond to a breach), a security property (e.g., confidentiality, integrity, availability, ID and authentication, accountability, and privacy), an asset, an actor, and action types (i.e., create, read, update, store, delete, and transfer). Criteria for cloud security and privacy with an emphasis on auditing are described by (Rizvi et al., 2015). A security metric tool is presented which generates a security index to evaluate a cloud provider's achieved security level. Threat modelling of cloud services and a methodology are detailed in (Kazim and Evans, 2016). It considers threats from various attackers to authentication, data processing, and data storage. A taxonomy of attacks in virtualized systems is provided by (Modi et al., 2013b; Sgandurra and Lupu, 2016), looking at possible attack paths, existing attacks and solutions. The focus is on hardware, virtualization, OS, and application layer.

The research by (Modi et al., 2013a) identifies security issues in the different layers of a cloud infrastructure, namely, on the application level, network level, data storage level, virtualization level, authentication and access control level, trust level, compliance, audit and regulations level. A classification model of security issues across the different cloud layers is presented by (Hussain et al., 2016). Attack types are associated with a risk level. The risk levels are further associated with security requirements, for example, data-encryption, multi-tenancy, data privacy, authentication, and authorization. For the various vulnerabilities, threats, attacks, and existing solutions are surveyed. An approach proposed by (Anand et al., 2016) uses Microsoft's STRIDE and DREAD model for threat assessment in cloud environments.

3. Method

Security requirements should be based on a valid interest of one of the stakeholders. Stakeholders in this specific context include: vendor, operator, supplier (cloud service provider – CSP), sender, receiver, society, government, and law enforcement. Hence, the stakeholders' interests are determined referring to the security properties confidentiality, integrity, and availability of the components that have interfaces to the software product. The scanner sends an image of the item to the software. The software then uses its OCR algorithms to identify the text on the image. Together with the data from the dictionary, the software produces an output of the identified personal data that contains instructions for the machine where to route the item.

Information was acquired through expert interviews and considering the requirements of logistics customers in the field of mail, parcel, and airport. The results are displayed in **Fehler! Verweisquelle konnte nicht gefunden werden.** Thus, threat scenarios are identified by applying STRIDE (Microsoft, 2017) and matching threats to the various deployment models. A further risk analysis of the threat scenarios can be elaborated, for example, by using DREAD. Based on these findings, security objectives that counter the threats can be formulated.

Table 1: Stakeholder interests regarding confidentiality (C), integrity (I), and availability (A)

Stakeholder	Scans of personal data			Data dictionary (DB)			Software/Algorithms			Identified personal data		
	C	I	A	C	I	A	C	I	A	C	I	A
Vendor												
Operator												
Supplier (CSP)												
Sender												
Receiver												
Law enforcement												
Society												

4. Future work

Extending the requirements mining based on laws, further work should address applicable rules and regulations covering all relevant jurisdictions. The impact of privacy and data protection needs to be evaluated as regards choice of deployment models, software architecture, and system operation. It would be beneficial to design the system architecture in a way that lends itself to fast, inexpensive, and potentially automated reconfiguration to ensure compliance with changing laws and regulations. Immediate considerations could be given to cost and performance aspects of different software architectures and deployment scenarios. What are the costs of necessary security controls and when do costs outweigh benefits of flexibility gained in the cloud?

As part of our future research, we also envision an investigation whether it makes more sense to analyze an existing configuration for potential security problems or if it is better to automatically create and tear down deployments to ensure they always conform to a desired configuration.

5. References

- Anand, P., Ryoo, J., Kim, H. and Kim, E. (2016). Threat Assessment in the Cloud Environment: A Quantitative Approach for Security Pattern Selection, *Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication*. New York, NY, USA, ACM, 5:1- 5:8.
- Hussain, S. A., Fatima, M., Saeed, A., Raza, I. and Shahzad, R. K. (2016). Multilevel classification of security concerns in cloud computing, *Applied Computing and Informatics*, vol. 13, no. 1, pp. 57–65 [Online]. DOI: 10.1016/j.aci.2016.03.001.
- Kazim, M. and Evans, D. (2016). Threat Modeling for Services in Cloud, *Service-Oriented System Engineering (SOSE), 2016 IEEE Symposium on*, pp. 66–72.
- Microsoft (2017). *The STRIDE Threat Model* [Online]. Available at [https://msdn.microsoft.com/en-us/library/ee823878\(v=cs.20\).aspx](https://msdn.microsoft.com/en-us/library/ee823878(v=cs.20).aspx) (Accessed 20 June 2017).
- Modi, C., Patel, D., Borisaniya, B., Patel, A. and Rajarajan, M. (2013a). A survey on security issues and solutions at different layers of Cloud computing, *The Journal of Supercomputing*, vol. 63, no. 2, pp. 561–592.
- Modi, C., Patel, D., Borisaniya, B., Patel, A. and Rajarajan, M. (2013b). A survey on security issues and solutions at different layers of Cloud computing, *The Journal of Supercomputing*, vol. 63, no. 2, pp. 561–592.
- Riaz, M., Stallings, J., Singh, M. P., Slankas, J. and Williams, L. (2016). DIGS: A Framework for Discovering Goals for Security Requirements Engineering, *Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*. New York, NY, USA, ACM, 35:1-35:10.
- Rizvi, S., Ryoo, J., Kissell, J. and Aiken, B. (2015). A Stakeholder-oriented Assessment Index for Cloud Security Auditing, *Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication*. New York, NY, USA, ACM, 55:1-55:7.
- Sgandurra, D. and Lupu, E. (2016). Evolution of Attacks, Threat Models, and Solutions for Virtualized Systems, *ACM Comput. Surv.*, vol. 48, no. 3, 46:1-46:38 [Online]. DOI: 10.1145/2856126.

Understanding Opportunities and Risks in Online Gaming Environments: An Ethnographic Approach

Dr B. Sanders
Faculty of Business, Law and Sport
University of Winchester
e-mail: ben.sanders@winchester.ac.uk

Keywords: Risk, Online Gaming, Internet Addiction

1. Introduction

Massively Multiplayer Online Role Playing Games (MMORPGs) are a popular form of entertainment used by millions of gamers worldwide. The popularity of MMORPGs and their exponential adoption in the community has raised concerns in the public and scientific communities, and a growing number of studies have been undertaken to examine the positive and negative impact on gamers' daily life and well-being (Billieux et al. 2014; Sanders and Marchang, 2016).

On the one hand, several positive outcomes were identified, such as the development of online social capital (e.g. new friendships, emotional support, civic engagement, problem solving and expression of identity). Preliminary data suggested that some cognitive functions (attentional, perceptual and visuomotor skills) could be enhanced by the practice of certain types of action-based video games (Green and Bavelier, 2008). Recent evidence asserted that an elevated commitment to online games (reflected by in-game rankings and achievements) does not necessarily interfere with daily life activities or social relationships (Billieux et al. 2013) and that most gamers do not play to fulfil unsatisfied basic psychological needs (Herodotou et al, 2014).

On the other hand, growing evidence suggests that the use of online games can become problematic and cause negative impacts on daily living (e.g. addiction, online bullying) (Starcevic and Aboujaoude, 2017; Kuss and Griffiths, 2012) and that online gaming overuse is often the main presenting issue in treatment-seeking self-identified problematic Internet users (Thorens et al, 2014). In recent years, many studies attempted to establish the prevalence of problematic online gaming and to elucidate the factors involved in its etiology (Kuss et al. 2017). In recent years, research in online gaming has predominately focused on risks associated with addiction (Markey and Ferguson, 2017; Griffiths et al. 2014; Ko, 2014). In 2013, the Diagnostic Statistical Manual for Mental Disorders (5th ed.; APA DSM-5) included "Internet Gaming Disorder" into Section 3, categorising it as an emerging measure warranting further research.

Preliminary research by the author (Sanders et al. 2011; 2010) classified 18 opportunities and 13 risks in MMORPGs. The findings were underpinned by two quantitative studies (Sanders et al. 2010) that used established, reliable and valid constructs. Player behaviour was measured using Bolino and Turnley's (2003) Impression Management Scale together with Deci and Ryan's (2000) Self Determination Theory constructs. Addiction was measured using a 22-item Game Addiction Scale (GAS) developed and validated by Lemmens et al. (2009). In addition to the quantitative method outlined above, participants were invited to provide further qualitative data on their personal experiences in online gaming environments, thereby gathering a subjective in-depth insight into their feelings, experiences and motivations of behaviour, which could not necessarily be captured by an objective quantitative approach.

2. Taxonomy of Opportunities and Risks in MMORPGs

A novel taxonomy of opportunities and risks incorporated the evidence from studies by Sanders et al. (2010; 2011), Griffiths and Hunt (2009) and Yee (2006). The transactional opportunities and risks were refined and structured into the taxonomy shown in Figure 1.

The horizontal axis of the taxonomy reflected three modes of online communication: one-to-many adult-to-child; and peer-to-peer. The vertical axis was divided into two categories: opportunities and risks, with four sub categories in each. Opportunities category encapsulated: Education learning and digital literacy, participation and civic engagement, creativity and self-expression and identity and social connection. Risks category incorporated: commercial, aggressive, sexual, values.

		One -> Many	Adult -> Child	Peer-to-Peer	
		Content: Vulnerable Individual as Recipient of Mass Distributed Content	Contact: Vulnerable Individual as Participant in an Interactive Situation Predominately Driven by Adults	Conduct: Vulnerable Individual as Actor in an Interaction in which He/She May be Initiator	Positive Consequences
	Education learning and digital literacy	Presents challenges, raises confidence, develop unique skill sets	Contact with other players who share one's interest	Peer-to-peer learning and team work	Develop skills & problem solving abilities
	Participation and civic engagement	Widening participation	Exchange of knowledge and ideas amongst players	Concrete forms of civic engagement	Participation & sharing of knowledge
	Creativity and self-expression	Diversity of roles, character types and guild memberships	Being inspired to participate in a socially equal environment	User generated add-ons, plug-ins and enhancements	Creation, engagement & exploration
	Identity and social connection	Information to enhance game play experiences	Networking with like-minded individuals	Expression and experimentation of identity	Identity, self-expression & social networking
					Negative Consequences
	Commercial	Addictive advertising, rewards and incentives	Tracking/ harvesting personal information	Account hijacking, virtual mugging & social engineering	Virtual Crime, financial & privacy loss
	Aggressive	Desensitising to violent/gruesome/ harmful content	Being bullied, harassed or stalked (including 'griefing')	Bullying or harassing another	Desensitisation, increased aggressiveness & harm
	Sexual	Desensitising to sexual/intimate scenes	Meeting strangers, being groomed	Erotic role play. Exposure to sexual conversation	Desensitisation and sexual harm
	Values, Morals, Norms & Lifestyle	Addiction and self destructive behaviour	Unwelcome persuasion & justification of obsessive and addictive behaviour	Social dependency to MMORPG's and real life social isolation	Time, Addiction & Dependency

Table 1: Taxonomy of Opportunities and Risks in MMORPGs

3. Ethnographic Study of Opportunities and Risks

The taxonomy presented in Section 2 was based upon user's perceptions of opportunities and risks juxtaposed with their personal experiences. The evidence base consisted of data from survey and focus group discussions together with conceptual clarity gained from a critical examination of current literature.

This paper presents the findings of an ethnographic study, which aimed to provide further validation of the taxonomy through participant observation and interaction within a naturalistic setting. The study built on previous findings to further understand the contextual influences that affect the probability of an individual experiencing risk.

Ethnographic research typically involves observing target users in their natural real-world setting rather than in the artificial environment of a lab or focus group. The aim is to gather insight into how people live; what they do; how they use things; or what they need in their everyday or professional lives (Bryman, 2013). Chen (2014) asserts, online ethnography differs from applying this methodology in real life settings as motions and interactions are all focused within the one domain.

A total of 20 MMORPG gamers were recruited through snowball sampling. Participants were regular gamers who agreed to record their video gaming sessions up to a maximum of 15 hours. Of the total, 67% were males, with 78% of the sample lying in the age group category of 18-21 years and the remainder 22% in the category of 22-35 years.

287 hours of recorded game play was coded and analysed using NVivo. The taxonomy of opportunities and risks was used as a lens in the coding process. Interaction with players facilitated insights into player behaviour in a way that would not have been possible through the previously discussed studies.

4. Contextual Influences Model

Based on the foregoing evidence, the Contextual Influences Model is presented (Figure 1) and builds upon the work of Livingstone and Haddon (2009). Establishing these contextual factors outlines the scope of the risk (Bryman, 2012), and enables the accurate assessment of the transactional opportunities and risks encountered.

The Contextual Influences Model provides an overarching overview of how the contextual influences affect the probability of an individual encountering the transactional opportunities and risks.

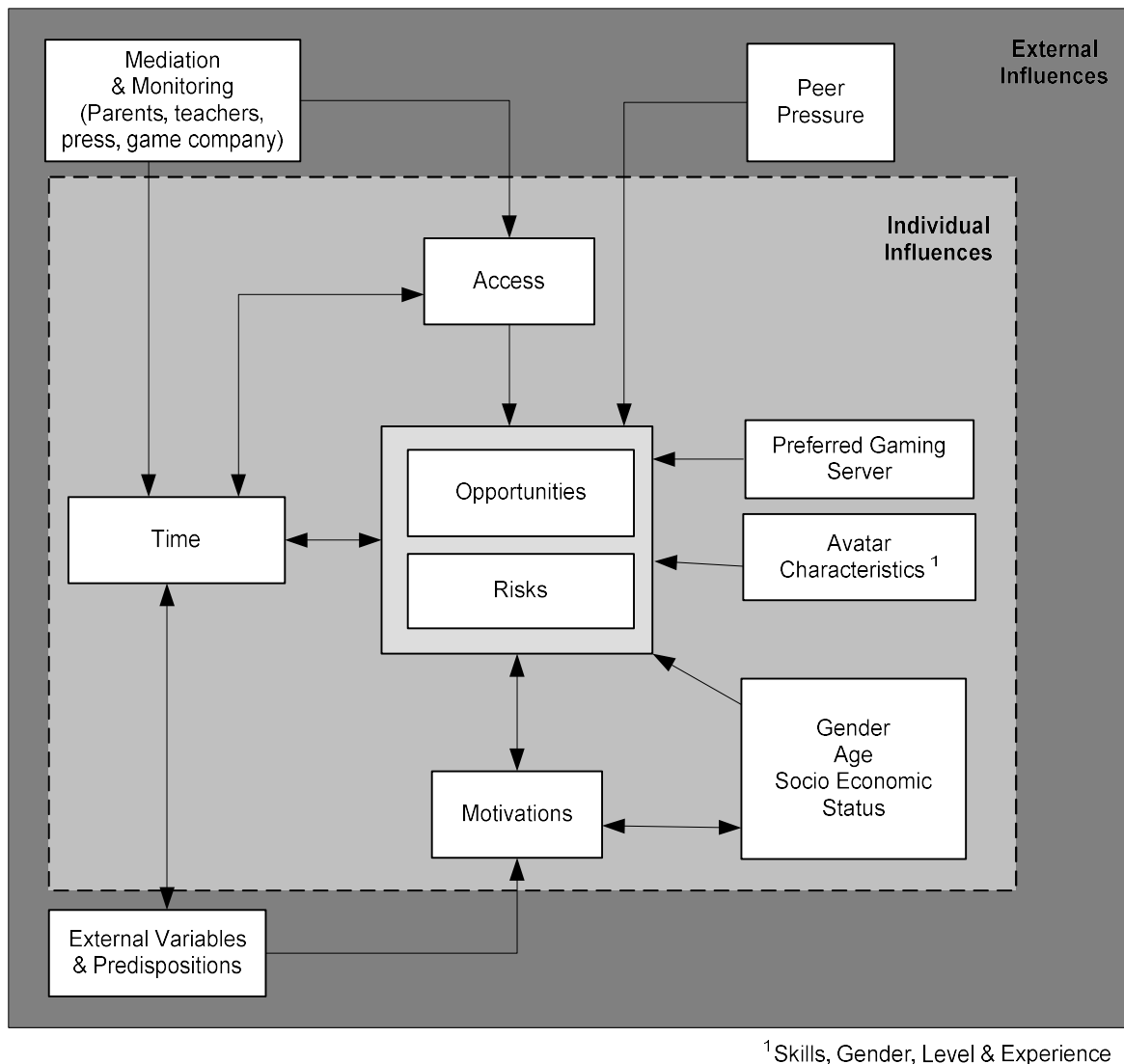


Figure 1: Contextual Influences Model

5. Conclusions

This study highlights the different types of opportunities and risks in MMORPGs. Validation is achieved using empirical data from a range of reliable and credible sources. The Contextual Influences Model makes a significant contribution to knowledge in online games through the identification of contextual influences on the probability of an individual encountering one or more of opportunities and risks presented in the taxonomy.

The models presented within this paper provide a sound basis on which to further explore the impact of encountering risks in online games. Future work will further examine the political, economic, sociological economic and environmental impact of online gaming.

6. References

- Billieux, J., Thorens, G., Khazaal, Y., Zullino, D., Achab, S., Van der Linden, M. (2015). Problematic involvement in online games: A cluster analytic approach. *Computers in Human Behavior* 43, 242–250.
- Billieux, J., Van der Linden, M., Achab, S., Khazaal, Y., Paraskevopoulos, L., Zullino, D., et al. (2013). Why do you play World of Warcraft? An in-depth exploration of self-reported motivations to play online and in-game behaviours in the virtual world of Azeroth. *Computers in Human Behavior*, 29, 103–109.
- Bolino M and Turnley W (2003) More Than One Way to Make an Impression: Exploring Profiles of Impression Management. *Journal of Management* April 2003. 29 (2): 141 - 160.
- Bryman. A. (2012) *Social Research Methods*. Oxford University Press. 4th Ed. ISBN 9780199588053
- Chen, V. (2014) *Virtual Ethnography*. Online: <http://www.ischool.drexel.edu/faculty/jrode/chen.pdf>
- Deci E L and Ryan R M (2000) The “what” and “why” of goal pursuits: Human needs and the self-determination of behaviour. *Psychological Inquiry*, 11: 227-268.
- Green, C. S., & Bavelier, D. (2008). Exercising your brain: A review of human brain plasticity and training-induced learning. *Psychology and Aging*, 23, 692–701.
- Herodotou, C., Kambouri, M., & Winters, N. (2014). Dispelling the myth of the socioemotionally dissatisfied gamer. *Computers in Human Behavior*, 32, 23–31.
- Ko, C. H., Yen, J. Y., Chen, S. H., Wang, P. V., Chen, C. S., & Yen, C. F. (2014). Evaluation of the diagnostic criteria for Internet gaming disorder in the DSM-5 among young adults in Taiwan. *Journal of Psychiatric Research*, 53, 103–110.
- Kuss, D. J., Griffiths, M. D., Karila, L., & Billieux, J. (2014). Internet addiction: A systematic review of epidemiological research for the last decade. *Current Pharmaceutical Design*, 20, 4026–4052.
- Kuss, D., Griffiths, M., Pontes, H. (2017) DSM-5 diagnosis of Internet Gaming Disorder: Some ways forward in overcoming issues and concerns in the gaming studies field. Response to the commentaries. *Journal of Behavioural Addictions*. 6 (2).
- Lemmens J S, Valkenburg P M and Peter J (2009) Development and validation of a game addiction scale for adolescents. *Media Psychology*, 12: 77-95.
- Livingstone, S. and Haddon, L. (2009). *Kids Online: Opportunities and Risks for Children*. Policy Press, ISBN: 1847424384
- Markey, P., & Ferguson, C. (2017) *Moral Combat: Why the War on Violent Video Games is Wrong*. BenBella, ISBN 978-1942952985
- Sanders B. & Marchang J. (2016). A Novel Taxonomy of Opportunities and Risks in Massively Multiplayer Role Playing Games. *Proceedings of the International ACM conference on management of Digital Ecosystems*. ACM Digital Library. ISBN 978-1-4503-4267-4
- Sanders B, Chen V, Zahra D, Dowland PS, Atkinson S, Papadaki M, Furnell SM (2010) Online Addiction: Privacy Risks in Online Gaming Environments. *Proceedings of the International Conference on Management of Emergent Digital EcoSystems (MEDES)*, October 26-29, Bangkok, Thailand, 2010
- Sanders B, Dowland PS, Atkinson S, Zahra D, Furnell SM, Papadaki M. (2010). Emerging risks in massively multiplayer online role playing games. *EU Kids Online Conference*, Friday 23 September, New Academic Building LSE, 2011.
- Starcevic, V., Aboujaoude, E. (2017). Internet addiction: Reappraisal of an increasingly inadequate concept. *CNS Spectrums*, 22(1), 7-13.
- Thorens, G., Achab, S., Billieux, J., Khazaal, Y., Khan, R., Pivin, E., et al. (2014). Characteristics and treatment response of self-identified problematic Internet users in a behavioral addiction outpatient clinic. *Journal of Behavioral Addictions*, 3, 78–81.
- Yee N (2006) The Psychology of MMORPG's: Emotional Investment, Motivations, Relationship Formation and Problematic Usage, In R. Schroeder & A. Axelsson (Eds.), *Avatars at Work and Play: Collaboration and Interaction in Shared Virtual Environments*, London: Springer-Verlag: 187-207.

IT Security Policy Enforcement in IoT Gateways

Ayhan Soner Koca
Faculty of Computer Science
Furtwangen University (HFU), Germany
e-mail: ayhan.soner.koca@hs-furtwangen.de

Keywords: Industrie 4.0, IoT gateway, policy enforcement

Introduction

Being influenced by the fourth industrial revolution the classic manufacturing process has to reorganize itself to keep up with the new reality of Industry 4.0. The consequently growing interconnection of machines and manufacturing plants with emerging data traffic requires vital integration of IoT and cloud computing. Companies now have to face various tasks which not only but particularly include security related challenges. The systems have to be capable of communicating without security issues and moreover controlling cross-company communications. Parts of the systems in use show limitations of memory capacity, computing power or energy consumption making them constrained. This is where IoT gateways being key element of Internet of Things (IoT) come into the picture. The usage of security policy enforcement on IoT gateways could multiply the communication possibilities with respect to the requirement of a sustainable Industrie 4.0. In this paper, a solution to enforcing security policy rules that addresses the highlighted above shall be described. This solution is grounded on a rule-based anomaly detection on a chosen IoT gateway. A worked-out concept with agreed regulations including entitlement rules will provide the basis for the implementation. Use cases offer the adaptation for the practically orientated engineering.

Related Work

Previous work in the area by Neisse et al. (2014) covered the enforcement of security policy rules at the MQ Telemetry Transport (MQTT) protocol layer. Their solution is part of the Model-based Security Toolkit (SecKit) which is implemented as an extension to the open source MQTT broker Mosquitto. Their work showed that the enforcement of complex security policies is possible with only marginal performance loss to the normal operation of the broker. However, this is constrained to the MQTT protocol only.

Li et al. (2016) present an IoT architecture based on Software-defined Networking (SDN) that supports deploying and adjusting dynamic security policies at fine-grained level for IoT services. In their approach, the central authorization and policy management are integrated in the SDN controllers. Since the IoT application developers build IoT services at the service layer by programming the SDN controllers it is possible to deploy security requirements in the network. Their test results have shown that the proposed architecture is feasible for providing secure IoT services and satisfy the needs from the constrained and heterogeneous IoT devices. The overhead of SDN with its additional controllers may not be suitable for all.

Another interesting approach is described by Sicari et al. (2017). The paper proposes a flexible policy enforcement framework for IoT that adopts a cross-domain policy specification language based on XML. The framework includes a Policy Enforcement Point (PEP), a Policy Decision Point (PDP), a Policy Administration Point (PAP) and a policies store, where the PEP intercepts the requests from users and makes a decision request to the PDP. To this end, the PDP queries the policies store. The policies in the policies store are administered through the PAP, which allows the runtime change of

policies. They demonstrate the feasibility of the proposed approach by running an example based on a smart health application. A drawback of such a framework is the extra effort that is needed to implement an additional layer which is not suitable for all infrastructures. The approach of this paper now describes a more general solution by using an IoT gateway.

Problem Statement

In the manufacturing process of Industrie 4.0 IoT gateways offer various possibilities but also face security related risks. Following use cases give an idea of the importance of having a security and privacy enhanced IoT gateway. Looking at the production side customers could benefit of a customized order tracking option. Subject to the condition that data can be separated accordingly. If a machine has to be repaired due to an error and requires a diagnosis, it shall be ensured that the technician only has access to the needed data but not to e.g. customer specific information. This customer specific information should be protected at all times. IT admins will have all tools and access needed to run a smooth production process without knowing any privacy related details of the customers that are not relevant to them. Lastly to mention are software updates which should not pose a risk. To realize the above outlined use cases, it must be ensured that security policies can be implemented allowing detailed separation and monitoring of data.

Methodology

The research methodology is to develop a rule-based anomaly detection on a selected IoT gateway, which checks the incoming data stream against defined rules. As a result, the data is either sent to the cloud, discarded, sent to the log file or written into a database to analyze them later. The initial step to develop a prototype was to find a suitable IoT gateway that meets the set requirements. The following evaluation criteria have been identified: device management (EC1), connectivity (EC2), development environment (EC3), IT infrastructure (EC4), interfaces (APIs) (EC5), security (EC6). In total eight different gateways were examined and evaluated based on a decision matrix (Table 1). Since not every criterion is equally relevant to this work, a weighting factor $W = \{1, 2, 3\}$ was defined according to the importance. For example, the access to the IoT platform via interfaces (EC5) is not as important as the support of different communication protocols (EC2). The IoT platforms receive point scores $P = \{1, 2, 3\}$ for each evaluation criterion, which are multiplied by the corresponding weighting factors.

	EC1	EC2	EC3	EC4	EC5	EC6	
Factor W	3	2	2	2	1	2	Result
Kaa	2	2	3	3	3	3	31
EclipseKura	3	3	2	2	3	3	32
OpenIoT	1	2	2	2	3	3	24
DeviceHive	2	2	3	3	3	2	29
SiteWhere	2	3	2	3	3	2	29
IoTivity	2	2	3	3	3	2	29
Lelylan	2	3	3	3	3	2	31
WSO2	2	3	2	3	3	3	31

Table 1: Decision matrix for selecting an IoT platform

Eclipse Kura is a Java OSGi based open-source platform for building IoT gateways. It serves as a container for Machine-to-Machine (M2M) applications running in service gateways and supports a large set of network protocols to communicate with lower-layer devices (Eclipse.org, 2017). The rule-based anomaly detection is implemented using Kura Wires. A new wire component *AnomalyDetection* has been developed and added. The wire component *AnomalyDetection* was implemented according to the adapter pattern (Figure 1). The simple adapter *SimpleAnomalyFilterAdapter* implements only the method *filter(List<WireRecord> records)*. In this, simple rules can be implemented as conditions. The second *VisualAnomalyFilterAdapter* adapter also implements the *Rule Execution API* from Visual Rules (Innovations, B, 2017).

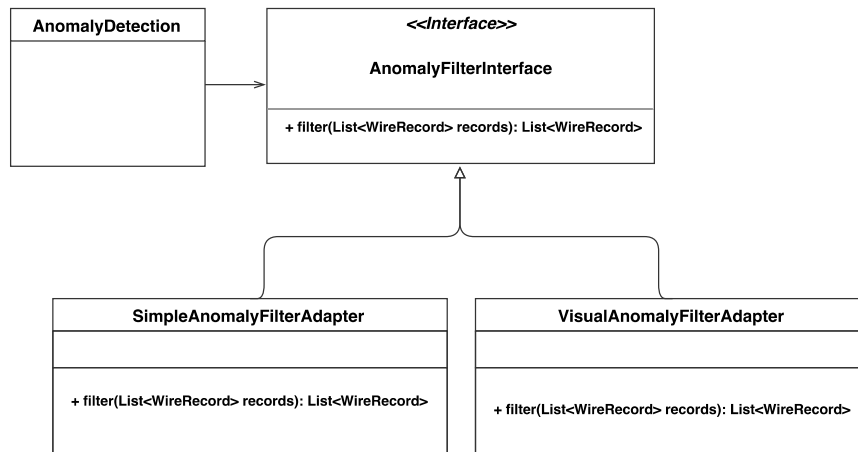


Figure 1: AnomalyDetection Wire Component

Kura itself is installed on the development board BeagleBone Black. At the current stage of development two producers are sending JSON files to the gateway where the anomaly detection module either allows to send the data to the AWS IoT cloud (Amazon Web Services, Inc., 2017) or blocks it and instead sends the data to the log file.

Conclusion

To determine the requirements of a privacy enhanced IoT gateway the above briefly mentioned use cases needed detailed definition at first. This formed the basis of the first steps of this work followed by the deduction of the rules which were implemented subsequently on the chosen IoT gateway Eclipse Kura. In the coming steps an evaluation will be carried out.

References

Amazon Web Services, Inc. (2017). AWS IoT – Amazon Web Services. [online] Available at <https://aws.amazon.com/de/iot/>, [Accessed 24. 06. 2017].

Eclipse.org. (2017). “Eclipse Kura™ - Open Source framework for IoT”, [online] Available at <http://www.eclipse.org/kura/>, [Accessed 24. 06. 2017].

Innovations, B. (2017). Visual Rules, [online] Available at <https://www.bosch-si.com/de/bpm-und-brm/visual-rules/business-rules-management.html>, [Accessed 24. 06. 2017].

- Li, Y., Björck, F., Xue, H., (2016). IoT Architecture Enabling Dynamic Security Policies. In: *Proceedings of the 4th International Conference on Information and Network Security - ICINS '16*. ACM Press, New York, NY, USA, pp. 50–54.
- Neisse, R., Steri, G., Baldini, G., (2014). Enforcement of security policy rules for the Internet of Things. In: *Proceedings of the 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. IEEE, 2014; pp. 165–172.
- Sicari, S., Rizzardi, A., Grieco, L. A., Piro, G. and Coen-Porisini, A. (2017). A Policy Enforcement Framework for Internet of Things Applications in the Smart Health, Smart Health.

Insecurity Refactoring As a Novel Method to Improve Manual Code Inspection Skills

Felix Schuckert
Department of Computer Science
HTWG Konstanz, Germany
e-mail: felix.schuckert@htwg-konstanz.de

Keywords: Vulnerabilities, Refactoring, Static Code Analysis

1. Introduction

Software development is a complex process. Currently, there is no fast, inexpensive automated way to prevent the occurrence of security issues. There are forms of effective formal methods to ensure correctness and security, but they are not efficient as of today. Static code analysers try to detect security issues in source code. As stated in the paper by Goseva-Popstojanova and Perhinschi (2015), the overall highest detection rate of known security vulnerabilities is 59%. The false positive rate still remains a problem because developers have to review each report. If developers have to fix possible vulnerabilities, they need software security skills - first of all to distinguish between false and true positive reports. The fixing process requires software security skills as well. Developers who try to fix vulnerabilities could create new security issues or just hide the existing one from static code analysers. Accordingly, the review process that is involved by using static code analysers requires software security skills.

A Software security development life cycle involves software security in the software development process. It includes different steps to prevent issues already in the development phase. One important part is that all developers require security training to prevent security issues. It is recommended to refresh the skills by participating in training events frequently (Howard et. al., 2006). A typical approach to learn software security skills is by using projects with existing vulnerabilities. These can be exploited to learn the effect and impact of such vulnerabilities. The vulnerabilities can be fixed by the developers and that will teach the prevention of such issues. Some projects exist which contain security vulnerabilities based on the most prevalent security issues. If developers require training frequently, the exercises should be different each time. Having a simple way of generating different exercises based on common security issues could help to teach developers software security skills and raise awareness.

2. Insecurity Refactoring

Insecurity refactoring uses refactoring methods to generate software vulnerabilities in source code. Opdyke coined the term Refactoring in his PhD thesis (Opdyke, 1992). Refactoring is a common approach in software development. For example, it is used to add new methods in source code written in an object-oriented programming language, which improves maintenance and readability of the source code. Refactoring methods do not change the external behaviour of a program. However, insecurity refactoring will introduce vulnerabilities using refactoring methods. The external behaviour does not change as long as the program is used in the normal benign way. If the vulnerability is exploited the external behaviour changes. This is the main difference of insecurity refactoring.

There are plenty of opportunities for Insecurity Refactoring. Figure 1 lists three opportunities. First of all, it will be examined, if the modified source code can be used as learning examples. Two

questions should be answered in this context. The first question is: Can it be used as learning examples to teach software security skills? This has the focus on manually reviewing source code samples. The second question is, can these examples can be used to teach the usage of static code analysis tools. Empirical studies will be used to get results on these questions. Insecurity refactoring can be used to create a data set for benchmarking static code analysis tools. The results can be compared to existing benchmarks. For example, it can be compared with the benchmark developed by Antunes and Vieira (2015).

3. Research method

There is currently no existing implementation of insecurity refactoring except for a prototype from our previous work (Schuckert, 2016). This prototype can generate SQL injection vulnerabilities in Java source code. The LAVA project by Dolan-Gavitt et al. (2016) injects buffer overflow in source code. The research about how insecurity refactoring can be developed to create software security exercises needs to survey prior research. Figure 1 shows an overview of the progress from our different research questions. Created vulnerabilities should not be artificial like existing learning examples.

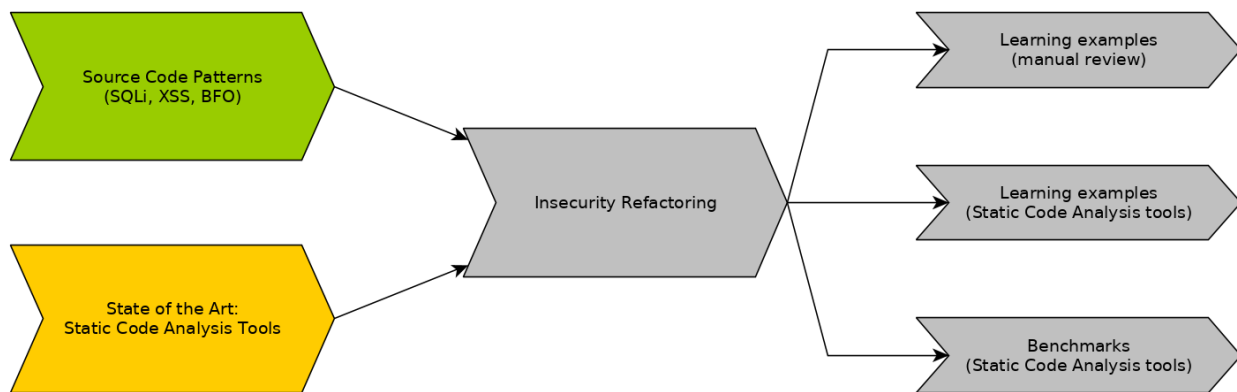


Figure 1: Overview of the research progress

3.1. Source Code Patterns of Software Vulnerabilities

The first research question is, what are the current patterns of software vulnerabilities in open source projects? We researched source code patterns of vulnerabilities from open source projects. We did classify patterns for SQL injection (OWASP A1), cross site scripting (OWASP A3) and buffer overflow vulnerabilities. These are vulnerabilities which occur in the top 3 of OWASP TOP 10 (OWASP 2017) and in the top 4 of CWE TOP 25 (CWE 2017). For each of these vulnerability categories we did choose 50 random CVE reports which are related to open source projects. Figure 2 shows the resulting taxonomy. The numbers are the occurrence of these patterns. For SQL injection, we discovered patterns for sources, concatenations, sinks, failed sanitization methods and fixes. Failed sanitization methods were methods used in the code samples to mitigate SQL injection attacks. Further explanation can be found in (Schuckert et. al., 2017). For cross site scripting patterns were documented for the same parts. The source code samples for the buffer overflow category were a bit different. We did choose 50 CVE reports who are related to Firefox. Source code pattern categories were created for types of error, sinks and fixes. Sources were missing in this part because within the bounds of the project we did not find a proper tool for doing data flow analysis in C++. These patterns

will be used for the insecurity refactoring method. Based on these patterns vulnerabilities will be created for training purposes.

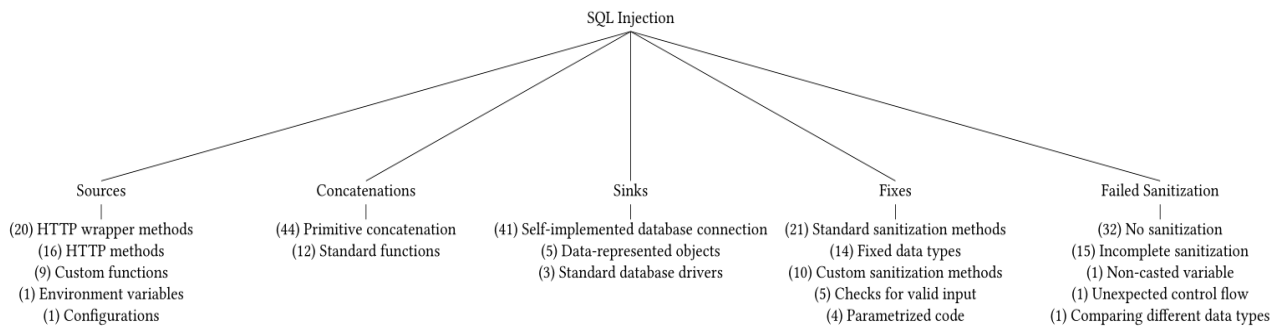


Figure 2: Taxonomy of SQL Injection vulnerabilities found in 50 CVE entries related to open source projects.

3.2. State of the art: Static Code Analysis tools

As seen in the work by Antunes and Vieira (2015) modern static code analysis tools still have a high false negative rate. Accordingly, tools do not detect all vulnerabilities. Our current research will answer the following question: What are the limitations and problems of current static code analysis tools? Figure 3 provides an overview of the choice of method for the static code analysis research question. We use static code analysis tools on our previously crawled source code samples. These samples are the same which were used to classify the source code patterns. Our data set provides the source code of the vulnerable version and the version which was patched by the developers. If the vulnerability is not found in the vulnerable code sample, a false negative will be reported. Accordingly, if the static code analysis tool reports a vulnerability in the patched source code sample, a false positive will be reported. The false negatives and false positives are investigated if there are any patterns found that allow conclusions why these vulnerabilities are not detected. It will be tested if these patterns can be reconstructed in a simplified version and are still not detected by static code analysis tools. These results will be stored in a database to be used later on for insecurity refactoring to create vulnerabilities, which create false negative results. Additionally, the false negative patterns can be used by insecurity refactoring to create samples, which can be used for benchmarking static code analysis tools.

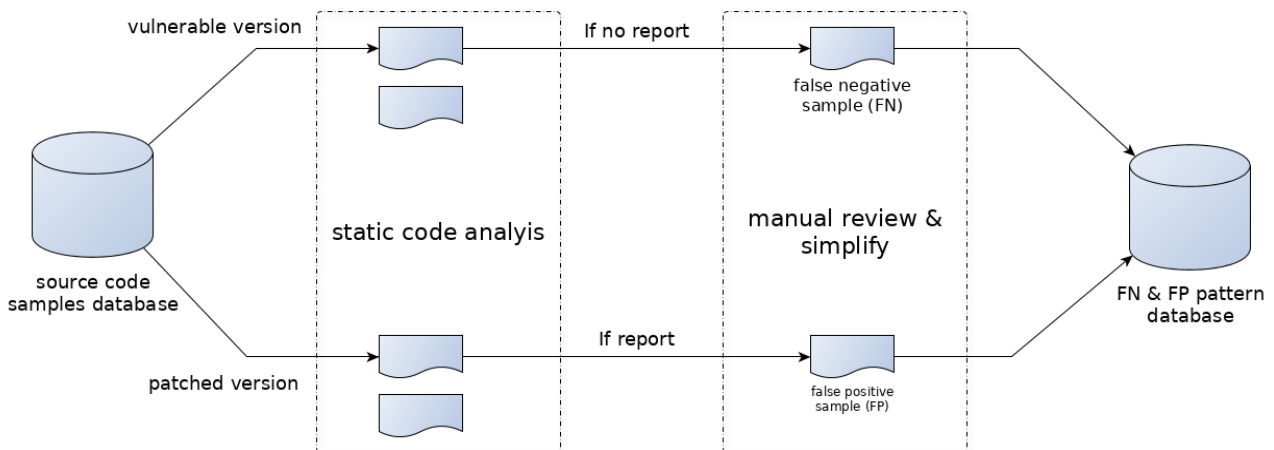


Figure 3: Choice of method for finding false positive and false negative patterns.

3.3. Insecurity Refactoring

After the patterns for vulnerabilities and the limitations of static code analysis tools are identified, the main research question can be answered, which is: How can static code analysis and refactoring methods be used to modify source code to create software vulnerabilities? Is it possible to create vulnerabilities which are not detectable by modern static code analysis tools? The work by Abadi et. al. (2011) did use refactoring methods to remove SQL injection vulnerabilities in Java source code. We have to choose a tool or framework that can be used to implement insecurity refactoring. In our previous work the tool Spoon (Spoon, 2017) was used. This tool allows to transform Java source code by modifying an abstract syntax tree. Mainly PHP was reviewed and Spoon does not support PHP, so an alternative has to be found or developed. An abstraction layer is required to be able to create the vulnerabilities in different programming languages. After the development of such a framework to create vulnerabilities in source code, our previous work will be included. The insecurity refactoring process should be able to create vulnerabilities which have the researched source code patterns. Also the possibility for the framework will be to create vulnerabilities which have a high chance to create false positive and false negative results on static code analysis. Further steps in the research process will be the evaluation of the results of the developed insecurity refactoring framework.

3.4. Learning examples

The first evaluation will be, if the insecurity refactoring helps to teach software security. The question is, if automatically generated exercises can be used to train software security skills? Do they have any advantages or disadvantages compared to manually created exercises? Also can such exercises be used to teach the use of static code analysis tools? A paper by Puhakainen et al. (2010) provides an overview of existing training literature for information security. They showed that only two of the reviewed articles did an empirical study to prove their usefulness. For our project an empirical study will be done to prove the usefulness of insecurity refactoring. Different target groups will be used, first of students that are currently learning software development. They will get different versions of refactored source code. Small and medium sized source code will primarily be used to teach how such security issues occur. The students will get source code with vulnerabilities from different categories. Large source code projects will also be used. They are used to teach the use of static code analysis tools. The generated vulnerabilities have to be detectable by static code analysis tools. Additionally the tools should report enough false positives that a learning effect is possible. The students have to use the tools to detect the vulnerabilities and correctly identify the false positives.

This evaluation will be done by questionnaires. Different questions will be constructed to represent the learning effect. Another criterion for the evaluation will be manual observation. For example, it will be observed, if students are motivated by these automatically generated exercises. Other learning platforms like WebGoat (WebGoat, 2017) or SEEDLabs (SEEDLabs, 2017) will be used as learning examples as well. The questionnaires will include questions that compare these platforms with the automatically generated exercises. The same empirical study can be done with developers from companies. For example, companies are doing refresher events that refresh software security skills of their developers. Another evaluation will be done with Capture the Flag events. The participants are primarily people with software security skills. Such an event can be hosted using tasks which are generated by insecurity refactoring. These events will provide different information about first exploit time and how many teams did exploit the vulnerability. If an attack/defense Capture the Flag event will be held, it also provides information on how many teams did fix the issues. Teams can provide the source code with the fix and the different fixes can be compared. Questionnaires can also be used after the Capture the Flag event. This probably requires a motivation for each team, that they will provide the answers.

3.5. Benchmarking static code analysis tools

The question is, can Insecurity Refactoring be used to create data sets which can be used to benchmark static code analysis tools? Are these results comparable to the results from existing research projects which benchmarked static code analysis tools? Research projects by Antunes et al. (2015, 2010, 2009) did benchmark static code analysis tools and penetration testing tools. They did choose open source and commercial tools for the benchmark tests. Both types of tools are reporting different results, for example, lines of code and input fields that probably contain issues. They used common benchmark metrics like F-measure by Van Rijsbergen (1979) for the comparisons. Another paper by Fonseca et al. (2007) did test scanning tools for SQL injections and cross site scripting attacks. The tested tools were using software fault injection techniques. They did evaluate the tools primarily on false positive and coverage metrics. They used as code base for the tests a self-created project and one existing project from the web. All of these research projects have no code base in common for the benchmarks. Benchmarks have to cover all kinds of security issues. Consequently the developed framework requires plugins which will cover all kinds. The permutations will basically include vulnerabilities categories having the categories from the classified source patterns in mind. Additionally, all types of code complexities are required. This will provide a wide range of security vulnerabilities for benchmark tests. Accordingly, plugins will be developed and improved to cover these types. The next part will be analysing the code bases from other evaluations and benchmark tests. All reported false positives will be noticed to add plugins that will create false positives instead of true positive vulnerabilities. This will improve the benchmark test because typical false positives will be also tested. The next step will be doing the same benchmarks from other research papers but using a refactored code base. The same metrics will be calculated and compared with the results from the papers. It will be checked if any metrics will diverge because of the code base. These tests have to use the same tool versions as used in the papers. After this comparison, the benchmarks with the refactored code base will be done with up to date static code analysis tool. It will be checked, if any improvements for the metrics did occur.

4. Conclusion

Insecurity Refactoring is a new approach to automatically create software vulnerability samples. The focus is to create samples that are not artificial like samples from existing learning platforms. Accordingly, source code patterns of software vulnerabilities are already classified. Current research is about the problems and limitations of static code analysis tools. The resulting false positive and

false negative patterns will be used for insecurity refactoring as well. How Insecurity Refactoring will be developed has to be decided. The focus will probably be on SQL injection and cross site scripting using the programming language PHP. Source code patterns for both of these vulnerability categories were already examined in PHP. Afterwards, a lot of opportunities for Insecurity Refactoring can be researched. For our research the main focus will be on using it as learning examples to learn software security skills and the use of static code analysis tools. Nevertheless, the opportunity to use these samples as a benchmark data set for static code analysis tools will be investigated. Overall the research projects aims at the benefits of Insecurity Refactoring.

References

- Abadi, A., Ettinger, R., Feldman, Y., A. and Shomrat, M. (2011). "Code-motion for API migration: Fixing SQL injection vulnerabilities in Java". In: *Proceedings - International Conference on Software Engineering*, pp. 1–7.
- Antunes, N. and Vieira, M. (2015) "On the Metrics for Benchmarking Vulnerability Detection Tools". In: *Proceedings of the International Conference on Dependable Systems and Networks*, pp. 505–516.
- Antunes, N., & Vieira, M. (2015). "Assessing and comparing vulnerability detection tools for web services: Benchmarking approach and examples". In: *IEEE Transactions on Services Computing*, 8(2), pp. 269-283.
- Antunes, N., & Vieira, M. (2010). "Benchmarking vulnerability detection tools for web services". In: *2010 IEEE International Conference on Web Services (ICWS)*, pp. 203-210.
- Antunes, N., & Vieira, M. (2009). "Comparing the effectiveness of penetration testing and static code analysis on the detection of sql injection vulnerabilities in web services". In: *2009. PRDC'09. 15th IEEE Pacific Rim International Symposium on Dependable Computing*, pp. 301-306.
- Common Weakness Enumeration (CWE) (2017) <https://cwe.mitre.org>.
- Dolan-Gavitt, B., Hulin, P., Kirda, E., Leek, T., Mambretti, A., Robertson, W. & Whelan, R. (2016). "Lava: Large-scale automated vulnerability addition". In: *2016 IEEE Symposium on Security and Privacy (SP)*, pp. 110-121.
- Fonseca, J., Vieira, M., & Madeira, H. (2007). "Testing and comparing web vulnerability scanning tools for SQL injection and XSS attacks". In: *PRDC 2007. 13th Pacific Rim International Symposium on Dependable Computing, 2007*. pp. 365-372.
- Goseva-Popstojanova, K. and Perhinschi, A. (2015). "On the capability of static code analysis to detect security vulnerabilities". *Information and Software Technology*, pp. 18–33
- Howard, M., & Lipner, S. (2006). *The security development lifecycle (Vol. 8)*. Redmond: Microsoft Press.
- Open Web Application Security Project (OWASP) (2017) <https://www.owasp.org>.
- Opdyke, W., F. (1992). *Refactoring Object-Oriented Frameworks*. PhD thesis, University of Wisconsin.
- Puhakainen, P., & Siponen, M. (2010). "Improving employees' compliance through information systems security training: an action research study". In: *MIS Quarterly*, pp. 757-778.
- Rijsbergen, V. (1979). *Information Retrieval*, 2nd edition, Butterworth-Heinemann Ltd
- Schuckert, F., Katt, B., & Langweg, H. (2017). „Source Code Patterns of SQL Injection Vulnerabilities". In: *Proceedings of the 12th ACM International Conference on Availability, Reliability and Security*. pp. 1 - 7.
- Schuckert, F. (2016). "Practitioners Track: Generating Security Vulnerabilities in Source Code". In: Gesellschaft für Informatik eV (GI), *Proceedings of Sicherheit 2016*
- SEEDLabs. (2017) <http://www.cis.syr.edu/~wedu/seed/index.html>.
- Spoon - Source Code Analysis and Transformation for Java. (2017) <http://spoon.gforge.inria.fr/>.
- WebGoat. (2017) <https://github.com/WebGoat/WebGoat>.

Expert System for Risk Analysis of Technical Systems

J. Single, J. Schmidt, J. Denecke
CSE Center of Safety Excellence, Pfinztal, Germany
e-mail: johannes.single@cse-institute.de

Keywords: Expert System, Risk analysis, Knowledge base

Introduction

Risk analyses are performed to identify and assess potential risks within technical, financial and organizational systems. With regards to the chemical process industry, facilities are assessed regarding potential hazards. Typical steps are hazard identification, consequence analysis, assessment of hazards and the consideration of protective measures. Risk analyses are usually conducted by an interdisciplinary team of experts. Therefore, it is time consuming and costly. Furthermore, subjective risk evaluations can lead to different results with different personnel. Also, laymen or even generalists (non-experts) can hardly perform risk analysis or contribute to the process. Due to an increased public awareness towards industrial accidents and increasing legal requirements, risk analyses become more important.

Literature review

For hazard identification purposes the HAZOP method (Hazard and Operability) is used most widely. A typical chemical process can take 1-8 weeks with a team of 4-8 members (Zhao et al., 2009). The estimated costs of hazard analysis in the process industry is 1 % of sales or 10 % of profits (Venkatasubramanian et al., 2000). Since the 1990s, various research groups are working on rationalization approaches. These can be categorized into automated HAZOPs, hybrid methods, improvements and screening. In automated model-based HAZOP approaches process simulators are extended to conduct a HAZOP while simulating processes (Janošovský et al., 2016). Another approach is the derivation of a meta-model based on a P&ID (Piping and instrumentation diagram). It is afterwards analyzed by an algorithm which automatically identifies potential hazards (Rodríguez et al., 2012). Another research group investigates possibilities, where hazards are automatically identified using a CAD system (Cui et al., 2010). Furthermore, multiple approaches use an expert system. These systems use knowledge bases which contain process-, material-, unit- and hazard specific knowledge (Zhao et al., 2009) and (Rahman et al., 2009). Other rationalization approaches combine methods e. g. HAZOP and FMEA (Failure mode and effect analysis) (Seligmann et al., 2012). Other approaches aim to identify potential hazards in a fast and superficial manner, see (Khan et al., 1999). So far, none of these research approaches made it into industrial practice. Currently, it is common in the industry to use pre-filled application forms to simplify documentation processes and checklists. Another aspect concerns the awareness of past accidents in companies. Egan describes a case, where in 2015 nobody was aware anymore of an accident that happened in 1966 (Egan, 2017).

Research approach

The research methodology is of deductive nature. It is based on the hypothesis, that with 20 % of the effort, 80 % of the relevant hazardous scenarios can be identified. Furthermore, an approach is explored which enables non-experts to estimate hazardous consequences.

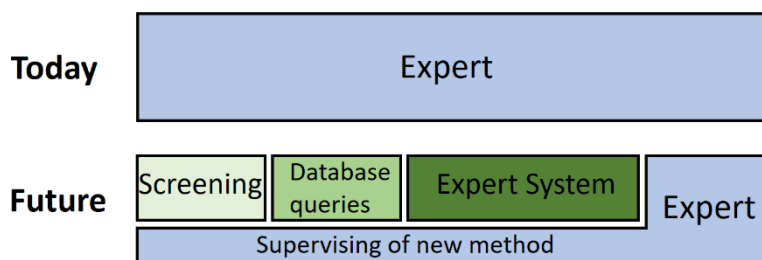


Figure 1: Aim of the research approach

To analyze the state of the research, an extensive literature review is conducted. Furthermore, experts from different companies are interviewed and their procedures are analyzed and commercial software is evaluated.

Today, risk analyses are fully carried out by experts (see Figure 1). In the future, a software approach is developed which supports experts. This can be achieved by automatically analyze simple plant sections, using an expert system. The main idea behind it is to conduct hazard analyses (HAZOPs) for individual units e. g. a vessel. These results are mapped in knowledge bases. To model an entire system these individual HAZOPs are concatenated. Afterwards, an algorithm investigates the propagation of hazardous events from one unit to another, crosswise and of the entire system.

To identify potential hazardous events in a fast manner, a screening method is developed. Thereby also non-experts should be enabled to roughly estimate the most hazardous consequences. Additionally, the awareness regarding potential accidents can be increased. To improve the situation where experts are not aware of past accidents, databases could be automatically queried. Therefore, databases like *eMARS* (online Major Accident Reporting System maintained by the Major Accident Hazard Bureau of the European Commission) are analyzed using *case-based reasoning* techniques. Thereby, the range of experience of experts can be complemented.

To investigate why rationalization approaches were not implemented in industrial practice, surveys for industrial experts are conducted. These findings will be used to develop a software to speed up the risk analysis process using an expert system. Hence, more plants can be reliably analyzed in the same time. Furthermore, experts are enabled to focus on complex cases, while less attention is demanded for routine procedures. The project is in the data collection phase, where literature evaluated, expert surveys are conducted and first results of a screening method are evaluated.

Preliminary results

The screening method is based on *case-based reasoning* (see Figure 2). Through the expert interface, expert knowledge can be manipulated or results approved or rejected. In Figure 3 as an example the expert knowledge regarding the hazardous event ‘jet fire’ is mapped. For this purpose, an ontology is used. Hazardous events can be identified via specific attributes e. g. ‘Flammable’ or ‘Pressurized’, which are interrelated causes (see Figure 3). Multiple hazardous events are stored in a knowledge base.

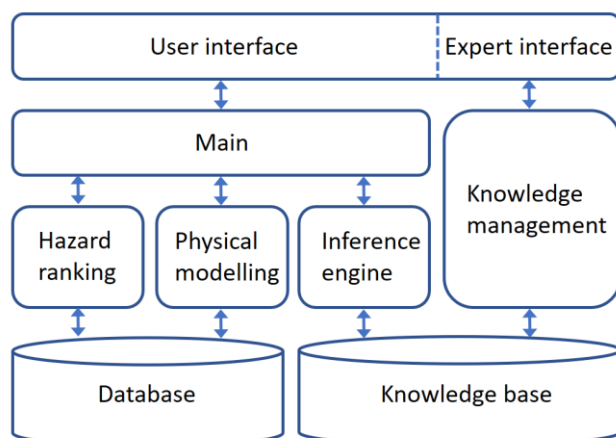


Figure 2: Schematic representation of the screening method

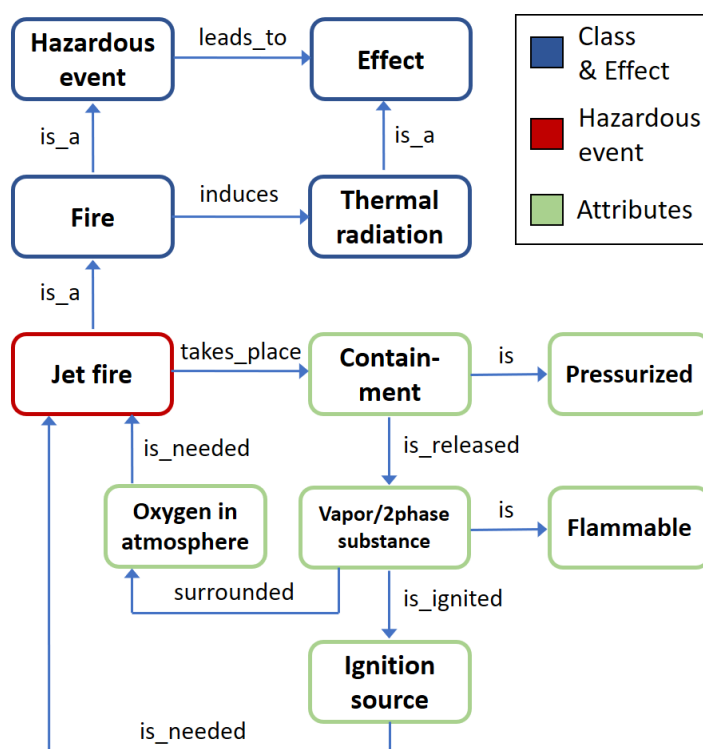


Figure 3: Schematic representation of an ontology used for knowledge representation of the hazardous event 'jet fire'

The user input consists of substance-, process- and environment specific data. It is used to build the attributes of the considered case using physical models, correlations and assumptions. For example, a database is queried regarding the substance properties and hence the attribute 'Flammable' is set. The inference engine determines the similarity measure between the considered case and the cases in the knowledge base. On that basis, potential hazardous events are predicted with a measure of confidence. Through the introduction of a hazardous index, units can be ranked relative to each other.

For now, knowledge is mapped in XML files. This could be improved using *Web Ontology Language* (OWL). This would need an adaption of the inference engine. The weights of the attributes which are

used for the similarity measurement have a strong impact on the resulting hazardous events. Therefore, they need further intense testing by benchmarking the results with case studies from literature and expert judgement.

Conclusion and Outlook

There are various research approaches to rationalize the process of a risk analysis in the process industry. Some of these, aim to overcome cognitive biases like overconfidence, deflected attention and varying results between experts. Until now, none of these research approaches are applied in industrial practice.

The risk analysis process must be understood as a socio-technical system to gain the approval from industrial experts. It consists of a technical- (software) and a social component (user/team) which needs to be considered together. This is lacking in other research approaches and probably the reason why methods could not be transferred into industrial practice. Furthermore, methods need to be compatible with the company internal workflow and policies.

Speeding up the process risk analysis is tackled from different sides. On one hand, a screening approach is developed to identify potential hazardous events in a fast manner. On the other hand, possibilities are investigated to use a database and a prediction algorithm to support experts. Furthermore, a software with a built-in expert system will be designed to rationalize the process of a risk analysis, while freeing resources for improved creativity and completeness. While using advanced software e. g. a self-learning system for decisions with ethic consequences, somebody needs to supervise the software.

References

- Cui, L., Zhao, J. & Zhang, R. (2010). The integration of HAZOP expert system and piping and instrumentation diagrams. In: *Process Safety and Environmental Protection*, 88(5), pp.327–334.
- Egan, S. (2017). Lessons from a fatal explosion in a pilot plant European DIERS June 2017. In: *European DIERS Users' Group Meeting (EDUG)*. Lyon, pp. 1–16.
- Janošovský, J., Labovský, J. & Jelemenský, L. (2016). Automated Model-based HAZOP Study in Process Hazard Analysis. In: *CHEMICAL ENGINEERING TRANSACTIONS*, 48(1994), pp.505–510.
- Khan, F.I. & Abbasi, S.A. (1999). MAXCRED – a new software package for rapid risk assessment in chemical process industries. In: *Environmental Modelling and Software*, 14, pp.11–25.
- Rahman, S. et al. (2009). Journal of Loss Prevention in the Process Industries ExpHAZOP+: Knowledge-based expert system to conduct automated HAZOP analysis. In: *Journal of Loss Prevention in the Process Industries*, 22(4), pp.373–380.
- Rodríguez, M., Luis, J. & Mata, D. (2012). Automating HAZOP studies using D-higraphs. In: *Computers and Chemical Engineering*, 45, pp.102–113.
- Seligmann, B.J. et al. (2012). Journal of Loss Prevention in the Process Industries A blended hazard identification methodology to support process diagnosis. In: *Journal of Loss Prevention in the Process Industries*, 25(4), pp.746–759.
- Venkatasubramanian, V., Zhao, J. & Viswanathan, S. (2000). Intelligent systems for HAZOP analysis of complex process plants. In: *Computers & Chemical Engineering*, 24, pp.2291–2302.
- Zhao, J. et al. (2009). Learning HAZOP expert system by case-based reasoning and ontology. In: *Computers and Chemical Engineering*, 33, pp.371–378.

On measuring the behaviour of users in the area of security-sensitive platforms

C. Schalles

Faculty of Business Informatics

Baden-Wuerttemberg Cooperative State University, Germany

e-mail: christian.schalles[at]mosbach.dhbw.de

Keywords: Software Security, Human Computer Interaction, Software Engineering

Abstract: In the past, the causality between the fields of Human Computer Interaction and Software Security have not been investigated very closely. That is to say that the techniques prevalent in HCI aim at improving users' effectiveness, efficiency or satisfaction, but they do not consider security aspects of software systems. The reason why security and user behaviour seem to be opposed is the conflict of interest that exists between the system owner and its users. This initial research paper proposes an evaluation framework for measuring this conflict generally. As a future result of this research, specific recommendations for future implementation of security standards in software systems can be defined.

1. Introduction

Increasing digitalization and IT-consumption lead to a specific need for analysing user behavioural aspects in the context of security and data protection needs. On the one hand, security-related and/or legal-based topics increase the security level of information, which is operated by information systems. On the other hand, traditional usability criteria come to the fore during the usage of information systems. If security concerns are not addressed in the design of an information system, users respond by circumventing them when using the software. Achieving a balance between the intrusiveness of security mechanisms and usability goals is an important consideration in designing successful software systems (Elahi & Yu, 2009).

In this publication, we aim to develop a model for measuring the behaviour of user groups in the context of security-sensitive systems. By deducing relevant influencing variables, psychological and behavioural aspects were considered and integrated in our research model. Thus, we aim to identify causal relations between user-based and security-based variables. For the upcoming research, we state the following general hypothesis:

H1: Security Properties of Software Systems affect User Behaviour negatively

The basic idea behind this is, that the number and intension of security layers implemented in software systems have a negative impact on the behaviour of individuals during the usage of the software system (see Figure 1).

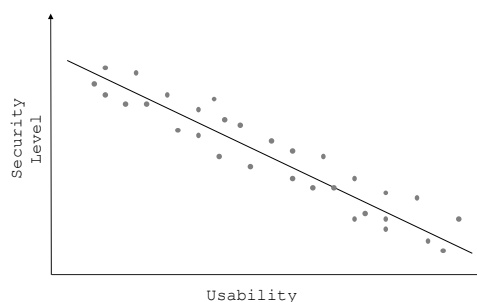


Figure 1: Hypothesized Relation

2. Evaluation Framework

In order to develop an evaluation framework, the relevant variables must be defined and structured. In general, three main types of variables concerning experimental data collection and analysis exist: dependent variables, independent variables and moderator variables. For deducing relevant variables, we underlay a framework, that was developed for evaluating the user behaviour during complex tasks in the area of software engineering (Schalles et al., 2015).

2.1. Usability Variables

Learnability describes the ease of learning the application (i.e. interpretation) of modeling languages. For this characteristic, the standard measure values are based on task completion rates and the task accuracy. In general, learnability is a development and can be graphically described by learning curves. Hence, learnability can be measured by the rate of difference when the user repeats evaluation sessions. Nielsen 2006 insists that highly learnable systems could be categorized as “allowing users to reach a reasonable level of usage proficiency (...)” (Nielsen, 2006, p.28).

The efficiency is the amount of human, economical and temporal resources. Measures of efficiency relate to the level of effectiveness achieved to the expenditure of resources (Bevan et al., 1994). Measure values of efficiency include time taken to complete tasks, i.e. duration time for performing an experimental task.

The individual satisfaction of a user is a subjective criterion that can be measured best by using standardized questionnaires. Therefore, questionnaires focusing on system and website usability have to be adapted. Prominent questionnaires are generally contributing to the Questionnaire for User Interaction Satisfaction (QUIS) and additionally the Software Usability Measurement Inventory (SUMI) (Kirakowski et al., 1993).

2.2. Security Variables

Software requirements can be defined as a condition or capability needed by a user to solve a problem or achieve an objective. Furthermore, the condition or capability must be met by the software to satisfy a contract, standard, specification or other formally imposed document (IEEE 729, 2017).

Furthermore, we analysed maturity models for measuring the security level of software and organizations. As a consequence, we mapped the Software Assurance Maturity Model (SAMM) to the categories of the IEEE 729 standard (Chandra, 2017).

Based on this, we derive backend properties, organizational properties and frontend properties as variables for measuring the security of a software system in organizations. Table 1 shows the idea of mapping and deriving security properties defined in various standards.

SAMM Criteria (Chandra, 2017)	Security Variables	Metrics (examples)
Deployment	Frontend Properties	<ul style="list-style-type: none"> • Authorization Scheme • Access Control
Construction	Backend Properties	<ul style="list-style-type: none"> • Encryption • Data Separation
Governance, Verification	Organizational Properties	<ul style="list-style-type: none"> • Trainings • Policies • Work Instructions

Table 1: Mapping of Security Properties

2.3. Moderator Variables

We analysed two moderate variables, which affect the causal relations in our evaluation framework. The experience of a user in working with a specific software and user's attitude towards information security. For measuring the security attitude several frameworks exist. Therefore, we underlay a typology of lifestyles developed by Otte (2005). The typology intends to provide a simple and resilient tool for the analysis of social structures by lifestyles. In general, the topology is based on the SINUS-Milieus (SINUS, 2015). Figure 2 shows the deduced research model as a basis for our work.

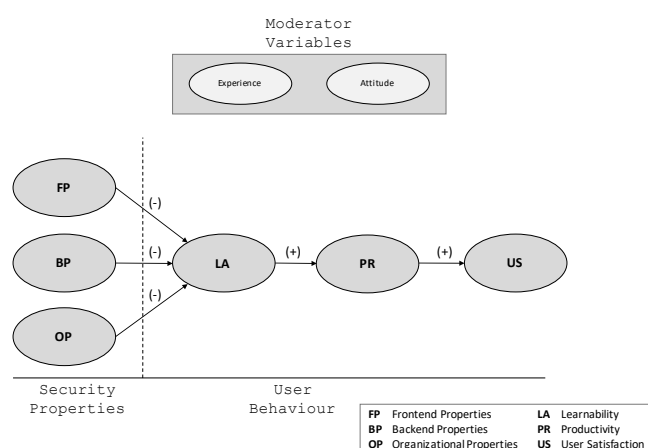


Figure 2: Proposed Research Model

3. Related and future work

Klein and Luciano (2016) analysed the influences of information security behaviour. This study contributes to understand a user's computer behaviour in an organizational context and made it possible to combine concepts from several behavioural theories. Tu and Yuang (2015) aim to understand the security-based behaviour of users in the area of mobile devices. Kobsa et al. (2009) found out that for a specific mobile device service different causal relations exist between usability attributes. Hence, an overall model measuring usability in the domain of secure platforms is missing yet. With our approach, we aim to close this research gap.

Next steps include the design of an empirical study to prove the research model developed and deduced from theory. For measuring the latent variables, we need to conduct multiple methods of data collection, e.g. questionnaire, experiment and observation. It is planned to include third year students of various study directions.

References

- Bevan, N. (1995). Measuring usability as quality of use. *Software Quality Journal*, 4, 115-130
- Chandra, P. (2017). Software Assurance Maturity Model - A guide to building security into software development. <http://www.opensamm.org>
- Elahi, G., Yu, E. (2009). Modeling and analysis of security trade-offs – A goal oriented approach. *Data & Knowledge Engineering*, 68, 579-598
- IEEE (2017). Standard Glossary of Software Engineering Terminology, *IEEE*

- Kirakowski, J., Corbett, M. (1993). SUMI: The Software Usability Measurement Inventory. *British Journal of Educational Technology*, 24, (1) 210-212
- Klein, R. H., Luciano, E. M. (2016). What influences information security behavior? A study with brazilian users, *Journal of Information Systems and Technology Management*, 13, 1775-1807
- Kobsa, A. (2009). Serial hook-ups: a comparative usability study of secure device pairing methods, in *Proceedings of the 5th Symposium on Usable Privacy and Security*. ACM: Mountain View, California. p. 1-12.
- Nielsen, J. (2006). *Usability Engineering*, Kaufmann
- Otte, G. (2005). Entwicklung und Test einer integrativen Typologie der Lebensführung für die Bundesrepublik Deutschland, *Zeitschrift für Soziologie*, 34, 442-467
- Schalles C., Creagh J., and Rebstock, M. (2015). Developing a causal model for analyzing the impact of graphical modeling languages on usability, *International Journal of Software Engineering and Knowledge Engineering*, p. 1337-1356
- SINUS (2015). Informationen zu den Sinus-Milieus 2015, http://www.sinus-institut.de/fileadmin/user_data/sinus-institut/Downloadcenter/Informationen_zu_den_Sinus-Milieus.pdf
- Z. Tu, Y. Yuan (2012). Understanding User's Behaviours in Coping with Security Threat of Mobile Devices Loss and Theft, *Proceedings of the 45th Hawaii International Conference on System Science*, p. 1393-1402.

Gamifying cyber security awareness via mobile training apps

F.Alotaibi¹, S.Furnell^{1,2,3}, I.Stengel^{1,4}, M.Papadaki¹

¹ Centre for Security, Communications and Network Research, University of Plymouth, Plymouth, UK

² Security Research Institute, Edith Cowan University, Perth, Western Australia

³ Centre for Research in Information and Cyber Security, Nelson Mandela Metropolitan University, Port Elizabeth, South Africa

⁴ Hochschule Karlsruhe, University of Applied Sciences Karlsruhe, Germany

{Faisal.Aalotaibi, Steven.Furnell, Maria.Papadaki}@plymouth.ac.uk, Ingo.Stengel@hs-karlsruhe.de

Keywords: Security, Awareness, Gamification

Introduction

Security awareness is now an important issue for all users of information technology and online systems, both on an individual basis and in the context of the workplace. Unfortunately, many users remain unaware of key issues and security itself is often seen as a chore or an overhead, rather than something that they seek to actively embrace. Moreover, while organisations essentially depend upon their staff to be security-aware in order to reduce the potential for incidents and breaches, the evidence consistently shows that only a minority devote attention towards supporting this amongst their staff. For example, the latest UK Cyber Security Breaches Survey indicates that only 30% of respondents claimed to be providing *user awareness and education*, far behind provisions such as *malware protection* and *network security*, with 90% and 89% respectively (Klahr et al. 2017). The premise of this paper is that users may be more accepting of security if it is presented to them in an engaging manner, and so examines the potential for fundamental awareness-raising to be fostered via gameplay, which could be used instead of, or alongside, traditional methods.

Prototype apps for security awareness raising

The concept of gamifying security awareness is by no means unique, and various prior studies have examined how games can be used to support and impart security lessons (e.g. Denning et al. 2013; Gondree et al. 2013; Nyeste et al. 2010). A difference in the approach taken by the current study is that the task was approached from the perspective of considering whether security awareness games could be approached in the similar manner to brain-training apps. Here the intention is for players to engage with short activities on a frequent basis, in order to gradually develop and hone a particular skill or ability. A set of 12 candidate game concepts were identified, and expressed in conceptual form. This included outlining the security learning/awareness objective(s) to be served, key elements of gameplay, and some storyboard level consideration of how the game might look. From this, two game concepts were taken forward for practical implementation and evaluation, as outlined below and illustrated in Figure 1.

- **Password Protector:** The aim of this app is to improve users' understanding of what makes a strong password, and to enable them to practice the task in a competitive context. The app provides them with a restricted set of characters to choose from, and they must attempt to make best use of the options available (based on standard password selection principles that longer passwords with more diverse character space will generally represent more secure choices). This is done against the clock, and selections are rated using a password meter, requiring users to reach at least Moderate for the password to be acceptable. Having chosen a password, users must then demonstrate that they can remember their choice by entering it again. They are given a time limit for doing this, based upon how long it took them to select the password in the first place.

- Malware Guardian:** The aim of this game is to acquaint users with the variety of guises in which they may encounter malicious software, and to emphasize the role of up-to-date anti-malware protection and backups as key means of safeguarding against problems. Within the game, players are required to defend their system by essentially playing the role of the anti-malware package, scanning files as they approach the system and blocking those that are found to be malicious. A variety of different malware types can be encountered (including viruses, Trojans, spyware, ransomware), each of which causes different effects if they manage to reach the system. Only clean/uninfected files should be permitted to get this far, and the gameplay challenge comes from the speed and frequency of objects to scan, combined with the parallel tasks of updating the malware signatures and ensuring regular backups.



Figure 1: The game app prototypes (a) Password Protector and (b) Malware Guardian

Experimental evaluation and early findings

The implemented games will now be the focus of experimental evaluation, with at least 50 participants being involved for each game. In each case, participants will be asked to complete pre-test questionnaires that assess their current knowledge and appreciation of the topic (i.e. password selection or malware protection, depending upon the game they are due to trial). They will then have a period of two weeks in which the related game app is installed on their device, and their usage of it will be tracked (i.e. counting the number of times they play it). At the end of the trial period they will be asked to complete a further survey, this time re-assessing their views on the topic area (i.e. to see if it has changed from the initial survey), as well as their feedback on the game (assessing aspects such as playability, enjoyment, and awareness-raising value).

A pilot of this process has already been conducted with nine participants, in order to validate the surveys and ensure that the game apps were meaningful to a wider audience. While it is not statistically meaningful to analyse the survey results from this stage in any depth, they did at least serve to give an illustration that the games themselves were considered effective by the participants. The themes of usability, learning content and enjoyment were each investigated via 4-5 related questions, with each rated on a 5-point scale (where 1 is most negative and 5 is most positive). Given the low number of respondents, rather than looking at each question individually, Table 1 presents the averages across all questions in each category for each of the games. This gives a broad indication

that the overall results in all cases was skewed towards the positive side. The notably lower ‘enjoyment’ score for the Password Protector game is perhaps to be expected, as it is a time-based memory game, whereas the apparently more enjoyable Malware Guardian is an action game and does not make the participants feel that they are being explicitly tested and rated in the same way.

Table 1: Averaged game feedback ratings from the pilot study

	Average ratings		
	<i>Usability</i>	<i>Learning content</i>	<i>Enjoyment</i>
Password Protector	3.8	3.9	2.9
Malware Guardian	3.6	3.8	3.7

As a more specific result, one of the activities in the pre- and post-study surveys for the Password Protector game is for users to create what they consider to be a strong password. In the pre-test surveys, this was yielding responses such as brief passwords and/or passwords composed of only one or two character types (e.g. alphabetic only, numeric only, etc), and scoring the passwords against the strength meter algorithm used in the app gave an average of just 45% across the participants. It was therefore encouraging to find that the average observed from the same task in the post-test survey had risen to 78%, and it was notable that all passwords were now alphanumeric and included at least one punctuation symbol, whereas none of the pre-test choices had included these at all.

Accompanying text-based comments were also collected and gave some early indications of areas for development. One notable example was the desirability of playing the games against other players as a form of competition. This is a theme that the authors had already identified as a potential route for deployment of such apps in an organisational setting (e.g. as a means of engendering a sense of competition amongst staff within and between departments).

Conclusions

Cyber security is a topic area in which those that need to hear about it are more likely to engage and accept it if the efforts are made to present the messages in a more effective manner. Gamification is one guise that is likely to improve the chances of reaching some of the audience, and short-use mobile apps is a format that is likely to appeal to the current generation of technology users. The initial findings from the prototyping and piloting of the two game apps discussed here are positive, and it is anticipated that the findings from the full trial will be able to further inform future research and development directions in this area.

References

- Denning T., Lerner A., Shostack A., and Kohno T. (2013). “Control-Alt-Hack: the design and evaluation of a card game for computer security awareness and education,” in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pp. 915–928.
- Gondree M., Peterson Z. N., and Denning T. (2013). “Security through play,” In: *Secur. Priv. IEEE*, vol. 11, no. 3, pp. 64–67.
- Klahr, R., Shah, J.N, Sheriffs, P., Rossington, T., Pestell, G., Button, M. and Wang, V. (2017). *Cyber Security Breaches Survey 2017*, Main report, April 2017, Department for Culture, Media and Sport, London, UK.
- Nyeste P. G. and Mayhorn C. B. (2010). “Training Users to Counteract Phishing,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 54, pp. 1956–1960.

Chapter 4

Business

Electricity Cost Reduction Potential of Industrial Processes using Real Time Pricing in a Production Planning Problem

M. Fleschutz¹, M. Bohlayer¹, A. Bürger², M. Braun²

¹ Institute of Applied Research

² Faculty of Management Science and Engineering
Karlsruhe University of Applied Sciences, Germany
markus.fleschutz@hs-karlsruhe.de

Keywords

Demand Response (DR), Demand flexibility, Industrial application, Numerical optimization, Machine scheduling, Time-Of-Use (TOU), Real-Time Pricing (RTP), Cement industry

Abstract

Within this paper a Mixed-Integer Linear Programming formulation for price-based demand response is proposed. Through modelling of multiple machines, storages and production sorts, the formulation can be used not only for optimizing the electricity purchase resulting in electricity cost savings but also for production planning problems. As a case study the model is applied to the machine scheduling of two cement mills comparing a time-of-use pricing scheme to a Real-Time Pricing scheme. The optimization resulted in 13 % annual electricity cost savings for the two cement mills.

Nomenclature

Indices / Sets		Variables	
$t \in T$	time steps from t^{start} to t^{end}	p_t^{el}	hourly electricity demand (kWh)
$s \in S$	production sorts	G_t^{pp}	production volume of pre-process (t/h)
$m \in M$	machines	$G_{t,s,m}^p$	production volume of process (t/h)
Parameter		SOC_t^α	α -storage filling level (t)
		$SOC_{t,s}^\beta$	β -storage filling level (t)
		$Y_{t,s,m}^{SU}$	binary start-up indicator
		$Y_{t,s,m}^{OP}$	binary operation indicator
		$Y_{q,s,m}^{SU}$	binary start-up indicator of subset
		$Y_{q,s,m}^{SD}$	binary shut-down indicator
		Abbreviations	
		DR	Demand Response
		RTP	Real-Time Pricing
		TOU	Time-Of-Use
		MILP	Mixed-Integer Linear Programming
		AMPL	A Mathematical Programming Language (software)
c_t	electricity price (€/kWh)		
c^{SU}	cost per start-up (€)		
$p_m^{el,max}$	maximum electric power of machine m (capacity)(kw)		
$g_t^{pp,max}$	maximum production volume of pre-process		
$g_{t,s}^{dem}$	demand (t/h)		
$\phi_{t,s}^{\alpha,cap}$	α -storage capacity (t)		
$\phi_{\alpha,min}$	minimum α -storage filling level (t)		
$\phi_{\alpha,init}$	initial and final α -storage filling level (t)		
$\phi_s^{\beta,cap}$	β -storage capacity (t)		
$\phi_s^{\beta,min}$	minimum β -storage filling level (t)		
$\phi_s^{\beta,init}$	initial and final β -storage filling level (t)		
ξ_s^{el}	sort-specific electricity demand (kW/t)		
$\tau_{min,UT}$	minimum up-time (h)		
$\tau_{min,DT}$	minimum down-time (h)		
w	sufficiently large number for Big M method		

1. Motivation and State of Research

The high penetration of fluctuating renewable energy sources in the electricity market due to the transition to a sustainable energy system increases the need for flexibility of market participants (EURELECTRIC, 2014). Demand Response (DR) programs are a promising and cost-effective alternative to other types of electrical flexibility measures like storages, interconnection or flexible generation (R.A. Verzijlbergh et al., 2016; Shoreh et al., 2016). Recent studies identified high economic potentials in industrial DR applications from both a system and a business perspective (Gruber, Roon and Fattler, 2016; Steurer, Klempp and Hufendiek, 2015). Whereas most of the energy intensive companies in Germany utilize peak load management, participation in spot or German control power markets are still seldom (Langrock et al., 2015). In the literature, different optimization approaches were presented to assess the industrial DR potential (Larsen et al., 2017; Molavi and Ardehali, 2016; Reka and Ramesh, 2016; Rodríguez-García et al., 2016; Wang, El-Farra and Palazoglu, 2017). DR programs can be categorized into price and incentive based programs, while each one has its own benefits taking advantage of different aspects of flexible demand like access time, duration of delay and temporal availability (Asadinejad and Tomsovic, 2017).

Price based DR programs such as Real-Time Pricing (RTP), where an energy-consumer responds to an hourly price signal, imply high requirements in load control infrastructure, communication and market access. However, especially energy intensive companies meet these challenges. Furthermore, compared with commercial and residential loads, industrial loads have the advantage of high saving potentials due to high magnitudes of power consumption.

2. Methodology

In this paper, a production planning model is proposed with the aim of reducing companies' electricity costs and contributing to a reliable power supply by optimizing the shape of the companies' electrical load curve according to a complex price signal. Serving a forecasted demand for finished or intermediate products, while adhering to technical, production-technological and process-related constraints, the machine load is preferably shifted to low price hours.

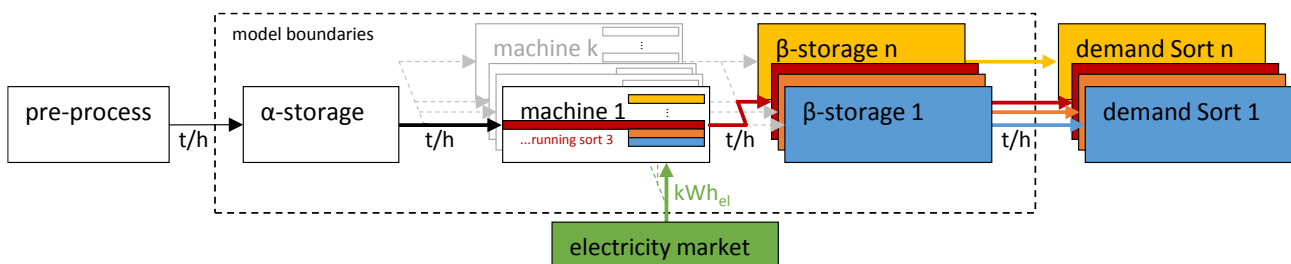


Figure 1: Schematic representation of DR-relevant components and its connections through flow of electricity and goods.

For the sake of simplicity, the model comprises only those components of a production process that are most relevant for DR. As outlined in Figure 1, there are two types of storages: The α -storage stores the semi-finished goods in order to decouple the process under investigation from the pre-process. Since we account for only one raw material for the observed process there is also only one instance of α -storage. There are a number of n β -storages that store the n different sorts of finished goods. There are a number of k machines that can produce every sort. However, not more than one sort can be produced per time step and machine.

3. Mathematical Modelling

The problem is formulated as a Mixed-Integer Linear Programming (MILP) problem. Thereby the combinatorial problem assigning the sorts to the machines can be formulated using binary decision variables while global optimality within acceptable computation time is ensured due to linear formulation. Furthermore, the model relies on a perfect foresight modelling approach in order to keep the model solvable. Optimization variables are denoted with an initial capital letter while parameters have initial lowercase letters.

The objective function consists of the minimization of the operation costs C^{op} caused by the process under investigation

$$\min C^{\text{op}} = \sum_{t \in T} (C_t^{\text{el}} + C_t^{\text{SU}}) \quad (1)$$

with the electricity costs

$$C_t^{\text{el}} = c_t P_t^{\text{el}} \quad \forall t \in T \quad (2)$$

and the costs for machine start-ups

$$C_t^{\text{SU}} = c^{\text{SU}} \sum_{m \in M} Y_{t,m}^{\text{SU}} \quad \forall t \in T \quad (3)$$

where t and T are the index and set for the time steps respectively, and m and M the index and set for the machines respectively. P_t^{el} is the hourly electricity demand, c_t the corresponding electricity price, c^{SU} the cost per start-up and $Y_{t,m}^{\text{SU}}$ the binary start-up indicator.

The optimization problem is subject to the following constraints, which derive from the abovementioned boundary conditions:

- The balance of electric power of the machines

$$P_t^{\text{el}} = \sum_{s \in S} \sum_{m \in M} Y_{t,s,m}^{\text{op}} \cdot p_m^{\text{el,max}} \quad \forall t \in T \quad (4)$$

where s and S are the index and set for the sorts, respectively; $p_m^{\text{el,max}}$ is the maximum electric power of machine m and $Y_{t,s,m}^{\text{op}}$ is the binary operation indicator which is defined through a Big-M formulation

$$G_{t,s,m}^{\text{p}} \leq Y_{t,s,m}^{\text{op}} \cdot w \quad \forall t \in T, s \in S, m \in M \quad (5)$$

where w is a sufficiently large number.

- The limit of volume produced on a machine per time step

$$G_{t,s,m}^p \leq Y_{t,s,m}^{\text{op}} \cdot \frac{p_m^{\text{el,max}}}{\xi_s^{\text{el}}} \quad \forall t \in T, s \in S, m \in M \quad (6)$$

where ξ_s^{el} is the sort-specific electricity demand.

- The production restriction of only one sort per machine and time step.

$$\sum_{s \in S} Y_{t,s,m}^{\text{op}} \leq 1 \quad \forall t \in T, m \in M \quad (7)$$

- The balance of β -storage

$$\text{SOC}_{t,s}^{\beta} = \text{SOC}_{t-1,s}^{\beta} + \sum_{m \in M} G_{t,s,m}^p - g_{t,s}^{\text{dem}} \quad \forall t \in [t^{\text{start}} + 1..t^{\text{end}}], s \in S, \quad (8)$$

where $\text{SOC}_{t,s}^{\beta}$ are the filling level of the β -storages that are defined for the first and last time step

$$\text{SOC}_{t,s}^{\beta} = \phi_s^{\beta,\text{init}} \quad \forall t \in \{t^{\text{start}}, t^{\text{end}}\}, s \in S \quad (9)$$

and restricted by a lower limit $\phi_s^{\beta,\text{min}}$ and the capacity $\phi_s^{\beta,\text{cap}}$.

$$\phi_s^{\beta,\text{min}} \leq \text{SOC}_{t,s}^{\beta} \leq \phi_s^{\beta,\text{cap}} \quad \forall t \in T, s \in S \quad (10)$$

- The balance of α -storage

$$\text{SOC}_t^{\alpha} = \text{SOC}_{t-1}^{\alpha} + G_t^{\text{pp}} - \sum_{m \in M} \sum_{s \in S} G_{t,s,m}^p \quad \forall t \in [t^{\text{start}}+1..t^{\text{end}}] \quad (11)$$

where G_t^{pp} is the volume produced by the pre-process and SOC_t^{α} is the filling level of the α -storage which values are also defined for the first and last time step

$$\text{SOC}_t^{\alpha} = \phi^{\alpha,\text{init}} \quad \forall t \in \{t^{\text{start}}, t^{\text{end}}\} \quad (12)$$

and restricted by the limits $\phi^{\alpha,\text{min}}$ and $\phi^{\alpha,\text{cap}}$.

$$\phi^{\alpha,\text{min}} \leq \text{SOC}_t^{\alpha} \leq \phi^{\alpha,\text{cap}} \quad \forall t \in T \quad (13)$$

- The definition of the start-up and shut-down indicators $Y_{t,s,m}^{\text{SU}}$ and $Y_{t,s,m}^{\text{SD}}$.

$$Y_{t^{\text{start}},s,m}^{\text{SU}} - Y_{t^{\text{start}},s,m}^{\text{SD}} = 0 \quad \forall s \in S, m \in M \quad (14)$$

$$Y_{t^{\text{start}},s,m}^{\text{SU}} - Y_{t^{\text{start}},s,m}^{\text{SD}} = y_{t^{\text{start}},s,m}^{\text{op}} - y_{t^{\text{start}}-1,s,m}^{\text{op}} \quad \forall s \in S, m \in M \quad (15)$$

- The restriction that a machine cannot start up and shut down in the same time step.

$$Y_{t,s,m}^{SU} + Y_{t,s,m}^{SD} \leq 1 \quad \forall t \in [t^{\text{start}+1}..t^{\text{end}}], s \in S, m \in M \quad (16)$$

- The restriction to a minimum up- and downtime $\tau^{\text{minUT}} / \tau^{\text{minDT}}$

$$\sum_{q=t-\tau^{\text{min},UT}+1}^t Y_{q,s,m}^{SU} \leq Y_{t,s,m}^{\text{op}} \quad \forall t \in [\tau^{\text{min},UT}..t^{\text{end}}], s \in S, m \in M \quad (17)$$

$$\sum_{q=t-\tau^{\text{min},DT}+1}^t Y_{q,s,m}^{SD} \leq 1 - Y_{t,s,m}^{\text{op}} \quad \forall t \in [\tau^{\text{min},DT}..t^{\text{end}}], s \in S, m \in M \quad (18)$$

where q is a subset of T .

4. Results of the Case Study (Cement Mills)

Within a case study the model is applied to the cement milling process of a cement plant with 9 sorts of cement (product types), two cement mills with 2.5 MW installed electric power each and a minimum up-time of 2 hours (see Figure 2). Two different pricing schemes are compared: A Time-of-Use (TOU) pricing scheme which comprises peak and off-peak rates and a RTP pricing scheme (hourly Day-Ahead Market price signal). The pre-process is the raw-mill, the α -storage is the clinker storage and the β -storages are the cement storages.

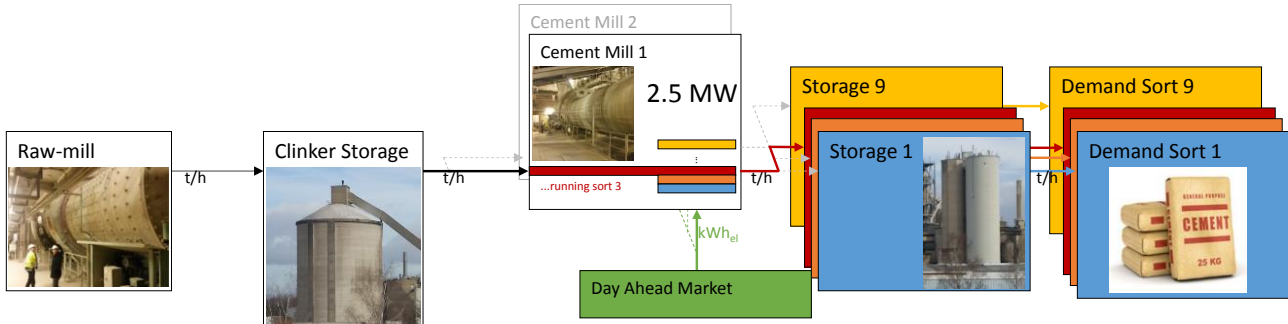


Figure 2: Case study cement milling process: Schematic representation of adopted DR-model (DR-relevant component and its connections through flow of electricity and goods).

Anonymised data of a German cement plant is used for TOU electricity prices, cement demand, production and storage capacities. Figure 3 shows the seasonal distribution of the cement demand over all cement sorts. With the cement demand peaking in summer it qualitatively fits with the annual course of photovoltaic generation (compare to Figure 4).

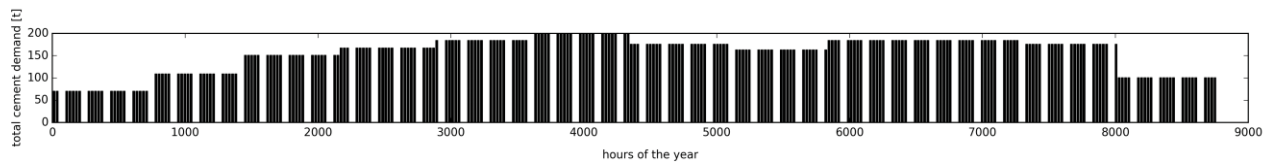


Figure 3: Bar chart of Seasonal distribution of total cement demand over all sorts

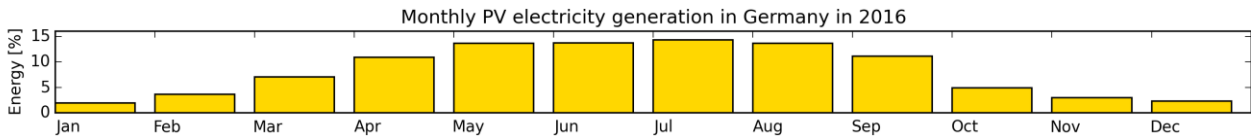


Figure 4: Bar chart of seasonal distribution of monthly PV electricity generation in Germany in 2016 (data from: energy-charts.de, 2017, EEX)

The problem is formulated in the AMPL modelling language. Using Gurobi 7.0.0 on a Laptop (2 · 2.50 GHz CPU and 11.9 GB available RAM) the problem with $|T| = 168$ can be solved within five minutes. The solution of the problem with $|T| = 8760$ is obtained on a server computer (12 · 2.40 GHz CPU and 126 GB available RAM) within one hour.

Figure 5 presents the results of the MILP for both pricing schemes. It shows the electric power consumption and the sorts produced at a time for both cement mills for one week (168 time steps of one hour). The week contains a public holiday on May 1st and a weekend. With the change from a TOU-based to a RTP-based machine scheduling, the resulting electrical loads are shifted from RTP peak hours to RTP off-peak hours. The graph shows that the TOU-based production program differs significantly from the RTP-based production program in the noon hours of business days and in times when RTP peaks coincide with the off-peak times of the TOU scheme (e.g. Saturday night).

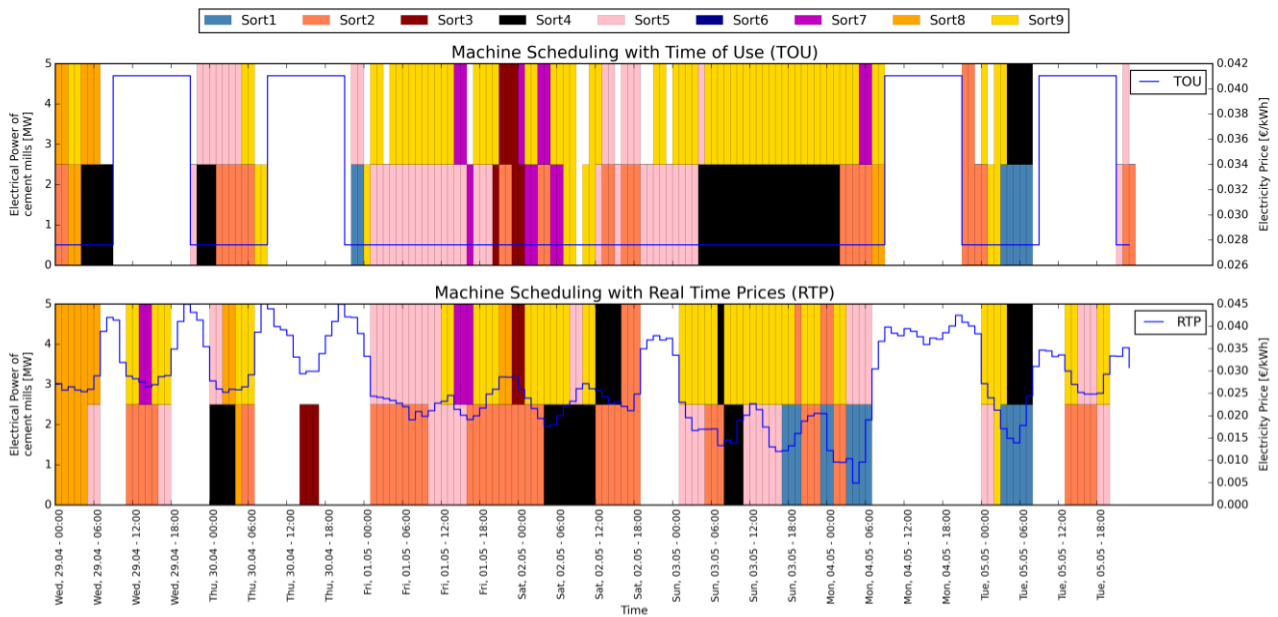


Figure 5: Comparison between TOU and RTP and according machine scheduling (TOU peak hours are from 8 a.m. to 8 p.m. on business days)

The Figures 6 and 7 show the results of the MILP for one year (8760 time steps) as heat maps. The upper heat maps show the pricing scheme, the lower heat maps the electric power consumption of both cement mills with the according optimized production plan.

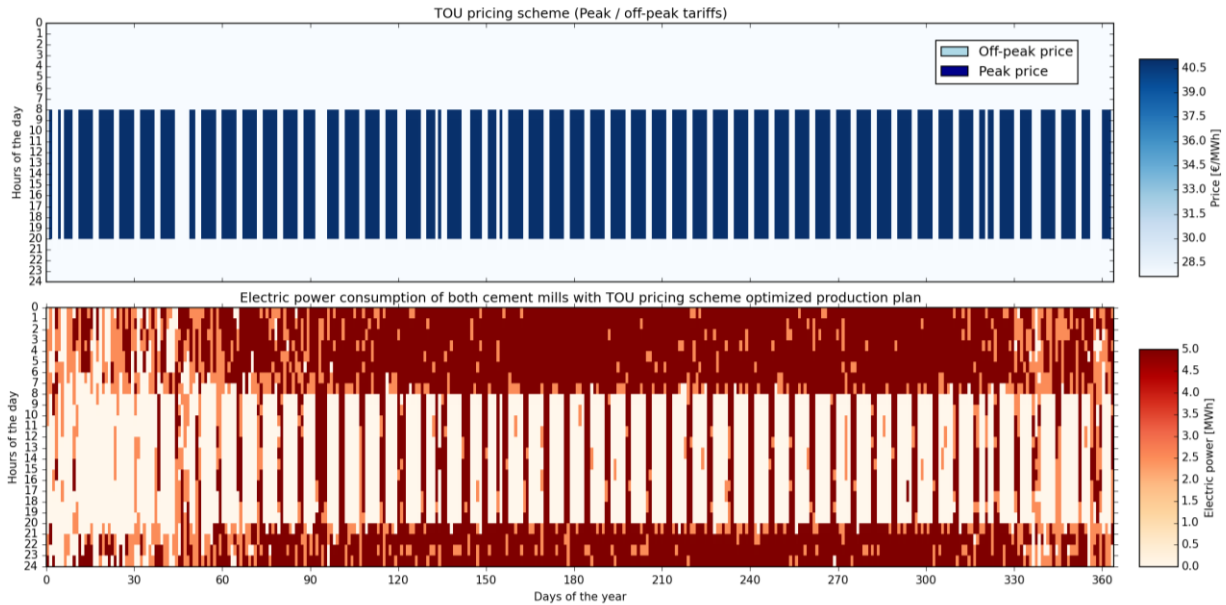


Figure 6: Peak / off-peak price scheme (top) and the according optimized machine scheduling (bottom)

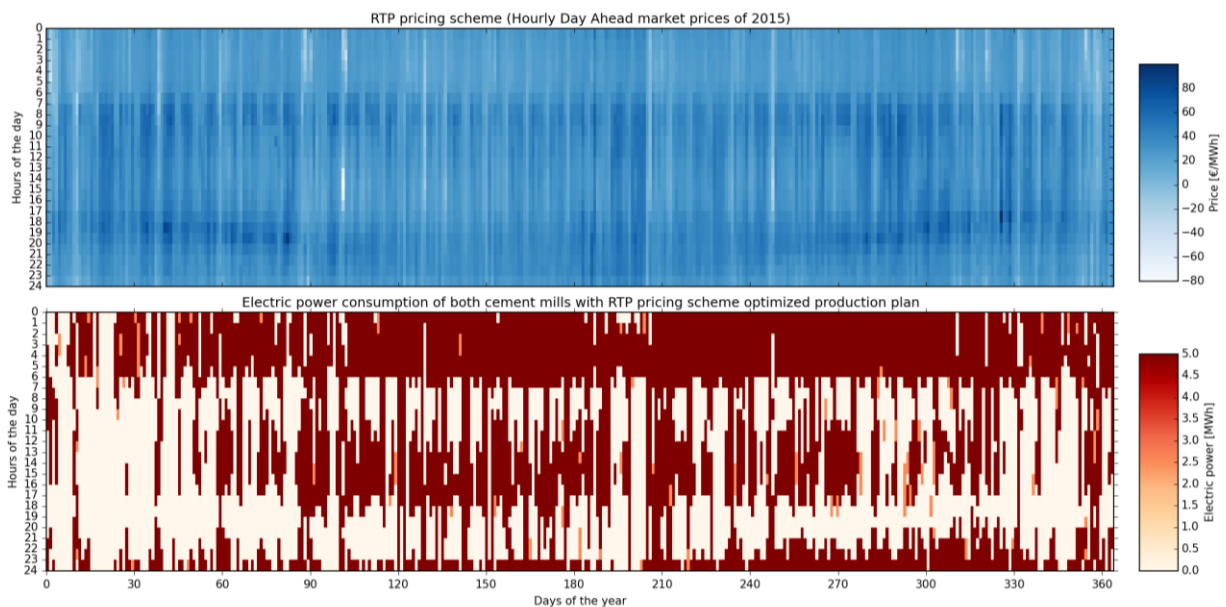


Figure 7: Day ahead electricity prices (top) and the according optimized machine scheduling (bottom)

Figure 8 shows the difference of both, the pricing schemes and the according production plans. In the lower heat map the load shifts from the TOU (purple) to RTP production plan (green) are illustrated. Over the entire year 5.8 GWh are shifted to times of lower day ahead market prices. That are 25 % of the annual energy consumption.

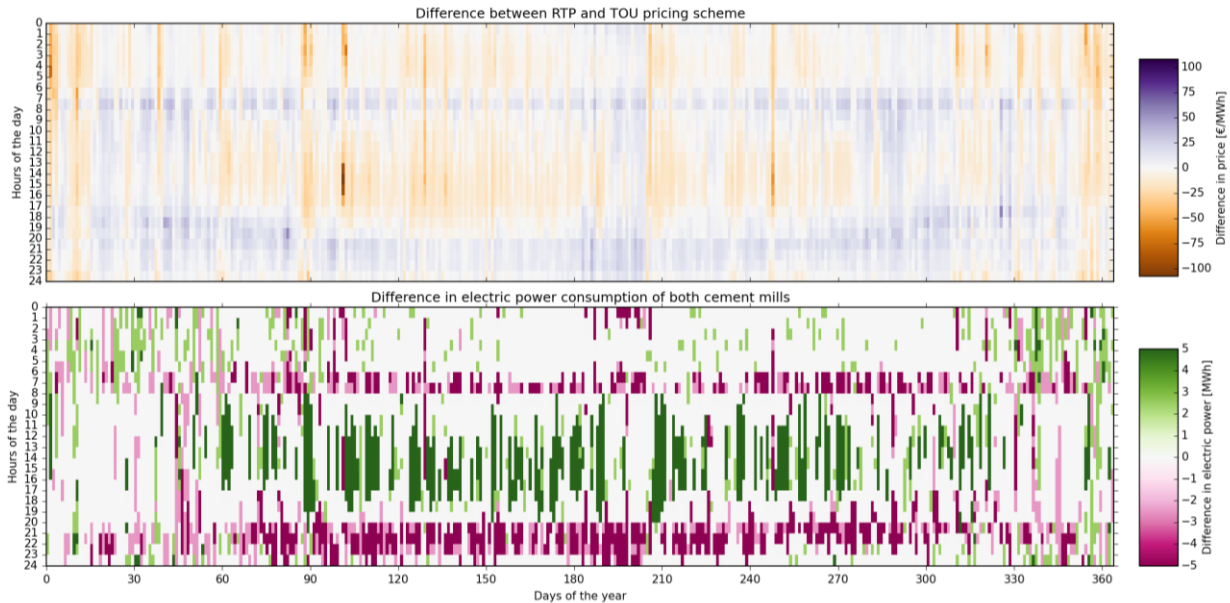


Figure 8: The difference between the RTP and TOU pricing scheme (top) and the difference between the according optimized production plans (bottom)

For this case study an hourly dissolved optimization of one year resulted in an electricity cost reduction of 13 % for the analysed process.

5. Conclusion and Outlook

Within this paper, a MILP formulation for price based DR was proposed and applied to a cement plant. Through integrated modelling of multiple machines and sorts, the formulation can be used as basis for dynamic machine scheduling.

The results revealed that through activation of a RTP based DR program to an electric load in the industry sector, one quarter of the energy could be shifted in a power grid-friendly way to times with low spot market prices and high shares of variable renewable energy sources without violating product-technical requirements. This approach leads to a win-win situation. On the one hand, flexibility of industrial processes could be used cost-effectively for the integration of renewable energies. On the other hand, energy costs for industrial companies could be lowered.

In future works, the model can be extended by other energy markets like the control reserve markets and be adapted to other applications in the industrial, commercial and residential sector.

Acknowledgment

This publication has been produced within the context of the Interflex4Climate project which is funded by the Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety of the Federal Republic of Germany (BMUB) as part of the National Climate Initiative.

References

- Asadinejad, A. and Tomsovic, K. (2017). "Optimal use of incentive and price based demand response to reduce costs and price volatility", In: *Electric Power Systems Research*, 144, pp. 215–223. Available at: <http://dx.doi.org/10.1016/j.epsr.2016.12.012> (Accessed: 23 October 2017).
- EURELECTRIC (2014). *Flexibility and Aggregation: Flexibility and Aggregation Requirements for their interaction in the market*.
- Gruber, A., Roon, S. von and Fattler, S. (2016). "Wissenschaftliche Projektbegleitung DSM-Bayern".
- Langrock, T., Achner, S., Jungbluth, C., Marambio, C., Michels, A., Weinhard, P., Baumgart, B. and Otto, A. (2015). *Potentiale regelbarer Lasten in einem Energieversorgungssystem mit wachsendem Anteil erneuerbarer Energien*.
- Larsen, E.M., Pinson, P., Leimgruber, F. and Judex, F. (2017). "Demand response evaluation and forecasting — Methods and results from the EcoGrid EU experiment", In: *Sustainable Energy, Grids and Networks*, 10, pp. 75–83. Available at: <http://dx.doi.org/10.1016/j.segan.2017.03.001> (Accessed: 23 October 2017).
- Molavi, H. and Ardehali, M.M. (2016). "Utility demand response operation considering day-of-use tariff and optimal operation of thermal energy storage system for an industrial building based on particle swarm optimization algorithm", In: *Energy and Buildings*, 127, pp. 920–929. Available at: <http://dx.doi.org/10.1016/j.enbuild.2016.06.056> (Accessed: 23 October 2017).
- R.A. Verzijlbergh, L.J. De Vries, G.P.J. Dijkema and P.M. Herder (2016). "Institutional challenges caused by the integration of renewable energy sources in the European electricity sector". Available at: <http://dx.doi.org/10.1016/j.rser.2016.11.039> (Accessed: 22 October 2017).
- Reka, S.S. and Ramesh, V. (2016). "Industrial demand side response modelling in smart grid using stochastic optimisation considering refinery process", In: *Energy and Buildings*, 127, pp. 84–94. Available at: <http://dx.doi.org/10.1016/j.enbuild.2016.05.070> (Accessed: 23 October 2017).
- Rodríguez-García, J., Álvarez-Bel, C., Carbonell-Carretero, J.-F., Alcázar-Ortega, M. and Peñalvo-López, E. (2016). "A novel tool for the evaluation and assessment of demand response activities in the industrial sector", In: *Energy*, 113, pp. 1136–1146. Available at: <http://dx.doi.org/10.1016/j.energy.2016.07.146> (Accessed: 23 October 2017).
- Shoreh, M.H., Siano, P., Shafie-khah, M., Loia, V. and Catalão, J.P.S. (2016). "A survey of industrial applications of Demand Response", In: *Electric Power Systems Research*, 141, pp. 31–49. Available at: <http://dx.doi.org/10.1016/j.epsr.2016.07.008> (Accessed: 23 October 2017).
- Steurer, M., Klempp, N. and Hufendiek, K. (2015). "Identifikation und Realisierung wirtschaftlicher Potenziale für Demand Side Integration in der Industrie in Deutschland". Management Summary.
- Wang, X., El-Farra, N.H. and Palazoglu, A. (2017). "Optimal scheduling of demand responsive industrial production with hybrid renewable energy systems", In: *Renewable Energy*, 100, pp. 53–64. Available at: <http://dx.doi.org/10.1016/j.renene.2016.05.051> (Accessed: 23 October 2017).

Multi-objective optimization of a waste heat driven heating supply system using Mixed-Integer-Programming

M. Bohlayer, M. Fleschutz, A. Bürger, M. Braun
Faculty of Management Science and Engineering
Karlsruhe University of Applied Sciences , Germany
e-mail: markus.bohlayer@hs-karlsruhe.de

Keywords: Design Optimization, Waste Heat, Heating Supply System

Introduction

Industrial waste heat can be used as an energy resource in distributed local heating networks. In many cases waste heat could be recuperated from hot exhaust air/gas streams using heat exchangers. Due to the volatile availability of waste heat and heating demands, the application of heat storages increases the shares of waste heat in the energy supply and leads therefore, both from an economic and ecologic point of view, to an advantageous system. Fluctuating heating demands, waste heat flow rates and heat loss as well as pressure loss occurring in the heating network make the design of the system a challenging task. Provided predetermined system temperatures, the optimal design variables (e.g. heat exchanger area, storage volume and boiler capacity) can be identified solving a Mixed-Integer-Linear-Programming (MILP) problem. However, taking system temperatures and mass flow rates as variables into account could have a major impact on the layout and therefore the total annual cost of the entire system. The consideration of system temperatures introduces nonlinearities to the model, which can be difficult to treat directly in the optimization problem.

Literature

There is a large body of literature on the design and operation of Distributed Energy Resources (DER) and the authors refer to Connolly et al. (2010), Karger and Hennings (2009) and Pepermans (2005) for comprehensive reviews. A lot of the recently published literature focuses on the optimization of the design of cogeneration technologies e.g. Li et al. (2016) or the integration of renewable energies e.g. Buoro et al. (2013). To the knowledge of the authors, there are currently no studies considering waste heat as an energy resource in DER planning. There exist few works tackling the problem of occurring nonlinearities in thermal energy system optimization, most of the works simply consider energy demands and energy supply in terms of kilo-watt-hours. Yokoyama et al. (2017) lately presented a model for the optimal operation of a heat supply system. They consider the nonlinearities in the fundamental equations and convert the nonlinear problem into a MILP using binary decision variables and discretization of mass flow rate.

Specifying the problem

Considering a heating network with a given layout (e.g. pressure loss and heat loss characteristics) and load profiles of waste heat as well as heat demand, determine the design (area of heat exchanger, storage volume, heat supply temperature) and operation (mass flow rates, temperatures, heat transfer) of a waste heat utilization centre, as represented in Figure 1, that minimizes the total annual cost and the CO₂ emissions.

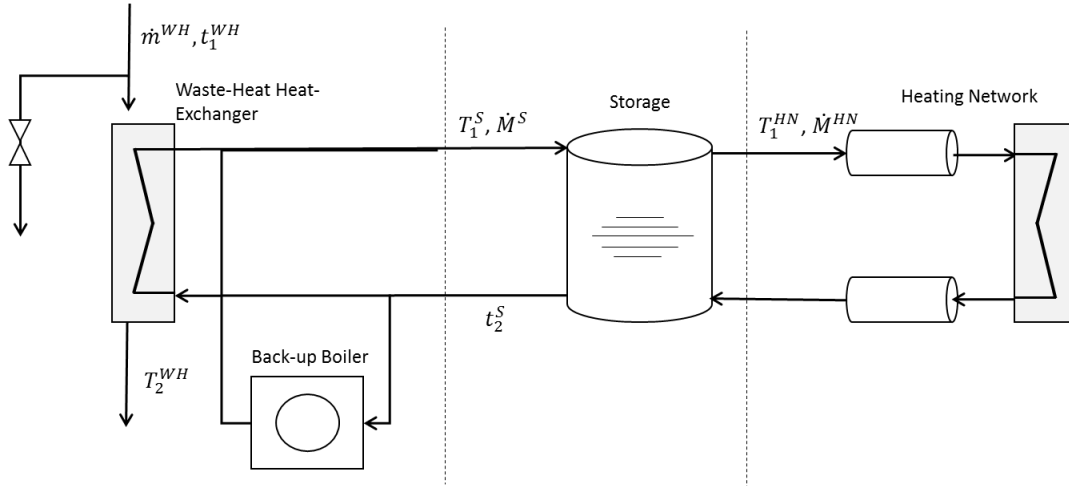


Figure 1: Schematic of the Waste Heat

Approach

The problem can be formulated as a Mixed-Integer-Nonlinear-Programming problem and has two objective functions.

The economic objective function consists of operating cost and annualized investment cost. Operating cost are determined by the fuel burned in the Back-up Boiler over all time periods. The investment cost are the sum of investment cost of all technologies. The investment cost of the heat exchanger is determined by the area of the heat exchanger, the cost of the heat storage by its volume and the cost of the boiler by its capacity.

The environmental objective function consists of the CO₂ emissions due to the natural gas consumption, which is determined by the carbon intensity of the fuel and the total amount of gas burned in the boiler.

To solve this multi-objective optimization problem one objective function is formulated as a weighted sum of the economic objective function and the environmental objective function.

The Waste Heat Utilization centre is represented by physically motivated grey box models which are formulated within the problem constraints. The most important ones are listed below:

1. The system supply temperature needs to be in a specific range to ensure heat transfer at local supply stations.
2. The network layout structure determines the range of possible mass flow rates.
3. The supply system meets the occurring demands.
4. The state of charge of the storage needs to be lower than its capacity and bigger than zero.
5. The heat flow in and out of the storage need to be within specific boundaries.
6. The pumping power is proportional to the mass flow rate.

7. The storage capacity is determined by its volume and the temperature difference of supply and return temperature.
8. The back-up boiler needs to work within its capacity limits.
9. The fuel burned in the boiler is calculated as a function of the supply and return temperature.
10. The heat supplied by the heat exchanger is a function of waste heat flow rate and temperature difference of the waste heat.
11. The heat supplied by the heat exchanger is a function of the area of the heat exchanger and the logarithmic mean temperature of the inlet and outlet streams.

Considering the supply temperature and the outlet temperature at the heat exchanger as variables leads to nonlinearities in constraints (7), (9), (10) and (11). As mentioned before it can be very time-consuming to solve nonlinear optimization problems and global optimality cannot be guaranteed (Yokoyama et al. 2017 and Ommen et al. 2014). Therefore we linearize the non-linear terms (7-11) by the introduction of temperature intervals (discretization) and binary decision variables Y_i as illustrated in Figure 2. Each decision variable indicates whether heat is transferred on the corresponding temperature interval or not. The non-linear relationship in constraint (6) is

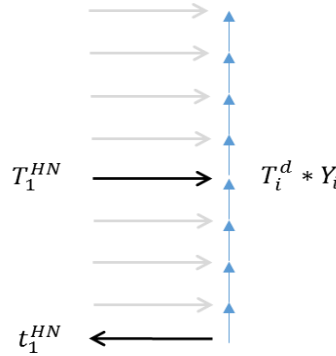


Figure 2: Schematic representation of temperature discretization

reformulated using piecewise linearization. This approach enables us to convert the model to a mixed integer linear programming problem. Therefore it can be guaranteed to be solved to global optimality and solved in reasonable time. The problem is formulated using the algebraic modelling language AMPL and is solved with GUROBI 7.0.0.

Preliminary results

Contrary to the current trend of decreasing supply temperatures in heating networks we can show that in local heating networks predominantly supplied by waste heat it can be favourable to increase the supply temperature. The effect of decreasing cost of investment for the heat storage and decreasing electricity cost for pumping outweighs the effects of increasing cost of investment for the heat exchanger and increasing temperature loss. Further research could include the integration of additional waste heat utilisation technologies such as heat pumps or Organic Rankine Cycles. The performance of these technologies is dependent on the occurring temperatures in a non-linear manner and could therefore be modelled within the framework of a MILP using the approach presented in this paper.

Acknowledgement

This work was funded by the Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety within the scope of the National Climate Initiative as part of the project INTERFLEX.

References

- Connolly, D., Lund, H., Mathiesen, B. and Leathy, M. (2010). A review of computer tools for analysing the integration of renewable energy into various energy systems. In: *Applied Thermal Engineering*, 87, pp. 1059–1082.
- Karger, C. and Hennings, W. (2009). Sustainability evaluation of decentralized electricity generation. In: *Renewable and Sustainable Energy Reviews*, 13(3), pp. 583 – 593.
- Pepermans, G., Driesen, J., Haeseldonckx, D., Belmans, R. and D'haeseleer, W. (2005). Distributed generation: definition, benefits and issues. In: *Energy Policy*, 33(6), pp. 787 – 798.
- Li, M., Hailin M., Nan L. and Baoyu M. (2016). Optimal design and operation strategy for integrated evaluation of CCHP (combined cooling heating and power) system. In: *Energy*, 99, pp. 202 – 220.
- Buoro, D., Casisi, M., De Nardi, A., Pinamonti, P. and Reini, M. (2013). Multicriteria optimization of a distributed energy supply system for an industrial area. In: *Energy*, 58, pp. 128 – 137.
- Yokoyama R., Kitano H., Wakui T. (2017). Optimal operation of heat supply systems with piping network. In: *Energy 2017*, doi: 10.1016/j.energy.2017.03.146.
- Ommen T., Markussen W. B., Elmegaard B. (2014). Comparison of linear, mixed integer and non-linear programming methods in energy system dispatch modelling. In: *Energy*, 74, pp. 109-118.

An Exploration of Experiential Marketing in the Fast Moving Consumer Goods Industry

Holly Barry, Rose Leahy, Pio Fenton
Department of Marketing and International Business
Cork Institute of technology Ireland
Email: holly-barry@mycit.ie

Keywords: Experiential Marketing, Customer Experience, Marketing, Fast Moving Consumer Goods, Return on Experience, Retailing, Marketing Metrics.

Introduction

Research Question: How do companies in the fast moving consumer goods industry implement experiential marketing, integrate it with other marketing methods and measure its impact?

Using a qualitative approach, this research will specifically examine the use of experiential marketing as a marketing communication's technique in the FMCG industry. Given the rise in popularity in the use of experiential marketing to communicate with a target market (Moderne Communications, 2014), it is timely to explore the implementation of this strategy in one of the main sectors where it is used, and examine its successes and failures therein. This research seeks to draw conclusions on why it is a chosen communication tool, how it is used and the Return on Experience (ROE) of experiential marketing in the FMCG industry.

Research Approach

The most basic definitions point to experiential marketing as a tactic which is utilised by a business to combine an experience with elements of entertainment into the offering, with the intention of reinforcing the product or service in the consumer's mind (Lee *et al.*, 2011; Atwal and Williams, 2009; Yuan and Wu, 2008; Pine and Gilmore, 1999). Companies that practice experiential marketing take a brand's essence and implement it in the form of an event, experience or interaction. This in turn will allow consumers to understand the brand at an intimate level by being active, rather than passive, participants (Gautier, 2003). A truly experiential event is not one that is simply branded with a company or branded name, but rather, an event which is created for and is owned by the brand or company (Gautier, 2003).

Many industries have witnessed a shift from traditional features and benefits marketing towards creating experiences for their customers (Pathak, 2014). Ernst and Young (2013) conclude that customer requirements have evolved and customers are now actively searching for marketing campaigns that deliver an experience. Organisations that have progressed to this method of marketing hold the view that their target audience should be involved in an active experience which portrays brand values (Strategic Direction, 2008). Rather than communicate through traditional methods, they create a brand environment in a public space and involve potential customers by encouraging them to take part in an activity (Strategic Direction, 2008). The main objective of these experiences is to generate strong feelings that the consumer may recall (Schmitt, 1999b). Consumers want products, communications, and campaigns that they can not only relate to but also incorporate into their lifestyles (Schmitt, 1999a).

Research in this area largely focuses on the consumer's experience of this marketing activity in addition to the methods of experiential marketing. Numerous research papers highlight the

consumer's modern day preference of brand engagement through experiential methods (for example, You-Ming, 2012; Smilansky, 2009 and Choou, 2009). However, from a review of the literature it can be stated that there is a lack of empirical qualitative data available on the advantages of experiential marketing to the company. Moreover, there is a distinct lack of empirical information on why experiential marketing is a chosen communication tool, how it is used and the return of experience (ROE) of experiential marketing in the FMCG industry.

ROE has emerged as an alternative to ROI for many professionals in marketing. The concept of Return of Investment (ROI) is losing relevance due to the ever-changing nature of marketing (Harris, 2016). This is primarily due to the digital revolution. Through the medium of digital channels, powerful tools have arisen which have created highly personalized and emotive relationships between a brand and their consumer which are measured through ROE (Frawley, 2015). ROE represents a long-term measurement of consumer satisfaction comprising of consumer experiences, interactions and reactions (Lenderman, 2005).

While some quantitative data exists, qualitative data pertaining to the FMCG industry is particularly lacking.

Research Methodology

Adopting a qualitative approach, in-depth, semi-structured interviews will be conducted with marketing and brand managers in leading FMCG companies. This research will identify the objectives and use of experiential marketing campaigns, the integration of experiential marketing with traditional marketing activities and if the set objectives are achieved. It will also provide information on the outcome of this method of marketing communication; specifically, in relation to how it performs and the ROE. Resulting from these outcomes, the research will highlight recommendations on the objectives and implementation of experiential marketing campaigns for both further academic research and marketing practice.

The sampling technique will be purposive. (Maxwell, 1997) defined purposive sampling as a type of sampling in which "particular settings, persons, or events are deliberately selected for the important information they can provide that cannot be gotten as well from other choices" (Maxwell, 1997). This method of research is preferred given the objective of investigating the effectiveness of experiential marketing in the FMCG sector. The individuals chosen to participate in the research will be marketing managers and brand managers in various companies in the FMCG sector. It is proposed that fifty managers will be interviewed using a semi- structured interview guide.

Preliminary Results

Although this research is still in the exploratory phase, it is proposed that when complete it will offer empirical insight into the heretofore under-researched area of experiential marketing in the FMCG industry. The research aims to highlight how companies successfully implement experiential marketing campaigns, how they integrate it with their other marketing activities and how they measure the campaign outcomes through ROE and other methods. In this way, the research will contribute to both academia and practice alike.

References

Atwal, G., Williams, A. (2009). Luxury Brand Marketing: The Experience is Everything!. *Journal of Brand Management* 16(5/6), pp. 338-346.

Choo, H.-J. (2009). The Effect of Experiential and Relationship Marketing on Customer Value: A case study of international American casual dining chains in Taiwan. *Society for Personality Research* 37(7), pp. 993-1008.

Ernst and Young (2013). The Journey Toward Greater Customer Centricity, Ernst and Young.

Frawley, A. (2015). *ROI is Dead. A New Metric is needed for Customer Relationships*. Available at <http://adage.com/article/digitalnext/brands-measure-experience-engagement/297426/> [Accessed 24 October 2016].

Gautier, A. (2003). Think Again: Why Experiential Marketing is the Next Big Thing. *New Zealand Marketing Magazine* 22(8).

Harris, M. (2016). Return on Experience' is the new ROI. Available at <http://digitalmarketingmagazine.co.uk/customer-experience/return-on-experience-is-the-new-roi/3315> [Accessed 24 October 2016].

Lee M.S. et al. (2011). The Study of Relationships Among Experiential Marketing, Service Quality, Customer Satisfaction and Customer Loyalty. *The International Journal of Organizational Innovation* 3(2), pp. 353-379.

Lenderman, M. (2005). Screw ROI. ROE is what really matters. Available at <https://maxlenderman.wordpress.com/2005/06/17/screw-roi-roe/> [Accessed 23 March 2017].

Maxwell, J. (1997). Designing a qualitative study. In: L. Bickman & D. J. Rog eds. *Handbook of applied social research methods*, Thousand Oaks, CA: Sage, pp. 69-100.

Moderne Communications (2014). Experiential Marketing Trends: 1-6.

Pathak, S. (2014). Just What Is Experiential Marketing, and How Can It Be Measured?. Available at <http://adage.com/article/agency-news/experiential-marketing-measured/291972/> [Accessed 1 November 2016].

Pine, J., Gilmore, J. (2011). The Experience Economy. *Harvard Business Press*.

Schmitt, B. (1999b). Experiential Marketing. *Journal of Marketing Management*, 15: 53-67.

Smilansky, S. (2009). Experiential Marketing: A Practical Guide to Interactive Brand Experiences, Kogan Page.

Strategic Direction (2008). Experiential Marketing: Advice on the potential and pitfalls of a growing trend. *Strategic Direction*, 24(10), pp. 24-26.

You-Ming, C. (2010). Study on the Impacts of Experiential Marketing and Customers' Satisfaction based on Relationship Quality. *The International Journal of Organizational Innovation* 3(1), pp.189-209.

Yuan, Y.H., Wu, C. (2008). Relationship Among Experiential Marketing, Experiential Value and Customer Satisfaction. *Journal of Hospitality and Tourism Research* 32(3), pp. 387-410.

Judging Competencies, Abilities and Expertise based on Enterprise Social Network Profile Pages

B. Schemmann [†], S. Behrendt ^{††}, S. Kinkel [†]
[†] ILIN Institute for Learning and Innovation in Networks
 Karlsruhe University of Applied Sciences, Germany
^{††} Cooperation Systems Center Munich (CSCM)
 Bundeswehr University Munich, Germany
 e-mail: brita.schemmann@hs-karlsruhe.de

Keywords: enterprise social network, profile page, expert search

1. Introduction and Problem Statement

The raise of the Internet and the social web has not only created new opportunities for people to maintain existing social ties but also to establish new relationships between people who did not know each other before (Guy, Ur, Ronen, Perer, & Jacovi, 2011). Social networking services (SNS) – in the form of enterprise social networks (ESN) – have also been introduced within companies and are becoming increasingly important for companies' knowledge management and the employees' communication and networking (Richter & Koch, 2008a). A study among German companies shows that 39% are using social networking services for their daily collaboration (Richter, 2014). Research has also shown that employees use ESN to stay in touch with close colleagues, to establish stronger bonds with their weak ties and to even broaden their networks by reaching out to colleagues they did not know before (DiMicco et al., 2008). In line with this, research has identified "expert search" to be one of the six core functionalities of social networking services (Richter & Koch, 2008a) as this is often required in the daily tasks of employees.

Important for the search for new contacts who possess certain competencies, abilities or expertise are the employees' personal profile pages within an ESN. Enterprise Social Networks normally enable the employees to manage a profile page that includes the employee's name and often photo, contact details and some basic information regarding the job role and possibly the employee's project history. While this basic information might be enough when looking for a colleague you already know, more information is needed when looking for "strangers" (Guy et al., 2011), e.g. when looking for potential members for a new project team. Hence ESN often provide the opportunity to include a range of other information concerning the profile owner's expertise, activities and relationships. A large amount of these information is not generated automatically (e.g. based on the activities in the ESN), but requires manual entry and updates. Maintaining detailed, reliable and up-to-date personal profile pages can thus be a very labour-intensive and expensive task (Braun, Kunzmann, & Schmidt, 2012; Koch et al., 2007). Hence the question arises: *Which of the information on an employee's profile page is used to judge whether the profile owner possesses certain (1) competencies, (2) abilities or (3) expertise?* In this context, competencies can be understood as the individual dispositional capability and readiness to act successfully and self-organized when facing novel, unstructured or complex situations or tasks. Competence is hereby based on the activation, combination and use of personal resources that can be developed (such as knowledge, networks, cognitive and practical abilities) as well as on social aspects or behaviour but not on a person's character. Cognitive and practical abilities as well as expertise based on knowledge are therefore necessary for the development of competencies (Kinkel et al., 2017). Or in other words: Knowledge or expertise enables employees to develop certain abilities. When these abilities are employed in a productive manner you gain relevant competencies (North 2002).

Based on former studies (Koch, Richter, & Schlosser, 2007; Richter & Koch, 2008b; Yarosh, Matthews, & Zhou, 2012; Yimam-Seid & Kobsa, 2003; Zhang, Tang, & Li, 2007) and on our analyses of such networking services we developed a framework of the different types of information that are likely to be provided on an employee's profile page (shown in table 1).

Table 1: Different types of information that can be provided on a SNS personal profile page

Explicit information		Implicit information	
self-declared	verifiable	activity-based	relationship-based
<ul style="list-style-type: none"> - expertise offered - (top-)skills offered - languages spoken - professional interests - personal interests (hobbies, etc.) - personal "mission" statement 	<ul style="list-style-type: none"> - current job/position - past jobs/positions - time at current and past position(s) - international experience - educational background - voluntary work - membership in organisations - awards - references - [endorsements of skills by other members of the SNS*] 	<ul style="list-style-type: none"> - group/themes followed* - [activity in groups*] - roles taken in groups (e.g. moderator)* - [projects worked on*] - events attended* - activity-level in SNS* - [amount of content produced*] 	<ul style="list-style-type: none"> - number of contacts* - diversity of contacts* - contacts of both the profile visitor and profile owner* - profile pages in other SNS

*dynamic information: (semi-)automatically generated based on employees' activities in the SNS.

[] Information in squared brackets was not a feature of the platform assessed.

2. Methodology

To gain first insights into the research question we use an exploratory research design, namely an informal qualitative approach. We therefore used the SNS Xing. The platform is a large career-oriented SNS that is especially popular in Germany and the German-speaking countries (Koch et al., 2007). Given the business and career-related context in which Xing operates, we consider the profiles provided in this SNS to be suitable for our analyses as they are comparable to profile pages in common ESN. Moreover the profiles provided in Xing can provide a wide variety of different sources of information that we are interested in (see table1). We therefore invited seven members of the project *ChampNet* who work in either academics or for one of the industry partners involved in the project to each take a detailed look at seven previously selected profile pages in Xing. The random seven profile pages were chosen based on their information richness on the one hand and the similarity of profile owners on the other. We made sure that none of the participants previously knew any of the profile owners.

The participants were asked to judge the profiles regarding the following questions: To what extent does the profile owner probably possess the following competencies: network competency, creative problem-solving competency, overview competency, integration competency and project management competency? To what extent does the profile owner probably possess the following abilities: ability to work in a team or under pressure, to deal with conflicts, to adapt quickly and flexibly to new (work-related) tasks and challenges? In which areas does the profile owner probably possess special expertise? For each question the participant had to explain which sources of information he or she used to make this judgement. The participants were all familiar with the different competencies and abilities, as the development of competencies is a core element of the *ChampNet* project. The content of all feedback was then analysed and coded to determine which information was used for the different competencies and abilities and to assess the profile owners' expertise.

3. First Insights

First results indicate that for the assessment of a profile owner's competencies, abilities and expertise different sources of information are likely to be considered and interpreted. First, we see that information regarding the current and past jobs and positions was heavily used to judge the profile owner's competencies, abilities and expertise. The same is true for information regarding the educational background. Especially detailed job descriptions and information concerning the duration of the different jobs were considered to be very helpful. Second, we see some differences in the use of explicit and implicit information: (1) to assess the profile owner's competencies all four sources of information (self-declared, verifiable, activity-based and relationship-based) were considered, (2) to assess the abilities only the two sources of explicit information were used, with a strong emphasis on information concerning personal interests and voluntary work as well as the awards received and the international experience gained, (3) and to assess the profile owner's expertise again only explicit information were used, especially self-declared information regarding expertise offered, languages spoken and areas of professional interest.

To analyse the competencies, abilities and expertise of a person solely based on his/her social network profile page different sources of information are used. Much attention seems to be paid to the profile owner's professional and educational background and some effort should be made to offer up-to-date and also detailed information on this. Moreover it seems as if self-declared information concerning the profile owner's expertise and interests are generally considered to be a useful and to some extent trustworthy source of information when judging his/her expertise. Finally, we see that sources of implicit information are only used to judge a person's competencies. Given the fact that competencies need to be seen in the context of a person's actions (North, Reinhardt, & Sieber-Suter, 2013) this makes sense. However, this activity and relationship-based information could also be useful to judge a profile-owners expertise (e.g. by looking at the number of substantial contributions made in a discussion group). Yet, to gain such insights a system needs to be in place that aggregates such information in meaningful way.

4. Acknowledgement

This research has been carried out as part of the research and development project *ChampNet*, which is funded by the German Federal Ministry of Education and Research (BMBF) within the Programme "Innovations for Tomorrow's Production, Services, and Work" (funding number 02L12A280 ff.) and managed by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the content of this publication.

References

- Braun, S., Kunzmann, C., & Schmidt, A. P. (2012). Semantic people tagging and ontology maturing: an enterprise social media approach to competence management. *International Journal of Knowledge and Learning*, 8(1-2), pp. 86-111.
- DiMicco, J., Millen, D. R., Geyer, W., Dugan, C., Brownholtz, B., & Muller, M. (2008). *Motivations for social networking at work*. Paper presented at the Proceedings of the 2008 ACM conference on Computer supported cooperative work.
- Guy, I., Ur, S., Ronen, I., Perer, A., & Jacovi, M. (2011). *Do you want to know? Recommending strangers in the enterprise*. Paper presented at the Proceedings of the ACM 2011 conference on Computer supported cooperative work.
- Kinkel, S., Schemmann, B., & Lichtner, R. (2017). Critical Competencies for the Innovativeness of Value Creation Champions: Identifying Challenges and Work-integrated Solutions. *Procedia Manufacturing*, 9, pp. 323-330.
- Koch, M., Richter, A., & Schlosser, A. (2007). Produkte zum IT-gestützten Social Networking in Unternehmen. *Wirtschaftsinformatik*, 49(6), pp. 448-455.

- North, K (2002): *Wissensorientierte Unternehmensführung. Wertschöpfung durch Wissen*. 3. Auflage. Gabler, Wiesbaden.
- North, K., Reinhardt, K., & Sieber-Suter, B. (2013). *Kompetenzmanagement in der Praxis. Mitarbeiterkompetenzen systematisch identifizieren, nutzen und entwickeln*. Wiesbaden: Springer Gabler.
- Richter, A. (2014). *Vernetzte Organisation*. München: De Gruyter Oldenbourg Wissenschaftsverlag.
- Richter, A., & Koch, M. (2008a). *Functions of social networking services*. Paper presented at the Proc. Intl. Conf. on the Design of Cooperative Systems.
- Richter, A., & Koch, M. (2008b). *Funktionen von Social-Networking-Diensten*. Paper presented at the Multikonferenz wirtschaftsinformatik.
- Yarosh, S., Matthews, T., & Zhou, M. (2012). *Asking the right person: supporting expertise selection in the enterprise*. Paper presented at the Proceedings of the sigchi conference on human factors in computing systems.
- Yimam-Seid, D., & Kobsa, A. (2003). Expert-finding systems for organizations: Problem and domain analysis and the DEMOIR approach. *Journal of Organizational Computing and Electronic Commerce*, 13(1), pp. 1-24.
- Zhang, J., Tang, J., & Li, J. (2007). *Expert finding in a social network*. Paper presented at the International Conference on Database Systems for Advanced Applications.

Practice-oriented Training as a Sales Engineer: Case Study Analysis of a Role Play

M. Murzin, M. Ben Amor

Faculty of Management Science and Engineering,
Hochschule Karlsruhe Technik und Wirtschaft, Karlsruhe, Germany,
e-mail: marion.murzin@hs-karlsruhe.de
e-mail: maha.ben-amor@hs-karlsruhe.de

Keywords: Practice-oriented Training, Sales Engineer, Role Play

The Concept of the Course

The engineering industry is one of the leading sectors in Germany that are internationally known for high-quality and innovative products with the "Made in Germany" label. Thus, the studies of engineering sciences are inherently popular receiving a great interest from professors as well as students. However, developing new and innovative products represents one side of success that must be complemented by suitable promotion and sales actions. Throughout their professional career, engineers are oftentimes responsible for the latter though they are not sufficiently qualified to derive appropriate marketing or sales strategies. Therefore, selling technical products must be learned in the same way the technical foundations for engineering are taught. Consequently, being able to successfully merge these two fields' results in higher profitability for the company.

For this purpose, the Karlsruhe University of Applied Sciences allows students to study the field of Business Administration and Engineering. In specialized courses, the students are given the right tools for a good technical sales engineer. The concept of the training is based on the B2B sales fundamentals, in which field sales representatives have to meet buyers, production managers, safety engineers or other contact persons.

The training includes the following areas:

- Psychological foundations in sales.
- The sale's phases: Demand analysis, obstruction treatment and closure techniques, Practice-oriented role plays and simulated purchases.
- Acquisition via telephone in a professional call center.
- Accompaniment of an external technical sales force in the technical area.

The Role Play

The role play behavior is reflected and practiced in different social situations. Hoffmann, B. and Langefeld, U., (1998), defined the role play as "a kind of preparation for the reality in "acting-as-if" or an attempt to put themselves in the role of others".

Recently, Schaefer, A. and Haytko, D. L., (2014) advanced that the role plays, as a teaching methodology, offer students an opportunity to experience selling different types of products and services as well as to different types of customers (e.g., final consumers, retailers, manufacturers, etc.). By participating in realistic role plays from various areas of the selling profession, students are in a better position to select sales positions in which they will potentially be successful. Their experience allows them to more accurately predict how real

world buyers would respond to particular questions and situations, adding to the realism of the role plays.

The role play has different intentions:

- Adaptation to realities and behavior patterns.
- Breaking conventional role behavior.

A characteristic feature of the role play is the fact that insights into personal behaviors and their changes are in the foreground. Accordingly, role playing can be used in different phases.

Practical Case

Who? This class experience is based on a simulation concept. Here, students get the opportunity to develop role-plays for training purposes on B2B sales talk. The participants are the High School of Karlsruhe's students from the 7th semester in a sales technics' course.

How? Workshops simulated with an industry office based on the followed topics: exchanges, formulation exercises, role-plays, practical exercises with video equipment (computer equipped with a camera and a microphone).

How Long? This approach of teaching in sales education is shared by 14 groups of 2 students during 4 months. Each role play takes approximately 5 minutes.

In order to give real effect to the conversation, every buyer (student) must follow many instructions in his interview with the seller student, such as price objection, incorrect features; product is not as expected, etc. Moreover, to avoid arbitrariness, some evaluation criteria should be established before the start of the activity and after the role play, for example we asked the buyer (our student) : " if that were the case you would buy?", and the seller (our student) and the rest of the class: "do you have any comment?". The assessment, the self-assessment and the clarification of criteria related to objectives will also help the teacher in his assessment and the return on the activity. Some evaluation methods may allow us to situate the student in their learning.

Sequence Analysis: In order to ensure an effective case analysis, we have selected the video that we consider the most illustrative and the most significant.

- Customer complains about the late coming of the seller.
- Seller answers that the traffic was the reason.
- Customer complains that the seller has not called ahead.
- The seller wants to sit. Customer offers him a place.
- Customer: the offer does not suit me. The last salesperson offered a wrong demonstration example color (0'58'').
- Seller asks what would not be exactly right.

The Customer shows dissatisfaction. The Seller remains calm, sitting at the table, but only on the front seat (escape position). The pattern is in the center of the table, prompting the customer to approach again

- Customer wants a powerful phone camera. The price is always away from his wishes (1'18'').
- The seller takes notes and answers to the customer objections. He explains the price differences and shows off the different services.

The seller retains the same sitting position.

- Customer replies that he would not assume that the Smartphone would go immediately broken.
- The seller: It depends to the maintenance.
- The customer explains again that he needs a reliable cell phone.

The seller looks - when he speaks - too often down and holds no eye contact with customers. The customer moves.

- The customer adds again that he would like that in the future the current salesperson comes and not another person (3'22'').
- The seller summarizes the requirements of the clients.
- The customer says that otherwise the offer would suit him.
- The seller summarizes the possibilities and announces a new offer, with about the same price.

Summary: The customer was very upset. Nevertheless, the seller remains calm and discusses the reasoning to convince the customer. He remains at the table, even if his chair is moved away. It lacks the seller to make eye contact with the customer. Despite the intense criticism of the customer, the seller manages to offer a new offer according the same attributes and the same price. The criticism of the high price is now exceeded. The seller, however, failed to agree on another appointment with the client to discuss the offer proposed. Overall, the chances are good that the customer buys. (It has also been confirmed by the customer so).

Discussion

The aim of this original course's concept is to provide students with a comprehensive theoretical and practical knowledge. The theoretical background is based on different aspects relating to the transmission of psychological sales foundations. The various sales stages are carried out in the practice in role-playing simulations. For this purpose, the students have to start the semester with the choice of a technical product that can be sold in the B2B sector. The student have to definite the products' characteristics, advantages and benefits and to imagine different competitive skills.

From many feedbacks of our students, we chose this student's testimony of the last summer semester: *"Die Rollenspiele sind eine gute Ergänzung zur theoretischen Wissensvermittlung. Insbesondere die Analyse der Videoaufzeichnungen bringen einen großen Lerneffekt mit sich."*

Conclusion

The Role play is an important element of all simulation-based procedures. An improved understanding for methodological aspects facilitates its goal-oriented use in education and training. In the present paper we have described how different a specific form of role play is used in different a sale simulated laboratory. We have discussed many skills demonstrating how may the introduction of role-playing increase perceived realism.

References

- Hoffmann B. and Langefeld U. (1998). Methoden-Mix Unterrichtliche Methoden zur Vermittlung beruflicher Handlungskompetenz in kaufmännischen Fächern, Darmstadt.
- Schaefer, A. and Diana L. Haytko (2014), "Corporate partnering for role play competitions in an advanced selling course," *Journal of Instructional Pedagogies*, Vol. 14 (March); pp. 1-8.
- Murzin M. (2017). Sales Engineering Course, International Management, Faculty of Management Science & Engineering, Hochschule Karlsruhe, Germany.

Measuring and Visualizing the value added through the Industrial Internet of Things

Dominik Gebhard¹, Tülin Ersari² and Andrea Wirth¹

¹Hochschule Karlsruhe – Technik und Wirtschaft, Germany
e-mail: dominik.gebhard93@web.de; andrea.wirth@hs-karlsruhe.de;

²Project BASF 4.0, BASF SE, Ludwigshafen, Germany
e-mail: tuelin.ersari@basf.com

Keywords: Industrial Internet of Things, Industry 4.0, digitization, value added, benefit

Introduction

The Industrial Internet of Things (IIOT), also known as Industry 4.0, is changing markets and businesses through digitization. The combination of key technologies, such as the Internet of Things, Big Data, and Cloud Computing enables countless opportunities. Market entrance barriers are shrinking and established global players compete with newly founded start-ups. (Albach, et al., 2015) “Today’s existing value chains and business models will come under increasing pressure.” (Kagermann, 2015) Leading to an immense need to keep pace with competitors and technology. Opportunities must be determined and implemented efficiently in running businesses. While the focus is on operational disruption and remaining competitive, however, measuring the value of digitization falls behind.

Without measuring the success of the digital transformation in a company, there is no way to adjust the outcome.

The advantages of Industry 4.0 are revolutionary, the implementation will be evolutionary. (Blanchet, et al., 2014) The area of measuring benefits in IT and digitization is complex. Due to a variety of project scopes and ecosystems, it is hard to calculate reliable figures. (Oks, et al., 2017) There is no commonly recognized method to quantify the benefits. Therefore, this paper will suggest a framework, that helps to categorize and understand the value levers of digitization. With this understanding, matching Key Performance Indicators (KPIs) can be identified and used to measure the value added by the IIOT. An innovative proposal, which is particularly designed for Industry 4.0, will be developed to help illustrate the benefits.

Research

The novelty of the IIOT aggravates the search for comparable approaches to measure its benefits. Experts and strategy consulting groups like PricewaterhouseCoopers are focused on assisting companies to identify their potentials (2016). A first attempt to link the opportunities to value drivers is done by McKinsey Digital (2016). They grouped Industry 4.0-levers like a faster experimentation through simulation to eight major benefits, i.e. faster time-to-market.

This paper aims for more and builds a framework, which can be used to assess digitization-initiatives with dedicated KPIs. Therefore, representative projects that are part of the digital transformation campaign of the chemical market leader, BASF SE, have been analyzed. They originate various operating divisions and are based on different enabling technologies like statistical models, platforms, and hardware. Furthermore, they aim for a broad range of objectives from increasing efficiency to implementing new business models.

Deducing the benefits of the Industrial Internet of Things

The IIOT enables countless use cases, that leverage the performance of companies, bringing them to a new level. This affects nearly all parts of the value chain, from suppliers, over the internal manufacturing process, to the interaction with customers. (Lasi, et al., 2014) The foundation of this revolution, that supports the gathering, processing and visualization of data, is Information Technology (IT). Thus, IT helps machines and people in the decision-making process and offers new opportunities to interact with customers. While it may seem apparent, that a benefit arises from an investment perspective, i.e. automation of factories, quantifying the value added in monetary terms proves difficult. The root of this problem is the pervasion of IT.

On one side, the costs of an Industry 4.0-initiative are relative simple to estimate and measure accurately. On the other side, predicting the value added is complex. Predictive Maintenance for example, is based on sensors that monitor and analyse different attributes of assets such as temperature and vibrations. If a predefined limit value is exceeded, the asset is proactively inspected, to prevent failures. The costs for this project consist mainly of purchased sensors, development of algorithms to track and process the data, and project personnel. The benefits are the leveraged knowledge of assets, increased safety, and the potential prevention of plant shutdowns, all of which lead to higher plant productivity. How does one quantify the value added through data? This complexity leads to a negligence of benefit tracking and a focus on quick and smooth implementation. To overcome this exact complexity issue, in the following paragraphs a suitable framework is proposed.

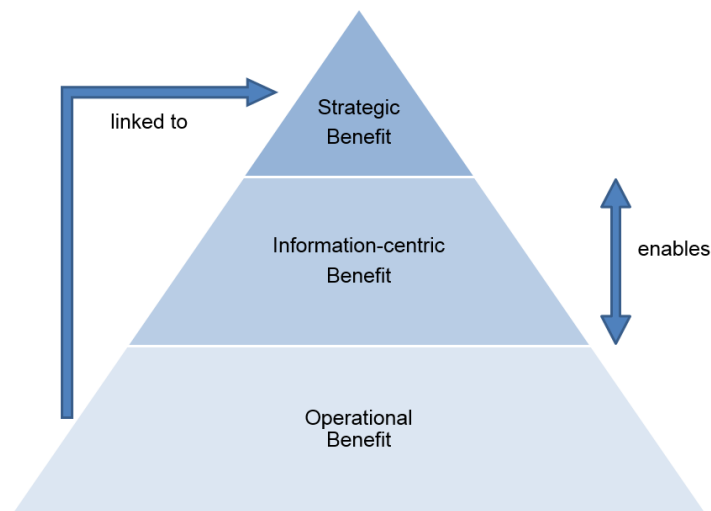


Figure 1: Relationship between benefit categories

The benefit of a digitization-project is a complex construct, adding value in several areas. Three general categories, referring to Mirani and Lederer (1998), have been identified to improve the ability to see the benefits of digitization. Those categories were validated by analyzing 45 representative Industry 4.0 initiatives.

This categorization helps break down an IIOT-initiative into its components. Figure 1 illustrates the relationship between the three categories: strategic benefits, information-centric benefits, and operational benefits. Every operational benefit is linked to a strategic benefit. In contrast to Mirani and Lederer, this paper suggests information-centric benefits as enablers and not quantifiable value added. This enabling role is also responsible for the majority of costs in digitization projects. Those costs are typically covered by benefits found in the operational and strategic benefits.

Results

The categorization is equivalent to the structure of the framework. For a better understanding and quantification, the benefits of the 45 initiatives have been broken down. After reviewing the projects and examining the targets and reasons that justify their respective investments, all benefits with similar intentions were clustered into subcategories. By rearranging and cross-checking, a preliminary grid has been developed. This grid can assist in gathering and understanding the benefits of digitization.

Strategic Benefits					
Customer satisfaction/ Loyalty	Safety	Quality	New revenue/ Safeguard sales	Strategic must	Awareness
easier to use/ easier processes	failure prevention	documentation quality	platform	regulatory	customers
agile approach	data security	product maturity	guarantee production/ service delivery	remain competitive	applicants
individulization			additional service offering		employees

Information-centric Benefits			
Data Insights	consistent data	statistic models	Infrastructure
correlations	labour data	plant simulation	IT-Services
customer related	manuals from OEMs	for research of new products	Platforms
product information/ learnings	assets and their lifecycle	company dependencies	Hardware
increased transparency		market prediction	
real-time information			
competing vendors			

Operational Benefits	
additional sales	cost reduction
increased productivity/ automation	reduced off-spec
better sales margins	HC reduction (time saved)
faster time-2-market	cheaper solutions
new service model	avoid cost

Figure 2: Developed grid to categorize benefits

By assigning appropriate KPIs, the grid can be used as framework, when trying to understand and measure the value added of an Industry 4.0-initiative. Examples therefore are benchmarks for productivity, time-2-market, or off-spec. To also highlight the qualitative benefits, measurement can be done by digital adoption rate, likelihood to recommend, customer reviews on platforms, and particularly internal expert statements, that show the need and impact of an Industry 4.0-project.

To revisit the complex quantification of the Predictive Maintenance-project, the information-centric benefits “increased transparency” and “real-time information” about the condition of an asset, enable strategic and operational benefits. Less unplanned outages and failures lead to an “increase in productivity”. This is linked to the strategic benefit “safeguarding sales” by “guaranteeing the required production”. The benefit calculation is based on prevented failures. Sensors that monitor the plant condition, feed an algorithm with data. If an anomaly is detected, preventive maintenance can be triggered to avoid losses through plant defects. The benefit of this project results from the output per time unit, multiplied with the value per output and expected inoperative time, that would have occurred without the preventive measure.

$$25\text{kg (output per min)} * 3.000\text{€ (value per kg)} * 35\text{ min(inoperative)} = 2.6\text{mn. € (benefit)}$$

To solely consider the benefit of Predictive Maintenance, the quantification should only be done, if a failure has been prevented in consequence of a hint by the algorithm. Otherwise the measurement can be based on misleading factors that bias the result. The understanding gained through the framework helps to identify underlying value drivers. Enabling the recognition of the most granular variables and their respective connections. Through these insights, a proper management and controlling of Industry 4.0-initiatives via measurement and adjustment of the correct factors can be guaranteed.

Conclusion

Matching each benefit of this framework with KPIs leads to the proposed solution. The assignment of key figures is based on literature review or the adoption of existing measures. I.e. the example of Predictive Maintenance, where a known Indicator is only considered, when triggered by Industry 4.0. To validate and refine the grid, additional projects will be analysed. The reporting system can be based on an app or web-app and is detached from conventional cockpits. Due to the novelty of the IIOT and the newly created framework, an innovative visualization is required.

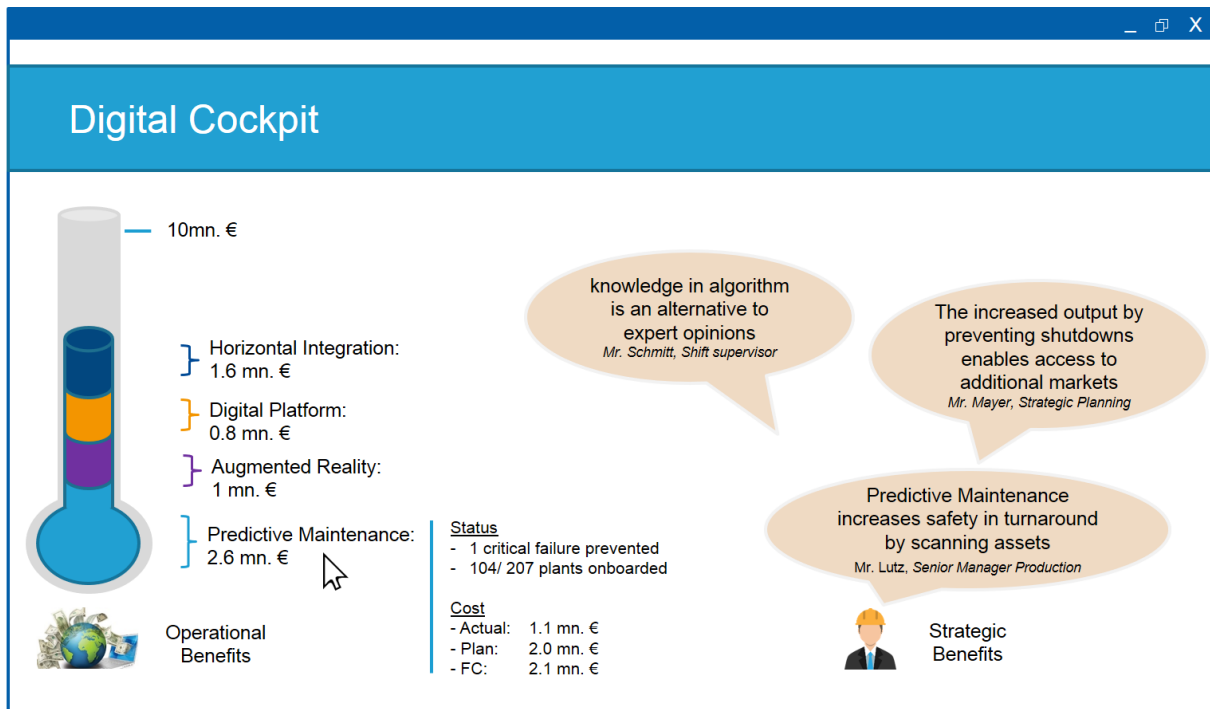


Figure 3: Mock-up of developed Digital Cockpit

The test tube on the left side fills up with the cumulative, operational benefits of each initiative. By clicking on a certain project, detailed information on cost and status are shown. The benefit of Predictive Maintenance is illustrated at the bottom of the test tube, with 2.6 mn. €, that were calculated in chapter 4. The cost details, like actual cost of 1.1mn. €, are extracted from an ERP-System. To sum up the information about this project, already one prevented failure covers the cost for this project. 104 of the 207 plants on site are provided with the sensors. Despite developing a framework with KPIs, some benefits remain unquantifiable. To highlight the importance of the strategic category, e.g. statements of internal experts, can be emphasized on the right side of the illustration. Strategic benefits like an increased safety in turnaround arise can be shown in the speech bubbles, that represent expert statements. In addition to the above shown reports, reporting certain benefits across all projects is possible. For example, the sum of all benefits due to reduced off-spec. This is similar to data-dicing and -slicing and allow further insights in benefit categories and value drivers. (van der Alst, 2013)

While there is a subjectivity due to the origin of the projects, they derive from various areas and implement hardware technologies, algorithms, statistical models, and new services. Thus, a broad area of Industry 4.0 is covered by the framework. Consequently, the grid is beneficial for the understanding of benefits and can be used as a starting point for a multitude of projects. The digital cockpit is a simple alternative to standard reporting systems, that visualizes the impact in a modern and intuitive way.

References

- Albach, H., Meffert, H., Pinkwart, A. & Reichwald, R. (2015). Management of Permanent Change - New Challenges and Opportunities for Change. In: H. Albach, H. Meffert, A. Pinkwart & R. Reichwald, Hrsg. *Management of Permanent Change*. Wiesbaden: Springer Fachmedien, pp. 1-22.
- Blanchet, M., Rinn, T., von Thaden, G. & de Thieulloy, G. (2014). *Industry 4.0. The new industrial revolution. How Europe will succeed*. [Online] Available at: https://www.rolandberger.com/publications/publication_pdf/roland_berger_tab_industry_4_0_20140403.pdf [Last access 2nd August 2017].
- Kagermann, H. (2015). Change through Digitization - Value Creation in the Age of Industry 4.0. In: H. Albach, H. Meffert, A. Pinkwart & R. Reichwald, Hrsg. *Management of permanent change*. Wiesbaden: Springer Fachmedien, pp. 23-48.
- Lasi, H. et al. (2014). Industry 4.0. *Business & Information Systems Engineering*, April, pp. 239-242.
- McKinsey Digital (2016). *Industry 4.0 after the initial hype*, s.l.: McKinsey&Company.
- Mirani, R. & Lederer, A. L. (1998). An Instrument for Assessing the Organizational Benefits of IS Projects. In: *Decision Sciences, Vol. 29, No. 4*, September, pp. 803-838.
- Oks, S. J., Fritzsche, A. & Möslin, K. M. (2017). An Application Map for Cyber-Physical Systems. In: S. Jeschke, C. Brecher, H. Song & D. B. Rawat, Hrsg. *Industrial Internet of Things*. Cham: Springer International, pp. 21-46.
- Pricewaterhouse Cooper (2016). *Industry 4.0: Building the digital enterprise*, Munich: PwC Strategy.
- van der Alst, W. M. P. (2013). Process Cubes: Slicing, Dicing, Rolling Up and Drilling Down Event Data for Process Mining. In: *Asia Pacific Business Process Management*, pp. 1-22.

The Effect of Cognitive and Individual-Level Factors in Wearable Computing Adoption – Empirical Evidence from Germany

Lena Gribel, Stefanie Regier, Ingo Stengel, Steven Furnell
Centre for Security, Communications and Network Research
Plymouth University, United Kingdom
e-mail: lena.gribel@plymouth.ac.uk

Keywords: Technology Acceptance, Wearable Computing, Information Systems Research

1. Introduction

The latest advances in communication and information technologies hold the promise of dramatically changing the way people live, work, and play. Wearable computers are emerging new technologies in the evolution of mobile devices, which introduce a paradigm shift in the field of human computer-interaction. By equipping the user with computational capabilities, ‘wearables’ provide context-aware and seamlessly integrated on-the-fly computing across heterogeneous circumstances and irrespective of place and time. Not least the very promising market prospects for wearable devices imply various unprecedented business opportunities for these socio-technological gadgets (Tractica, 2016). Driven by the enormous economic potential of ‘smart wearware’, large consumer brand companies such as Samsung, Apple and Google, but also a variety of innovative start-ups engage in the rapid-growth wearable market by tapping new business segments and launching own wearable products. For instance, the ‘Pebble Technology Corporation’, which largely financed its business venture via crowdfunding, greatly succeeded with its e-paper based smartwatch (Smartwatch Group, 2015). In the meantime, there is a high degree of product diversity in the area of wearable computers, ranging from smart wristwatches, glasses, jewellerys, headgears and e-textiles to digital contact lenses, consumer EEGs and even smart tattoos (Amft, 2017). All these wearable products promise substantial efficiency gains on the end-user’s side by creating seamless and convenient access to critical electronics and information services. Given the added value through augmented reality applications and a more unobtrusive design in comparison to conventional hand-held devices, wearables are currently influencing a multitude of non-consumer niches including those in healthcare, education, transportation and manufacturing. For instance, smart glasses are beneficial to heavy industry by providing interactive, hands-free manuals and up-to-the-minute information. These features foreshadow the tremendous impact such technologies will have not only on industry, but also on a variety of consumer sectors in the near term.

Nonetheless, the current wearable technology market is still a niche, characterised by low public awareness and a high level of turbulence and uncertainty (Tremblay & Yagoubi, 2017). Since numerous efforts in the area of innovation already failed due to a lack of consumer acceptance, it becomes clear that facilitation of acceptability is a key issue for entrepreneurship. However, up to now, there is only little scientific research on the acceptance of wearable computing. In particular, an integrated approach that exhausts the subject matter from a behavioural perspective is still lacking due to the emerging nature of the novel concept of wearable computing. At the same time, it is also significant that personality variables have seldom been examined within the scope of Information Systems research (Svendsen, et al., 2013). Therefore, the overall aim of this study is to deepen understanding of latent psychographic factors that lead to either acceptance or resistance towards wearable technologies. Specifically, a new behavioural model is introduced, which extends the well-established Technology Acceptance Model (TAM) (Davis, 1985) by explicitly incorporating a dispositional perspective into the conceptual framework.

2. Research Approach

The theoretical foundation of the study was laid by an extensive literature review that combines different streams of research including innovation diffusion and Information Systems research (inter alia, Davis, 1985; Rogers, 2010). Based on the findings from the conceptual work together with the results from a subsequent exploratory study (see Gribel, et al., 2016), salient acceptance factors were identified. These potential predictors were then integrated into a coherent system of hypotheses, the Wearable Technology Acceptance Model (Wearable TAM).

2.1. Towards a New Wearable Technology Acceptance Model

Drawing on the field of innovation acceptance, IT security and personality psychology, the derived cause-effect model aims at holistically explaining the acceptance behaviour in the case of wearable computing. Figure 1 depicts the overall system of hypotheses of the Wearable TAM. The underlying assumption of the novel conceptual framework is that the target construct ‘behavioural intent to use wearables’ in terms of a proxy of acceptance behaviour has a two-fold nature: it is jointly determined by cognitive beliefs and personality-related correlates. While cognitive beliefs relate to the knowledge and thoughts an individual has about a specific attitude object, personality variables are defined as relatively enduring psychological predispositions. In line with the common conception in personality psychology, it is assumed that the domain of personality is completely captured by the ‘Big Five’ superordinate constructs, which has come to be known as the Five Factor Model (FFM) of personality (McCrae and Costa, 1997). This descriptive model both comprehensively and parsimoniously reflects the most salient, psychometrically sound dimensions of personality along five bipolar trait concepts (i.e. ‘extraversion – introversion’, ‘agreeableness – disagreeableness’, ‘conscientiousness – lack of direction’, ‘neuroticism – emotional stability’, ‘openness to experiences – closedness to experiences’) in a common classification scheme.

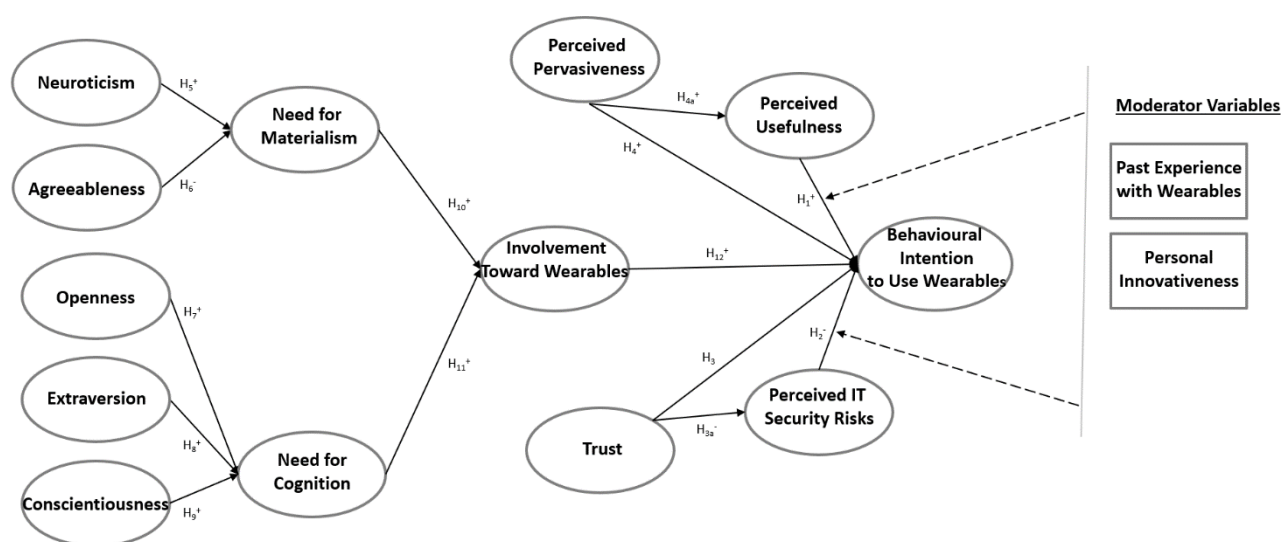


Figure 1: System of hypotheses of the overall Wearable TAM to explain the adoption of wearable computing

Furthermore, perceived usefulness and perceived security risk are both hypothesised to directly affect the intention to use. Borrowing from a quantitative social study among over 4,000 UK and US adults (Brauer & Barth, 2013), perceived usefulness is conceptualised as a formative index that reflects several utilitarian beliefs. In particular, this construct is theorised to be formed by five distinctive

facets which primarily relate to enhancements of living conditions, i.e. ‘enhancement of personal abilities’, ‘improvement of health and fitness’, ‘enhancement of self-confidence’, ‘more control of life’, and ‘enhancement of social relations’. In contrast, perceived security risk is considered a multidimensional construct. More precisely, the three first-order dimensions ‘perceived confidentiality’, ‘perceived integrity’ and ‘perceived availability’ are viewed to mutually span the conceptual domain of risk perceptions. In doing so, the current research bridges the gap between the traditional multifaceted conceptualisation of IT risk in security literature and its unidimensional measurement in most empirical studies (Hartono et al., 2014). At the upstream level of explanation, perceived pervasiveness and trust in wearable technologies are deemed to shape the individual acceptance judgment indirectly through the intermediary effect of usefulness and risk perceptions. To calibrate the research model to the actual research subject, perceived pervasiveness is also defined as a higher-order construct, consisting of the subdimensions ‘ubiquity’, ‘unobtrusiveness’ and ‘context awareness’. By explicitly accounting for the characteristics of future pervasive information systems, this new factor supersedes the original model’s ease of use construct.

In addition, based on the so-called 3M model by Mowen (2000), the theoretical framework proposes a hierarchical model of dispositional predictors of adoption behaviour, which is fully mediated by the involvement towards wearables. The research thereby responds to the call for the inclusion of individual-level variables in Information Systems models (Svendsen et al., 2013). Finally, to allow for a differentiated analysis of the effect of cognitive beliefs on usage intention, the two moderator variables ‘past experience with wearables’ and ‘personal innovativeness’ were also integrated into the explanatory model. In a technology acceptance context, both adopter-specific characteristics appear to be particularly informative exogenous factors (Rogers, 2010 p. 268; Gefen, et al., 2003 p. 307; Venkatesh, et al., 2003 p. 442).

2.2. Sample Structure

The validity of the entire structural model and its measurement instruments was empirically tested by means of a web-based survey. The majority of participants ($n = 425$) were recruited via the online panel Toluna Inc. (Toluna USA, Inc., 2016a), i.e. the sample was bought from a panel provider. Due to the widespread availability of the Internet among various groups, online research panels are increasingly utilised as a valid, cost-effective and efficient mean of data collection (Toft, et al., 2014). However, to further mitigate issues of over- and underrepresentation, additional participants were recruited via postings in diverse discussion boards. For the present research, several Internet forums were targeted, some of which focussed on technical topics and others which were thematically open.

In total, the online questionnaire was distributed to a sample of 641 participants from Germany, of which 474 cases were accepted. As can be seen from Figure 2, the gender split was 57.00 % females and 43.00 % males. Since it has to be assumed that females are increasingly interested in consumer electronics, this slight prevalence is considered acceptable. In addition, Figure 2 gives an overview of the percentage share of age categories compared to the German population. The age distribution indicates that younger adults aged 20-29 were somewhat over-represented in comparison to German census statistics (Statistisches Bundesamt, 2017). From a supplier’s point of view, this is clearly a worthwhile age group in terms of technology adoption behaviour at early stages of the product introduction process (Yun, 2013, p. 50). With particular regard to the elderly adults aged 50 and more, the overall age-profile of the sample corresponds largely to the target population. This is important since an under-representation of elderly (which is frequently observable in studies on technology acceptance) would result in a neglect of a promising future consumer market (Gartner, Inc., 2013).

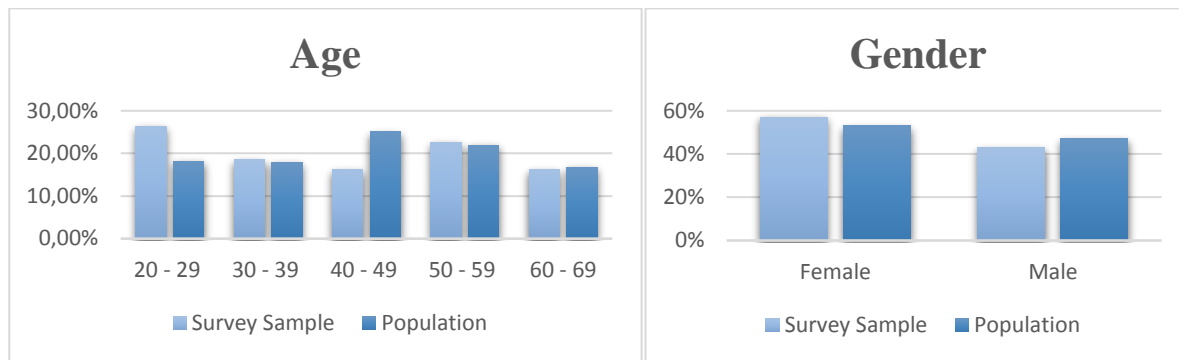


Figure 2: Comparison of gender and age distributions

Furthermore, the descriptive analysis shows that the educational level of the survey respondents can be regarded as far above-average. Nearly 40% of respondents state that they hold a university degree, while 20% have a general qualification for university entrance. This group distribution corresponds well with the socio-economic structure of innovators and early adopters, who are considered to be more qualified and in the higher income group (Yun, 2013, p. 91).

3. Empirical Findings

In order to quantify the constructs and their causal links, the Partial Least Squares (PLS) method was used as the best-fitting prediction-oriented technique for the analytical research objectives (Hair et al., 2016). After having assured that the psychometric properties of the study-inherent constructs meet all relevant standard criteria in terms of reliability and construct validity (the employed criteria can be found in Table 1), the structural model was estimated.

Reflective Measurement Model Criteria		Formative Measurement Model Criteria	
Evaluation Criterion	Requirement	Evaluation Criterion	Requirement
Item Reliability		Indicator Relevance	
<i>Factor Loadings</i>	> 0.7	<i>Outer Weights</i>	No minimum level; interpretation of the magnitude of contribution of each formative indicator
Internal Consistency Reliability			
<i>Composite Reliability</i>	≥ 0.7	Statistical Significance	
<i>Cronbach's Alpha</i>	≥ 0.6	<i>t-value</i>	> 1.98 (two-tailed test; ~ error of 5%)
Convergent Validity		Multicollinearity	
<i>Eigenvalue</i>	EV1 > 1; EV2 < 1	<i>Variance Inflation Factor</i>	< 5.0
<i>Average Variance Extracted (AVE)</i>	> 0.5		
Discriminant Validity			
<i>Fornell-Larcker criterion</i>	AVE > squared correlations of construct with other constructs		
<i>HTMT</i>	HTMT confidence interval does not include 1.0		
Statistical Significance			
<i>t-value</i>	> 1.66 (one-tailed test; ~ error of 5%)		

Table 1. Reflective and formative measurement model evaluation criteria

Overall, the empirical data confirm the predictor structure implied by the proposed model as shown in Figure 3. The predictive effects are visualised as arrow paths and the strength and significance of the pertaining structural coefficients are given for each hypothesised relationship. In addition, R-squared values are reported, which indicate the portion of a dependent variable's explained variance.

Furthermore, the Stone-Geisser's Q^2 -test criterion is displayed as a measure for the predictive relevance of the reflectively measured, endogenous constructs. Non-significant relationships are illustrated by dotted lines. As shown, most of the obtained structural parameters suggest strong empirical relations. Only the proposed dependency structures denoted by H3, H4, and H9 fail to meet the established requirement level ($\beta > 0.1$). To verify the significance of the hypothesised causal links, two-tailed t-tests ($n = 3000$) were used due to the directional nature of hypotheses. The sample distribution was derived by means of bootstrapping, whereby subsamples are randomly drawn with replacement from the original survey data. Again, all three hypotheses H3, H4, and H9 can be considered as being not supported by the data, since the error probability for the respective path coefficients is greater than 0.1.

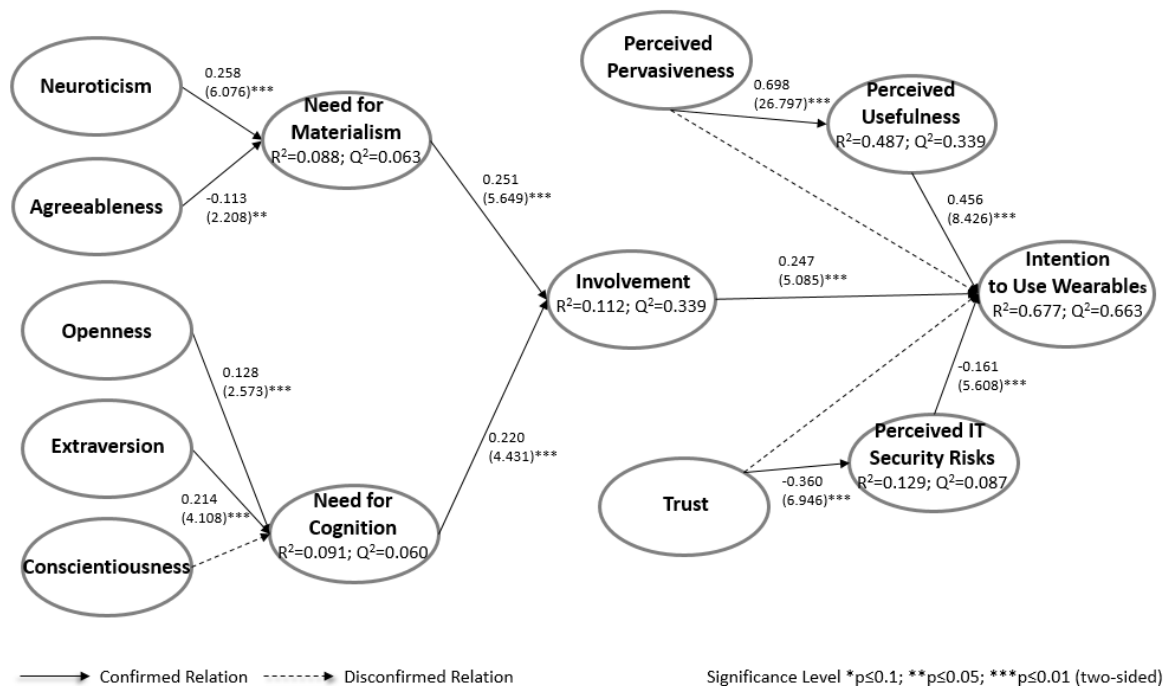


Figure 3: Standardised path coefficients, significance values and predictive capacity of the integrative structural model

Consequently, the effect sizes of the concerning predictor constructs (i.e. conscientiousness, perceived pervasiveness and trust) on their related dependent variables (i.e. need for cognition and intention to use) indicate a substantively meaningless explanatory contribution, having f^2 -values far below 0.02. The effect size is defined as the increase in R^2 relative to the proportion of unexplained variance of the dependent variable when estimating the path model twice – once with the exogenous latent variable excluded, and once more with the exogenous variable included. This measure should be above 0.02 to suggest at least a weak predictive relationship (Hair, et al., 2016). In light of the estimation results, hypotheses H3, H4, and H9 had to be rejected. As suggested by the findings of the mediation analysis, pervasiveness perceptions and affect-based trust are rather mediated by their theoretically linked cognitive beliefs. The results from the bootstrapping procedure indicate that both indirect pathways as denoted by H3a and H4a are significant at an alpha 0.01 level. It can therefore be concluded that a mediating effect is present in either causal relationship. Yet, the former hypothesis is only partially supported due to a small total effect of the mediated relationship.

Furthermore, the results of the moderation analysis lead to the conclusion that the structural relation between perceived usefulness and behavioural intention is strongly moderated by innovativeness and

prior experience with wearable technologies. The pertaining interaction terms are significant with an error probability of $p < 0.01$. On the contrary, no moderating effect could be proven for the perceived risk \rightarrow intention relationship. Hence, further interacting third variables may be assumed, which account for the variability in response patterns when it comes to risky adoption decisions.

4. Discussion of Results

Consistent with the apriorically formulated conceptual framework, the parameter estimates provide evidence that both cognitive and affective responses strongly covary with usage intention. Still, cognitive information processing comes more to the fore in the context of wearable computing adoption. In line with numerous prior Information System studies, the empirical investigation thereby corroborates the dominance of cognitive beliefs in attitude formation. It may be argued that the usage of a specific information system often represents a means-end behaviour, which emphasises the instrumental nature of technology adoption decisions (Bhattacharjee, 2001). Remarkably, the computed R-squared value of 0.677 for the dependent variable 'intention to use' underlines the predictive capacity of the research model.

More specifically, the results of the quantitative study identify the perceived support of health and fitness as well as the perceived enhancement of personal abilities as strongest drivers of usefulness perceptions that affect the intention to use wearable technologies. Once again, the analytical results hereby implicate that instrumentality considerations may override affective motives in stimulating adoption decisions. In line with the conceptual framework, the analysis of mediating effects moreover underpins the theoretic assertion that the influence of perceived pervasiveness on behavioural intention is completely mediated by individual usefulness perceptions. This means that the unique technological features that distinguish pervasive computing from the classical desktop paradigm only take effect on device usage if they are translated into concrete applications that are found to be useful. This result fits earlier conceptual and empirical works, which indicate that motivational variables entirely mediate system design features, which in turn have no additional direct effect on system use (Davis, 1993 p. 482).

On the contrary, perceived privacy risk was found to be a major barrier to adoption. Even though less predictive of behaviour, personal predispositions qualify as another essential source of technology acceptance. Empirical data indicates that need for material resources and need for cognition jointly evoke personal commitment towards innovative information systems. In spite of the seemingly noisy, high-variability residuals, the predecessor constructs of involvement still provide valuable information about tendencies in consumer responses – particularly, as this study does not aim at precisely predicting intrapsychic predispositions, but rather at determining which explanatory traits are statistically significant and how one-unit increases in these individual difference variables relate to changes in the emotional attitude towards wearables. Besides, the analysis of moderator effects shows that both the level of personal innovativeness and past experience with wearables indirectly influence benefit expectations. In other words, innovative individuals and those who are familiar with these technologies are more likely to develop positive attitudes towards the use of wearable devices.

5. Conclusions

In view of the significant business opportunities the wearable technology market is expected to create, understanding user adoption behaviour becomes an economic imperative for management of companies. However, due to the emerging nature of the novel concepts of ubiquitous and wearable computing, extensive theoretical and empirical gaps still exist in Information Systems research. The present research work thus contributes to the body of knowledge in that it provides a new behavioural

model that is capable of explaining holistically the interpersonal decision to adopt wearable computing.

Furthermore, motivated by the actual contrast between objective measures and subjective perceptions of IT security, the current research delivers empirical insights into consumer concerns about concrete security risks associated with wearables. Based on a consumer value perspective, it further advances knowledge on usefulness perceptions by combining a series of utilitarian beliefs into an index representing the wearables' positive usage consequences. In addition, by incorporating personality-related correlates within the synthesised nomological net, the study contributes to contemporary research on the role psychological predispositions play in the technology adoption process. Especially, the effect of the well-recognised Big Five personality dimensions on the intention to use an innovative technology is still an underdeveloped area. In an effort to deepen the established understanding of attitudinal dependencies, this research project moreover investigated the contingencies that underlie the relationships between usage intention and its cognitive causes.

The empirical findings not only contribute to the existing body of knowledge in Information Systems research, but also have several important implications for marketing practitioners. Given the dominance of cognitive beliefs in attitude formation, companies in the wearable sector should focus more on informative issues in their communication mix to educate consumers about the main benefits of wearables. For wearables to be perceived as useful, they should operate even more naturally and unobtrusively than preceding mobile innovations, what clearly reinforces the significance of human-centred design principles. Since the study findings reveal that potential breaches in data privacy represent the greatest IT security concern, vendors should aim at improving consumer attitudes towards their privacy practices. Moreover, considering that risk perceptions are heavily affected by trusting beliefs, building consumer trust appears to be key in reducing latent uncertainties and resistance to adoption. As the survey data underscore the importance of personal involvement with the new type of technology, marketers should also take account of the emotional nature of information systems usage. To attract those consumers who tend to be more involved with wearables, promotional applications ought to be sensitive to individual differences as regards materialistic values. In view of the personality traits and group differences that were proved to be of behavioural relevance, it appears furthermore worthwhile to divide the consumer market psycho-graphically into relevant personality profiles: In the case of wearable computing, especially male consumers with neurotic tendencies will act as early adopters. By developing target-group oriented communication strategies, marketers may efficiently approach the key segment of current and prospective wearable computing users.

References

- Amft, O. (2017). "Smart Eyeglasses, e-Textiles, and the Future of Wearable Computing". In *Advances in Science and Technology*, Volume 100, pp. 141-150.
- Bhattacharjee, A. (2001). "Understanding Information Systems Continuance: An Expectation-Confirmation Model". In: *MIS Quarterly*, 25(3).
- Brauer, C. & Barth, J. (2013). *The Human Cloud: Wearable Technology from Novelty to Production*. [Online] Available at: http://www.rackspace.co.uk/sites/default/files/whitepapers/The_Human_Cloud_-_June_2013.pdf
- Davis, F. D. (1985). *A Technology Acceptance Model for Empirically Testing New End-user*, Massachusetts: Wayne State University.
- Davis, F. D. (1993). "User acceptance of information technology: system characteristics, user perceptions and behavioral impacts". In: *International journal of man-machine studies*, Volume 38.

- Gartner, Inc. (2013). *Gartner Says Silver Surfers Represent a More Important Technology Market than Generations X and Y*. [Online]. Available at: <http://www.gartner.com/newsroom/id/2615818>
- Gefen, D., Karahanna, E. & Straub, D. W. (2003). "Inexperience and experience with online stores: The importance of TAM and trust". In: *IEEE Transactions on Engineering Management*, 50(3).
- Gribel, L., Regier, S. & Stengel, I. (2016). "Acceptance Factors of Wearable Computing: An Empirical Investigation". In: *Proceedings of the Eleventh International Network Conference (INC 2016)*, pp. 67-72.
- Hair, J. F., Hult, G. T. M., Ringle, C. M. & Sarstedt, M. (2016). *A Primer on Partial Least Squares Structural Equation Modeling*. 2nd ed. s.l.:Thousand Oaks: Sage.
- Hartono, E. et al. (2014). "Measuring perceived security in B2C electronic commerce website usage: A respecification and validation". In: *Decision Support Systems*, Volume 62, pp. 11-21.
- Maisto, M. et al. (2017). "Evaluation of wearable haptic systems for the fingers in Augmented Reality applications". In: *IEEE Transactions on Haptics*.
- McCrae, R. R. & Costa, P. T. (1997). "Personality trait structure as a human universal". In: *American psychologist*, 52(5).
- Mowen, J. C. (2000). *The 3M model of motivation and personality: Theory and empirical applications to consumer behavior*. s.l.:Springer Science & Business Media.
- Rogers, E. M. (2010). *Diffusion of Innovations, 4th Edition*. 4 ed. s.l.:Simon and Schuster.
- Smartwatch Group (2015). *Pebble - Smartwatch Group*. [Online] Available at: <http://www.smartwatchgroup.com/pebble/>
- Statistisches Bundesamt (2017). *Destatis*. [Online]. Available at: <https://www.destatis.de/DE/Startseite.html> [Accessed 09 01 2017].
- Svendsen, G. B., Johnsen, J. A. K., Almås-Sørensen, L. & Vittersø, J. (2013). "Personality and technology acceptance: the influence of personality factors on the core constructs of the Technology Acceptance Model". In: *Behaviour & Information Technology*, 32(4), pp. 323-334.
- Toft, M. B., Schuitema, G. & Thøgersen, J. (2014). "Responsible technology acceptance: Model development and application to consumer acceptance of Smart Grid technology". In: *Applied Energy*, Volume 134.
- Toluna USA, Inc. (2016). *toluna-group*. [Online]. Available at: <http://www.toluna-group.com/en-gb/>
- Tractica (2016). *Wearable device sales revenue worldwide from 2015 to 2021 (in billion U.S. dollars)*. [Online] Available at: <https://www.statista.com/statistics/610447/wearable-device-revenue-worldwide/> [Accessed 08 06 2017].
- Tremblay, D. G. & Yagoubi, A. (2017). "Business Model and Wearables: What Convergence and Collaboration in the Area of Connected Objects and Clothing?". In: *Journal of Human Resource and Sustainability Studies*, 5(1), p. 101.
- Venkatesh, V., Morris, M. G., Davis, G. B. & Davis, F. D. (2003). "User acceptance of information technology: Toward a unified view". In: *MIS quarterly*, 27(3).
- Yun, Y. (2013). *Understanding Product Attachment and Expected Product Lifetime by Extending Technology Acceptance Model (TAM) with Product Personalization and Innovation Diffusion Theory (IDT)*, East Lansing, USA: Michigan State University.

A review of challenges in blockchain adoption for supply chains

F. Buchner, I. Stengel
Department of Business Information Systems
Karlsruhe, University of Applied Sciences
e-mail: f.buchner@outlook.com

Keywords: Blockchain, Distributed Ledger, Supply Chain

1. Introduction

Blockchain is one of the up-and-coming technologies and is predicted to be merely years from broad market adoption. (Stephan Janssens *et al.*, 2017) The blockchain originated as a crypto currency in 2008 with the idea of creating a distributed ledger that works without any central authority allowing everyone to participate anonymously. (Nakamoto, 2008) After the original concept proofed to be functional new use cases have been developed which extend the initial technology. There are two different ideas behind modern blockchain systems. One aspect is the bundling of transactions which are then accepted via a consensus protocol. The other aspect is to include computer code in these transactions which effectively transforms a blockchain from a distributed data ledger into a distributed computer (Wood, 2017).

Another field currently facing many challenges are supply chains. In the global economy supply chains have become more and more complex. Contracts between suppliers or between suppliers and manufacturers or retailers are often valid in different countries with different jurisdictions. The amount of interactions is growing as international goods transfer increases. Additionally, consumers play an important part as well, demanding transparency of the life cycle of the product they are buying. (Lehmacher, 2016)

Various works have studied the application of blockchain technology in supply chains. In this paper, we will give an overview of relevant existing research and the use-cases they describe. Since the technology is evolving fast and due to it being still in the development stage, there are still issues with the implementation of use-cases in real-life application. We will discuss current challenges for each use-case both from a business and a technology point of view. This will give an overview of the present progress and challenges in bringing blockchain technology into production.

2. Blockchain Technology

Blockchain systems combine several technologies and protocols. A blockchain is an ordered set of transaction lists. Each transaction list combines a selection of transactions, metadata and a consensus proof into a block of data. All block data is then hashed. Each new block links to the predecessor by referencing the hash value of the latter which creates a chain of blocks (Nakamoto, 2008). Since hashing functions are one-way methods and each block in the blockchain contains the link to its predecessor an altering of any block of the blockchain requires to alter all subsequent blocks. Adding new blocks to a blockchain by following the consensus protocol is known as mining. The aforementioned consensus proof is required to prevent participants from generating a new version of a blockchain after changing data in an existing block. Currently the most popular form of consensus is the proof-of-work where hashing power must be invested costing the participants real resources to create virtual transactions.

Modern blockchain systems do not only allow to send and receive transactions by command but also to run computer code in the form of so-called smart contracts. Smart contracts are touring-complete programs, i.e. they can be used for any deterministic computer program. (Turing, 1937, 1938)

The next layer of a blockchain system after the data structure itself is the participant network. Each server holding a copy of the blockchain and mining new blocks is known as a node. Those nodes use standardized protocols to discover each other and allow communication in the distributed network. With another protocol clients can then publish transactions within the network. The whole process is illustrated in Figure 1.

Due to the network being distributed rather than centralized it does not have a single point of failure. The consensus mechanism is another measure of risk mitigation since it allows the blockchain to reach an agreed upon state even when a minority of participants try to sabotage or undermine it.

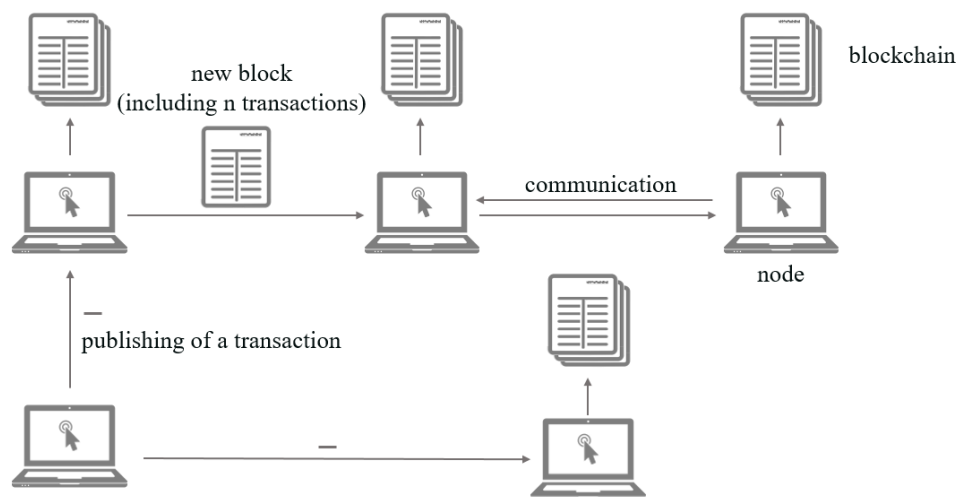


Figure 1- A blockchain network consists of various nodes whereas each node communicated with a number of other nodes to obtain the current valid copy of the blockchain. New transactions are broadcasted into the network and then mined into new blocks.

This blockchain system can be used to support business use-cases in various fields.

3. Supply chain use-cases

A supply chain consists of all the companies and their processes connected to the production and shipment of a good or service (Voß, 2007). In the context of this paper the scope will be mostly narrowed down to the aspect of the journey of goods in between their production life cycle. This begins with the production of intermediate goods by a supplier, and ends with the purchase of the finished good by the customer.

Use-cases for blockchain technology cover various aspects of supply chain and its management. A number of current challenges and possible improvements can be found within the following use-cases.

An often sought case is providing customers with information about the production history of a product. This includes the various stations a product comes along before a customer finally acquires it. This history is especially relevant for agricultural goods since numerous scandals have weakened the trust of buyers. Providing consumers with the information of where e.g. a chicken was raised, where it was slaughtered and processed and which companies handled the processing can make them feel more in control and will force companies to adhere to certain standards. The latter is due to

consumers being able to easily identify products where a certain supplier was involved in the production.

Linked to consumers' trust as well is the certification of goods. Nowadays there are many certification marks trying to aid the consumer in their buying decisions. For consumers, it can be hard to distinguish the various seals and the meaning behind them. Additionally, buyers need to put in a rather great effort to verify the certification information by visiting the different websites and trying to find audit information about the company carrying the respective seal. Here one immutable ledger containing certification information for all products can improve usability for customers.

Another area where supply chains can still be improved is tracking of good delivery. While the production history outlined above is focused on providing information about the composition of a finished good and the companies involved in making it, tracking is focused on enabling a client to receive up-to-date information on the location of a delivery. Therefore, data provided by the shipping companies has to be impossible to alter and should be provided as automated as possible. Ideally the delivery process of a shipment can be integrated with the fulfilment of the associated delivery contracts. This would mean that payment of a delivery can be linked with the tracking process allowing the parties involved a more efficient execution of their contracts.

4. Review of current research

One of the most detailed works regarding a framework for blockchain for supply chains is "An Agri-food Supply Chain Traceability System for China Based on RFID & Blockchain Technology" by (Tian, 2016). Tian develops the idea of using Internet of Things (IoT) technology, namely RFID chips, to unambiguously identify goods along the geographical route of their transport along the supply chain. He specifically focusses on the advantages of that system to reduce fraud in agricultural goods processing. This works by assigning each product unit a unique key which is then stored on a blockchain where it is impossible to alter the data without consent. Additionally, the proposed system is using a hardware sensor system to include real life metadata like temperature associated with the respecting goods. Altogether Tian aims for three advantages. One is to provide information tracing to find perpetrators after an incident. The other is a guarantee of freshness of the traced agricultural goods by providing their production date and circumstances. As a third advantage transparency should be brought to consumers allowing them to access product related supply chain information.

There are two main points Tian acknowledges as challenges. On the one hand, the blockchain technology is not market ready as of now. It is hard for businesses to adopt a technology which is still in the making and changing daily. On the other hand, RFID chips are still cost intense especially in comparison to the barcodes that are currently in use. Contrary to many technologies sinking costs of RFID chips cannot be expected within the very near future since the antennas of the included microchips must be made from expensive metals. In addition to the challenges the author talks about there are still several points one can make against the agri-cultural blockchain based supply chain as described by the author. One point is the usage of pallets in goods distribution. Since individual agricultural goods like fruits and vegetables cannot be individually equipped with RFID chips the solution is to attach the RFID chips to the transport pallets instead. The issue herein lies with the reusing of pallets by suppliers. Usually goods are transported on these pallets whereby old pallets are returned to the delivering supplier in exchange for the new pallets included in the delivery. (Kuhn et al., 2005, p. 48) This complicates the system substantially since pallet management needs to be a part of it. Another point regarding the pallets is, that these cannot be tampered with when recorded in a blockchain system but nevertheless the goods stored on the pallet can easily be moved from one pallet to another. This can neither be noticed by the system nor by the consumer which destroys the

whole trust and unambiguity of the system. Another problem comes with the use of the temperature sensors the author wants to include. A system where temperature sensors are directly connected to a blockchain is possible but even if the measured data of the sensors is correct the system can still easily be tampered with. When placing cooling cartridges close to those sensors the sensors still technically measure the correct temperature but the assumption of the system does not apply anymore (McHugh and Steadman Jr., 1997). This kind of loophole is a hard to fix issue with many blockchain systems. A blockchain ledger can guarantee the immutability of the data but it cannot control the data before it enters the system. One last important point Tian did not include is the transformation of units of production goods. To visualize that one can look at olive oil and the according supply chain. In the beginning olives are farmed and measured in weight. In the end olive oil comes out as a result which cannot be measured in the same unit, as in one load of olives transforms into one load of olive oil. Instead a transformation of the supply chain goods is necessary. The more processed the agricultural good is the more important and complicated is the provenance process. Integrating a system like that into a blockchain system was already discussed by (Kim and Laskowski, 2016).

In “Towards an Ontology-Driven Blockchain Design for Supply Chain Provenance” by (Kim and Laskowski, 2016) the authors lay out a system for bringing ontologies, formalized conceptualizations, onto a blockchain. Kim and Laskowski focus on provenance, which means tracing the origin of a product providing a footprint of that product for consumers. The specific ontology described by the authors tries to establish provenance for a good along the supply chain even though the good may change units along the way. In the publication, the authors use the existing TOVE ontology and implement the smart contracts necessary to model said ontology on the Ethereum blockchain. Due to using a formalization this system can vastly improve data quality by providing a standard for the data transformation process. As a conclusion of their prototype the authors note two hardships in bringing their system to production. They argue that the programming language, in this case solidity, is still in a very immature stage which makes it hard to implement anything more than a prototype. Another point the authors make is that the whole system naturally depends on being able to formalize the ontology for the respective process. This requires efforts from all parties involved and compromises may need be made. Going further from there, several more issues can be stressed. One is, that as described a mutual ontology needs to be found. When there are many parties involved this can lead to accepting the smallest common denominator instead of the best solution. This applies to all alternative systems. In addition an ontology based system needs to be used as conceptualization is the focus there. Again, another issue here is how to stop manipulation. The data in the system is controlled by its clear structure and processes. Nevertheless, this does not stop participants abusing the system by entering malicious data. To combat this issue, it is necessary to look beyond the blockchain ontology as an isolated system e.g. seeing it as a part of a framework for supplies chain provenance.

There is another publication focussing on the whole field of manufacturing in relation to blockchain technology called “Blockchain ready Manufacturing” by (Abeyratne and Monfred, 2016). In this work the authors look at a set of use-cases for blockchain application and focus on manufacturing, especially regarding opaque supply chains. Abeyratne and Monfred make the point of current supply chains facing issues in regard to traceability of goods. Contrary to many other works this paper not only focusses on one or two parties involved in supply chains but includes the whole set of different actors from standards organizations, to producers, retailers and many more. The authors draw the image of an ecosystem where a digital profile of each good is created and enriched with data by several third parties along the product production life cycle. Here the authors focus on the issue of including external data in the blockchain by including third party audits for product certification and thereby intertwining real world with the digital blockchain. While the system itself is very comprehensive there are still a number of open questions left to be discussed. One of those questions

is how a system like this can be adopted in a real-life scenario. The extent of the system including the large amount of third parties makes it hard to find a common ground and incentives for everyone to join that system. Additionally, it is left open by the authors who should operate and maintain such a complex system. Since it relies on blockchain technology there will be a blockchain currency and thought still needs to be put in as on how to include said blockchain tokens.

5. Conclusion and future work

As the review of the current literature regarding blockchain systems, especially in the area of supply chains, shows technology still faces challenges on its way to production deployment. So far, many thoughts have been put into the general system on how a supply chain can be brought onto a blockchain system in general. Nevertheless, there is still work to do, in designing the details of such a system. Many works focus on the big picture and general processes but details are often omitted. Due to the immature nature of the blockchain technologies there are many theoretical works but there is still a lack in actual implementations of prototypes.

With continuous development of the blockchain technology prototypes and implementations of processes that now have already been described are becoming more and more sensitive. Another important area for future work is on how to build the participant network of such a blockchain supply chain system, not from a technology side, but from an actual business network perspective. When looking at the various advantages such a system can bring for both consumers and producers, it can be seen how important further research into the topic is.

6. References

- Abeyratne, S.A. and Monfared, R.P. (2016). "Blockchain ready Manufacturing Supply Chain using Distributed Ledger", *International Journal of Research in Engineering and Technology*, Vol. 05 No. 09, pp. 1–10.
- Kim, H.M. and Laskowski, M. (2016). *Towards an Ontology-Driven Blockchain Design for Supply Chain Provenance*.
- Kuhn, E., Lange, V. and Zimmermann, P. (2005). *Paletten-Management: Leitfaden für die Praxis*, 1. Aufl., Vogel, München.
- Lehmacher, W. (2016). *Globale Supply Chain*, Springer Fachmedien Wiesbaden, Wiesbaden.
- McHugh, T.K. and Steadman Jr., A.L. (1997). *Setpoint limiting for thermostat, with tamper resistant temperature comparison* No. US5697552 A.
- Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*.
- Stephan Janssens, Sebastian Steger, Urs Arbter, Wolfgang Hach, Michael Hilbert, David Lecea, Kevin de Patoul, Manuel Keller and Michael Müller (2017). *Enabling decentralized, digital and trusted transactions: A comprehensive guide to reinventing companies Why blockchain will transform the financial services industry*.
- Tian, F. (2016). "An agri-food supply chain traceability system for China based on RFID & blockchain technology", in Yang, B. (Ed.), *2016 13th International Conference on Service Systems and Service Management (ICSSSM): June 24-26, 2016, KUST, Kunming, China, Kunming, China*, IEEE, Piscataway, NJ, pp. 1–6.
- Turing, A.M. (1937). "On Computable Numbers, with an Application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society*, s2-42 No. 1, pp. 230–265.
- Turing, A.M. (1938). "On Computable Numbers, with an Application to the Entscheidungsproblem: A Correction", *Proceedings of the London Mathematical Society*, s2-43 No. 1, pp. 544–546.

Voß, H.-P. (2007), *Horizontale Supply-Chain-Beziehungen*, Deutscher Universitäts-Verlag, Wiesbaden.

Wood, G. (2017), *Ethereum: A Secure Decentralised Generalised Transaction Ledger: EIP-150 Revision*.

Chapter 5

Learning

Thoughts on Presenting Learning Recommendations in the Context of the Project SensoMot

O. Schneider

TBA – Zentrum für technologiebasiertes Assessment
DIPF – Deutsches Institut für Internationale Pädagogische Forschung
e-mail: Oliver.Schneider@dipf.de

Keywords: Adaptive learning, sensors, motivation, recommender systems, UI

1. Motivation for Learning

Motivation is a key factor in learning, both in terms of understanding and memorising learning contents. A high degree of motivation will benefit the joy of learning and interest, leading to good learning outcomes. However, motivational problems possibly hinder the learning process and may result in a barrier to learning. Comparable to real-life learning situations, where teaching staff adjust the teaching process to the current learning situation, adapting the presentation of learning content appropriately before the learner's motivation decreases should strengthen learning motivation and therefore learning success.

As one main subject to the SensoMot (Schneider et al., 2017) project, it is assessed how motivationally caused learning blockades can be detected early by means of unobtrusive sensors, and learning content can be adapted appropriately. If learning motivation could thus be increased, higher learning success might be effected as well as lower dropout rates in many technology-based learning and teaching scenarios.

2. Objectives and Procedures

SensoMot targets the detection of critical motivational conditions by collecting sensory data without disturbing the learning context. In this regard, for example, filling in a form for the purpose of formatively assessing motivation like shown in Figure 1 distracts from learning, and results of such assessments are generally falsified. To detect learners' motivation without distracting them from the

learning context itself, in SensoMot sensors are evaluated that are commonly available to consumers, implemented in fitness tracking devices, smartphones and smart watches etc. Such so-called wearables will be used to assess physical data of the learners which might indicate stress or boredom. By deduction of adequate mechanisms for adaptation, learning processes will be governed according to the learners' motivation. The learning software algorithm will, for example, adjust the learning path and therefore recommend alternative content-sections.

The screenshot shows a Moodle 2.6 web interface. The page title is 'Learning Tools Interoperability™'. Below it, the section is 'Motivatoring im Kurs'. There is a table with 10 rows of statements and 5-point Likert scales for rating. The statements are in German and relate to the learner's current motivation and engagement in the course. The form is titled 'Motivatoring' and has a 'Motivatoring' button at the top right.

Statement (Left)	Rating Scale (Center)	Statement (Right)
Im Moment bin ich an der Arbeit in unserer Veranstaltung interessiert.	○ ○ ○ ○ ○	Im Moment würde ich lieber an etwas anderem arbeiten.
Ich fühle mich in der Lage, in unserer Veranstaltung gut mitzuarbeiten.	○ ○ ○ ○ ○	Ich fühle mich nicht in der Lage, in unserer Veranstaltung gut mitzuarbeiten.
Ich finde, dass wir in unserer Gruppe gut zusammenarbeiten.	○ ○ ○ ○ ○	Ich finde, dass wir in unserer Gruppe nicht gut zusammenarbeiten.
Bei unserer Arbeit können wir genug eigene Entscheidungen treffen.	○ ○ ○ ○ ○	Bei unserer Arbeit können wir nicht genug eigene Entscheidungen treffen.
Meine Arbeit in unserer Veranstaltung finde ich im Moment spannend.	○ ○ ○ ○ ○	Meine Arbeit in unserer Veranstaltung finde ich im Moment langweilig.
Für die Arbeit in unserer Veranstaltung habe ich ausreichende Kenntnisse und Fähigkeiten.	○ ○ ○ ○ ○	Für die Arbeit in unserer Veranstaltung habe ich keine ausreichenden Kenntnisse und Fähigkeiten.

Figure 1: Formative assessment via web-form

Prototypically, respective learning scenarios will be developed and evaluated for university teaching in the field of nanotechnology, and distance learning in vocational education and training of technicians in engineering. This learning system will be duly made available to educational practice by means of applications.

3. Obstacles

The two main objectives are measurement of motivation and content adaptation. To attract users, common wearables like fitness trackers should be used, which is more problematic than expected: To be able to react on the users' motivational changes, the sensor-data need to be read immediately. However, none currently available commercial wearable can be addressed directly. All these devices use the vendor's cloud for submitting data, making latency and vendor's data pre-processing unpredictable. Additionally, it is a security hazard to store body measurement data for learning purposes in a cloud service outside of the EU.

While the project members still seek for and discuss the best adaptation methods that are based on motivation, in the last time the work focused on the presentation of the adaption.

4. Presentation of Adaptive Learning Content

For presenting adaptive learning content, some requirements should be considered:

- Learners should focus on learning content, the system should not distract
- Learners decide what is best
 - The learners should have the freedom to choose to learn linearly, via random access, or adaptively
 - The adaptive system should recommend changes, but not enforce them

To take in to account these points, SensoMot's GUI is based on the learning environment NanoTecLearn (TUI, 2017) that already provides linear and random access, and additionally provides different learning methods.

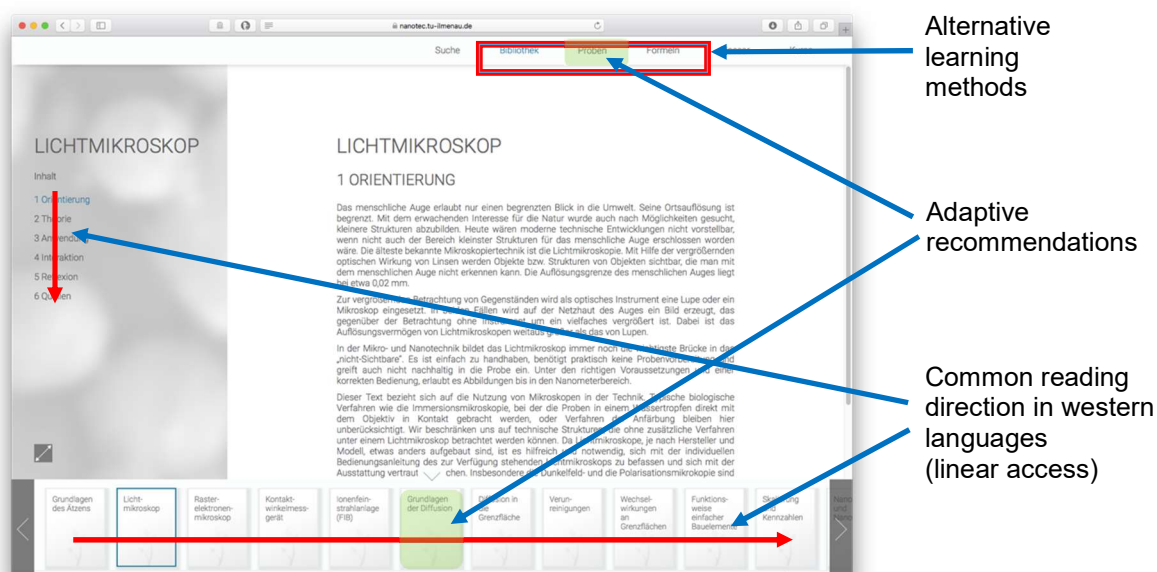


Figure 2: Screenshot of the intended SensoMot UI

To present the adaptive algorithm's suggestions an appropriate adaptation method (Knutov et al., 2009) needs to be selected that integrates into the NanoTecLearn platform. Due to the complexity of content interrelationships, the content itself is not adapted. Because distraction from the content should be avoided, no visual elements are added or deleted. Thus, an adaptive navigation presentation technique has been chosen, namely the link annotation.

In NanoTecLearn and therefore also in SensoMot, content-links are presented as books, chapters, or menu entries. As soon as SensoMot's algorithm detects a remarkable change in motivation, it suggests to alter the learning method or to change to another content-part. This suggestion is presented by highlighting the learning method or content part by colourising the appropriate link (Figure 2 gives an overview). The colour will not be changed suddenly but transitions colours smoothly. This even more avoids distraction and gives users a better chance to follow changes of the suggestions.

5. Early Conclusions and Short Outlook

As SensoMot will facilitate detecting motivational obstacles in the way of learning with the help of unobtrusive sensors, also the presentation of suggestions should not distract from learning. However, no distraction does not mean that the suggestions should be unnoticed. This paper presents the usage of link annotation for SensoMot by colourising links as adaptive technique integrated into the NanoTecLearn-UI, so that the suggestions should be noticed in time but will not distract the learner.

As further work, the real user experience needs to be evaluated by usability tests. Additionally, the adaptation itself is currently developed and will be integrated into SensoMot. As soon as the motivation-indicators have been detected, hopefully the market has changed so that wearables will be available that can be addressed directly without being forced to use any vendor's cloud.

6. Acknowledgments

SensoMot is funded by the Bundesministerium für Bildung und Forschung – BMBF (Federal Ministry of Education and Research).

References

- E. Knutov, P. De Bra, and M. Pechenizkiy (2009). "AH 12 Years Later: A Comprehensive Survey of Adaptive Hypermedia Methods and Techniques", in P. De Bra and P. Brusilovsky, editors: *Special Issue: Adaptive Hypermedia, volume 15 of New Review of Hypermedia and Multimedia*, Bristol, PA, USA. Technische Universiteit Eindhoven, Taylor & Francis, Inc. ISSN 1361-4568, DOI 10.1080/13614560902801608, pp. 5–38.
- O. Schneider, T. Martens, M. Bauer, A. Ott-Kroner, U. Dick, and M. Dorochevsky (2017). "SensoMot – Sensorische Erfassung von Motivationsindikatoren zur Steuerung adaptiver Lerninhalte", in: *DeLFI 2017 – Die 15. E-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V.*, Universität Chemnitz (in press)
- Technische Universität Ilmenau (2017). "NanoTecLearn – Die Wissensplattform für die Aus- und Weiterbildung in der Mikro-Nano-Integration", <https://www.tu-ilmenau.de/nanoteclearn/> (accessed 21 July 2017)

Mobile Learning in Nature: A digital field guide to plants

Peter A. Henning, Jacqueline Henning, Kordula Manegold, Kevin Fuchs
Department of Computer Science and Business Information Systems
HSKA – Karlsruhe U of Applied Sciences, Germany
e-mail: peter.henning@hs-karlsruhe.de

Keywords: e.g. Mobile Learning, Semantic Web Technology,

Saving the environment and protecting biodiversity is an important educational issue not only in school, but also for adult life-long learning. We have to admit that knowledge of plant and animal biodiversity is very poor and, even more so, further declining in the normal population.

In particular urban citizens do have problems naming species they encounter. For the majority of individuals, perception of nature during recreational walks is mainly focused on green areas as colour, on the possible beauty of a flowering field, but not as an appreciation of the single plants encountered while walking, trekking or hiking by. Conversely, also close to their normal living place areas, urban citizens tend to ignore even the greatest variety of organisms – and have almost no conception of the fact that plants can form complex communities

Consequently, in order to save the environment and biodiversity, a behavioural change towards more biodiversity awareness is needed. This is a severe and future relevant problem, because the decline in precise knowledge has already started to transgress into conservationist organizations. Some have been found to distribute “native seed mixtures” that are very far from being native.

Most programs attempting to increase knowledge of plant and animal biodiversity are targeted on school or university students, for example <http://www.fmnh.org/biodiversity>, not on adults, which are the decision makers in community affairs. Alternatively they concentrate on exotic habitats as for example the program of <http://www.worldwildlife.org>. As welcome as these initiatives are: through media coverage and specific programs adult people learn and know more about foreign species than about the green life in their own neighbourhood.

Even when hiking the forest, fewer and fewer people have a book available to look up the name and properties of a given plant. On return home, this loses importance and is hardly remembered. At most, nowadays photographs of plants or animals are taken by smartphone and posted in one of the knowledge forums on the internet to ask the community what species it is, or in a social media channel just for its beauty.

Smartphones have become the dominant device to collect and share information on the fly, and also to browse through available data (Henning 2017). Today, in almost any major city of the world tourists may use augmented reality for tourist information on selected places. The recognition of the objects is based either on photographs, on local markers or on GPS data.



A smartphone application following the first recognition paradigm is Pl@ntNet, where one takes a photograph of a plant blossom and then is provided with suggestions what this could be. Unfortunately, species recognition from photographic pictures has a very limited recognition rate and is bound to favourite environmental conditions. GPS data for single plants is recorded only in very few applications (Henning et. al. 2011, see also <http://www.wb3-project.de>) and local markers such as QR codes are also not commonly found at plants.

In our project we follow a semantic approach for plant recognition, including information on time of the year (month), place of growth (like e.g. forest, field, wayside), growth form (tree, bush, herb), flowering colour, shape of leaves etc. and place these in a Semantic Media Wiki (SMW) on common plants of Middle Europe.

A Semantic Media Wiki is a semantic wiki engine that enables users to add semantic data to wiki pages. This data can then be used for better searching, browsing, and exchanging of information. Semantic data describe the kinds of relationships between the various data elements, such as <is part of>, <is located in>, <has kin to>. The semantic data is based on an ontology written in the language OWL, i.e. on a formalized knowledge collection, therefore enabling knowledge

sharing and reuse of information in different context (Zander and Henning 2016). It is worthwhile to note that existing and widely know ontologies on plant life, such as *Plant Ontology* <http://www.obofoundry.org/ontology/po.html> are focused on genetics and description of the individual plant parts. They do not contain ecological metadata, have only limited taxonomical capabilities and have no concepts of location and appearance aspects relevant for the general public we are addressing. In the ongoing project we are currently establishing corresponding ontology extensions that may be imported in such existing schemes.

To outline this for a specific example, consider the common wood anemone, or *Anemona nemorosa*. Its most common factual attributes are: it blooms in March/April in certain forests, it is of herbal growth, the blossoms are white, the leaves are pinnate and it is protected.

In our SMW, this factual information is presented in a so called fact box, see figure.

However, the average person strolling through the forest is hardly able to browse through fact boxes. Rather, in obtaining a description, any recognition system is facing far less precise information like: a white to pinkish flower, which blooms in early spring, found in a forest. Or flowers are gone completely, only the typical pinnated leaves turning yellow in May are left and confuse the finder by presenting fall colours in spring.

In our semantic information system, all these improper information collections are related back to the facts to narrow the result, until finally a decision is reached and the fact box is presented to the

questioning system. To this end, our ontology and the SMW contain several categories and properties to connect the information and to help the user to find the required information.

This approach has turned out to be quite difficult to follow, because most of the knowledge we are collecting appears to be rather fragmented. On the other hand we find it very successful in increasing public biodiversity knowledge. Moreover, it has opened further possibilities for modern digital learning systems:

- We produced a smartphone application for Android devices which is coupled to the SMW data base. The user enters the unprecise information detected by observation into this app and receives suggestions from the SMW. This may comprise pictures from typical specimen available, to which the newly found plant may be compared, and progresses to additional information about changes according to season, flower or fruit and what the characteristics of location are. An extension to plant communities is currently being developed.
- It was shown in our FP7 project INTUITEL, how the semantic annotation of knowledge data may be used to produce user-adaptive E-Learning systems. Combining the cognitive map (= the ontology) developed for our biodiversity SMW with existing pedagogical and domain dependent ontologies therefore allows a novel quality of technology enhanced learning

Our SMW pages also contain QR code pictures that may be copied and attached to real plants, linking back into the SMW. While we do not envisage forests where each little flower carries a QR code, we experience great interest for this public knowledge approach from the public administration of some cities.

References

- Fritz, F. et al (2005). Enhancing Cultural Tourism experiences with Augmented Reality Technologies, *The 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST 2005*.
- Fuchs, K, Henning, PA (eds) (2017). Computer-Driven Instructional Design with INTUITEL: An Intelligent Tutoring Interface for Technology-Enhanced Learning. River Publishers Series in Innovation and Change in Education (River Publishers, Delft), ISBN 978-8793519510.
- Henning PA. (2017). Learning 4.0. In: NORTH, K (Ed): Managing Knowledge in Digital Change. Heidelberg: Springer.
- Henning, PA. et. al. (2011). Konzepte eines kommunalen Wissensmanagementsystems. In: *HMD – Praxis der Wirtschaftsinformatik* 277, pp. 86-97.
- Zander, S., Henning, PA. et. al. (2016). A Semantic MediaWiki-Based Approach for the Collaborative Development of Pedagogically Meaningful Learning Content Annotations. In: P. Molli, J. G. Breslin, M-E. Vidal (Eds.), *Semantic Web*.

Machine Learning for User Learning: Supporting Users in Different Stages of Expertise

J. Schuhmacher, M. Kaiser
SAP SE
e-mail: janina.schuhmacher@sap.com

Keywords: Digital Assistance System, Machine Learning, User Modeling

Introduction

Guiding users through their business applications' functionality requires an intelligent digital assistance system that adapts to the user's stage of expertise. Drawing on Event Segmentation Theory (Zacks *et al.* 2007) and Knowledge Space Theory (Falmagne *et al.* 1990), we propose to model the users' domain specific knowledge and their learning process dynamically in the interaction between system and user. In the support process, the system retrieves, from a hierarchically organized case base, the support content that matches the user's knowledge state. Using Case-Based Reasoning (Aamodt and Plaza 1994) as a psychologically inspired machine learning method facilitates the incorporation of user's feedback in the interaction: the system continuously updates its user model to learn how to support the user most efficiently and effectively.

Psychological Background

Psychological research indicates that the way users tackle tasks with their business software depends on their level of expertise (for example, Falzon 1990). Therefore, to give efficient support, an assistance system needs to adapt to the user's state of knowledge rather than providing generic solutions. Anderson (1982) characterizes the movement from novice to expert in skill acquisition as a process of proceduralization. Novices use declarative knowledge to a large extent when they carry out tasks. As learning proceeds, this declarative knowledge becomes implicit. This means novice users may need very specific guidance, for example click-by-click instructions, to solve tasks effectively, whereas for expert users, high-level recommendations are sufficient. An efficient support system should offer detailed guidance only at the exact points where the user struggles.

Business software supports users in a variety of domains, for example HR self-services or finance processes. Tasks such as claiming travel expenses or managing incoming payments can be understood as goal-directed, hierarchical events (Zacks *et al.* 2017). According to Event Segmentation Theory, the way users segment these tasks is determined by their stage of expertise. Users can segment very fine-grained tasks. When they cannot find a certain function in the system, hitting one single button is perceived as one task. At the same time, they can perceive a whole sequence of clicks and mouse movements as a sequence of events that eventually leads to a goal, for instance the goal of buying product from the catalogue. A user may need to complete several fine-grained sub-tasks in order to achieve one super-ordinate coarse-grained task. An appropriate proposal for guidance should match the level on which events are segmented by the user and how this changes throughout the learning process.

Approach

For the digital assistance system to support the user in a comprehensible and personalized manner, knowledge about the system's functionality needs to be represented in a way that is understandable and transparent to enable learning. To this end, we propose a Case-Based Reasoning approach.

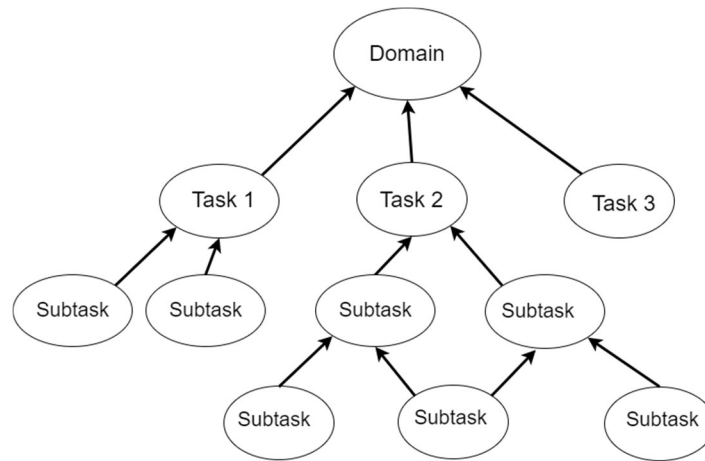


Figure 1: The system's domain knowledge is stored in a hierarchical case base.

Individual differences in expertise are reflected by storing solutions to each domain-relevant task on different levels of granularity in the case base. Thus, the case base is structured as a hierarchical graph, going from a highly abstract description of a given problem to explanations of all related sub-tasks (Figure 1).

To provide personalized support, it is crucial that the computer has available a model of the user's expertise. Knowledge Space Theory provides a framework for modeling the user's expertise: the domain-specific knowledge of the user is represented as the subset of tasks that the individual is capable of solving (Falmagne *et al.* 1990). In the user model, the tasks that compose a domain are organized in a partial ordering graph according to task complexity. We represent the individual's knowledge in terms of the tasks they can solve. Thus, they will only be provided with solutions for the tasks they cannot yet solve on the level of abstraction that matches their expertise.

In each session of interaction with the user, the system provides support according to the user's needs (Figure 2). First, the system interprets the interaction with the user extracting features, explicitly, for example via Natural Language Processing, or implicitly, for example by tracking the user's navigation behavior. In this way, the system determines that the user needs assistance. Based on the features, the system infers the user's goal in terms of the most similar case from the case base. As the system observes how the users utilize the proposed solutions, it continuously optimizes the case retrieval by adapting the feature weights.

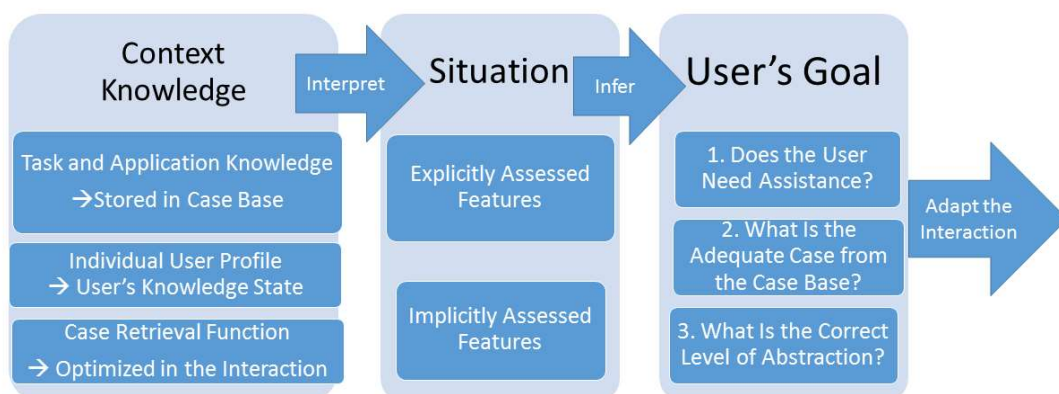


Figure 2: Providing support according to the user's needs

Once the system has successfully identified the appropriate case, it draws on its user model to infer the user's current knowledge state. From the interaction history, the system models which subtasks the user has already solved successfully. Then it guides the user through the completion of their tasks by traversing the case base graph on the level of abstraction that meets the user's current stage of expertise. Imagine the user wants to buy a product from their company's online catalogue (Figure 3). If the user completes the task up to the point of adding the product to the cart but struggles at the point of sending the product order, the system would explain the subtask of sending a product order but not the whole process of buying a product. After each user session, the user model is updated. The computer captures what type of assistance the user needed in the specific context of the interaction and which tasks they achieved successfully so as to adapt dynamically to the user's evolving expertise. To continue our example, if, next, the user requests guidance to complete a more complex task that involves buying a product (for example, ensuring that the team's IT equipment is up-to-date), the assistance system will adapt to the user's advanced knowledge by offering them a high-level recommendation on this point (for example, "You can order the required items from the company catalogue"), rather than guiding them through the whole process.

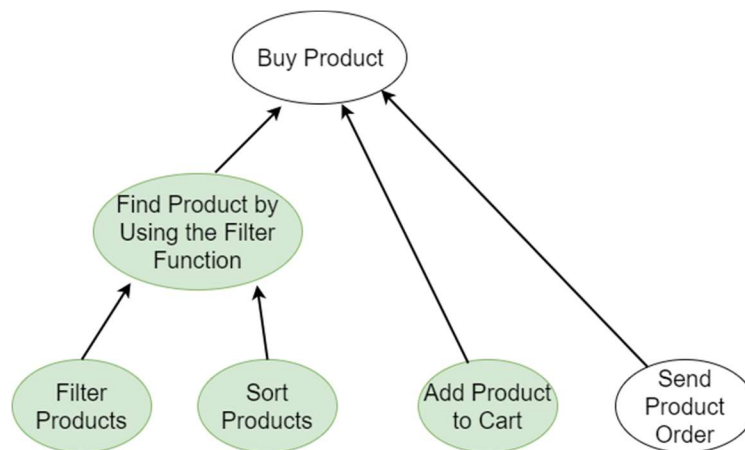


Figure 3: Example of buying a product. Completed subtasks are colored in green.

Originality/Value

Based on psychological theory, the technical support system presented in this paper learns from the user how to propose effective, personalized solutions. At the same time, the user learns from the system's recommendations. As they interact, the system's efficiency increases as it adapts to the user's needs, that is, their current knowledge state. Consequently, introducing such a system offers considerable economic potential by increasing users' performance and satisfaction.

References

- Aamodt, A. and Plaza, E. (1994) 'Case-Based Reasoning: foundational issues, methodological variations, and system approaches', *AI Communications*, 7 (1), 39-59.
- Anderson, J.R. (1982) 'Acquisition of Cognitive Skill', *Psychological Review*, 89 (4), 369-406.
- Falmagne, J.-C., Doignon, J.-P., Koppen, M., Villano, M. and Johannesen, L. (1990) 'Introduction to knowledge spaces: how to build, test and search them', *Psychological Review*, 97 (2), 201-224.

Falzon, P. (1990) *Cognitive Ergonomics: Understanding, Learning and Designing Human-Computer Interaction*, London, GB: Academic Press.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T.S. and Reynolds, J.R. (2007) 'Event perception: a mind-brain perspective', *Psychological Bulletin*, 133 (2), 273-293.

Supporting 21st Century learning through digitally-enabled collaborative learning spaces

Nirosha Holton
South Devon College, Paignton, UK
e-mail: Nirosha.Holton@southdevon.ac.uk

Keywords: Computing, Learning, Collaboration

Introduction and aims

It is often suggested that digital devices are a classroom distraction that hinder the student learning, as well as being heavily linked to negative behaviour in class (Fisher et al. 2011). Reed (2016) states that students tend to not only use the devices individually to stream videos, music or play games in classrooms during a session, but also share these among neighbouring peers which in return distracts others. Equally, we live in an era where most if not all students can be considered so-called Digital Natives, whereas many teachers are Digital Immigrants (i.e. not born into the digital era, but was introduced to various technologies at a later point in life) (Prensky, 2010). This generation gap can create a barrier between educators and students, where the education system is seemingly not compatible with current educational needs of learners (Rosen, 2010; Tapscott, 2009).

Even though it is accepted that learners are from a different era, it is safe to say that technology will play even a bigger part in the future both within and outside of classrooms (Drago, 2015). Hallett (2017) claims that students are highly attracted to multi-platform research resources where culture of online assembly and exhibition of collectively produced knowledge are readily available via applications such as YouTube, Prezi, Tumbler etc. Using such applications, information can be found instantaneously as oppose to following lecture notes or more conventional methods such as reading books or taking notes. These sometimes become a challenge where students and teacher's expectations clash which is confirmed by Robinson et al (2010) where it is stated that "*digital technology creates the biggest generation gap since rock and roll.*" Therefore, it is useful to understand how learners react to the use of technologies, along with collaborative learning spaces, as opposed to traditional classroom environments, in order to determine ways to embrace the new advances to optimise learning. Therefore, the purpose of this evidence-based research is to:

- understand how learners react to the introduction of digital devices within their preferred collaborative learning spaces;
- identify any improvements that can be made, and whether digital technologies become classroom distractions;
- change some of the negative perception of educators around the potential of technologies to enhance learning experience.

Research Methodology and evaluation of current setup

The chosen research methodology was to observe learners in their natural learning environment and then to introduce various environmental and device-related adjustments to a random sub-group in order to observe any changes of behaviour and performances that may result. A group of 20 students (aged 14-16) were chosen for the experiment, on the basis that it was possible to directly affect their

learning environment. Students were not initially informed of the purpose of the research to avoid any behaviour changes, such as Hawthorne Effect, where they will react as expected as opposed to their natural reactions to the experiment (Allen et al, 2010). At the beginning of the experiment, a questionnaire was given to analyse students' initial perceptions of the current traditional classroom setup, and another after the experiment, along with six weekly formative assessments to monitor their academic performance.

The impression of the current classroom set-up was revealed to be mixed with more than half indicating a negative response, and only 40% being explicitly positive (Figure 1). It was considered interesting to contrast this perception of the provided environment with the learning spaces created by students for themselves. To this end, participants were asked to briefly describe their study area at home and reasons why they liked it. The responses here revealed that mention of technology-related aspects was a recurrent theme, which contrasted with the current style of classroom usage (where students may have personal devices with them, but were not using them as a standard element of the lessons). Students were also asked whether they felt digital devices would be a distraction in the classroom (Figure 2), with half explicitly suggesting not, and a third indicating that this was sometimes the case (in fact only one respondent explicitly felt the devices were a regular distraction to them that impeded their learning). The final pre-experiment question asked students whether permitting the use of their own devices (e.g. smart phones, tablets etc) would enable them work better than at present. The response here was overwhelmingly positive, with two thirds indicating yes, and again only one respondent believing their would be distracted.

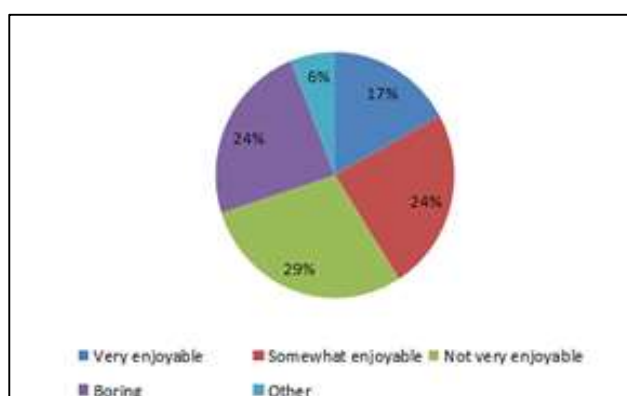


Figure 1: Student perceptions of current classroom setup

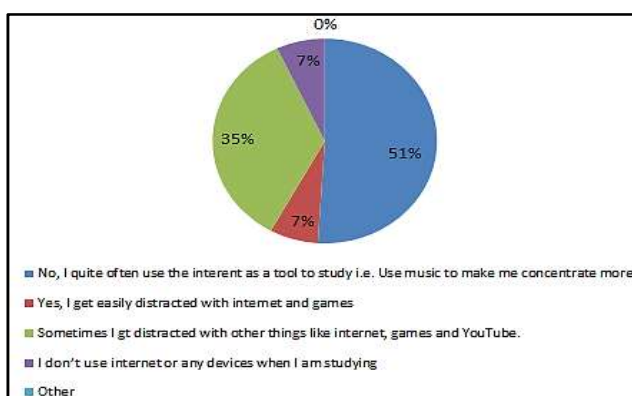


Figure 2: Student perceptions on using digital devices in learning spaces

Analysis of data and evaluation

These findings led into an experiment in which the classroom set-up was changed from more the traditional approach (with a formal classroom layout and lack of technology) to a more casual environment with fewer restrictions. Then the students were introduced to various hi-tech devices including access to BYOD (Bring Your Own Device) and online applications. Students were encouraged to sit anywhere they liked, and the teacher changed the conventional 'teacher's desk' to a more central position of choice. Students were also encouraged to bring any preferred device into classroom and use them, plus a number of additional devices were also made available (tablets and laptops). It was notable that, during the experiment, all the participants opted to use either one or more digital device regardless of any initial reservations about devices being a potential distraction.

From the data collection conducted after the experimental period, all participants found the introduction of digital devices in their learning spaces to have been positive. The most enjoyable features indicate a common theme of '*ownership*', '*own choices*' and '*comfort*', which ties in with the cognitive learning theories where less precision is forced upon students and the learning process

(Gould, 2009). However, the most interesting finding is that even though the main focus of the experiment and the adaptation done was to introduce students to various technologies, none of the students explicitly commented on those technologies in their feedback. This suggests that the devices were immediately assimilated into the learning experience, becoming a natural/transparent part of the learning environment that was “unnoticeable” as a result.

Of course, students welcoming the use of technology is of little benefit if it serves to undermine their learning experience, and it is encouraging to see positive results in this respect as well. Indeed, whereas initial results from the group indicated mediocre performance ranging credit pass to ungradable/fail, the second review (after six weeks) showed an increase across the board in attainments, with only one student scoring the simple pass and three students performing at a distinction level. The third and the final review showed further improvement, with the lowest mark being a Credit pass and some students performing at the highest level (A*). Of course, this could simply have been down to improvements within the group over time, and so it is relevant to add that, when compared to some random samples taken from a similar group in a traditional set-up, the upward progression pattern was not mirrored.

The results show that digital devices can be a useful tool and It would be useful to further the research by broadening the demography of participants to other subjects to take a more meaningful and holistic approach. Consequently, it is also worth thinking about how to change the perceptions of the teachers and the overall educational organisation ethos from conventional methods to embrace the current technological advancements and learner needs instead of discarding technologies as a hindrance.

References

- Allen, R., Davis, A. (2011), Hawthorne Effect. In: S. A. N. J. Goldstein, ed. *Encyclopaedia of Child Behavior and Development*. New York: Springer US, pp. 731- 732.
- Drago, E., (2015). The Effect of Technology on Face-to-Face Communication. *The Elon Journal of Undergraduate Research in Communications*, 6(1), p. 13.
- Fisher, D., Cheung, H., Pickard, V., Chen, T., Cheung, T., Wong, A. (2011), Integrating ePortfolios into teaching and learning:10 CityU case studies, City University of Hong Kong Available at: <http://www.cityu.edu.hk/edge/eportfolio>
- Gould, J. (2009), Cognitive Approaches. In: *Learning Theory and Classroom Practice in the Lifelong Learning Sector*: Exeter: Learning Matters, pp. 43 - 75.
- Hallett, R.,(2017). Surfing, Assembling & Curating – Nurturing Digital Scholarship (SDC Staff development day lecture), Available at: <https://prezi.com/9t1bawnjx8hp/surfing-assembling-curating-nurturing-digital-scholarship/>
- Prensky, M. (2010), Introduction: Our Changing World: Technology and Global Society. In: *Teaching Digital Natives*. London: SAGE Ltd., pp. 1 - 7.
- Reed, L. (2016), *Study: Digital distraction in class is on the rise*, Nebraska: Nebraska Today: University of Nebraska-Lincoln.
- Robinson, K and Aronica, L., (2010). The Element: The Pace of Change. In: *The Element: How Finding Your Passion Changes Everything*. 1 ed. London: Penguin Books, pp. 17 - 21.
- Rosen, L.D. (2010), *Rewired: Understanding the iGeneration and the Way They Learn*, Palgrave Macmillan.
- Tapscott, D. (2009), *Grown up digital: How the net generation is changing your world*. McGraw-Hill, New York.

Operationalizing Learning Behavior as Spatio-Temporal Trajectories

Kevin Fuchs, Peter A. Henning
Institute for Computers in Education
University of Applied Science, Karlsruhe, Germany

Keywords: learning analytics, spatio-temporal database, e-learning, learning behavior

1. Introduction

We introduce a system called the "Hypercube Database". Its purpose is the transformation of learning behavior into spatio-temporal trajectories and their analysis with methods that origin from the field of Geo Informational Systems (GIS). Our main goal behind the Hypercube Database is the provision of real-time analysis of learning behavior within a learning environment. Based on this analysis the Hypercube Database provides a novel approach for the implementation of adaptive learning environments. Each learner is represented by a spatio-temporal trajectory. To model these trajectories, we use a spatio-temporal database approach as it used in GIS. This technique provides two advantages: First, complexity reduction is achieved in the sense that all data is abstracted to trajectories that are purely geometrical objects. These trajectories can then be analyzed only with respect to their geometrical relations. Second, spatio-temporal databases are capable of processing large amounts of geometrical data with high processing speed and therefore allow for real-time analysis.

2. Enhancing Adaptive Learning Environments

Besides learning analytics, the field of adaptive learning environments has attracted a lot of attention. Adaptivity in learning environments aims to emulate intelligent systems that instruct a learner on the basis of pedagogical and didactic knowledge. A human teacher is able to reflect on her theoretical knowledge and to adapt it to practical experience with her students. Moreover she has the ability to take over the learner's perspective. Taking over another person's perspective is the key for teaching and instructing and it is a human privilege because it requires what we commonly call consciousness. Taking over another person's perspective first requires being conscious of oneself in difference to the other person.

When implementing adaptive systems we should respect that the emulation of consciousness is a barrier that we cannot cross with contemporary methods. Even the definition of consciousness can provide grounds for age-long debates. We suggest an economical view restricting ourselves to what computers can actually do and what they cannot. In fact we can give a very clear definition of this by identifying two fields which are deduction and pattern recognition.

Deduction means that we can represent human knowledge in a machine-processable form and then have a machine infer on that knowledge. For example the INTUITEL project (Henning et al., 2014b; Fuchs, Henning, 2017; Fuchs et al., 2016) uses human-defined ontologies to didactically annotate learning material and then infers on them to create learning recommendations in a digital learning environment (Swertz et al., 2014; Henning et al., 2014a). Pattern recognition – as the second field – describes the discovery of similarities. For adaptive learning environments this becomes most interesting in case similar learning behaviors or learning histories are to be found.

With the Hypercube Database – in the following abbreviated as HCDB – we created a system providing the following: didactic experts can define similarities between learning objects for example in a wiki or a Learning Management System (LMS) by using meta-data annotations. Those meta-data are transferred into a spatial model, such that learning histories can be analyzed on the basis of spatio-temporal similarity. The leading idea behind this is, that we leave the meta-data-based definition of didactics to the didactic expert whereas our algorithms discover similarities based on these definitions.

In principle we could treat learner data as time series data for which well-established analysis solutions exist. Nevertheless, we explicitly decided to use a representation with spatio-temporal trajectories instead because it is in particular the concept of spatial nearness that provides a novel way of human-machine symbiosis. We expect the highest benefit from an arrangement in which humans and computers practice a division of labor – with humans defining knowledge and machines inferring on that knowledge and/or discovering similarities and patterns. For this symbiosis machines and humans need a common ground to communicate. Ontologies are an option for this. However, we claim that the transformation of meta-data into a spatial model is another option. This is because thinking in space and time is a natural trait of human imagination and with spatio-temporal database technology we can also have computers operate on such models. Moreover our spatio-temporal model includes time as an inherent dimension of learning processes in a way that ontologies cannot do.

In the sequel we explain the data model of HCDB and how we intent to exploit it for the implementation of adaptive learning environments. Especially, we focus on the question how human knowledge about pedagogy and didactics can be transferred into this model by using meta-data annotations.

3. The Hypercube Model

Basically, we need a method for data abstraction that gives us the ability to join heterogeneous data sources and reduce complexity. HCDB is based on a spatio-temporal database (STDB) with which we model students' behavior in the form of spatio-temporal trajectories. The foundation of HCDB is inherited from the "Hypercube Model" that was developed with the former INTUITEL project (Henning et al., 2014b; Fuchs, Henning, 2017). This Model describes a n -dimensional space with n being the number of learning objects (LO) in a learning environment. LOs represent items like text documents, exercises, test modules or video items. Alternatively, a LO may contain other LOs – like a lesson containing other content. The granularity depends on the learning environment and the content authors.

Each of the n Hypercube dimensions is assigned a numeric value representing the state of progress a learner has performed on the according LO at a certain point in time. This is expressed by a value from the interval $[0,1]$. A learner's position in this space is a vector $L = (l_1, \dots, l_n)$ evolving over time and thus drawing a trajectory in the n -dimensional space. (Fuchs et al., 2016). Figure 1 shows a 4-dimensional hypercube with $n=4$. The bold arrows describe exemplary transitions of a learner between LOs. If the learner finishes one LO by the other one in a disciplined way the learner's behavior results in a movement along the edges of the cube. In case the learner switches between LOs and/or does not process them completely, her movements draw a trajectory in the inner space of the hypercube.

For HCDB we enhanced the Hypercube Model adding k arbitrary dimensions to represent not only a learner's progress on LOs but also any kind of LO-specific metadata or indicators that may be interesting for learning analytics (Fuchs et al., 2016). This creates a k -dimensional subspace within the hypercube describing the individual features of LOs. In this subspace each individual is assigned a k -dimensional time-evolving vector drawing a k -dimensional trajectory in that space. Because all elements of that vector are treated equally HCDB is indifferent of what each dimension represents. A vector element may be a variable indicating a student's position within a learning environment – for example a course that a student has visited, a file she has downloaded or a video he has watched. It may also be a score a student has achieved on a specific item or it may represent an influencing factor like the student's location, the device she is using or even stress measurement data. These time-evolving vectors are predestined to be modeled with a spatio-temporal database (STDB).

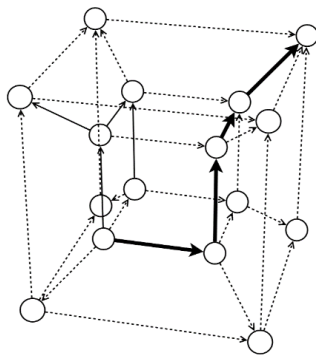


Figure 1: 4D Hypercube

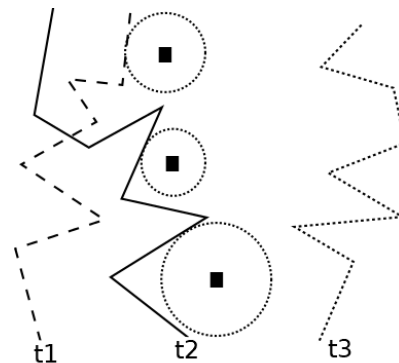


Figure 2: IkNN-Query

4. Implementation of a Spatio-Temporal Database

The core element of our approach is the usage of a spatio-temporal database (STDB) instead of a classical database. Within the Hypercube model, learning behavior is represented by hyperpolylines. Having transformed learners' behaviors into such purely geometric trajectories, we are able to model similarities only through spatio-temporal nearness. Only STDB are able to perform the respective queries efficiently as they are especially designed to index spatio-temporal objects.

Before we developed the design of HCDB, we evaluated existing research in the field of STDB. A particular challenge arises from the fact that our spatio-temporal trajectories are high-dimensional objects that cannot be handled by common STDB. Most STDB are designed for real-world objects with two or three spatial dimensions. Purely temporal databases are for example the ARCADIA database for clinical applications (Combi et al. 1995), Calanda for time series with financial data (Schmidt et al., 1995), ChronoLog running on top of a standard Oracle database (Böhlen, 1994), HDBMS (Sarda, 1987), TDBMS (Tansel, 1997) and TimeDB for general purpose which is based on the ATSQL2 query language (Steiner, 1998; Carvalho et al., 2006a; Carvalho et al., 2006b).

The field of spatio-temporal databases is mostly dominated by Geographical Information Systems (GIS), Network and Facility Management, Land Information Systems (LIS) and Image Processing (Abarham, Roddick, 1999). For example GRASS GIS (Neteler et al., 2012) and GeoToolKit (Balovnev et al., 2000) are Geographical Information Systems while the CONCERT database focuses on management of raster images (Relly et al., 1998; Relly et al., 1997). The SECONDO database is a multi-purpose system for spatio-temporal data (Güting et al., 2004; Almeida et al., 2006).

Due to the nature of their subject these systems mostly provide support for only two or three spatial dimensions. The DEDALE database is capable of dealing with higher dimensions and is based on a constraint database technique that describes spatio-temporal objects as point sets defined by logical constraints (Rigaux et al., 2003; Grumbach et al., 1997). All databases dealing only with two or three spatial dimensions are not an option for the Hypercube Database due to its high-dimensional space. The aforementioned DEDALE system appears to be an interesting candidate because of its constraint approach that can be exploited for high numbers of dimensions (Grumbach et al., 1998; Grumbach et al., 1999). However, the constraint database approach is most appropriate for querying geometric objects containing infinite point sets whereas it is less suitable for querying continuous trajectories. We therefore developed our own database but we used the temporal database TimeDB as its back end building the spatio-temporal functionality upon it.

A crucial element for any database is an efficient indexing mechanism. For spatio-temporal indexing there are various methods available most of which are based on the R-Tree family (Guttman, 1984; Balasubramanian, Sugumaran, 2012). Unfortunately, the performance of those indexing techniques declines significantly with higher dimensions. This is because spatial data is organized with bounding boxes in a search tree. The more dimensions these trees have, the more the respective bounding boxes tend to overlap which results in traversing more sub trees. The X-Tree (BKK96) was therefore designed for higher dimensions. However, HCDB organizes trajectories with a grid index (Chakka et al., 2003) that is most appropriate for indexing trajectories. It provides fast implementation of range and k nearest neighbor (kNN) queries. Within a learning environment we can track the n latest KOs of a learner. We can then perform an incremental kNN based query (IkNN) as proposed in (Z.Chen et al., 2010) which in principle is a sequence of kNN queries. By this we can find the k learning pathways either of other learners or predefined ideal pathways that fit best the learner's previous history of LOs. Figure 2 shows three sample trajectories and three sample LOs (dark squares). The circles around each LO represent the distance to the nearest neighbor trajectory representing either another learner or an ideal learning pathway.

5. Visualization of Students' Behavior

By now we have described how the HCDB can find similar learning trajectories based on spatio-temporal nearness. However, we have not yet made clear how spatial nearness is supposed to be modeled. Let us therefore illustrate the concept of spatio-temporal nearness for learning objects and learning histories. We captured experimental data from real learning environments with the HCDB. We then visualized the data as it is kept by the Hypercube Model using a special coloring technique.

In our first test we identified cooperating students in a course. This test included only a small sample of 15 students, but it is sufficient to demonstrate how in principle we can discover cooperation among students with the concept of spatial nearness. The 15 students edited 47 articles in a Semantic MediaWiki. The articles were grouped by topics. The logging data of the MediaWiki included two data records indicating when a student viewed or edited an article. To transfer the logging data into the Hypercube Model we enumerated the articles whereby each topic group was given a continuous interval of numbers. For the visualization the numbers were associated with colors in the HSB spectrum. So the visualization of each topic corresponded to a subarea in the color spectrum, which further visually indicates spatial nearness in the hypercube space.

In this example we could identify students working on articles of the same topic group. Illustration shows one trajectory for each student. The first row of each trajectory contains the normalized histogram of the student's activities. The second row indicates the student has viewed an article and

the third row shows articles that were edited. The two variables view and edit represent a 2D-Hypercube. Figure 3 shows three topic groups: bluish, greenish and orange-colored. Trajectory 6 represents a student that quit the course very early and therefore shows no more activity. Trajectory 2 and 8 represent students that joined the course later.

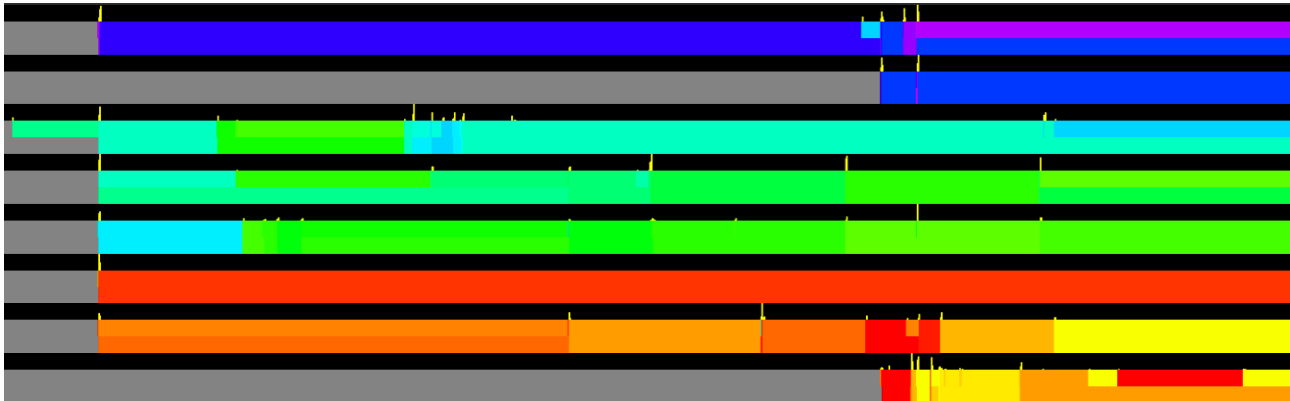


Figure 3: cooperating students (grey means no data available)

Another – more representative example – is students' behavior on a video streaming Platform: That platform contained 55 different video files and 200 users. It provided logging data over a period of four semesters. The logging data was extracted and processed from the internal log files of the platform and it resulted in the following data items for each user: the operating system of the user (Windows, Mac OS, Linux, Android, etc.), his subnet ip (net mask 255.0.0.0), video title and seen percentage. This conforms to a 4D-Hypercube.

Again, the student's activity is represented as a normalized histogram in the first row of each track. Each operating system is represented by one in 10 different colors, each subnet IP by one in 33 colors and each video title by one in 55 colors. The seen-percentage value is represented by a color scale that equates to the spectrum of light running from purple (0%) to red (100%). This enables us to categorize students' behavior by the following criteria:

regular users repeatedly used the video service during periods of several months. Figure 4 visualizes three representative individuals of that group.

occasional users watched the lecture recordings only occasionally. A visualization example of three occasional learners is illustrated by figure 5.

dense users were inactive during longer time periods but used the video platform intensively during short time periods which indicates that they used the lecture recordings for specific purposes like the preparation for an exam. See figure 6 where the outlined areas represent a time period of 10 days.

mobile users were indicated by a fluctuation of the "system" and "subnet ip" variable. Mobile learners were mostly present among the group of regular users. A visualization of three representative mobile users is shown in figure 7.

stationary users rarely or never changed their location or system. For example some dense users showed stationary behavior (see the third dense user in figure 6). One reason might be that when students prepare for exams they often do this in a ritualized way to discipline themselves. This often

involves learning in a particular place like in a library or at home. A visualization of three representative stationary users is shown in figure 8.

Note that the visualizations hold a value for a variable until it is updated by a new event. This means that if a student watches for example 25% of a video and if he is inactive for the following two weeks, the visualization will show the respective color for a period of two weeks.

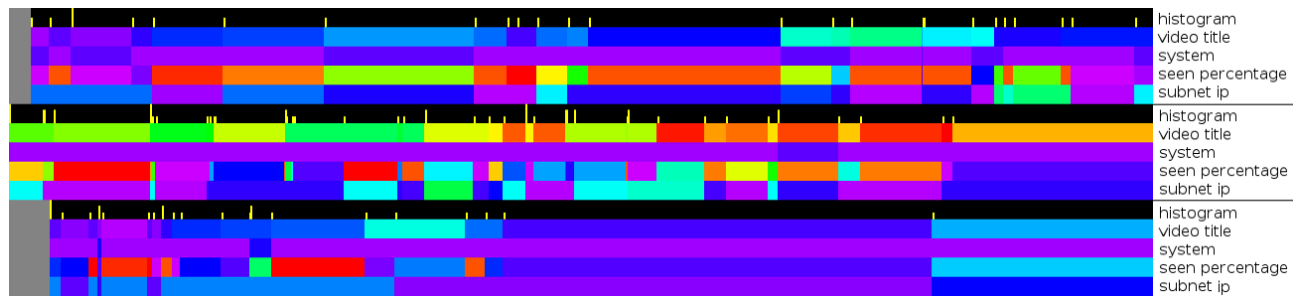


Figure 4: regular users corresponding to a time period of 6 months (gray: no data available)

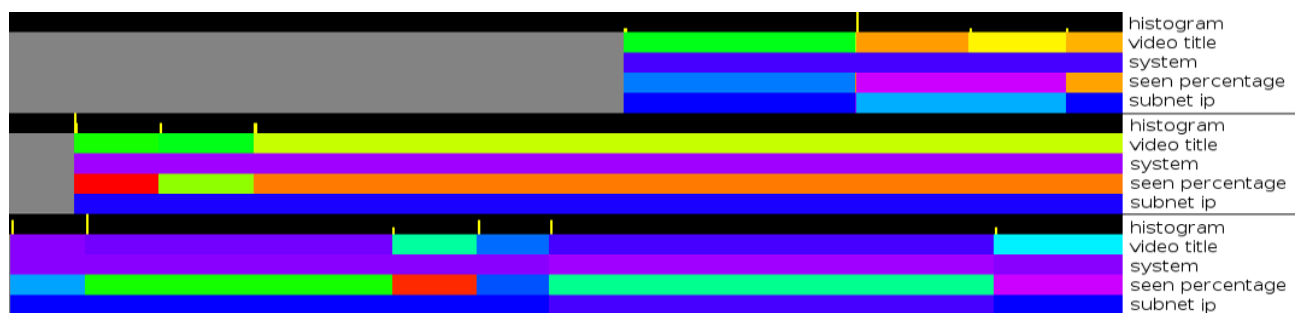


Figure 5: occasional users corresponding to a time period of 6 months (gray: no data available)

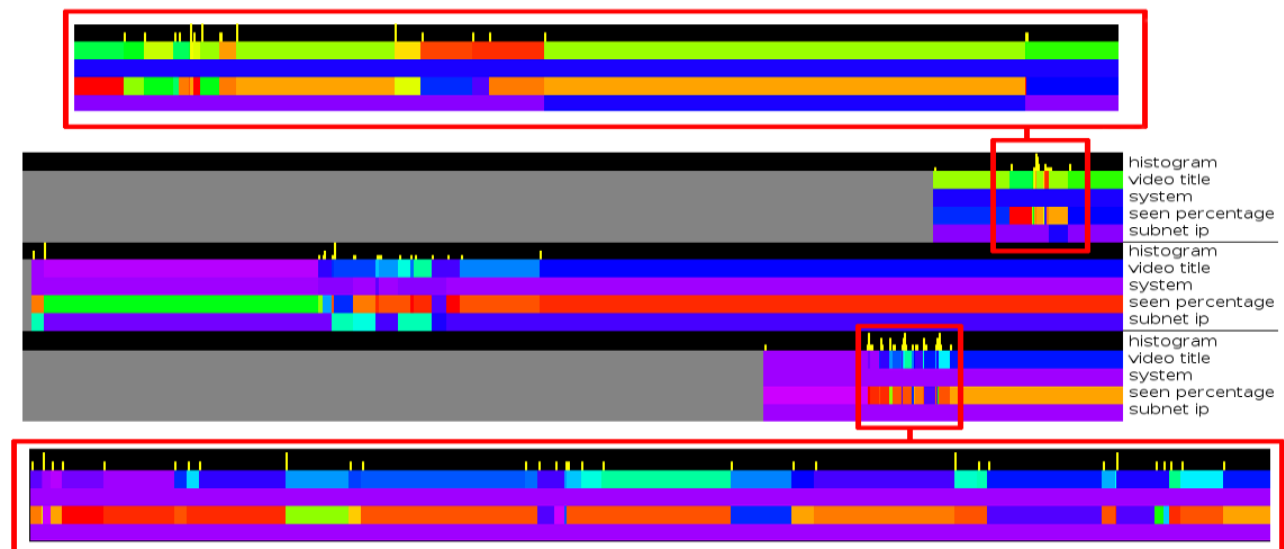


Figure 6: dense users corresponding to a time period of 5 months (gray: no data available)

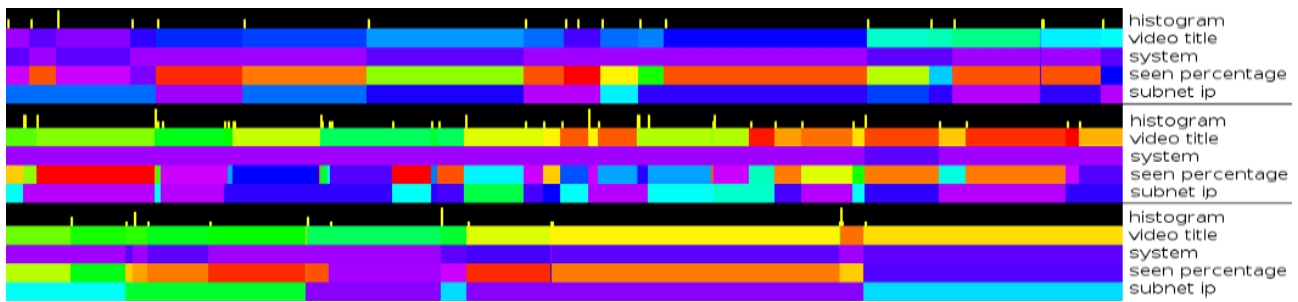


Figure 7: mobile users corresponding to a time period of 4 months (gray: no data available)

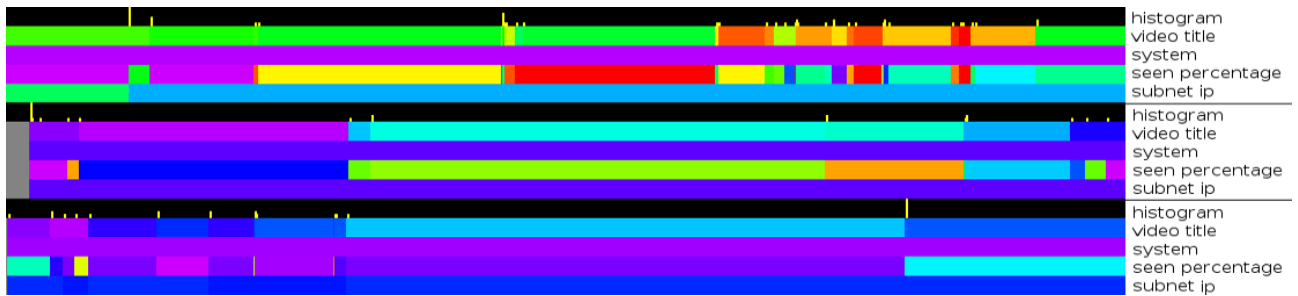


Figure 8: stationary users corresponding to a time period of 4 months (gray: no data available)

The visualizations reveal two types of patterns: spatial and temporal ones. Temporal patterns are constituted by periods and frequencies of activity and the change rate of variables. Spatial patterns refer to measured variables which in these examples are subnet ips, seen percentages, operating systems etc. Together they form spatio-temporal patterns which HCDB operationalizes with data structures and algorithms from the field of spatio-temporal databases.

In the above examples spatio-temporal nearness might be understood as similar wiki articles, the same subnet ip or the same operating system of a user within the same time period. But how can we utilize such nearness for learning environments and what kind of analysis can be useful? First we need a concept to define similar objects in a learning environment. Second we need to define how these similarities have to be transformed into the Hypercube Model so we can utilize it for our spatio-temporal database.

6. Spatio-Temporal Nearness in the Hypercube

In learning environments, the similarity of learning objects can be modeled by meta-data. Most learning environments provide functionality for meta-data annotation. For each piece of meta-data a learning object (LO) has a specific value. In the hypercube model, each meta-data item is represented by a dedicated dimension. If a learner attends to a LO, this is equivalent to a particular position within the hypercube. This spatial position is determined by the entirety of the coordinates that are represented by the respective meta-data. As the learner attends to multiple KOs, she draws a trajectory within the space of the hypercube.

Basically, the content author of a learning environment is responsible for the definition of a meta-data set that is appropriate for the modeling of spatial nearness. The following listing gives a rather arbitrary example of such a LO-based meta-data set:

- ***general items***
 - suitability for beginners (on a scale from 0-10 points)
 - suitability for advanced (on a scale from 0-10 points)
 - suitability for profession group X (on a scale from 0-10 points)
 - suitability for blind people (0-7 points)
 - difficulty level on a scale (from 0-5 points)
 - topic (nominal data represented by numbers)
 - portion of video content (0-100%)
 - portion of text content (0-100%)
- ***learner specific items on this LO***
 - completion level (0-100%)
 - score (1-10 points)
 - grade (1-5)

All items having an ordinal scale or higher are appropriate for modeling spatial nearness. Nominal scales like the topic item can represent spatial information only in the sense of “is the same” or “is not the same”.

7. Finding Similar Learning Trajectories

Suppose we annotate learning material in a learning environment like a Wiki or a Learning Management System the way as we described it in the previous section. When transferred into the HCDB the similarity of LOs with respect to their meta-data equates to spatial nearness in the Hypercube. The more similar meta-data LOs share the closer they are in that space. When a learner works on a sequence of LOs she draws a trajectory inside the Hypercube space. Learners who work on familiar LOs and progress in a similar manner will draw trajectories that are close to each other spatially.

Finding similar trajectories in that sense will offer new opportunities for enhancing adaptive learning environments: adaptive systems can perform learning pathway prediction taking a learner's previous learning history and finding the most similar pathways of other learners. Such a system could also recommend learning mates for pair or group learning. Moreover, we may perform cluster analysis in order to find categories of learning behavior. In another scenario a teacher may define a representative sequence of LOs to give learners a rough orientation in the learning material. The system could then compare this representative sequence with those of the learners and generate according recommendations. Because HCDB also allows adding arbitrary sensor data we could also correlate such categories with this sensor data. This way we may obtain sets of indicators to measure in a learning environment and predict learning behavior.

As we elaborated previously, we implemented an Incremental kNN based query (IkNN) similar to the work in (Z.Chen et al., 2010). Of course this is just one possible way of finding similarities among spatio-temporal trajectories. In contrast to many other solutions the IkNN algorithm does not calculate similarities along all segments of a trajectory. Instead it approximates nearness along a set of query points and a kNN query for each of them. We decided to implement the IkNN algorithm firstly for its simplicity and secondly because it provides good approximation with a minimum of processing effort.

To give an impression of the various issues in the field of trajectory clustering we briefly introduce other researchers' work in this area. Byoung-Kee Yi et al. introduced a method to discover similar

patterns in time sequences (Yi et al., 1998). They especially considered time warping techniques to investigate similarities that occur in different time sequences at different time periods. Nanni and Pedreschi worked on density clustering focusing on the time dimension (Nanni, Pedreschi, 2006). Their approach is based on the OPTICS algorithm (Ordering Points To Identify the Clustering Structure) (Ankerst et al., 1999) which is an enhanced version of the DBSCAN algorithm (Density-Based Spatial Clustering of Applications with Noise) (Ester, 1996). Furthermore, Jae-Gil Lee et al. addressed the problem of similar sub-trajectories instead of treating trajectories only as a whole (Lee, 2007).

By the date of this publication, the entire insertion and indexing of data as well as range, kNN and 1kNN queries have been implemented. Small test scenarios with generic and real data have been conducted to proof the functionality of the Hypercube Database. However, larger field tests with bigger learner populations are still in progress.

References

- Z. Chen et al. (2010). "Searching Trajectories by Locations – An Efficiency Study". In: *Proc. ACM International Conference on Management of Data (SIGMOD)*. 255-266, 2010.
- Almeida, V.T. De; Güting, R. H. Behr, T. (2006). "Querying Moving Objects in SECONDO". In: 7th International Conference on Mobile Data Management, 47-51.
- Ankerst, M. et al. (1999). "Ordering points to identify the clustering structure". In: *Proc. ACM SIGMOD International Conference on Management of Data*.
- Abraham T., Roddick, J. F. (1999). "Survey of Spatio-Temporal Databases". In: *GeoInformatica vol 3, no. 1, pp. 61–99*.
- Balovnev, O. et al. (2000). "Extending GeoToolKit to Access Distributed Spatial Data and Operations". In: *Proc. 12th International Conference on Scientific and Statistical Database Management*.
- Berchtold, S., Keim, D. A.; Kriegel, H. (1996). "The X-Tree : An Index Structure for HighDimensional Data". In: *Proc. 22nd International Conference on Very Large Data-Bases*.
- Balasubramanian, L., Sugumaran, M. (2012). "A State-of-Art in R-Tree Variants for Spatial Indexing". In: *International Journal of ComputerApplications, vol. 42 no.20* 975–8887.
- Böhlen, M.:Managing (1994). *Temporal Knowledge in Deductive Databases*. Dissertation submitted to the Swiss Federal Institute of Technology Zürich.
- V.P. Chakka, A.C. Everspaugh, J.M. Patel (2003). "Indexing large trajectory data sets with SETI". In: *Proceedings of the Conference on Innovative Data Systems Research (CIDR)*.
- Combi ,C. et al. (1995). "Querying temporal clinical databases with different time granularities: the GCH-OSQL language". In: *Proc. Annual Symposium on Computer Applications in Medical Care*, pp. 326–330.
- Carvalho, A., Ribeiro, C., Sousa, A. (2006). "A Spatio-temporal Database System Based on TimeDB and Oracle Spatial". In: *Research and Practical Issues of Enterprise Information Systems, vol. 205*, pp. 11-20.
- Carvalho, A., Ribeiro, C., Sousa, A. (2006). "Spatial TimeDB - Valid Time Support in Spatial DBMS". In: *Proc. 2nd International Advanced Database Conference*.
- Ester, M. (1996). "A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise". In: *Proc. 2nd International Conference on Knowledge Discovery and Data Minin*.
- Fuchs, K., Henning, P.A. (2017). *Computer-Driven Instructional Design with Intuitel: An Intelligent Tutoring Interface for Technology-Enhanced Learning*, River Publishers Series in Innovation and Change in Education.
- Fuchs, K., Henning, P.A., Hartmann, M. (2016). "INTUITEL and the Hypercube Model -Developing Adaptive Learning Environments". In: *Journal of Systemics, Cybernetics and Informatics*.
- Grumbach, S., Segoufin, L., Rigaux, P. (1998). *Efficient multi-dimensional data handling in constraint databases*. BDA, 1998
- Grumbach, S., Segoufin, L., Rigaux, P. (1999). *On the Orthographic Dimension of Constraint Databases*. ICDT, pp. 199-216.

- Grumbach, S. et al. (1997). *DEDALE, A Spatial Constraint Database*. DBPL, pp. 38-59.
- Güting, R. H. et al. (2004). *SECONDO: An Extensible DBMS Architecture and Prototype*. Fernuniversität Hagen, Informatik-Report 313.
- Guttman, A. (1984). "R-trees: a dynamic index structure for spatial searching". In: *Proc. ACM SIGMOD International Conference on Management of Data*, pp. 47–57.
- Henning, P.A. et al. (2014). "Learning Pathway Recommendation based on a Pedagogical Ontology and its Implementation in Moodle". In: *Proc. DeLFI 2014 conference, GI Lecture Notes in Informatics*.
- Henning, P.A. et al. (2014). "INTUITEL – Intelligent Tutorial Interface for Technology Enhanced Learning". In: *Proc 22nd UMAP Conference*.
- Lee, J-G; Han, J., Whang, K-Y. (2007). "Trajectory clustering: a partition-and-group framework". In: *Proc. ACM SIGMOD International Conference on Management of Data*.
- Neteler M. et al. (2012). "GRASS GIS: A multi-purpose open source GIS". In: *Environmental Modelling & Software 31*, pp.124–130.
- Nanni M.; Pedreschi, D. (2006). "Time-focused clustering of trajectories of moving objects". In: *Journal of Intelligent Information Systems vol.1 27, no. 3*, pp. 267-289.
- Relly, L. et al. (1997). "Physical Database Design for Raster Images in Concert". 5th International Symposium on Spatial Databases.
- Rigaux, P. et al. (2003). "Building a constraint-based spatial database system: model, languages and implementation". In: *Inf. Syst. 28(6)*: pp. 563-595.
- Relly, L., Schuldt, H., Schek, H. (1998). "Exporting Database Functionality - The CONCERT Way." In: *IEEE Data Eng. Bull 01/1998*, 21, pp. 43-51.
- Sarda, N. L. (1987). "Design of an Information System Using a Historical Database Management System". In: *Proc. 8th. Annual International Conference on Information Systems*, pp. 86-96.
- Schmidt, D. et al. (1995). *Calanda - a Complete Solution for Time Series Management in Banking*, UBS, Zurich.
- Steiner, A. (1998). *A Generalisation Approach to Temporal Data Models and their Implementations*. Dissertation submitted to the Swiss Federal Institute of Technology Zurich.
- Swertz C. et al. (2014). "A SMS with a KISS. Towards a Pedagogical Design of a MetadataSystem for Adaptive Learning Pathways". Ed Media. *World Conference on Educational Media & Technology*.
- Tansel, A. U. (1997). "Temporal Relational Data Model". In: *IEEE Transactions on Knowledge and Data Engineering*, vol. 9 no. 3.
- Yi, B-K., Jagadish, H.V., Faloutsos C. (1998). "Efficient retrieval of similar time sequences under time warping". In: *Proc. 14th International Conference on Data Engineering*.

Chapter 6

Media

Acoustic Reading Experience: Aligning Sound Events to Text Using Gaze Tracking to Improve Immersion in Reading

David Hill¹ and Matthias Wölfel²

¹ Isle Audio GbR, Brunnenstr. 10, 71679 Asperg, e-mail: david@isle-audio.com

² Department of Digital Media, Furtwangen University, e-mail: matthias.woelfel@hs-furtwangen.de

Keywords: Text Augmentation, Digital Reading, Acoustic Experience, Beyond Gutenberg

The way we present written information has not been essentially changed since the introduction of movable types in Europe by Johannes Gutenberg around 1450s (Kinross 2010). While industrialization was necessitating standardization of type, digitalization offers a liberation of these stringent constraints. Various input parameters can now have an influence on the output parameters. Depending on the design goal these changes have to be performed in real time or can be pre-calculated. Figure 1 summarizes these relationships. In this publication, we propose to augment reading with synchronized sound events.

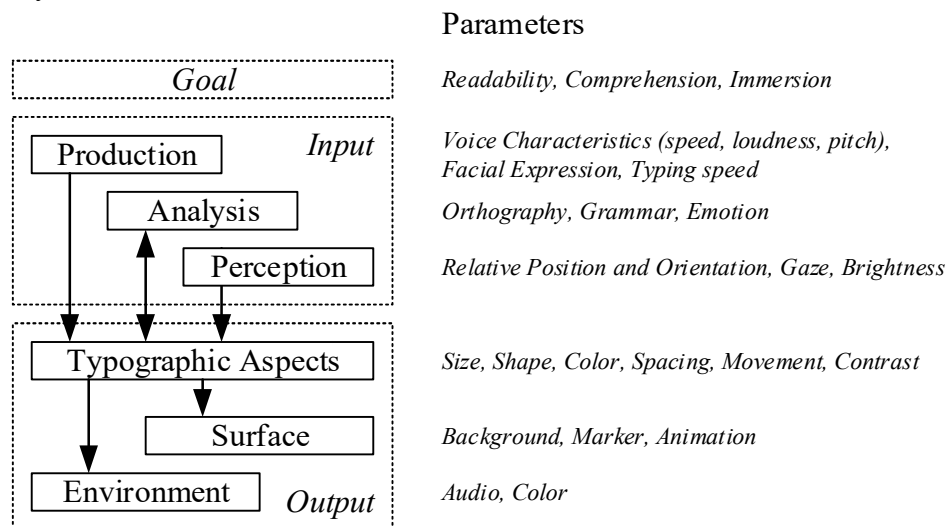


Figure 1: Relationship between the design goal, input and output parameters

1. Goals of Text Augmentation

This section briefly describes the three different goals of text augmentation; namely readability, comprehension and immersion.

1.1. Readability

Readability is defined in various ways. For our purpose readability is limited to the form of presentation and not include content. Therefore, *readability* can be defined as how easily a reader can distinguish individual letters from each other (legibility) as well as words, sentences, and paragraphs depending on typographic aspects.

Examples in supporting readability are reading aids and responsive type. Reading aids include reading guide strips and magnifiers. *Responsive type* (Wölfel and Stitz 2015) changes the shape of typographic characters by adjusting the free parameters according to the physical environment such as angle of view or by compensating for visual constraints; e.g. myopia of a particular reader.

1.2. Comprehension

Reading *comprehension* is defined as the ability to read text, process it, and understand its meaning. Comprehension can be improved for instance by grammar and spell checkers to avoid ill-formed sentences and typos. Another approach in this direction is *voice driven type design* (Wölfel et al. 2015) which adjusts the shape of each single character according to particular acoustic features in the spoken reference. It thus transfers additional information present in spoken language into a visual representation. Because this additional information provides important hints to detect e.g. irony, it can help to prevent misinterpretation.

1.3. Immersion

In this context, we define *immersion* as the deep mental involvement in something and thus the definition used here is broader than the definition of immersion in relation to virtual environments. While immersion is mainly established by the narration it can be supported by additional events.

2. Acoustic Reading Experience

To support the reading experience additional events can be triggered according to the current read word or word sequence. To know what word is currently looked at gaze tracking (Baluja and Pomerleau 1994) is used. The gaze data is post processed by Bayesian filtering to improve the position estimate of the current read word sequence. In addition, it is important to consider the provided font size because fixation durations are significantly longer for smaller font sizes (Beymer et al. 2008).

Similar to the development from silent movies to movies with sound, text can now be augmented by adding synchronized sound events. The *aBook*, the acoustic book is ‘born’. It allows to make scenes, locations, characters, moods, emotions and even every single action of a story experienceable in sound. It provides a creative and innovative medium, that opens a new market for composers, audio designers and writers. It offers a new perspective on how authors can enhance their stories as well as new possibilities of literary expression.

3. Evaluation

In this section, we evaluate the proposed application. To carry out the user study a situation was established, which corresponded to a possible real reading situation. The user was sitting in a chair holding a tablet running the *aBook* application in his hands and wearing headphones. The user study was conducted at Frankfurt Book Fair 2016¹. A total of 201 people, 35% males and 65% females between 8 and 65 years (average 28), participated in our user study. 57 people dealt with the book trade. None of them had ever been exposed to such kind of system before. After testing the system for 5 to 10 minutes the subjects were asked to fill out a questionnaire.

For all closed-ended questions we used a 1-to-5 Likert scale, ranging from (1) strongly disagree, (2) somewhat disagree, (3) undecided, (4) somewhat agree to (5) strongly agree. The results are shown as median scores, average scores and variances to provide a convenient overview. To test for statistical significant differences between females and males as well as people working in the book industry and those who do not we used Cohen’s *d* (Cohen 2012). Because we couldn’t find any relevant difference between the different groups, we don’t present numbers for the individual groups here.

¹ www.buchmesse.de/en/ [15.06.2017].

Question “Do you like the application”: $M=5$ (median); $\mu=4.34$ (average); $\sigma=0.87$ (standard deviation). Thus, most participants liked reading with aligned sound events. Question “Sound and text harmonize well”: $M=5$; $\mu=4.34$; $\sigma=0.66$. This confirms that the sounds are well selected according to the given story. Question “Number of sounds were appropriate”: $M=4$; $\mu=4.07$; $\sigma=1.06$. This indicates that not all people were satisfied with the number of sounds and that a better balance might be possible. Question “Sound supported the reading experience”: $M=4$; $\mu=3.92$; $\sigma=1.16$. Most people agreed that the reading experience could be improved by synchronized sound events. Question “Sound is disturbing”: $M=2$; $\mu=2.33$; $\sigma=1.66$. This low number is in particular interesting because it indicates that the proper sounds at the right moments are not perceived as disturbing. Usually, ambient sound and music—which usually not correlates to the story—is perceived as disturbing while reading (Rugg and Andrews 2010). Question “Sound was well synchronized”: $M=5$; $\mu=4.30$; $\sigma=0.87$. The sound was perceived as well synchronized by most participants. Having a look at Table 1 we see that a clear correlation or negative correlation (in case of ‘sound is disturbing’) exists between the qualities how well the sound was perceived as ‘well synchronized’. Thus, a good alignment seems to be important for the success of the proposed application.

	Sound was well synchronized				
	1	2	3	4	5
Number of votes	4	7	22	59	109
Sound is disturbing	2.00	3.29	3.36	2.26	2.09
Sound supported the reading experience	3.25	2.43	3.36	4.19	4.23
Do you like the application	3.25	2.86	3.61	4.44	4.57
Sound and text harmonize well	2.75	2.71	3.68	4.33	4.60
Number of sounds were appropriate	3.75	2.29	3.05	4.03	4.41

Table 1: Relationship between perceived synchronization and other impressions.

4. CONCLUSION & OUTLOOK

We have introduced and investigated a novel way to combine audio with text. Our evaluation demonstrates that people liked the application, a good alignment seems to be important for its success. So far given text has been augmented in a postprocessing step. To write scripts particular for the aBook might reveal an even higher potential.

References

- Baluja, S., & Pomerleau, D. (1994). Non-intrusive gaze tracking using artificial neural networks.
- Beymer, D., Russell, D., & Orton, P. (2008). An eye tracking study of how font size and type influence online reading. In Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction-Volume 2, pp. 15-18. British Computer Society.
- Cohen, C. (2012). Statistical Power Analysis for the Behavioral Sciences. Routledge; 2 edition, Erlbaum, Hillsdale, NJ.
- Kinross, R. (2010). Modern Typography, vol. 2, Hyphen Press.
- Rugg, M. and Andrews, M. A. (2010). How does background noise affect our concentration. Scientific American Mind.
- Wölfel, M., Schlippe, T., & Stitz, A. (2015). Voice driven type design. In International Conference on Speech Technology and Human-Computer Dialogue (SpeD).
- Wölfel, M. & Stitz, A. (2015). Responsive Type-Introducing Self-Adjusting Graphic Characters. In International Conference on Cyberworlds (CW), pp. 298-305.

Sensory Modalities Influence on Musicians' Latency Measurements

Jorge Medina Victoria
University of Applied Sciences Darmstadt
Jorge.medina.victoria@h-da.de

Keywords: Latency, Metronome, Musical Instruments

1. Introduction

Musical collaboration over networks (e.g. The Internet) between musicians at different places at the same time are nowadays possible. Virtual reality musical instruments and other immersive technologies are growing further. However, those technological advances share the same drawback: latency (Fonseca and Monteiro, 2005), (Mäki-Patola, 2005).

Based on the works of Chafe et al. (Chafe and Gurevich, 2004), Barbosa (Barbosa; Barbosa, 2006) and Boley et al. (Lester and Boley, 2007) the author developed a measure to quantify the latency tolerance range of different musical instruments (Medina Victoria, 2014). This paper presents how the different sensory control mechanisms (auditive and visual metronome) employed to measure the latency tolerance range have an influence in the latency values obtained.

2. Related Work

Previous research established a linear relationship between latency and musical tempo (Schuett, 2002). Musicians performing at higher tempi in beats per minute (BPM) are not able to cope with latency as well as when the performance is, rather slow.

For collaborative performances, results show that some amount of latency up to 20 to 40ms is necessary to interact musically (Schuett, 2002), lower latencies (delays) are also normal when rehearsing in big rooms such as concert halls.

Different factors such as the musical genre, the performed musical instrument and the tactile feedback may interact regarding the ability to cope with latency (Mäki-Patola and Hämäläinen, 2004). On the other hand, experiments on detection of asynchronies between audio and haptic content highlight the importance of considering a holistic view when dealing with human perception (Kim *et al.*, 2006).

In this work, the sensory modalities such as sound and light play a decisive role regarding the control mechanisms (metronome) used in the listening experiment. The metronome (visual or auditive) may have an influence on the results, when measuring the ability to cope with latency.

3. Experimental Setup for Measuring Latency Effects on Musicians

A research design to measure the ability of musicians to perform music in the presence of latency has been previously presented (Medina Victoria, 2014). Participants were music students and professionals. Twelve test subjects were evaluated. Western musical instruments were performed. For this study 3 western guitars, 2 snare drums and one musician for each instrument: alto sax, classical guitar, violin, trombone, trumpet in b, tenor sax and traverse flute. Musicians played a two bars musical score, which is recorded through an audio interface and a digital audio workstation (DAW).

The recorded signal is played back to the musician through headphones and delayed with the DAW. The metronome is used to guide the musician while the returned audio signal from his performance is delayed. The Metronome consists of bar and beats. The bar is the stronger accent, the beat the weaker. The visual configuration are two circles one bigger (bar) than the other (beat). The circles blink indicating bar or beat. By the aural configuration the principle is the same. However, a sound beep is used. Both beeps have the same amplitude. The bar has a slightly different sound as the beat. The third option is an audio-visual metronome, where both metronomes work parallel visual and aural.

Three different measurements for each BPM are taken, indicating the latency value corresponding to the disruption of the performance. Two independent variables are controlled by the experimenter. The latency values range from 0 to 300ms in 10ms increments, while the BPM values are presented randomly from 60 to 240 in 30BPM increments. These 30BPM steps are not very precise. However, an accurate scale implies a longer duration of the experiment, which translates in poor validity due to fatigue of the test subjects. The evaluated BPM values (60, 90, 120, 150, 180, 210, 240) cover the range of the relevant tempo markings in music, such as *largo/adagio* (40-60BPM), *andante/moderato* (56-90BPM), *allegro* (84-132BPM), *presto* (100-180BPM) and *prestissimo* (over 200BPM). The BPM values are approximations made by electronic metronomes and are not standardized.

The control variables are the musical score to be performed and the metronome (3 different configurations). For better internal validity of the experiment, the BPM values are randomized to avoid carry over effects. After measuring different performances on instruments the average of all measurements have been computed as can be seen in Figure 1.

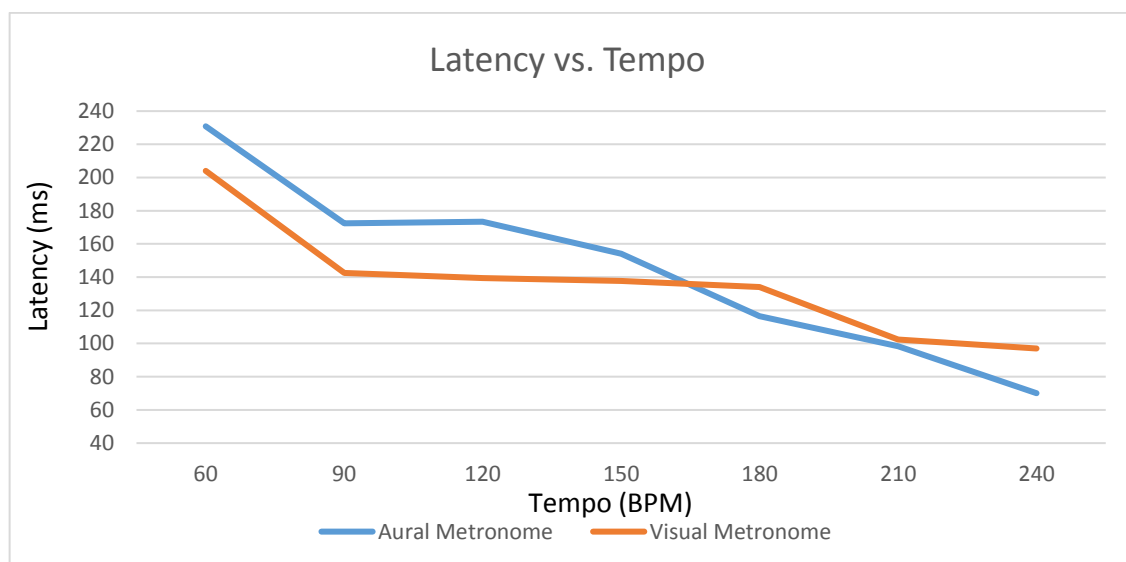


Figure 1: Latency vs. Tempo for measurements controlled with an aural and visual metronome.

The linear negative relationship between tempo and latency confirms previous research (Barbosa, 2006). Some relevant issues are identifiable by observing the graph:

- Between 150BPM and 180BPM, test subjects may cope better with latency when using the visual metronome.

- Up to 150BPM the use of the aural metronome enable musicians to cope better with latency.

The graph in Figure 1 displays the measurement with two metronome settings (aural and visual). In contrast, the graph in Figure 2 shows the differences arising upon using the aural and the audio-visual metronome.

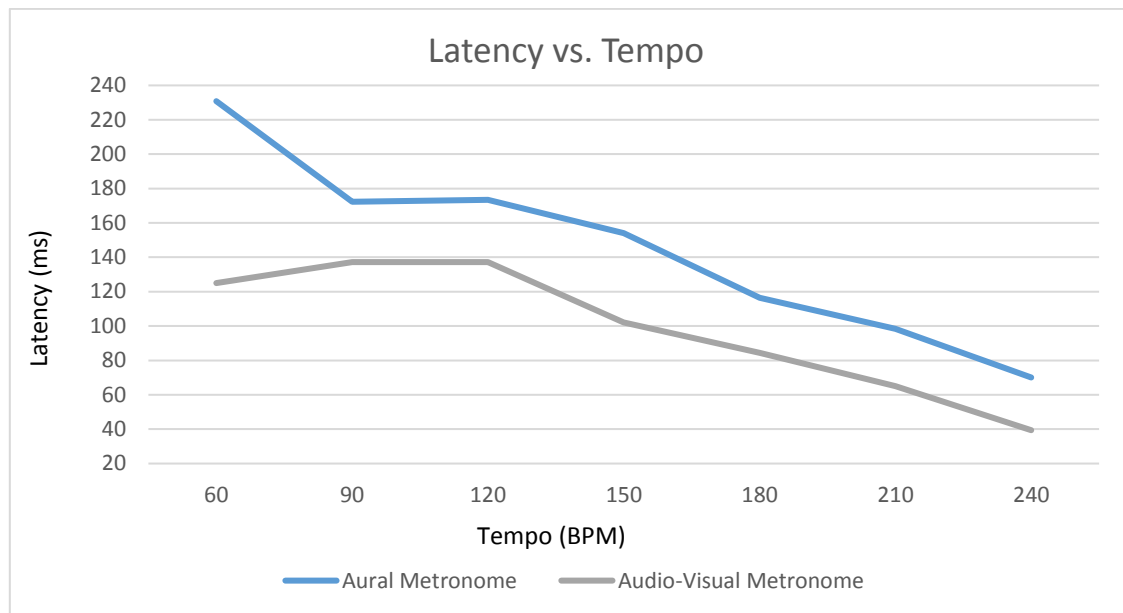


Figure 2: Latency vs. Tempo for measurements controlled with an aural and audio-visual metronome.

It is easy to notice the similarity, despite the different values. Here it is also possible to make some assumptions:

- Regardless of the latency values. The measurement with the assistance of the aural and the audio-visual metronome shows a negative linear relationship.
- There is no crossing point as in the Figure 1.

4. Conclusions and Outlook

In this paper sensory modalities such as light and sound present in a control mechanism (metronome) deliver different results when evaluating the ability to cope with latency in musicians.

The amount of information allows only to observe some trends. More data is necessary for a better interpretation of the results. However, there is a clear influence of the metronome configuration (aural, visual or audio-visual).

It could be assumed that test subjects only use the visual information, when the musical tempi are fast, beginning at 150BPM. This could explain the crossing point for the two curves (aural and visual). On the other hand, when both stimuli (audio-visual) are presented, the test subjects seem to rely only on the aural information.

5. References

- Barbosa, A. (2006). *Displaced Soundscapes. Computer-Supported Cooperative Work for Music Applications*, PhD Thesis, Pompeu Fabra University, Barcelona, Spain, April 2006.
- Chafe, C. and Gurevich, M. (2004). Network Time Delay and Ensemble Accuracy: Effects of Latency, Asymmetry, In: *AES Audio Engineering Society Convention Paper, 117th San Francisco California, October 28-31*, pp. 1–7.
- Fonseca, N. and Monteiro, E. (Eds.) (2005). *Latency Issues in Audio Networking*.
- Kim, S., Martens, W.L. and Walker, K. (2006). *Perception of simultaneity and detection of asynchrony between audio and structural vibration in multimodal music reproduction*.
- Lester, M. and Boley, J. (Eds.) (2007). *The Effects of Latency on Live Sound Monitoring*.
- Mäki-Patola, T. (2005). Musical Effects of Latency, In: *Proc. of Suomen musiikintutkijoiden 9. valtakunnallinen symposium*, 17-19 March, 2005, Jyväskylä, Finland.
- Mäki-Patola, T. and Hämäläinen, P. (2004). Effect of Latency on Playing Accuracy of two Gesture Controlled Continuous Sound Instruments without Tactile Feedback, In: *Conference on Digital Audio Effects*.
- Medina Victoria, J. (2014). An Approach to Quantifying the Latency Tolerance Range in Non-collaborative Musical Performances. In: *AES Audio Engineering Society Convention Paper 136th*, Berlin, Germany, pp. 1–6.
- Schuett, N. (2002). *The Effects of Latency on Ensemble Performance*, Undergraduate Honors Thesis, Stanford, California 94305.

Virtual Reality – Quo Vadis? How to Address the Complete Audience of an Emerging Technology

Andreas Sieß, Sandra Beuck, Matthias Wölfel
Department of Digital Media, Furtwangen University
Robert-Gerwig-Platz 1, 78120 Furtwangen, Germany
e-mail: {siea, beus, woma}@hs-furtwangen.de

Keywords: Virtual Reality, Gender Divide, Content Preferences, Interactivity, Motion Sickness

Abstract

With the introduction of the first affordable high-quality head mounted displays such as the HTC Vive and Oculus Rift in 2016 the “old technology” (Sutherland, 1965) of virtual reality (VR) has experienced quite a boost: it seems foreseeable that the once “geek and nerd” (Mizer, 2013) technology will attract a growing consumer audience and is going to find its place within the universe of entertainment besides its already frequent use in engineering. However, when consumer VR applications are presented the two focused key audiences are on the one hand the cliché young male (intended use: gaming) (Hern, 2014) on the other hand high end professionals (intended use: engineering, etc.). If this trend continues not only a majority of potential users will be excluded but rather the applications and their content cannot unfold the technology’s full potential at all.

1. Introduction

Head mounted displays (HMD) and virtual reality (VR) are absolutely no new phenomena. Already in the second half of the 20th century, there were first prototypes. As early as 1969 Ivan Sutherland developed a “virtual reality head mounted display system” (Burton, 2017). This system was so heavy, it had to be mounted on the ceiling. There were even some older technologies (Sherman, 2002) but the systems were not ready for the consumer market. A big step in the VR-development was the affordable and relative light weighted system “Oculus Rift DK I” (Luckey, 2012) that paved the way for further progression towards a mass market like the “HTC Vive” (HTC 2017).

With a growing number of technologies ready for the consumer market chances are good that sales are also increasing from 2.5 in 2016 to 25.5 million US Dollar in 2021 (Statista, 2017). Forecasts also predict that hardware will only represent a major portion of sales until the market's zenith is reached. After that, the essential part will be the software and actual content (Bitkom, 2016). Thus, VR needs a new approach of content since it is foreseeable that a plain game orientated focus (Mizer, 2013) will only address a fraction of the potential target audience.

The investigation “Gender Differences and VR: A Non-User Survey of What Women Want” from 1994 by Carrie Heeter showed that women were less likely to use VR: “[...] Women also do not play today's video games. Although it is clear that women are not attracted to the current battle-oriented VR experiences, what women DO want from VR has received little attention.” Heeter postulated that women are less likely to play games in general—therefore, VR was considered a male dominated pleasure with focus on gaming—at least in 1994. A newer investigation however analyzed the distribution of computer and video gamers in the United States from 2006 to 2017, by gender and showed a rising number of female gamers in the past years: Almost half of the gamers are female today (Statista, 2017b). How can this trend to digital content be utilized to open VR to a wider audience? And furthermore: Is gaming, the only consumer experience that can be established as the most essential VR content?

2. Study and Hypothesis

We re-evaluated Heeter's study published in 1994 that questioned content preferences from users who have not experienced VR and probably had only a vague idea what it might be (since the study was conducted in 1993 it is not surprising that this technology had not yet found a widespread distribution).

We conducted a similar study that questioned 168 test persons between the age of 17 to 50 (mean age: 27 years) of which 89 were female and 76 were male. In contrast to the cited study by Heeter the majority (more than 66%) of our test population had already experienced applications in VR at least once—whereby we noticed a significant gender difference—maybe due to the mentioned cliché: over 48% of our female test population has not tried VR compared to 21% of males. What needs to be done in order to delight women to give VR a chance in their media consumption habits? Are Heeter's results after almost 25 years—what is an aeon in computer technology—and with up-to-date media perception habits still correct?

2.1. Method

The survey was conducted via an online questionnaire that was online from June 25th to 28th 2017. The participation was voluntary, the used language for the questionnaire was German. We did not track any other data other than the direct questions. The participants' IP-address was anonymized and not saved.

2.2. Results: Gender & VR

Participants were given the same seven hypothetical scenarios as defined by Heeter for the sake of comparison. Each hypothetical scenario could be rated from 1 (= not interested) to 7 (= highly interested). As shown in Table 1 the interest figures look quite balanced—yet there are still some typical male clichés (i.e. the interest in Games and Combat), but there are as well scenarios where females feature, also in 1994, significant higher interest rates than men (i.e. Fitness and Travel).

Gender	Virtual Fitness	Virtual Presence	Virtual Sex	Virtual Travel	Virtual Games	Virtual Learning	Virtual Combat
female 2017	4.75	4.33	3.57	5.41	4.98	4.90	3.03
male 2017	3.67	4.58	4.38	5.18	5.78	4.35	4.78
female 1994	5.6	5.3	4.4	6.3	4.9	5.4	3.0
male 1994	4.9	5.7	5.7	6.1	6.5	5.9	5.6

Table 1: Mean values of interest rates (1 = not interested; 7 = highly interested) of different activities in relation to gender in VR. These values were calculated using the mean of each individual gender's average as normalization. The rows "female 1994" and "male 1994" cite the figures from Heeter 1994.

With this finding we can state, that although the technology needs to advance the salient point of gender differences in VR is the differing demand of content and application scenarios that women ask for. This effect is similar to conventional media: i.e. the significant gender differences in TV content preferences (IfD Allensbach, 2016) that show that males and females have different media consumption habits in general.

It is interesting to note that the average interest in VR in 1994 was in average 5.38 (male: 5.77; female 4.99) while in 2017 it was 4.55 (male: 5.67; female 4.42). If this effect, however, is caused by a cross-

cultural self-report measure difference¹ or by an actual lower interest in VR cannot be determined by the given results.

Gender	VR Experience	#	Virtual Fitness	Virtual Presence	Virtual Sex	Virtual Travel	Virtual Games	Virtual Learning	Virtual Combat
female	with	43	4.93	4.54	3.86	5.46	5.20	4.96	3.37
female	without	46	4.57	4.12	3.27	5.37	4.75	4.84	2.69
male	with	63	3.73	4.57	4.37	5.32	5.87	4.37	4.84
male	without	13	3.61	4.58	4.39	5.03	5.68	4.33	4.72

Table 2: Mean values of interest values in different activities in VR (1 = not interested; 7 = highly interested)

In comparison to Heeter's results there seems to be a change in gender differences: The relative higher interest for males in comparison to females² in 1994 is 16% while in 2017 is only 6%. As shown in Table 1 females reached in our study higher mean values in the categories "Virtual Learning", "Virtual Fitness" and "Virtual Travel". Only in the mentioned category "Virtual Fight" and in "Virtual Games" as well as "Virtual Sex" men are still more interested than women. Almost identical results were noted at "Virtual Presence". This gender-gap gets even more close when the individual experience with VR of the test subjects is put into consideration: Table 2 subdivides both gender groups into parts with and without experience. As we can see: Having tried VR at least once increases the interest in each scenario and brings both genders closer together (male: 4.72; female: 4.62; relative difference: 2%).

Regarding the statistical relevance of the impact of gender or VR experience respectively on the interest figures of the particular scenario, we conducted an analysis of variance (ANOVA). As Table 3 shows we can state that with a few exceptions gender as well as VR experience can have a significant influence³ on the different scenario's ratings.

Attribute	Virtual Fitness	Virtual Presence	Virtual Sex	Virtual Travel	Virtual Games	Virtual Learning	Virtual Combat
VR Experience	0.39	0.13	<0.001	0.07	<0.001	0.96	<0.001
Gender	<0.001	0.065	<0.001	0.25	<0.001	0.08	<0.001

Table 3: p values of an ANOVA on the impact of the participant's attribute on the preference of different scenarios.

2.3. Paradigm shift

Looking beyond the cliché confirmations of gender stereotypes, it could be stated that one of the most interesting phenomena is not gender-specific at all: While the study by Heeter was only able to question the interest in VR in general without the ability to differentiate between experience levels we can now present these results. In Table 2 and 3 we can easily observe that interest in every category is not only correlated to the state of the test person's gender but rather also on his/her individual VR experience: persons with VR experiences are more likely to be interested in all kind of investigated

¹ These cultural determined differences can have a significant impact on the findings (Harzing, 2006): I.e. Dolnicar/Grün (2007) found out that over 6% of cross-test-group differences can be attributed to cultural differences in survey response patterns.

² In the 90ties computer and media technology was still considered as a male dominated culture (Bjorkman et al. 1998). This gender gap was closed at the latest with the pervasion of smartphones in popular media usage habits and even reversed in 2015 when females used their smartphone longer than their male counterparts on average (Andone et al., 2016).

³ Small p-values (< 0.05) are indicators for statistical significance (Rice, 1989).

activities, regardless of gender. If this effect, however, is caused by VR experience, a-priori VR interest or can be explained with a general higher interest in new technologies stays unclear⁴.

With this results in mind, chances seem good that the remaining 48% (female) and 21% (male) unexperienced-VR-users would have a higher interest in VR if they would try at least once.

Used VR	Virtual Fitness	Virtual Presence	Virtual Sex	Virtual Travel	Virtual Games	Virtual Learning	Virtual Combat
never	3.74	4.23	3.42	4.57	4.61	4.54	3.16
once	4.24	4.75	4.34	5.16	5.23	4.62	3.71
more often	4.40	4.47	4.01	5.53	5.73	4.71	4.32

Table 4: Mean values of interest rates (1 = not interested; 7 = highly interested) of different activities in relation to VR experience, regardless of gender. These values were calculated using the mean of each individual gender's average as normalization.

The values suggest that if a person has tried VR at least once the rates of interest grow significantly in every given scenario (excluding "Learning"). With this new paradigm in mind the question arises what has to be done in order to address non-VR users to get in touch with this technology. What kind of content and environment do they prefer and what kind of interactions within VR should be possible to lower the threshold for this target audience?

2.4. Desired content approach of different test groups

In this section the participants were asked what kind of visual content they would prefer in general. One of the most significant findings here was the declining crave for visual realistic environments with inclining VR experience. Also, the desire of world-like physical paradigms (e.g. gravitational force) is declining when users can draw on prior VR experience. When asked about the desired look of their own avatar inside the virtual environment non-VR users preferred a rather realistic approach (mean: 3.71 on a scale from 1 (=realistic) to 7 (=fantasy)) whereas frequent VR-users went after a fantasy-like design (4.54). In terms of the visual approach of the given environment non-users also preferred a more realistic (mean: 4.32 on a scale from 1 (=fantasy) to 7 (=realistic)) design compared to test persons with VR experience (3.70). When questioned on the importance of realistic behavior of virtual environments (e.g. gravity) it can be stated that the desire for nature-like paradigms also decreases with VR experience: Non-users answered with a mean of 3.82 (on a scale from 1 (=very important) to 7 (=not important at all)) while test persons with VR experience answered with a mean of 4.54.

In summary the non-VR users tried to picture the material world within the virtual environment. In order to open up the potential user group of inexperienced persons we can state that realistic environments, paired with realistic-looking avatars and realistic physics are elements that non-VR-users expect. Once they are familiar with this technology their tolerance to different approaches of virtual environments rises. What also should be mentioned is the fact that the demand of high quality images is a little lower (mean: 5.23 on a scale from 1 (=unimportant) to 7 (=highly important)) than the average of experienced users (mean: 5.79)).

⁴ The question after their tech-savviness was answered by non-VR-Users with a mean of 4.17 (1 = not tech-savvy at all; 7 = very tech-savvy) in comparison to a mean of 5.66 for frequent VR-users.

2.5. Desired activities of different test groups

The test population was asked what kind of activities they preferred to have an impact on their experience in VR. They were given 16 activities that could be ranked in their subjective importance from 1 (= very little) to 7 (= very much).

When looking at Table 5 it can be stated that the most important activity that should influence the virtual experience is *Exploring* that achieved in every user group the highest values and its overall mean of 6.02 outnumbers every other activity's measured value. With inclining experience in VR *Exploring* gains in relevance and can be considered as *very important* in every user group.

Activity	Often used VR <i>n</i> =71	Once used VR <i>n</i> =39	Never used VR <i>n</i> =57	Mean	Male, normalized Mean	Female, normalized Mean
Visual Quality	5.58	5.77	5.23	5.53	5.74	5.53
Being a Scientist	4.86	4.44	4.39	4.56	4.87	4.50
Exploring	6.24	6.08	5.75	6.02	6.03	5.98
Leadership	3.72	4.10	3.95	3.92	4.28	3.89
Dexterity	4.85	5.15	5.09	5.03	4.92	5.27
Luck	3.18	3.59	3.96	3.58	3.72	3.68
Fighting	3.24	3.31	3.02	3.19	3.93	2.86
Communication	4.92	5.15	4.70	4.92	4.93	5.12
Creativity	5.52	5.69	5.47	5.56	5.53	5.77
Physical Fitness	3.83	3.67	4.32	3.94	4.02	4.26
Risk-Taking	4.48	4.56	4.65	4.56	4.84	4.51
Solving a Mystery	5.03	4.92	4.67	4.87	4.87	5.10
Learning while Playing	5.41	5.54	5.02	5.32	5.72	5.25
Teamwork	4.86	4.85	4.77	4.83	4.87	4.97
Prior Knowledge	3.24	3.23	3.58	3.35	3.73	3.31
Competition	3.51	3.97	3.44	3.64	3.98	3.52

Table 5: Mean values of different given activities in virtual environments across user groups. (1=not important; 7=very important)

Regarding the question which activity is least desired in VR the measured values differ slightly between the user groups: Experienced users who have often used VR disliked the activity *Luck* whereas probands who never tested VR before showed the largest disapproval at the activity *Fighting*. The overall least desired activities were *Fighting* (mean: 3.19), *Prior Knowledge* (mean: 3.35), and *Luck* (mean: 3.58)—although the category *Prior Knowledge* is among the activities that was ranked higher by unexperienced users in relation to experienced user groups (other activities that ranked higher among probands that never used VR: *Physical Fitness*, *Luck*, and *Risk Taking*).

Since our study featured vast differences in VR experience between male and female test persons we normalized the results by calculating the mean of each experience level among each gender separately and used this values as basis for the final mean calculation.

The largest difference between each gender's preference could be measured at the activity *Fighting* with a distance of 1.06, followed by *Learning while Playing* (distance: 0.47), and *Competition* (distance: 0.46). The smallest distances that could be measured across the genders are also among activities that produced the most extreme results in general: Males as well as females correspond with their favorite activity *Exploring* (overall mean: 6.02, distance between male and female: 0.05) as well as their denial of *Luck* (overall mean: 3.58, distance: 0.04).

2.6. Desired genres of different user groups

In her study Heeter (1994) also included the question which genres—in terms of movie/TV genres—the test persons preferred. Although this question seems a little bit outdated on first sight because VR can offer a lot more interactivity than sheer movies by now we however decided to include this question as well in our recent study because it offers known evaluation benchmarks that even test persons with zero VR experience (still roughly $\frac{1}{3}$ of our study's test population) can easily relate.

The test persons were given 16 different genres that they could rank from 10 (= would enjoy very much) to 0 (= would not enjoy at all).

TV genre	Often used VR n=71	Once used VR n=39	Never used VR n=57	Mean	Male, normalized Mean	Female, normalized Mean
Adventure	8.21	7.49	7.04	7.58	7.66	7.53
Comedy	4.03	4.92	3.79	4.25	3.42	4.74
Documentary	7.52	7.41	7.26	7.40	7.89	7.25
Drama	4.41	4.85	4.44	4.57	4.71	4.79
Erotic	4.56	5.49	3.26	4.44	5.81	3.70
Horror	4.90	3.69	2.56	3.72	3.81	3.43
International Culture	6.20	5.33	5.25	5.59	5.48	5.97
Detective/Cop	6.04	5.74	5.86	5.88	5.83	6.11
Live Events	6.28	6.95	6.53	6.59	6.37	7.00
MTV	5.82	6.79	6.84	6.48	6.58	6.66
Science Fiction	7.10	6.44	4.72	6.09	6.79	5.33
Soap Opera	1.59	2.10	1.98	1.89	1.70	2.37
Splatter	3.59	3.08	1.93	2.87	3.53	2.44
Sports	4.44	5.36	5.02	4.94	5.11	5.24
Talk	1.63	2.21	1.60	1.81	1.51	2.05
Science & Education	7.28	7.08	6.86	7.07	7.61	7.01

Table 6: Mean values of different movie/TV genres across user groups. The participants could rank the given experiences from 10 (would enjoy very much) to 0 (would not enjoy at all).

As seen in Table 6 the most interesting genres are *Adventure*, *Documentary* and *Science & Education*. This correlates also with the learnings about activity preferences in paragraph 2.5 where *Exploring* being the most favorite activity. The least desired genres were *Talk* (mean: 1.81) and *Soap Opera* (mean: 1.89) although a small gender gap emerges with a distance of 0.54 (*Talk*) respectively 0.66 (*Soap Opera*). The largest distances between measured preferences could be observed at the genres *Science Fiction* (distance between normalized male and female mean: 1.46), *Erotic* (distance: 2.11), and *Comedy* (distance: 1.32).

2.7. Desired sensations of different user groups

The participants were asked about their interest in four different sensations that they could rank on a scale from 1 (= very little interested) to 7 (= very much interested). The sensations could be classified in *Physical Sensations*, *Visual and Auditory Sensations*, *Intellectual Stimulation*, and *Emotional Experiences*. Following the style of Heeter's (1994) study each sensation was questioned by providing a small example in order to facilitate the inner visualization. *Physical Sensations* were introduced by the example "Imagine being able to fly like a bird or jump down a cliff without worrying about injury." *Visual and Auditory Sensations* were questioned with the example of "breathtaking landscapes, detailed facial expressions of an alien", *Intellectual Stimulation* with

“content and experiences that make you think or learn”, and *Emotional Experiences* with “experiences that evoke emotional responses”.

Sensations	Often used VR	Once used VR	Never used VR	Mean	Male, normalized Mean	Female, normalized Mean
Physical sensations	5.96	5.69	5.89	5.87	5.95	5.80
Visual and auditory sensations	5.80	5.59	5.12	5.52	5.39	5.57
Intellectual stimulation	5.27	5.28	5.00	5.18	5.45	5.13
Emotional experiences	4.86	4.95	4.46	4.74	4.98	4.69

Table 7: Mean values of different sensations across different user groups. The interest in each sensation could be ranked from 1 (= very little interested) to 7 (= very much interested)

As Table 7 shows: There is a visible difference between interest values of frequent VR users and non-VR users—however an ANOVA reveals that these figures differ in their statistical significance: The p-values vary from 0.8 (*Physical sensations*), 0.3 (*Intellectual Stimulation*), 0.2 (*Emotional Experiences*) to 0.02 (*Visual and Auditory Sensations*)—therefore only *Visual and Auditory Sensations* feature significant differences between users and non-users. All other sensations show comparable interest values, whereby *Physical Sensations* are desired most, and *Emotional Experiences* are desired least within all user groups.

2.8. Motion Sickness: Prevalence and Causes

Divergent from Heeter’s study we also asked our participants if they ever experienced physical discomfort while using a VR application. If that was the case, they could select from five given answers or had the opportunity to add a free text as well. Since our study featured 33% test persons that had never tested VR we only put participants into consideration that used VR at least once for the following paragraph.

The majority (about 55%, $n = 60$) of the study’s participants with VR experience already went through a form of physical discomfort when using a HMD (Head Mounted Display). Of these 60 probands 56% were male and 44% were female. When asked what effect or feature of VR actually caused the experienced discomfort, 59.8% of this test group answered that there was too much parallel information they had to handle and criticized the excessive speed of the provided content making it unable to process adequately. 53.7% found fault with improper mapping between the displayed image and the underlying technics, 46.3% criticized technical issues (i.e. distortion, insufficient lens quality, bad wear comfort), whereas only 17.1% deprecated insufficient graphics (i.e. bad textures or wrong shadow rendering). 19.5% answered that they are sensitive to multimedia applications in general and their discomfort is not exclusively related to HMDs.

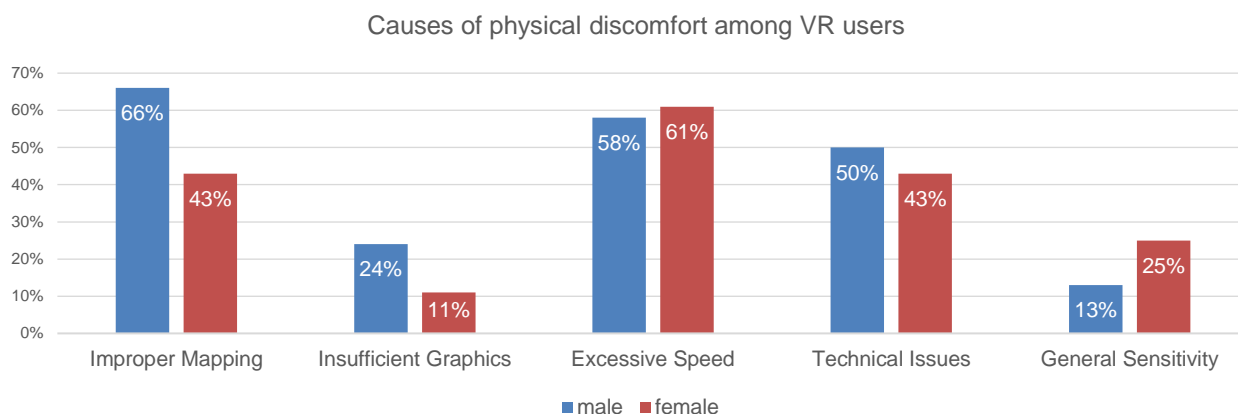


Figure 1: Causes of physical discomfort (multiple choice question) across the genders.

As Figure 1 indicates: There are some subtle gender differences in the cause of the individual discomfort: On the one hand males seem to be more sensitive to improper mapping and insufficient graphics than females, who face on the other hand more problems with general sensitivity in VR.

In the free text area, the participants could point out additional factors, if needed. 19 probands gave additional information here, such as the lack of motion liberty, the absence of fixed points in the virtual world, the physical discomfort when using glasses in combination with a HMD, too strong motion blur effects and problems with first person perspective in general—regardless of the used output device.

Since this study's focus is on content we tried to find relationships between content preferences and prior experiences of discomfort related to VR. By using an ANOVA, we found out that there is a correlation between the desire for the *“virtual world feels and behaves like the real world”* is higher, when the user has experienced in-use-discomfort before ($p = 0.03$). Also the activity *Physical Fitness* (see paragraph 2.5) is ranked significantly higher ($p = 0.05$), when the test person had problems with discomfort before. A comparable relationship was found with the activity *Competition*, that is ranked lower ($p = 0.08$) by participants with VR-related uneasiness.

3. Conclusions and Outlook

VR is a rapidly growing market (Statista 2017). When this technology wants to continue its triumph, it must not exclude large potential user groups. On the one hand our re-evaluation paired with the new introduced parameter of VR experience has shown the different needs of women for virtual environments, on the other hand it proposes a strategy how to address an audience that has not yet experienced VR at all. Regarding this finding we believe a rethinking of gender stereotypes in VR is necessary. Our results also show that the VR market is far from being covered, as the forecasts and the unfulfilled wishes of many users show there is a great potential for this technology in the future. It is meanwhile possible to use VR without spending a lot of money, but this achieves nothing if the right contents are lacking.

As already mentioned, the future VR market will no longer consist of offering hardware, but from suitable applications. Additionally, one should be aware that potential technical problems are not on the long-term, since technologies such as VR are constantly improving. Thus target-group-oriented VR applications are the new challenge in this area. These groups are not just women and men, target groups should be also separated by user's interests. VR consumer technology today primarily stands

for "gaming". As long as this is the case, users with other interest may be less interested in using or buying a VR-system in general.

Acknowledgements

This project was partly funded by MFG Medien- und Filmgesellschaft Baden-Württemberg.

References

- Andone I, Blaszkiewicz K, Eibes M, Trendafilov B, Montag C, and Markowetz A (2016). "How age and gender affect smartphone usage". In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp '16)*. ACM, New York, NY, USA
- Bjorkman C, Christoff I, Palm F, Vallin A (1998). "Exploring the Pipeline: Towards an Understanding of the Male Dominated Computing Culture and Its Influence on Women". In: *SIGCSE Bull., Vol. 30, Issue 2*.
- Burton, Robert (2017). "Ivan Sutherland". In: url: http://amturing.acm.org/award_winners/sutherland_3467412.cfm [28.10.2017]
- Bitkom (2016). "Prognose zum Umsatz mit Virtual Reality in Deutschland von 2016 bis 2020 (in Millionen Euro)". Statista - Das Statistik-Portal, Statista, de.statista.com/statistik/daten/studie/604199/umfrage/prognose-zum-umsatz-mit-virtual-reality-in-deutschland/ [28.10.2017].
- Dolnicar S, Grün B, (2007) "Cross- cultural differences in survey response patterns". In: *International Marketing Review, Vol. 24 Issue: 2*.
- Google (2017). "Google Cardboard". In: url: <https://vr.google.com/cardboard/> [28.10.2017].
- Harzing A (2006) "Response styles in cross-national survey research: A 26-country study". *International Journal of Crosscultural Management, vol. 6/2*.
- HTC (2017). "HTC Vive". In: url: <https://www.vive.com/> [28.10.2017].
- Heeter, C (1994). "Gender Differences and VR: A Non-User Survey of What Women Want". In: *Virtual Reality World, Issue March 1994*. Available at <http://commtechlab.msu.edu/randd/research/gender.html> [26.06.2017].
- Hern, A (2014). "Oculus Rift: 'virtual reality is still niche – it's hard to service millions'". In: The Guardian, September 23 2014. Available at: <https://www.theguardian.com/technology/2014/sep/23/oculus-rift-virtual-reality-niche-max-cohen-morpheus> [27.06.2017]
- Luckey, Palmer (2012). "Oculus Rift: Step Into the Game". In: url: <https://www.kickstarter.com/projects/1523379957/oculus-rift-step-into-the-game> [28.10.2017].
- Mizer, N. J. (2013). "Nomadic tactics in post Revenge geek culture". In: *Proceedings of the Pop Culture Association/American Culture Association (PCA/ACA)*. Washington, DC: Wiley.
- Rice, W (1989). "Analyzing Tables of Statistical Tests". In: *Evolution 43, no. 1*.
- Sherman, W R and Craig, A B (2002). *Understanding Virtual Reality: Interface, Application, and Design* (The Morgan Kaufmann Series in Computer Graphics). Morgan Kaufmann.
- Statista (2017). Prognose zum Umsatz mit Virtual Reality in Deutschland von 2016 bis 2020 (in Millionen Euro). Statista. Available at: <https://de.statista.com/statistik/daten/studie/604199/umfrage/prognose-zum-umsatz-mit-virtual-reality-in-deutschland/> [28.06.2017].
- Statista (2017b). Distribution of computer and video gamers in the United States from 2006 to 2017, by gender. Statista. Available at: <https://www.statista.com/statistics/232383/gender-split-of-us-computer-and-video-gamers/>. [20.10.2017]
- Sutherland, IE (1965). The ultimate display. Multimedia: From Wagner to virtual reality.

IfD Allensbach. (2016). Ranking der beliebtesten Fernsehformate und -genres in Deutschland nach Geschlecht im Jahr 2016. Statista. Available at: <https://de.statista.com/statistik/daten/studie/291182/umfrage/umfrage-in-deutschland-zu-den-beliebtesten-fernsehformaten-nach-geschlecht/> [30.06.2017].

Fostering Human-to-Human Communication and Collaboration in Physical Space by Playful Augmentation and Information Inequality

Manuel Conzelmann, Sabrina Lange, Matthias Wölfel
Department of Digital Media, Furtwangen University
Robert-Gerwig-Platz 1, 78120 Furtwangen, Germany
e-mail: matthias.woelfel@hs-furtwangen.de

Keywords: Computer-Assisted Human Interaction, Playful Augmentation, Information Layer, Blending Virtuality and Reality

Today public or semi-public spaces are becoming less and less a place for non-digital communication, collaboration or social exchange. This change might be mainly caused by novel handheld devices which allow us to be physically present but not mentally. This behavior is not only unsocial, it can also lead to potential dangerous situations (Haga 2015). Instead of understanding communication devices as tools to communicate over long distances, we are interested in developing devices supporting onsite face-to-face communication in public or semi-public spaces. In this publication, we propose a digital adaptation of the famous children game *blind man's buff* as an installation to foster communication and collaboration between people in physical space.

1. Interaction in Public Space, Behavior Changes and Cooperative Games

Interactive screens and public displays have become commonplace in public or semi-public spaces (Müller et al. 2010). The goal of such systems is very often to provide information (interactive maps, location) or advertisement (Wölfel 2014). The focus of such developments is on human-to-computer communication. Thus, today's use of interactive public displays is reducing personal social interaction.

Another important goal of digital devices is to encourage changes in human behavior. For instance, the project *Piano Stairs* transformed the stairs in a metro station in Stockholm into a piano. By playful augmenting the physical environment the use of stairs has been promoted. The experiment was quite successful and helped that the stairs were used more often in comparison to moving stairway than before augmentation (Arroyo et al. 2012). While the feedback of the stairs was prominent and took the focus, alternate methods have been investigated where the attention stays on the main task; e.g. by mirroring the seating posture on anthropomorphic (Wölfel 2017).

In cooperative games players are enforced to cooperative behavior. While such type of games is commonplace in online/digital games where players communicate exclusively over technical devices, only a couple of applications exist where players have to communicate (verbally) to solve a task next to each other. An excellent example for a two-person cooperation is the game *Keep Talking and Nobody Explodes*. One player is trapped with a ticking time bomb. The other player gives instructions how to defuse the bomb without seeing the bomb itself.

2. System Design

In this section, we describe the system design of our installation. The goal was to develop an interactive installation which fosters face-to-face communication between people. Games can be seen as a neutral field for actions without consequences in real life. This liberates our communication concept from any serious situation and reduces constraints that people may have.

We adapted the main idea of the well-known game *blind man's buff* (Encyclopædia Britannica 2010). Similar to the original idea we have used “information divide” as one of the key elements of the gameplay. While some players are moving within a declared area, others could see the position of these players (tracked by a *Kinect* sensor) as cows on the screen including different obstacles. Those obstacles are not visible for the players on the physical game field and thus they have to rely on the provided information from the second group. The German version of blind man's buff is called *Blinde Kuh* which can be translated into English as *blind cow*. We used this metaphor for our visualization, see Figure 1.



Figure 1: This storyboard (rows from left to right) describes the user experience in the proposed adaptation of the game Blinde Kuh.

The different steps of the gameplay are sketched in Figure 1: Before entering the game field passer-by reads an information sign “enter the lawn and listen to voices.” The passer-by enters the gaming area and does not know what to do because he is missing important information. Another person next to the game field is addressing the passer-by and provides instructions what to do. The information screen is only visible for the person next to the game field. Only by cooperation between people on and next to the game field the game can be mastered.

3. Evaluation

The user test took place at *ZKM | Center for Art and Media*¹ at *KAMUNA*² in August 5, 2017. *KAMUNA* is an annual event where all museums in Karlsruhe are open to the public one night. At this event, the general public was attracted and kids up to older people played the game. Impressions of the game are given in Figure 2. The results of the user evaluation (not all participants have answered the questionnaire) are given in Table 1.



Figure 2: The final setup of the test at ZKM Karlsruhe.

¹ <http://www.kamuna.de/> [16.6.2017]

² <http://zkm.de/> [16.6.2017]

	#	communicated with known people ⁺	communicated with unknown people ⁺	easiness of communication [*]	Communication element [*]	joyfulness of physical element [*]
median	30	2	1	3	4	4
average	30	2.53	1.83	3.42	4.03	4.37
variance	30	3.98	5.59	2.09	1.13	0.45
male [°]	16	3.13	2.19	3.40	3.75	4.19
female [°]	14	1.86	1.43	3.45	4.35	4.57
board gamer [°]	23	2.65	2.00	3.22	4.13	4.30
computer gamer [°]	13	3.08	2.62	3.38	3.92	4.23
non-gamer [°]	4	2.25	0.50	1.75	3.25	4.75

Table 1: The result of the evaluation at the ZKM | Center for Art and Media.

⁺ free number ^{*} Likert scale from 1 (lowest) to 5 (highest) [°] median

The easiness of starting a communication in real life with foreigners has been answered in average with 3.93 (median 4) and in the virtual world with 2.72 (median 3) which compares to 3.42 on average though the game.

4. Conclusion & Outlook

We have introduced and investigated a novel way to foster human-to-human communication and collaboration. We found out that starting verbal communication between persons not knowing each other has been fostered. We assume that solving (or helping to solve) the game has a bigger priority for participants than the barrier to start a conversation. This can be confirmed because in average visitors communicated with 1.83 foreigners while playing the game.

References

- Arroyo, E., Bonanni, L., & Valkanova, N. (2012). Embedded interaction in a water fountain for motivating behavior change in public space. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 685-688). ACM.
- Chou, K. T., Hou, H., Shen, M., Chen, T. Y., Liu, C., & Wu, P. J. (2016, May). Human and Dog: Explore the Game Design of Unequal Communicative Patterns. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 156-159). ACM.
- Encyclopædia Britannica. 15th edition. 2. Chicago: Encyclopædia Britannica, Inc. 2010. p. 283.
- S. Haga, S., Sano, A., Sekine, Y., Sato, H., Yamaguchi, S. and Masuda, K. (2015) "Effects of using a smart phone on pedestrians' attention and walking," *Procedia Manufacturing*, vol. 3, pp. 2574-2580
- Heeter, C. (1994). Gender Differences and VR: A Non-User Survey of What Women Want. *Virtual Reality World*, Issue March 1994. Available at <http://commtechlab.msu.edu/randd/research/gender.html> [26.06.2017].
- Hern, A. (2014). Oculus Rift: 'virtual reality is still niche – it's hard to service millions'. *The Guardian*, September 23 2014. Available at: <https://www.theguardian.com/technology/2014/sep/23/oculus-rift-virtual-reality-niche-max-cohen-morpheus> [27.06.2017]
- Mizer, N. J. (2013). *Nomadic tactics in post Revenge geek culture*. Proceedings of the Pop Culture Association/American Culture Association (PCA/ACA). Washington, DC: Wiley.
- Müller, J., Alt, F., Michelis, D., & Schmidt, A. (2010, October). Requirements and design space for interactive public displays. In *Proceedings of the 18th ACM international conference on Multimedia* (pp. 1285-1294). ACM.
- Wölfel, M. (2014). Interacting with Ads in Hybrid Urban Space. In *Cyberworlds (CW), 2014 International Conference on* (pp. 190-197). IEEE.
- Wölfel, M. (2017) Acceptance of Dynamic Feedback to Poor Sitting Habits by Anthropomorphic Objects. In *Proceedings of PervasiveHealth*.

Chapter 7

Interdisciplinary

Experimental Raman spectra of hydrocarbons up to flame relevant temperatures

Kevin Marc Dieter, Dirk Geyer

Department of Thermodynamics and Alternative Propulsion Systems (TAPS)

h_da – University of Applied Sciences Darmstadt, Germany

e-mail: kevin.dieter@stud.h-da.de

Keywords: combustion, Raman spectroscopy, gas-concentrations, hydrocarbons

Introduction

The emission of greenhouse gases, such as carbon dioxide, methane and nitrogen dioxide, cause an increase in the temperature in the earth's atmosphere (Pachauri and Mayer, 2015). By far the largest source of emissions is the burning of fossil fuels and the associated emission of carbon dioxide and other gaseous pollutants (International Energy Agency, 2015). In order to counteract the emission of greenhouse gases and to provide an alternative to fossil fuels, nowadays biofuels are added to the fossil fuels in amounts ranging from 5 to 85%. At present, ethanol is of the greatest importance as bio-generous fuel and is therefore also the most widespread (Bergthorson and Thomson, 2015). The turbulent combustion of renewable fuels is therefore one of the most important fields of research in combustion technology today (Law, 1994).

By knowing the processes prevailing in the turbulent flow field and the chemical-turbulence interaction arising during combustion, the efficiency and the reduction of carbon dioxide emissions as well as other pollutants can be optimized (Dreizler, 2006). Mathematical models, in the form of numerical simulation of turbulent flames, become increasingly important. However, this requires the validation of the models made under simplified assumptions based on characteristic target variables (Seffrin et al., 2010).

The objective of the research presented is to develop a novel measuring system for the analysis of the chemical composition of gas mixtures in high-dynamic systems using a laser-diagnostic method. Spontaneous Raman/Rayleigh scattering, which is based on the interaction of light with matter, is to be used as measuring principle (Raman and Krishnan, 1928). The Raman effect allows the simultaneous diagnosis of a wide variety of gas species by means of the molecular-specific Raman scattered radiation, which can be identified by the wavelength of the measurement signal (Smith and Dent, 2008). Because of a very weak Raman signal, the exciting laser energy must be as high as possible, but without causing any optical breakdowns in the probe volume (Hanson et al., 2016).

The Raman spectra of the intermediate products and their concentrations have already been detected in the past, but they are strongly temperature-dependent (see figure 1). In order to be able to quantitatively determine these species with spontaneous Raman scattering in ethanol flames, the temperature depended spectral response of each individual hydrocarbon species has to be quantified. Unfortunately, the spectra, unlike those of the di- and tri-atomic molecules, cannot be simulated with quantum mechanics because of their complexity (Magnotti et al., 2015). The species,

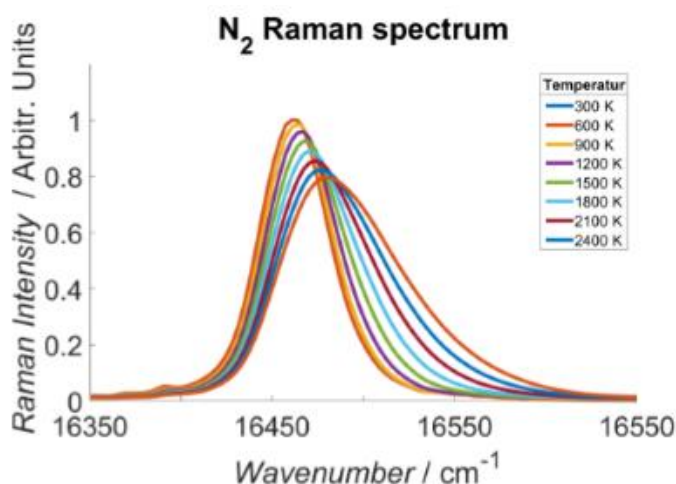


Figure 1: Temperature dependence of a simulated nitrogen Raman signal normalized to temperature 300 K

occurring in significant mole fractions (up to 0,55%) during ethanol combustion, in addition to di- and triatomic molecules are CH₄, C₂H₄, C₂H₆, CH₂O, CH₃CHO and C₂H₅OH. Therefore, the experimental determination of every individual intermediate hydrocarbon Raman spectra is necessary over a flame relevant temperature range in order to be able to measure species concentrations in ethanol flames at all.

Main Part

In order to measure and record the respective Raman signals, the intermediates are required in the gas phase at various temperatures. If one were to attempt to heat a gaseous hydrocarbon species via an electrical heating element, the highest temperatures would occur at a small distance from the heating elements surface in order to achieve the required heat transfer to the gas. At a certain temperature, this results in thermal decomposition, which adversely affects the measurement result (Magnotti et al., 2015).

A novel gas heater, shown in figure 2, solves this problem by using helium as an inert carrier gas. The helium is first heated to temperatures of up to 1380 K and then mixed with a mixture of 50 mole% of the preheated hydrocarbon and 50 mole% of nitrogen. Helium has no Raman cross section and is therefore not visible in the detected signals. Furthermore, helium has a high thermal conductivity, which allows the hydrocarbons to be heated up to high temperatures by homogeneous mixing, enhanced through various static mixers. The Nitrogen serves as reference and carrier gas for the vaporized, gaseous alcohols with a well-known Raman response. The heated mixture then flows from the gas heater, controlled by mass controllers, and hits the probe volume in the middle of the multipass cell (compare figure 3). In order to reach the maximum possible temperatures, the gas heater is coated with micro-porous insulation material.

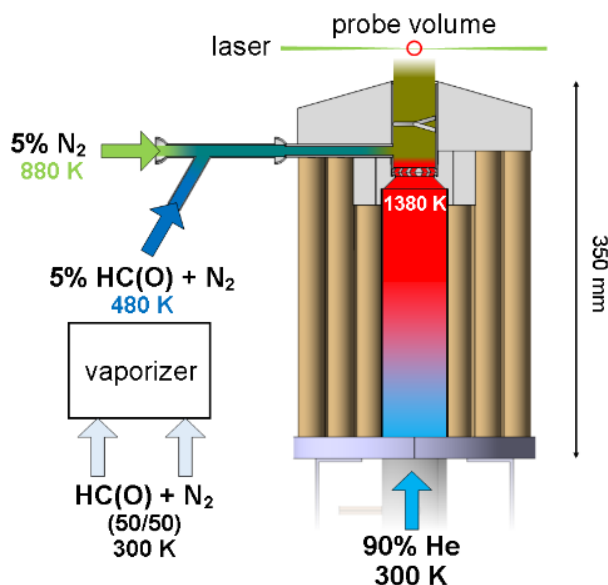


Figure 2: Schematic Helium Gas Heater setup with colored gas flows. blue to red: helium, green: oxygen, dark blue: oxygen and hydrocarbons

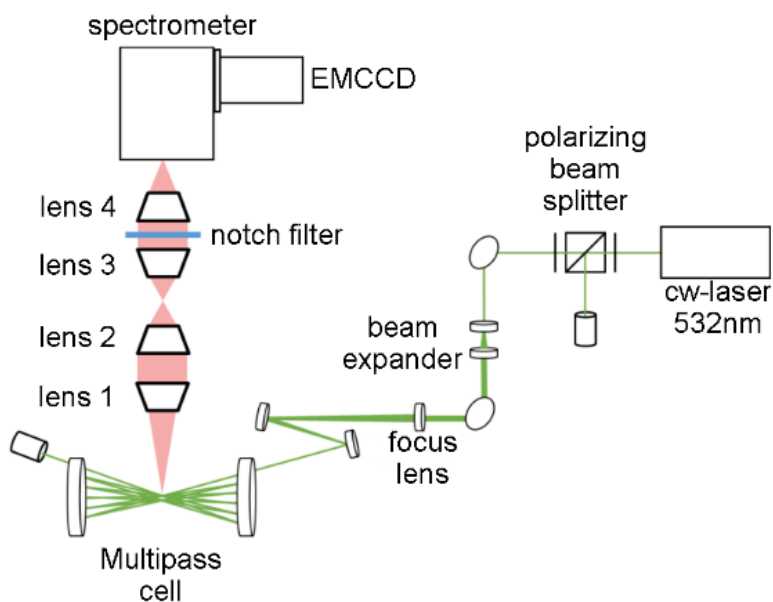


Figure 1: Schematic optical and detection setup of the Multipass Raman Spectrometer (birds eye view)

A spectrally high-resolution Raman spectrometer was developed especially for measurements in stationary gas flows, shown in figure 3. The Raman scattering is excited by a CW-fiber-laser. Its beam diameter is enlarged with beamforming optics, focused and coupled into a multipass cell. Approximately 18 passages of the laser beam take place in the cell in order to enhance the laser energy in the focus. Using a CW-laser in combination with a multipass cell avoids optical breakdowns but enhances the detectable Raman signal. The gas heater is placed underneath the multipass cell. For detection, the scattered radiation is determined by an Apo-chromatically corrected lens group. It is then focused by a commercially available objective-lens onto the input of a Rayleigh filter cell in which a holographic notch filter suppresses the excitation wavelength of the laser. In a transmission spectrometer the sampled signal is spectrally resolved over a range of 26 nm and recorded by an EMCCD camera.

Conclusions and Outlook

In this work, the temperature-dependent Raman spectra of selected hydrocarbon species, relevant for ethanol combustion are experimentally characterized up to temperatures of 1000 K (preliminary result for methane shown in figure 4).

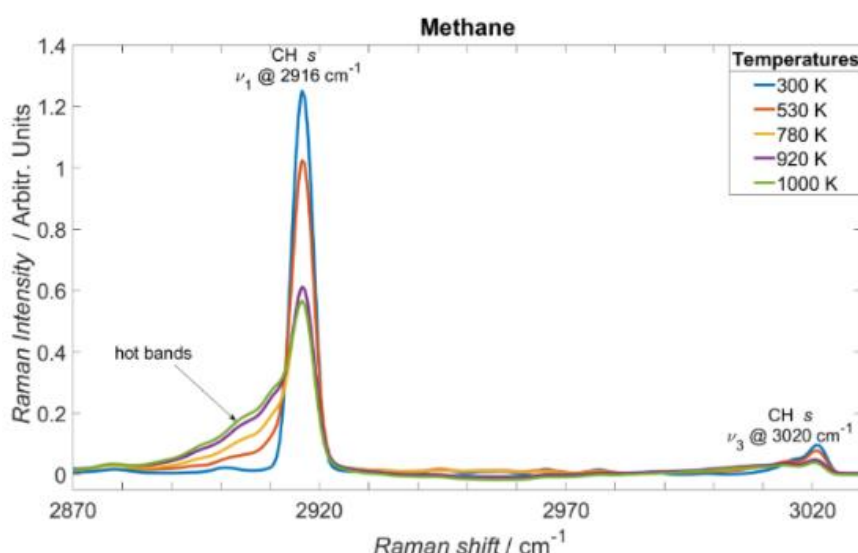


Figure 2: Preliminary Raman spectra from methane for various temperature setpoints. You can see symmetric CH stretchmodes at 2916 cm^{-1} and 3020 cm^{-1}

For this purpose, a new type of gas heater has been developed, which enables the species to be heated up to very high temperatures, preventing thermal decomposition. The heated gases are then detected by a high resolution multipass Raman spectrometer. The gained information is mandatory for the evaluation of Raman/Rayleigh measurement data and will be processed in a spectral library in a way that they can be made available to other research groups as a basis for the measurement data analysis. A future goal is the first comprehensive study of the Raman spectra of formaldehyde. The determination of the Raman spectra of the formaldehyde over a wide temperature range is an essential requirement for the quantitative measurement of formaldehyde concentrations in flames by Raman scattering. For this purpose, an apparatus needs to be developed which continuously feeds formaldehyde to the novel gas heater in a gaseous state without polymerizing it.

References

- Bergthorson, J. M. and Thomson, M. J. (2015). A review of the combustion and emissions properties of advanced transportation biofuels and their impact on existing and future engines, *Renewable and Sustainable Energy Reviews*, vol. 42, pp. 1393–1417.
- Dreizler, A. (2006). Verbrennung im Laserlicht: Methoden zur Validierung numerischer Simulationen [Online], Darmstadt, Wiss. Buchges. Available at http://deposit.ddb.de/cgi-bin/dokserv?id=2718069&prov=M&dok_var1&dok_ext=htm.
- Hanson, R. K., Spearrin, R. M., Goldenstein, C. S. (2016). *Spectroscopy and Optical Diagnostics for Gases*, Heidelberg, Springer International Publishing
- International Energy Agency (2015). CO₂ Emissions from fuel Combustion: Highlights [Online], Paris. Available at <https://www.iea.org/publications/freepublications/publication/CO2EmissionsFromFuelCombustionHighlights2015.pdf>.
- Law, C. K. (1994). Combustion Science and Technology: Accomplishments and Challenges, *Combustion Science and Technology*, vol. 98, 4-6, pp. 295–300.
- Magnotti, G., KC, U., Varghese, P. L. and Barlow, R. S. (2015). Raman spectra of methane, ethylene, ethane, dimethyl ether, formaldehyde and propane for combustion applications, *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 163, pp. 80–101.
- Pachauri, R. K. and Mayer, L., eds. (2015). *Climate change 2014: Synthesis report*, Geneva, Switzerland, Intergovernmental Panel on Climate Change.
- Raman, C. V. and Krishnan, K. S. (1928). A New Type of Secondary Radiation, *Nature*, vol. 121, no. 3048, pp. 501–502.
- Seffrin, F., Fuest, F., Geyer, D. and Dreizler, A. (2010). Flow field studies of a new series of turbulent premixed stratified flames, *Combustion and Flame*, vol. 157, no. 2, pp. 384–396.
- Smith, E. and Dent, G. (2008). *Modern raman spectroscopy: A practical approach*, Chichester, Wiley.

Bioactives from *Saccharina latissima* and Gut Health

Orlaith O'Connell, Helen O'Shea, Hugh McGlynn
Department of Biological Sciences
Cork Institute of Technology, Rossa Avenue, Bishopstown, Ireland
E-mail: orlaith.oconnell@mycit.ie

Introduction

Seaweeds are classified into Rhodophyta (red), Chlorophyta (green) and Phaeophyta (brown) marine macroalgae and have typically been used by the food and cosmetic industry as sources of thickeners, gelling agents and stabilisers such as alginate, agar and carrageenan. However in recent years, these marine plants have been identified as rich sources of bioactive compounds. Brown macroalgae in particular contains a series of bioactive compounds not found in red and green macroalgae such as fucoidan, phlorotannins and fucoxanthin. These compounds have been found to possess potent anti-oxidant, anti-inflammatory and anti-cancer properties which could contribute to the maintenance of gastrointestinal health (Brown *et al*, 2014), (Wijesinghe *et al*, 2012). The prevalence of gastrointestinal disorders such as inflammatory bowel disease (IBD), Coeliac disease and alcoholic liver disease has increased across Europe as well as incidence of colorectal and pancreatic cancers (Farthing *et al*, 2014). As such effective measures to promote gastrointestinal health are necessary. The purpose of this research project is to investigate the immunomodulatory properties of extracts from brown macroalgae species', particularly that of *Saccharina latissima*, and their potential role in gastrointestinal health. The anti-bacterial properties of these extracts will also be investigated using selected bacterial species relevant to the food industry. This project is a component of a large ERAnet funded project i.e. 'SeaRefinery', which aims to develop processing technologies to extract high value-added bioactive components from cultivated seaweed species in an integrated biorefinery.

Research Approach

1. Seaweed extracts

Extracts from the brown macroalgae species *Saccharina latissima* and *Ascophyllum nodosum* were provided by SeaRefinery project partner CyberColloids. *Saccharina latissima* extracts were isolated by means of hot water extraction. *Ascophyllum nodosum* extracts from different locations and different seasons were isolated by means of methanol, ethanol or water extraction. Dried extracts were dissolved in Phosphate Buffered Saline solution (PBS)

2. In-vitro model of gastrointestinal inflammation

In order to assess the anti-inflammatory potential of extracts, the production of pro-inflammatory cytokines Interleukin 8 (IL-8) and Tumour Necrosis Factor- α (TNF- α) will be measured using ELISA assay kits & RT-PCR. Both IL-8 and TNF- α play causative roles in inflammation, making inhibition of these pro-inflammatory cytokines a key therapeutic target for many anti-inflammatory compounds (Harada *et al*, 1994), (Popa *et al*, 2007). The assessment of the anti-inflammatory properties of brown macroalgae extracts is currently underway. If any anti-inflammatory properties are observed in these extracts, the influence of varying concentrations of the extracts on inflammation will be assessed using a co-culture system. This co-culture system is comprised of human intestinal epithelial cells and murine macrophages, as described by Tanou *et al.*, (2008), with minor adjustments. As seen in Figure 1, the human intestinal epithelial cell line CaCo-2 cells and the murine macrophage cell line J774.2 cells are seeded in a co-culture system. The apical side of the co-culture system is treated

with extracts and the basolateral side is stimulated with 1 µg/ml LPS in order to initiate the process of inflammation. Supernatants are collected for measurement of IL-8 and TNF- α production. This *in-vitro* model was chosen in order to simulate the process of inflammation at the cellular level.

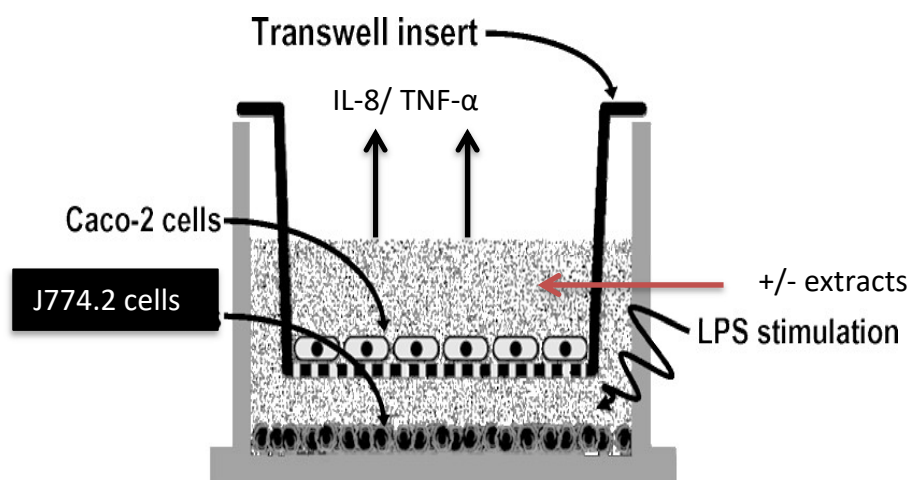


Figure 1: In-vitro model of gastrointestinal inflammation.

Modified from Tanou et al., (2008)

3. Anti-bacterial activities

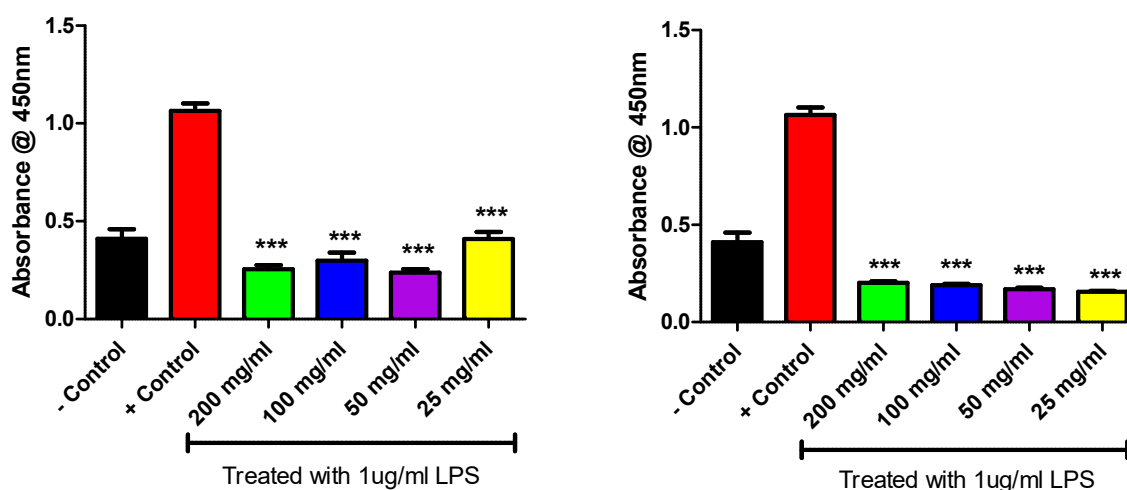
Six species of bacteria were selected for this project based on their importance to the food industry. These bacteria include *Staphylococcus aureus*, *Escherichia coli*, *Enterococcus faecalis*, *Salmonella enterica*, *Pseudomonas aeruginosa* and *Streptococcus agalacticae*. The influence of varying concentrations of extracts on the growth of the selected bacterial isolates was assessed using 96 well plates. Anti-bacterial assay was conducted as described by Gupta et al., (2010) with modifications. Extract obtained from *Saccharina latissima* (1g) was dissolved in Tryptic Soy Broth (1ml). All incubations took place at 37°C. Should any inhibitory effect be observed, further testing will be carried out to determine if the extracts are bacteriostatic or bactericidal.

Results & Discussion

Anti-inflammatory effects of seaweed extracts

Seaweed and seaweed extracts, particularly those from brown seaweed, have potent anti-inflammatory and immunomodulatory properties which have largely been tested within *in vitro* or animal models. For example, several *in vitro* studies have determined the inhibitory effect of the marine carotenoid fucoxanthin on inflammatory mediators and pro-inflammatory cytokines using lipopolysaccharide stimulated RAW 264.7 macrophages (Shiratori *et al*, 2005), (Heo *et al* 2010). However, insufficient research has been carried out on the anti-inflammatory effects of seaweed extracts using intestinal epithelial cell lines. Therefore in this study CaCo-2 cells were treated with seaweed extracts in order to determine their influence on the production of pro-inflammatory cytokine by intestinal epithelial cells during inflammation. The main focus for this study was the production of IL-8. IL-8 is a key target in many anti-inflammatory treatments due to its role in the recruitment of neutrophils to sites of acute inflammation (Harada *et al*, 1994). As seen in Figures 2 & 3 IL-8 production was significantly

inhibited in both *Saccharina* and *Ascophyllum* extracts at concentrations of 25mg/ml – 200mg/ml. When compared to the positive control (i.e. LPS treated cells), IL-8 production of cells treated with *Saccharina* and *Ascophyllum* extracts were inhibited by 70% and 83% respectively. Similar suppressive effects of brown seaweed extracts on IL-8 expression have been observed during inflammatory response of an *ex vivo* porcine colon model (Bahar *et al*, 2013).

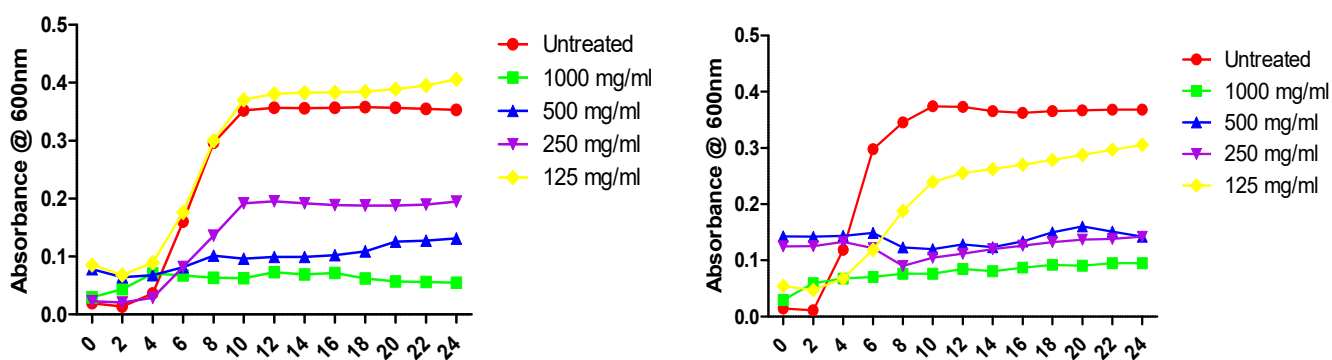


Figures 2 & 3: IL-8 production of human intestinal epithelial cells after 24hr incubation with seaweed extracts and lipopolysaccharide. CaCo-2 cells were treated with indicated concentrations (25mg/ml – 200mg/ml) of *Saccharina* or *Ascophyllum* extracts plus 1µg/ml LPS. After 24hrs IL-8 production was measured by IL-8 ELISA. Values are expressed as mean +SEM of triplicate experiments. ***p < 0.001 indicates a significant difference from the LPS-stimulated control group

Anti-bacterial effects of Saccharina latissima extracts

Microbial contamination is a major obstacle to food safety and quality. Conventional preservation methods, including the addition of chemical preservatives and pasteurization has allowed food producers to extend the shelf-life of their products far beyond what would have been achieved naturally. However, with consumer preferences moving towards natural, minimally-processed and microbiologically safe products, food producers have started to seek out preservatives from natural sources. Bioactive extracts from several seaweed extract have been shown to possess potent anti-microbial properties which could be of benefit to the food industry. Brown macroalgae extracts in particular demonstrates anti-microbial activity against a number of Gram negative and Gram positive bacteria, including *Escherichia coli* and *Staphylococcus aureus* (Sandalsen *et al*, 2003). In order to determine the anti-bacterial activity of *Saccharina* extracts over time, growth of selected bacterial isolates in the presence of the extracts was sampled over 24hrs. As seen in Figures 4 & 5 the growth of *Staphylococcus aureus* and *Enterococcus faecalis* in the presence of *Saccharina* extracts resulted in varying levels of inhibition at concentrations of 125mg/ml – 1000mg/ml. Cox (2010) observed similar anti-bacterial properties of *Saccharina* extracts against *Enterococcus faecalis* which was linked to the extracts total phenolic content. Growth inhibitory results at these concentrations were also observed for *Salmonella enterica*, *Streptococcus agalacticae*, *Escherichia coli* and *Pseudomonas aeruginosa* (data not shown). Future work will include investigating whether

growth inhibitory activities of extracts are bactericidal or bacteriostatic, in order to determine the suitability of the extracts as natural antibacterial compounds in the food industry.



Figures 4&5: Growth kinetics of *Staphylococcus aureus* and *Enterococcus faecalis* in the presence of different concentrations of *Saccharina latissima* extracts.

Conclusions

Saccharina latissima can be considered a promising source for effective bioactives. The extracts from this marine plant demonstrated evident anti-inflammatory properties in response to LPS-induced inflammation in intestinal epithelial cells. Further investigation of these anti-inflammatory properties will be conducted in an *in vitro* model which better mirrors to process of inflammation in the body. *Saccharina* extracts also display effective growth inhibitory activities against various food spoilage and food pathogenic bacteria of importance to the food industry. Further studies will determine if inhibitory activities are bactericidal or bacteriostatic.

Acknowledgements

This work was funded by MarineBiotech, Innovationsfonden, Innovasjon Norge, Agentschap Innoveren en Ondernemen, Technology Development Fund, Marine Institute and ERA-MBT Joint Call Secretariat. Thanks to all partners involved in the project: Danish Technological Institute, Centexbel, Marinox ehf, CyberColloids Ltd, Sioen Industries NV, Hortimare AS, and ViVoX ApS.

References

- Bahar, B., O'Doherty, J., Hayes, M. and Sweeney, T. (2012). Extracts of brown seaweeds can attenuate the bacterial lipopolysaccharide-induced pro-inflammatory response in the porcine colon ex vivo. *Journal of Animal Science*, 90(Supplement 4), pp.46-48.
- Brown, E., Allsopp, P., Magee, P., Gill, C., Nitecki, S., Strain, C. and McSorley, E. (2014). Seaweed and human health. *Nutrition Reviews*, 72(3), pp.205-216.
- Cox, S., Abu-Ghannam, N. and Gupta, S. An Assessment of the Antioxidant and Antimicrobial Activity of Six Species of Edible Irish Seaweeds. *International Food Research Journal* 17: 205-220 (2010)
- Farthing, M., Roberts, S., Samuel, D., Williams, J., Thorne, K., Morrison-Rees, S., John, A., Akbari, A. and Williams, J. (2014). Survey of digestive health across Europe: Final report. Part 1: The burden of gastrointestinal diseases and the organisation and delivery of gastroenterology services across Europe. *United European Gastroenterology Journal*, 2(6), pp.539-543.

Gupta, S., Cox, N., Rajauria, G., Jaiswal, A.K. & Abu-Ghannam, N. (2010). Modelling the growth inhibition of common food spoilage and pathogenic micro-organisms in the presence of brown seaweed extracts. Food and Bioprocess Technology,

Harada, A., Sekido, N., Akahoshi, T., Wada, T., Mukaida, N. and Matsushima, K. (1994). Establishment of essential involvement of IL-8 in acute inflammation. Journal of leukocyte biology, 56(5), pp.559–564.

Heo, S., Yoon, W., Kim, K., Ahn, G., Kang, S., Kang, D., affan, A., Oh, C., Jung, W. and Jeon, Y. (2010). Evaluation of anti-inflammatory effect of fucoxanthin isolated from brown algae in lipopolysaccharide-stimulated RAW 264.7 macrophages. Food and Chemical Toxicology, 48(8-9), pp.2045-2051.

Popa, C., Netea, M., van Riel, P., van der Meer, J. and Stalenhoef, A. (2007). The role of TNF- α in chronic inflammatory conditions, intermediary metabolism, and cardiovascular risk. Journal of Lipid Research, 48(4), pp.751-762.

Sandsdalen, E., Haug, T., Stensvåg, K. and Styrvold, O. (2003). The antibacterial effect of a polyhydroxylated fucophlorethol from the marine brown alga, *Fucus vesiculosus*. World Journal of Microbiology and Biotechnology, 19(8), pp.777-782.

Shiratori, K., Ohgami, K., Ilieva, I., Jin, X., Koyama, Y., Miyashita, K., Yoshida, K., Kase, S. and Ohno, S. (2005). Effects of fucoxanthin on lipopolysaccharide-induced inflammation in vitro and in vivo. Experimental Eye Research, 81(4), pp.422-428.

Tanoue, T., Nishitani, Y., Kanazawa, K., Hashimoto, T. and Mizuno, M. (2008). In vitro model to estimate gut inflammation using co-cultured Caco-2 and RAW264.7 cells. Biochemical and Biophysical Research Communications, 374(3), pp.565-569.

Wijesinghe, W. and Jeon, Y. (2012). Biological activities and potential industrial applications of fucose rich sulfated polysaccharides and fucoidans isolated from brown seaweeds: A review. Carbohydrate Polymers, 88(1), pp.13-20.

Protection of chemical reactors with partial-least-square-modeling (PLS) for safety devices

J. Biernath, F. Goerlich, J. Denecke, J. Schmidt,
CSE Institute – Center of Safety Excellence
e-mail: johannes.biernath@cse-institut.de

Keywords: Process control system, Partial-least-square-modeling, Computer Science

Introduction:

A huge proportion of chemical processes can be categorized as exothermic chemical reactions. A possible hazard of these chemical reactions is a pressure rise due to malfunction of technical apparatuses like heat exchangers. In this case a temperature rise occurs a pressure rise, comparable with a cooking pot. This hazard is called *runaway reaction*. If the pressure rise is too high, plastic deformation occurs, the vessel deforms irreversible and can burst. Therefore, an overpressure protection of the vessel, like *process-control-system-safety-devices* (PCS), is essential (Schmidt and Giesbrecht, 1997).

The task of a PCS-system is to control process variables like temperature and initiate countermeasures to prevent an inadmissible build up in pressure. To determine when and which countermeasure must be initiated, a *programmable logic controller* (PLC) is necessary. The PLC constitutes the connection between causes (action) and countermeasures (reaction). For coding the PLC, it requires a full understanding of the chemical reactions and the mutual influence between any process variable, like temperature, pressure, concentration or enthalpy.

To describe a process, balance, constitutive equations and constraints must be defined. Due to the complexity, even of a simple chemical batch process, a large number of equations is needed. For example, the simulation of a polymerization process needs about 300 equations (Fischer, 2001). An implementation in a PLC is often not possible, because the solution of the system of equations is too time-consuming. For a safety application it's important, that every process variable can be determined quickly and reliably, so that also rapid reactions can be safeguarded. An appropriate solution to reduce the number of equations is a *partial-least-square-modeling* (PLS), because compared to other multiple regressions, PLS manifests itself through a good accuracy and stability (Cramer, 1993).

This PLS-modeling enables a reduction of process variables due to linear regressions as demonstrated for a non-safety-related system (Adebiyi and Corripio, 2003). The next, logical step is to develop a PLS-system for safety applications. This development is a part of the PhD thesis *SmartHIP* at the CSE (Center of Safety Excellence) Institute, which focuses on the protection of chemical reactors with PCS-Systems. By omitting mechanical safety devices emissions will be prevented and the productivity of the process will be increased.

Theory:

PLS is a statistical technique, which reduces the variables of a system to a smaller set within a good accuracy due to linear regression. Fig. 1 demonstrates the functional principle of a PLS-Modeling. PLS-Modeling distinguishes between two datasets, *X-Data* and *Y-Data*. *X-Data* includes every process variables of the reactor, which can be measured for a certain number of batches. The result is a cloud of data points, which are dependent of different process variables.

In Fig. 1 the cloud, symbolized by green points, is dependent of three process variables. In reality, there are much more variables and the number of dimensions is much higher. The next step is a reduction of dimensions. Therefore, a new multilinear coordinate system is created, which is called *latent variable*. These *latent variables* must be adopted to the data, so that the covariance is maximum and there is an acceptable agreement with the original data. The same procedure must be repeated with *Y-Data*. In a safety application *Y-Data* would include the safety parameters to protect the reactor. The last step is to link the new coordinate systems together. The method has been developed for process optimization already (Adebiyi and Corripio, 2003), but it has never been applied and certified as a safety technology.

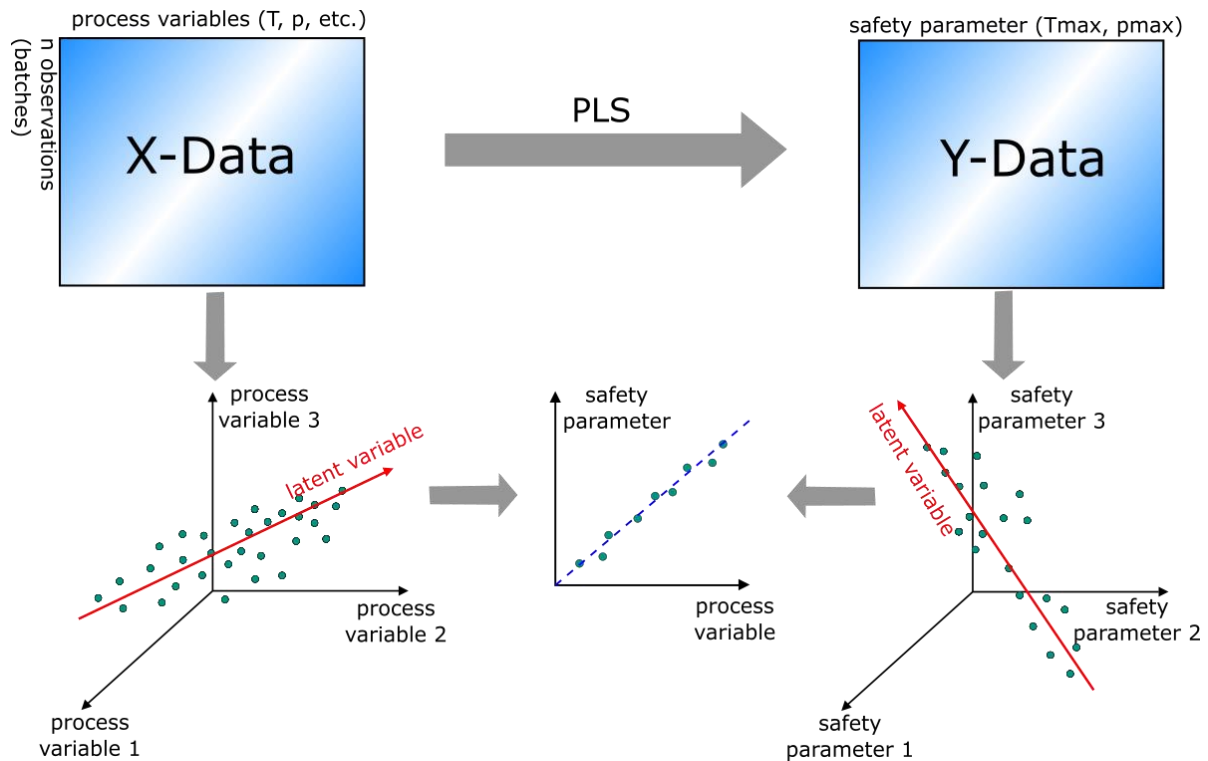


Figure 1: Functional principle of PLS-Modeling

At this point, the question arises why PLS isn't used as a safety technology. The main reason is, that in many cases a reliable model of the process or experimental data in all operating areas doesn't exist due to missing physical properties or reaction kinetics. There is a possible hazard, that missing or wrong PLS parameters may lead to a failure of the safety device. An evaluation of the quality of inter- or extrapolation capabilities of the PLS model is missing. Therefore, it must be shown, that these problems can be solved, so that a certification as a safety technology is possible.

Research approach, methodology and current status:

The hypothesis is, that the complexity of a model for a *process-control-system* as a safety technology can be reduced by *partial-least-square-modeling*. For its certification, following points must be fulfilled (ISO 61508):

- High availability of production using the safety system, so that a single deviant measurement doesn't lead to a shutdown.
- Failsafe-states must exist, if the measured parameters lead to invalid results.

- Transparent verification of the PLS model parameters for each conceivable/ numerically possible condition.
- Reliable traceability of the numerical accuracy
- Adaptability of complexity for different processes
- Certification of the development process and resulting PLS model

In order to use a PCS with complex models the chosen research approach is to develop a PLS within a PCS, which fulfills the above-named prerequisites. Therefore, a deductive, quantitative research methodology has been selected, which draws a conclusion to the logical consequence by recognizing logical rules and given conditions (Bryman et al., 2011; Greener, 2008). Transferring this to the research approach it means, that instead of using rigorous models based on energy and material balances, multivariate data analysis can help to build a simple model. This analysis is realized by the *partial-least-square-modeling*.

Furthermore, it is essential to study the influence of uncertain measured values on the safety distance of the model to a required shut-down. This analysis is the basis to define different failsafe-states. For this purpose, a *sensitivity analysis* must be performed. The aim of a sensitivity analysis is to examine the behavior of individual parameters among each other, those combined effect and the impact of parameter fluctuations on the target function. A sensitivity analysis involves a series of methods that allow to establish and understand the relationship between the variance of a single parameter and the variance of the model function. A generally classification of different methods is the distinction between local and global sensitivity analysis (Dangendorf et al., 2012).

The local sensitivity analysis describes the behavior of a single parameter with regard to the target function. In this case, a single parameter is varied, while the remaining parameters are fixed. The disadvantage of this method is, that parameter interactions are not taken into consideration. This is especially dangerous, because in a chemical reaction many process variables have an enormous impact on other variables. For example, a variation of the temperature leads to an immediate variation of the pressure, but also has an influence on the reaction velocity due to the Arrhenius approach. Therefore, a further method, the global sensitivity analysis, is necessary

In contrast to the local sensitivity analysis, the remaining parameters can also be kept variable and because of that a consideration of interdependencies and their impact is possible. Hereby, the variability of each parameter is described by a distribution function. A Monte Carlo simulation allows to test any number of different parameter combination. With regard to this project, a generalized sensitivity analysis (GSA) will be used, which is based on a stochastic approach to compare different parameters. Hereby, the prerequisite is, that the variability of each parameter can be represented by an individual distribution function. Thereby, each distribution function is described by the whole parameter space, which can be assumed by each single parameter. By using Monte Carlo simulation random combinations of different parameters can now be drawn.

The main disadvantage of global sensitivity analysis is, that local effects of single parameter are ignored due to the global consideration. Thereby, a critical value, e.g. a local maximum, of a single parameter at a specific time is not recognized, because there is a higher critical value, e.g. a global maximum, over the whole time. But the local critical value can also be important to describe the risk potential of the runaway reaction. Therefore, it is necessary to conduct both, a local and a global sensitivity analysis. A detailed sensitivity analysis is the condition for a reliable *uncertainty analysis*.

Additionally, a *multi-sensor data fusion* offers a possibility to generate a high availability of the measured input data. It is a technology to enable combining information from several sources in order to form a unified picture (Khaleghi et al., 2013). For example, if one sensor fails, by combining data from several and different sensors, conclusions about the missing sensor data can be drawn. The applications of multi-sensor data fusion range from military problems such as automatic target identification, the determination of occurrence of natural resources to industrial applications such as the control of complex reactors or automated production. Data fusion is generally a mixture of mathematical and heuristic techniques in the field of statistics, digital signal processing and decision theory (Klaus, 1999), nonetheless there exist many different definitions. A popular example is Joint Directors of Laboratories (JDL), which is grounded more on input and output data than on processing, because it has a military background. Alternatives are Dasarath's or Kokar's framework, which are more focused on the processing aspect.

To decide, which data fusion algorithm should be chosen, it is important to define the data-related challenges. Therefore, it distinguished between different data-related fusion aspects. They are imperfection, correlation, inconsistency and disparateness, as shown in Fig. 2. Especially data imperfection is relevant for chemical processes, because data provided by sensors are subject to an inaccuracy as well as an uncertainty in the measurements. The task of the data fusion algorithm must be now to express this imperfection and to use the data redundancy to reduce these effects. Therefore, there are provided different frameworks of imperfect data treatments, which constitute a further classification of the imperfection in uncertainty, ambiguity, vagueness, incompleteness and granularity. For the beginning, the most important aspect is the uncertainty of the measurements, so that a probabilistic method will be chosen, because it based on the probability distribution function to express data uncertainty. This method relies on the Bayesian inference, which derives the posterior probability from a prior probability and a likelihood function, which dates back to a statistical model for the observed data.

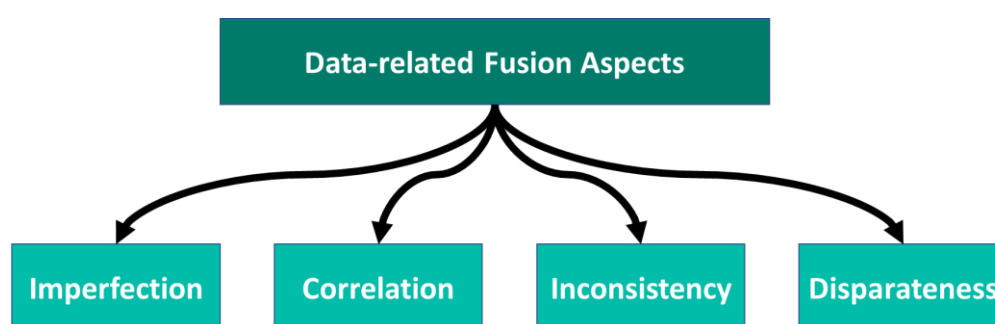


Figure 2: Taxonomy of data fusion methodologies

To validate the quality of inter- or extrapolations a process with well-known physical properties and reaction kinetics in all operating areas can be used, which serves as a basis to develop a general approach for a reliable inter- or extrapolation. Therefore, a part of the known process variables is removed and an algorithm will be developed, which reconstructs these missing variables with an inter- or extrapolation. After that, a verification of transmission this algorithm to other processes will be conducted. If there isn't a possibility to transmit it, further well-known processes can be examined, which a more similar with the unknown process.

The project is still at the beginning. The first step is to create an example data set, the cloud of *X- and Y-Data*. For this purpose, an example polymerization process has been chosen. Rigorous and detailed parametric simulations of different batch conditions will be performed in order to produce a huge cloud of data. In next step, this cloud will be transfer into a PLS-System and it will be implement into a PLC. Afterwards, it will be verified whether the prerequisites are fulfilled. After that, *sensitivity analysis* and *multi-sensor data fusion* will be carried out.

Conclusion:

The protection of reactors with PCS is challenging because of the large number of equations and the involved time-consuming calculations. PLS-Modelling is a promising option to reduce the numerical effort. Similar systems exist already only for process optimization as non-safety-related systems. Therefore, a PLS-Model is developed, which can be used as a safety technology and can be implemented in a PCS-system.

References

- Schmidt, J. and Giesbrecht, H. (1997). „Absicherung von Semi-Batch-Reaktoren gegen unzulässige Drücke“, in: *Chemie Ingenieur Technik*, vol. 68, no. 9, pp. 1144-1145.
- Fischer, N. (2001). *Absicherung von Alkoxylierungsreaktoren und Untersuchung des Stofftransporteinflusses auf den maximalen Druck beim Durchgehen der Reaktion*. Fachhochschule Mannheim.
- Cramer, R. D. (1993). “Partial Leas Squares (PLS): Its strengths and limitations”, in: *Perspectives in Drug Discovery and Design*, vol. 1, no. 2, pp. 269-278.
- Adebiyi, O. A. and Corripio, A.B. (2003). “Dynamic neural networks partial least square (DNNPLS) identification of multivariable processes”, in: *Computers and Chemical Engineering*, vol. 27, no. 2, pp. 143-155.
- Greener, S. (2008). *Business Research Methods*. Ventus Publishing ApS, Frederiksberg.
- Bryman, A., Bell, E., Mills, A.J. and Yue, A.R. (2011). *Business Research Methods*. Oxford University Press, Toronto.
- Dangendorf, S., Burzel, A., Wahl, T., Mudersbach, C., Jensen, J. and Oumeraci, H. (2012). *Unsicherheits- und Sensitivitätsanalyse im Rahmen einer integrierten Risikoanalyse*, Technische Universität Braunschweig.
- Khaleghi, B. Khamis A., Karray F.O. and Razavi S.N. (2013). “Multisensor data fusion: A review of the state-of-the-art”, in: *Information Fusion*, vol. 14, no.1, pp. 28-44.
- Klaus, F. (1999). *Einführung in Techniken und Methoden der Multisensor-Datenfusion*. University of Siegen.

Chapter 8

Traffic and Transportation

A Vehicular Trajectory-Driven Method for Signalized Corridor Control under Connected Vehicles Environment

Slobodan Gutesa, Joyoung Lee, Ph.D, Dejan Besenski, Ph.D, Lazar Spasovic, Ph.D

Department of Civil and Environmental Engineering
New Jersey Institute of Technology, Newark, NJ. U.S.A
e-mail: jo.y.lee@njit.edu

Keywords: Intersection Management; Connected Vehicles; Vehicle to Infrastructure (V2I) Communication; Signal Phase and Timing (SPaT)

In December, 2015, the Fixing America's Surface Transportation (FAST) Act (FHWA 2016) was introduced with the goal of setting the course for transportation investment in the highway. The FAST Act includes provisions to improve the performance of nation's surface transportation in terms of mobility, job creation, and promotion of innovations. On the other hand, private auto manufacturers are moving quickly by offering various self-driving solutions readily available. Such a rapid advancement in the private sector has led the USDOT and other public sectors to develop a Connected Vehicle Reference Implementation Architecture (CVRIA) (Iteris 2016) to support the field implementation of connected vehicle concept.

Improving fuel efficiency and reducing emissions through the provision of signal phase and timing information to individual vehicles gained interest worldwide (Rakha and Kamalanathsharma, 2011; Xia et al., 2013; Barth, et al., 2011; Zhang and Yao, 2015). The main idea behind the concept is to provide the advisory speed through an on-board device in order to improve safety (Jimenez et al. 2008; Paine et al. 2007) environmental and mobility performance of an intersection (Katsaros et al. 2011; Seredynski et al. 2013). In-vehicle Signal Assistance (ISA) Paine et al. (2007) is one of the Connected Vehicles (CV) applications utilizing the vehicle to infrastructure (V2I) communications. Vehicles equipped with the ISA application receive real-time signal phase and timing (SPaT) (RITA 2016) data from the intersection where the vehicles approach. With the SPaT data, the ISA application conveys the current signal status to the driver via a graphical display unit (e.g., an opt-in LCD panel or head-up display, an external smartphone or tablet PC).

As a game-changer, the primary idea presented in this paper proposes a new signal control paradigm where the traffic streams are manipulated to conform to the signal control devices. The ISA concept is utilized to convey the most desirable speed to individual road users, based on the current state of traffic streams, the state of signalization, and the position of the individual vehicles in real time. Signal status parameters such as cycle length and remaining green/red time are continuously captured. At the same time, the in-vehicle unit provides vehicle position information through cell-phone GPS receiver. Both inputs are then used by predictive, trajectory-driven, control algorithm to provide optimal vehicle speed for each ISA equipped vehicle.

The majority of signal timing applications are evaluated through the time-space diagram as the primary measure of the control strategy operation. To that end, a control algorithm was developed to generate optimal, time-distance, vehicle trajectories for all vehicles in the system equipped with ISA on-board device. The methodology assumes that the vehicle trajectory can be defined as a cubic interpolated spline allowing flexible accommodation of the trajectory to the given signal timing obstacles in the time-distance searching space. An example of such trajectory is illustrated in Figure

1, where control points $p_1(X_1, Y_1) \dots p_M(X_M, Y_M)$ were used for the trajectory interpolation. By respecting interpolation and monotonicity rules, the trajectory T will be produced in the field of real numbers giving the sequence of coordinates in the defined coordinate space:

$$T: x^T = (x_1, \dots, x_n): y^T = (y_1, \dots, y_n)$$

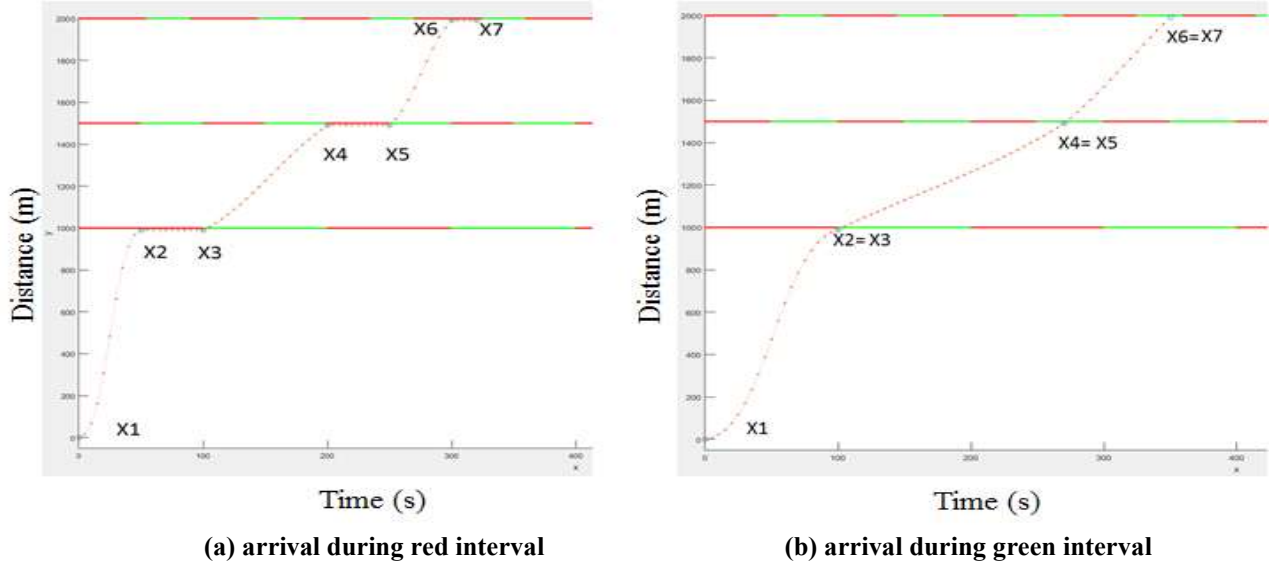


Figure 1: Trajectory defined using cubic interpolation

The main objective of the trajectory optimization is to minimize the sum of all the trajectory curves for vehicles in the control space C . The length of the curve is therefore calculated as the sum of Euclidean distances between successive x_i and y_i elements of the trajectory for N number of vehicles in the control environment.

$$Length = \sum_{j=1}^N \sum_{i=1}^{n-1} \sqrt{\Delta x_i^2 + \Delta y_i^2} = \sum_{j=1}^N \sum_{i=1}^{n-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}$$

In order to maintain cruising condition along the green-bend of the corridor, and comply with the posted speed limit conditions the optimization model is formulated for the corridor illustrated in Figure 1 as follows:

$$\begin{aligned} MIN \quad & \sum_{j=1}^N \sum_{i=1}^n Length(p_{1+(2N+1)j}, p_{2+(2N+1)j}, p_{3+(2N+1)j}, p_{4+(2N+1)j}, p_{5+(2N+1)j}, p_{6+(2N+1)j}, p_{M+(2N+1)j}) \\ & = \sum_{j=1}^N \sum_{i=1}^{n-1} \sqrt{\Delta x_i^2 + \Delta y_i^2} = \sum_{i=1}^n \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \end{aligned}$$

It is assumed that the vehicles are being indexed in the following fashion: $j=1$ vehicle closest to the intersection 1 with current distance to the intersection d_1 , $j=N$ the last vehicle, with the greatest distance from the first intersection d_N .

The defined problem was solved using Genetic Algorithm, and the optimization framework for a given corridor described above is illustrated in Figure 2. Optimal solutions with respect to different corridor configurations are presented in Figure 3.

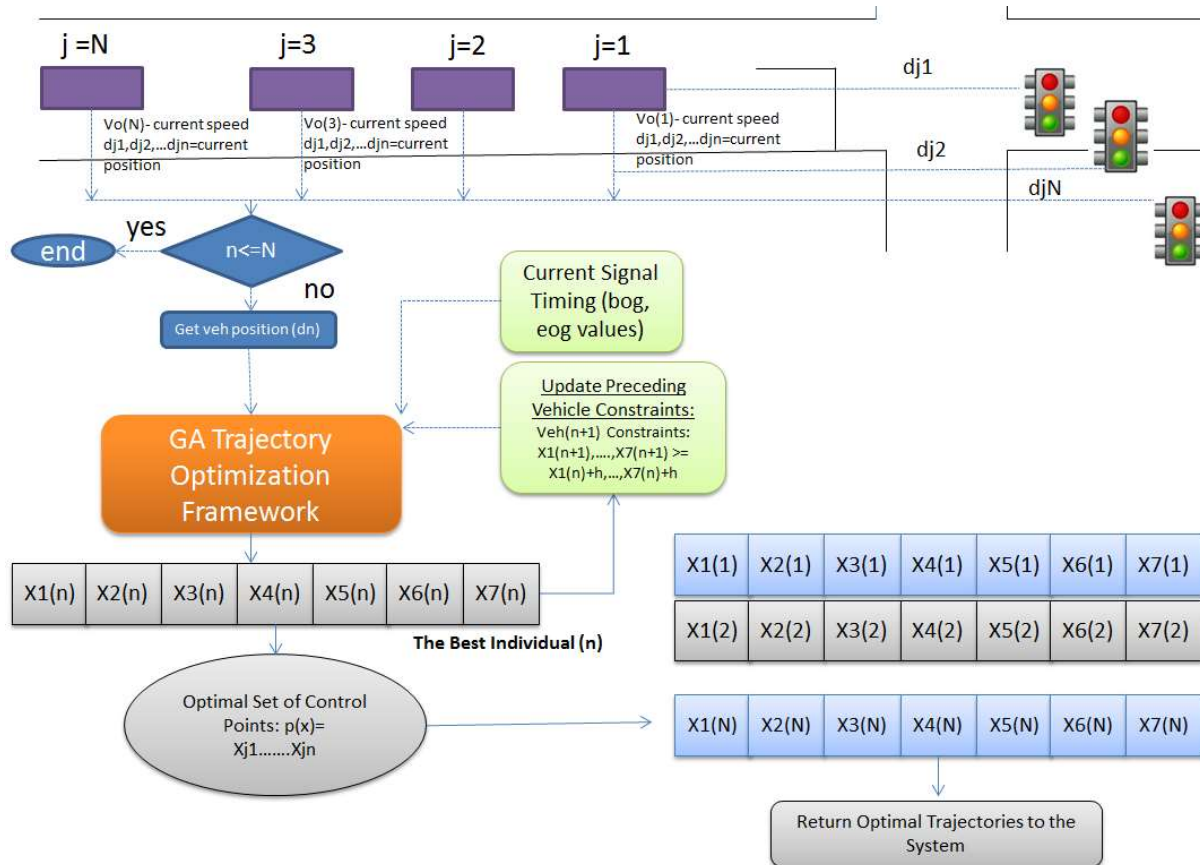


Figure 2: Optimization Framework for the Corridor Control Strategy

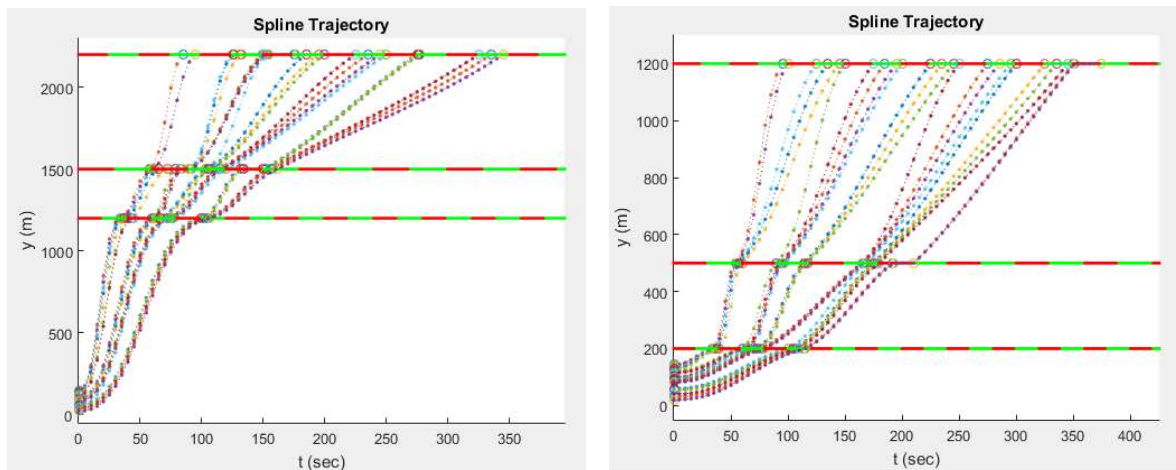


Figure 3: Numerical Analysis Results

As the proposed control strategy was developed to manipulate the prevailing traffic flow, rather than adjusting the signal timing and configuration, simple pre-timed devices are sufficient for the successful system operation.

Proof-of-Concept (POC) tests were conducted by using VISSIM (PTV, 2013)-based microscopic traffic simulation models. Besides the elimination of the complex signal control systems, the preliminary results for a signalized corridor in New Jersey indicate 1.2% to 5.4 % reduction in overall corridor travel time depending on different market penetration and traffic volume conditions.

References

- Barth, M., Mandava, S., Boriboonsomsin, K. and Xia, H., (2011), "Dynamic ECO-driving for arterial corridors". In *Integrated and Sustainable Transportation System (FISTS)*, 2011 IEEE Forum pp. 182-188.
- Federal Highway Administration (2016), "Fixing America's Surface Transportation Act or the FAST Act," <http://www.fhwa.dot.gov/fastact/>, (Accessed 30 July 2016)
- Iteris, "Connected Vehicle Reference Implementation Architecture," <http://www.iteris.com/cvria> (Accessed 30 July 2016)
- Jimenez, F., Aparicio, F. and Paez, J., (2008). "Evaluation of in-vehicle dynamic speed assistance in Spain: algorithm and driver behaviour". *IET Intelligent Transport Systems*, 2(2), pp.132-142.
- Katsaros, K., Kernchen, R., Dianati, M. and Rieck, D., (2011), "Performance study of a Green Light Optimized Speed Advisory (GLOSA) application using an integrated cooperative ITS simulation platform". In *Wireless Communications and Mobile Computing Conference (IWCMC)*, 2011 7th International, pp. 918-923.
- Paine, M., Paine, D., Griffiths, M. and Germanos, G., (2007), "In-vehicle intelligent speed advisory systems". In *Proceedings of the 20th International Conference on the Enhanced Safety of Vehicles*.
- PTV, A., (2013). PTV Vissim 6 user manual. Karlsruhe, Germany.
- Rakha, H. and Kamalanathsharma, R.K., (2011), "Eco-driving at signalized intersections using V2I communication". In *Intelligent Transportation Systems (ITSC)*, 2011 14th International IEEE Conference, pp. 341-346.
- RITA (2016), "Knowledge Resources.", <http://www.itsknowledgeresources.its.dot.gov/ITS/benecost.nsf/0/73DE374084895A2985257EF10065B157?OpenDocument>, (Accessed 30 July 2016)
- Seredynski, M., Mazurczyk, W. and Khadraoui, D., (2013), "Multi-segment green light optimal speed advisory". In *Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW)*, 2013 IEEE 27th International, pp. 459-465.
- Xia, H., Boriboonsomsin, K. and Barth, M., (2013). "Dynamic eco-driving for signalized arterial corridors and its indirect network-wide energy/emissions benefits". *Journal of Intelligent Transportation Systems*, 17(1), pp.31-41.
- Zhang, R. and Yao, E., (2015). "Eco-driving at signalised intersections for electric vehicles". *IET Intelligent Transport Systems*, 9(5), pp.488-497.

Towards Automated Driving: Development of an Unmanned Driving Vehicle for Road Maintenance

T. Reuber, C. Holldorb, J. Wachsmann
 Institute of Traffic Planning and Infrastructure Management
 Karlsruhe University of Applied Sciences, Germany
 e-mail: tim.reuber@hs-karlsruhe.de

Keywords: automated driving, road maintenance, safeguarding road works

Task & Motivation

Road works on the hard shoulder of motorways during moving traffic are common in Germany, they are called short-term road works. This everyday operational scenario carries a high risk for employees of the road maintenance service. Despite safety precautions, serious accidents caused by third parties happen frequently. This risk especially applies to the maintenance personnel in charge of safeguarding their colleagues during road maintenance.



Figure 1. A protective vehicle safeguarding a road maintenance vehicle in front of itself (Hessen Mobil)

As can be seen in Figure 1, short-term road works generally consist of two vehicles. There is one vehicle, the road maintenance vehicle, working at the site and another, the protective vehicle that follows up, in a distance of 50 to 100 meters. This protective vehicle safeguards short-term and mobile road works against moving traffic (Roos et. al., 2008). In the case of rear-end collisions the drivers of the protective vehicle risk their lives.

One approach to reduce this risk is the unmanned operation of the protective vehicle. For this purpose a consortium of eight partners, consisting of industrial partners, academic research institutions and German road authorities has been founded. These partners are working on the project “Automated Unmanned Protective Vehicle for Highway Hard Shoulder Road Works” (aFAS), partially funded by the German Federal Ministry of Economics and Technology (BMWi). In the aFAS project a prototype of an unmanned protective vehicle is being developed. This will be the very first unmanned operation

of a vehicle on German roads in public traffic without any supervision. Due to low speed, mostly straight roads and separate driving lanes the prototype has great potential for a first introduction of unmanned vehicles into public traffic.

Research Approach

The automation of driving is one of the most challenging fields of research in the automotive and traffic context. The level of automation can be defined after SAE, NTHSA and BAST (see Figure 2). SAE International is an international association of engineers, who developed a framework for describing the overall capabilities of vehicles. It is divided into five levels of automation, starting from “no automation” (level 0) to “full automation” (level 5). Systems of Level 4 and Level 5 are operated without human supervision like the engineered prototype of the protective maintenance vehicle.

SAE Level	SAE Name	SAE Narrative Definition	Execution of Steering/ Acceleration/ Deceleration	Monitoring of Driving Environment	Fallback Performance of Dynamic Driving Task	System capability (driving modes)	BAST Level	NTHSA Level
Human Driver monitors the driving environment								
0	No Automation	the full-time performance by the human driver of all aspects of the dynamic driving task	Human Driver	Human Driver	Human Driver	N/A	Driver only	0
1	Driver Assistance	the driving mode-specific execution by a driver assistance system of either steering or acceleration/deceleration	Human Driver and Systems	Human Driver	Human Driver	Some Driving Modes	Assisted	1
2	Partial Automation	Part-time or driving mode-dependent execution by one or more driver assistance systems of both steering and acceleration/deceleration. Human driver performs all other aspects of the dynamic driving task.	System	Human Driver	Human Driver	Some Driving Modes	Partially Automated	2
Automated driving system ("system") monitors the driving environment								
3	Conditional Automation	driving mode-specific performance by an automated driving system of all aspects of the dynamic driving task - human driver does respond appropriately to a request to intervene	System	System	Human Driver	Some Driving Modes	Highly Automated	3
4	High Automation	driving mode-specific performance by an automated driving system of all aspects of the dynamic driving task - human driver does not respond appropriately to a request to intervene	System	System	System	Some Driving Modes	Fully Automated	3/4
5	Full Automation	full-time performance by an automated driving system of all aspects of the dynamic driving task under all roadway and environmental conditions that can be managed by a human driver	System	System	System	Some Driving Modes		

Figure 2: Levels of driving automation according with SAE, BAST and NTHSA (Winner et. al., 2015)

Within the aFAS project, many different aspects have to be processed. One of these aspects is the characterization of the hard shoulders of motorways. For this purpose, 8.113 km carriageways in three federal states of Germany were analysed regarding their width, the condition of the carriageway markings and the condition of the pavement. In addition, the extremes of occurring weather conditions were considered.

Furthermore, operation scenarios of the road maintenance service were analysed. Typical works in short-term road works on the hard shoulder of motorways are maintaining the nature that surrounds highways, cleaning marker posts and sweeping carriageways.

These results are the basis for the general requirements of the vehicle, and further for the required sensor technology, as well as the selection of the test tracks.

State of research

The first milestone has been reached, and the development of the vehicle is nearly completed. The next step, executed by Karlsruhe University of Applied Sciences, is to evaluate the impacts of an automated and unmanned vehicle on traffic flow and traffic safety, as well as, on the road maintenance service. This will be accomplished by the usage of laser devices, ultrasonic units, radar units, GPS and several video cameras. The following parameters will be recorded:

- Localization of the protective vehicle on the hard shoulder
- Distance measurement between protective vehicle and moving traffic
- Traffic volume and density
- Speed of moving traffic
- Driving behaviour

Especially driving behaviour contains simultaneous recording with up to 6 video cameras, mounted on the protective vehicle and the road maintenance vehicle, filming in different directions. Everything in range of short-term and mobile road works is of importance.

The examination will take place for 20 days, in different operational areas. This will be split up into two parts of 10 days each. In autumn 2017, the examination of operation with the conventional, manned vehicle will be conducted on 10 days. In spring 2018, the operation with the automated, unmanned vehicle will be examined on 10 days.

The comparison should determine if differences exist. Lane change manoeuvres, changes of speed, time headway and the driving behaviour at weaving sections of the moving traffic are prioritised in the comparison.

Alongside the evaluation of the impacts, a benefit-cost analysis will be carried out. This will be analysed for the federal state Hesse, as well as nationwide. Several research studies, for example Roos et. al. (2008), Cindric-Middendorf, Zimmermann (2011), Kemper (2010) examined the incidence of accidents involving short-term and mobile road works. The outcomes of these studies can be used to quantify the costs of the accidents. Following the reduction of accident hazard must be compared to investment costs and running costs of the automated protective vehicle. In this process personal injuries, social and medical aspects are also taken into consideration.

Conclusion & outlook

The first goals have been reached successfully. At this point, the project is ahead in its progress. For example, the operating environment has been specified and the development of the automated, unmanned vehicle is completed. Upcoming operations of the road maintenance service are carried out and the calculation of benefit-cost analysis has begun. The analyses will be completed July 2018. The results of this research can also be the basis of a new project called “Test Area Autonomous Driving Baden-Wuerttemberg”.

References

- Cindric-Middendorf, D., Zimmermann, M. (2011). *Implementierung von Informationen zu Arbeitsstellen kürzerer Dauer zur Gefahrenwarnung in On-Board-Informationssystemen, Untersuchung im Auftrag der BMW AG in Zusammenarbeit mit dem Hessischen Landesamt für Straßen und Verkehrswesen, Karlsruhe. (unveröffentlicht)*
- Kemper, D. (2010). *Vergleichende Betrachtung der Sicherheit und Wirtschaftlichkeit von Arbeitsstellen kürzerer Dauer auf Autobahnen bei Tag und Nacht, Aachener Mitteilungen Straßenwesen Nr. 52, Erd- und Tunnelbau, Institut für Straßenwesen, RWTH Aachen.*
- Roos, R., Zimmermann, M., Riffel, S. B. and Cypra, T. (2008). *Verbesserung der Sicherheit des Betriebspersonals in Arbeitsstellen kürzerer Dauer auf Bundesautobahnen, BASt, Tech. Rep. Heft V170.*
- Winner, H., Hakuli, S., Lotz, F. and Singer, C. (2015), *Handbuch Fahrerassistenzsysteme. Grundlagen Komponenten und Systeme für aktive Sicherheit und Komfort, Wiesbaden, Springer Vieweg.*

Development of a concept of need-based de-icing salt storage capacities and the implementation in a dimensioning guideline

J. Wachsmann, C. Holldorb, J. Schappacher
Institute of Traffic Planning and Infrastructure Management
Karlsruhe University of Applied Sciences, Germany
e-mail: jan.wachsmann@hs-karlsruhe.de

Keywords: road maintenance, winter maintenance, extreme situations

Task

Efficient winter maintenance on the classified road network is a prerequisite for safe driving in wintery road conditions and furthermore for ensuring winter serviceability to the road network as an essential part of infrastructure and economy. Alongside efficient staff, vehicles and technology it is necessary to have sufficient spreading materials for clearing and preventing ice.

Generally, barns and silos are available for the storage of spreading material. In Germany differentiated standards and recommendations for dimensioning de-icing salt storage spaces for the winter service do not yet exist. In the past, the dimensioning was oriented strongly on experience, available funds and benchmarks, so long-lasting and especially transregional wintery periods partly led to massive shortages in the salt supply.

Research Approach

In the framework of the completed 04.0243/2011/LRB “Key Figures in Operational Service – KENNBET” R&D project (Holldorb and Streich, 2013), a calculation model (Badelt, 2011) for the determination of the necessary de-icing salt storage capacities was developed and utilised on a general basis. By means of historic weather data and consistent spreading scenarios, the resulting salt usage quantities to maximise the security of supply for a so-called extreme winter (severest winter of the last 50 to 100 years) has been calculated. In this context, the influence of different concepts for storage and subsequent delivery on the necessary storage capacity has been researched and monetarily evaluated.

The calculation model was based on an allocation of the salt requirements and storage capacity specific to each road maintenance depot. However, specific local requirements, e.g. special route characteristics, porous asphalt, etc. have not been considered. Possible changes of the demand for spreading material as a result of climate change (Holldorb *et al.*, 2016a) and due to technological developments in the application of spreading material (use of FS 100¹) have also not been considered in the model. Therefore, the developed calculation model was an available approach with which the necessary de-icing salt storage capacities could be estimated on a state level based on detailed data, but which only insufficiently takes into account the local requirements and their changes.

In the framework of the 04.0288/2015/KRB “Concept of need-based de-icing salt storage capacities – TAUSALA” R&D project (Holldorb *et al.*, 2016b), a dimensioning procedure has now been further

¹ 100 % brine instead of prewetted salt

developed to ascertain the necessary de-icing salt storage capacities for the individual road maintenance depot. In order to do so, the dimensioning concept (see Figure 1) has been further refined mainly in order to consider the differentiations stated by the road administrations of the federal states, suggested as a reaction to the final report of the “KENNBET” R&D project. Adaptions of the dimensioning model in the following points have been examined in points of necessity and purpose:

- Adaption of the allocation to climate regions
- Examination of spreading material scenarios (consideration of preventive action, increased spreading density at lower temperatures, consideration of pure salt brine spreading)
- Examination of spreading areas to be set (consideration of incline sections or sections on ridges, sections with asphalt surfaces made from porous asphalt, consideration of the hard shoulder, consideration of rest areas on federal motorways, differentiation of the scope of winter service according to road classification)
- Consideration of climate change

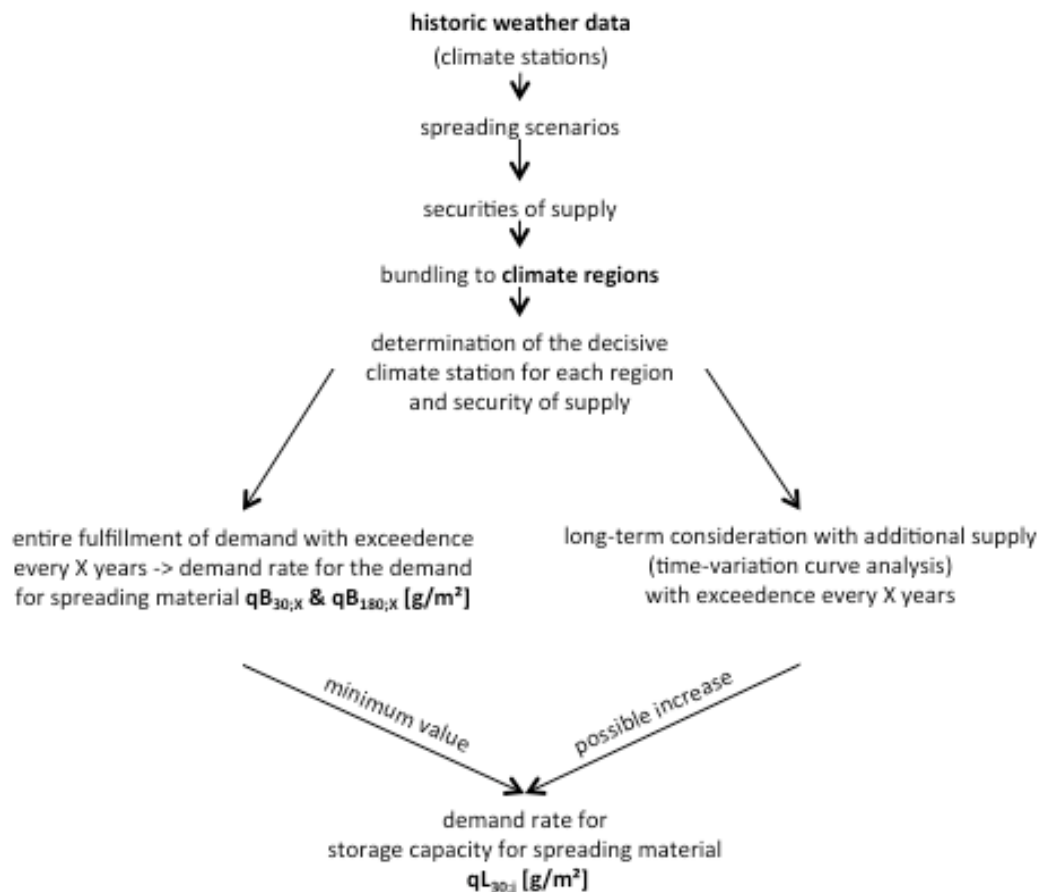


Figure 1: Determining the demand rate for the necessary storage capacity for spreading material

Ascertaining annual costs and capital costs on a general basis for different storage and logistic concepts and the juxtaposition with a consideration of economics should answer which security of supply (maximum security, 30-year security or 10-year security) to be set is the most sensible in terms

of economics. Therefore the costs for the storage of spreading material were juxtaposed with the costs as a consequence of insufficient security of supply, which consist of

- economic damages through additional travel times as a consequence of reduced speed
- economic damages through additional travel times as a consequence of traffic jams
- economic damages through additional accident costs
- economic damages through a breakdown of road infrastructure.

The evaluation of the economic damages through additional travel times as a consequence of reduced speed is based on the determination of the mileage on an average winter day, the decrease of speeds based on percentage, as well as the adapted time cost rates of the “recommendations for economic feasibility studies” (EWS). The determination of the economic damages through additional travel times as a consequence of traffic jams follows a global approach depending on lost time and affected network share. The economic costs in consequence of road traffic accidents are based on data from the Federal Office of Statistics from 2009 to 2013. The breakdown of road infrastructure is mirrored in a decrease of the gross national product. However, it is difficult to quantify this so for this general approaches were considered.

The examination and further development of the dimensioning model, as well as the development of a dimensioning guideline ensued in close coordination with a project specific committee, which consisted of representatives of the Federal Ministry of Transport and Digital Infrastructure (BMVI) and the Federal Highway Research Institute (BAST), as well as representatives of four federal states. An ancillary Data Processing (DP) application was developed in coordination with the committee, as well as with pilot users from the four federal states involved. The pilot users were involved in the framework of a workshop, as well as first users in the DP development to guarantee functionality and practical manageability.

Conclusions for practice

The overall target of the “Concept of need-based de-icing salt storage capacities – TAUSALA” R&D project was the implementation of the developed dimensioning model into a dimensioning guideline, as well as a simple, practical Data Processing (DP) application so that the necessary spreading material storage requirements can be ascertained on the basis of consistent dimensioning principles for every motorway and road maintenance depot, however, while considering specific requirements and constraints.

The target has been reached successfully. The blueprint of the dimensioning guideline, as well as the DP application, is now in review by all Federal States in order to be launched nationwide.

References

- Badelt, H. (2011). Ein neues Modell für einen Winterindex zur Abschätzung und Bewertung des Salzverbrauchs, *Kolloquium Straßenbetrieb 2011*, FGSV-Verlag, Köln.
- Holldorb, C., Rumpel, F., Gerstengarbe, F.-W., Österle, H. and Hoffmann, P. (2016a). *Analyse der Auswirkungen des Klimawandels auf den Straßenbetriebsdienst (KliBet)*, *Berichte der Bundesanstalt für Strassenwesen. Verkehrstechnik*, Heft V 270, Fachverlag NW in der Carl Schünemann Verlag GmbH, Bremen.
- Holldorb, C., Schappacher, J. and Wachsmann, J. (2016b). *Konzeption bedarfsgerechter Tausalzlagerkapazitäten: Schlussbericht zum FE-Vorhaben 04.0288/2015/LRB im Auftrag des Bundesministeriums für Verkehr und digitale Infrastruktur*, Karlsruhe.

Holldorb, C. and Streich, M. (2013). *Kennzahlen im Betriebsdienst: Schlussbericht zum FE-Vorhaben 04.0243/2011/LRB im Auftrag des Bundesministeriums für Verkehr, Bau und Stadtentwicklung*, Karlsruhe.

Assessing the Value of GPS Travel Data for Travel Forecasting Models

Jeffrey M. Casello

School of Planning and Department of Civil and Environmental Engineering

Xiameng Xu

Pedram Fard

Andrea Mikkila

School of Planning

University of Waterloo, ON Canada

e-mail: jcasello@uwaterloo.ca

Keywords: GPS data collection, travel forecasting, data validation

Introduction

A key component to travel modeling is trip generation – the expected number of trips to begin in a Traffic Analysis Zone (TAZ) (Ortúzar & Willumsen, 2014). Normally, generation rates are estimated using empirically gathered data on household and person trip activity over time. When sufficiently large samples are gathered, the generation rates can be estimated as a function of the travelers' socio-economic data (gender, age, income, household composition, number of vehicles) (McNally, 2000).

This study aims to develop, implement and evaluate a passive data collection method to inform trip generation for a regional travel forecasting model in the City of Edmonton, Alberta Canada. In this research, data are gathered on traveler behavior using two approaches. First, respondents self-report their travel activities using a web-based survey. In addition to the web-based survey, GPS data are gathered using a customized smartphone “app” – EdmoTrack – developed at the University of Waterloo. The specific foci of this research are to assess and improve the functionality of GPS data collection and analysis as a tool to improve the development of travel forecasting models.

Related Research

Paper-based surveys on travel behavior are known to produce low response rates and poor accuracy in those responses received (Hooper, 2017). Various, improved recruitment methods have been studied (Greaves, et al., 2014; Millar and Dillman, 2011). Studies such as Alsnih (2006) and Sills and Song (2002) have shown that compared to paper-based surveys, web-based surveys tend to have higher response trip rates at lower costs. Yet, there remains a tendency to underreport trips, most notably local walking, cycling and other short trips.

The widespread use of smartphones that are equipped with Global Position Systems (GPS) now offers a low cost, passive way to gather traveler information (Shen and Stopher, 2014). The primary benefits of GPS are the ease of data collection and detailed trip records produced. Possible challenges are both technological – the reliability of GPS signals and batter consumption, for example – and social – the potential sacrificing of traveler privacy through their participation in a GPS study.

Methods

This study attempts to answer three major sets of research questions:

1. Is GPS an effective way to validate the self-reported trip activities when building a travel forecasting model? Under what circumstances will the GPS data produced be of sufficient quality to add value to the travel forecasting model development?

2. Can the GPS data be used to correct errors in self-reported trips? More specifically, is it possible to compute the rate of underreported trips such that trip generation rates can be made more accurate?

To answer the first two questions, it is necessary to identify the number of trips made by a traveler. This observation leads to the third set of research questions:

3. What attributes of the GPS can be used in an automated algorithm to identify travel as opposed to conducting an activity? Under what circumstances will this algorithm work well?

We begin by defining a trip as travel occurring between two activities. The most obvious differentiating factor between activities and travel is very low travel speed. As such, the algorithm begins by computing speeds between consecutive points. Speeds that exceed a “stationary” threshold are initially identified as “travel” points, while speeds below the threshold are labeled as possible activities. This initial classification is insufficient to correctly sort travel from activities. Consider the case where an automobile traveler is stopped at a traffic signal. The speed data alone would incorrectly indicate an activity. On the other hand, sufficient “noise” exists in the GPS data that while stationary – for example sitting at one’s desk in an office – the GPS data can reflect constant motion at speeds that may be construed as walking velocities.

To further refine the classification, the algorithm next considers the trajectory of movement observed. During an activity, GPS data tend to be circuitous whereas GPS data during travel are typically more direct. To differentiate activities from travel, the directness of travel can be measured as the ratio of total cumulative travel distance to Euclidean distance between a subset of n points. Equation 1 expresses this calculation mathematically, while Figure 1 shows the calculation graphically.

$$Directness = \frac{\sum_{i=2}^n Distance(x_{i-1}, y_{i-1}; x_i, y_i)}{EuclideanDistance(x_1, y_1; x_n, y_n)} \quad \text{Equation 1}$$

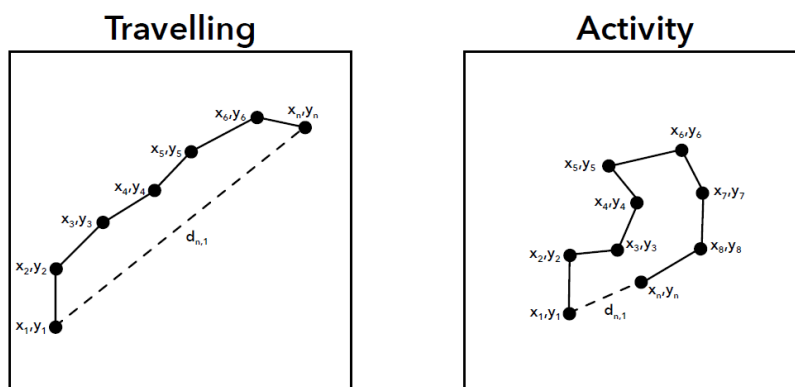


Figure 1 Contrasting GSP data records (trajectories) while travelling or conducting activities

When traveling, equation 1 generates values near to 1; while conducting an activity, results are near to 0. In our work, empirical approaches were taken to generate appropriate thresholds for equation 1, and for the value of n to produce the best results.

To differentiate travel stops from activities, a final step is conducted in a GIS. Potential activities – subsets of points with low speeds and circuitous travel – are analyzed in a spatial context. When these sets of points occur within a buffer distance to a roadway or travel facility, the points are re-labeled as travel. The results of these three steps are a set of activities and trips for a participant that can then be compared to the self-reported data.

We now have two records for the number of trips made by respondent – the self-reported data and the GPS analysis data. To determine the actual number of trips taken, we take the following approach. For a given participant, we begin with the first self-reported trip. We compare the reported origin, destination and time to the first record in the GPS file. If these two trips match (within acceptable error tolerances), we consider the trip to be valid and we iterate to the next trip on the respondent's list. If the two data sources do not match and valid GPS data exist, then we assume that the GPS data are correct and the traveler made an error in recording a trip, or failed to report a trip taken. If no valid GPS data exist for the time period when the trip began, then we assume that the traveler made the reported trip, but without gathering the concurrent GPS data. We again iterate through this process until all the self-reported trips have been tested. The result of this step is a record for each participant with the three data points on trip activity: self-reported, GPS identified, and truth. Figure 2 shows the logic.

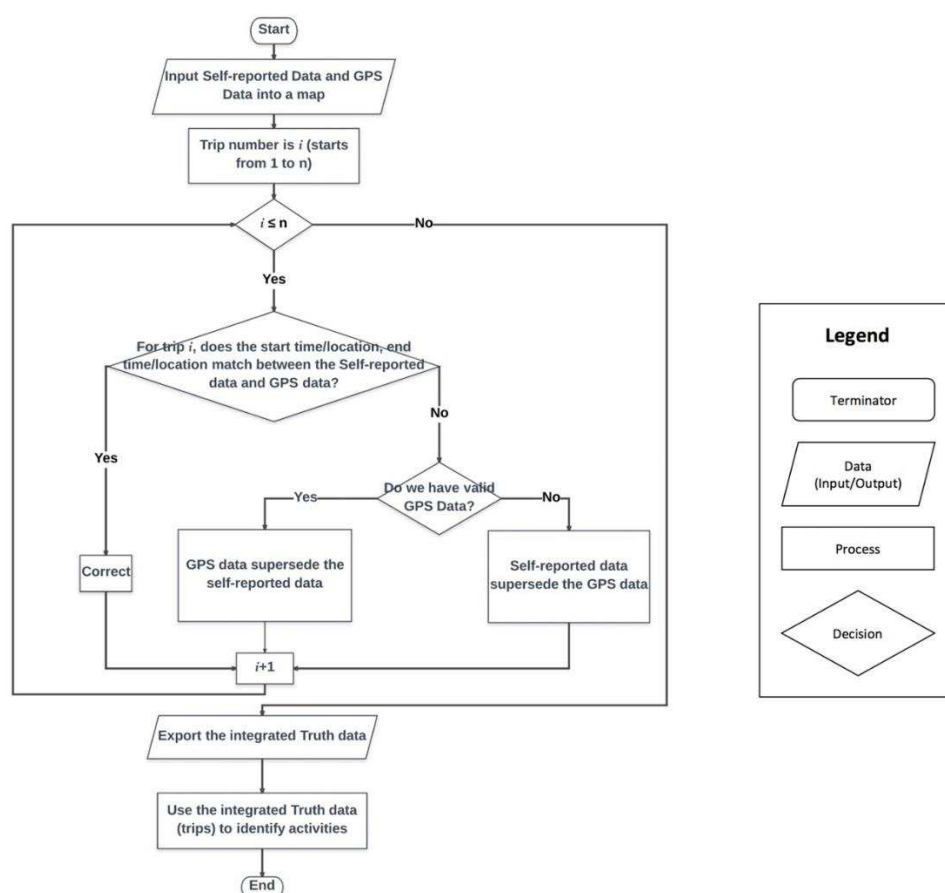


Figure 2: The Logic

Results

The analysis demonstrates that about 24% of respondents underreported their trips. Only 8% of respondents overreported their trips. The integrated data set also indicates that the GPS analysis alone underestimates the number of trips made by 65% of respondents, while the GPS analysis generates an overestimate of trip activity for 14% of respondents. The majority of the GPS errors are a result of participants failing to keep the app running on their smartphones. In many cases, respondents turned off the app while beginning an activity and never restarted the program to capture travel. Beyond user error, other sources of GPS error include entry into tunnels where signal obstructions can prevent GPS signals reaching the smartphone; travel by Light Rail Transit, where presumably the electric powered vehicles produced interference which precluded reliable GPS. The GPS algorithm developed also produced erroneous results when travelers engaged in a uniquely North American travel activity – entering “drive thrus” – as these activities were commonly misclassified as travel stops.

Conclusions

The results of this research suggest that additional training and education may need to be done in order for GPS to become the sole data source for a large-scale, professional studies. But, this work suggests that GPS can be used effectively to validate the self-reported trips, and adjust the generation rates to better inform travel models.

References

- Alsnih, R. (2006). Characteristics of Web Based Surveys and Applications in Travel Research. *Travel Survey Methods. Quality and Future Directions*, pp. 569-592.
- Greaves, S. et al. (2014). A Web-Based Diary and Companion Smartphone App for Travel Activity Surveys. Leura, Australia, s.n.
- Hooper, K. G. (2017). *Trip generation handbook: an ITE recommended practice*. Washington, D.C., Institute of Transportation Engineers.
- McNally, M. (2000). The four-step model (ch. 3). In: Hensher, D., Button, K. (Eds.), *Handbook of Transport Modeling*. Pergamon, Amsterdam, pp. 35–69.
- Millar, M. M. & Dillman, D. A. (2011). Improving Response to Web and Mixed-Mode Surveys. *Public Opinion Quarterly*, 75(2), pp. 249-269.
- Ortúzar, J. D., & Willumsen, L. G. (2014). *Modelling transport*. Chichester, Wiley.
- Parsons, C. (2007). Web-based Surveys – Best Practices Based on the Research Literature. *Visitor Studies*, pp.13-33.
- Shen, L., & Stopher, P. R. (2014). Review of GPS Travel Survey and GPS Data-Processing Methods. *Transport Reviews*, 34 (3), pp. 316-334. doi:10.1080/01441647.2014.903530
- Sills, S. J., & Song, C. (2002). Innovations in survey research: An application of web-based surveys. *Social Science Computer Review*, 20(1), pp. 22–30.

Dealing With the Rise of Transportation Network Vehicle Service: Rethinking the Regulatory Regime for TNVS Towards Sustainable Urban Mobility

Mark Anthony M. Gamboa
School of Urban and Regional Planning
University of the Philippines Diliman, Quezon City, Philippines
e-mail: mmgamboa@up.edu.ph

Keywords: urban mobility policy, social inclusion, ride-hailing services

Introduction

Sustainable urban mobility has been a key concern in the Philippines as urbanization continues to increase the demand for an inclusive, safe, efficient, and reliable transport services. The rise of a new business model in transportation as a result of rapid advancements in mobile technology has tested the policy landscape in the Philippines. In 2015, the Philippines' Department of Transportation (DOTr) issued Department Order No. 2015-11 updating the standard classifications for public transport conveyances in the country by introducing a new type of service known as the transportation network vehicle service (TNVS)- a pre-arranged transportation through the use of internet-based digital technology application on an internet-connected device.

The rise of Transport Network Companies (TNCs) and TNVS has slowly transformed the country's taxi industry. More people have been choosing TNVS over the traditional taxi services. Of late, however, the operation of the TNCs and TNVS has been put to question. Relying on existing transportation regulations such as the Commonwealth Act No. 146 (1936), the Land Transportation Franchising and Regulatory Board (LTFRB), the government agency with the primary mandate to regulate public land transportation services in the Philippines, suspended the operation of some TNVS. This suspension has led to a gridlock and a debate between public welfare and the existing regulatory mechanisms for transportation.

In this context, this paper focuses on the review of the regulatory regime for TNCs and TNVS with the objective of providing an understanding of how the legal and policy landscape can adapt to a dynamic and advancing industry environment. This paper explores the emergence of TNVS as a result of the impact of advancements in mobile technology on transport services in the Philippines in the context of regulatory issues that come with it with the aim of proposing policy reforms.

Research Approach

The main contribution of this paper to relevant literature is both practical and theoretical as it will help identify emerging regulatory regimes for TNCs/TNVS which the Philippine Government can consider as it embarks on formulating pieces of legislation and regulations to ensure that this emergent business model is integrated in the system of sustainable urban mobility. Likewise, the use of an initial regulatory impact assessment (RIA) framework (OECD, 2008) will add to the literature on the emerging application of the RIA framework to produce relevant and responsive public policies.

The paper starts with mapping through the conduct of systematic content analysis (Stemler, 2001) of legal, policy, and regulatory instruments in the Philippines including the evolution/timeline of legislation and policy initiatives relevant to the operations of TNVS and TNCs in the context of social inclusion. The capacity of these instruments will be assessed in terms of design, implementation, and outcomes with the primary goal of identifying issues and problems. The instruments will be assessed individually and collectively to determine its design coherence and consistency to inclusive mobility indicators. The assessment will also include the determination of whether the Philippine TNVS/TNC legal infrastructure reflects current realities and conditions or whether these policies lead to social exclusion.

Based on the assessment, policy and institutional reforms shall be recommended to develop and strengthen the capacity of the legal framework for sustainable and inclusive urban mobility in the Philippines with particular emphasis on the regulatory regime for TNVS/TNCs.

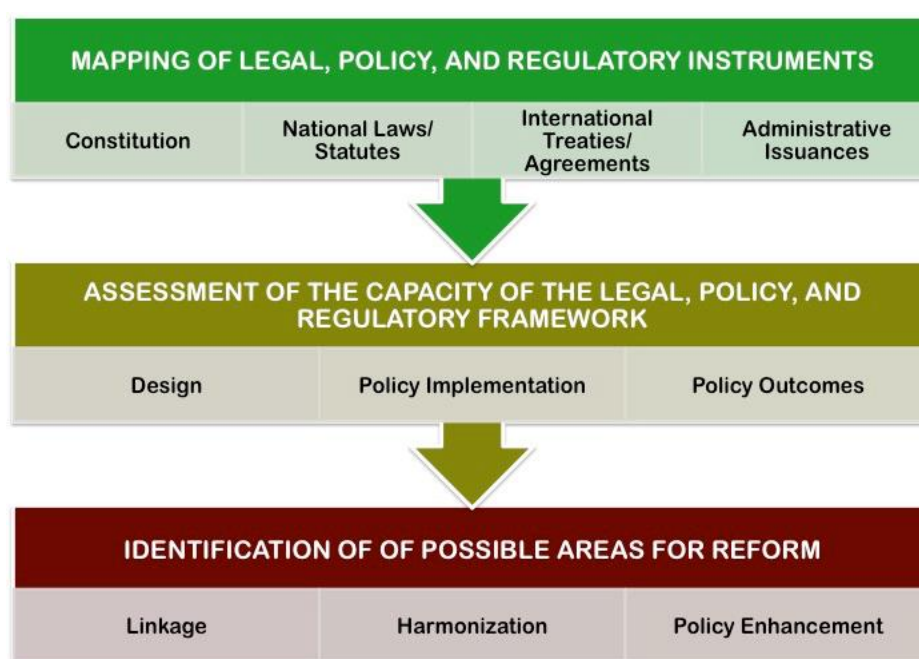


Figure 1: Methodological Framework of the Paper

Preliminary Results

TNVS has recently emerged as a new mode of transportation in the country. Uber and Grab, which are the largest TNCs in the Philippines, started operations in the country in 2015. The DOTr issued Department Order No. 2015-11 as a response and recognition of the role that these TNCs play in urban mobility systems. In turn, the LTFRB issued several memorandum circulars, which, among others, set the guidelines for the acceptance of applications for a certificate of public convenience (CPC) to operate a TNVS. GrabCar was the first to receive full accreditation as a TNC in the Philippines (Grab, 2015).

In 2016, the LTFRB has suspended the processing of applications of TNVS and has even threatened to go after TNVS operators that lack CPC. Uber and Grab were fined by LTFRB with Php5 million each for their failure to abide by government regulations on accreditation of drivers (Yee, 2017). The LTFRB further ordered the suspension of more than 50,000 ‘colorum’ Uber and Grab drivers (Amilhamja 2017). According to Francisco (2017), “[T]he issue sparked fierce debate on both sides – with commuters saying Grab and Uber are essential to metro life

and much more reliable than ordinary taxis, while the government insists that these need to be regulated to ensure accountability.”

People find TNVS a much better alternative to an inefficient traditional taxi industry service. In a study by Paronda et al. (2016) it was concluded that Uber and GrabCar provide better quality of service, more convenience, and flexibility than conventional taxi. This is confirmed by a survey conducted by Nistal and Regidor (2016) on Uber as a TNVS, which showed that “commuters have embraced Uber” and that even if a traditional taxi would offer the same price and service, “most users will still prefer Uber”.

A preliminary review of the current instruments regulating the TNCs and TNVS in the Philippines shows the inadequacy of current regulatory regime to address the emerging modes of transport services as a result of advancement in mobile computer technology towards urban social mobility. The legal and institutional arrangements, which deal mainly with a range of formal laws, policies, and administration systems, are a key element in providing an enabling environment for sustainable urban mobility. According to Macario (2000), “[T]he lack of a clear and well-structured regulatory and organizational framework is a determinant factor that may hinder the successful definition and implementation of a coherent mobility system.”

This new business model is under the category of collaborative economy that now proliferates in the Philippines necessitating “the government to revisit its regulation mechanisms and explore how to integrate businesses arising from it into the broader economy and make them legitimate economic entities” (Valencia, 2017). However, the Philippine regulatory regime is currently “dedicated to the ownership-based economy, and not the sharing economy” (Weston et al., 2015 as cited in Valencia, 2017).

There are now pending bills in the Philippine Congress to regulate the operation of these TNCs/TNVS. This indicates the necessity of having a new and comprehensive regulation as the current regime is not adequate, or worse inappropriate, to address the emergence and rise of TNCs and TNVS in the country.

Conclusions and Recommendations for Further Research

There is an inadequacy of the current regulatory regime for TNCs/TNVS in the country. The current regime does not address the distinct character of TNCs/TNVS as a public conveyance. Even the regulatory agencies are very adamant to stick to the existing regulatory regime that is outdated and does not respond to emergent and new modes of transport services. The current policies, such as the Public Service Act also known as Commonwealth Act No. 146- a law that was passed in 1936- are not sufficient to address several issues, such as fare “surge” rates or the higher pricing scheme that these TNCs implement which, if practiced by traditional taxi services would be a violation of their franchise and the laws. In addition, there is no clear guideline as to the treatment of ride-hailing services that operate in the same way as traditional taxi service. It however has to be emphasized that regulations should not serve as a hindrance to the development of public transportation services. As such, it is important to establish an enabling policy environment that will harness the benefits that ride-hailing services provide to the public in general.

References

- Amilhamja, C. (2017). 'LTFRB stands firm vs. Grab, Uber but clarifies services not suspended', *Inquirer.net*, 18 July, [online] available: <http://newsinfo.inquirer.net/914805/ltfrb-stands-firm-vs-grab-uber-but-clarifies-services-not-suspended> [accessed 05 August 2017].
- Francisco, K. (2017). 'What's the fuss about the Grab, Uber regulation issue?' *Rappler*, 28 July, [online] available: <http://www.rappler.com/newsbreak/iq/176933-you-need-to-know-fuss-grab-uber-ltfrb-regulation-explainer> [accessed 05 August 2017].
- Grab (2015). 'GrabCar first and only Transport Network Company legalized by the Philippine Government', [online] available: <https://www.grab.com/ph/press/business/grabcar-first-and-only-transport-network-company-legalized-by-the-philippine-government/> [accessed 05 August 2017].
- Macario, R. (2000). 'Upgrading quality in urban mobility systems', *Total Quality Management*, 11(4-6), pp. 747-753.
- Marletto, G. & Mameli, F. (2012). 'A participative procedure to select indicators of policies for sustainable urban mobility. Outcomes of a national test', *European Transport Research Review*, 4(2), pp. 79-89.
- Nistal, P.D. and Regidor, J.R.F. (2016). 'Comparative study of Uber and regular taxi service characteristics,' [online] available <http://ncts.upd.edu.ph/tssp/wp-content/uploads/2016/08/Nistal-Regidor.pdf> [accessed 05 August 2017].
- OECD (2008). 'Introductory Handbook for Undertaking Regulatory Impact Analysis (RIA)', [online] available: www.oecd.org/gov/regulatory-policy [accessed 05 August 2017].
- Paronda, A.G.A., Regidor, J.R.F. and Napalang, M.S.G. (2016). 'Comparative Analysis of transportation network companies (TNCs) and conventional taxi services in Metro Manila'. presented at 23rd Annual Conference of the Transportation Science Society of the Philippines, 08 August.
- Stemler, S. (2001). 'An overview of content analysis', *Practical Assessment, Research & Evaluation*, 7(17), [online] available: <http://pareonline.net/getvn.asp?v=7&n=17> [accessed 05 August 2017].
- Valencia, R.G. (2017). 'The rise of collaborative economy in the Philippines', *Economic Issue of the Day*, 17(1), [online] available: <https://pidswebs.pids.gov.ph/CDN/PUBLICATIONS/pidseid1701.pdf> [accessed 05 August 2017].
- Yee, J. (2017). 'LTFRB to impound Uber, Grab cars lacking franchise', *Philippine Daily Inquirer*, 16 July, [online] available: <http://newsinfo.inquirer.net/914214/ltfrb-to-impound-uber-grab-cars-lacking-franchise> [accessed 05 August 2017].

Innovation Coaching in Foresight Processes for Distributed Product Development

Carsten Hahn¹, Albert Albers², Markus Stöckner¹, Manuel Niever¹, Benjamin Walter², Roman Kerres¹, Nikola Bursac²

¹Karlsruhe University of Applied Sciences, Germany

²Karlsruher Institut für Technologie (KIT), Germany

e-mail: manuel.niever@hs-karlsruhe.de

Keywords: Innovation Coaching, Live-Lab, Design Research, Distributed Product Development, Virtual Collaboration, ProVIL, ASD - Agile Systems Design

Abstract: Between globalized competition and shortened lifecycles on the one hand and the new possibilities of digitalization on the other, companies are increasingly confronted with the challenge of optimizing their innovation processes. While the implementation of approaches, such as open innovation, co-creation or agile development, often leads to the desired success, they require additional competencies and resources on the part of the correspondent product developers. This applies in particular in cases where the persons involved do not work together at a common location and therefore depend on the use of media such as web conferences and innovation platforms. This situation results in the increasing practical need for effective support, as would be possible, for example, through the use of appropriately trained experts as enablers in the sense of innovation coaching. However, there are currently few approaches to innovation coaching. Existing approaches also have the weakness that they are too late in the innovation process and tend to focus on purely economic aspects, while technical aspects take a back seat. Against this background, a research project has recently been implemented at the institutes of the participating authors with the aim of making innovation coaching usable for the early phases of innovation processes and treating economic and technical perspectives in the same way. This paper presents the results of two empirical studies carried out in innovation projects with 42 and 64 participants. It can be shown which core activities and responsibilities are decisive for successful innovation coaching in early innovation phases and which problems and unused potentials exist in this field. In addition, hypotheses can be derived that will guide future research in the area of innovation coaching in the coming years.

1. Motivation

Due to the Globalization companies of all sizes and branches increasingly face the challenge of distributed collaboration across locations and countries including people of different cultural backgrounds across different hierarchy positions (Eppinger/Chitkara, 2006). Especially in the discipline of product innovation globally distributed teams are getting more popular. These teams can be easily set up in order to create an ideal combination of expertise and different thinking. In order to speed up and increase the efficiency of innovation processes, companies build up virtual teams on the basis of information and communication technologies (ICT). The ICT facilitates the geographic distribution of team members and potentially allows for extensive blurring of team boundaries (Haines, 2014). These technologies strive to achieve an equal level of quality in virtual teams than in face-to-face teams. Even though there are lots of ICT tools to support distributed team work there still are differences between working virtually and face-to-face. These differences can be experienced for instance in communication because of time delays, quality issues or the missing transmission of body language.

Furthermore, the classic approach of an innovation process is hitting its boundaries because of the

increasing speed in innovation cycles. Customer needs as well as technological possibilities are changing quickly and offer great chances to create new value for companies which are able to adopt these information and trends. In the foresight process companies are merging all kind of internal as well as external information in order to get a clear understanding of future requirements. This information is then to be used to derive innovation impulses in the present. In order to meet these market requirements many companies adopt different styles in managing innovation processes. Methods like agile development, open innovation and co-creation experience an increasing popularity (Chesbrough, 2006). Even in globally distributed teams, companies are aiming to use or are using these methods. The combination of these strategies and the working environment of distributed teams leads to manifold challenges when it comes to the management of an innovation process, especially in the development of new product generations (Albers et al., 2015a).

In both, the usage of ICT tools for distributed teams as well as in using agile methods of product development companies are facing new challenges. This paper will discuss a possible approach based on a new concept of innovation coaching in order to meet these challenges.

2. State of the Art

Product development in distributed teams

In order to integrate worldwide distributed internal as well as external expertise in product development the use of modern software tools from the area of information and communication technology (ICT) is necessary. These tools can not only consist of e-mail, instant messaging or telephone but also of more sophisticated software. Examples are video-conferencing, shared and digital workspaces or communication and working platforms for group working and idea sharing. On the one hand this work environment has advantages such as a very flexible team size and team composition. Therefore, it is easy to integrate different people at different phases in the product development process based on their particular knowledge and experiences. The team boundaries can be adjusted to the specific needs of the process step. (Haines, 2014) On the other hand, the use of ICT leads to great challenges as well. Those have different origins and can be separated into technical, organisational and human challenges (Bullinger et al., 1998).

Technical challenges include the access to appropriate information and communication technologies as well as the technical usability of these tools. In order to achieve good outcome within distributed teams it not only take an appropriate infrastructure but also different tools than in face-to-face teams. For instance, software tools need to be collaborative. This means different members are able to work together at that same task anytime and from anywhere – in real time. Technical problems, a bad user-experience and a not clearly structured workflow of the ICT tools can lead to great efficiency losses (Walter et al., 2016a).

Organisational challenges emerge especially in complex processes such like an agile product development process including several different participants from different cultural backgrounds. This calls for a new set of routines which has to be developed, implemented and executed. (Bessant, 2003). This begins with different time zones the team is located in and continues with different processes for instance in the identification of who is responsible for delivering a task. Also, there may occur different organizational frameworks like the structure of task allocation, coordination and supervision. These organizational differences between distributed team members were found to negatively affect communication and therefore effectiveness (Alzoubi et al., 2015).

Human challenges arise because positive team developments are based on personal communication and interaction in the team (Kotlarsky/Oshri, 2005). Because traditional methods of social control that rely on face-to-face interaction are not as powerful in building up a positive team atmosphere leaders of distributed teams need different tools for inducing trust. Haines (2014) suggests that leaders of virtual teams need to be aware that a group development relies on cultivating a virtual workspace

that is similar to real world team environments. Thus he says that team members should have clear, specific goals and should be encouraged to communicate with each other.

Innovation methods in the foresight phase

Especially in the foresight process of a product development project, not only every team member needs to be integrated, but also their knowledge. In addition to that it is advisable to integrate external knowledge into this phase. This can be described by concepts like co-creation or even open innovation. Co-creation is about joint creation of value by the company and the customer. It relates to topics like defining a problem, but in the very core of co-creation also to the co-construction of the product or service in question (Albers et al., 2017). Also, it is important to merge external and internal knowledge in order to get a clear understanding of future technologies. Combined with methods like the scenario technique, this should allow to recognize the technological and scientific requirements as well as general societal relevance in the future in order to derive and implement innovation impulses for the present (Walter et al., 2017a). With these insights, the team aims to understand the customer problem in order to derive customized and detailed product profiles. Only with a deep understanding of a problem worth solving and a defined challenge in the format of a product profile, one is able to find fitting solutions. (Walter et al., 2016b) After identifying the market potentials of this problem, the development team will need to create ideas how to solve it. This step can be supported with several innovation methods which enable the creativity and leads to new ideas (Albers et al., 2016). Depending on the team members, the technical framework and further factors, the right innovation method needs to be chosen and conducted in an efficient way.

Innovation Coaching

The outcome of these creativity methods as well as the integration of the co-creation approaches can be increased by the effort of a coach (Albers et al., 2016). Within these methods and furthermore in the overall innovation process the different abilities as well as the intrinsic and extrinsic motivations of the actors have to be considered (Folkerts, 2001). The success of a development project therefore depends on the actors and their individual factors. Accordingly, these aspects should be in the focus during the development of innovation processes and methods (Albers et al. 2013 und Albers et al. 2015b). Especially because the human factor increases in importance as projects increase in complexity, risk and innovativeness (Lechler, 1997). This leads to the need of more emphasis on organizational and human aspects in development projects, which could be complied by coaching (Berg/Carlsen 2007). This should provide organisational and communicational support and enhanced strategic thinking across the whole process (Stern, 2004). The idea behind coaching is to develop the capabilities of a high-performing team in order to enable it to succeed in its project (Berg/Carlsen 2007). The problem within the coaching term is that there is no consistent definition and hence a variety of definitions for specific areas of application exist. That is why there is an ongoing development or change of coaching activities and responsibilities (Boening, 2005). But although the concept of coaching is still seeking for a definition, it is gaining legitimacy as a business tool for teams (Coutu/Kauffman, 2009).

3. Research objective, research methodology and research environment

As shown in the State of the Art, a large number of collaboration tools and innovation methods are existing. Furthermore coaching is used in a variety of ways in the business area. Therefore, a long-term research objective is to establish a consistent approach to innovation coaching for agile innovation processes in distributed product development. In order to reach this overarching goal a corresponding set of hypotheses shall guide the future research projects. As a first step, this contribution focuses on the following research questions:

- What are core activities and responsibilities for innovation coaching in foresight processes for distributed product development?
- Which problems and unused potential do exist in this field?

To explore these research questions, two empirical studies with 42 and 64 participants are conducted. As a research environment the Live-Lab *ProVIL – Product Development in a Virtual Idea Laboratory* is used. In this, teams of master students in mechanical engineering are working in a three-month product development project. They are accompanied by innovation coaches, which are master students in the fields of business administration and engineering as well as international management. The project is structured according to ASD – Agile Systems Design into the analysis phase, the foresight phase, the conception phase and the specification phase (Albers et al., 2017). During the project all teams co-create with changing international project partners (Albers et al., 2016a). In doing so they partly collaborate on an innovation platform (Walter, 2016a) e.g. for the application of stimulus based creativity methods (Walter et al., 2017b) and the evaluation of product concepts.

In order to answer the research questions, an innovation coaching concept was developed and implemented into ProVIL in summer semester 2016 (with project partner Dr. Ing. h. c. F. Porsche AG) and 2017 (with project partner Fiat Chrysler Automobiles N.V.). Core element of this concepts is the description of the roles and responsibilities of the mechanical engineering students and the innovation coaches. The mechanical engineering students mainly concentrate on the project work, the deliverables of every project phase and the interaction with the project partner. In contrast to this, the innovation coaches focus on process coaching, moderation of physical and virtual meetings, recognition of social, personal or technical barriers and the valuation of degree of maturity of intermediate results.

In particular, the results presented in this paper were achieved accompanying eight students' teams in 2017 which developed products and concepts in the field of mobility solutions for future sharing economies:

- Case 1: A mobile unit which charges e-cars
- Case 2: A product which automatically registers the cleanliness status of carsharing cars
- Case 3: A product to provide a germ-free car to each carsharing user using a disinfection unit.
- Case 4: An app which combines ride sharing and hitchhiking
- Case 5: An app which synchronizes settings of the infotainment system over different car-sharing cars
- Case 6: A carsharing franchise concept
- Case 7: A cart for elderly to carry groceries by car without having to bend down at all
- Case 8: A rentable trailer which enables cyclist to carry more groceries spontaneously

The evaluation of the innovation coaching concept was conducted with an extensive data collection. During the entire project there were weekly online project surveys with specific questions for the ProVIL participants, the Innovation Coaches and further project participants. In addition to this, the Innovation Coaches composed individual reports about their role during the project.

4. Results

Core Activities and Responsibilities

Within the empirical studies, all ProVIL participants were asked about the activities and responsibilities of a perfect innovation coach. Based on these studies the core activities and

responsibilities for innovation coaching in foresight processes of a Live-Lab for distributed product development are presented in figure 1. They are established in the areas of process organisation, communication, team-enablement and the introduction of a multidisciplinary know how.

The responsibility of the process organization includes a close interaction with the project management as well as a good understanding of the innovation process and its phases. The knowledge about the procedure and goals of the different ASD-phases allows the innovation coaches to coordinate meetings and workflows of their team to effectively reach the respective deliverables. Especially the role as the interface between the development team and the project management is important in distributed development projects to ensure a transparent documentation of the current project progress for the management.

Furthermore, innovation coaching improves the outcome of foresight processes through a professional communication management, which includes guidance during research activities and the moderation of virtual team meetings and creativity sessions. The selection and set up of an appropriate ICT to foster productive meetings and sessions, depending on the team composition and tasks, is a core competency of innovation coaching. To the competency additionally belongs the provision of feedback as well as critical questions on ideas and (milestone-) presentations. As an example, the implementation of quality checks in the sense of an evaluation of results regarding the maturity level is particularly helpful.

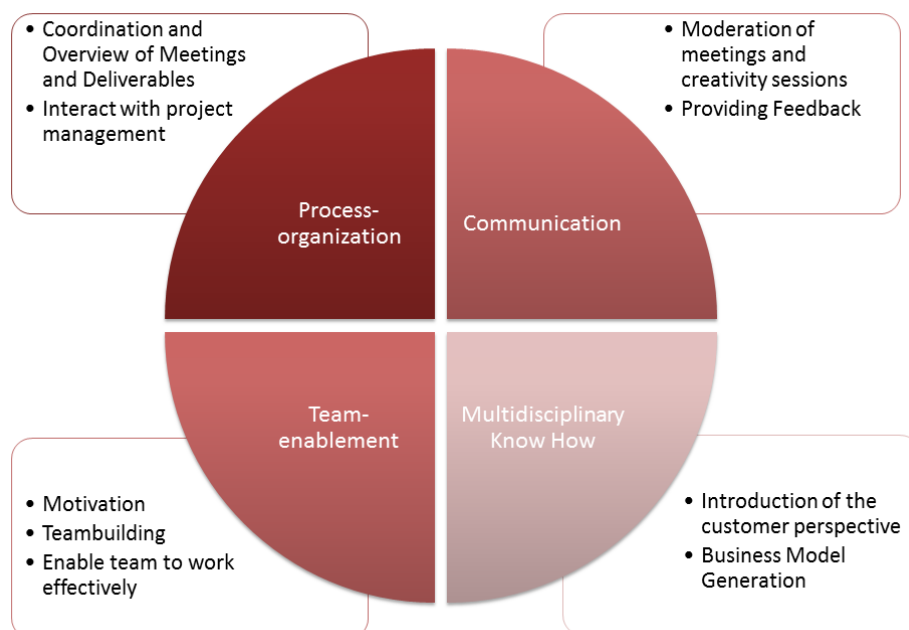


Figure 1: Core activities and responsibilities for innovation coaching

Moreover the teams need to be enabled to work effectively through motivation and teambuilding. Depending on the team's situation, which needs to be recognized through the corresponding work atmosphere, body language or short weekly surveys, it is important that innovation coaches are able to apply a set of possible action to encourage the teamwork. Especially in teams where all members are on the same hierarchy level, coaching increases the efficiency of arrangements concerning development tasks within the team.

Due to the interdisciplinarity of the innovation coaches in ProVIL they additionally focused on the human desirability of the ideas by adopting the customer perspective. Next to this, they ensured the economic viability by the generation of business models for the product profiles in the early stages of product development. With these activities, the innovation coaches enable the teams to get a clear

understanding of future requirements and associated challenges in order to derive innovation impulses and develop promising invention respectively product generations.

Problems and unused Potential

A problem in innovation coaching is the lack of distinction between the area of responsibilities and tasks. This occurs if innovation coaches begin to actively engage in the results beyond the representation of the customer perspective or the business model generation as intended within the framework of the feedback process. This can lead to misunderstandings or even drawbacks within the team and thus constrain the development process. Therefore the activities of innovation coaching and the integration into the innovation process need to be clarified in a systematic manner. The issue is the clear demarcation between coaching and self-development, consulting and leadership (see figure 2). In the product development this challenge increases in consideration of all ASD-phases in the innovation process. Moreover the question of how to successfully integrate innovation coaching respectively innovation coaches in the development teams. Within the ProVIL projects a need for an integration as part of the team was highlighted, to ensure a consistent and intensive knowledge sharing to enable a proactive and effective collaboration. Findings of the surveys showed that the presence or constant availability as well as the intensive cooperation with the team are important characteristics.

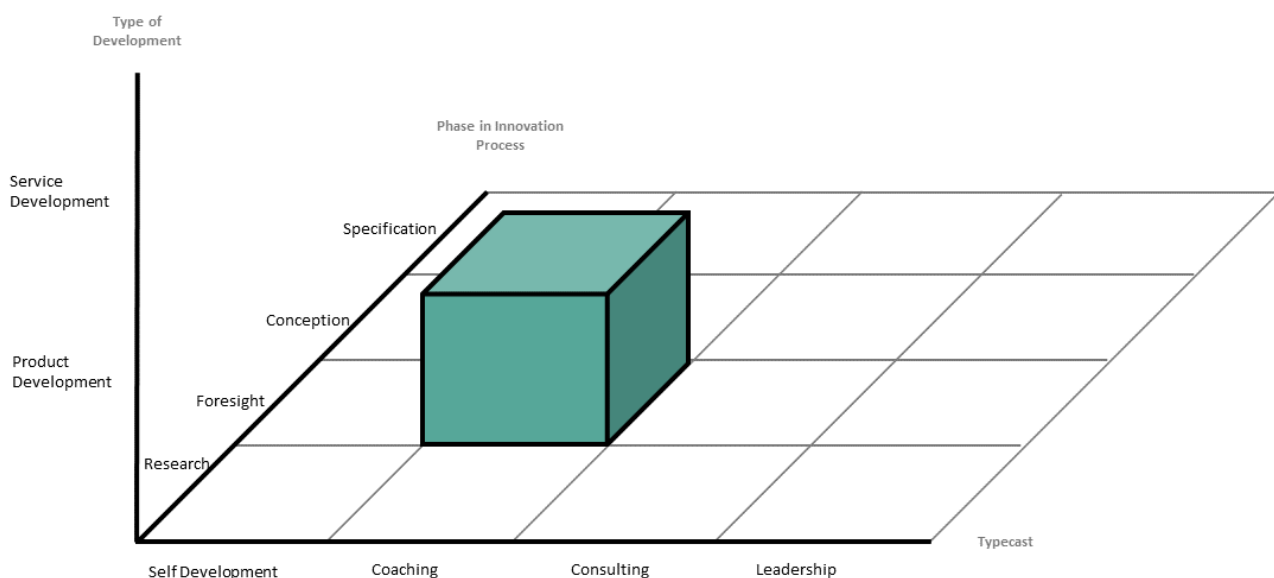


Figure 2: Lack of distinction between the area of responsibilities and tasks

Furthermore, this leads toward the unused potential of filling the gap between the business- and the engineering know-how. The empirical studies showed, that the integration of business model generation and the focus on the human desirability (for example through the persona method) strongly improved the outcome of the foresight phase. Therefore the innovation coaching could be improved with interdisciplinary innovation coaches, which focus on the business view as well as on the self-responsible implementation of methods of PGE - product generation engineering like the estimation of development risks or the definition of the design space in relation to the next product generation.

Derived Hypotheses

Based on the knowledge about the core activities and responsibilities as well as the problems and unused potentials of innovation coaching in foresight processes for distributed product development in the Live-Lab ProVIL in addition to the practical experience through organizing and managing the Live-Lab we put up a set of hypotheses to guide future research projects. The research hypotheses are as follows:

Hypothesis 1: *A clear, transparent role description for innovation coaching in a Live-Lab enables a good collaboration with the development team.*

In ProVIL it was empirically determined that the lack of transparency in the tasks and responsibilities of innovation coaching led to misunderstandings and to a lack of acceptance. Due to the ambiguous hierarchy in the beginning, the innovation coaches and the development teams did not know about the different responsibilities which negatively affected the collaboration as well as the teambuilding. This was solved with a clearly defined description of the responsibilities and the demonstration of interactions within the project work. Since these areas of responsibility can change in the different ASD-phases of the innovation process, there is a demand for constant transparency.

Hypothesis 2: *Innovation Coaching requires a clear demarcation from project management, project consulting and team coaching.*

As described within the problems of innovation coaching in ProVIL, the aim of innovation coaching is to enable the development teams to achieve the given goals efficiently, to operationalize them and to wisely use the scope of action. This can be accomplished by performing the above explained core tasks. The allocation of tasks needs to be communicated in a precise structure to all participants.

Hypothesis 3: *There is a demand for innovation coaching, which is integrated directly into the development teams and therefore does not require any additional, external person.*

The results of the weekly surveys show, that the integration of the innovation coaches into the development teams is a very important aspect in regard to the team performance. In particular, a thorough, intensive exchange of knowledge enables proactive and efficient action by the innovation coaches. The presence or constant availability as well as the intensive cooperation with the team were mentioned as one of the most important characteristics in the surveys. Additionally this effects the long-term acceptance of an innovation coach. The weekly surveys in ProVIL showed, that the acceptance was highest before important deadlines or milestones. This was due to the many feedback or coordination iterations between the innovation coaches and development team which displayed the added value to the team.

Hypothesis 4: *The long-term cooperation between innovation coaches and development team promotes successful innovation coaching.*

The long-term support of the innovation coaches increases the trust relationship with the development team as well as its acceptance. Added to this is the knowledge about characteristics and expertise of the individual team members, which leads to an efficient assignment of tasks and organization in the team. The long-term cooperation makes it possible to achieve the so-called performance phase of the entire team, according to Tuckman (Tuckman, 1965).

Hypothesis 5: *Innovation coaching allows an organizational ambidexterity, which means the ability to be efficient and flexible at the same time.*

Applying innovation coaching in foresight processes for distributed product development enables to be efficient in the management of the innovation process and still be adaptable for coping future trends or challenges. Due to the close interaction with the team, the innovation coach manages to focus on the given process but still enables the team to act flexible. An example is that they constantly integrate the fast changing customer requirements. Innovation coaching brings in methodical knowledge which allows to integrate approaches of exploitation and exploration into the development process. This is realized with guidance to an intensive research about the challenge, its framework and customers to exploit the existing. Furthermore the support to generate future scenarios, analyze survey results and conduct creativity methods in order to generate product profiles, leads to the exploration of new knowledge.

Hypothesis 6: *Innovation coaching enables the establishment of a structure for decision-making in which the desired transparency and participation in (corporate) networks are realized for all levels.*

The close interaction between project management and innovation coaches, which are integrated into the individual development teams, allows a decentralized management, so called holacracy. This ensures a transparent documentation of the current project progress for the management as well as for the development teams and enables the participation of each actor. The resulting value of a well-structured network especially profits in new approaches for managing innovation processes like agile development, open innovation and co-creation.

Hypothesis 7: *A (data-based) warning system, e.g. a project monitoring for the team performance and motivation, helps the innovation coach to recognize the need for interventions in time.*

Empirical results showed, that innovation coaches can compensate motivation fluctuations in the team by applying situation-related action. For example in very work intensive phases, such as analysing the problem and generating future scenarios, the innovation coaches kept the teams motivation high by offering individual support and organizing joint events during breaks to strengthen the teambuilding. Therefore a set of intervention strategies for critical situation is needed, which implements a core competency of an innovation coach. In order to recognize the need for interventions in time, an innovation coach, especially in the case of distributed teams, needs a corresponding, data-based warning system, e.g. a project monitoring or barometer. This helps to proactively take action for a highly motivated and well performing team.

5. Conclusion and Outlook

Empirical studies within the Live-Lab ProVIL indicate that the core activities and responsibilities for innovation coaching in foresight processes for distributed product development are established in the areas of process organisation, communication, team-enablement and the introduction of a multidisciplinary know how. Problematic is the lack of distinction between the area of responsibilities and tasks of innovation coaching. The issue is the clear demarcation between coaching and self-development, consulting and leadership which needs to be clarified in a systematic manner to successfully use the potential of innovation coaching.

The showed results of this paper about innovation coaching in foresight processes for distributed product development shall guide to the long-term research objective which is to establish a consistent

approach to innovation coaching for agile innovation processes in distributed product development. Therefore the set of hypotheses will be validated through qualitative and quantitative field studies as well as practical research in the Live-Lab ProVIL.

Accompanying research applied the approach of innovation coaching for the area of decision preparation and making in strategic transport infrastructure planning. In terms of the economy, transport infrastructure plays an important economic role in the task of ensuring the basic conditions for adequate and reliable mobility systems. In comparison to business-related innovation processes, strategic infrastructure planning is characterized by a large number of different stakeholders with strongly contradictory objectives in most cases. Planning processes run for very long time periods, during which social and legal framework conditions generally change. Considering technical tasks, a wide variety of expertise must be taken into account in a complex coordination process. Ensuring a transport infrastructure with high-availability is therefore caught between technical, economic, environmental and social needs. Current project examples show that today's methods and processes are no longer able to cover future demands on strategic decision-making processes. For this reason, various projects assessed whether the use of innovation coaching in such strategic planning processes is successfully applicable. First attempts in the BMBF project "Foresight Innovation Communities" as well as in two planning projects basically show an application potential. In addition to the research activities in the Live-Lab ProVIL, it is therefore planned to launch a second project, "Innovation Communities for Strategic Decision Making", in order to demonstrate the potential of the approach in the complex area of strategic transport infrastructure development.

References

- Albers, A., Bursac, N., Heimicke, J., Walter, B., Reiß, N. (2017). *20 years of co-creation using case based learning. An integrated approach for teaching innovation and research in Product Generation Engineering*.
- Albers, A., Bursac, N., Walter, B., Hahn, C., Schröder, J. (2016). "ProVIL – Produktentwicklung im virtuellen Ideenlabor". In: *Entwerfen Entwickeln Erleben– Methoden und Werkzeuge in Produktentwicklung und Design 2016*.
- Albers, A., Bursac, N., Wintergerst, E. (2015a). "Product Generation Development – Importance and Challenges from a Design Research Perspective". In: *Proceedings of INASE Conferences 2015*. INASE. Wien, Austria.
- Albers, A., Maul, L., Bursac, N. (2013). "Internal Innovation Communities from a User's Perspective. How to Foster Motivation for Participation". In: M. Abramovici, R. Stark (Eds.): *Smart Product Engineering. Proceedings of the 23rd CIRP Design Conference*: Springer (Lecture Notes in Production Engineering, 5), pp. 525–534.
- Albers, A., Maul, L., Bursac, N., Heismann, R. (2015b). "Connected Creativity – A Human Centered Community Platform for Innovation Impulses". In: DS79: *Proceedings of The Third International Conference on Design Creativity*, Indian Institute of Science, Bangalore. 2015.
- Alzoubi, Y. I., Gill, A. Q., Al-Ani, A. (2016). *Empirical studies of geographically distributed agile development communication challenges: A systematic review*. In Information and Management, Jan. 2016
- Berg, M. E., Karlsen, J. T. (2007). "Mental Models in Project Management Coaching". In: *Engineering Management Journal*, 19(3), pp. 3-13.
- Bessant, J., (2003). "Challenges in Innovation Management". In: Shavinina, L. (Ed.): *The International Handbook on Innovation*. Kidlington, Oxford: Elsevier Science Ltd.
- Boenning, U. (2015). "Coaching: Der Siegeszug eines Personalentwicklungs-Instruments. Eine 15-Jahres-Bilanz". In: Rauen, C. (Hrsg.). *Handbuch Coaching*. Göttingen: Hogrefe Verlag, pp. 21-54.
- Bullinger, H.-J., Wörner K., Prieto J. (1998). *Wissensmanagement – Modelle und Strategien für die Praxis*. In B.H.D., Wissensmanagement – Schritte zum intelligenten Unternehmen, pp.. 21-39.
- Chesbrough, H. (2006). "Open Innovation: A New Paradigm for Understanding Industrial Innovation". In: Chesbrough, H., Vanhaverbeke, W., West, J. (2006). *Open Innovation. Researching a New Paradigm*. Oxford University Press, pp.1-15.
- Coutu, D., Kauffman, C. (2009). "What Can Coaches Do for You?". In: *Harvard Business Review*, January 2009 Issue.

- Eppinger, S., Chitkara, A. (2006). "The New Practice of Global Product Development". In: *MIT Sloan Management Review*.
- Fichter, K., Beucker, S. (2012). *Innovation Communities. Teamworking of Key Persons – A Success Factor in Radical Innovation*. Heidelberg: Springer Verlag.
- Folkerts, L. (2001). *Promotoren in Innovationsprozessen: empirische Untersuchung zur personellen Dynamik*. Wiesbaden: Deutscher Universitäts-Verlag.
- Haines, R. (2014). "Group development in virtual teams: An experimental re-examination". In: *Computers in Human Behavior* 39, 2014.
- Kotlarsky, J., Oshri I., (2005). "Social ties, knowledge sharing and successful collaboration in globally distributed system development projects". In: *European Journal of Information Systems* 14, 2005.
- Layzell, P., Brereton, O.P., French, A. (2000). *Supporting Collaboration in Distributed Software Engineering Teams*.
- Lechler, T. (1997). *Erfolgsfaktoren des Projektmanagements*. Frankfurt am Main: Lang.
- Stern, L.R., (2004). "Executive Coaching: A Working Definition". In: *Consulting Psychology Journal: Practice and Research*. Issue 56 (3), pp. 154-162.
- Tuckman, B.W., (1965). "Developmental sequence in small groups". In: *Psychological Bulletin*. Issue 63, pp. 384–399.
- Walter, B., Albers, A., Haupt, F., & Bursac, N. (2016b). "Produktentwicklung im virtuellen Ideenlabor–Konzipierung und Implementierung eines Live-Lab". In: *DFX 2016: Proceedings of the 27th Symposium Design for X*. 5-6 October 2016, Jesteburg, Germany.
- Walter, B., Albers, A., Heck, M., Bursac, N. (2016a). "ProVIL – Produktentwicklung im virtuellen Ideenlabor: Anpassung einer kollaborativen Innovationsplattform zur Realisierung eines communityorientierten Innovationsprozesses". In: *12th Symposium für Vorausschau und Technologieplanung*, Paderborn, Germany.
- Walter, B., Klippert, M., Kunz, M., Albers, A., Reiß, N. (2017b). *Kreativitätsmethoden im digitalen Umfeld – Aktive Förderung von Innovationsimpulsen in standortverteilten Entwicklungsteams*.
- Walter, B., Wilmsen, M., Albers, A., Bursac, N. (2017a). "Zukunftsmanagement in Zeiten der Digitalisierung: Die Szenario-Technik als Innovationsmethode in der standortverteilten Produktentwicklung". In: *13. Symposium für Vorausschau und Technologieplanung*, Berlin, Germany.

Real-World Laboratory: Promoting Pedestrians

E. Haeussler

Institute of Traffic Planning and Infrastructure Management
Hochschule Karlsruhe – University of Applied Sciences
e-mail: elke.haeussler@hs-karlsruhe.de

Keywords: walkability, participation, experiments

GO Karlsruhe is a real-world laboratory (RwL) in which issues and their possible solutions are investigated and scientifically evaluated in the real world. The overarching project aim is to promote walking. Another objective is the comparison between offline (face-to-face) and online (digital) citizen participation. The RwL follows the concept of participatory research by including citizens in the development of the research agenda, conduction of the research, and the implementation of the research findings. Assisted by the research team pedestrians in Karlsruhe identify measures to improve walkability and develop solutions for existing problems in the city. As part of the research, real-world experiments are conducted in order to identify the drivers which promote walking. In real-world experiments, previously developed solutions are implemented and it is subsequently analyzed whether the desired improvement has been achieved and which aspects (still) remain to be considered. GO Karlsruhe was launched in November 2015 and is due to run until October 2018. The research project is funded by the Federal State of Baden-Württemberg.

Why walkability?

The promotion of walking takes a key role in the transformation of our cities towards sustainable mobility and social inclusion (Southworth, 2005). Cars dominated urban spaces for decades, but over the past years municipalities in Germany undertake great efforts to promote biking and public transport by building bike lanes and by the expansion of public transport systems. Now walking is seen as an attractive alternative in urban spaces. In Germany today's situation is best described by the term *renaissance of walking* (Gehl/Svarre, 2013). There are compelling reasons to promote walking as the improvement of the environment for pedestrians goes along with the reduction of exhaust gases and thus air pollution, less traffic noise, and more space for trees, green spaces and benches. Regular walking has been shown to reduce blood pressure, diabetes, obesity, and mental stress as well as the overall level of traffic danger (Dittrich-Wesbuer/Erl, 2004; ITF, 2012; Southworth, 2005). Moreover, there are benefits for the local economy given that pedestrians are loyal customers (ITF, 2012). However, providing more space for pedestrians is nowhere near enough.

Research Approach

So far preferences of pedestrians have not been comprehensively investigated due to the lack of data bases or survey instruments. Through participatory research and a targeted use of digital tools, socially robust knowledge will be generated in the RwL (Wagner/Grunwald, 2015). The project should provide answers to the following research questions: Why do people choose to walk and what are the deciding parameters? What are the appropriate forms of participation for pedestrians depending on the area? Which groups are reached through face-to-face and which groups through digital participation respectively? It is important to understand how walking works as a system, therefore the RwL pursues a holistic approach. The expected findings of the survey cover a wide

range, including an evaluation of attributes of the pedestrian network such as surface type and width of pedestrian sidewalks as well as the social framework that has to be established to make walking more popular.

Participation on a new level

Conventional face-to-face participation usually addresses people living in a specified district where an infrastructure measure is planned. However, in the city center and other heavily-attended places users of the traffic infrastructure are rarely residents. Users are tourists, employees, and customers who live outside the city center or even outside the city. Conventional citizens' participation methodology therefore has inadequacies. With the help of digital participation instruments such as interactive posters and smartphones pedestrians are involved during the problem analysis, the development of measures, the decision on implementation and the assessment of the measures. By using the Android app GO Karlsruhe pedestrians have the comfortable opportunity to give spontaneous feedback on the way or from home. The app enables users to report what they perceive as good or bad on their everyday paths. In doing so, they actively improve the walkability in their city.

Interactive posters, which are currently being developed, will involve pedestrians in Karlsruhe on walkways, public transport stops, crossroads, or traffic lights. This low-threshold approach is used to reach as many people as possible.

On- and Offline Comparison

Criteria have been developed in order to enable a comparison between user (online) and citizen (face-to-face) participation. These criteria apply to both digital and conventional participation. A suitable approach for qualitative comparison are so-called *micro scenarios* (Postma et al., 2004). To ensure the high quality of the engagement process they were adapted to the conditions of participatory research based on existing work in stakeholder engagement. These micro scenarios can also help to find out which approaches of participatory process are more suitable in terms of target groups, areas and problems.

Stakeholder Analysis: who is in and why?

The RwL draws on the knowledge and expertise of local stakeholders (Moura et al., 2017). To develop an overview of the potential stakeholders and to select the stakeholders relevant for the conventional participation process a stakeholder analysis was conducted (Aichholzer et al., 2016). It summarizes the main interests and concerns of the identified key stakeholders and provides lessons learned from previous engagement processes. The results from the stakeholder analysis are used to tailor the engagement to the identified needs. Actors were identified, grouped into categories (key stakeholders, veto-player e.g), and a profile was created for each stakeholder. In addition a stakeholder map was drawn up for each district, which provides an overview on the relation between the stakeholders and helps to recognize conflicts at an early stage.

First results and Outlook

As mentioned before, the stakeholder analysis helps to understand who is participating in the RwL and reveals that only certain citizens are actually participate. There is a self-selection of participants so in GO Karlsruhe events mainly older people are found. This is a well-known problem. Economically powerful groups and the educated middle class are often overrepresented in stakeholder engagement processes (Aichholzer et al., 2016). Young people, parents of small children, poorer groups, migrants etc. are usually missing (Baek et al., 2012). The RwL tries to reach those groups that are usually underrepresented in conventional participation processes by digital participation. The approach is to be transparent about the participation bias so that this could be taken into account when interpreting the results.

As expected, activating citizens has proven to be the most difficult task so far. This is true for both forms of participation but with different emphasis. Although almost everyone is affected by the absence of walkability the public perception is relatively low. Publicity activities are necessary in the further course of the RwL.

The on- and offline identification of possible places for real-world experiments will be completed in the foreseeable future. The next step will be the development of the experiments followed by their implementation. New challenges are expected such as resistance due to safety considerations or removal of parking spaces.

References

- Aichholzer, G., Kubicek, H. and Torres, L. (2016). *Evaluating e-Participation. Frameworks, Practice, Evidence*, Springer International Publishing, Cham.
- Baek, Y. M., Wojcieszak, M. and Delli Carpini, M. X. (2012). "Online versus face-to-face deliberation. Who? Why? What? With what effects," in: *New Media & Society* vol.14, no.3, pp. 363–383.
- Dittrich-Wesbuer, A. and Erl, E. (2004). "Zu Fuß unterwegs. Wissenswertes und Wünschenswertes zu einer unterschätzten Verkehrsart," in: *Handbuch der kommunalen Verkehrsplanung*.
- Gehl, J., and Svarre, B. (2013). *How to study public life*, Island Press, Wasington, DC.
- ITF (2012). *Pedestrian Safety, Urban Space and Health*. OECD Publishing, Paris. Available at <http://www.oecd-ilibrary.org/docserver/download/7511011e.pdf?expires=1495202981&id=id&accname=ocid56023212&checksum=42DF6DFA241BF7CC8B0E8924B3933669> [19.05.2017].
- Moura, F., Cambra, P. and Gonçalves, A. B. (2017). "Measuring walkability for distinct pedestrian groups with a participatory assessment method. A case study in Lisbon," in: *Landscape and Urban Planning* 157, pp. 282–296.
- Postma, L., James, A. J. and van Wijk, W. (2004). "QIS: A New Participatory Management Tool to Assess and Act in Field Reality," Vientiane, Laos (30th WEDC International Conference). Available at <https://www.ircwash.org/sites/default/files/Postma-2004-QIS.pdf> [19.05.2017].
- Southworth, M. (2005). "Designing the Walkable City," in: *J. Urban Plann. Dev.* 131 (4), pp. 246–257.
- Wagner, F. and Grunwald, A. (2015). "Reallabore als Forschungs- und Transformationsinstrument. Die Quadratur des hermeneutischen Zirkels," in: *GALA - Ecological Perspectives for Science and Society* 24 (1), pp. 26-31.

Chapter 9

Civil Engineering

The Successful Integration of Health and Safety in the Engineering Curriculum of the Graduate Engineer

M Moloney, D Walsh
Department of Civil, Structural & Environmental Engineering
CIT – Cork Institute of Technology, Ireland
e-mail: mary.moloney@cit.ie

Keywords: Project-based Learning; engineering education ; curriculum design

Introduction

Health and safety (H&S) is one of the key ethical and competency requirements for the professional practice of the engineer (Godfrey and King 2011). In line with the Bologna process and international trends, the Department of Civil, Structural and Environmental Engineering (DCS&EE), at Cork Institute of Technology (CIT) sought to infuse H&S and risk management within the Level 7 & 8 engineering curricula it delivers, with a focus on the ‘student-centred and outcome-centred approach’ (Kennedy 2007). However, the Department went further, asking: can H&S education be successfully converted into a Health & Safety Culture in engineering Students? This paper reports on the methodology adopted to address this research question, and the successes achieved, with examples provided where appropriate.

The influences - Legal responsibly and the accrediting Professional bodies

There is considerable legislation covering the practice of professionals in the built environment, most of which emanates from the Health and Safety at work Act 2005 (Irish Government 2005) and regulations applicable to the built environment (Irish Government 2016, 2013). Whether practicing in design, construction, maintenance, operation or decommissioning of facilities, all practicing engineers must adhere to such legal requirements.

Thus the two key influences in the department’s H&S curriculum development process were: the legal responsibilities that the students would have upon graduation when practicing as Engineers; and the requirements of the accrediting body for engineering degrees in Ireland, Engineers Ireland (EI). They require that Students can ‘*Demonstrate ... knowledge and understanding of the **health, safety and legal issues** and responsibilities in your engineering practice*’ (Engineers Ireland 2014).

Compliance with EI accreditation criteria in turn enables compliance with the EUR-ACE system and international trends (Godfrey and King 2011, ENAAE (European Network for Engineering Accreditation) March 2015).

The challenge and methodology adopted – successfully develop a H&S culture

However, the Department took the view that the importance of H&S merited a deeper learning than traditional teaching techniques alone could achieve. In particular, it considered that merely educating students in the relevant subjects through traditional lectures and assignments would be insufficient to achieve the desired outcome of embedding H&S principles in the ‘everyday’ mindset of the students; in other words, to form a H&S culture within the student body. Therefore, an innovative approach was adopted to address the research question of ‘if and how’ a H&S culture could be developed and

integrated successfully, by including risk management procedures as core operating principles within the engineering department, across all student cohorts.

The Department's vision led to the following 3-pronged process of curriculum development as illustrated in Figure 1: compliance with and continued development of the departmental H&S management system; the development and introduction of a new mandatory H&S module at undergraduate level in 2012; and the integration of applied H&S within subsequent modules.

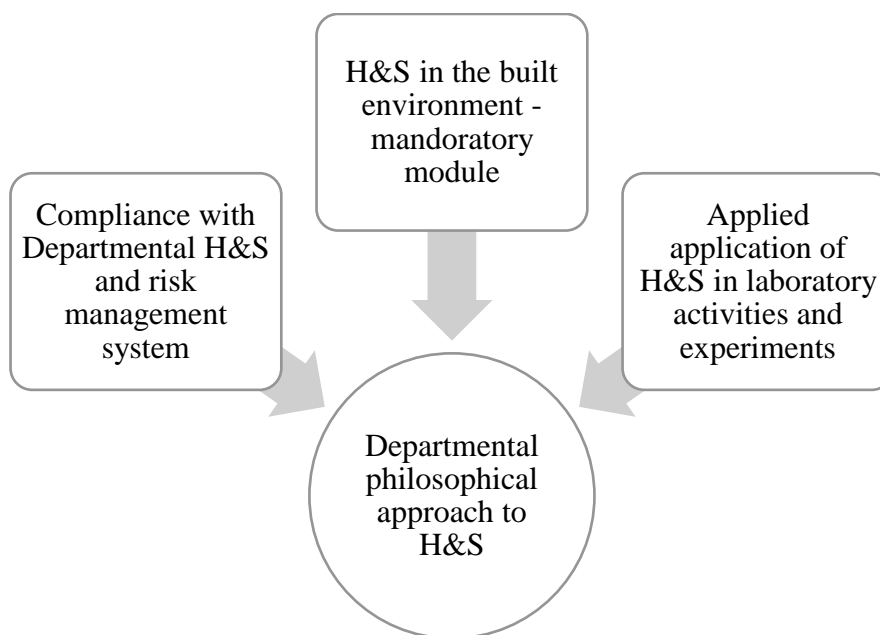


Figure 1: Departmental approach to the integration of H&S within the curriculum of the graduate engineer

The novel aspect of the adopted approach included the shift away of theoretical assignments, to real life engineering challenges; the integration of H&S into their working practice within the department; to ultimately students using the principles of risk management and H&S to design their own research experiments in final year.

For the Student research projects, no laboratory or field work is permitted without Students first designing and documenting their plans; carrying out a risk assessment, and safe work plan. This embeds the theoretical learning from lectures and assignments; and ensures the critical nature of of these stages of the project development, worked through at the outset, become the 'norm' in their approach to design and construction projects.

An overview of each of the elements is discussed in the following sections.

1. Element 1: Compliance with Departmental H&S management system

Under national health and safety legislation, the Department has a legal responsibility and duty of care to all persons working and studying within, or visiting the Department. Thus the department has a H&S management system which is reviewed annually, with an online document system for staff and students.

2. Element 2: Health and Safety in the Built Environment module

In keeping with engineering education in general, the department takes the view that it is vital for students to obtain a firm grasp of the fundamentals of H&S. Consequently, a specific module dedicated to such H&S fundamentals and their application in construction technology was introduced in 2012.

The pedagogical approach to the module was influenced by the legalistic nature of the content, and the very important ethical responsibilities of the future graduate engineer. The department recognised the shift in focus of engineering education from the theory-based practice of a 'chalk and talk' pedagogy, to a more applied delivery and assessment of the student based on project assisted and project based learning activities (Mills and Treagust 2003). In addition, professional and accrediting bodies are calling for the development of the more human skills in the engineers' formation; hence a paradigm shift towards active learning pedagogies, including problem-based learning and project-based learning, is increasingly evident (Mills and Treagust 2003, Fernandes 2014).

The module is therefore based on continuous assessment with formative feedback. For example, one of the project assignments is based on a real site, which the student visits, requiring detailed note taking and hazard identification of that particular site; risk assessment; and identification of the preferred task-appropriate construction methodologies. Students are thus forced to consider the H&S and risk management issues of construction first, and not as the last item on site.

3. Element 3: Applied application of H&S systems within the Departmental activities

Combining the use of H&S policy processes that are in place within the department; and the mandatory H&S introductory module; students are equipped with the necessary scaffolding to support their further learning in the department and the application of H&S. The innovative aspect of the department approach is the infusion of various aspects of H&S and risk management within subsequent modules in supervised environments e.g.: soils laboratory classes require students to prepare a method statement and risk assessment of the experiment; students on surveying activities undergo an induction and are required to identify the risks and controls of the particular area they plan to survey.

Final year students are required to deepen this H&S application through independent design of research experiments by assessing and mitigating the H&S risks; while the design office module requires students to prepare a design risk assessment. All these steps are in line with practicing engineers.

Validation of the department's approach - Conclusions

Ultimately the validation of the department's approach is judged by the accrediting professional bodies, employers and state agencies. Students of the department have successfully participated in national competitions organised by the Health and Safety Authority of Ireland (the state agency with responsibility of health and safety in the workplace), achieving second place in the 2015 competition, and national winners in 2016 (Health and Safety Authority, 2016). The department offers student work placement opportunities. Many of these employers have recognised the student's readiness to take on H&S tasks in both the design office and construction sites, over the past 5 years since the infusion of this H&S culture.

Such success on the national stage, along with positive employer feedback, is taken as a mark of success in the approach the department has developed, and will continue to expand.

References

ENAAE (European Network for Engineering Accrediation) (2015). "EUR-ACE Framework standards and guidelines."

Engineers Ireland (2014). "Regulations for the Registered professional title of Chartered Engineer."

Fernandes, S. R. G. (2014). "Preparing Graduates for Professional Practice: Findings from a Case Study of Project-based Learning (PBL)." In: *Procedia - Social and Behavioral Sciences* 139:219-226. doi: <http://dx.doi.org/10.1016/j.sbspro.2014.08.064>.

Godfrey, E. and King, R. (2011). "Curriculum specification and support for engineering education: understanding attrition, academic support, revised competencies, pathways and access." *Strawberry Hills NSW: Australian Learning & Teaching Council*.
Irish Government. 2005. Health and Safety at Work Act.

Irish Government. 2013. SAFETY, HEALTH AND WELFARE AT WORK (CONSTRUCTION) REGULATIONS 2013. In *S.I. No. 291 of 2013*. Dublin.

Irish Government. 2016. SAFETY, HEALTH AND WELFARE AT WORK (GENERAL APPLICATION) (AMENDMENT) (NO. 3) REGULATIONS 2016. In *S.I. No. 370 of 2016*. Dublin.

Kennedy, Declan. 2007. *Writing and Using learning outcomes A practical guide*. Cork: UCC.

Mills, Julie, and David Treagust. 2003. *Engineering Education, Is Problem-Based or Project-Based Learning the Answer*. Vol. 3.

Conference Awards

Best Presentation Award

presented to :

Fuchs K., Henning P.

Operationalizing Learning Behavior as Spatio-Temporal Trajectories

Most Innovative Paper Award

presented to :

Bohus, Deuschel, Gimbel, Görg, Humm, Schüller, Turan

Predicting Premature Termination of Treatment in Psychotherapy for Borderline Personality Disorder

Best Paper Award - Traffic and Transportation

presented to :

Gamboa

Dealing with the Rise of Transportation Network Vehicle Service: Rethinking the Regulatory Regime for TNVS towards Sustainable Urban Mobility