

**Martin-Luther-Universität Halle-Wittenberg**  
Juristische und Wirtschaftswissenschaftliche Fakultät  
Wirtschaftswissenschaftlicher Bereich

**Beiträge zum 22. Interuniversitären  
Doktorandenseminar Wirtschaftsinformatik  
(IUDS 2018)**

Stefan Sackmann, Stephan Kühnel, Johannes Damarowsky



Beitrag Nr. 35

**Diskussionsbeiträge zu  
Wirtschaftsinformatik und Operations Research**

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

**Sackmann, Stefan; Kühnel, Stephan; Damarowsky, Johannes:**

Beiträge zum 22. Interuniversitären Doktorandenseminar Wirtschaftsinformatik / Stefan Sackmann, Stephan Kühnel, Johannes Damarowsky

Martin-Luther-Universität Halle-Wittenberg, Juristische und Wirtschaftswissenschaftliche Fakultät, Wirtschaftswissenschaftlicher Bereich.- Halle (Saale): Martin-Luther-Universität Halle-Wittenberg, 2018.

(Diskussionsbeiträge zu Wirtschaftsinformatik und Operations Research; Beitrag Nr. 35)

ISSN: 1611-6879

ISBN: 978-3-86829-977-9

**Schlagworte:** Wirtschaftsinformatik, Interuniversitäres Doktorandenseminar, Forschungsbeiträge

Korrespondenzanschrift:

**Prof. Dr. Stefan Sackmann**

Martin-Luther-Universität Halle-Wittenberg

Juristische und Wirtschaftswissenschaftliche Fakultät, Wirtschaftswissenschaftlicher Bereich

Institut für Wirtschaftsinformatik und Operations Research

06099 Halle (Saale)

Tel.: +49 345 55 23 471

Fax: +49 345 55 27 374

E-Mail: [im@wiwi.uni-halle.de](mailto:im@wiwi.uni-halle.de)

WWW: <https://informationsmanagement.wiwi.uni-halle.de>

Halle (Saale), 2018

# Inhaltsverzeichnis

- Seite III Stefan Sackmann, Stephan Kühnel, Johannes Damarowsky  
**Vorwort**
- Seite 1 Michael Aleithe, Universität Leipzig  
**Methode zur Entwicklung von Patienten-Monitoringsystemen**
- Seite 9 Sven Gehrke, Friedrich-Schiller-Universität Jena  
**Data Mining for Skills/ Qualifications in Social Networks**
- Seite 10 Tobias Goblirsch, Universität Leipzig  
**Urbanes Umweltmonitoring - Eine kollaborative und interdisziplinäre Herausforderung am Beispiel der individuellen thermischen Exposition**
- Seite 16 Alexandra Herzog, Universität Leipzig  
**Entwicklung eines Modells zur Abbildung von Transformationsprozessen von Servicesystemen am Beispiel von Energiemanagementsystemen**
- Seite 22 Christoph Jobst, Universität Leipzig  
**Qualitätssicherung elektronischer Prüfungsverfahren. Integration der Item-Response-Theorie zur Testauswertung in Lern-Management-Systeme und Abgleich mit der klassischen Testtheorie**
- Seite 32 Sandra Niemz, Friedrich-Schiller-Universität Jena  
**Prozessorganisation in ungeplanten Notfallsituationen mithilfe von Workflowmanagement**
- Seite 38 Tobias Seyffarth, Martin-Luther-Universität Halle-Wittenberg  
**Business Process Compliance and Change**
- Seite 40 Manuel Siekmann, Universität Leipzig  
**RTML – Eine Modellierungssprache zur Wissensentwicklung im Kontext klinischer Befundberichterstellung**
- Seite 48 Norman Spangenberg, Universität Leipzig  
**Development of a Decision Support System for Online Surgery Scheduling**
- Seite 58 Anja Tetzner, Technische Universität Chemnitz  
**Potentiale künstlicher neuronaler Netze zur Identifikation von Ironie im Rahmen der Sentiment-Analyse**



## Vorwort

Das Interuniversitäre Doktorandenseminar der Wirtschaftsinformatik kann inzwischen auf eine langjährige Tradition zurückblicken. Bereits zum 22. Mal finden sich Doktorandinnen und Doktoranden, Professorinnen und Professoren aus Mitteldeutschland zusammen um Promotionsvorhaben und aktuelle Forschungsergebnisse zu präsentieren und zu diskutieren.

Seitdem das erste Doktorandenseminar im Jahre 2000 von den Universitäten Leipzig und Halle-Wittenberg ins Leben gerufen wurde, erweiterte sich der Kreis seither um die Lehrstühle der Wirtschaftsinformatiker von der Technischen Universität Chemnitz, der Technischen Universität Bergakademie Freiberg, der Technischen Universität Dresden, der Friedrich-Schiller-Universität Jena und der Technischen Universität Ilmenau.

In diesem Jahr hat wieder die Martin-Luther-Universität Halle-Wittenberg die Ehre das Interuniversitäre Doktorandenseminar auszurichten. Zuletzt durften wir vor sechs Jahren, zum 16. IUDS im Jahre 2012, nach Halle an der Saale einladen. Mit insgesamt elf Beiträgen, einem Gastvortrag und knapp 50 Gästen verspricht das Programm den ganzen Tag über spannend zu werden. Dies nicht zuletzt durch die thematische Vielfalt der Einreichungen, welche die fachliche Breite der Wirtschaftsinformatik widerspiegelt. Vom Informationsmanagement über das Geschäftsprozessmanagement, E-Health und Operations Research bis hin zu Data Mining und Machine Learning erstrecken sich die diesjährigen Vorträge. Trotz dieser thematischen Breite finden sich aber auch deutliche Gemeinsamkeiten, welche spannende und konstruktive Diskussionen sowie einen intensiven Wissensaustausch über das Doktorandenseminar hinaus ermöglichen.

Wir bedanken uns sehr herzlich beim Institut für Unternehmensforschung und Unternehmensführung an der Martin-Luther-Universität Halle-Wittenberg e.V. für die Finanzierung des Caterings. Ebenfalls danken wir sehr herzlich allen Professorinnen und Professoren, die mit Ihrer reichhaltiger Expertise den Doktorandinnen und Doktoranden den Diskurs anbieten. Zuletzt gilt ein besonderer Dank natürlich auch den Vortragenden des Doktorandenseminars, die den wissenschaftlichen Diskurs erst mit Leben füllen.

Wir wünschen allen Teilnehmern ein erfolgreiches Seminar, spannende Vorträge sowie gute und konstruktive Diskussionen, auch außerhalb des Programms.

Stefan Sackmann  
Stephan Kühnel  
und Johannes Damarowsky

Halle im November 2018



# Methode zur Entwicklung von Patienten-Monitoringsystemen

Michael Aleithe  
Institut für Wirtschaftsinformatik  
Universität Leipzig

Leipzig, Deutschland  
aleithe@wifa.uni-leipzig.de

**Zusammenfassung**—Die Entwicklung von Patienten-Monitoringsystemen gewinnt in Anbetracht der Digitalisierung der Gesundheitsbranche zunehmend an Bedeutung. Nicht zuletzt geht dieser Trend einher mit dem rasanten Aufkommen des Internet of Things (IoT), wodurch eine Vielzahl neuer technischer Anwendungsmöglichkeiten gerade im Bereich des mobile health (mHealth) zu verzeichnen sind. Gleichzeitig erscheinen jedoch mit dieser Zunahme an technologischen Anwendungsmöglichkeiten neue Herausforderungen. Neben den technischen Aspekten spielt ebenso der Datenschutz eine zentrale Rolle, der auf keinen Fall vernachlässigt werden darf. Weiterhin müssen die medizinischen sowie technischen Anforderungen in der Art und Weise synchronisiert werden, sodass ein zielgerichteter organisatorischer Entwicklungsablauf möglich ist. Demnach steht diese Entwicklung gewaltigen technischen, rechtlichen sowie organisatorischen Herausforderungen gegenüber. Ziel dieser Forschung ist demnach eine Methode zur Entwicklung von Patienten-Monitoringsystemen bereitzustellen, um den genannten Herausforderungen hinreichend Rechnung zu tragen, sodass ein effektiver und möglichst reibungsfreier Entwicklungsablauf eingehalten werden kann.

**Keywords**—*Internet of Things; Mobile Health; Patienten-Monitoring-System; Selbst-Management; Method Engineering*

## I. EINLEITUNG

Der unaufhaltsame Zuwachs der Digitalisierung hält auch Einzug in den medizinischen Sektor, womit ein wirtschaftlicher Bedeutungsanstieg einhergeht [1]. Besonderen Stellenwert hat in diesem Zusammenhang das Internet of Things (IoT), welches im medizinischen Anwendungsfeld als mHealth bezeichnet wird. Die Bandbreite der realisierbaren Anwendungsmöglichkeiten [2] vergrößert sich somit immens. Insbesondere der Bereich der mobilen Patientenüberwachung stellt ein enorm großes Anwendungsfeld dar. Gerade im Zusammenhang mit der Überlastung des derzeitigen Pflegepersonals in Deutschland [3] stellt das mobile Monitoring von Patienten, insbesondere mit der Eigenschaft des Selbstmanagements, eine zukunftsweisende Alternative dar, um dieser Überlastung entgegenzuwirken. Durch den demographischen Wandel in Deutschland [4] wird diese Überlastung zusätzlich verstärkt. Die Eigenschaft des Selbstmanagements charakterisiert hierbei das selbstständige Erheben von Daten über den eigenen Krankheitsverlauf. Dieses

Monitoring erfolgt anhand mobiler körpernaher Sensorik, welche biologische Parameter messen (z. B. Blutzuckerspiegel) und Aufschluss über den aktuellen Krankheitszustand geben. Patienten-Monitoringsysteme sind im Gegensatz zur menschlichen Ausführung der Überwachung weniger fehleranfällig und führen letztendlich auch zu einer Kosteneinsparung. Die Entwicklung benannter Systeme unterscheidet sich jedoch von der Entwicklung bisheriger Systeme [2] [5] [6] [7]. Es bestehen Entwicklungsarbeiten, welche die Beachtung softwareseitiger sowie hardwareseitiger Kriterien erfordern, wodurch die Verwendung klassischer Entwicklungsansätze aus der Softwareentwicklung ungeeignet ist. Diese Entwicklung ist geprägt von hoher Interoperabilität der untereinander vernetzten Komponenten, wodurch unter Zuhilfenahme klassischer Entwicklungsansätze die Entwicklungsdauer erheblich ansteigt [2] [5]. Weiterhin erfolgt bisweilen die Entwicklung von Patienten-Monitoringsystemen fast ausschließlich technisch getrieben, wodurch der Aspekt der krankheitsbedingten Beeinträchtigung durch die Einschränkungen der Patienten bei initialen Analysen im Entwicklungszyklus nahezu unbeachtet bleibt. Dieser Aspekt führt unter Zuhilfenahme bisheriger Ansätze ebenso zu einem rapiden Anstieg der Entwicklungsdauer. Der beschriebene Engpass geeigneter Ansätze zur Entwicklung IoT-geprägter Patienten-Monitoringsysteme stellt den Fokus der vorliegenden Arbeit dar.

Um den Bereich der Patienten-Monitoringsysteme in den wissenschaftlichen Kontext des mHealth einzuordnen wird die *Taxonomie des mHealth* [2] herangezogen, welche in Abb. 1 grafisch dargestellt wird. Diese Taxonomie unterteilt sich grundsätzlich in die drei Dimensionen *medizinischer Anwendungsfall*, *technische Modalitäten* und *politische Erwägungen*, wonach jede mHealth Applikation kategorisiert werden kann. Gemäß dieser Taxonomie stellt die Gruppe der Patienten-Monitoringsysteme eine Gruppe der medizinischen Anwendungsfälle dar, die sich laut [2] weiterhin in folgende vier Untergruppen aufteilen lassen:

- klinische Systeme, welche auf das Monitoring von Patienten im stationären Bereich ausgerichtet sind
- Monitoringsysteme für Patienten im Heim-Bereich

- mobile Monitoringsysteme, die durch den Aspekt des Selbstmanagements seitens des Patienten gekennzeichnet ist
- Patienten-Monitoring innerhalb von Communities

Innerhalb dieses Beitrages stehen die mobilen Patienten-Monitoringsysteme im Fokus, welche im Selbstmanagement durch den Patienten bedient werden. Infolgedessen stellt diese Gruppe den lukrativsten Anteil dar, welcher momentan und zukünftig der erheblichen Überlastung des derzeitigen Pflegebedarfs insbesondere in Deutschland entgegenwirkt.

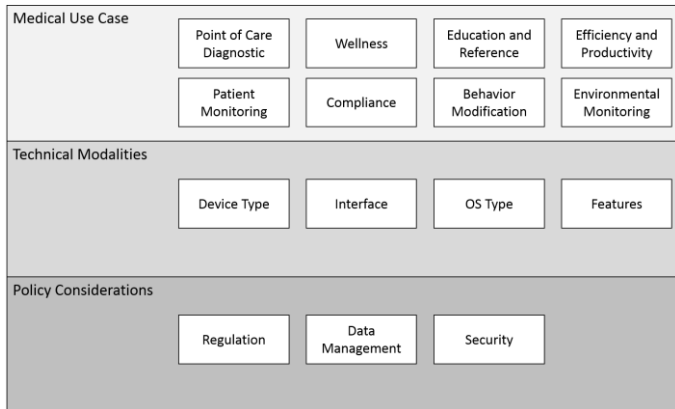


Abbildung 1: Taxonomie mHealth [2]

## II. FORSCHUNGSFRAGEN

Um den soeben benannten Engpass hinreichend zu untersuchen, ergibt sich die folgende übergeordnete Forschungsfrage: **Wie muss eine Methode gestaltet sein, um ein Patienten-Monitoringsystem zu entwickeln, welches charakterisiert ist durch mobile körpernahe Sensorik sowie die Eigenschaft des Selbstmanagements?** Der Aspekt der körpernahen Sensorik wird an dieser Stelle adressiert, um den Problemraum einzugrenzen. Die Rolle des Selbstmanagements für diesen Beitrag wurde intensiv in der Einleitung erläutert. Diese Forschungsfrage untergliedert sich in folgende Unterforschungsfragen:

RQ1: Welche Ansätze existieren bereits bezüglich der Entwicklung von Patienten-Monitoringsystemen?

RQ2: Welche Anforderungen charakterisieren eine Methode zur Entwicklung eines Patienten-Monitoringsystems, welches gekennzeichnet ist durch mobile körpernahe Sensorik sowie die Eigenschaft des Selbstmanagements?

RQ3: Welche Optimierungen können durch den Einsatz einer diesbezüglichen Methode erzielt werden, welche diese Anforderungen erfüllt?

In Anbetracht der übergeordneten Forschungsfrage beschränkt sich diese Arbeit auf mobile körpernahe Sensorik mit dem Zweck, den zu untersuchenden Problemraum einzuschränken und demnach zielgerichteter und tiefgreifender untersuchen zu können.

## III. FORSCHUNGSRAHMEN

Der in dieser Arbeit zugrunde liegende Forschungsrahmen folgt dem Ansatz der gestaltungsorientierten Wirtschaftsinformatik nach Österle et al. (2010), welcher in [8] detailliert beschrieben wird. Dementsprechend erfolgt eine Unterteilung in die Phasen Analyse, Entwurf, Evaluation sowie Diffusion, welche iterativ durchlaufen werden. Innerhalb dieser Phasen können Konstrukte, Modelle, Methoden oder Instanzen als Ergebnisse (Artefakte) generiert werden, welche gemäß dem Memorandum der gestaltungsorientierten Wirtschaftsinformatik [9] folgenden Prinzipien unterliegen:

- Abstraktion
- Originalität
- Begründung
- Nutzen

Die Beschreibung der soeben benannten Prinzipien wird an dieser Stelle direkt aus [9] zitiert. Demnach wird ein Ergebnis als abstrakt bezeichnet, sofern es auf eine Klasse von Problemen anwendbar ist. Weiterhin gilt ein Artefakt als original, sobald ein innovativer Beitrag zum publizierten Wissensstand geleistet wird. Ebenso muss jedes Artefakt nachvollziehbar begründet sowie validierbar sein. Letztlich müssen Ergebnisse für bestimmte Anspruchsgruppen einen Nutzen leisten. Unter Zuhilfenahme dieser Prinzipien unterscheidet sich die gestaltungsorientierte Wirtschaftsinformatik maßgeblich von der konkreten Problemlösung aus der Praxis.

## IV. METHODIK

Initial wurde eine systematische Literaturanalyse nach dem Framework von vom Brocke et al. (2009) [10] [11] durchgeführt. Diese Methode wurde mit dem Ziel angewendet, um einerseits Probleme und Herausforderungen in der Literatur zu identifizieren, welche zur Entwicklung von mHealth Anwendungen speziell von Patienten-Monitoringsystemen bestehen sowie andererseits verwendete Ansätze und Methoden aufzufinden, dessen man sich in der Entwicklung diesbezüglicher Systeme in Wissenschaft und Praxis bereits bedient. Mit Hilfe dieser systematischen Literaturanalyse ließ sich eine Vielzahl von Quellen identifizieren, die sich in folgende drei Bereiche unterteilen lässt:

- Technologie als Enabler für Patienten-Monitoringsysteme
- Probleme, Herausforderungen und Chancen im Bereich des mHealth
- Ansätze und Methoden zur Entwicklung von Patienten-Monitoringsystemen

Zusätzlich zur systematischen Literaturanalyse wurden Experteninterviews [12] durchgeführt, um weitere Probleme, Herausforderungen und bestehende Ansätze und Methoden bei der Entwicklung von mHealth Applikationen (insbesondere Patienten-Monitoringsystemen) vor Allem im praktischen Umfeld herauszufinden. Es wurden Interviews mit Experten



durchgeführt, welche bereits aktiv an einer Entwicklung im mHealth-Bereich mitwirkten oder diese steuerten.

Anhand der identifizierten Probleme, Herausforderungen und Chancen, welche durch die systematische Literaturanalyse sowie die Experteninterviews identifiziert werden konnten, wurde unter Zuhilfenahme des Requirements Engineering [13] [14] ein Anforderungskatalog abgeleitet. Die Anforderungen lassen sich gemäß dem Requirements Engineering in funktionale und Qualitätsanforderungen sowie Rahmenbedingungen untergliedern. Die Rahmenbedingungen können weiterhin technologisch, organisatorisch, rechtlich sowie ethisch charakterisiert sein. Der in dieser Arbeit abgeleitete Anforderungskatalog für eine Methode zur Entwicklung von Patienten-Monitoringsystemen setzt sich zusammen aus allgemeinen und spezifischen Anforderungen. Allgemeine Anforderungen definieren demnach jene Bestandteile, sodass es sich demnach gemäß [15] um eine Methode handelt. Weiterhin handelt es sich bei den spezifischen Anforderungen allesamt um Rahmenbedingungen gemäß [14], welche hier technologisch, organisatorisch und rechtlich geprägt sind.

Die mit Hilfe der durchgeführten systematischen Literaturanalyse identifizierten Ansätze und Methoden zur Entwicklung von mHealth Anwendungen wurden anschließend hinsichtlich der abgeleiteten Anforderungen evaluiert. Der Bewertungsprozess erfolgte hierbei in Anlehnung an [16]. Das Ergebnis dieser Evaluation ergab, dass die identifizierten Ansätze lediglich partiell die Anforderungen erfüllten. Somit konnte kein Ansatz identifiziert werden, welcher die Gesamtheit der spezifischen Anforderungen sowie die Anforderungen einer Methode gemäß [15] erfüllen.

Aufgrund der Tatsache, dass keine adäquaten Ansätze und Methoden in der Literatur ausfindig gemacht werden konnten, welche der Gesamtheit der abgeleiteten Anforderungen gerecht werden, erfolgt die Entwicklung einer Methode. Hierbei kommt das sogenannte Methoden Engineering [15] [17] [18] zum Einsatz. Entsprechend dieser Methodik verfügt eine Methode über ein Vorgehensmodell (Phasen mit untergeordneten Aktivitäten), ein Metamodell, ein Rollenmodell, ein Ergebnismodell sowie ein Technikmodell. Diese strukturierte Methodenbeschreibung wird zusätzlich in Abb. 2 grafisch dargestellt.

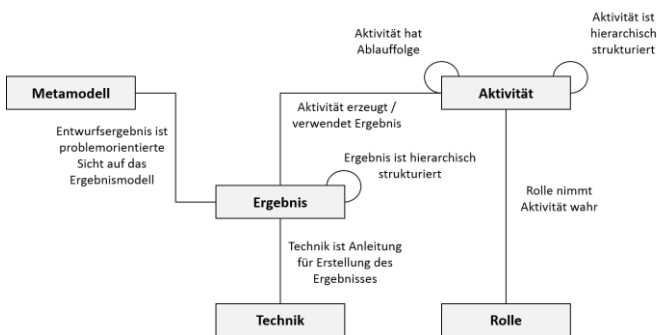


Abbildung 2: Struktur der Methodenbeschreibung (in Anlehnung an [15])

Gemäß der Evaluation aus dem in [9] beschriebenen Forschungsrahmen wird letztlich die entwickelte Methode

validiert, ob der Entwurf seinem Zweck entspricht und sich wie erwartet verhält [9] [19] [20]. Dementsprechend erfolgte die Validierung der entwickelten Methode anhand von zwei Fallstudien, welche innerhalb von zwei Forschungsprojekten durchgeführt wurden. Weiterhin existieren einzelne Bestandteile dieser Methode gemäß [9] als Modelle sowie Instanzen.

## V. ANALYSEPHASE

In diesem Kapitel werden initial die konkreten Probleme und Herausforderungen erläutert, welche mit Hilfe der systematischen Literaturanalyse sowie Experteninterviews identifiziert wurden. Anschließend werden auf dieser Grundlage Anforderungen abgeleitet sowie bestehende Ansätze und Methoden diesbezüglich evaluiert.

### A. Probleme und Herausforderungen

Bei der Entwicklung von mHealth Anwendungen insbesondere von Patienten-Monitoringsystemen bestehen technische, organisatorische, datenschutzrechtliche sowie ethische Probleme und Herausforderungen.

Eine essentielle technische Problemstellung ist die hardware- und softwareseitig geprägte Entwicklung sowie die bestehende hohe Interoperabilität der Vielzahl möglicher partizipierender Komponenten [21]. Detailliert wird die Interoperabilität in [22] und [23] untersucht. Hierbei werden die Energienutzung batteriebetriebener Komponenten, Datenübertragungsgenässe aufgrund der Bluetooth-Reichweite, die Variabilität der Smartphone-Plattformen sowie die Security und Privacy hinsichtlich des Datenmanagements als zentrale technische Probleme erläutert. Aufgrund der zunehmenden Ausprägung des Selbstmanagements seitens der Patienten werden die Patienten-Monitoringsysteme zukünftig der Herausforderung einer erhöhten Benutzerfreundlichkeit gestellt, wodurch die Security und Privacy als technisches Problem zunehmend Beachtung gewinnt [23].

Neben den technischen existieren ebenso organisatorische Problemstellungen bei der Entwicklung von Patienten-Monitoringsystemen. Laut [21] ist der Großteil der Patienten von mHealth Anwendungen sich unklar über ihren derzeitigen Krankheitszustand sind benötigen ein Tool, welches ihre gesundheitlichen und medizinischen Bedürfnisse organisiert. Jedoch ist ein Großteil dieser Patienten negativ gegenüber Online-Registrierungen, Internet-Problemen und Online-Bezahlungen eingestellt ist wodurch eine ganzheitliche Befriedigung der organisatorischen und medizinischen Bedürfnisse seitens eines Patienten-Monitoringsystems eingeschränkt wird. Weiterhin wird die Entwicklung dieser Systeme organisatorisch fast ausschließlich technische getrieben [24], wodurch eine unzureichende Einbeziehung des medizinischen Personals, vor Allem in den späteren Phasen der Entwicklung, erfolgt.

Ebenso werden in [25] [26] [27] datenschutzrechtliche Probleme und Herausforderungen erläutert, welche vor Allem im Zusammenhang mit der Verarbeitung personenbezogener Daten sowie damit zusammenhängender Restriktionen und Pflichten gemäß der Datenschutzgrundverordnung (EU DSGVO). Forschungsbezogene Tätigkeiten mit medizinischen

Hintergrund stehen außerdem vor der Problematik eines Forschungsantrages. Ein stellvertretendes Beispiel wird in [28] angeführt.

### B. Ableitung der Anforderungen

Aus den soeben beschriebenen Problemen und Herausforderungen in technischen, organisatorischen sowie datenschutzrechtlichen und ethischen Aspekten, werden in diesem Kapitel Anforderungen abgeleitet, welche eine Methode zur Entwicklung von Patienten-Monitoringsystemen charakterisieren. Folglich wird mit diesen Anforderungen auch RQ2 beantwortet.

Die Anforderungen unterteilen sich grundsätzlich in allgemeine Anforderungen (A1, A2, A3, A4) und Anforderungen, die spezifisch für die Entwicklung von Patienten-Monitoringsystemen (P1, P2, P3, P4) sind und werden zusammenfassend im Anforderungskatalog in Tab. 1 aufgelistet. Die allgemeinen Anforderungen sind in diesem Kontext inhaltlich an [15] angelehnt.

Die Flexibilität (A1) ist ein entscheidendes Kriterium bezüglich der Anwendbarkeit eines Ansatzes. Hierbei muss ein Ansatz generisch und unabhängig vom zu überwachenden Krankheitsbild sein und infolgedessen vom umzusetzenden Anwendungsszenario. Weiterhin müssen Teilschritte übersprungen werden können, um sich jeglichen praktischen Anwendungssituationen anpassen zu können, ohne dabei das zu produzierende Artefakt zu verfälschen. Die Übersichtlichkeit (A2) definiert die strukturierte Abfolge von Aktivitäten eines Ansatzes. Dahingehend existiert für jeden Aktivität ein zu nutzendes sowie zu erreichendes Artefakt. Weiterhin muss die Abhängigkeit zu anderen Artefakten definiert sein. Die strategische Abdeckung (A3) definiert die Abdeckung der Prozesse eines Ansatzes bezüglich der verschiedenen zu beachtenden Aspekte. Neben technischen Aspekten gilt es auch nicht-technische insbesondere medizinische Aspekte zu beachten. Hierbei soll auch der Engpass der unzureichenden Einbeziehung des medizinischen Personals, vor Allem in den späteren Phasen der Entwicklung gemäß [24] bedient werden. Mit der Rücksprungmöglichkeit (A4) kann während der Durchführung zu vorherigen Phasen oder Aktivitäten zurückgekehrt werden. Die Anforderung der Hardware und Software (P1) bedient die hardware- sowie softwareseitige Systembetrachtung, sodass eine ganzheitliche Systembetrachtung erzielt wird, welche beide Sichten möglichst zielgerichtet auf den Monitoring-Zweck abstimmt. Auf der Grundlage der Interoperabilität (P2), welche im vorigen Kapitel als essentielles Problem adressiert wurde, gilt es diese adäquat im Entwicklungsprozess einzukalkulieren sowie deren Auswirkungen auf die vorhandenen Komponenten im System so früh wie möglich kalkulieren zu können. Dadurch können einzelne Entwicklungsiterationen erspart beziehungsweise verkürzt werden wodurch zeitliche und personelle Ressourcen im Entwicklungsprojekt minimiert werden können. Die Patienten-Compliance (P3) adressiert die Umgangscharakteristika seitens der Patienten eines Monitoringsystems, welche mit großer Wahrscheinlichkeit auch krankheitsbedingt geprägt sind. Diese können in Abhängigkeit vom Krankheitsbild sehr verschieden sein und sollten in der frühen Phase des Entwicklungsstadiums mit

einbezogen werden, sodass Patienten-Compliance bedingte Bedienungscharakteristika frühzeitig in das Design der Systemarchitektur einfließen. Weiterhin gilt es auch datenschutzrechtliche Aspekte (P4) mit in den Entwicklungsprozess einzubeziehen, wobei an dieser Stelle insbesondere der Umgang mit personenbezogenen Daten betont sei. Ebenfalls werden dieser Anforderung der datenschutzrechtlichen Aspekte (P4) auch ethische Aspekte zugeordnet, welche den Umgang mit Ethikkommissionen fordern.

TABELLE 1: ANFORDERUNGSKATALOG

Allgemeine Anforderungen	
A1	Flexibilität
A2	Übersichtlichkeit
A3	Strategische Abdeckung
A4	Rücksprungmöglichkeit
Spezifische Anforderungen	
P1	Hardware und Software
P2	Interoperabilität
P3	Patienten-Compliance
P4	Datenschutzrechtliche Aspekte

### C. Evaluation bestehender Ansätze

Die Ansätze, welche mit Hilfe der systematischen Literaturanalyse [10] [11] ausfindig gemacht werden konnten, wurden in diesem Kapitel bezüglich den Anforderungen, die in Tab. 1 zusammengefasst sind, gemäß der Vorgehensweise in [16] evaluiert. Diese Evaluation dient gleichwohl der Beantwortung von RQ1. Folgende Ansätze wurden somit identifiziert:

- **Ignite** [29] [30]
- **IoT-Meth** [30] [31]
- **IoT-AD** [30] [32]
- **ELDAMeth** [30] [33] [34] [35]
- **SPLP-IoT** [30] [36] [37] [38] [39]
- **GSEM-IoT** [30] [40] [41]
- **IoT-Health DMA** [42]

Zusammenfassend ist festzuhalten, dass fast die Gesamtheit der Ansätze flexibel (A1) und somit unabhängig vom zu überwachenden Krankheitsbild einsetzbar ist. Weiterhin wird die Übersichtlichkeit (A2) mit einer Ausnahme von jedem dieser Ansätze erfüllt. Die strategische Abdeckung (A3) wird jedoch von einem Großteil der Ansätze entweder gar nicht oder nur unzureichend erfüllt. Ebenso ist die Möglichkeit von Rücksprüngen (A4) bei vielen Fällen nicht gegeben. Die Einbeziehung hardware- sowie softwareseitiger Perspektiven im Entwicklungsprozess (P1) wird bis auf eine Ausnahme von allen Ansätzen weitestgehend erfüllt. Der Aspekt der

Interoperabilität (P2) wird jedoch nur in seltenen Fällen beachtet. Ebenso werden die Patienten-Compliance (P3) sowie datenschutzrechtliche Aspekte (P4) in Anbetracht der existierenden Ansätze nur sehr selten und mit einer schwachen Ausprägung beachtet.

Letztlich lässt sich zusammenfassen, dass keine passende Methode zur Entwicklung von Patienten-Monitoringsystemen bisher vorhanden ist. Es existieren eine Reihe von Ansätzen, welche einige Anforderungen adressieren, jedoch kein Vertreter alle notwendigen Anforderungen ganzheitlich erfüllt. Weiterhin bestehen bei allen bisher bekannten Ansätzen Engpässe hinsichtlich der Beachtung der Interoperabilität der im System vorhandenen Komponenten, der Patienten-Compliance sowie die Einbeziehung datenschutzrechtlicher Aspekte, wodurch die Erkenntnisse der Probleme und Herausforderungen in diesem Bereich bestätigt werden.

## VI. BESCHREIBUNG DER METHODE

Auf der Grundlage der im letzten Kapitel beschriebenen Erkenntnis, dass bislang kein Ansatz vorhanden ist, welcher die Gesamtheit der Anforderungen aus Tab. 1 erfüllt, wird unter Zuhilfenahme des Methoden Engineerings [15] [17] [18] eine Methode entwickelt, welche genau diesen Forderungen gerecht wird. Wie in Abb. 2 grafisch dargestellt wird eine Methode durch die Komponenten Metamodell, Vorgehensmodell, Ergebnismodell, Rollenmodell und Technikmodell beschrieben. Das Metamodell dient als Beschreibungsmodell der Methode. Mit dem Vorgehensmodell wird die zeitliche Abfolge der einzelnen Aktivitäten definiert. Das Ergebnismodell ordnet darauf aufbauend jeder Aktivität konkret zu nutzende oder zu generierende Ergebnisse zu. Die Abarbeitung dieser Aktivitäten erfolgt dann durch Rollen mit zugeordneten Kompetenzen, welche im Rollenmodell festgelegt sind. Diese Rollen generieren somit Ergebnisse während der Aktivitäten unter Anwendung einer jeweils zugeordneten Technik. Die Zuordnung der Technik zu den Aktivitäten erfolgt im Technikmodell.

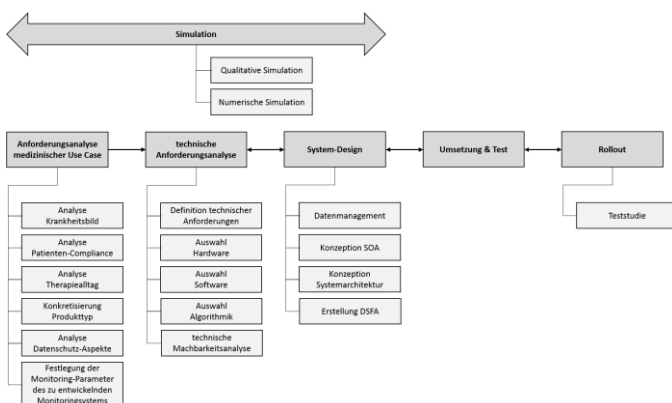


Abbildung 3: Vorgehensmodell der Methode

Die Darstellung der Methode beschränkt sich in diesem Beitrag auf die Beschreibung des Vorgehensmodells sowie deren wesentlichsten Phasen und Aktivitäten, da hier der Schwerpunkt der Arbeit liegt. In Abb. 3 wird das entwickelte Vorgehensmodell grafisch dargestellt. Es beinhaltet die

Vorgehensschritte *Anforderungsanalyse des medizinischen Use Case, technische Anforderungsanalyse, System-Design, Simulation, Umsetzung und Test* sowie *Rollout*. Diesen Phasen sind die in Abb. 3 zugewiesenen Aktivitäten untergeordnet. Einige wesentlichste Aktivitäten werden im Folgenden detaillierter beschrieben.

### A. Analyse der Patienten-Compliance

Die Analyse der Patienten-Compliance ist eine Aktivität der Analyse des medizinischen Anwendungsfalls. Sie beinhaltet die Einstufung der Kompetenzbeeinträchtigung bei der Benutzung des Monitoringsystems durch die Patienten, wobei diese Beeinträchtigung insbesondere durch das zu überwachende Krankheitsbild geprägt ist. Es werden die zu benutzenden Sensortypen sowie die Einschränkung während der Benutzung von Patienten mit strukturierten Tests evaluiert. Weiterhin können Kompetenzbeeinträchtigungen im Umgang mit oberflächenbasierten Bedienelementen durch die Zuhilfenahme des Design-Space eingeordnet werden. Detailliert wird der Design-Space in [43] erläutert, wobei die Grundstruktur in Abb. 4 dargestellt ist. Die beeinträchtigte Systembedienung bei dieser Einordnung wird durch vier grundlegende Dimensionen (vgl. Abb. 4) adressiert. Hierzu zählen die Einschränkung der Fähigkeiten des Benutzers sowie die Anpassungszeit, die Granularität und das Abstraktionsniveau der Bedienelemente.

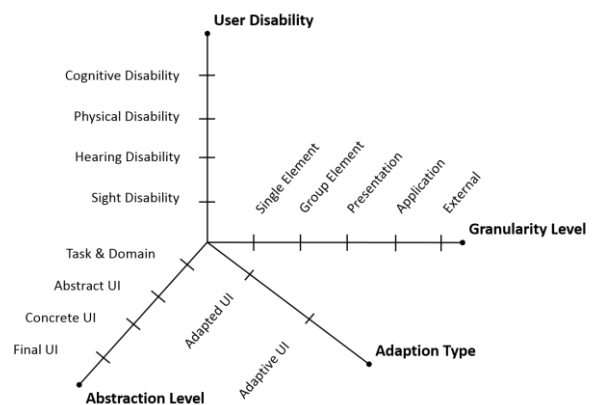


Abbildung 4: Design Space zur Klassifizierung der Patienten-Compliance [43]

### B. Datenmanagement

Das Datenmanagement ist eine Aktivität des Systemdesigns, wobei die Konzeption des Patienten-Monitoringsystems aus der Datenperspektive erfolgt. Die Daten entstammen der mobilen körpernahen Sensorik, welche gemäß des Vorgehensmodells (siehe Abb. 3) in der Phase der technischen Anforderungsanalyse ausgewählt werden, nachdem eine Begutachtung aller medizinischen Anforderungen, insbesondere aus der Sicht des zu überwachenden Krankheitsbildes, erfolgte. Konkret wird beim aktuellen Schritt des Datenmanagements die Modellierung und Speicherung der Daten festgelegt. Des Weiteren werden der Datenfluss sowie Datensicherheitsaspekte adressiert. Innerhalb dieser Methode kommt an dieser Stelle das sogenannte *Concept Mapping Paradigma* zum Einsatz, da dieses Vorgehen

besonders für die interdisziplinäre Modellierung von Datenstrukturen zwischen verschiedenen Domänen geeignet ist. In [44] wird das *Concept Mapping Paradigma* generell erläutert sowie in [45] speziell für Sensor-Zeitreihen erweitert, wobei Letzteres in Abb. 5 schematisch illustriert wird. Im vorliegenden Fall handelt es sich folglich um eine Datenmodellierung zwischen medizinischen und technischen Akteuren, womit unter Zuhilfenahme dieser Technik auch die medizinische Seite in späteren Phasen mehr in die Entwicklung einbezogen kann [44], um somit die Entwicklung nicht vollkommen technisch zu treiben. Eine Möglichkeit zur Speicherung dieser Graphen-basierten Daten erfolgt durch eine eigens dafür vorgesehene Graph-Datenbank. Dieser Ansatz zur Speicherung von Sensordaten wird in [46] vorgestellt und wurde in diesem konkreten Fall speziell für vernetzte Datenstrukturen konzipiert.

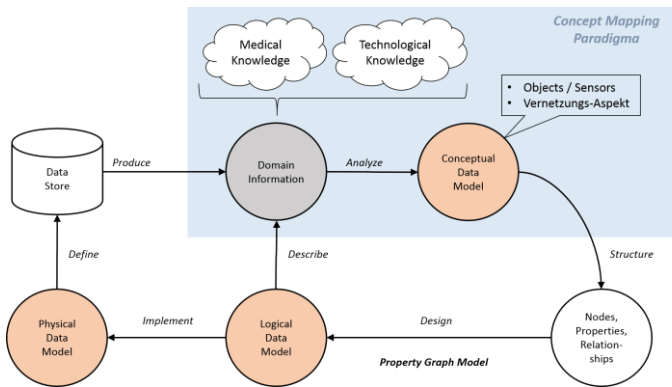


Abbildung 5: Prozess der interdisziplinären Datenmodellierung (in Anlehnung an [44] [45])

### C. Simulation

Der Vorgehensschritt der Simulation erfolgt parallel zu der medizinischen und technischen Anforderungsanalyse sowie dem Systemdesign (vgl. Abb. 3). Die Funktion der Simulationsphase ist die auftretende Interoperabilität zwischen den verschiedenen Systemkomponenten frühzeitig und strukturiert zu identifizieren, wodurch anschließend daraus resultierende Entwicklungsiterationen reduziert werden können.

Die zwei grundsätzlichen Simulationsansätze von Systemen werden in [47] beschrieben und unterscheiden sich hinsichtlich der Abstraktion des realen Simulationsproblems. Einerseits existieren *qualitative Simulationsmodelle*, welche globale Zusammenhänge zwischen den maßgebenden Systemgrößen beliebig abstrahieren und somit grundlegendes Systemverhalten annähernd beschreiben. Eine komplette zeitliche Systembeschreibung wird bei diesem Ansatz vernachlässigt. Andererseits existieren *quantitative (numerische) Simulationsmodelle*, welche das Systemverhalten mit hoher Genauigkeit beschreiben. Die quantitative Beschreibung erfolgt mit Hilfe von mathematischen Konstrukten wie Differentialgleichungen, Differenzgleichungen oder Übergangsfunktionen. Stellt man die beiden Ansätze gegenüber so ergeben sich je nach Modellierungsansatz einige Vorteile und Nachteile. Die qualitativen Simulationsmodelle können flexibel und schnell

komplexes Systemverhalten abstrahieren, wobei diese Abstraktion die Genauigkeit vernachlässigt. Die quantitative (numerische) Beschreibung modelliert den Systemcharakter mit einer höheren Genauigkeit. Jedoch ist diese Modellierung erst unflexibler und zeitaufwendiger sowie in den früheren Phasen der Entwurfsphasen nicht möglich, da die komplette Komponenten- und Parameterspezifikation der verwendeten Systemkomponenten vorhanden sein muss.

Die Simulation der vorliegenden Methode verwendet einen hybriden Ansatz. Initial wird eine qualitative Modellierung des Patienten-Monitoringsystems durchgeführt, um zunächst eine gesamtheitliche Abdeckung des Systemverhaltens zu gewährleisten. Dadurch können sensible Engpässe im Gesamtsystem identifiziert werden, die anschließend quantitativ simuliert werden, um diese Einzelaspekte detailliert untersuchen zu können. Die quantitative Modellierung erfolgt durch das strukturierte Vorgehen einer *Failure Mode and Effects Analysis (FMEA)* [48] [49] [50], welche in Abb. 6 schematisch dargestellt wird. Hierbei wird das Gesamtsystem in Elemente zerlegt, wobei anschließend Funktionen, Fehlermöglichkeiten, Ursachen und Handlungsempfehlungen strukturiert abgeleitet werden. Die anschließende numerische Simulation erfolgt mit einem implementierten Framework in MATLAB/Simulink. Hierbei werden bisher die Stromversorgung der batteriebetriebenen Komponenten sowie die Algorithmelaufzeit von Berechnungsroutinen als typische Engpässe im Monitoringsystem abgeschätzt und können durch den generischen Aufbau stets erweitert werden.

Der hybride Simulationsansatz vereint somit die Schnelligkeit und Flexibilität der qualitativen Systembeschreibung mit der hohen Genauigkeit der numerischen Variante. Letztere wird jedoch erst im späteren Verlauf der ersten drei Phasen (vgl. Abb. 3) durchgeführt, nachdem die akuten Engpässe bekannt sind.

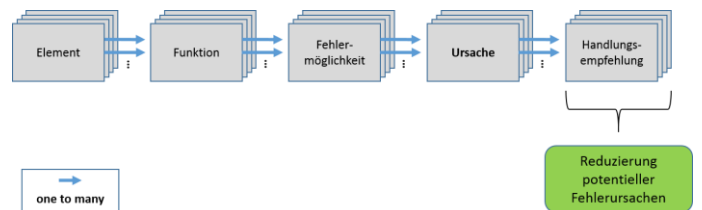


Abbildung 6: Struktur der FMEA (in Anlehnung an [48] [49] [50])

## VII. EVALUATION

Die Evaluation erfolgt durch die Zuhilfenahme der in diesem Beitrag vorgestellten Methode bei zwei Forschungsprojekten, wobei diese als Fallstudie fungieren. Demnach dient diese Methode dem Anwender als Schablone um eine strukturierte Entwicklung unter Einbeziehung der zu beachtenden Anforderungen (siehe Tab. 1) zu ermöglichen. Die erste Fallstudie ist ein Forschungsprojekt zur Entwicklung eines Depressions-Managementsystems, welches durch den Aspekt des Selbstmanagements gekennzeichnet ist. Ein weiteres Forschungsprojekt dient hinsichtlich der Evaluation als zweite Fallstudie. Hierbei ist das Ziel die Gestaltung der Umgebung, um die Mobilität älterer Personen zu optimieren. Ein Teil dieser Zielstellung fokussiert sich auf geriatrische

Krankheitsbilder (beispielsweise altersbedingte Sehbeeinträchtigungen), welche die Überwachung des Krankheitsbildes der Person einschließen. Die Entwicklung dieses Krankheitsbildes erfolgt ebenfalls unter der Anwendung der in diesem Beitrag beschriebenen Methode.

Bezugnehmend zur untergeordneten Forschungsfrage (RQ3), werden folglich Optimierungen erläutert, welche durch den Einsatz der hier vorgestellten Methode im praktischen Einsatz der Fallstudien bisher evaluiert werden konnten. Ein wesentlicher Vorteil dieser Methode im praktischen Einsatz ist die sequentielle Abgrenzung der initialen Analysephasen vor der technischen Umsetzung, da bei beiden medizinisch geprägten Projekten die Tendenz besteht, dass die Analysephase sich über die komplette Projektlaufzeit erstreckt und folglich stetig neue Anforderungen generiert werden, welche unter begrenzten Ressourceneinsatz nicht mehr realisiert werden können. Ein konkretes Beispiel für die extrem verspätete Einbeziehung medizinischer Anforderungen stellt die Patienten-Compliance dar, welche bei der hier beschriebenen Methode relativ zeitig in die Entwicklungslaufzeit einbezogen wird. Diese Methode bietet keine Garantie alle Anforderungen im Vorhinein identifizieren zu können, erhöht jedoch die Tendenz, dass wesentliche medizinische Aspekte gleich zu Beginn des Entwicklungsablaufes beachtet werden und in die Anforderungsanalyse mit einfließen. Das Resultat ist demnach eine Reduzierung von unnötigen Entwicklungsiterationen, welches einen minimierter Ressourceneinsatz von Personal und Entwicklungszeit nach sich zieht.

Weiterhin erwies sich der Einsatz des hybriden Simulationsansatzes als sehr effizientes Werkzeug bei der Entwicklung von Patienten-Monitoringsystemen. Durch die strukturierte Vorgehensweise bei diesem Simulationsansatz, wird es den partizipierenden medizinischen und technischen Partner während der Entwicklung gestattet, sich mit dem aktuellen Stand der Systemarchitektur zu synchronisieren sowie aktiv dazu beizutragen, sodass alle anderen Partner sich diesbezüglich synchronisieren können.

## VIII. ZUSAMMENFASSUNG UND AUSBLICK

In diesem Beitrag wurde der Entwurf einer Methode zur Entwicklung eines Patienten-Monitoringsystems beschrieben, welches durch körpernahe Sensorik sowie den Aspekt des Selbstmanagements charakterisiert ist. Es wurde die Forschungsmethodik beschrieben, wonach zunächst durch eine systematische Literaturanalyse sowie Experteninterviews Probleme und Herausforderungen bei bisherigen Entwicklungsprojekten im Bereich mHealth (insbesondere bei Patienten-Monitoringsystemen) aufgedeckt wurden. Von dieser Grundlage ausgehend wurden anschließend Anforderungen abgeleitet, welche eine diesbezügliche Methode beschreiben. Diese Anforderungen wurden im Anforderungskatalog zusammengefasst. Bisher existierende Ansätze in diesem Bereich, die mit Hilfe der systematischen Literaturanalyse sowie den Experteninterviews identifiziert werden konnten, wurden anschließend bezüglich dieser Anforderungen evaluiert. Kein Ansatz bot eine ganzheitliche Abdeckung dieser Anforderungen, sodass unter Zuhilfenahme des Methoden-Engineering eine Methode entwickelt wurde, welche diesen

Anforderungen gerecht wird. Diese Methode besteht im Wesentlichen aus einem Metamodell, einem Vorgehensmodell, einem Ergebnismodell, einem Technikmodell sowie einem Rollenmodell. Das entwickelte Artefakt ermöglicht letztlich die Einbeziehung technischer, organisatorischer sowie datenschutzrechtlicher und ethischer Aspekte in den Entwicklungszyklus, wodurch ein minimierter Ressourceneinsatz erzielt wird.

## REFERENCES

- [1] statista: mHealth (mobile health) industry market size projection from 2012 to 2020 (in billion U.S. dollars)\*. Online verfügbar unter <https://www.statista.com/statistics/295771/mhealth-global-market-size/>, zuletzt geprüft am 11.06.2018.
- [2] Olla, Phillip; Shimskey, Caley (2015): mHealth taxonomy. A literature survey of mobile health applications. In: Health Technol. 4 (4), S. 299–308. DOI: 10.1007/s12553-014-0093-8.
- [3] Bertelsmann Stiftung (2017): Pflegepersonal im Krankenhaus (6). Online verfügbar unter [http://faktencheck-gesundheit.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/VV\\_SpotGes\\_Pflegepersonal-KH\\_dt\\_final.pdf](http://faktencheck-gesundheit.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/VV_SpotGes_Pflegepersonal-KH_dt_final.pdf), zuletzt geprüft am 11.06.2018.
- [4] Statistisches Bundesamt: Bevölkerung in Deutschland nach Altersgruppen in den Jahren 2013, 2030 und 2060. Online verfügbar unter <https://de.statista.com/statistik/daten/studie/71539/umfrage/bevoelkerung-in-deutschland-nach-altersgruppen/>, zuletzt geprüft am 12.06.2018.
- [5] Zhang, Hehua; Zhang, Han; Wang, Xiaoning; Yang, Zuosen; Zhao, Yuhong (2017): Analysis of Requirements for Developing an mHealth-Based Health Management Platform. In: JMIR mHealth and uHealth 5 (8), S. e117. DOI: 10.2196/mhealth.5890.
- [6] Hassan, Qusay F.; Khan, Atta ur Rehman; Madani, Sajjad A. (2017): Internet of Things. Challenges, Advances, and Applications. 1st ed. Milton: CRC Press (Chapman & Hall / CRC Computer and Information Science Series). Online verfügbar unter <https://ebookcentral.proquest.com/lib/gbv/detail.action?docID=5191113>.
- [7] Miah, S. J.; Gammack, J.; Hasan, N.; Hoque, R. (2016): Design methodologies for m-health innovations: a content analysis. In: Australasian Conference on Information 59.
- [8] Österle, Hubert; Winter, Robert; Brenner, Walter (2010): Gestaltungsorientierte Wirtschaftsinformatik: Ein Plädoyer für Rigor und Relevanz. In: infowerk, Nürnberg. Online verfügbar unter <https://www.alexandria.unisg.ch/213292/1/ATTD05CN.pdf>, zuletzt geprüft am 07.08.2018.
- [9] Österle, Hubert; Becker, Jörg; Frank, Ulrich; Hess, Thomas; Karagiannis, Dimitris; Kremar, Helmut; Lous, Peter; Mertens, Peter; Oberweis, Andreas; Sinz, Elmar J. (2010): Memorandum zur gestaltungsorientierten Wirtschaftsinformatik. In: Zeitschrift für betriebswirtschaftliche Forschung. Bd. 6. S. 664-672
- [10] vom Brocke, Jan; Simons, Alexander; Niehaves, Bjoern; Riemer, Kai (2009): Reconstructing the Giant: On the Importance of Rigour in Documenting the LiteratureSearch Process. In: ECIS 2009 Proceeding 161.
- [11] vom Brocke, Jan; Simons, Alexander; Niehaves, Bjoern; Riemer, Kai (2015): Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research. In: Communications of the Association for Information Systems 37.
- [12] Mayer, Horst Otto (2018): Interview und schriftliche Befragung: Grundlagen und Methoden empirischer Sozialforschung. de Gruyter, Oldenbourg.
- [13] The Institute of Electrical and Electronics Engineers Inc. (1998): IEEE Recommended Practice for Software Requirements Specifications. In: IEEE-SA Standards Board.
- [14] Patig, Susanne; Dibbern, Jens: Requirements Engineering. Online verfügbar unter <http://www.enzyklopaedie-der>

- wirtschaftsinformatik.de/lexikon/is-management/Systementwicklung/Hauptaktivitäten-der-Systementwicklung/Problemanalyse-/Requirements-Engineering/index.html, zuletzt geprüft am 08.08.2018.
- [15] Gutzwiller, T. A. (1994): Das CC RIM-Referenzmodell für den Entwurf von betrieblichen, transaktionsorientierten Informationssystemen 1. Physica-Verlag, Heidelberg.
- [16] Meier, Markus (2002): Auswählen und bewerten, the key to innovation. ETH, Eidgenössische Technische Hochschule Zürich, Institut für Mechanische Systeme, Zentrum für Produkte-Entwicklung.
- [17] Brinkkemper, J. N.; Brinkkemper, S. (1996): Method Engineering: Engineering of Information Systems Development Methods and Tools. *Information and software technology* 38(4), S. 275-280. DOI: 10.1016/0950-5849(95)01059-9.
- [18] Rolland C. (2008): Method Engineering: Towards Methods as Services. In: Wang Q.; Pfahl D.; Raffo D. M. (eds) *Making Globally Distributed Software Development a Success Story. ICSP 2008. Lecture Notes in Computer Science 5007*. Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-540-79588-9\_2.
- [19] Institute of Electrical and Electronics Engineers (1990): IEEE Std. 610.12-1990: IEEE Standard Glossary of Software Engineering Terminology. IEEE Standards Association, Piscataway.
- [20] Marwedel, P. (2007): *Eingebettete Systeme*. Springer, Berlin.
- [21] Zhang, H.; Zhang, H.; Wang, X.; Yang, Z.; Zhao, Y. (2017): Analysis of requirements for developing an mhealth-based health management platform. In: *JMIR Mhealth Uhealth* 5(8).
- [22] Baig, M. M.; GholamHosseini, H.; Connolly, M. J. (2015): Mobile healthcare applications: system design review, critical issues and challenges. In: *Australas Phys Eng Sci Med*. 38(1), S. 23-38. DOI: 10.1007/s13246-014-0315-4.
- [23] Kher, R. K. (2016): Mobile and E-Healthcare: Recent Trends and Future Directions. In: *Journal of Health & Medical Economics* 2(3), S. 2-10. DOI: 10.21767/2471-9927.100010.
- [24] Miah, S. J.; Gammack, J.; Hasan, N.; Hoque, R. (2016): Design methodologies for m-health innovations: a content analysis. In: *Australian Conference on Information Systems* 59.
- [25] EU-Datenschutz-Grundverordnung. Online verfügbar unter [https://www.lida.bayern.de/de/datenschutz\\_eu.html](https://www.lida.bayern.de/de/datenschutz_eu.html), zuletzt geprüft am 09.08.2018.
- [26] Friedewald, M.; Obersteller, H.; Nebel, M.; Bieker, F.; Rost, M. (2016): White Paper: Datenschutz-Folgenabschätzung: Ein Werkzeug für einen besseren Datenschutz. In: *Forum Privatheit und Selbstbestimmtes Leben in der digitalen Welt*.
- [27] Bachmann, Y. (2018): *How To Leitfaden: zur Vorbereitung auf die Datenschutzgrundverordnung (DSGVO)*. Händlerbund.
- [28] Ethikkommission an der Medizinischen Fakultät der Universität Leipzig (2016): *Satzung der Ethikkommission an der Medizinischen Fakultät der Universität Leipzig*.
- [29] Slama, D.; Puhlmann, F.; Bhatnagar, R. (2015): *Enterprise IoT: Strategies and Best Practices for Connected Products and Services*. Chapter 7-8. O'Reilly Media.
- [30] Giray, G.; Tekinerdogan, B.; Tüzün, E. (2018): *IoT System Development Methods*. In: *Internet of Things: Challenges, Advances, and Applications*, S. 141-159. CRC Press.
- [31] Collins, T. (2017): *A methodology for building the internet of things*. Online verfügbar unter <http://www.iotmethodology.com/>, zuletzt geprüft am 09.08.2018.
- [32] Patel, P.; Cassou, D. (2015): Enabling high-level application development for the internet of things. In: *The Journal of Systems and Software* 103(C), S. 62-84.
- [33] Fortino, G.; Russo, W. (2012): Eldameth: An agent-oriented methodology for simulation-based prototyping of distributed agent systems. In: *Information and Software Technology* 54(6), S. 608-624.
- [34] Fortino, G.; Rango, F.; Russo, W. (2014): Eldameth design process. In: *Handbook on Agent-Oriented Design Processes*, S. 115-139.
- [35] Fortino, G.; Guerrieri, A.; Russo, W.; Savaglio, C. (2015): Towards a development methodology for smart object-oriented iot-systems: A metamodel approach. In: *IEEE International Conference on Systems, Man, and Cybernetics*, S. 1297-1302.
- [36] Ayala, I.; Amor, M. (2012): Self-configuring agents for ambient assisted living applications. In: *Ubiquitous Computing* 17(6), S. 1159-1169.
- [37] Ayala, I.; Amor, M.; Fuentes, L. (2012): An agent platform for self-configuring agents in the internet of things. In: *3<sup>rd</sup> International Workshop on Infrastructures and Tools for Multi-Agent Systems (ITMAS)*, S. 65-78.
- [38] Ayala, I.; Amor, M.; Fuentes, L. (2014): Towards a cvl process to develop agents for the iot. In: *8<sup>th</sup> International Conference (UCAml 2014)*, S. 304-311.
- [39] Ayala, I.; Amor, M.; Fuentes, L. (2015): A software product line process to develop agents for the iot. In: *Sensors* 15(7), S. 15640-15660.
- [40] Zambonelli, F. (2016): Towards a discipline of iot-oriented software engineering. In: *17<sup>th</sup> Workshop From Objects to Agents*, S. 1-7.
- [41] Tambonelli, F. (2017): Key abstraction for iot-oriented software engineering. In: *IEEE Software* 34(1), S. 38-45.
- [42] Dziak, D.; Jachimczyk, B.; Kulesza, W. J. (2017): Iot-based information system for healthcare applications: Design methodology approach. In: *Applied Science* 7(6).
- [43] Minon, R.; Paterno, F.; Arrue, M. (2013): An Environment for Designing and Sharing Adaption Rules for Accessible Applications. In: *EICS'13*.
- [44] Frisendal, T. (2016): *Graph Data Modeling for NoSQL and SQL 1*. Chapter 2, S. 7-74. Technics Publications.
- [45] Aleithe, M.; Skowron, P.; Franczyk, B.; Sommer, B. (2017): Data modeling of smart urban object networks. In: *Proceedings of the International Conference on Web Intelligence*, S. 1104-1109.
- [46] Aleithe, M.; Hegerl, U.; Ivanova, G. (2017): An Architectural Model for High Performance Pattern Matching in Linked Historical Data. In: *Lecture Notes in Business Information Processing* 263, S. 323-331. Springer. DOI: 10.1007/978-3-319-52464-1\_29.
- [47] Lunze, J. (1995): *Künstliche Intelligenz für Ingenieure – Technische Anwendungen*. Chapter 6, S. 182-192. Oldenbourg Verlag GmbH, München.
- [48] MIL-P-1629 (1949): *Procedures for Performing a Failure Mode, Effects and Criticality Analysis*
- [49] Carlson, C. S. (2012): *Effective FMEAs*. Hoboken, NY, USA. John Wiley & Sons, Inc.
- [50] Tietjen, T.; Decker, A.; Müller, D. H. (2011): *FMEA Praxis: Das Komplettpaket für Training und Anwendung* 3. Hanser.

## “Data Mining for Skills/ Qualifications in Social Networks”

Sven Gehrke (sven.gehrke@uni-jena.de)  
Prof. Johannes Ruhland (Johannes.ruhland@ uni-jena.de)  
Lehrstuhl für Wirtschaftsinformatik  
Friedrich Schiller Universität  
Jena

**Abstract.** Increasing digitalization of our environment leads to increasing complexity and tight coupling of business processes. That this increases the risk of crises or ‘natural accidents’ has been known for some time. An answer to this is a holistic mindfulness culture. This mindfulness culture has a horizontal and a vertical dimension. The vertical dimension has a permanent focus on improvements and the horizontal dimension a holistic approach, i.e. the inclusion of multiple aspects. The requirements of this holistic approach should be reflected in the self-assessment of company's employees. In the presented work it is attempted to clarify whether social media allow an insight into how it is distributed in the current enterprises. Social networks are particularly suitable for this because the members want to present themselves in the best possible way to potential employers. In doing so, they reflect the perceived demands of the companies, while at the same time the characteristics presented are „quality assured” by connected members of the network. For this purpose, more than 3500 profiles of senior IT managers in Germany were examined in a social network. When categorizing the properties, Cobit has been used as a proven IT governance tool. On the one hand, the study confirms that the clusters can be identified by generalists and specialists, on the other hand, the work shows that even the specialists have a holistic perception and demonstrate a broad knowledge of their skills.

**Keywords:** qualification distribution, business continuity, social media, COBIT, senior IT management

# Urbanes Umweltmonitoring - Eine kollaborative und interdisziplinäre Herausforderung am Beispiel der individuellen thermischen Exposition

Tobias Goblirsch *Universität Leipzig*  
*Institut f. Wirtschaftsinformatik*  
 Leipzig  
 tobias.goblirsch@uni-leipzig.de

**Zusammenfassung**—Stadtentwickler aus Wissenschaft und Praxis stehen im Zuge zunehmender globaler Urbanisierungsprozesse vor vielfältigen Herausforderungen. Klimawandel, Mobilität in Städten sowie steigende Ressourcenknappheit sind weitere Aspekte, die in diesem Zusammenhang eine wichtige Rolle spielen. Für eine nachhaltige Entwicklung gilt es, den Transformationsprozess hin zur Stadt der Zukunft mit innovativen Technologien und Strategien zu begleiten und zukunftsfähig mitzugestalten. Das in der Gesellschaft gestiegene Interesse an individuellen aber auch echtzeitnahen Informationen spielt bei diesen Transformationsprozessen eine ebenso wichtige Rolle. Durch die heterogenen Strukturen und die sehr dynamischen Prozessgrößen in urbanen Gebieten sind die Aggregation, Verarbeitung und Bereitstellung derartiger Informationen eine große Herausforderung. Konventionelle Ansätze im Bereich der Umweltbeobachtung sind aufgrund ihrer geringen räumlichen Auflösung nicht in der Lage, die lokalen Besonderheiten des urbanen Raumes in ihrer Komplexität abzubilden. Die übergeordnete Forschungsfrage lautet demnach, wie derart komplexe (Öko-)Systeme in ihrer Gesamtheit und Heterogenität erfasst, bewertet und besser verstanden bzw. geschützt werden können. Basierend auf Entwicklungen und innovativen Technologien im Bereich der Informations- und Kommunikationstechnik lassen sich bestehende Monitoringkonzepte für urbane Gebiete um z.B. Datenintegrations-, Datenverarbeitungs-, Dateninfrastruktur-, oder Analysekomponenten ergänzen, die es dann ermöglichen, individuelle echtzeitnahe Informationen dem Anwender zur Verfügung zu stellen. Eingeordnet in einem interdisziplinären Vorgehensmodell wird am Beispiel der individuellen thermischen Exposition ein Lösungsansatz gezeigt. Die individuelle thermische Exposition beschreibt die thermische Belastung, der ein Individuum in seiner Umgebung ausgesetzt ist. Anforderungsanalysen haben dargelegt, dass klassische meteorologische Methoden durch eine skalenerübergreifende Monitoringstrategie erweitert werden müssen. Hierbei besteht die große Herausforderung darin, verschiedenartige Datenquellen mit unterschiedlicher räumlicher und zeitlicher Auflösung zu integrieren, zu verarbeiten und zu analysieren, um die erwünschten individuellen Informationen zu erhalten. Für derartige Problemstellungen zeigen Forschungszweige aus anderen Bereichen der Wirtschaftsinformatik wie z.B. der Logistik oder der Energiewirtschaft bereits vielversprechende Lösungsansätze, wie derartige Informationssysteme konstruiert werden können. Der vorliegende Beitrag geht neben der bereits angedeuteten Problemstellung kurz auf das Vorgehensmodell (Raster Model Exposure Pattern) ein und fokussiert sich dann auf die Beschreibung und Integration von Datenquellen, die für ein skalenerübergreifendes Monitoring von großer Bedeutung sind. Als Datenquellen sind in diesem Zusammenhang stationäre und mobile Messsysteme, Daten aus der Fernerkundung, offene Datenschnittstellen und Daten, die durch die Bevölkerung (Crowd) erhoben werden, zu nennen. Neben der Entwicklung

eines mobilen Messsystems zeigt der vorliegende Beitrag erste Analyseergebnisse der Datenquellen, die notwendig sind, um die Anforderungen an das Informationssystem aufzustellen. Aufbauend auf diesen Ergebnissen wird eine Grobarchitektur für ein mögliches Informationssystem beschrieben und diskutiert.

**Index Terms**—Urbanes Umweltmonitoring, Skalenübergreifende Monitoringstrategie, Datenfusion, Crowd Sensing, Stationäre und Mobile Datenerfassung, Remote Sensing, Datenintegration, Open Data

## I. INTRODUCTION

Das Leben, die Wirtschaft aber auch die Arbeitswelt haben sich infolge der Digitalisierung stark gewandelt. Daten rücken immer mehr in den Mittelpunkt unserer Gesellschaft und ziehen durch Entwicklungen in Bereichen wie Smart Home, Smartphones oder auch E-Commerce immer mehr in den Alltag jedes Menschen ein. Mittlerweile sind allgemeine Daten dem Anwender aber nicht mehr ausreichend genug. Sie sind daran interessiert, dass Informationen individuell auf die eigenen Bedürfnisse oder passend zu der aktuellen Umgebungssituation abgestimmt sind. Dieser Prozess der Individualisierung ist bereits in sehr vielen Branchen ausgeprägt. Schaut man in die Bekleidungsindustrie, sieht man eigene T-Shirt-Kreationen, in der Lebensmittelindustrie selbst kreierte Müslis oder in der Automobilindustrie individuell einzustellende Bedienelemente. Diese drei Branchen zeigen nur exemplarisch, wie das Bedürfnis der Menschen an individuellen Produkten steigt. Richtet man den Blick auf den Bereich der Umweltinformationen ist festzustellen, dass eine Individualisierung von Daten für Anwender noch nicht so stark ausgeprägt ist. Zwar zeigen Wetterdienste mittlerweile standortspezifische Informationen an, kommt man aber in sehr heterogene Regionen, ist nur eine geringe räumliche Differenzierung möglich. Dies ist darauf zurückzuführen, dass die stationären Mess- und Beobachtungsnetze, wie z.B. das Netz des Deutschen Wetterdienstes oder die Ländermessnetze zur Luftüberwachung für andere Fragestellungen konzipiert wurden, als für die der Stadtklimatologie. Diese Messnetze bilden die Grundlage für numerische Wettervorhersagen, Warnmanagement oder den Schutz der Ökosysteme vor Luftschadstoffen [1], [2]. Um kleinräumige Gebiete untersuchen zu können, müssen diese Messnetze durch zusätzliche Messinfrastrukturen erweitert werden. Urbane Gebiete zeichnen sich besonders



durch ihre heterogenen Strukturen und ihre sehr dynamischen Systemkenngrößen aus. Um derartige Bedingungen erfassen zu können, müssen skalenübergreifende Monitoringstrategien zusammen mit einem Umweltinformationssystem entwickelt werden.

## II. PROBLEMRAUM

Auch wenn Städte bezogen auf die Gesamtfläche der Staaten einen sehr geringen Anteil einnehmen, steigt das Interesse der Bevölkerung in Städten zu leben kontinuierlich. Dieser Prozess der Urbanisierung ist in sehr vielen Ländern ersichtlich. In der Abbildung 1 wird die Bevölkerungsentwicklung in Städten für Nordamerika, Indien, Deutschland und der gesamten Welt aufgezeigt [3] [4].

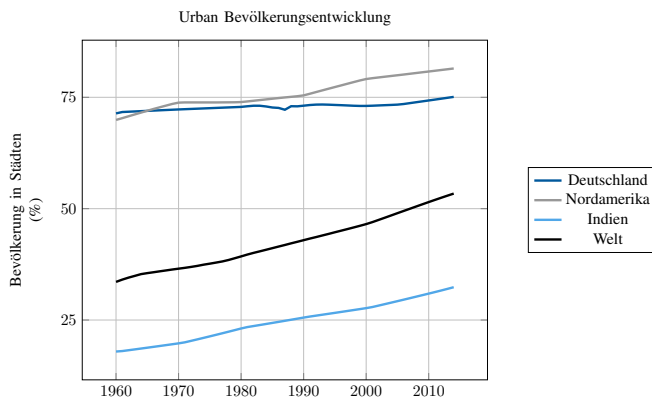


Abbildung 1. Anteil der in Städten lebenden Bevölkerung in Deutschland, Nordamerika, Indien und der gesamten Welt im Zeitraum von 1960 bis 2018 mit Prognose bis 2030

Neben dem Trend der Urbanisierung zeigen urbane Gebiete heterogene Strukturen und sehr dynamische Systemkenngrößen auf. Um die Heterogenität von Städten zu verdeutlichen, zeigt die Abbildung 2 zwei Luftbildaufnahmen der Stadt Leipzig. Die Daten stammen vom Satelliten Landsat8. Auf der linken Seite der Abbildung 2 sind die Satellitendaten für den RGB-Farbraum ausgewertet; die rechte Seite zeigt eine Auswertung hinsichtlich der Oberflächentemperatur (Land Surface Temperature - LST). Diese beiden Abbildungen verdeutlichen sehr stark die Heterogenität städtischer Strukturen. Innenstadtbereiche, Parkanlagen, Wohngebiete oder Industriegebiete lassen sich auf der linken Abbildung unterscheiden. Richtet man den Blick auf die Oberflächentemperaturen sind, bezogen auf die städtischen Strukturen, Unterschiede erkennbar.

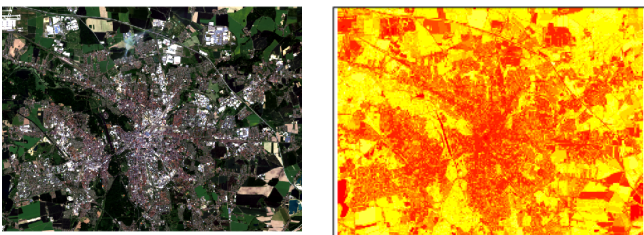


Abbildung 2. Luftbildaufnahme des Satelliten Landsat8. Auswertung der Satellitendaten hinsichtlich RGB-Farbraum (links) und Oberflächentemperatur (rechts)

Infolge dieser heterogenen Strukturen und den oft anthropogen verursachten dynamischen Systemkenngrößen in der Stadt bilden sich innerhalb des Mesoklimaraumes Stadt verschiedene unterschiedliche räumlich und zeitlich variierende Mikroklimata aus [5]. Ein Erscheinungsbild dieser Mikroklimata sind städtische Wärmeinseln. Städtische Wärmeinseln bilden sich vornehmlich bei windschwachen strahlungsreichen Sommertagen aus. Infolge dieser klimatischen Situation können sich in kleinräumigen Gebieten Temperaturunterschiede von bis zu 10K höher als im Umland einstellen. Die Ausbildung städtischer Wärmeinseln kann durch städtebauliche Maßnahmen wie zum Beispiel Grünflächen [6] unterbunden werden. Es besteht heutzutage kein Zweifel mehr daran, dass diese Erhöhung der Temperatur für die Gesundheit des Menschen bedeutsam ist [7]. Trotz des ausgeprägten klimatologischen Wissens über das Phänomen der Wärmeinsel existiert immer noch ein Defizit zur praktischen Anwendung. Aus diesem Grund ist es notwendig Daten leichter zugänglich zu machen bzw. anwendbare Tools zu entwickeln, mit deren Hilfe politische Entscheidungsträger aber auch Stadtplaner die grundlegenden Ziele wie z.B die Schaffung von gesunden Wohn- und Arbeitsbedingungen zu erreichen [8]. Zusätzlich ist aber auch die Bevölkerung daran interessiert, individuelle Informationen in ihren Lebensablauf einfließen zu lassen, um somit ihren Alltag besser zu gestalten, Abläufe zu optimieren oder sich einer Belastung zu entziehen. Um diesen Bedürfnissen gerecht zu werden, müssen innovative Monitoringstrategien in Verbindung mit echtzeitfähigen Informationssystemen und neuartigen Algorithmen zur Datenintegration für den urbanen Raum untersucht werden. In diesem Zusammenhang ist das Konzept des Raster Model Exposure Pattern entstanden.

## III. RASTER MODEL EXPOSURE PATTERN

Das Raster Model Exposure Pattern verfolgt den Lösungsansatz, urbane Gebiete in einzelne Rasterelemente zu unterteilen; zusammen bilden alle Rasterelemente dann eine Rasterkarte über das urbane Gebiet. Aggregiert man diese Rasterkarten dann über einen längeren Zeitraum, lassen sich Sequenzen erstellen. Dieser Lösungsansatz ist in der Abbildung 3 dargestellt. Die Grundidee Umweltsystem

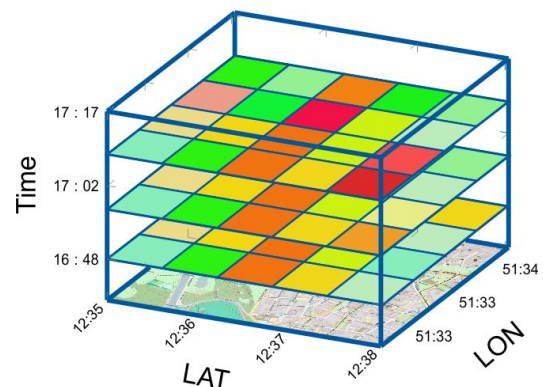


Abbildung 3. Lösungsansatz des Raster Model Exposure Patterns für urbane Gebiete

mit Hilfe von Rasterelementen zu erfassen, zu verarbeiten

und auszuwerten ist kein neuartiger Ansatz. Im Bereich der Geoinformatik haben sich Rasterdaten wie auch Vektordaten als Standard etabliert [9] [10]. Aufbauend auf dieser Betrachtungsweise wird der Lösungsansatz des Raster Model Exposure Pattern untersucht, mit deren Hilfe echtzeitnahe individuelle Expositionen für urbane Gebiete abgeleitet werden können. Individuelle Exposition beschreibt die Summe aller chemischen (z.B. Luftverunreinigungen, Partikel), biologischen (z.B. Milben) und physikalischen (Lärm, Temperatur) Einflüsse, denen ein Organismus ausgesetzt ist. Je nach Qualität, Intensität und Häufigkeit können akute oder chronische Gesundheitsschädigungen hervorgerufen werden. Bezogen auf das Raster Model Exposure Pattern liegt der Fokus auf der thermischen Exposition (TE). Bezieht man die thermische Exposition auf das einzelne Individuum, spricht man von individueller thermischer Exposition (ITE), die wie folgt zusammengefasst werden kann.

$$ITE = \left\{ \begin{array}{c} \text{Individuum } i \\ \text{Ort } (x, y, z) \\ \text{Zeit } t \end{array} \right\} \cap \left\{ \begin{array}{c} TE \\ \text{Ort } (x, y, z) \\ \text{Zeit } t \end{array} \right\}$$

Sie beschreibt die Schnittmenge der thermischen Exposition mit der des Individuums an einem bestimmten Ort und zu einer bestimmten Zeit. Erweitert man diese allgemeine Beziehung auf das Raster Model Exposure Pattern, kann die individuelle thermische Exposition mit Hilfe des Frobenius-Skalarproduktes bestimmt werden. Dieser Zusammenhang ist in der Gleichung 1 dargestellt,

$$\begin{aligned} ITE &= \langle TE_{Vektor}(t), t_{Exposition} \rangle_F \\ &= \sum_{i=0}^n \sum_{j=0}^m TE_{Vektor_{i,j}} t_{Exposition_{i,j}} \end{aligned} \quad (1)$$

wobei  $TE_{Vektor}(t)$  die zugrundeliegende Rasterkarte der thermischen Exposition beschreibt und  $t_{Exposition}$  die Zeit, wie lange ein Individuum in dem jeweiligen Rasterelemente verweilt. Die große Herausforderung beim Raster Model Exposure Pattern besteht nun darin, die thermischen Expositionen für die einzelnen Rasterelemente zu erheben. Aufgrund der eingangs erläuterten Problematik hinsichtlich Heterogenität und dynamischer Systemkenngrößen in urbanen Gebieten wird als Lösungsansatz eine skalenübergreifende Monitoringstrategie untersucht. Die thermische Exposition ist neben der Temperatur auch sehr stark von der Umgebung (Kontext) geprägt; in stark bebauten Gebieten fühlen sich z.B. 32°C für Individuen anders an als in Grünanlagen [11]. Dieser Zusammenhang muss bei der Betrachtung der thermischen Exposition Berücksichtigung finden. Die Rasterelemente in den Rasterkarten der thermischen Exposition sind somit nicht nur von der Temperatur (In Situ Map) beeinflusst, sondern es sind auch die Umgebungsbedingungen (Kontext Map) in den einzelnen Rasterelementen von Bedeutung. Die Fusion dieser beiden Rasterkarten ergibt dann die gewünschte Rasterkarte

für die thermische Exposition, die in der Abbildung 4 dargestellt ist.

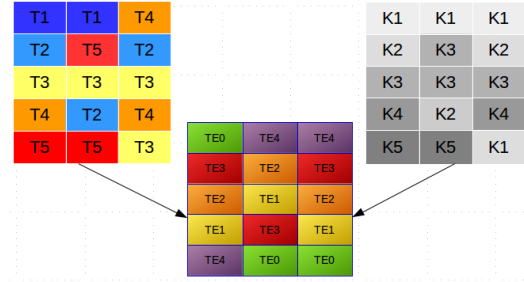


Abbildung 4. Prinzip der Fusion von In Situ und Kontext Karte zur thermischen Expositionskarte

In den ersten Untersuchungen ist davon auszugehen, dass die Rasterelemente gleiche Dimensionen und Geopositionen aufweisen, um somit den Prozess der Datenfusion zu vereinfachen. Für zukünftige Untersuchungen soll dieser Aspekt aber auch variabel gestaltet werden können.

#### IV. SKALENÜBERGREIFENDES MONITORING

Ein skalenübergreifender Monitoringansatz beschreibt, dass zur Bestimmung der Zielkenngröße unterschiedliche Messsysteme mit unterschiedlicher zeitlicher und räumlicher Auflösung zum Einsatz kommen [12]. Durch diesen Ansatz lassen sich heterogene Strukturen aber auch dynamische Systemkenngrößen, wie sie in urbanen Gebieten vorherrschen, abbilden. Die große Herausforderung besteht dann aber, darin die erhobenen Daten wieder miteinander zu verknüpfen. Für das skalenübergreifende Monitoring zur Bestimmung der thermischen Exposition, finden fünf verschiedenartige Messsysteme Anwendung. Für erste Feldexperimente wird als Untersuchungsgebiet die Stadt Leipzig herangezogen. Leipzig ist mit einer Fläche von ca. 300km<sup>2</sup> und einer Einwohnerzahl von ca. 580.000 die zehntgrößte Stadt Deutschlands. Die nachfolgenden Gliederungspunkte gehen auf die Messsysteme ein und erste Ergebnisse werden vorgestellt.

##### A. Stationäre Messsysteme

Stationäre Messsysteme bilden die Grundlage für Langzeitbeobachtungen, um kontinuierliche Informationen über den Zustand von Ökosystemen zu erhalten [13]. Für das Monitoring in Leipzig sind zwei Arten von stationären Messsystemen entwickelt worden. Zum einen befindet sich auf dem Gelände des Helmholtz-Zentrum für Umweltforschung (UFZ) eine traditionelle Klimastation (siehe Abbildung 5 links) und zum anderen sind kleinere stationäre Messsysteme für den Betrieb an Stadtbeleuchtungen (siehe Abbildung 5 rechts) entwickelt worden.

Neben der Temperaturmessung werden an der traditionellen Klimastation mittels 3D-Anemometer die genauen Strömungsverhältnisse aufgezeichnet. Diese Strömungsdaten können in der Zukunft für prädiktive Modelle eingesetzt werden.



Abbildung 5. links: Traditionelle Klimastation am Helmholtz-Zentrum für Umweltforschung (UFZ), rechts: Messsysteme an Stadtbeleuchtungen

### B. Mobile Messsysteme

Mobile Messsysteme sind in der Lage, profilierte Messungen durchzuführen; raum- und zeitbezogene Daten werden über das Untersuchungsgebiet erhoben. Neben dem Datenerhebungsprozess ist bei mobilen Anwendungen der Geräteträger von großer Bedeutung. Der Geräteträger vereint zum einen die Integration des Messsystems und zum anderen die Montage an das Transportgerät wie z.B. Fahrrad, Auto oder Straßenbahn. Die fertige Lösung für den Geräteträger zur Bestimmung der thermischen Exposition ist in der Abbildung IV-B rechts zu sehen. Ein aus Polyethylenterephthalat mit Glycol (PETG) gefertigtes Gehäuse kann mittels Wetterhüte die Temperatur messen und eine passende Bodenplatte ermöglicht es, das mobile Messsystem leicht auf dem Gepäckträger eines Fahrrades zu installieren. Die linke Seite der Abbildung IV-B zeigt die elektrotechnische Implementierung des mobilen Messsystems. Auf Basis einer mikrocontrollerbasierten Grundarchitektur, können die Temperatur, die Zeit und die Geoposition über die WLAN-Datenschnittstelle auf einen Server übertragen werden. Erste Anwendungen in Abbildung 7 haben gezeigt, dass sich durch das mobile Messsystem zuverlässig und reproduzierbare Daten erzeugen lassen.

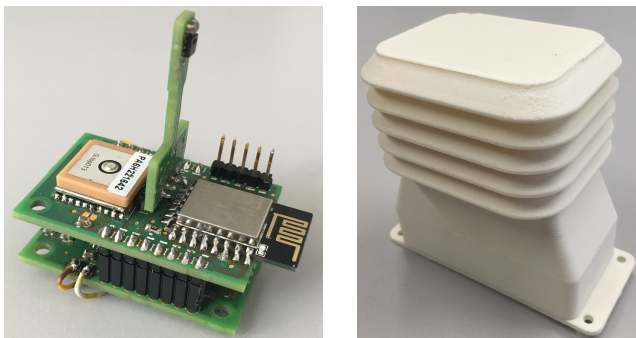


Abbildung 6. links: Messsystem, rechts: Gehäuse eines Messsystems

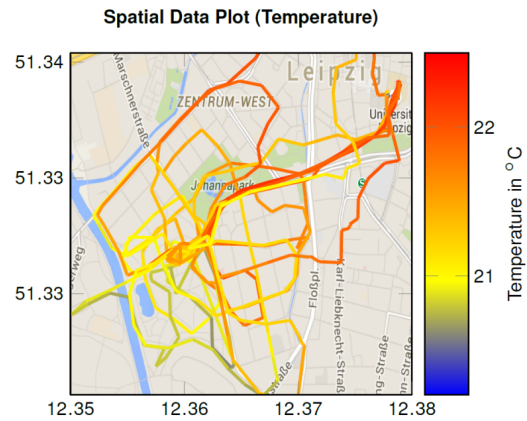


Abbildung 7. Mobile Temperaturmessungen Leipzig.

### C. Fernerkundungsdaten - Remote Sensing

Erdbeobachtungen aus der Luft ermöglichen es großskalige Informationen über den Zustand der Erdoberfläche zu erhalten. Die Fernerkundung verfolgt das Ziel für verschiedenartige (z.B. geografische, geodätische oder geologische) Anwendungsgebiete, bildhafte Daten zu erheben. Bezogen auf den jeweiligen Anwendungsfall werden spezielle Algorithmen eingesetzt, um entsprechende Zielkenngrößen abzuleiten. Für den Bereich der thermischen Exposition sind die Oberflächentemperatur (LST) und der normierter differenzierter Vegetationsindex (NDVI) wichtige Zielkenngrößen. Die Erdbeobachtungssatelliten Landsat8 und die Sentinel-Satelliten besitzen die notwendigen Messsysteme, um die aufgeführten Kennzahlen zu bestimmen. Für die aktuellen Forschungsaufgaben wird auf die Daten des Landsat8 Satelliten zurückgegriffen. Der Landsat8 Satellit ist mit zwei Sensoren (OLI, TIRS), die insgesamt 11 Frequenzbänder abdecken, ausgestattet [14]. Er liefert seit über 40 Jahren diese hoch aufgelösten kalibrierten Daten der Erdoberfläche, die dann einer breiten Nutzergemeinde zur Verfügung gestellt werden. Aufbauend auf dieser Datenbasis lassen sich die Zielkenngrößen NDVI:

$$NDVI = \frac{NIR - Rot}{NIR + Rot}$$

mit  $NIR$  = Nahes Infrarot Band (Band 5) (2)

$Rot$  = Rotem sichtbaren Band (Band 4)

und LST in °C:

$$LST = \frac{Bt}{1 + (\omega * \frac{Bt}{14388}) \ln(e)} - 273,15$$

mit  $Bt$  = Beleuchtungstemperatur (3)

$\omega$  = Wellenlänge der emittierten Strahlung)

$e$  = elektromagnetischer Emissionsgrad

bestimmen. Diese Kennzahlen fließen in die Charakterisierung urbaner Gebiete mit ein und sollen Erkenntnisse über Umgebungsinformationen und somit für die Kontextkarte liefern.

D. Offene Datenquellen - Open Data

Die Bereitstellung offener Datenschnittstellen für den Bereich des Umweltmonitorings steigt kontinuierlich. Neben offenen Daten von Messinfrastrukturen, wie sie z.B. der Deutsche Wetterdienst bereitstellt, spielen Daten von Umgebungsinformationen eine wichtige Rolle. Fragen hinsichtlich der Klassifizierung oder Charakterisierung urbaner Gebiete lassen sich zum Beispiel mittels OpenStreetMap beantworten. OpenStreetMap verfolgt das Ziel offene Weltkarten mit vielen zusätzlichen Informationen über z.B. Häuser, Wälder oder Straßen zur Verfügung zu stellen. Betrachtet man diese Informationsbasis von OpenStreetMap in Bezug auf die thermische Exposition, kann festgestellt werden, dass Informationen über die Verteilung von Gebäuden, Straßen und Grünflächen herangezogen werden können, um Umgebungsinformationen zu erhalten. Die Abbildung 8 zeigt für das gesamte Stadtgebiet von Leipzig Informationen über die Straßenzüge (rot) und Gebäudestrukturen (gelb). In dieser Darstellung ist sehr



Abbildung 8. Auswertung von OpenStreetMap Daten hinsichtlich Straßenzüge (rot) und Gebäudestrukturen (gelb) für die Stadt Leipzig

deutlich die Struktur der Stadt erkennbar. Innenstadtgebiete, Wohngebiete in der Randlage aber auch unbebaute Gebiete lassen sich identifizieren.

E. Crowd Daten

Crowd Sensing, das Erheben von Daten durch eine Menschenmenge, wird in Zukunft einen immer wichtigeren Beitrag für die Bereiche Smart Cities, Klima und Umwelt oder auch beim Erfassen von Extremereignissen leisten. Crowd Sensing kann entweder durch einzelne Personen, die mitgeführte Sensorik besitzen, erfolgen (objektive Messung) oder aber die Person fungiert selbst als Sensor (subjektive Messung) [15]. Durch diesen Ansatz ergeben sich neue umfassende Chancen im Bereich des urbanen Umweltmonitorings. Um Daten auf Basis von Crowd Sensing für die thermische Exposition zu erheben, ist die in der Abbildung 9 dargestellte Applikation für mobile Endgeräte entstanden. Für den Anwender besteht die Möglichkeit, seine individuelle thermische Exposition mittels vier Auswahlbuttons (Low, Medium, High, Extreme) zu bestimmen. Zusätzlich besteht die Möglichkeit zusätzliche Kontextinformationen in einen Freitext einzutragen. Die Daten werden dann über die HTTP-Schnittstelle an einen entsprechenden Server weitergeleitet.

V. URBANES UMWELTINFORMATIONSSYSTEM FÜR INDIVIDUELLE THERMISCHE EXPOSITIONEN

Wie der Abschnitt IV zeigt, werden beim skalenergreifenden Monitoring für die Bestimmung

Individual Thermal Exposure

Type Context Information



Abbildung 9. App für Adroid und iOS zur Bestimmung der individuellen thermischen Exposition mit Kontextinformationen.

der thermischen Exposition verschiedenartige Datenquellen benötigt. Neben der räumlichen- und zeitlichen Auflösung, unterscheiden sich die Datenquellen aus Abschnitt IV auch hinsichtlich ihrer Datenerhebungsmethode (Typ) und dem eigentlichen Messwert (Wert). Dies hat zur Folge, dass auch die Datenflüsse der Messsysteme sehr unterschiedlich hinsichtlich Volumen, Geschwindigkeit und Vielfalt sind. Eine Übersicht dieser Zusammenhänge ist in der Tabelle I dargestellt.

Tabelle I

PROPERTIES OF DIFFERENT TYPES OF ENVIRONMENTAL OBSERVATION TECHNOLOGIES REGARDING THE VALUE, TIME, SPATIAL AND INFORMATION ABOUT THE DATA TYPE.

	Stationär	Mobil	Remote	Open	Crowd
Typ	quantitativ	quantitativ	quantitativ	qualitativ	qualitativ
Zeit	kontinuierlich	kontinuierlich	diskontinuierlich	diskret	diskontinuierlich
Raum	Punkt	Profil	Areal	Areal	Profil
Wert	direkt	direkt	proxy	indirekt	indirekt
Volumen	MB	MB	TB	GB	MB
Geschwindigkeit	periodisch	adaptiv	periodisch	batch	batch
Vielfalt	strukturiert	strukturiert	strukturiert	strukturiert	unstrukturiert

Desweiteren muss hinzugefügt werden, dass der Prozess der Datenerhebung durch ein sehr interdisziplinäres und kollaboratives Zusammenspiel vieler Domänenexperten geprägt wird. Für jede Datenquelle haben sich Vorgehnsmodelle, Algorithmen aber auch Entwicklungstools etabliert, die sich sehr stark unterscheiden. Die große Herausforderung für das Informationssystem besteht deshalb darin, einen Lösungsansatz zu identifizieren, in dem diese Anforderungen Berücksichtigung finden. In der Softwarearchitekturfrage steht man oft zwischen den beiden Extrema monolithischer vs. serviceorientierter Ansatz. In monolithische Softwarearchitekturen ist die Funktionalität zentralisiert implementiert und verfolgt keine explizite Gliederung in Teilsysteme. Bei einem serviceorientierten Ansatz wird die Idee verfolgt komplexe Zusammenhänge in kleine, lose gekoppelte und eigenständige Softwarekomponenten zu unterteilen. Durch das Kombinieren und Zusammenfügen dieser Service entsteht dann das Anwendungssystem. Durch die Eingangs beschriebenen Anforderungen an das Informationssystem zur Bestimmung thermischer Expositionen wird

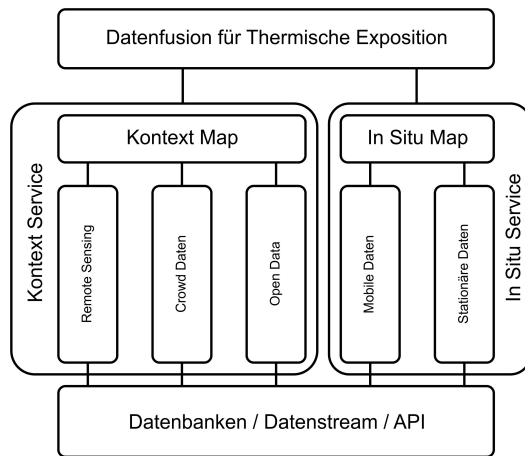


Abbildung 10. Schichtenmodell einer Microservice-Architekturen mit monolithischem Prozessierungskernen zur Bestimmung individueller Expositionen

im folgenden ein hybrider Lösungsansatz vorgestellt. Die Abbildung 10 zeigt eine Microservice-Architektur mit monolithischem Prozessierungskernen. Die Softwarearchitektur besteht aus vier Schichten. In der untersten Schicht wird das Thema Datenmanagement bearbeitet. Es werden Datenbanken, Datenstreams oder Schnittstellen zu APIs organisiert und bilden somit die Datenbasis für die übergeordneten Schichten. In der zweiten Schicht findet eine vertikale Dekomposition der unterschiedlichen Datenquellen und damit auch der Domänen statt. Jede vertikale Komponente verwaltet ihre eigenen Algorithmen, Deployment- oder Monitoringprozesse; Shared Code ist zu vermeiden. Die beschriebenen domänengetriebenen Prozessierungskomponenten bilden die Grundlage für die übergeordnete Kontext- und In Situ Map. In dieser Schicht findet die Aggregation und Fusion der Daten zur Kontext- und In Situ Map statt. Die oberste Schicht im Architekturmodell beinhaltet die Fusion der Kontext- und In Situ Map, mit dem Ergebnis (siehe Abbildung 4) einer Karte, die die thermische Exposition beschreibt.

## VI. ZUSAMMENFASSUNG

Die vorliegende Arbeit gibt einen Einblick in den Bereich des urbanen Umweltmonitorings und beschreibt den Problemraum urbaner Gebiete; Heterogenitäten und sehr dynamische Systemkenngrößen spielen dabei eine übergeordnete Rolle. Basierend auf einer skalenergreifenden Monitoringstrategie wird das Konzept des Raster Model Exposure Pattern erläutert. Das Raster Model Exposure Pattern bildet dabei die Grundlage zur Bestimmung individueller Expositionen. Mit dem speziellen Anwendungsfall der individuellen thermischen Exposition werden Messsysteme vorgestellt, die für das skalenergreifende Monitoring zum Einsatz kommen. In diesem Zusammenhang wurden zum einen neuartige Messsysteme (stationär, mobile) und zum anderen Softwarekomponenten (Open Data, Remote Sensing, Crowd) entwickelt. Aufbauend auf diesen Datenquellen, wird eine erst Grobarchitektur vorgestellt, die die notwendigen Systemanforderungen abbildet.

## LITERATUR

- [1] Dwd.de. (2018). Wetter und Klima - Deutscher Wetterdienst - Messnetz. [online] Available at: [https://www.dwd.de/DE/derdwd/messnetz/messnetz\\_node.html](https://www.dwd.de/DE/derdwd/messnetz/messnetz_node.html) [Accessed 10 Sep. 2018].
- [2] Umweltbundesamt. (2018). Luft. [online] Available at: <https://www.umweltbundesamt.de/themen/luft> [Accessed 10 Sep. 2018].
- [3] World Urbanization Prospects. (2018). [online] Available at: <https://population.un.org/wup/> [Accessed 25 Sep. 2018].
- [4] In Städten lebende Bevölkerung in Deutschland und weltweit bis 2030 — Statistik, Statista, 2018. [Online]. Available: <https://de.statista.com/statistik/daten/studie/152879/umfrage/in-staedten-lebende-bevoelkerung-in-deutschland-und-weltweit/>. [Accessed: 25-Sep- 2018].
- [5] W. Kuttler, "Klimawandel im urbanen Bereich, Teil 1, Wirkungen Climate change in urban areas, Part 1, Effects," Environ. Sci. Eur., vol. 23, no. 1, p. 11, 2011.
- [6] F. Koch, L. Bilke, C. Helbig, and U. Schlink, "Compact or cool? The impact of brownfield redevelopment on inner-city micro climate," Sustain. Cities Soc., vol. 38, pp. 31–41, Apr. 2018.
- [7] S. Breitner, K. Wolf, A. Peters, and A. Schneider, "Short-term effects of air temperature on cause-specific cardiovascular mortality in Bavaria, Germany.," Heart, pp. 1272–1280, 2014.
- [8] Endlicher, W. and Gerstengarbe, F. (2007). Der Klimawandel - Einblicke, Rückblicke und Ausblicke. [ebook] Potsdam, p.111. Available at: [https://www.pik-potsdam.de/services/infothek/buecher\\_broschueren/.images/broschuere\\_cms\\_100.pdf](https://www.pik-potsdam.de/services/infothek/buecher_broschueren/.images/broschuere_cms_100.pdf) [Accessed 11 Sep. 2018].
- [9] R. Bill, Grundlagen der Geo-Informationssysteme, 6., Völlig. Berlin: Wichmann.
- [10] N. de Lange, Geoinformationssysteme. 2013.
- [11] E. Banzhaf et al., "A conceptual framework for integrated analysis of environmental quality and quality of life," Ecol. Indic., vol. 45, pp. 664–668, 2014.
- [12] S. Reis et al., "Integrating modelling and smart sensors for environmental and human health," Environ. Model. Softw., vol. 74, pp. 238–246, 2015.
- [13] H. Mollenhauer et al., "Long-term environmental monitoring infrastructures in Europe: observations, measurements, scales, and socio-ecological representativeness," Sci. Total Environ., vol. 624, pp. 968–978, May 2018.
- [14] U.S. Geological Survey (USGS), Landsat 8 Data Users Handbook, 2016
- [15] R. Ganti, F. Ye, and H. Hei, "Mobile crowdsensing: Current state and future challenges," IEEE Commun. Mag., vol. 49, no. 11, pp. 32–39, 2011.

# Entwicklung eines Modells zur Abbildung von Transformationsprozessen von Servicesystemen

## am Beispiel von Energiemanagementsystemen

Alexandra Herzog  
Institut für Wirtschaftsinformatik  
Universität Leipzig  
Leipzig  
herzog@wifa.uni-leipzig.de

**Abstract**— Die Transformation vom Produkt- zum Dienstleistungsunternehmen rückt stetig in den Fokus von Wissenschaft und Praxis. Neben den Unternehmen durchlaufen aber auch Kunden Transformationsprozesse. Mit dieser Forschung soll ein Modell vorgestellt werden, mit welchem die Transformationsprozesse darstellbar sind. Im vorliegenden Beitrag wird das Vorgehen ausgehend von einer Literaturrecherche zur Identifikation der Anforderungen an das entsprechende Artefakt in der Anwendungsdomäne der Energiewirtschaft beschrieben. Es schließt sich die Vorstellung des weiteren Vorgehens zum Design und der Validierung des Artefakts an. Der Beitrag zeigt die Forschungsschwerpunkte und Artefakte auf. Zentraler Gegenstand ist das Modell, welches mithilfe eines weiteren Artefakts verprobt wird. Mit dem entstehenden Service Konfigurator wird das Modell angewendet.

**Keywords**— Servicesysteme, Energiemanagement, Transformationsprozesse, Service Configurator

### I. EINLEITUNG

Die Transformation vom Produkt- zum Dienstleistungsunternehmen ist in den letzten Jahren zunehmend in den Mittelpunkt von Wissenschaft und Praxis gerückt. Durch steigende Wettbewerbsintensität, Angleichung von Produktangeboten und voranschreitender Digitalisierung sind Unternehmen gefordert, neben ihren Produkten auch Services zu gestalten und zu vertreiben. Dadurch ergibt sich auch die Möglichkeit zum Aufbau neuer Kundenbeziehungen. [1] Der Fokus liegt in der Entwicklung innovativer servicebasierter und vor allem kundenorientierter Geschäftsmodelle. Auch der Kunde selbst durchlebt eine Transformation durch die voranschreitende Dezentralisierung und Digitalisierung und wird zum sogenannten Prosumer. Damit ist der Kunde nicht mehr nur allein Konsument, sondern wird auch zum Produzent. Durch die Energiewende findet in der Energiebranche ein Paradigmenwechsel statt. Der Ausbau der erneuerbaren Energien führt zu einer Dezentralisierung, da mittlerweile ein Großteil der regenerativen Erzeugung durch Privatpersonen und Gewerbetunden erfolgt. Die bisherigen Energieversorgungsunternehmen (EVU) tragen nur einen kleinen Anteil. Der vormals reine Abnehmer von Energie wird zum Erzeuger und damit zum Prosumer. [2] Ausgehend von diesen beiden Transformationsprozessen bestehen

Herausforderungen sowohl auf Unternehmens- aber auch auf Kundenseite. [3] Es bedarf einer Unterstützung für den Prosumer bei der Identifikation und Strukturierung der angebotenen Services, um die für ihn relevanten Services aus den angebotenen Vorrat an Services zu identifizieren, zu nutzen und damit die für ihn und seine Bedürfnisse erforderlichen Services zu konsumieren. Aufgrund der Serviceorientierung der Unternehmen muss der Prosumer nun die Services von verschiedenen Unternehmen mit deren spezifischen Ausprägungen identifizieren, die Bestandteile analysieren und die für ihn relevanten Services erwerben.

### II. ZIELSETZUNG

#### A. Der Energiesektor durchlebt Transformationsprozesse

Der Energiesektor erlebt seit Jahren verschiedene Transformationsprozesse. Transformation lässt sich als der Prozess der Überführung von einem bestehenden in einen neuen Zustand definieren. Dabei findet die Umwandlung einer Form, Struktur oder Gestalt mit oder ohne Inhalts- und Substanzverlust statt [4]. Dolesky (2017) klassifiziert diese im Bereich der Energiewirtschaft in System-, Branchen- und Unternehmenstransformation. [5] Unter Systemtransformation werden die Prozesse verstanden, die aufgrund der Ablösung der klassischen Energieträger auf erneuerbare Energien entstehen. Zudem beschreibt es den Wandel von zentralen hin zu dezentralen Erzeugungsstrukturen. Die Branchentransformation beschreibt den Wandel des klassischen Geschäfts zur Dienstleistungs- und Kundenorientierung. Der bisherige Letztverbraucher wird als Kunde erkannt und muss mit neuen Angeboten bedient werden. Als weiterer Transformationsprozess findet die Unternehmenstransformation statt. Diese bedeutet eine branchenweite Umgestaltung der klassischen Energieverteilung zu einer innovativen Energieversorgung. [5] Zusätzlich zu den genannten Transformationsprozessen erfährt die Energiebranche auch einen enormen Einfluss durch die Digitalisierung und damit einhergehend auch eine Technologietransformation. Mit den Einsatz neuer Technologien wie Smart Meter oder Internet of Thing (IoT) werden alle drei genannten Transformationsprozesse ebenfalls unterstützt oder beeinflusst. Neue Technologien wie

Blockchain oder BigData verändern die Möglichkeiten im gesamten Prozess.

Neben diesen, auf die Unternehmen abzielenden Transformationsprozesse, durchläuft der Endkunde ebenfalls einen Wandel. Durch voranschreitende Digitalisierung erfährt auch er neue Möglichkeiten in der Energiebeschaffung. So kann er bspw. durch den Einsatz von Solarzellen auf Dach einfach selbst Strom erzeugen. Aus dem „Konsumer“ wird ein Prosumer. Ein Prosumer „ist eine aktive oder passive Instanz in der Wertschöpfungskette, welche Energie nicht nur konsumiert, sondern zu einem gewissen Grad auch produziert oder mit seinen Assets durch Dienstleistungen Dritter am Energieversorgungssystem teilnimmt.“ [2]

### B. Die Dezentralisierung findet auf allen Ebenen statt

Der Wandel von zentralen hin zu dezentralen Erzeugungsstrukturen findet auf allen Ebenen statt. Neben der branchenseitigen Ausprägung wächst auch der Bedarf an einer IT-seitigen Unterstützung. TABLE I. zeigt eine Übersicht über die Branchen- und IT-seitigen Einflüsse.

TABLE I. DEZENTRALISIERUNG

Dezentralisierung in der Energiebranche	
Branchenseitig	IT-seitig
Die Förderung regenerativer Energien führt zu dezentraler Energieerzeugung [6]	Bedarf dezentraler Energiemanagementsysteme wächst [8], [9], [10]
Zukünftig werden dezentrale Energielieferanten am Energiemarkt teilnehmen [7]	

Als neuer Akteur in der Prozesslandschaft muss der Prosumer durch geeignete IT-Systeme und Applikationen unterstützt werden. Als wesentliche Systeme zur Unterstützung werden Energiemanagementsysteme (EMS) genannt. Unter einem EMS wird ein ganzheitlicher Ansatz verstanden, der wirtschaftliche und technische Kriterien gemeinsam betrachtet [11] und den Endkunden in diesen Prozess durch Nutzung von smarten Services integriert.

### C. Servicetransformation durch Digitalisierung

Durch die Digitalisierung und der damit einhergehenden Technologietransformation ändern sich auch die bisher angebotenen Services. Auch diese durchlaufen Transformationsprozesse. In der Literatur wird diese Änderung als Servicetransformation bezeichnet. Dabei behandelt die Servicetransformation durch Digitalisierung die Generierung von Kundennutzen durch Vernetzung von Unternehmen, Produkten und Kunden. Das beinhaltet aber nicht ausschließlich das Angebot von Dienstleistungen, sondern auch die Entwicklung innovativer servicebasierter Geschäftsmodelle [1].

### D. Zielsetzung der Dissertation

Aus den genannten Treibern der Transformationsprozesse ergeben sich Herausforderungen, welchen mit den Ergebnissen der Dissertation begegnet werden soll. Wesentliche Herausforderungen stellen für den Prosumer die zu erfüllenden Prozesse dar. Aufgrund der Teilnahme an einzelnen Prozessen des Gesamtprozesses der Energiewirtschaft, müssen diese Prozesse identifiziert und die erforderlichen Aufgaben in diesen Prozessen gelöst werden. Einige dieser Prozesse unterliegen gesetzlichen Vorgaben, welche in regelmäßigen Zeitabständen geändert werden. Der Prosumer muss somit nicht nur die Prozesse, sondern auch ihre Änderungen kennen, verstehen und anwenden.

Die Dissertation liefert für den Prosumer eine Serviceübersicht, im Folgenden Servicelandkarte genannt, um eine Einordnungshilfe in den Gesamtprozess zu erhalten. Als Ausbaustufe wird ein Service Konfigurator erstellt, welcher IT-gestützt den Auswahlprozess der benötigten Services unterstützt. Zur Beschreibung dieser Services und zur Darstellung der Verbindungen wird die Forschung im Bereich der Service System Beschreibung erweitert. Dies ist erforderlich, da die Cluster für den Konfigurator einheitlich beschrieben werden müssen.

## III. FORSCHUNGSDESIGN

Die Dissertation folgt dem Design Science Research Prozess [12], welcher einen iterativen Prozess vorzieht. Zunächst erfolgt die Motivation in das Thema und die Problemdefinition. Anschließend werden die Ziele der Lösung definiert. Nach dieser Zieldefinition erfolgen das Design und die Entwicklung des Hauptartefakts. Die erarbeiteten Ergebnisse werden validiert und veröffentlicht.

### A. Forschungsfragen

Mit der vorliegenden Dissertation sollen folgende Forschungsfragen beantwortet werden:

FF1: Welche Treiber führen zu Transformationsprozessen in der Energiebranche?

Zur Unterstützung bei der Identifikation und Strukturierung der Services in der Energiebranchen müssen zunächst die wesentlichen Treiber der Transformationsprozesse analysiert werden.

FF2: Wie können die Transformationsprozesse in der Servicedarstellung abgebildet werden?

Die Darstellung der Services bzw. Service Systeme ist komplex und muss die entsprechenden Akteure und die Beziehungen untereinander beinhalten, um Aussagen über betroffene Services tätigen zu können.

Als Artefakt zur Validierung soll ein Service Konfigurator eingesetzt werden. Zur Erarbeitung des Konfigurators muss der Aufbau geklärt werden. Daher wird dies in einer dritten Forschungsfrage betrachtet.

FF3: Wie ist der Service Konfigurator aufgebaut?

## B. Forschungsmethodik

Zur Beantwortung der gestellten Forschungsfragen werden verschiedene Forschungsmethoden eingesetzt.

Zur Erarbeitung der Treiber der Transformationsprozesse wird eine strukturierte Literaturanalyse nach vom Brocke et al. (2009) vorgenommen [13]. Als Ergebnisse neben den identifizierten Transformationsprozessen und Treibern stellen die Anforderungen an eine IT-Unterstützung dar. Diese Anforderungen münden in einem Anforderungskatalog. Durch Interviews werden diese Anforderungen zusätzlich mit Anforderungen aus der Praxis angereichert.

Zur Erarbeitung der Service System Beschreibung erfolgt eine Literaturrecherche zu bestehenden Beschreibungsmodellen. Diese werden hinsichtlich ihrer Vor- und Nachteile eruiert. Durch Interviews mit Praxispartnern (Requirements Engineers, IT Consultants) werden diese Modellierungen hinsichtlich ihrer Praxistauglichkeit untersucht. Dieser iterative Prozess soll zu einem Modell führen, aus welchem die Servicemap erstellt werden kann.

Als Mittel zur Evaluation der Lösung wurde das Prototyping Verfahren ausgewählt. Es soll ein Service Konfigurator prototypisch und anhand eines größeren komplexen Anwendungsfalls aufzeigen, ob die Modellierung der Service Systeme anhand des in der Dissertation entwickelten Beschreibungsmodells den Anforderungen entspricht. Der Prototyp wird durch eine Feldanalyse geprüft.

## C. Forschungsschwerpunkte und Artefakte

Die Arbeit soll einen Beitrag zur Unterstützung bei der Identifikation und Strukturierung von Services leisten. Als Hauptartefakt soll ein Modell erarbeitet werden, welches die Abbildung von Transformationsprozessen unterstützt. Zur Erarbeitung dieses zentralen Artefakts werden weitere Artefakte benötigt.

Es wird ein Anforderungskatalog mit den zentralen Anforderungen an die Transformationen in den zentralen Ausprägungen (Branche, IT, System) erstellt. Als weiteres Artefakt wird eine Beschreibungsschablone für die Services bzw. Service Systeme erarbeitet und validiert. Auf Basis dieser Beschreibungsschablone soll es möglich sein, für den jeweiligen Anwendungskontext Servicemaps aufzubauen, die die jeweiligen Zusammenhänge abbilden. Zudem sollen diese Schablonen die Transformationspunkte aufzeigen.

Für die Prototypische Umsetzung wird ein Prototyp als weiteres Artefakt erzeugt werden. Dieser Prototyp prüft eine exemplarische Umsetzung an einem konkreten, komplexen Anwendungsfall aus der Energiewirtschaft.

## D. Die Dissertation verbindet Forschungsbereiche

Die Dissertation verbindet die Service Science und das Service Systems Engineering.

Peters et al. (2016) definieren Service Systems als "are configurations of people, technologies, organizations, and information that create and deliver value to all stakeholders in the system" [14]. Konkret sind Service Systeme komplexe, sozio-technische Systeme, welche eine interaktive

Wertschöpfung ermöglichen [14]. Forschungsherausforderungen in diesen Bereich liegen in der Darstellung der Komplexität, bspw. bei der Darstellung vom Service System zum Service Ökosystem. Die Abbildung muss die dynamische Anforderung der Service Systeme berücksichtigen. Service Systeme sind relativ selbständige, sich selbst anpassende Systeme ressourcenintegrierender Akteure, die durch gemeinsame Logiken und gegenseitige Wertschöpfung durch einen Serviceaustausch verbunden sind. [15]

Der zweite Bereich ist das Service Systems Engineering. „Service-Systems-Engineering oder die systematische Servicesystementwicklung (SSE) stellt auf die systematische Gestaltung und Entwicklung von Servicesystemen ab“ [16]. Forschungsherausforderungen im Bereich SSE liegen ebenfalls in der Komplexität von Servicesystemen, die nicht nur Informationen und physische Artefakte vereinen, sondern auch Kommunikation und vernetzte Akteure. Der gewählte Anwendungsbereich der Energiewirtschaft erfüllt diese Komplexität. [16]

## E. Das entstehende Artefakt erzeugt Nutzen für die Wissenschaft

Das Hauptartefakt liefert einen Beitrag für die Gestaltung multidisziplinärer, komplexer und soziotechnischer Systeme. Es wird ebenfalls ein Beitrag für die Service System Beschreibung und Modellierung geleistet. Hauptaugenmerk soll dabei auf der Komplexitätsreduktion liegen. Das Modell soll einen Überblick über die Services, deren Akteure und der Kommunikation der beteiligten Services liefern. Zudem soll es die Abbildung der Transformationsprozesse unterstützen.

## F. Das entstehende Artefakt erzeugt Nutzen für die Praxis (speziell Energiebranche)

Der Prosumer soll auf Basis des Modells seine Energiemanagementtätigkeiten besser verstehen, einordnen und für den eigenen Vorteil nutzen können. Es soll das Aufzeigen der Energiemanagementservices, die benötigt werden, unterstützen und Vernetzungspotenziale aufzeigen. Zudem sollen Kooperationsmöglichkeiten erkannt und genutzt werden. Der Prosumer soll in der Lage sein, seine eigenen Energiemanagementtätigkeiten selbständig in die Wertschöpfungskette einzuordnen und die dafür erforderlichen Services kennen und nutzen.

## IV. VORGEHEN ZUR ENTWICKLUNG DES ARTEFAKTS

Die Erarbeitung des zentralen Artefakts folgt dem folgenden Vorgehen. Zunächst werden die Anforderungen in einem Anforderungskatalog zusammengetragen und klassifiziert. Dabei erfolgt das Clustering der Anforderungen in zwei Hauptcluster: allgemeine Anforderungen und spezifische Anforderungen (Transformationsperspektiven).

## A. Das Artefakt muss verschiedene Sichten berücksichtigen

Das Artefakt muss die Transformationsperspektiven unterstützen. Dazu neben der Servicetransformation aus



Kundenperspektive auch die Technologietransformation. Wesentliche Herausforderungen sind dabei die Kundentransformation (aus Konsument wird Prosumer) sowie die Frage, wie neue Technologien in die Modelldarstellung integriert werden können (AG01). Generell ist die Fragestellung der allgemeinen Transformationsdarstellung eine wesentliche Anforderung an das Artefakt. Wie können Änderungen in der Wertschöpfungskette im Artefakt abgebildet werden (AG02)?

Eine weitere Anforderung besteht in der Berücksichtigung von Veränderungszyklen. Der Einfluss von kurzfristigen Änderungen soll nicht zum Systembruch führen (AG03).

Hinsichtlich der Kundenperspektive spielt neben den branchenspezifischen Änderungen generell eine Umbruch statt. Der Paradigmenwechsel weg vom Anbieter hin zum Kunden (anbieterzentriert zu kundenzentriert) ist ein zentraler Aspekt, welcher in allen Branchen Auswirkungen hat ([17], [18]). Aufgrund der Transformationsprozesse steigen auch die Anforderungen der Kunden an ein Energiemanagementsystem (EM-System) [19]

Dieser Paradigmenwechsel führt auch zu Anforderungen an EM-Systeme, sowohl in Business als auch Research Perspektive. Neben den Mindestanforderungen an die Gestaltung von EM-Systemen wie Datenauswertung, Visualisierung von Energieverbräuchen, Berichtswesen und Alerting [20] ergänzen sich Anforderungen hinsichtlich der Transformation. TABLE II. zeigt einen Auszug der identifizierten Anforderungen.

TABLE II. ANFORDERUNGEN AN EM-SYSTEME (AUSZUG)

Lfd. Nr.	Anforderungen
AK01	Integration und Kombination der Prozesse der Energielieferanten mit den Bedürfnissen der Endkunden
AT01	Digitale Unterstützung der Endanwender im Hinblick auf neue für den Endkunden relevante Prozesse aufgrund der neuen Rollen
AG05	Vollständige Abbildung und Unterstützung der Geschäftsprozesse entlang der Wertschöpfungskette
AT02	EMSs must provide advanced and versatile functionality while keeping the installation simple and running cost low [9]
AT03	The systems should integrate with users' daily activities and offer actionable feedback [9]
AS01	...EMS which can measure and manipulate essential system parameters [21]
AK02	Modern EMS's may also be capable of communications with utilities, wheater forecasters, and other [21]

Auf Basis der allgemeinen Anforderungen an das Modell sowie der konkreten Anforderungen an IT-Systeme, im Anwendungsbeispiel an EM-Systeme, werden wesentliche generelle Anforderungen an die zu erarbeitenden Schablone formuliert.

### B. Identifikation der Energieservices

Neben der Ermittlung der Anforderungen ist ein wesentlicher Bestandteil zur Erarbeitung der Schablone die Identifikation der Services. Um die Anforderungen an Service Systeme zu erfüllen, müssen die Services einem Prozess der

Wertschöpfung zugeordnet werden. Bei der Recherche zu Energieservices zeigt sich, dass unterschiedliche Definitionen und unterschiedliche Verständnisse des Begriffs in der wissenschaftlichen Literatur existieren. Fell (2017) fasst diese zusammen und charakterisiert energy services folgendermaßen:

- provide 'benefits' to people,
- are services provided by energy,
- they are what is actually demanded by and delivered to the consumer,
- they result, at least in part from energy conversion in combination with technology (and some definitions only consider them to be energy services when energy is a major input in the provision of a service) [22]

Neben des unterschiedlichen Verständnisses zeigt die Recherche auch einen unterschiedlichen Grad der Beschreibungsebene wie bspw. Energiebeschaffung (Oberbegriff) und Energy Contracting Services (als spezielle Ausprägung). Für das Artefakt muss daher eine geeignete Beschreibungsebene für die Einordnung in der Wertschöpfungskette sowie die Darstellung der Vernetzung aufweisen.

Im Zuge der Arbeit wurden aus der Literatur zunächst Oberbegriffe für die Bezeichnung der Energieservices identifiziert. Diese Aufgabengebiete des betrieblichen Energiemanagements wurden mit Bestandteilen der DIN EN ISO 50001 (Energiemanagementnorm) angereichert. Um die Anforderungen an Prosumer zu erfüllen, wurden diese betrieblichen Aufgaben um Aufgaben von Privatpersonen ergänzt bzw. aufgeteilt.

### C. Identifikation von Designkriterien für Service Konfiguratoren

Da neben dem zentralen Artefakt eine prototypische Umsetzung erfolgen soll, wurden im Rahmen einer Literaturanalyse Designkriterien für Service Konfiguratoren identifiziert:

- address problems, i.e. project risks, both internal and external ones,
- take into account uncertainty and the co-creative nature of the project front-end
- be based on a modular service architecture. [23]

Als zweiten Schritt wird das Vorgehen für die Konstruktion des Service Konfigurators festgelegt. Dabei erfolgt das Design des Konfigurators einem iterativen Prozess. Zunächst wird der bestehende finale Anforderungskatalog an die Schablone um weitere Kriterien ergänzt. Auf Basis dieser Kriterien erfolgt die Auswahl der Service Bündel. Neben den für den Anwendungsfall relevanten Services müssen auch Services speziell für den Konfigurator umgesetzt werden. Nach detaillierter Beschreibung des komplexen Anwendungsfalls und vordefinierten Ereignissen, die eine Transformation

erzeugen sollen, erfolgt die Umsetzung des Service Konfigurators.

Während der Umsetzung erfolgt die konkrete Beschreibung des Feldexperiments.

**D. Erarbeitung der Schablone zur Beschreibung der Service Systeme**

Zur Erarbeitung der Schablone werden die bestehenden Ansätze zur Beschreibung von Service Systemen betrachtet und hinsichtlich der erarbeiteten Anforderungen geprüft. Als Ausgangspunkt der Erstellung der Schablone wird von einem Metamodell ausgegangen. Dies basiert auf den Eigenschaften von Metamodellen (*“integrated view of value constellations, service systems, and service activities”*), welche die Service Analyse und das Design der Prozesse erleichtert [24].

Aus diesem Grund wurde in der ersten Iteration das Metamodell von Alter (2012) mit weiteren Eigenschaften detailliert.

Single Service System			
Customers (Alter, 2012a)		Product/Services (Alter, 2012b)	
<ul style="list-style-type: none"> <li>- External Customers</li> <li>- Internal Customers</li> <li>- Direct Customers</li> <li>- Indirect Customers</li> <li>- Paying Customers</li> <li>- Non-paying Customers</li> <li>- Voluntary Customers</li> <li>- Involuntary Customers</li> <li>- Nonparticipating Executives</li> <li>- Managers in other departments</li> <li>- Automated Agents</li> </ul>		<ul style="list-style-type: none"> <li>- Characteristics that can be measured along separate dimensions that range from product-like to service-like (e.g., degree of customization and extensiveness of customer interaction).</li> <li>- Other attributes for product/service entities include directly measurable performance indicators and also subjective assessments such as quality or customer value.</li> </ul>	
Major Processes and Activities (Alter, 2012b)			
Business, society and government (Barile et al., 2016)			
Resources human: Participants (Alter, 2012b)	Resources informational: Information (Alter, 2012b)	Resources technological: Technologies (Alter, 2012b)	Other types of resources: (Alter, 2012a)
<ul style="list-style-type: none"> <li>- Provider</li> <li>- Customers</li> <li>- Capabilities</li> <li>- Competences</li> <li>- Incentives</li> </ul>			

Fig. 1. Metamodell erste Iterationsstufe

Als ein Element des Gestaltungsrahmen für die Schablone und das zu erarbeitende Modell wird ein Service Konzept gefordert, welches definiert was der Service tun soll und welche Kundengruppe dadurch bedient werden soll. [25]

Nach dieser Detaillierung erfolgt die weitere Iteration und Erweiterung des Modells. Dazu wird das Modell zum einen an zwei Softwareprojekten in der Energiewirtschaft eingesetzt; zum anderen werden Interviews mit Requirements Engineers und IT-Consultants geführt. Die erhaltenen Ergebnisse werden in das Modell eingearbeitet und es folgen weitere Interviews und Anwendungen des Modells.

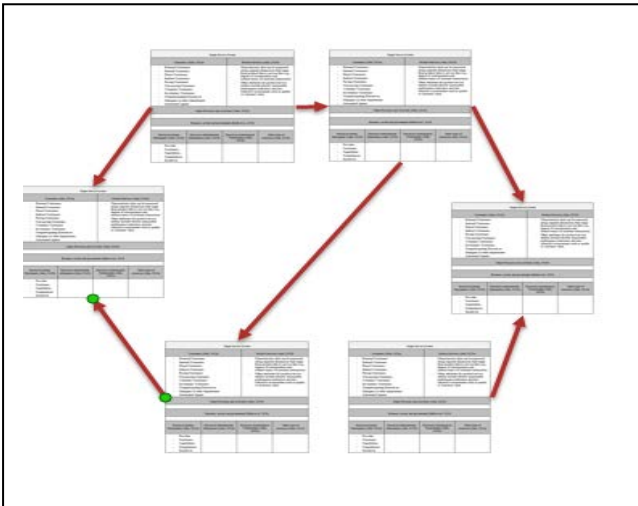


Fig. 2. Beispieldarstellung Abhängigkeit Service Systeme

Ein weiterer Betrachtungsgegenstand ist die Etablierung von Klassifikationsmerkmalen, welche eine Einordnung in die Servicelandkarte hingehend der Architekturschicht (Layer) ermöglichen. Diese Einordnung soll die Abbildung der technischen Transformationsprozesse unterstützen. Als mögliche Einordnung kann die Struktur von Yoo et al. (2010) genutzt werden, die eine Einordnung in devices, networks, services und content vorschlagen [26].

Die Einordnung in verschiedene Betrachtungswinkel stellt eine wesentliche Anforderung an das Modell dar. Es ist von hoher Relevanz einen angemessenen Umfang an Merkmalen aufzuspannen, um die Komplexität beherrschbar zu halten. [27]

Als Abschluss der Designphase erfolgt mithilfe der finalen Schablone die Beschreibung des Anwendungsfalls für den Service Konfigurator. Die nachfolgende Abbildung zeigt schematisch die Artefakte, welche in den Service Konfigurator einfließen.

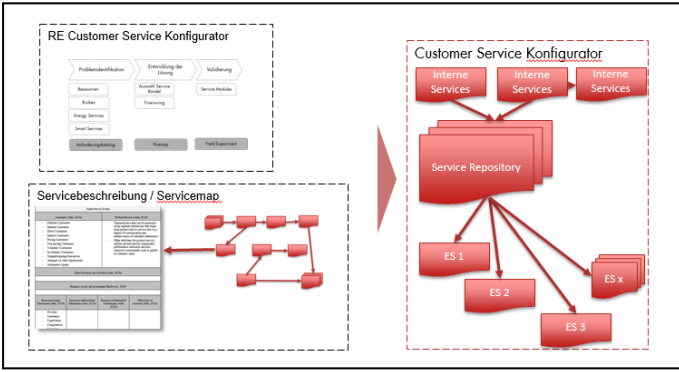


Fig. 3. Metamodell erste Iterationsstufe

Die Ergebnisse aus der Validierung über den Service Konfigurator werden analysiert und entsprechend in der Dissertation berücksichtigt. Auf Basis der Validierung anhand des Prototyps soll das Modell weiterentwickelt und die Anwendbarkeit hinsichtlich weiterer Anwendungsfälle untersucht werden.

## V. ZUSAMMENFASSUNG UND AUSBLICK

In diesem Beitrag wurde das Vorgehen des Dissertationsprojekts beschrieben. Zunächst wurde das Thema in den Gesamtkontext eingeordnet und das Forschungsvorgehen beschrieben.

Zunächst erfolgte die Vorstellung des Vorgehens zur Entwicklung des Artefakts. Basierend auf der Ermittlung der Anforderungen an verschiedene Sichten, wurde im zweiten Schritt die Identifikation der Services beschrieben. Das Vorgehen zur Entwicklung der Schablone für das Modell zeigt den iterativen Prozess der Erstellung.

Das Promotionsvorhaben befindet sich aktuell in der Designphase. Aufgrund der noch auszuwertenden zweiten Iterationsphase wurden für die Ergebnisrepräsentation in diesem Beitrag die initialen Modelle verwendet.

## REFERENCES

- [1] Bruhn, M., & Hadwich, K. (2016). *Service transformation*. (M. Bruhn & K. Hadwich, Eds.) *Service transformation*. Wiesbaden: Springer Fachmedien Wiesbaden. <http://doi.org/10.1007/978-3-658-11097-0>
- [2] Huener, U., & Bez, M. (2015). Erneuerbare Energien als Grundlage für Prosumer-Modelle. In C. Herbes & C. Friege (Eds.), *Marketing Erneuerbarer Energien: Grundlagen, Geschäftsmodelle, Fallbeispiele* (pp. 335–358). Wiesbaden: Springer Fachmedien Wiesbaden. [http://doi.org/10.1007/978-3-658-04968-3\\_15](http://doi.org/10.1007/978-3-658-04968-3_15)
- [3] Warg, M., Weiß, P., & Engel, R. (2015). Service Dominierte Architektur (SDA): die digitale Transformation erfolgreich meistern
- [4] Alt, R., & Puschmann, T. (2016). *Digitalisierung der Finanzindustrie: Grundlagen der Fintech-Evolution* (Vol. 85). <http://doi.org/10.1007/978-3-662-50542-7>
- [5] Doleski, O. D. (2017). Die Energiebranche am Beginn der digitalen Transformation: aus Versorgern werden Utilities 4.0. In O. D. Doleski (Ed.), *Herausforderung Utility 4.0: Wie sich die Energiewirtschaft im Zeitalter der Digitalisierung verändert* (pp. 3–27). Wiesbaden: Springer Fachmedien Wiesbaden.
- [6] Bardt, H., Chrischilles, E., Growitsch, C., Hagspiel, S., Schaupp, L., Hagspiel, S., ... Chrischilles, E. (2014). Eigenerzeugung und Selbstverbrauch von Strom – Stand, Potentiale und Trends. *Z Energiewirtschaft*, 38, 83–99. <http://doi.org/10.1007/s12398-014-0133-0>
- [7] Krickel, F. (2015). Digitalisierung in der Energiewirtschaft. In W. Hecker, C. Lau, & A. Müller (Eds.), *Zukunftsorientierte Unternehmenssteuerung in der Energiewirtschaft* (pp. 41–73). Wiesbaden: Springer Fachmedien Wiesbaden. [http://doi.org/10.1007/978-3-658-07816-4\\_2](http://doi.org/10.1007/978-3-658-07816-4_2)
- [8] Winkels, L., Schmedes, T., & Appellrath, H. J. (2007). Dezentrale Energiemanagementsysteme. *Wirtschaftsinformatik*, 49(5), 386–390. <http://doi.org/10.1007/s11576-007-0083-9>
- [9] Aman, S., Simmhan, Y., & Prasanna, V. K. (2013). Energy management systems: state of the art and emerging trends. *IEEE Communications Magazine*, 51 (January), 114–119.
- [10] Schwieters, N., Hasse, F., von Perfall, A., Maas, H., Willms, A., & Lenz, F. (2016). Deutschlands Energieversorger werden digital.
- [11] Herzog, A., Hrach, C., Schubert, N., & Alt, R. (2014). Aufgabengebiete und Entwicklungspotenziale von Energiemanagementsystemen. In D. Kundisch, L. Suhl, & L. Beckmann (Eds.), *MKWI 2014 - Multikonferenz Wirtschaftsinformatik: 26. - 28. Februar 2014 in Paderborn* (pp. 921–933). Paderborn
- [12] Peffers, K., Tuunanen, T., Rothenberger, M. a., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45–77. <http://doi.org/10.2753/MIS0742-1222240302>
- [13] vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A., ... Reimer, K. (2009). Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. In *17th European Conference on Information Systems* (Vol. 9, pp. 2206–2217). <http://doi.org/10.1108/09600031211269721>
- [14] Peters, C., Maglio, P., Badinelli, R., Harmon, R. R., & Maull, R. (2016). Emerging Digital Frontiers for Service Innovation. *Communications of the Association for Information Systems*, 39(Article 8), 136–149. Retrieved from <http://aisel.aisnet.org/cais/vol39/iss1/8>
- [15] Lusch, R. F., Vargo, S. L., & Gustafsson, A. (2016). Fostering a trans-disciplinary perspectives of service ecosystems. *Journal of Business Research*, 69(8), 2957–2963. <http://doi.org/10.1016/j.jbusres.2016.02.028>
- [16] Böhm, T., Marco Leimeister, J., Möslein, K., Leimeister, J., & Möslein, K. (2014). Service Systems Engineering. *Business & Information Systems Engineering*, 6(2), 73–79. <http://doi.org/10.1007/s12599-014-0314-8>
- [17] Fragkidakis, G. (2016). *The value perspective in the analysis and design of service systems*. Université de Genève. Retrieved from <https://archive-ouverte.unige.ch/unige:89306>
- [18] Sachse, S. (2017). Customer-centric Service Management Conceptualization and Evaluation of Consumer-induced Service Composition. Dissertation Universität Leipzig
- [19] Bunse, K., Vodicka, M., Schönsleben, P., Brühlhart, M., & Ernst, F. O. (2011). Integrating energy efficiency performance in production management - Gap analysis between industrial needs and scientific literature. *Journal of Cleaner Production*, 19(6–7), 667–679. <http://doi.org/10.1016/j.jclepro.2010.11.011>
- [20] Schellong, W. (2016). *Analyse und Optimierung von Energieverbundsystemen*. Springer Vieweg, Berlin, Heidelberg. <http://doi.org/10.1007/978-3-662-49463-9>
- [21] Treado, S. J., & Holmberg, D. (2010). Holmberg „Energy Systems Management and Greenhouse Gas Reduction“, *ASHIRE Transactions*, 2010, 358-364
- [22] Fell, M. J. (2017). Energy services: A conceptual review. *Energy Research & Social Science*, 27, 129–140. <http://doi.org/10.1016/j.erss.2017.02.010>
- [23] Luotola, H., & Wikstro, R. (2016). The value of project execution services: a problem and uncertainty perspective. *Construction Management and Economics*, 34(4–5), 272–285. <http://doi.org/10.1080/01446193.2016.1151062>
- [24] Alter, S. (2012). Metamodel for service analysis and design based on an operational view of service and service systems. *Service Science*, 4(3), 218–235. <http://doi.org/10.1287/serv.1120.0020>
- [25] Chew, E. K. (2014). Service Innovation for the Digital World. *Enterprise Modelling and Information Systems Architectures*, 9(1), 70–89. <http://doi.org/10.18417/emisa.9.1.5>
- [26] Yoo, Y./Henfridsson, O./Lyytinen, K. (2010): Research Commentary - The New Organizing Logic of Digital Innovation- An Agenda for Information Systems Research, in: *Information Systems Research*, Vol. 21, No. 4, pp. 724-735
- [27] Senn, T. (2017). Service Transition in the Context of Digitized Service Ecosystems. In M. Bruhn & K. Hadwich (Eds.), *Dienstleistungen 4.0: Geschäftsmodelle - Wertschöpfung - Transformation. Band 2. Forum Dienstleistungsmanagement* (pp. 241–267). Wiesbaden: Springer Fachmedien Wiesbaden. [http://doi.org/10.1007/978-3-658-17552-8\\_10](http://doi.org/10.1007/978-3-658-17552-8_10)

# Qualitätssicherung elektronischer Prüfungsverfahren

## Integration der Item-Response-Theorie zur Testauswertung in Lern-Management-Systeme und Abgleich mit der klassischen Testtheorie

Christoph Jobst  
Institut für Wirtschaftsinformatik  
Universität Leipzig  
cjobst@wifa.uni-leipzig.de

**Abstract**—Die Item-Response-Theorie ist der klassischen Analyse von Persönlichkeits- und Leistungstests messtheoretisch überlegen, findet aber in der allgemeinen Testpraxis kaum Anwendung. Ihr empirischer Nutzen wird in einer Literaturanalyse bestätigt und ihre Integrierbarkeit in Lern-Management-Systeme aus psychometrischer und technischer Sicht untersucht. Auf Basis der Resultate wird ein Prototyp erstellt und validiert.

**Keywords**—E-Assessment, Item-Response-Theorie, Lern-Management-System

### I. EINLEITUNG

In Bildungseinrichtungen und Unternehmen finden zunehmend Lern-Management-Systeme (LMS) für Lehre und Prüfung Anwendung. Die integrierte teststatistische Unterstützung elektronischer Prüfungsverfahren in diesen und anderen prüfungstauglichen Systemen basiert, bis auf spezialisierte Ausnahmen, auf der klassischen Testtheorie (KTT). Die modernere Item-Response-Theorie (IRT) ist der KTT messtheoretisch überlegen, findet aber insbesondere aufgrund ihrer Komplexität und unzureichenden Integration in die Prüfungssysteme unter psychometrischen Laien kaum Anwendung.

Es wird untersucht, ob eine einheitliche Integration der IRT für verschiedene Testformen in LMS aus psychometrischer und technischer Perspektive sinnvoll und möglich ist. Ein auf Basis eines Integrationsrahmenwerks entwickelter Prototyp dient der Validierung des dabei erstellten Konzepts und soll IRT-basierte Auswertungen für Endanwender zugänglich machen sowie der KTT gegenüberstellen. Dabei wird gezeigt, dass die IRT, sofern die Bedingungen für ihren Einsatz erfüllt sind, detailliertere Informationen liefern kann oder zumindest mit der KTT vergleichbare Erkenntnisse generiert, wobei sie methodeninhärent ein messtheoretisch besseres Fundament für die weitere Datenverarbeitung bietet.

Die Schaffung und Dokumentation einer quelloffenen Basis für IRT-basierte Auswertungen in LMS ist eine erweiterbare Grundlage, welche die Nutzung messtheoretischer Vorteile der IRT, etwa hinsichtlich adaptiver Tests, der Aggregation von Kennwerten in Itempools und der Vergleichbarkeit von Probanden über variierende Testformen hinweg vereinfacht.

### II. TESTTHEORIEN

#### A. Einführung

Viele physikalische Eigenschaften, wie beispielsweise die Temperatur, können deterministisch vermessen werden. Psychische Eigenschaften hingegen, etwa Depression oder Intelligenz, bieten diese Möglichkeit nicht und benötigen statistische Messmethoden, die von den psychologischen Testtheorien beschrieben werden.

Eine Testtheorie beschreibt sowohl die Verbindung zwischen manifesten und latenten Variablen als auch die Anforderungen und folglich aus diesen ableitbare Gütekriterien an den zur Erhebung genutzten Test [1,2,3]. Manifeste Variablen sind die beobachtbaren Reaktionen von Probanden auf Items (Fragen, Aufgaben und Stimuli). Latente Variablen, auch Konstrukte genannt, liegen den beobachteten Werten zugrunde und sollen deren Kontingenz erklären. Sie können nicht direkt gemessen, sondern nur durch Schlussfolgerung aus einem angenommenen, statistischen Zusammenhang mit den manifesten Variablen gewonnen werden. Es existieren mit der KTT und der IRT zwei testtheoretische Strömungen.

#### B. Klassische Testtheorie

Testvorgänge vor der KTT waren vorwiegend durch Pragmatismus geprägt und kaum mathematisch fundiert [1]. Die KTT wurzelt im beginnenden 20. Jahrhundert und basiert auf einem naturwissenschaftlich geprägten Messverständnis. Jeder Messvorgang spiegelt demnach die wahre Merkmalsausprägung wider und wird durch einen zufälligen Messfehler verunreinigt. Das Ergebnis sollte daher umso präziser werden, je häufiger gemessen wird, da sich der angenommene Messfehler herausmittelt. Die erreichte Anzahl korrekt beantworteter Items zeigt somit die reale Merkmalsausprägung in Relation zu den anderen Probanden des Tests. Die Annahmen der KTT lassen sich in fünf Axiomen formulieren [2]:

- Existenzaxiom – Es existiert ein wahrer Wert pro Proband und Item als Erwartungswert der Messung.
- Verknüpfungaxiom – Der Messwert setzt sich aus dem wahren Wert und einem Messfehler zusammen.

- Unabhängigkeitsaxiom – Wahrer Wert und Fehler sind unkorreliert.
- Lokale stochastische Unabhängigkeit der Items – Die Beantwortung eines Items hat keinen Einfluss auf die Beantwortung eines anderen Items.
- Lokale stochastische Unabhängigkeit der Personen – Der Messfehler für ein Item ist zwischen unterschiedlichen Personen unkorreliert.

Bewusst oder auch unbewusst wird die Erfüllung dieser Axiome üblicherweise vorausgesetzt, wenn beispielsweise Prüfungen in Bildungseinrichtungen ausgewertet werden. Die KTT hat allerdings gravierende messtheoretische Mängel, die sie als korrektes Messmodell disqualifizieren [1,2,4]:

- Die Axiome der KTT sind nicht falsifizierbar.
- Die KTT geht durch ihre Axiomatik von stabilen oder linear veränderlichen Merkmalen aus, die unbeeinflusst von Zeit und Situation sind.
- Unabhängig von den Daten geht die KTT von einem Intervallskalenniveau und einer Normalverteilung aus.
- Sämtliche Kennzahlen, von der Itemanalyse bis hin zu den Gütekriterien, sind stichproben- und testabhängig.

Die KTT hat die psychologische Diagnostik auf den Weg eines objektiven und mathematisch begründeten Vorgehens geführt. Es gibt Hilfestellungen für die Bestimmung der Gütekriterien, die Testkonstruktion, die Normierung und die Testauswertung. Die meist einfache Anwendbarkeit der Methoden und die pragmatischen, messtheoretisch problematischen, aber dennoch verwertbaren Ergebnisse, scheinen die bekannten Nachteile oft zu überwiegen. Die mehrheitliche Zahl von Tests wurde und wird durch die KTT bestimmt [1,5,6].

### C. Item-Response-Theorie

Die IRT kam in Europa [7] und den USA [8] insbesondere seit den 1960er Jahren auf und wurde von den Nachteilen der KTT angetrieben. Sie ist ein Sammelbegriff für beliebige Messmodelle, die berücksichtigen, dass der Ausgang einer Messung von unüberblickbar vielen Randbedingungen abhängt und somit probabilistisch ist. Selbst bei exakter Kenntnis des latenten Merkmals folgt jede Manifestation lediglich einer Auftretenswahrscheinlichkeit und im umgekehrten Fall erlauben etwaige Testwerte lediglich eine Wahrscheinlichkeitsaussage zur latenten Variable [1].

Das Rahmenwerk der IRT bietet damit gegenüber der KTT vorteilhafte Messeigenschaften. Es wird angenommen, dass jeder Proband eine bestimmte Merkmalsausprägung besitzt, deren Schätzwert nicht von den vorgelegten Items beeinflusst wird. Auch die Eigenschaften der Items werden als stabil innerhalb einer Population angenommen, da diese nicht in Abhängigkeit der Merkmalsverteilung in der jeweils getesteten Probandenstichprobe beispielsweise leichter oder schwerer werden, sondern entsprechend ihrer Eigenschaften jeweils eigene Messinformationen für einen bestimmten Bereich der Merkmalsausprägung liefern. Schreibe also ein Proband in einer anderen Probandengruppe oder erhalte er einen anderen

Itemmix, sollten sich weder die Schätzung seines latenten Merkmals, noch die Eigenschaften der genutzten Items ändern. Diese Entkoppelung der Personen- und Itemeigenschaften vom konkreten Test erreicht die IRT durch Parametrisierung. Das ermöglicht für einige Modelle die Stichprobenunabhängigkeit der Parameterschätzungen innerhalb einer Population (spezifische Objektivität), was die skalenkonforme Zusammenführung und fortlaufende Präzisierung von Itemeigenschaften in einem Itempool über unterschiedliche Tests und Probandengruppen hinweg vereinfacht. Die Merkmalsausprägungen von Probanden und Eigenschaften von Items können einzeln auf ihre Plausibilität geprüft werden (*person-fit*, *item-fit*), statt, wie in der KTT, durch einen stichprobenabhängigen Wert gemeinsam beurteilt zu werden und sie sind zudem zwischen verschiedenen Testformen und Probandengruppen vergleichbar. Die Gültigkeit der IRT-Modelle ist auch für die damit auszuwertenden Daten insgesamt falsifizierbar (*model-fit*). Durch die Auswahl der je nach individuellem Antwortmuster informativsten Items können adaptive Tests auch ohne Einbußen bei der Reliabilität kürzer ausfallen.

Je nach vermessenem Konstrukt, Datenlage und Messsituation bietet die IRT verschiedene und kombinierbare Modellklassen an. Beispielsweise für:

- Ein- oder mehrdimensionale Konstrukte
- Dominanz- (korrekte Antwortwahrscheinlichkeit verläuft monoton) oder Idealwertkonstrukte (korrekte Antwortwahrscheinlichkeit ist bei mindestens einer bestimmten Merkmalsausprägung erhöht)
- Dichotome (in oder gegen die Schlüsselrichtung geantwortet) oder polytome (Teilpunktvergabe) Antwortformate
- Nominal-, ordinal- oder intervallskalierte Daten
- Lokal unabhängige oder abhängige Items (Testlets)
- Längs- oder Querschnittsdiagnosen
- Berücksichtigung der Antwortzeit oder sonstiger Einflussfaktoren

Mithilfe eines Schätzalgorithmus werden Parameter zur Beschreibung einer Modellgleichung berechnet, die den funktionalen Zusammenhang der Itemeigenschaften zur Antwortwahrscheinlichkeit herstellt. Die bekanntesten Modelle nutzen für die so berechneten *item characteristic curves* (ICC) eine logistische Funktion. Diese ist leichter verarbeitbar als Polynome oder gar Polynomzüge und folgt im Gegensatz zu stufenförmigen oder linearen Gleichungen der psychologischen Erkenntnis, dass markante Änderungen in den Randbereichen eines Messspektrums schwerer herbeizuführen sind, als in dessen Mittelbereich [1]. Durch das Antwortmuster kann die Verortung von Probanden auf dem Merkmalskontinuum ebenfalls mittels eines Schätzalgorithmus festgestellt werden.

Eines der frühesten IRT-Modelle ist das Rasch-Modell von 1960 [7]:

$$P(X_{iv} = 1 | \theta_v, \beta_i) = \frac{\exp(\alpha(\theta_v - \beta_i))}{1 + \exp(\alpha(\theta_v - \beta_i))} \quad (1)$$

Es erfordert einen eindimensionalen Test mit unabhängigen Items und einem dichotomen Antwortformat. Bei ihm hängt die Lösungswahrscheinlichkeit eines Items  $i$  durch Proband  $v$  von der Fähigkeit des Probanden  $\theta_v$  und dem Lokationsparameter  $\beta_i$  des Items ab. Der Parameter  $\beta_i$  entspricht der Schwierigkeit und zeigt in diesem Modell den Punkt des Merkmalskontinuums, an dem die korrekte Antwortwahrscheinlichkeit für dieses Item 50% beträgt. Der Parameter  $\alpha$  bestimmt den Anstieg der Kurven und entspricht in seiner Interpretation der Trennschärfe. Im Rasch-Modell wird dieser Parameter auf  $\alpha = 1$  fixiert. Die grafische Aufbereitung der Modellformel stellt sich beispielsweise für drei Items mit Schwierigkeiten von  $\beta_i = \{-1, 0, 1\}$  wie folgt dar:

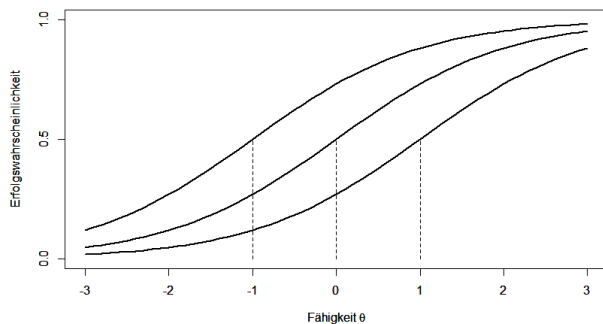


Fig. 1. Beispieldarstellung für drei Items im Rasch-Modell

Sofern die Modellgleichung einen passenden Zusammenhang zwischen der Antwortwahrscheinlichkeit und der Merkmalsausprägung herstellt, kann diese beliebig gewählt werden. Populäre Varianten für dichotome und eindimensionale Formate sind Birnbaums PL-Modelle (*parameter logistics*) [8], die neben der Schwierigkeit auch Analogien zur Trennschärfe und zur Ratewahrscheinlichkeit bieten. Für polytome Daten sind insbesondere die folgenden Modelle populär:

- *Graded response model* (GRM) [9]
- *Nominal response model* (NRM) [10]
- *Rating scale model* (RSM) [11,12]
- *Partial credit Model* (PCM) [13]
- *Generalized partial credit model* (GPCM) [14]

Die Modelle unterscheiden sich im angenommenen kognitiven Prozess, in den hintergründigen Berechnungsmethoden, im Skalenniveau der Messdaten und in den berechneten Itemparametern.

Viele große Testprogramme gingen im Verlauf der Jahrzehnte von der KTT zur IRT über [15]. Ein prominentes Beispiel für ihren Einsatz sind die PISA-Studien, die dadurch Daten verschiedener Länder auf einer gemeinsamen Skala abbilden und somit einfacher vergleichen können [15,16]. Zumindest innerhalb der Psychologie gehöre die IRT inzwischen zum „Mainstream“, würde aber vom eigenen Erfolg behindert, da dem Lehrbetrieb kaum Absolventen erhalten bleiben [17]. Eine Befragung von Leitern psychologischer

Studienprogramme im Jahr 2017 ergab, dass die Lehrangebote in Verantwortung von 70% der Befragten keine IRT beinhalten, wobei der erstgenannte Grund fehlende Expertise beim verfügbaren Personal war [18]. Die Befragung erstreckte sich auch auf Psychometriker im praktischen Einsatz. Im Schnitt schätzten alle Befragten die IRT als wichtiges Instrument ein. Kritiker gaben an, dass die IRT in Relation zu ihren Vorteilen ebenso wie die dafür gedachte Software zu kompliziert wäre. Der Schlüssel für eine weitere Verbreitung der IRT sei daher ein Ausbau der Lehrangebote und nutzerfreundliche Software.

Für weitere Erläuterungen zur IRT im Allgemeinen und den Modellen im Speziellen sei auf die grundlegenden Überblickswerke [17,19] verwiesen.

### III. VORGEHEN

Die Promotionsziele werden mit einem empirisch-normativen Vorgehen in sechs Schritten umgesetzt:

- **Grundlage** – Untersuchung des empirischen Nutzens der IRT gegenüber der KTT mittels einer strukturierten Literaturanalyse
- **Funktionale Lücke** – Untersuchung von LMS hinsichtlich ihrer teststatistischen Unterstützung
- **Psychometrische Machbarkeit** – Diskussion und Auswahl einer einheitlichen IRT-gestützten Auswertung für Tests in LMS
- **Technische Machbarkeit** – Analyse bestehender Software hinsichtlich ihrer Integrierbarkeit in Websysteme
- **Prototyp** – Instanziierung des Integrationsrahmens für die endanwendertaugliche Bereitstellung von IRT-gestützten Auswertungen
- **Validierung** – Durchführung einer Fallstudie mit Dozenten zur Prüfung des praktischen Nutzens der Analysen für die Testoptimierung anhand von echten Klausuren

### IV. LITERATURANALYSE

Die einzig auffindbare Literaturanalyse mit Fokus auf empirische Arbeiten, in denen IRT und KTT gleichzeitig angewandt und deren Ergebnisse verglichen wurden, liefert [20]. Ihre Forschungsfrage ist, in welchem Ausmaß die Eigenschaften der Items und Probanden übereinstimmen und ob es empirische Belege für praktische Unterschiede zwischen den Theorien gibt. Die Analyse umfasst 16 Publikationen, davon 15 mit echten Daten und eine mit simulierten Daten. Die Literatursammlung wurde nicht dokumentiert. Es wird resümiert, dass die Testanalyseergebnisse von IRT und KTT beinahe deckungsgleich sind und es abseits der theoretischen Unterschiede keinen empirischen Beweis für die Überlegenheit des einen Ansatzes über den anderen gibt.

Da die Analyse in [20] lediglich 16 scheinbar willkürlich gewählte Publikationen berücksichtigt, soll eine eigene mit weiterer Orientierung angefertigt werden. Dies erfolgt über eine Methodentriangulation aus Stichwort-, Journal- und

Referenzsuche. Um als gültiger Treffer eingestuft zu werden, muss eine Publikation die KTT und mindestens ein IRT-Modell anwenden und die Ergebnisse interpretieren. Die Suche umfasst vergleichende Studien mit echten und simulierten Daten sowie die Auswertung konkreter Tests ohne vordergründige Vergleichsintention mithilfe beider Theorien. Erfasst wird, ob es sich um echte oder simulierte Daten handelt, welche IRT-Modelle genutzt werden, welche Theorieeigenschaft untersucht oder Auswertungsintention verfolgt wird, welche Methoden für die Gegenüberstellung zum Einsatz kommen und welches Fazit hinsichtlich der Theorien gezogen wird.

Die Stichwortsuche verwendet drei Datenbanken: Die kuratierte Datenbank des *Education Resources Information Center*, die Datenbank von Springer aufgrund der Indexierung von zahlreichen bildungswissenschaftlichen Journals und den aggregierten Datenbestand der Leipziger Universitätsbibliothek. Der Suchterm lautet:

*compar\* AND (IRT OR "item response theory") AND (CTT OR "classical test theory")*

Der Suchbereich wurde jeweils auf die Bildungswissenschaften und die englische Sprache beschränkt, um den Analyseaufwand zu begrenzen und gleichzeitig den inhaltlichen Fokus der Gesamtarbeit zu wahren. Insbesondere eine Ausweitung auf die allgemeine Psychologie würde die Anzahl der Treffer stark erhöhen. Auch in medizinischen Journals finden sich zahlreiche Publikationen, in denen die IRT bei der Instrumentenanalyse, der Diagnose von psychischen und physischen Erkrankungen sowie der Therapieevaluation zum Einsatz kommt.

Die Auswahl für die Journalsuche basiert auf einer Liste qualitativer und quantitativer Publikationskanäle der *American Psychological Association* [21]. Die Vorwärts- und Rückwärtsreferenzsuche beschränkt sich auf drei Arbeiten mit grundlegender Theorievergleichsabsicht aus dem Zeitraum der 1990er, 2000er und 2010er.

Es wurden 172 Treffer der Stichwortsuche gesichtet, acht Journals durchsucht und für drei Arbeiten eine Referenzsuche durchgeführt. Insgesamt erfüllten 104 Publikationen die Suchkriterien. Unter den gültigen Treffern nutzen 22 simulierte und 80 echte Daten. Zwei Publikationen greifen auf beide Datenformen zurück. Die positiv beschiedenen Treffer pro Jahr verteilen sich wie folgt:

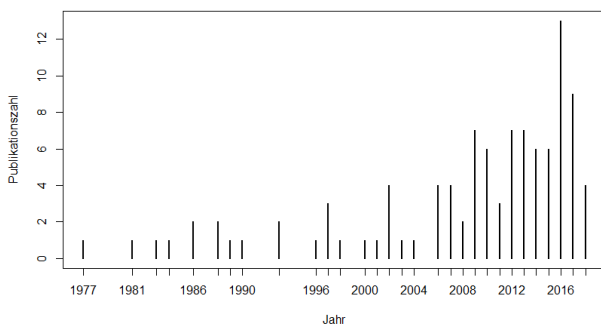


Fig. 2. Zeitliche Einordnung der berücksichtigten Publikationen

Die Simulationsstudien untersuchen die Richtigkeit der Messergebnisse, die Parameterinvarianz zwischen generierten Stichproben, die Robustheit gegenüber modellverletzenden Situationen sowie die Übereinstimmung der IRT-Analyse mit der KTT-Analyse. Ihr Fazit lautet weitgehend, dass die IRT- und KTT-Ergebnisse allgemein eine hohe Korrelation aufweisen, allerdings die mittels IRT berechneten Merkmalsausprägungen weniger Klassifikationsfehler aufweisen und die Trennschärfe der Items präziser ist. Eine Invarianz der Itemeigenschaften konnte auch für die KTT festgestellt werden, wobei die IRT robuster gegenüber anomalen Messsituationen ist. Es wird von einigen empfohlen, IRT und KTT parallel oder als Hybrid anzuwenden, um einerseits zusätzliche Evidenz für die einzelnen Analyseresultate zu erhalten und andererseits die mittels KTT auf Testebene erlangten Ergebnisse durch die IRT zu bestätigen sowie auf Personen- und Itemebene zu detaillieren.

Auch unter den Studien mit echten Daten sind Theorievergleiche enthalten, deren Ergebnisse jene der Simulationsstudien stützen. Die Testanalysen ohne Vergleichsabsicht nutzen die KTT meist zur Voruntersuchung, um besonders auffällig schwere oder leichte Items herauszufiltern und die sich anschließende Untersuchung des *model-fit* vorzubereiten. Die IRT kommt dabei als Hauptanalyseinstrument zum Einsatz. Weiterhin werden beide Theorien genutzt, um mit ihren jeweiligen Methoden die Reliabilität des Tests zu untersuchen. Die KTT liefert einen Wert für den Test als Ganzes und die IRT erzeugt Kurven, welche die verfügbare Messinformation und den Messfehler im Merkmalskontinuum aufzeigen. Dadurch wird deutlich, um welche Items ein Test gekürzt werden könnte und in welchem Merkmalsbereich neue Items für eine präzisere Messung hinzugefügt werden sollten. Weiterhin kommen beide Theorien auch unabhängig voneinander zum Einsatz, um jeweils potentiell problematische Items zu entdecken. Auch im praktischen Einsatz sprechen sich viele Autoren für eine Parallelanwendung aus. Ebenfalls konstatieren einige, dass aufgrund der Ähnlichkeit des Erkenntnisgewinns die KTT zu bevorzugen wäre, weil die IRT zu aufwändig und die bislang verfügbare Software nicht nutzerfreundlich genug sei.

Die Literaturanalyse verstärkt die Evidenz, dass die IRT eine sinnvolle Ergänzung der Testanalysemethoden darstellt und präzisere sowie umfangreichere Informationen als die KTT bereitstellen kann.

## V. FUNKTIONALE LÜCKE IN LMS

Der Einsatz von LMS in Bildungseinrichtungen nimmt stetig zu, wie eine internationale Statistik von [22] seit dem Jahr 2000 zeigt. Demnach nutzten 74% der 2017 erfassten höheren Bildungseinrichtungen Deutschlands ein LMS. Unter den erfassten Ländern Europas nimmt Deutschland damit den zweiten Platz nach Großbritannien (77%) ein. Die 2018 mit 50% der von den gleichen Studienmachern erfassten LMS-Installationen international verbreitetste Plattform ist Moodle, gefolgt von Blackboard mit 19%. In Europa ist hinter diesen ILIAS präsent, das bei einer Erhebung im Jahr 2013 das meistinstallierte LMS in Deutschland war [23,24].

Ein modernes LMS soll keine „PDF-Schleuder“ sein, sondern die Lehre in möglichst vielen Aspekten unterstützen.

Dazu zählen neben der Bereitstellung von Lernmaterial auch Organisation, Durchführung und Kontrolle von Lernaktivitäten. Zudem sind interaktive und kollaborative Elemente möglich. Bei allen Aktionen im LMS können Daten der Nutzer gesammelt werden, sofern der Datenschutz dies nicht untersagt, und im Rahmen von *learning analytics* Verwendung finden. Die daraus gewonnenen Erkenntnisse dienen der Verbesserung des Lehrangebotes und der individuellen Förderung. Tests stellen das Hauptinstrument zur Lernerfolgsmessung dar und sollten bei der Konstruktion, der Durchführung und der Auswertung einer entsprechenden Qualitätssicherung unterliegen. Für alle drei Bereiche spielt die Teststatistik eine wichtige Rolle, da sie die Auswahl von Items für einen Test, in adaptiven Szenarien den konkreten Testablauf und nach dem Test die Bewertung der Probanden beeinflussen kann.

Der funktionale Stand der teststatistischen Auswertungsunterstützung wurde in den drei genannten LMS untersucht. Sie unterscheiden sich nur geringfügig. Die Gemeinsamkeiten belaufen sich auf:

- Anzahl der Testdurchläufe
- Median und Durchschnitt der erreichten Punkte
- Schwierigkeitsindex
- Trennschärfe
- Standardabweichung
- Antwortübersicht zur Distraktorenanalyse

Moodle sticht durch zusätzliche Informationen hervor, die über eine bloße Datenaggregation hinausgehen. Blackboard bereitet die Zahlen in einer Zusammenfassung mit einer verbalisierten Einteilung auf und ermöglicht die Anpassung problematischer Items im aktiven Test, was allerdings hinsichtlich der Vergleichbarkeit von Testdurchläufen problematisch sein kann. ILIAS überträgt als einziges System zumindest einen geringfügigen Teil der Informationen in den Itempool, wo sie für die Zusammenstellung neuer Tests nutzbar sind. Durch die fehlende Verwendbarkeit für Such-, Filter- oder Sortiervorgänge sind aber selbst diese wenigen Daten kaum praxisrelevant. ILIAS besitzt damit aber zumindest eine ausbaufähige Grundstruktur zur Übertragung teststatistischer Daten.

Alle drei Systeme ermöglichen für Items die manuelle Zuteilung von Metadaten, Tags oder Taxonomien, was in Moodle lange Zeit ohne Nutzen war, da der Itempool vor der Veröffentlichung von Version 3.5 im Mai 2018 weder Suche noch Filter bot und auch seitdem nur die Filterung von selbst gesetzten Tags ermöglicht. Dennoch besitzt Moodle im Vergleich die besten strukturellen Voraussetzungen für eine plattformweite Itempoolverwaltung, weil Items zu Kategorien, die den Verzweigungen und Blättern des baumartigen Aufbaus der Lernressourcen im LMS entsprechen, zugeordnet werden können. Da ein komplexes Rechtesystem auf Item- und Kategorieebene, ein nutzbringender Einsatz von Metadaten und eine Integration von Qualitätssicherungsmaßnahmen über die Präsentation der Teststatistik hinaus fehlen, kann der systemweite Itempool jedoch nicht ausgereizt werden. In ILIAS befinden sich beliebig viele Itempools als separate Lernobjekte

in Kursen und Kategorien, deren Verwendbarkeit über das Zugriffsrecht auf die enthaltende Struktur geregelt wird. In Blackboard können pro Kurs mehrere Itempools angelegt werden, die zur Wiederverwendung in andere Kurse importiert werden müssen.

Die Untersuchung der drei LMS ergab, dass sie ausschließlich klassisch orientierte Testanalysen anbieten, diese nicht oder nur marginal mit dem Itempool verbinden und auch allgemein nur grundlegende Unterstützungsfunktionen für den Testprozess bieten. Die Weiterentwicklung von ILIAS wird unter Mitwirkung der Universität Leipzig vorangetrieben. Items sollen eine installationsübergreifend eindeutige Identifikationsnummer und eine Versionierung erhalten, an welche die (bisher klassischen) teststatistischen Kennzahlen gebunden werden können. Parallel sollen die Items zusätzliche Metadaten zur Abbildung des Itemlebenszyklus erhalten, worauf ein Review-Verfahren aufsetzen kann. Um für die Testzusammenstellung, den Ablauf und die Auswertung auch die Vorteile der IRT nutzbar zu machen, ist ihre Integration eine sinnvolle Ergänzung.

## VI. UNTERSUCHUNG DER PSYCHOMETRISCHEN MACHBARKEIT

Tests in LMS sind für die konfigurationsgemäße Präsentation von Items, die geregelte Entgegennahme von Reaktionen der Probanden und die Unterstützung der Auswertung im Rahmen einer Querschnittsdiagnose zuständig. Der Vergleich mehrerer Durchläufe kann Grundlage einer Längsschnittsdiagnose sein. Mit LMS-Tests können diagnostische, formative, summative und qualitätssichernde Szenarien abgedeckt werden. Eine einheitliche psychometrische Herangehensweise erleichtert die Zusammenführung von Analysedaten in einem Itempool und den Vergleich verschiedener Tests. Neben dem Szenario sind für die Auswahl der Methoden auch die Konfiguration des Tests und die Beschaffenheit der Items ausschlaggebend.

Grundlage der Detailbetrachtungen ist ILIAS, weil es für die Universität Leipzig durch den Einsatz als elektronische Prüfungsplattform lokale Relevanz besitzt und da bereits einschlägige Praxis- und Entwicklungserfahrung in dessen „*Test & Assessment*“-Komponente vorliegt. Verschiedene IRT-Modelle werden anhand der psychometrisch relevanten Testkonfigurationsoptionen diskutiert, die für eine einfachere Übertragbarkeit auf andere Systeme gruppiert wurden.

### A. Steuerung des Testablaufs

Mehrere Durchläufe pro Person werden im Bildungsbereich vor allem zu Übungszwecken genutzt und verzerren durch den Lerneffekt bei fester Itemauswahl oder zufälliger Wahl aus einem zu kleinen Pool die Parameterschätzung. Für die Vermessung der Items und Probanden ist daher in Abhängigkeit der Durchführungsbedingungen der erste Durchlauf ausschlaggebend. Durch die messtheoretischen Eigenschaften der IRT können fortlaufende Übungen mit voneinander abweichenden Item-Sets zur vergleichbaren Messung des Kenntnisstandes genutzt werden, da die Parameter auf einer gemeinsamen Skala liegen. Dies erfordert jedoch, dass alle Items zumindest einmal gemeinsam oder mit Ankeritems in Gruppen ohne Beeinflussung durch den Lerneffekt



parametrisiert wurden oder die kompromittierten Items in Folgedurchläufen gesondert berücksichtigt werden.

Eine Beschränkung der Bearbeitungsdauer ermöglicht die Erstellung von Geschwindigkeitstests, die aus einer großen Anzahl etwa gleichschwerer Items bestehen, für deren komplette Inangriffnahme die gegebene Zeit aber nicht ausreicht. Dem gegenüber stehen Niveautests, die sich durch einen stetig ansteigenden Schwierigkeitsgrad und, falls überhaupt vorhanden, eine großzügige Zeitvorgabe auszeichnen [5]. Die vorgestellten IRT-Modelle betrachten Tests als Niveautests und es gibt Anpassungen für deren Anwendung auf Geschwindigkeitstests. Je nach Schwierigkeitsmix sowie Intention und Wirkung der Zeitvorgabe kommen daher verschiedene IRT-Modelle infrage.

Der überwiegende Teil der Leistungstests stellt eine Mischform dar [5]. Die Items unterscheiden sich im Schwierigkeitsgrad und es gibt (mindestens aus praktischen Gründen) eine Zeitbegrenzung. Bei Prüfungen im Bildungsbereich wird die Bewertung üblicherweise nicht auf Basis der korrekten Antworten in Relation zu den bearbeiteten Items in einer bestimmten Zeitspanne berechnet, sondern alle im Test enthaltenen Items tragen zur maximal erreichbaren Punktzahl bei. Dies verdeutlicht, dass trotz Zeitbegrenzung der Niveautestcharakter überwiegt. Um die Komplexität des Prototyps für LMS zu begrenzen soll daher auf Zeitkomponentenmodelle verzichtet und für den Testalltag die vorwiegende Verwendung von Niveautests angenommen werden. Die bereits genannten Modelle sind mit dieser Rahmenbedingung ausreichend.

#### *B. Auswahl der Items*

Die Auswahl der Items bestimmt in Niveautests maßgeblich das manifeste Ergebnis der Probanden. Sie können fest vorgegeben oder zufällig gewählt sein. Je größer im zweiten Fall die Anzahl der Probanden und die Überschneidung der genutzten Items in den Durchläufen, desto besser können die Parameter verschiedener Testformen mit Ankeritems zusammengeführt werden. Diese Einstellung beeinflusst also nicht die Modellwahl, sondern in Kombination mit der Datenmenge die Qualität der Ergebnisskala.

Die Zusammenstellung der Items bestimmt aber auch den inhaltlichen Aufbau eines Tests und beeinflusst damit dessen Dimensionalität. Das Idealbild der IRT (und ebenso der KTT) eines homogenen Tests wurde wegen der entgegenstehenden empirischen Realität durch die Forderung nach einem „dominanten Faktor“ [25] ersetzt. Haben mehrere Konstrukte Einfluss auf die manifesten Ergebnisse, so sind mehrdimensionale IRT-Modelle angebracht. Diese erfordern für stabile Parameterschätzungen gegenüber eindimensionalen Modellen etwa die doppelte Datenmenge. Für den Prototyp soll davon ausgegangen werden, dass pro Test ein vorrangig eindimensionales Konstrukt gemessen wird und geringe Abweichungen in der Faktorenstruktur als verschiedene Facetten des Konstrukts interpretiert und damit unter einem dominanten Faktor subsumiert werden können.

#### *C. Verhalten der Items*

Hinweise mit einhergehender Punktreduktion sowie sofortige Rückmeldungen sind für eine aussagekräftige Testanalyse zu unterlassen, sofern die erste Antwort nicht festgehalten wird. Die Hinweisgabe kann ignoriert werden, wenn die Punktreduktion genau dem hinzugewonnen psychometrischen Abstand entspricht. Dieser ist aber einerseits schwer zu bestimmen und andererseits könnte in dem Fall auch schlicht auf den Hinweis verzichtet werden. Wurde der Umfang der Reduktion nicht korrekt bemessen, sollten betroffene Items als neue Items angesehen werden, da die lokale stochastische Unabhängigkeit verletzt wird. In Analysen könnte einfließen, welche Probanden einen Hinweis anfordern und was sie, die Reduktion ausgeklammert und soweit im Test erfasst, ohne und mit Hinweis geantwortet hätten. Statt Hinweisen mit Punktreduktion könnte die Abfrage der Selbstsicherheitseinschätzung pro Item sinnvollere Erkenntnisse über den Antwortprozess erzielen, etwa mit Hybriden aus der IRT und der Signalentdeckungstheorie.

Items des Tests können als verpflichtend gekennzeichnet werden. Die Testabgabe kann dann nur erfolgen, wenn alle obligatorischen Items beantwortet wurden. Sofern eine systemseitige Zeitvorgabe besteht verliert sich die Wirkung, da der Test bei Zeitablauf automatisch eingezogen wird. Antworten auf alle Items zu erzwingen, damit keine lückenhaften Datensätze vorliegen, ist in Leistungstests nicht zwingend nötig. Auch unbeantwortete Items können in Niveautests als falsch beantwortet gewertet werden, sofern beispielsweise die Bearbeitungszeit allgemein zur Inangriffnahme aller Items des Tests reicht oder die Items bei knapper Zeitbegrenzung in zufälliger Reihenfolge präsentiert werden. Auf diese Weise müssen keine Antworten erzwungen werden und es gibt keine fehlenden Daten pro Testdurchlauf. Die Option für verpflichtende Items ist in üblichen Leistungstests daher unnötig. Eine Besonderheit ergibt sich aus der Implikation, dass es in Leistungstests auch fakultative Items gibt. Prinzipiell trifft dies auf alle Items zu, da kein Proband zu einer Antwort gezwungen werden muss und wie bei Papiertests eine leere Abgabe haben darf. Fakultative Items im engeren Sinne sind Zusatzitems und Wahlitems. Zusatzitems sind in der IRT nicht nötig, da sie bestenfalls zu einer weiteren Verringerung des Schätzfehlers der Merkmalsausprägung führen würden. Sie könnten daher ebenso fester Testbestandteil mit hohem Anspruch sein, um zwischen den besonders fähigen Probanden noch etwas genauer unterscheiden zu können. Wahlitems erzeugen unterschiedliche Testformen, die über Ankeritems zur skalenkonsistenten Parametrisierung der Wahlbestandteile zu verbinden sind. In ILIAS kann diese Form der Testgestaltung nicht konfiguriert, sondern nur informell über die Testinstruktionen vorgegeben werden. Es ist nicht eindeutig festzustellen, welche Wahlitems beantwortet wurden. Bis ILIAS eine integrierte Konfigurationsmöglichkeit für Testsektionen bietet, sollen Wahlaufgaben als Anwendungsfall aufgrund der Unwägbarkeiten nicht weiter berücksichtigt werden.

#### *D. Itemtypen*

Tests können beliebige Mischungen der Itemtypen enthalten. Geschlossenen Items wird größere Objektivität und Reliabilität unterstellt, aber auch ein vordergründiger

Reproduktionscharakter, während offene Items bei meist geringerer Reliabilität valider testen sollen [26]. Wenn das zu messende Konstrukt tatsächlich eindimensional wäre, könne auf den Itemmix auch verzichtet und nur jener Itemtyp genutzt werden, der das Merkmal am besten erfasst – da perfekte Eindimensionalität jedoch unrealistisch ist, sind auch Mischtests vertretbar.

Kommen offene und geschlossene Items innerhalb des gleichen Tests vor, bieten sich einige Alternativen für die Testauswertung. Es ist zu entscheiden, ob die Daten nach Typ getrennt oder gemeinsam, eindimensional oder mehrdimensional und dabei die Typen beachtend oder die Typen ignorierend ausgewertet werden. Untersuchungen der Dimensionsstruktur von Itemtypen zeigen, dass eine domänenabhängige Schwankung besteht. Die Komplexität der Dimensionsstruktur von offenen Formaten ist etwa in linguistischen Fächern höher als in naturwissenschaftlichen Fächern [27,28,29]. Das bedeutet, der Verlust von Testinformation durch die Verwendung von Mehrfachauswahlaufgaben statt Freitextaufgaben macht sich beispielsweise in Sprachtests stärker bemerkbar als in einem Chemietest. Die Wirkung sei allerdings gering genug, um sie zur Vereinfachung der Messung zu ignorieren.

Eine in [30] durchgeführte Literaturanalyse bestätigte unterschiedliche Messeigenschaften der Itemtypen, zeigte aber auch, dass die Konstruktäquivalenz vorrangig vom Itemdesign abhängt und Korrelationen von durchschnittlich 80% bei nicht inhaltsgleichen Items und 95% im Extremfall eines übereinstimmenden Itemstamms erreicht werden. Die gemeinsame Auswertung von offenen und geschlossenen Items wird daher ebenfalls als plausibel angesehen. In [31] diskutierte der gleiche Autor die Auswahl der Itemtypen für Tests: Dass viele Tests gemischte Formen darbieten führt er auf spezielle Eigenheiten des zu erfassenden Merkmals, aber auch auf Vorurteile aller Beteiligten (inklusive der Geprüften) und eine vorgegebene Testgestaltung zurück. Wenn verschiedene Typen möglich wären, sollten ihm zufolge geschlossene Formate bevorzugt werden, da sie einfacher auszuwerten sind. Sofern offene Items in geringem Maße auch andere Fähigkeiten messen, werde diese Wirkung in einem gemischten Test oft verschleiert.

Wie offene Items in Relation zu geschlossenen Items zwischen und innerhalb von Tests behandelt werden sollten, ist noch nicht ausreichend erforscht [32]. Der IRT-Prototyp für Tests in LMS wird gemäß der bisherigen Recherche den Itemtyp ignorieren und alle Items gemeinsam und eindimensional auswerten.

#### E. Punktvergabe

Sofern Items unterschiedlich viele Punkte vergeben, muss dies psychologisch begründet sein. Beispielsweise ist eine höhere Punktvergabe bei besonders schwierigen Items grundsätzlich fragwürdig, da fähige Probanden zusätzlich belohnt, schwächere Probanden zusätzlich abgestraft und damit Boden- und Deckeneffekt der Bewertung befördert werden. Gleiches gilt für Negativpunkte, die obendrein den Messvorgang durch Einbezug der individuellen Risikofreudigkeit verzerren.

Die Punktevergabe sollte sich an einem Punkt pro korrekter Reaktion orientieren und nicht auf Negativpunkte zurückgreifen. Die IRT bildet Punkte oder Punktintervalle als Kategorien ab und verortet diese mit jeweils einer ICC im Merkmalskontinuum. In Modellen, die lediglich einen Schwierigkeitsparameter pro Antwortkategorie schätzen, gibt es keine (beziehungsweise eine einheitliche) Gewichtung. In höher parametrisierten Modellen wird das Gewicht von Items zur Bestimmung der Merkmalsausprägung implizit durch die itemspezifischen Parameter realisiert (Trennschärfe, Ratewahrscheinlichkeit usw.), welche die Form der ICC beeinflussen. Eine Gewichtung über die vergebenen Punkte ist somit nicht erforderlich, da sie aus dem Informationsbeitrag des Items abgeleitet wird.

Wie schon bei der Beachtung unterschiedlicher Itemtypen sind auch bei gemischten Punktvergabeverfahren verschiedene Herangehensweisen denkbar. Die Items können gemäß dem Punktvergabeverfahren oder der maximal erreichbaren Punktzahl gruppiert, gemeinsam dichotom oder polytom und eindimensional oder mehrdimensional ausgewertet werden. Da die IRT die psychometrische Gewichtung und die Merkmalskala gemeinsam abbildet, spräche in eindimensionalen Tests nichts gegen eine ebenso gemeinsame Auswertung aller Items [26]. Die Gewichtung von Items nach ihrer Punktzahl wird dabei ignoriert. Der Prototyp wird pro vergebenem Punktwert, ungeachtet der Punkthöhe, eine Antwortkategorie abbilden. Es ist daher ein polytomes IRT-Modell zu bevorzugen, dass sich bei entsprechender Datenlage zu einem dichotomen Modell vereinfachen kann und keine starren Restriktionen (beispielsweise Äquidistanz) für die Positionierung einzelner Antwortkategorien pro Item vorgibt.

#### F. Fazit

Selbst beschränkt auf Leistungstests im konfiguratorischen Rahmen eines LMS deckt die Diskussion der Modelle nicht alle darin denkbaren Szenarien ab. Die Umsetzung des Prototyps kann nicht alle Auswertungsvarianten und Integrationsmöglichkeiten nutzerfreundlich konfigurierbar darbieten, sondern wird sich auf einfache Leistungstestszenarien im Bildungsbereich konzentrieren. Daher muss sich der Prototyp auf bestimmte Annahmen und Einschränkungen stützen, die einen technischen Durchbruch für den weiteren Ausbau ermöglichen:

- Es gibt kein Vorwissen über die Probanden oder die Items.
- Der Test dient der Querschnittsdiagnose.
- Der Test soll nur den dominanten Faktor messen.
- Items sowie Probanden sollen lokal stochastisch unabhängig sein (keine Testlets, kein Betrug).
- Der Test nutzt für alle Probanden ein festes Item-Set.
- Fehlende Antworten werden als falsche Antworten gewertet.
- Offene und geschlossene Items dürfen im gleichen Test verwendet werden.

- Dichotome und polytome Items dürfen im gleichen Test verwendet werden.
- Jede korrekte (Teil-)Antwort wird ungewichtet mit einem Punkt bewertet beziehungsweise in der IRT-Auswertung so behandelt.
- Es gibt keine Negativpunkte.

Die Betrachtung der verschiedenen Konfigurationsoptionen und deren Implikation auf die Modellwahl führen zum Schluss, dass für den allgemeinen Anwendungsfall ein eindimensionales und polytomes Modell genutzt werden sollte. Der allgemeinste Fall für die gegebene Messsituation ist das GRM [11], welches daher das Hauptmodell für den Prototyp werden sollte.

Hinsichtlich der Integrierbarkeit von Teststatistiken in einen Itempool scheint das GRM zunächst problematisch. Ein solcher Pool sollte eine sortierbare Datenbank sein, was auch auf die Schwierigkeitseigenschaft der Items zutreffen muss. Doch lassen sich Items mit unterschiedlich vielen Antwortkategorien, die jeweils einen eigenen Schwierigkeitsparameter haben, nicht direkt in dieses Schema einfügen. Eine Lösung ergibt sich aus dem Skalenniveau der IRT-Parameter, welches unter anderem die Verwendung des arithmetischen Mittels zur Bildung eines einzigen Schwierigkeitswertes für ein Item erlaubt [33].

Da die Anwendbarkeit der Modelle im Prototyp letztlich von den Funktionen der nutzbaren IRT-Software abhängig ist, sei eine absteigende Präferenzreihenfolge gegeben:

- GRM – passend für die meisten allgemeinen Leistungstests in LMS
- GPCM – passend für Tests mit sequenziellem Antwortverhalten
- PCM – wie GPCM, Trennschärfe gleich Eins für alle Items
- 2-PL – wie GRM, für dichotome Daten
- 1-PL – wie 2-PL, identische Trennschärfe für alle Items
- Rasch-Modell – wie 1-PL, Trennschärfe gleich Eins für alle Items

Aufgrund seiner Genese liefert das GRM zum 2-PL-Modell identische Ergebnisse, wenn es auf dichotome Daten angewandt wird. Im Fall der einheitlichen Anzahl und auch in den Daten vollständig vertretener Kategorien über alle Items wäre das RSM zwischen 2-PL-Modell und PCM einzufügen: Falls keine bessere Alternative genutzt werden kann und die Daten es zulassen, erzeugt eine polytome Schätzung mit starren Kategorieabständen weniger Informationsverlust als eine Dichotomisierung aller Daten. Ein weiterer Sonderfall ist das 3-PL-Modell, welches für dichotome Daten unter Berücksichtigung der Ratewahrscheinlichkeit anwendbar ist und sich damit besonders für Tests mit geschlossenen Items eignet. Das NRM bietet sich für die Distraktorenanalyse an. Prinzipiell käme selbst das 4-PL-Modell in Betracht, das einen Unachtsamkeitsparameter einfügt. Es sei aber auf die pro hinzukommenden Parameter ansteigende und ohnehin schon große Datenmenge hingewiesen, die für stabile

Parameterschätzungen benötigt wird und bereits für die präferierten Modelle nicht in jedem Test erreicht werden dürfte.

Diese Vorauswahl kann weder *model-fit*, *item-fit* noch *person-fit* berücksichtigen. Das GRM mag als allgemein passendes Modell im hier betrachteten Szenario eingeschätzt werden, dennoch könnten die konkreten Daten nicht mit den Modellannahmen übereinstimmen. Der allgemeine Vergleich zwischen infrage kommenden Modellen wird in [17] auf ähnliche Weise diskutiert und konstatiert, dass die Forschung zur mathematisch begründeten Modellwahl der IRT noch fortgeführt werden muss und die Auswahl bislang letztlich doch meist aus praktischen Erwägungen erfolgt.

## VII. UNTERSUCHUNG DER TECHNISCHEN MACHBARKEIT

Es gibt diverse Softwareangebote zur Durchführung einer IRT-gestützten Testauswertung (einen Auszug listet [34]). Viele davon sind technisch veraltet, teuer und für ungeschulte Anwender schwer zugänglich, da sie Programmierkenntnisse, Arbeit auf der Kommandozeile oder eine intensive Einarbeitung in Details der IRT-Modelle erfordern.

Auch wurde schon ein Vorgehen für die IRT-Integration in LMS-Testauswertungen in [35] beschrieben. Diesem liegt ein Prototyp für das LMS DOKEOS zugrunde, der in der Installation der Universität Makedonien vorlag. Das darin umgesetzte Rahmenwerk hat allerdings mehrere Schwachpunkte, die eine Adaption und weitere Verwendung beeinträchtigen: Es basiert auf dem Programm „*IRT Command Language*“ (ICL), das bereits seit 2002 durch den Tod des verantwortlichen Forschers nicht mehr weitergeführt wird [36]. Die IRT-Software befindet sich lokal auf dem Webserver des LMS und die Schnittstelle zum LMS bilden Textdateien, die von den involvierten Systemen abwechselnd zu schreiben und einzulesen sind. Zudem ist die Auswertung auf Einfachauswahlaufgaben und das 3-PL-Modell beschränkt.

Das Rahmenwerk des existierenden Vorstoßes einer IRT-Integration zur Testauswertung in LMS ist aus den genannten Gründen unattraktiv. Es bedarf eines psychometrisch flexibleren und technisch moderneren Vorgehens, wobei durch die Komplexität der IRT eine Eigenimplementierung nicht in Betracht kommt. Daher wurde das Angebot vorhandener IRT-Software untersucht. Kommerzielle Software und eingeschränkte Demonstrationsversionen wurden nicht berücksichtigt, da eine freie Integration angestrebt wird. Nur auf persönliche Anfrage zugängliche Programme wurden ebenfalls nicht betrachtet. Die übrige frei zugängliche, kostenlose und idealerweise quelloffene Software wurde hinsichtlich ihrer Nachhaltigkeit und Integrierbarkeit sowie, sofern in Betracht kommend, ihres Funktionsumfangs untersucht.

Die Analyse ergab, dass es kaum geeignete Software für eine Integration in Websysteme gibt. Die meisten Angebote besitzen nur eine grafische Oberfläche für die Verwendung auf dem Desktop oder lesen Textdateien mit Skripten ein. Es gibt teilweise aktuell gehaltene Klassenbibliotheken in C, C++, Python und Java, für die allerdings erst passende Umsysteme für den angedachten Einsatzzweck entwickelt werden müssten. Die Wahl fiel letztlich auf die Statistikumgebung R [37].

Als freies, kostenloses und quelloffenes System bietet R zahlreiche psychometrische und zudem regelmäßig gepflegte Erweiterungen, darunter für die IRT beispielsweise „*eRm*“, „*mirt*“ oder „*ltm*“ (eine Übersicht listet [38]). Die Pakete bieten alle benötigten Funktionen einer IRT-gestützten Testanalyse im hier beschriebenen Szenario. Für den Prototyp wird das Paket „*ltm*“ genutzt, das mehrere dichotome und polytome Modelle, die Erstellung von *fit*-Statistiken und die Durchführung von Testkopplungsverfahren unterstützt.

Weiterhin liefert R zahlreiche Alternativen für Integrationsszenarien. Darunter beispielsweise Plumber, OpenCPU [39] oder Shiny. Bei näherer Betrachtung der Integrationsmethoden erweist sich die Serveranwendung OpenCPU als die geeignetste. Sie erfordert weder Entwicklungsarbeiten, um vorhandenen R-Code über Annotationen (siehe Plumber) netzwerkfähig zu machen, noch unterliegt sie kommerziellen Einschränkungen, wie Shiny, das beispielsweise nur in der Bezahlversion eine verschlüsselte Datenübertragung oder den Zugriff auf eine Datenbank ermöglicht. OpenCPU besitzt keinen Fertigbaukasten für Webanwendungen, ermöglicht aber dafür eine größere Flexibilität. Es basiert auf ApacheR und bietet eine REST-Schnittstelle zu sämtlichen auf dem Server installierten R-Paketen. Die Installation von OpenCPU kann sowohl lokal erfolgen, als auch per APT oder selbst kompiliert als dedizierter Server und letztlich auch als Docker-Container.

Durch die skalierbare Auslagerung der Berechnungen auf einen oder mehrere OpenCPU-Server wird der Betrieb des LMS nicht von prozessorlastigen Rechenoperationen beeinträchtigt. Die einheitliche Schnittstelle macht die Anbindung unabhängig vom integrierenden System, in welchem für die initiale Bereitstellung der IRT-Analyseergebnisse lediglich der Versand und die Entgegennahme der Daten geregelt sein müssen.

## VIII. PROTOTYP

Nachdem die psychometrische und technische Machbarkeit geklärt wurde, sind die Erkenntnisse in einem Prototyp für ILIAS umzusetzen. Dieser soll IRT-Analysen in möglichst simplifizierter Form für ungeschulte Nutzer darbieten können und, im Zusammenspiel mit den von der Universität Leipzig forcierten Kernentwicklungen, das Fundament dafür legen, die IRT-Ergebnisse auch in Itempools überspielen zu können.

Um den Prototyp nicht zu eng mit dem LMS zu verbinden soll kein Patch entwickelt werden, obwohl das die gestalterische Freiheit erhöhen würde. Stattdessen wird ein Plugin erstellt, wofür ILIAS verschiedene Möglichkeiten bietet. Für das Szenario naheliegend wären Cron-Plugins, Testexport-Plugins und User-Interface-Plugins.

Ein erstes Testexport-Plugin wurde als Vorläufer der IRT-Integration bereits 2015 entwickelt und lieferte klassische Teststatistiken als Excel-Datei mit Feldfunktionen für eine bessere Nachvollziehbarkeit und Anpassbarkeit der Berechnungen. Ein Jahr später wurde es von einem User-Interface-Plugin der Universität Erlangen („*ExtendedTestStatistics*“) abgelöst, das diese Werte im Browser anzeigt und mittels PHP-Implementierung berechnet. Dieses Plugin soll fest in den LMS-Kern übernommen werden, weshalb es bereits für die Untersuchung der LMS einbezogen

wurde und beinhaltet auch eine Erweiterungsschnittstelle für teststatistische Berechnungen. Da diese Schnittstelle für den mit der IRT-Integration angestrebten Einsatzzweck gedacht ist, wird sie trotz begrenzter Entwicklungsfreiheit für den Prototyp genutzt. Der IRT-Prototyp ist die erste Entwicklung dafür, weshalb gleichzeitig auch Verbesserungsmöglichkeiten der einbindenden Strukturen entdeckt und beigesteuert werden konnten. Es waren Änderungen nötig, um konkret den ersten statt nur den letzten oder besten Testdurchlauf auszuwerten, verschiedene Konfigurationseingaben zuzulassen sowie Grafiken und HTML-Templates zu verwenden.

Der Prototyp erstellt nicht nur IRT-Analysen, sondern steuert auch KTT-Auswertung bei. Er soll es später auch ermöglichen, die in PHP umgesetzten Berechnungen von *ExtendedTestStatistics* zu ersetzen. Abhängig von der Größe des Datensatzes und der mathematischen Komplexität zeigten sich die alternativen Berechnungen via OpenCPU bislang um ein Vielfaches performanter als ihre PHP-basierten Gegenstücke. Bei kleineren Berechnungen wird der Vorteil jedoch vom Overhead der Netzwerkoperationen und der Datenaufbereitung getilgt.

Bislang wird aus der KTT folgendes angeboten:

- Interne Konsistenz (Cronbach's Alpha/ $\lambda_3$ /KR-20)
- Einfluss der Löschung eines Items auf die interne Konsistenz
- Faktorenanalyse (Scree-Plot und Faktorladungen)
- Rohwertanalyse (Verteilung, Schiefe, Wölbung)

Die IRT-Analyse umfasst bislang:

- Rasch-Modell
- 1/2/3-PL-Modell
- GRM
- GPCM

Pro Modell werden die ICC pro Item beziehungsweise Antwortkategorie, die Iteminformationskurven und die Testinformationskurve angezeigt. Die Schwierigkeitswerte der Antwortkategorien pro Item von GRM und GPCM werden zu einem sortierbaren Wert zusammengefasst. Sollen polytome Daten explizit mit einem dichotomen Modell ausgewertet werden, wird eine Dichotomisierung durchgeführt, wofür als Trennkriterien der Median (Standardeinstellung), der Modus, der Mittelwert und die Hälfte der maximal erreichbaren Punktzahl pro Item implementiert sind. Weiterhin wird eine Texteingabe für R-Befehle angeboten, in der von Nutzern simple Datenanalysen via OpenCPU selbst angestoßen werden können. Alle von OpenCPU verarbeiteten Daten werden vor dem Versand vom Plugin aufbereitet und enthalten ausschließlich die Antwortmatrix mit den IDs der Items und den Punktwerten.

Die Funktionalität des Prototyps muss noch durch *fit*-Statistiken und kurze Hilfetexte vervollständigt werden. Das Plugin wird quelloffen und kostenlos bereitgestellt und bei Abschluss der Arbeiten in die vom ILIAS e.V. bereitgestellte Liste von Plugins aufgenommen. Ein dedizierter OpenCPU-

Server wird empfohlen. Sofern nicht zu viele Anfragen gleichzeitig ausgelöst werden, kann auch der offizielle Demoserver ([cloud.opencpu.org](http://cloud.opencpu.org)) genutzt werden.

## IX. VALIDIERUNG

Eine Fallstudie mit Dozenten der Universität Leipzig und deren Klausuren soll den Nutzen und die Benutzbarkeit der IRT-gestützten Auswertung des Prototyps prüfen. Die Arbeiten daran beginnen, sobald der Prototyp abgeschlossen ist.

- [1] Guthke, J., Böttcher, H. R. und Sprung, L. (Hrsg.) (1990) Psychodiagnostik Band 1. Berlin: Deutscher Verlag der Wissenschaften.
- [2] Pospeschill, M. (2010) Testtheorie, Testkonstruktion, Testevaluation. München: Ernst Reinhardt Verlag.
- [3] Eid, M. und Schmidt, K. (2014) Testtheorie und Testkonstruktion. Göttingen: Hogrefe.
- [4] Fischer, G. (1974) Einführung in die Theorie psychologischer Tests. Bern: Huber.
- [5] Lienert, G. A. und Raatz, U. (1994) Testaufbau und Testanalyse. 5. Auflage. Weinheim: Beltz.
- [6] Rost, J. (1999) „Was ist aus dem Rasch-Modell geworden?“, *Psychologische Rundschau*, 50(3), S. 140–156. doi: 10.1026//0033-3042.50.3.140.
- [7] Rasch, G. (1960) Probabilistic models for some intelligence and attainment tests. Kopenhagen: Paedagogiske Institut.
- [8] Lord, F. M. und Novick, M. R. (1968) Statistical theories of mental test scores. Oxford: Addison-Wesley.
- [9] Samejima, F. (1969) Estimation of Latent Ability Using a Response Pattern of Graded Scores, Psychometric Monograph No. 17. Richmond, VA: Psychometric Society. Verfügbar unter: <https://www.psychometricsociety.org/sites/default/files/pdf/MN17.pdf>.
- [10] Bock, R. D. (1972) „Estimating item parameters and latent ability when responses are scored in two or more nominal categories“, *Psychometrika*, 37(1), S. 29–51. doi: 10.1007/BF02291411.
- [11] Andrich, D. (1978) „A rating formulation for ordered response categories“, *Psychometrika*, 43(4), S. 561–573. doi: 10.1007/BF02293814.
- [12] Andrich, D. (1978) „Application of a Psychometric Rating Model to Ordered Categories Which Are Scored with Successive Integers“, *Applied Psychological Measurement*, 2(4), S. 581–594. doi: 10.1177/014662167800200413.
- [13] Masters, G. N. (1982) „A rasch model for partial credit scoring“, *Psychometrika*, 47(2), S. 149–174. doi: 10.1007/BF02296272.
- [14] Muraki, E. (1992) „A Generalized Partial Credit Model: Application of an EM Algorithm“, *Applied Psychological Measurement*, 16(2), S. 159–176. doi: 10.1177/014662169201600206.
- [15] Wendt, H., Bos, W. und Goy, M. (2011) „On applications of Rasch models in international comparative large-scale assessments: a historical review“, *Educational Research and Evaluation*, 17(6), S. 419–446. doi: 10.1080/13803611.2011.634582.
- [16] OECD (2009) PISA Data Analysis Manual: SPSS, Second Edition. OECD Publishing (PISA). doi: 10.1787/9789264056275-en.
- [17] Embretson, S. E. und Reise, S. P. (2000) Item Response Theory for Psychologists. Mahwah, New Jersey: Erlbaum Associates.
- [18] Foster, G. C., Min, H. und Zickar, M. J. (2017) „Review of Item Response Theory Practices in Organizational Research“, *Organizational Research Methods*, 20(3), S. 465–486. doi: 10.1177/1094428116689708.
- [19] Hambleton, R. K. und van der Linden, W. J. (Hrsg.) (1997) Handbook of Modern Item Response Theory. doi: 10.1007/978-1-4757-2691-6.
- [20] Bichi, A. A. u. a. (2015) „Comparison of Classical Test Theory and Item Response Theory: A Review of Empirical Studies“, *Australian Journal of Basic and Applied Sciences*, 9(7), S. 549–556.
- [21] APA (2016) Journals in Assessment, Evaluation, Measurement, Psychometrics and Statistics. Verfügbar unter: <http://apadivisions.org/division-5/resources/journals/index.aspx> (Zugegriffen: 29. September 2018).
- [22] LISTedTECH (2017) What Percentage of HigherEd Institutions have a LMS? Verfügbar unter: <http://listedtech.com/percent-highered-institutions-lms/> (Zugegriffen: 6. Oktober 2017).
- [23] LISTedTECH (2018) LMS Coverage - Two-year anniversary!!! Verfügbar unter: <http://listedtech.com/lms-coverage-two-year-anniversary/> (Zugegriffen: 19. Juni 2018).
- [24] e-Literate (2016) „European LMS Market Dynamics - Fall 2016 Report“. Verfügbar unter: <http://mfeldstein.com/academic-lms-market-share-view-across-four-global-regions/>.
- [25] Hambleton, R. K., Swaminathan, H. und Rogers, D. J. (1991) Fundamentals of Item Response Theory. Newbury Park: SAGE.
- [26] Wainer, H. und Thissen, D. (1992) „COMBINING MULTIPLE-CHOICE AND CONSTRUCTED RESPONSE TEST SCORES: TOWARD A MARXIST THEORY OF TEST CONSTRUCTION“, *ETS Research Report Series*, 92–93, S. i-13. doi: 10.1002/j.2333-8504.1992.tb01454.x.
- [27] Wang, W., Drasgow, F. und Liu, L. (2016) „Classification Accuracy of Mixed Format Tests: A Bi-Factor Item Response Theory Approach“, *Frontiers in Psychology*, 7(February). doi: 10.3389/fpsyg.2016.00270.
- [28] Thissen, D., Wainer, H. und Wang, X.-B. (1994) „Are Tests Comprising Both Multiple-Choice and Free-Response Items Necessarily Less Unidimensional than Multiple-Choice Tests? An Analysis of Two Tests“, *Journal of Educational Measurement*, 31(2), S. 113–123.
- [29] Ercikan, K. u. a. (1998) „Calibration and scoring of tests with multiple-choice and constructed-response item types“, *Journal of Educational Measurement*, 35(2), S. 137–154. doi: 10.1111/j.1745-3984.1998.tb00531.x.
- [30] Rodriguez, M. C. (2003) „Construct equivalence of multiple-choice and constructed-response items: A random effects synthesis of correlations“, *Journal of Educational Measurement*, 40(2), S. 163–184. doi: 10.1111/j.1745-3984.2003.tb01102.x.
- [31] Rodriguez, M. C. (2002) „Choosing Item Format“, in Tindal, G. und Haladyna, T. M. (Hrsg.) Large-Scale Assessment Programs for all Students. Mahwah, New Jersey: LAWRENCE ERLBAUM ASSOCIATES, S. 213–231.
- [32] Kolen, M. J. und Brennan, R. L. (2014) Test Equating, Scaling, and Linking. 3. Auflage. New York, NY: Springer New York. doi: 10.1007/978-1-4939-0317-7.
- [33] Ali, U. S., Chang, H.-H. und Anderson, C. J. (2015) Location Indices for Ordinal Polytomous Items Based on Item Response Theory. Princeton, N.J.: Educational Testing Service. doi: 10.1002/ets2.12065.
- [34] Zhao, Y. und Hambleton, R. (2009) Software for IRT Analyses: Descriptions and Features, Center for Educational Assessment Research Report. University of Massachusetts Amherst.
- [35] Fotaris, P. u. a. (2010) „Extending LMS to Support IRT-Based Assessment Test Calibration“, *Technology Enhanced Learning. Quality of Teaching and Educational Reform. Communications in Computer and Information Science*. Herausgegeben von M. D. Lytras u. a. Berlin: Springer, 73, S. 188–194. doi: 10.1007/978-3-642-13166-0.
- [36] openirt.com (2014) IRT Command Language. Verfügbar unter: <http://www.openirt.com/b-a-h/software/irt/icl/> (Zugegriffen: 18. März 2018).
- [37] R Core Team (2018) „R: A language and environment for statistical computing“. Wien: R Foundation for Statistical Computing. Verfügbar unter: <https://www.r-project.org>.
- [38] Mair, P. (2018) *CRAN Task View: Psychometric Models and Methods*. Verfügbar unter: <https://cran.r-project.org/web/views/Psychometrics.html> (Zugegriffen: 28. September 2018).
- [39] Ooms, J. (2014) The OpenCPU System: Towards a Universal Interface for Scientific Computing through Separation of Concerns. Verfügbar unter: <http://arxiv.org/abs/1406.4806>.

# Prozessorganisation in ungeplanten Notfallsituationen mithilfe von Workflowmanagement (working title)

Vorstellung eines Dissertationsvorhabens

Sandra Niemz

Wirtschaftswissenschaftliche Fakultät – Lehrstuhl Wirtschaftsinformatik  
Friedrich-Schiller-Universität Jena  
Jena, Deutschland  
sandra.niemz@uni-jena.de

**Abstract** – Durch Menschen verursachte oder naturbedingte unvorhergesehene Ereignisse oder Abweichungen können extreme Auswirkungen haben. Diese können insbesondere einen Effekt auf Prozesse, seien es Standard oder geplante oder ungeplante Nicht-Standardprozesse, in Unternehmen haben. Unternehmen greifen zur Organisation ihrer Prozesse oft auf Workflowmanagement zurück. Dabei gelingt es diesem jedoch nur schwer, auf unvorhergesehene Situationen adäquat zu reagieren. Deshalb ist es wichtig, unter anderem die Abläufe von Nicht-Standardprozessen proaktiv organisieren zu können.

Ein neuerer Ansatz, der erhöhte Flexibilität von Workflows gewährleisten kann, besteht aus der Kombination der Modellierungsmethoden BPMN (Business Process Model and Notation) und CMMN (Case Management Model and Notation). Die Kombination beider Methoden wurde bisher kaum im Zusammenhang von ungeplanten Nicht-Standardprozessen im Rahmen eines Notfalls oder einer Krise betrachtet. Eine weitere Ergänzung dazu kann außerdem die Integration der Modellsprache DMN (Decision Model and Notation) bilden.

Die angestrebte Dissertation möchte deshalb mithilfe einer Verbindung von BPMN, CMMN und DMN ein Methode vorschlagen und nach bestimmten Kriterien evaluieren, die die Organisation dieser Prozesse in Notfallsituationen besser unterstützen soll. Insgesamt möchte die Dissertation untersuchen, inwieweit sich ungeplante Nicht-Standardprozesse, insbesondere im Zusammenhang mit dem Notfallmanagement durch die vorgeschlagenen Methoden organisieren und durchführen lassen.

**Keywords** — Workflowmanagement; BPMN; CMMN; DMN; Notfallmanagement

## I. EINLEITUNG

Ungeplante Ereignisse können eine verheerende Wirkung auf Unternehmen und die Gesellschaft haben. Statistiken der Swiss Re Economics Research & Consulting zeigen, dass die Anzahl (seien durch Natur oder Menschen bedingte) Katastrophen in den letzten sieben Jahren etwa stabil geblieben ist. Den

niedrigsten Stand von Vorfällen gab es im Jahr 2017 mit 301 Zählungen, den höchsten 2015 mit 355 Vorkommnissen. [22] Andere Untersuchungen zeigen sogar einen Anstieg von Naturkatastrophen seit den 1940er Jahren. [18] Ein bekannter, durch die Natur bedingter Vorfall aus den letzten Jahren ist das Elbhochwasser aus dem Jahr 2006. Als eine von Menschen verursachte Katastrophe in einem Unternehmen kann der Abgasskandal von VW aus dem Jahr 2012 genannt werden. Ereignisse, die zu diesen unerwünschten Ergebnissen führen, können nicht immer vorbeugend verhindert werden. Daher ist es wichtig, auch beim Wegfallen gewohnter Strukturen und Arbeitsabläufen adäquat reagieren zu können um schlimmere Schäden zu vermeiden. Mithilfe von Workflowmanagement (WFM) können Prozesse vorgedacht und dann (halb-)automatisiert durchgeführt werden. Eine Schwäche des klassischen WFMs lässt sich aber erkennen, wenn Prozesse ad hoc angepasst werden müssen, um auf veränderte Bedingungen zu reagieren. Mit dem Adaptive Case Management (ACM) und dessen Methoden können diese Defizite ausgeglichen werden. [6] Diese Arbeit möchte eine Kombination der Methoden des Business Process Managements (BPM) und des ACMs im Zusammenhang mit dem Notfallmanagement untersuchen und aufzeigen, welche Anpassungen oder Erweiterungen vorgenommen werden müssen, damit diese effektiv zum Notfallmanagement beitragen können.

Zunächst soll eine Aufbereitung der relevanten Aspekte erfolgen. Im Fokus der Arbeit stehen dabei vor allem die Modellierungsmethoden, die Charakteristika eines Notfalls und die Einbettung des dessen in einem Tool. Die theoretische Ausarbeitung soll auf ihre Anwendbarkeit initial mithilfe von Simulationen überprüft werden, ehe im zweiten Schritt eine praktische Validierung mithilfe eines Experiments erfolgt. Insgesamt möchte die Dissertation damit untersuchen, in wie weit sich ungeplante Nicht-Standardprozesse, insbesondere im Zusammenhang mit dem Notfallmanagement durch die vorgeschlagenen Methoden organisieren und durchführen lassen.

## II. PROZESS- UND UMWELTABGRENZUNG

In Unternehmen gibt es eine Vielzahl von Prozessen, die tagtäglich ablaufen. Bei einem Großteil handelt es sich um Standardprozesse, welche vor allem im operativen Tagesge-

schaft anfallen. Dadurch, dass Prozesse im Geschäftsprozessmanagement (halb-)automatisiert und wiederholend auftreten, können diese in passender Weise von Workflowmanagementsystemen (WFMS) unterstützt werden. [6] Auch geplante Nicht-Standardprozesse können durch diese organisiert werden. Die einzige Eigenschaft, die Standard- von Nicht-Standardprozessen im Wesentlichen unterscheidet, ist die Tatsache, dass Nicht-Standardprozesse weniger häufig durchgeführt werden. Dennoch ist der Inhalt des Prozesses bekannt und kann deshalb mithilfe von WFMS vorgedacht werden.

Eine andere Situation ergibt sich bei ungeplanten Nicht-Standardprozessen. Unter ungeplanten Nicht-Standardprozessen fallen Prozesse, welche aufgrund eines durch Mensch, Natur oder Technik bedingten Vorfalls (fortan als Notfall bezeichnet) nicht wie ursprünglich geplant durchgeführt werden können. Al-Dahash et al. haben sich mit verschiedenen Definitionen der Begriffe Disaster, Krise und Notfall auseinandergesetzt und aufgezeigt, dass es bisher noch keine einheitliche Beschreibung dieser gibt und sie synonym füreinander verwendet werden. [1] Eine relativ passende Definition des Begriffs des Notfalls hat Alexander formuliert. Im folgenden Verlauf dieser Arbeit soll diese Definition weiterverwendet werden. Alexander versteht unter einem Notfall:

*„an imminent or actual event that threatens people, property or the environment and which requires a coordinated and rapid response. Emergencies are usually unanticipated, at least in terms of exactly what happens and when and where they take place. However, they can, and should, be planned for.“* [2]

In dieser Definition geht er sowohl auf mögliche Auswirkungen eines Notfalls ein, als auch auf bestimmte Eigenschaften, wie dass ein Notfall meist unerwartet auftritt. Dies kann die Ebenen Zeit und Ort umfassen. Darüber hinaus sagt die Definition aus, dass eine Vorbereitung auf Notfälle notwendig ist, um den Auswirkungen dessen entgegenzuwirken. Über den Schweregrad der Auswirkung eines Notfalls wird in dieser Definition nicht gesprochen. Im Rahmen dieser Arbeit wird ein leichter bis mittlerer Schweregrad angenommen, da Al-Dahash et al. mit den Begriffen der Krise und des Disasters zwei Szenarien voranstellt, die sich aus einem unbehandelten Notfall entwickeln können. [1] Darüber hinaus soll sich der Ansatz auch an Prozesse richten, die durch bestimmte Umstände nicht mehr wie vorgedacht durchgeführt werden können und eine Änderung benötigen. In schwerwiegenderen Szenarien, bei denen die vorherige Struktur kaum oder gar nicht mehr vorhanden ist, ist der im Folgenden vorgestellte Ansatz weniger hilfreich.

Das Notfallmanagement lässt sich nach einem der geläufigsten Modelle in vier Phasen aufteilen, welche in Abbildung 1 zu sehen sind. Generell geht dieses Modell von einer stets erfolgreichen Notfallbehandlung aus. In diesem Fall wird angenommen, dass alle betroffenen Einheiten wieder in den Normalzustand oder sogar in einen verbesserten Zustand übergehen. [5] Dieses Modell bildet nur einen Teil der Komplexität ab, es gibt aber durchaus Modelle mit weiteren oder überlappenden Phasen. [11] Eine Unterscheidung z.B. zwischen Recovery und Lessons Learned sind in diesem Modell nicht vorgesehen. Ein Auseinandernehmen dieser beiden Punkte kann

aber durchaus sinnvoll sein, da der Notfall womöglich nicht erfolgreich behoben werden konnte, es aber dennoch möglich ist, Erkenntnisse für die Zukunft zu gewinnen. Besonders im Hinblick auf die Abgrenzung des Begriffs gegenüber Krisen und Disaster ist es vorstellbar, dass ein Unternehmen aufgrund des geringen Auswirkungsgrades des Notfalls auch bei einer nicht erfolgreichen Behandlung weiter bestehen bleibt.

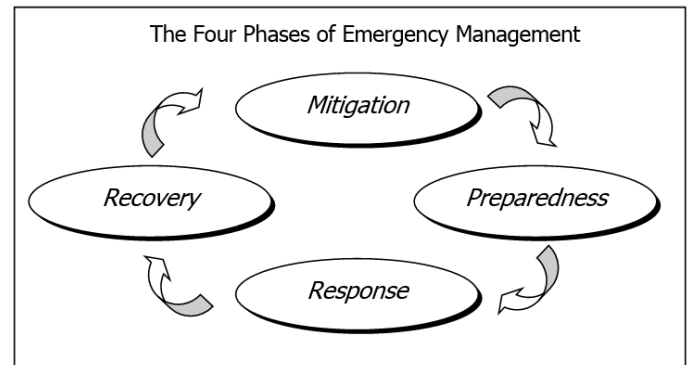


Abbildung 1: Die Phasen des Emergency Managements [5]

Unabhängig davon muss sich ein flexibles WFM im Notfall besonders auf die Phasen Response und Recovery konzentrieren. Hier muss zwischen zwei verschiedenen Prozessen unterschieden werden. Bedingt durch das unerwartete Ereignis wird zunächst der normale Ablauf von bestehenden Prozessen gestört. Diese müssen an die geänderten Umstände angepasst werden. Das zweite Szenario umfasst Maßnahmen zur Überwindung des Notfalls. [8] Ein Beispiel für eine solch zweigeteilte Situation würde aufgrund eines externen Ereignis entstehen, welches z.B. dazu führt, dass es in einem Unternehmen kein Internet mehr gibt. Es muss sichergestellt werden, dass das Tagesgeschäft möglichst reibungslos fortgeführt werden kann. Dafür müssen schon bestehende Prozesse angepasst werden (= geplante Standardprozesse). Darüber hinaus muss aber auch ein neuer Prozess angestoßen werden, der das Problem (kein Internet) behebt (= ungeplanter Nicht-Standardprozess). Bei diesen ist nicht nur die Anzahl der Wiederholungen kritisch zu betrachten, sondern auch die (Halb-)Automatisierung muss in diesem Kontext als schwierig oder schlecht erfüllbar angesehen werden. [6]. Des Weiteren haben ungeplante Nicht-Standardprozesse erweiterte Anforderungen an ein WFMS als geplante Prozesse. Diese Anforderungen wurden von Sell et al. [19] als folgende identifiziert:

1. Ressourcenmanagement
2. (Fortschritts-)Anzeige bzw. Prozessfortschritt
3. Anpassung des Workflows vor und während des Prozesses
4. Rechtsmanagement
5. Ausführung des Workflows

Insgesamt kann ein Notfall jede dieser Anforderungen beeinträchtigen und Prozesse benötigen dort eine Anpassung. Eine der zentralsten und auch am schwierigsten zu bewältigenden Probleme besteht aber bei der dritten Anforderung. Abhilfe dafür kann das Adaptive Case Management bieten. Prozesse, die mithilfe des Adaptive Case Managements bearbeitet werden, müssen nicht die Eigenschaft besitzen, vordefiniert und wiederholbar sein müssen, so wie es das klassische Work-

flowmanagement fordert. Damit kann besonders auf ungeplante Nicht-Standardprozesse reagiert werden, die beide Eigenschaften erfüllen, da es eine Stärke des Adaptive Case Managements ist auf sich verändernde Bedingungen einzugehen. [7]

### III. THEORETISCHE GRUNDLAGEN FÜR MODELLIERUNGSANSÄTZE

Zur Gestaltung und Organisation eines Workflows gibt es zahlreiche Modellierungsmethoden. Eine der am weitesten verbreiteten Methoden ist die Business Process Model Notation (BPMN) der Object Management Group (OMG) [3]. Diese Methode ist vor allem zur Prozessdokumentation geeignet. Ihre Schwächen lassen sich zum größten Teil in den (allgemeinen) Schwächen des Business Process Managements wiederfinden. Fallbasierte Szenarien oder entscheidungsstützende Mechanismen bietet die Modellierungsmethode BPMN nicht hinreichend an. Die OMG hat dafür zwei andere Methoden entwickelt, die Abhilfe schaffen können. Die CMMN wurde 2014 und die DMN 2015 vorgestellt. [14], [15] Problematisch an dieser Situation kann jedoch sein, alle Methoden passend miteinander zu verbinden, ohne dass dadurch Doppelstrukturen oder gar Widersprüche auftreten. [12] In Tabelle 1 sind die drei Methoden mit unterschiedlichen Eigenschaften abgebildet. Nachfolgend soll kurz auf wichtige Punkte der Modellierungsmethoden eingegangen und erläutert werden, inwieweit diese bei Notfällen eine besondere Bedeutung aufweisen.

TABELLE 1

	BPMN	CMMN	DMN
Core Concepts	Processes	Cases	Decisions
	Activities	Events	Rules
Focus	Transitional	Contextual	Applied
	Data	Information	Knowledge
Semantics	Procedural	Declarative	Functional
	Token	Event Condition Action (ECA)	First Order Logic (FOL)
	About Processing	About Managing	About Deciding

Unterschiede der Modellierungsmethoden BPMN, CMMN; DMN in Kernkonzepten, Fokus und Semantik [12]

#### A. Business Process Model and Notation (BPMN)

Die BPMN verknüpft Elemente wie Ereignisse (Events) und Aktivitäten (Activities) mit einem Sequenz- oder Nachrichtenfluss (Flow) und kann zwischen verschiedenen vordefinierten Situationen mithilfe logischer Operatoren (Gateways) unterscheiden und diese Aktionen auch bestimmten Rollen (Pools und Swimlanes) oder Artefakten (Data Objects) zuteilen.

Wie in Tabelle 1 ersichtlich ist, konzentriert sich die BPMN auf Prozesse und die Durchführung von Aktivitäten in dieser. Der Fokus der Methode liegt darauf, dass Daten von einer Aktivität zur nächsten durch Token weitergegeben werden. [12], [13]

Um die Granularität eines Prozesses anpassen zu können, besitzt die BPMN Subprozesse. Dort können, zusammengefasst in einem Element, weitere Aufgaben hinterlegt werden. [13] Mit diesen Subprozessen und entsprechenden Gateways kann eine Modellierung von Notfallsituationen erfolgen. Dafür müssen aber die Notfälle weitestgehend bekannt sein und entsprechend unflexibel oder unfähig ist die Methode, auf ungeplante Ereignisse zu reagieren. Auch wenn durch die Subprozesse ein Prozess in erster Linie nicht stark beeinflusst wird, können der zusätzliche Modellierungsaufwand und die daraus folgende Unübersichtlichkeit zu Missverständnissen und Unklarheiten führen. Allein weil es in einigen Notfallsituationen schwierig ist, eine Reihenfolge für die Aufgaben festzulegen, erscheint dieser Ansatz nicht praktikabel, da dies dazu führen kann, dass das Modell viele Gateways aufweist um jedwede Alternative zu berücksichtigen. Dadurch wird das Modell schnell groß und unübersichtlich wird und auch die Alternativen bereithalten, die nicht oder nie gewählt werden, weil sie sehr außergewöhnliche Zustände beschreiben oder weil sie nicht zur vorliegenden Notfallsituation passen. [6]

Dadurch ist die Flexibilität eingeschränkt, sodass im Notfall möglicherweise weder Standardprozesse angepasst noch Nicht-Standardprozesse passend erstellt werden können. Eine Erstellung von Nicht-Standardprozessen mit der BPMN ist nicht ausgeschlossen, aber sobald dort eine Änderung auftritt, ist eine Änderung im Prozess wieder schwierig zu erreichen.

#### B. Case Management Model and Notation (CMMN)

Eine Methode, die die Anpassung während der Runtime erlaubt, ist die CMMN. Der Fall wird von einem Plan umschlossen, welcher die Kontextinformationen vorgibt (Case Plan Model). So wie in der BPMN gibt es auch bei dieser Methode Events, welche durch bestimmte vorher eingetretene Bedingungen aktiviert werden können. Weitere wichtige Elemente sind Aufgaben (Tasks), welche in drei verschiedene Kategorien fallen können (human task, case task, process task). Aufgaben können auch die Eigenschaft besitzen wählbar zu sein (Discretionary Task) und von einem Mitarbeiter während der Runtime optional hinzugefügt werden zu können. Weitere Elemente sind Planfragmente (Plan Fragments) und Stufen (Stages). In einem Fall kann es Meilensteine (Milestones) geben, die bestimmte Zwischen- oder Endziele anzeigen. Weitere Elemente sind Plantabellen (Planning Tables), Schranken (Sentries) und Falldokumente (Case Files). Um Abhängigkeiten zwischen Aufgaben herzustellen, wird ein Konnektor (Connector) verwendet. Von der Methode ist es aber nicht vorgesehen, zu viele Aufgaben mit Konnektoren zu verbinden. [14]

Inhaltlich konzentriert sich diese Methode besonders auf Ereignisse (Events), welche während eines Geschäftsvorfalles auftreten können. Der Fokus liegt dabei auf Daten, die einem bestimmten Kontext zugeordnet werden und daraus Informationen generierenden und diese damit zu managen. [12], [14]

Im Zusammenhang mit Notfällen ergibt sich hier eine valide Möglichkeit sowohl ungeplante Nicht-Standardprozesse als auch Standardprozesse bearbeiten zu können. Isoliert betrachtet können ungeplante Nicht-Standardprozesse auch allein mithilfe von CMMN bearbeitet werden. Im größeren Kontext eingebunden und im besonderen Hinblick auf die Synchronisation



des Nicht-Standardprozesses ist aber eine Verbindung mit der BPMN sinnvoll. Eine Verbindung beider Methoden ist ebenso bei Standardprozessen sinnvoll, die in Teilen noch funktionieren, aber an bestimmten Stellen eine ad hoc Anpassung während der Laufzeit benötigen.

### C. Decision Model and Notation (DMN)

DMN besitzt zwei große Gruppen von Elementen. Zunächst gibt es Entscheidungselemente (Decision) und Inputfaktoren (Input), welche einer Entscheidung zugeordnet werden. Entscheidungen können außerdem von Wissensquellen (Knowledge Sources) abhängen. Diese Quellen können beispielsweise gesetzliche Regelungen, Entscheidungsregeln oder eine zugrundeliegende Prozedur sein. Einer Entscheidung kann eine Entscheidungstabelle zugeordnet werden, in der bestimmte Regeln mit beispielsweise Wahr- und Falschaussagen hinterlegt werden können. [15] Im Rahmen des WFMs werden Entscheidungstabellen von Nutzern im Vergleich zu anderen Methoden gut angenommen und verhältnismäßig intuitiv und leicht verständlich. [17]

Die dargestellten Fakten in Tabelle 1 umrahmen DMN im Kernkonzept durch Entscheidungen, die nach bestimmten Regeln getroffen werden. Der Fokus liegt auf dem der Anwendung der Entscheidungsregeln im Zusammenhang mit dem Kontext, sodass daraus Wissen generiert wird. [12], [15]

In einer Notfallsituation können einfache Entscheidungsregeln dazu beitragen, den Überblick zu behalten oder auch eine komplexe Situation übersichtlich aufzuspalten und verschiedene Ergebnisse darzustellen. Es kann sinnvoll sein, einige allgemeine Entscheidungsregeln vor dem Eintritt des Notfalls festzulegen. Insbesondere die Rollenverteilung im Sinne der Verantwortlichkeitsübertragung oder auch eine Umgestaltung der Entscheidungsgewalt von Rollen können in diesem Zusammenhang vorüberlegt werden.

## IV. ZUSAMMENFÜHRUNG DER MODELLIERUNGSMETHODEN

Nachdem die Methoden vorgestellt wurden, soll in diesem Bereich aufgezeigt werden, wie eine allgemeine gemeinsame Verwendung aussehen kann. In der Literatur gibt es unterschiedliche Meinungen, ob die BPMN mit Methoden des ACMs kompatibel ist, mit Gegenstimmen [21] und auch Fürsprechern. [4], [6], [20]

Wenn man zwischen den angesammelten Informationen aus Tabelle 1 Relationen herstellt, dann ist zu erkennen, dass die BPMN als Basis dienen muss, da sie daten- und prozessgetrieben ist. Daran schließt sich die CMMN mit den Kontextinformationen und darauf aufbauend die DMN mit dem hinzugefügten Wissen an. [4]

Auf dieser Grundlage und auch aufgrund der Tatsache, dass Unternehmen ihre Geschäftsprozesse in der Regel mit der BPMN modellieren bzw. schon modelliert haben, kann man sehr gut auf einen existierenden Prozess mit der CMMN aufsetzen. Wie bereits dargestellt, können in der BPMN aus Gründen der Übersichtlichkeit Prozesse in Subprozesse ausgelagert werden. Genau an dieser Stelle kann man mithilfe der CMMN ansetzen und einen Subprozess durch einen Fall organisieren. [4]

Vertiefend dazu können in einem Fall Entscheidungen der DMN hinterlegt werden um bestimmte Regeln zu definieren. [4] Dies kann in Notfällen nicht nur in den Phasen der Antwort (Response) oder Erholung (Recovery) von Nutzen sein, sondern auch in der Phase der Abschwächung (Mitigation), wie Riedel et al. am Beispiel eines Hochwasserwarnsystems zeigen [17]

Die OMG schlägt als Best Practice für diese Methoden in Verbindung mit bestimmten Elementen eine bestimmte Modellierungsmethode zur Verbesserung vor (Empfehlung). Die Zusammenfassung ist in Tabelle 2 zu sehen. Nachfolgend soll kurz darauf eingegangen werden. Weist eine BPMN viele logische Operatoren auf kann es empfehlenswert sein, anstelle der Gateways die DMN zur Modellierung des Sachverhalts zu verwenden. Dies kann insgesamt zu einer Vereinfachung des Modells führen. Befinden sich im Modell der BPMN viele Events, kann es ebenfalls sinnvoll sein, diese durch Fälle der CMMN abzulösen, da die Modellierung der verschiedenen Kontextinformationen, mit denen Events verbunden sind, dadurch effektiver dargestellt werden können. Wie schon zuvor erläutert, können Subprozesse der BPMN – insbesondere, wenn es sich dabei um ad hoc Prozesse handelt, bei dem die Reihenfolge oder der zeitliche Ablauf nicht klar ist – mit einem CMMN-Fall modelliert werden. In der CMMN können einzelne Elemente mit Konnektoren verbunden werden. Sollte es aber zu viele Verbindungen zwischen diesen Elementen geben kann dies ein Zeichen sein, dass eine Modellierung einzelner Abhängigkeiten mit der BPMN empfehlenswerter ist. [4], [6]

TABELLE 2

Method	Elemente	Empfehlung
BPMN	Gateways	DMN
BPMN	Events	CMMN
BPMN	Adhoc	CMMN
CMMN	Task Interdependencies	BPMN

Best Practice Ansätze zur Kombination von BPMN, CMMN und DMN [4]

## V. KRITERIEN ZUR ERSTELLUNG EINER GEMEINSAMEN MODELLIERUNG

Wie aus Kapitel IV ersichtlich, gibt es insbesondere von der OMG Bemühungen und Ansätze, diese drei Methoden miteinander zu kombinieren. Wie man anhand von Tabelle 2 erkennen kann, gibt es bereits Best Practice Empfehlungen, welche aufzeigen können, ob ein Prozess mit der richtig gewählten Methode gestaltet wurde.

In Anlehnung an das Notfallmodell aus Abbildung 1 und der in Kapitel II erwähnten Ergänzung einer Lessons Learned, sowie den im selben Kapitel unterschiedenen Prozessarten kann eine Umsetzung der einzelnen Schritte so aussehen, wie es in Tabelle 3 vorgeschlagen wird. Beim Eintreten einer Notfallsituation müssen zunächst Sofortmaßnahmen (in Tabelle 3 Response) ergriffen werden. Diese beziehen sich sowohl auf geplante Standardprozesse, als auch auf die Erstellung der zur Krisenbewältigung eingeführten ungeplanten Nicht-Standardprozesse. Durch den ad hoc Charakter der Notfallsituation wird

dafür die CMMN favorisiert. Im Fall der geplanten Standardprozesse wird eine Kombination der bereits vorhandenen BPMN mit CMMN und/oder DMN angestrebt. Wie im vorhergehenden Abschnitt erwähnt, könnte ein verändertes Rollenkonzept mithilfe der DMN umgesetzt werden. In der Recovery-Phase erfolgt dort weiterhin die Kombination der beiden Methoden, da es möglich ist, dass der Prozess noch nicht vollständig normalisiert durchgeführt werden kann. Bei ungeplanten Nicht-Standardprozessen kann es möglich sein, dass diese in der Recovery-Phase nicht mehr ausgeführt werden müssen, da der Notfall beseitigt ist und sich hier eher um die geplanten Standardprozesse gekümmert wird. Dennoch sollten die gewonnen Erkenntnisse in einer Lessons Learned verarbeitet werden. In dieser Phase sollte für geplante Standardprozesse unbedingt eine Kombination aus allen Methoden in Betracht gezogen werden, da diese in allen vorherigen Phasen Anwendung fanden.

TABELLE 3

	<b>Response</b>	<b>Recovery</b>	<b>Lessons Learned</b>
Anpassung der Standardprozesse	BPMN + CMMN + DMN	BPMN + CMMN	BPMN, DMN, CMMN
Erstellung eines Nicht-Standardprozesses	CMMN		CMMN

Mögliche Anwendung der Sprachen BPMN, CMN, DMN in Bezug auf Situation und Prozesse

Hinkelmann et al. schlagen deshalb Kriterien für eine gemeinsame Sprache bzw. Modellierungsmethodik von BPMN und CMMN vor. Die Kriterien beinhalten unter anderem die Beibehaltung wichtiger und übersichtlicher Elemente wie z.B. der Swimlanes der BPMN, aber auch eine Integration von den verschiedenen Aufgaben (Tasks) der CMMN in die BPMN und auch ein System zur besseren Entscheidungsfindung. [6] Eine Untersuchung von Marin et al. ergab außerdem eine hohe Ähnlichkeit der beiden Sprachen. [10] Bei der Verbindung mehrerer Modelle und Sprachen können Probleme auftreten, im Rahmen dieser Arbeit noch untersucht werden müssen. [16] Aufgrund der Tatsache, dass alle drei Methoden die OMG als gemeinsamen Entwickler haben und es bereits Ansätze gibt, die diese Methoden zu verbinden versuchen, wird erwartet, dass eine Umsetzung dieses Konzepts in einem eigenen Tool möglich sein sollte.

## VI. METHODISCHE UMSETZUNG

Als Rahmenwerk und methodisches Vorgehen möchte die Arbeit sich auf das Design Science Research Modell von Peffers et al. stützen [15]. Innerhalb der Methode findet die Bearbeitung eines Problems in sechs Aktivitäten statt (siehe Abbildung 2).

Die erste Aktivität beschäftigt sich mit der Identifikation des Problems und der Motivation. Diese wurde in Kapitel I umrissen.

Die zweite Aktivität befasst sich mit der Zieldefinition. Diese soll zunächst mit einer Ausarbeitung der theoretischen Grundlagen stattfinden. Erster Grundstein dessen ist eine systematische Literaturanalyse nach vom Brocke et al. [23] Die bisherige Literaturanalyse hat ergeben, dass es im Rahmen des Notfallmanagements noch keine Ansätze gibt, die eine Verbindung der in Kapitel III und IV vorgestellten Methoden anstreben. Weiterhin soll die Literaturrecherche ergründen, welche anderen oder vergleichbaren Ansätze es gibt und in wie weit diese relevant für die weitere Arbeit sind.

Weiterhin beinhaltet das Vorgehen von Peffers et al. das Entwerfen und Entwickeln. Angewandt auf das Dissertationsvorhaben ist damit die konkrete Erstellung und Verbindung der in Kapitel III ausgearbeiteten Methoden gemeint, so wie die in Kapitel IV mögliche Integration der einzelnen Sprachen. Diese Verbindung soll mithilfe eines eigens erstellten Tools stattfinden. Genauere Spezifikationen der Umsetzung wurden bisher noch nicht festgelegt.

Im vierten Schritt des Design Science Researchs ist eine Demonstration vorgesehen. Diese soll im Dissertationsvorhaben durch den theoretischen Test des Tools und des Ansatzes umgesetzt werden. Generell besteht ein Interesse, diesen Ansatz mit anderen Notfallbewältigungsmechanismen zu vergleichen. Der Vergleich kann mithilfe einer Kriterienliste durchgeführt werden. Diese theoretische Validierung des Ansatzes kann mithilfe von Simulationen erfolgen. Das Erkenntnisinteresse in diesem Bereich liegt vor allem darin zu vergleichen, wie sich die Performance des geplanten Standardprozesses mit den im Notfall angewandten Überbrückungsmechanismen gegenüber dem geplanten Standardworkflow im Normalbetrieb gestaltet.

Weiterhin schlagen Peffers et al. eine Evaluation vor. In dieser Phase der Arbeit ist zur Überprüfung der theoretischen Grundlagen und der Erkenntnissen der Simulation ein Experiment angedacht. Hier soll im letzten Schritt überprüft werden, inwieweit in einem praxisähnlichen Umfeld das Tool angenommen wird und ob in einer Notfallsituation adäquat damit umgegangen werden kann. In diesem Experiment ist geplant, unter zwei Gruppen einer Organisation eine Notfallsituation hervorzurufen. Einer der Gruppen erhält vor dem Notfall Zugriff und eine Einführung in das Tool, die Vergleichsgruppe soll ohne das Tool versuchen, den Notfall zu bewältigen (Kontrollgruppe). Aus dem Experiment ermittelt die Arbeit sowohl Nutzerfeedback als auch eine Aussage über die praktische Wirksamkeit des Tools. Ein besonderes Augenmerk liegt auf der zeitlichen Nutzung des Tools. Im Notfall muss es möglich sein, innerhalb eines kurzen zeitlichen Rahmens eine Lösung zu finden.

Die letzte Aktivität beschäftigt sich mit der Kommunikation des Arbeitsfortschrittes und der Ergebnisse. Im Rahmen des Dissertationsvorhabens werden mehrere Zwischenüberprüfungen vor entsprechendem Fachpublikum stattfinden.

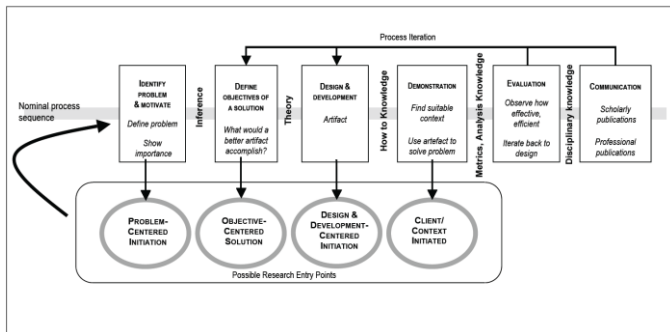


Abbildung 2: Design Science Research Methodology (DSRM) Prozessmodell [15]

## VII. ZUSAMMENFASSUNG UND AUSBLICK

Das durchzuführende Dissertationsvorhaben befindet sich bisher noch im Anfangsstadium. Die Arbeit beschäftigt sich zum jetzigen Stand noch mit dem Eingrenzen und Gestalten der Rahmenbedingungen und ebenso mit den drei Modellierungssprachen und dem Vorhaben, diese ineinander zu integrieren. Des Weiteren wird zurzeit an der Erstellung der Kriterienliste für den Vergleich der verschiedenen Methoden gearbeitet. Im nächsten Schritt steht an, diese Sprachen in einem ersten Ansatz im Hinblick auf Notfälle zielführend miteinander zu verbinden, damit die Grundlagen für einen Prototypen des angestrebten Tools geschaffen werden können.

In welchen Rahmen das praktische Experiment durchgeführt werden kann, ist zum jetzigen Zeitpunkt noch ungeklärt. Es wäre denkbar das Experiment mit einem KMU aus der Region durchzuführen. Ein Experiment im Rahmen einer Lehrveranstaltung mit Studierenden der Universität wäre aber auch denkbar.

Auch wenn es durch die OMG bereits Ansätze gibt, die drei Sprachen BPMN, CMMN und DMN zu verbinden, geschah dies noch nicht im Zusammenhang eines Notfalls. Wie durch Abbildung 1 erkenntlich ist, betrachtet das Notfallmanagement verschiedene Phasen, in denen unterschiedliche Ziele vorherrschen können und durch die eine Variation der Methoden nötig sein kann. Ein erster möglicher Entwurf zeigt Tabelle 3. Die Arbeit hofft damit die Problemstellung beantworten zu können, ob eine Verbindung dieser Methoden ein erfolgreiches Notfallmanagement unterstützen kann.

## VIII. LITERATUR

- [1] Al-Dahash, H., Thayaparan, M., Kulatunga, U.: Understanding the Terminologies: Disaster, Crisis and Emergency. Proceeding of the 32th Annual ARCOM Conference, 5.-7. September 2016, in: Association of Researchers in Construction Management (2016), S.1191–1200.
- [2] Alexander, D.: Towards the development of a standard in emergency planning, in: Disaster Prevention & Management: An International Journal, 2005. Jg., S.158–175.
- [3] Allweyer, T. BPMN 2.0 - Business Process Model and Notation (German Edition), 2015.

- [4] Berardi, T., Baudoin, C., Gagné (Hrsg.). OMG Webinar: How to use BPMN, CMMN and DMN together, 06.07.2016.
- [5] Bird, M. E.: The "Phases" of Emergency Management. Background Paper (Abrufdatum: 25.07.2018).
- [6] Hinkelmann, K., Pierfranceschi, A.: COMBINING PROCESS MODELLING AND CASE MODELLING. Online im Internet. URL: [https://irf.fhnw.ch/bitstream/handle/11654/10620/Case\\_Process\\_Modelling\\_MeTTeG14\\_Final.pdf?sequence=1](https://irf.fhnw.ch/bitstream/handle/11654/10620/Case_Process_Modelling_MeTTeG14_Final.pdf?sequence=1), Stand: 14.10.2015 (Abrufdatum: 27.09.2018).
- [7] Kirsch-Pinheiro, M., Rychkova, I.: Dynamic Context Modeling for Agile Case Management, in: On the Move to Meaningful Internet Systems: OTM 2013 Workshops, 2013. Jg. (2013), S.144–154.
- [8] Kushnareva, E., Rychkova, I., Le Grand, B.: Modeling Business Processes for Automated Crisis Management Support: Lessons Learned, in: IEEE Ninth International Conference on Research Challenges in Information Science (18.05.2015).
- [9] Marin, M. A., Lotriet, H., Van Der Poll, J. A.: Measuring Method Complexity of the Case Mesuring Method Complexity of the Case Management Modeling and Notation (CMMN), in: SAICSIT '14, September 29-October 1 2014, Centurion, South, 2014. Jg., S.209–216.
- [10] Neal, D.: Reconsidering the Phases of Disaster, in: International Journal of Mass Emergencies and Disasters Vol 15, No. 2, 2009. Jg. (1997), S.239–264.
- [11] OMG (Hrsg.): BPMN, CMMN AND DMN SPECIFICATIONS AT OMG. Online im Internet. URL: <https://www.omg.org/intro/TripleCrown.pdf> (Abrufdatum: 26.09.2018).
- [12] OMG (Hrsg.): Business Process Model and Notation (BPMN). Version 2.0. Online im Internet. URL: <https://www.omg.org/spec/BPMN/2.0/About-BPMN/> (Abrufdatum: 27.08.2018).
- [13] OMG (Hrsg.): Case Management Model and Notation (CMMN). Version 1.1. Online im Internet. URL: <https://www.omg.org/spec/CMMN/> (Abrufdatum: 27.09.2018).
- [14] OMG (Hrsg.): Decision Model and Notation. Version 1.2. Online im Internet. URL: <https://www.omg.org/spec/DMN/1.2/Beta1/>, Stand: 01.06.2018 (Abrufdatum: 27.09.2018).
- [15] Peffers, K., Tuunanen, T., Rothenberger M. A., Chatterjee, S.: A Design Science Research Methodology for Information Systems Research, in: Journal of Management Information Systems (2007), S.45–78.
- [16] Pottinger, R. A., Bernstein, P. A.: Merging Models Based on Given Correspondences, in: Proceedings of the 29th VLDB Conference, Berlin, Germany, 2003, 2003. Jg. (2003), S.862–873.
- [17] Riedel, F., Chaves, F.: Workflows and Decision Tables for Flexible Early Warning Systems, in: Proceedings of the 9th International ISCRAM Conference – Vancouver, Canada, April 2012 (2012), S.1–5.
- [18] Ritchie, H., Roser, M.: Natural Catastrophes. Online im Internet. URL: <https://ourworldindata.org/natural-catastrophes> (Abrufdatum: 14.10.2018).
- [19] Sell, C., Braun, I.: Using a Workflow Management System to Manage Emergency Plans, in: Proceedings of the 6th International ISCRAM Conference – Gothenburg, Sweden, May 2009, 2009. Jg.
- [20] Silver, B.: BPMN and CMMN Compared. Online im Internet. URL: <https://methodandstyle.com/bpmn-cmmn-compared/>, Stand: 14.05.2014 (Abrufdatum: 27.09.2018).
- [21] Swenson K. D.: Position: BPMN Is Incompatible with ACM, in: International Conference on Business Process Management; BPMN 2012: Business Process Management Workshop (2012), S.55–58.
- [22] Swiss Re (Hrsg.): Natural catastrophes and man-made Disasters in 2011-2017. Online im Internet. URL: <http://www.swissre.com/library/archive/> (Abrufdatum: 16.06.2018).
- [23] Vom Brocke, J., Simons, A., Niehaves, B., Niehaves, B., Reimer, K., Plattfaut, R., and Cleven, A.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process, in: European Conference on Information Systems (ECIS) (2009).

# Business Process Compliance and Change (Extended Abstract)

Tobias Seyffarth

Chair of Information Management  
Martin Luther University Halle-Wittenberg  
06108 Halle (Saale), Germany  
tobias.seyffarth@wiwi.uni-halle.de

**Keywords—** *Business Process Adaption; Business Process Compliance; Change; IT Component*

## I. MOTIVATION

Business Process Compliance (BPC) denotes the execution of business processes in adherence to applicable compliance requirements [1]. A compliance requirement is an assertion resulting from the interpretation of compliance sources, such as laws, norms, standards or internal policies [2]. In addition to business processes, compliance requirements may also place demands against IT components which are sometimes necessary to execute business activities ([3, 4]).

In the literature, various approaches to ensure BPC during the design time of a business process are discussed. One approach is the separate modeling and integration of compliant process fragments into the business process [5, 6]. We call such a compliant process fragment a compliance process which is a process part that consists of at least one compliance-related activity that satisfies at least one compliance requirement and thus ensures BPC [7].

However, many factors such as business process improvements, new technologies or outsourcing decision can lead to business process change, including changed IT components or compliance requirements ([8, 9]). Thus, the goal is to ensure BPC during the design time of the business process with changes to compliance requirements, business activities and IT components. Therefore, we raise the following three research questions:

- RQ1: How can compliance requirements, business processes, compliance processes and IT components be modelled in a common model?
- RQ2: What are the interactions between BPC and business process change and how can they be determined?
- RQ3: How can a business process be adapted to ensure BPC despite business process change?

In order to reach the research goal, we develop artifacts by following the design research paradigm proposed by Hevner et al. [10]. Further, the knowledge contribution of the developed artifacts is each an exaptation which extends known solutions to

new problems [11]. In Section 2, we propose a method that ensures BPC despite business process change. Finally, Section 3 concludes this contribution.

## II. TOWARDS A METHOD TO ENSURE BPC WITH CHANGES TO VARIOUS PROCESS PERSPECTIVES

We propose a three-step method that ensures BPC despite changes to meet the objectives of our solution. First the compliance requirements, business processes, and IT components have to be modelled into one common model. Second the interactions between BPC and business process change have to be analyzed. In case of a detected compliance violation, the business process have to be adapted through the integration of alternative compliance processes in a third step. Each step consists of different artifacts and will be briefly described in the following sections. We also refer to already existing artifact demonstrations.

### A. Model a Common Model

A precondition for analyzing the interrelation between BPC and business process change is the modeling of compliance requirements, business processes, compliance processes and IT components within a common model. Therefore a formal representation of each model type is required. In [12] we proposed a method that transforms and integrates each formal representation into one common model. The common model is a directed graph whereby each single element of the separate models is represented by a node that is connected to other nodes by directed edges. The efficacy of this method is demonstrated by the software prototype BCIT [13]. Thus, RQ1 is answered by these method.

### B. Analyze the Interactions between BPC and Business Process Change

In the literature various business process change patterns are discussed (e.g. [14]). In [12], we proposed a method that adapts graph search techniques in order to analyzes the interrelation between BPC and business process change considering the business process change patterns “delete element” and “replace element”. Thus, this method is used to answer RQ2.

The pattern “delete element” removes an element which can be either a compliance requirement, business activity, compliance process or IT component. The pattern “replace element” replaces one of the aforementioned elements by a new one.

The change pattern “delete element” has two different impacts on BPC. First, there is a compliance violation and second a compliance requirement can become obsolete. The change pattern “replace element” has two different effects, too. The replaced element must satisfy direct and indirect demands by compliance requirements. The efficacy of this method is demonstrated by the software prototype BCIT, too [13].

### C. Adapt Business Process through the Integration of Compliance Processes

In case of a compliance violation due to a change of compliance requirements, business activities, compliance processes and IT components, the business process may be adapted to avoid a compliance violations. An appropriate method for the business process adaption answers RQ3.

Since a compliance requirement can be satisfied by more than one compliance process [7], the separate modelling and integration of alternative compliance processes in the business process is a promising approach to ensure BPC when changing compliance requirements, business activities and IT components.

An example of alternative compliance processes is the control of the number of bacteria in drinking water. The number of bacteria can be controlled by alternative compliance processes and at different points in the supply process e.g., by analyzing the water sources, the water depots, the hand-over to the consumer or by the consumer [15]. Each of these compliance processes can have different properties as well as different effectiveness and cost. The compliance process taxonomy in [7] gives an overview over possible property dimensions and characteristics.

## III. CONCLUSION

Many factors such as business process improvements, new technologies or outsourcing decision can lead to business process change, including changed IT components or compliance requirements. However, to avoid compliance violations, the interactions between BPC and business process change have to be determined and business processes may have to be adapted to be compliant. Consequently, we presented a method towards ensuring BPC with changes of compliance requirements, business activities, and IT components.

The proposed method has a narrow focus on business processes and compliance requirements. There are other views on business processes besides the control flow perspective, compliance requirements and IT components (e.g. an organizational and data perspective). Additionally, compliance requirements place more demands against the business process than the control flow perspective of a business process. Furthermore, we assume that the existence or absence of an IT component is the only prerequisite to execute a compliance process [7].

In a next step, the proposed method to adapt business processes through the integration of alternative compliance processes will be further developed and demonstrated by a software prototype. Furthermore, an evaluation of the proposed method is planned, since the evaluation of artifacts is a main guideline in design science research [10].

## REFERENCES

- [1] G. Governatori and S. Sadiq, “The Journey to Business Process Compliance,” *Handbook of research on business process modeling*, pp. 426–454, 2009.
- [2] O. Turetken, A. Elgammal, W.-J. van den Heuvel, and M. Papazoglou, “Enforcing compliance on business processes through the use of patterns,” *19th ECIS 2011*, 2011.
- [3] Committee of Sponsoring Organizations of the Treadway Commission (COSO), *Internal Control - Integrated Framework: Framework and Appendices*, 2012.
- [4] *The Audit of Financial Statements in an Information Technology Environment: IDW AuS 330*, 2002.
- [5] D. Schumm, F. Leymann, Z. Ma, T. Scheibler, and S. Strauch, “Integrating Compliance into Business Processes: Process Fragments as Reusable Compliance Controls,” *Proc. of the MKWI'10*, 2010.
- [6] S. Sackmann and K. Kittel, “Flexible Workflows and Compliance: A Solvable Contradiction?!,” in *Management for Professionals, BPM - Driving Innovation in a Digital World*, J. Vom Brocke and T. Schmiedel, Eds., Cham: Springer International Publishing, 2015, pp. 247–258.
- [7] T. Seyffarth, S. Kühnel, and S. Sackmann, “A Taxonomy of Compliance Processes for Business Process Compliance,” *15th International Conference on Business Process Management, Business Process Management Forum. In: Lecture Notes in Business Information Processing (LNBIP)*, 2017.
- [8] P. Rudzajs and I. Buksa, “Business Process and Regulations: Approach to Linkage and Change Management,” in *Lecture Notes in Business Information Processing, Perspectives in Business Informatics Research*, J. Grabis and M. Kirikova, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 96–109.
- [9] W. Fdhila, C. Indiono, S. Rinderle-Ma, and M. Reichert, “Dealing with change in process choreographies: Design and implementation of propagation algorithms,” (eng), *Information Systems*, vol. 49, pp. 1–24, 2015.
- [10] A. R. Hevner, S. T. March, J. Park, and S. Ram, “Design Science in Information Systems Research,” *MIS Quarterly*, vol. 28, no. 1, pp. 75–105, <http://dl.acm.org/citation.cfm?id=2017217>, 2004.
- [11] A. R. Hevner and S. Gregor, “Positioning and Presenting Design Science Research for Maximum Impact,” *MIS Quarterly*, vol. 37, no. 2, <http://misq.org/positioning-and-presenting-design-science-research-for-maximum-impact.html>, 2013.
- [12] T. Seyffarth, S. Kühnel, and S. Sackmann, “Business Process Compliance and Business Process Change: An Approach to Analyze the Interactions,” *Business Information Systems. BIS 2018. Lecture Notes in Business Information Processing*, no. 320, pp. 176–189, [https://link.springer.com/chapter/10.1007/978-3-319-93931-5\\_13](https://link.springer.com/chapter/10.1007/978-3-319-93931-5_13), 2018.
- [13] T. Seyffarth and K. Raschke, “BCIT: A Tool for Analyzing the Interactions between Business Process Compliance and Business Process Change,” *Proceedings of the Dissertation Award and Demonstration, Industrial Track at BPM 2018*, pp. 81–85, <http://ceur-ws.org/Vol-2196/>, 2018.
- [14] B. Weber, M. Reichert, and S. Rinderle-Ma, “Change patterns and change support features – Enhancing flexibility in process-aware information systems,” *Data & Knowledge Engineering*, vol. 66, no. 3, pp. 438–466, [http://dbis.eprints.uni-ulm.de/335/1/DKE\\_WRR08.pdf](http://dbis.eprints.uni-ulm.de/335/1/DKE_WRR08.pdf), 2008.
- [15] K. Kittel, S. Sackmann, H. Betke, and M. Hofmann, “Achieving Flexible and Compliant Processes in Disaster Management,” *46. Hawaii Internat. Conference on System Sciences (HICSS'13)*, pp. 4687–4696, 2013.

# RTML – Eine Modellierungssprache zur Wissensentwicklung im Kontext klinischer Befundberichterstellung

Manuel Siekmann  
Institut für Wirtschaftsinformatik  
Universität Leipzig  
Leipzig, Deutschland  
siekmann@wifa.uni-leipzig.de

**Kurzfassung** — Ein erheblicher Teil der Arbeitszeit deutscher Klinikärzte ist reine Verwaltungsarbeit ohne Patientenkontakt. Zentrale Bestandteile dabei sind die medizinische Dokumentation und die Befundberichterstellung. Insbesondere durch den stetig steigenden Kostendruck gewinnen die Faktoren Zeit und Qualität bei der Patientenversorgung immer mehr an Bedeutung. Aktuelle Informationssysteme unterstützen den Anwender nicht ausreichend im Dokumentationsprozess. Zielsetzung dieser Forschungsarbeit ist der Entwurf und die Evaluation einer Modellierungssprache zur kollaborativen Entwicklung und zum herstellerunabhängigen Austausch klinikübergreifend genutzter Befundvorlagen. Durch diese Vorlagen soll die Qualität der Dokumentation verbessert und der Aufwand bei der Befunderstellung sowie bei der Informationsrückgewinnung reduziert werden. Der Nutzen für Autoren von Befundvorlagen, die an der Behandlung beteiligten Ärzte und Hersteller klinischer Informationssysteme soll in einem anschließenden Feldversuch evaluiert werden.

**Keywords** — Medical Report, Semantic Web, Ontology, OWL, RDF, NLG, E-Collaboration, Social Software, Design Science

## I. PROBLEMBEREICH UND MOTIVATION

Die ökonomische Entwicklung im Gesundheitswesen steht seit Jahren im Fokus wissenschaftlicher Forschung. Die Einflussfaktoren sind vielfältig, häufig werden der demografische Wandel, der medizinische Fortschritt und die Lohnentwicklung angeführt [1]. Der dadurch entstandene Rationalisierungsdruck, beispielsweise durch die seit 2004 in Deutschland eingeführte leistungsorientierte Abrechnung, führt dazu, dass Kliniken und infolgedessen auch das medizinische Personal verstärkt auf ein Gleichgewicht im Spannungsfeld zwischen Qualität und Kosten bei der Patientenversorgung achten müssen.

Die sorgfältige und umfangreiche Dokumentation der medizinischen Behandlung ist ein zentraler Bestandteil der Patientenversorgung. Neben dem Zeit- und Kostendruck, haben, insbesondere durch die Vielzahl der beteiligten Akteure, auch die planmäßige Durchführung, die Vollständigkeit und die Qualität einen hohen Stellenwert bei der medizinischen Berichterstellung [2]. Die Dokumentation dient der Patientenversorgung, der Leistungsabrechnung, der rechtlichen Absicherung und als Basis für wissenschaftliche Studien [3]. Für

eine qualitative und ökonomische Versorgung sollen Befunddaten sorgfältig und dennoch schnell erfasst werden. Die Dokumentationspflicht beansprucht jedoch einen erheblichen Teil der ärztlichen Arbeitszeit. Verschiedene Studien haben dieses Problem quantifiziert und kamen dabei zu dem Ergebnis, dass der Dokumentationsaufwand täglich mehr als drei Stunden pro Arzt beträgt [4–6]. Elektronische Systeme zur Verbesserung der medizinischen Dokumentation werden bereits seit mehr als 25 Jahren erforscht [7]. Ärzte haben häufig den Wunsch mehr Zeit in die eigentliche Patientenversorgung zu investieren und den Verwaltungsaufwand durch unterstützende Systeme zu reduzieren [8].

Für die Beschreibung häufiger Beobachtungen gibt es bereits jetzt Informationssysteme, die Vorlagen anbieten, um typische Fälle zeiteffizient zu erfassen. Das Wissen hinter diesen Vorlagen wird, meist in Zusammenarbeit mit medizinischen Fachgesellschaften, redaktionell erstellt. Im Rahmen dieser Arbeit soll eine offene Beschreibungssprache entwickelt werden, die aktuelle Anforderungen der Klinikärzte an eine individuelle und flexible Dokumentation erfüllt und gleichzeitig die maschinelle Weiterverarbeitung und Auswertung der Daten ermöglicht. Die Beschreibungssprache soll neben der Individualisierung auch den herstellerunabhängigen Austausch und die Mitarbeit einer breiten Anwenderbasis direkt am Erstellungsprozess von Befundvorlagen ermöglichen [9].

## II. ANFORDERUNGEN AN DIE BEFUNDBERICHTERSTELLUNG

Die Befunddokumentation ist eine abstrakte Abbildung der Behandlung und enthält alle nötigen Informationen für die Patientenversorgung. Um später Nutzen aus der Dokumentation zu ziehen, muss diese im Vorfeld geplant und vom behandelnden Arzt sorgfältig und systematisch durchgeführt werden [3]. Die folgenden zwei Abschnitte sollen zeigen wofür Befundberichte benötigt werden und welche Besonderheiten bei der medizinischen Dokumentation beachtet werden müssen.

### A. Zweck der medizinischen Befunddokumentation

Die Dokumentation hat administrative und medizinische Empfänger mit entsprechend unterschiedlichen Anforderungen an den Befundbericht [10]. Die Forschung im Gesundheitswesen führt dazu, dass Fachbereiche immer weiter spezialisiert werden um das Wissen zu beherrschen. Die

Kommunikation zwischen den beteiligten Akteuren wie Ärzten, Fachärzten, Krankenkassen und dem Pflegepersonal wird auch deshalb immer wichtiger [11]. Die Dokumentation ist hierbei eine wichtige Kommunikationshilfe. Neben der Kommunikation dient die Dokumentation auch als Erinnerungshilfe, um ein Gesamtbild des Falls und dessen chronologischer Entwicklung zu erhalten.

Ein weiterer Grund für genaue ärztliche Aufzeichnungen ist die Leistungsabrechnung. Leistungen und Diagnosen werden in Fallkosten zusammengefasst und unter Berücksichtigung von Nebendiagnosen, wie beispielsweise Komplikationen oder Komorbidität, in Pauschalen abgerechnet [12].

Auch aus juristischer Sicht muss sich ein Arzt gegen straf- und zivilrechtliche Klagen absichern. Die Dokumentation dient dabei der nachträglichen Rechtfertigung des Vorgehens. Die Dokumentationspflicht ist bereits in der Berufsordnung deutscher Ärzte vorgegeben. Im Behandlungsvertrag zwischen Arzt und Patient wird die Aufzeichnungspflicht festgelegt. Das Bürgerliche Gesetzbuch definiert die Art und den Umfang der medizinischen Dokumentation [13].

Darüber hinaus werden Untersuchungsdaten für die Forschung generalisiert, um daraus Erkenntnisse abzuleiten, die dabei helfen, die medizinische Versorgung zu verbessern. Neben der Forschung liefert die Befunddokumentation auch retrospektiv Erkenntnisse für die ärztliche Ausbildung [3].

### B. Anforderungen und Besonderheiten

Befundberichte haben generalisierbare Strukturen. Umso spezifischer aber eine Fallbeobachtung beschrieben wird, desto individueller sind die zu erfassenden Merkmale und auch identische Ausprägungen würden unabhängige Beobachter subjektiv in unterschiedlicher Art und Tiefe dokumentieren. Erwartet wird von Klinikärzten ein entsprechend flexibles System zur Befunddokumentation [14]. Auch die Wiederverwendbarkeit der Daten spielt eine große Rolle, zum einen in der Informationsrückgewinnung, zum anderen für Automatisierung, beispielsweise von Berechnungen oder der automatischen Übernahme von Messwerten in den Bericht. Um die Anforderungen erfüllen zu können, müssen die Daten miteinander vernetzt sein und auch implizites Wissen abgebildet werden.

Die sorgfältige Befunddokumentation ist die Basis für eine bedarfsgerechte Informationsversorgung und Grundlage der erfolgreichen Kommunikation zwischen den an der Behandlung beteiligten Akteuren. Medizinische Ordnungssysteme ermöglichen die semantische Interoperabilität. Solche Ordnungssysteme können Glossare, Thesauri, Taxonomien oder Ontologien sein [15]. Das in Deutschland verwendete Klassifikationssystem zur Leistungsabrechnung ist bereits gesetzlich standardisiert und kann so beispielsweise für die Kommunikation zwischen Kliniken und Krankenkassen genutzt werden. Das Deutsche Institut für Medizinische Dokumentation und Information (DIMDI) hat hierfür die „Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme“ (ICD) der Weltgesundheitsorganisation (WHO) und die „Operationen- und Prozedurenschlüssel“ (OPS) angepasst. Neben diesen Klassifikationssystemen gibt es im Bereich der medizinischen Dokumentation einige weitere

domänenspezifische Modelle mit unterschiedlicher Granularität und Ausrichtung [16].

Eine weitere Besonderheit der Befundberichterstellung ist die Qualitätssicherung. Klinische Daten werden für die Auswertung der Behandlungserfolge genutzt und umgekehrt entwickeln Fachgesellschaften auf Grundlage klinischer Studien praxisorientierte Leitlinien, um medizinischem Personal eine systematisch erstellte Entscheidungshilfe für die Patientenversorgung zur Verfügung zu stellen. Einige Fachgesellschaften, wie beispielsweise die Deutsche Gesellschaft für Gastroenterologie (DGVS), veröffentlichen umgekehrt auch Leitlinien für eine vollständige, qualitätsgesicherte Dokumentation. Hierfür werden die zu dokumentierenden Eigenschaften, die Terminologie und die Struktur von Fachgesellschaften formal definiert und können von Systemanbietern in geschlossenen Systemen weiterverarbeitet werden [17].

### III. METHODIK UND VORGEHENSWEISE

Die Forschungsarbeit ist gestaltungsorientiert nach den Richtlinien von Hevner aufgebaut [18] und wird methodisch nach Peffers [19] in einem iterativen Vorgehensmodell organisiert. Die folgende Darstellung zeigt die verschiedenen Phasen und die dort verwendeten Methoden und Techniken [20].

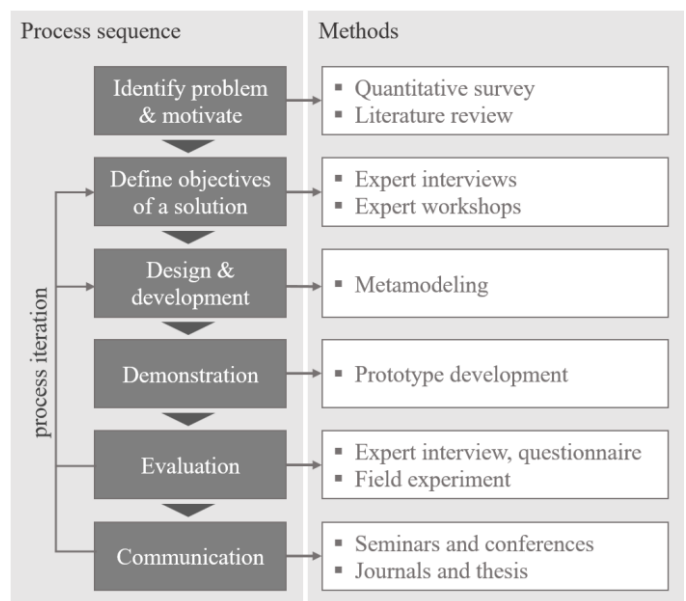


Abb. 1: Vorgehensmodell basierend auf Peffers (2007)

In der ersten Phase wurden eine Literaturrecherche [21] und eine quantitative Befragung durchgeführt, um den Forschungsbedarf und die Problemstellung zu analysieren. Es konnte gezeigt werden, dass Bedarf nach flexiblen Systemen zur Unterstützung der strukturierten Befunddokumentation besteht und insbesondere die Wissensentwicklung in bisherigen Arbeiten kaum berücksichtigt wurde.

In der zweiten Phase werden die Anforderungen und Forschungsziele definiert. Für diesen Zweck werden Experteninterviews [22] und Expertenworkshops [23] eingesetzt. Die Experteninterviews werden mit Klinikärzten

verschiedener Fachbereiche der Inneren Medizin durchgeführt, um die individuellen und die fachbereichsspezifischen Anforderungen an die Befunddokumentation zu erfassen und zu generalisieren. Als Techniken werden hierfür ein Leitfaden und Feldnotizen eingesetzt. Die Anforderungen an die Wissensentwicklung werden in Expertenworkshops mit Mitarbeitern medizinischer Redaktionen analysiert. Technisch werden hierfür Mockups eingesetzt, um das Werkzeug für die Wissensmodellierung zu entwickeln und die Anforderungen an die Modellierungssprache zu erfassen.

Auf der Anforderungsanalyse aufbauend werden eine Beschreibungssprache für Befundvorlagen und Handlungsempfehlungen für die Wissensrepräsentation im Kontext der medizinischen Befunddokumentation und die technische Umsetzung in einem Informationssystem modelliert, prototypisch umgesetzt und demonstriert.

Das Artefakt wird gegen die Forschungslücke und gegen die Realwelt evaluiert [24]. In der ersten Phase wird der entwickelte Prototyp mit Experten qualitativ in einer kontrollierten Umgebung bewertet. In der zweiten Phase wird das System in einem Feldexperiment eingesetzt. Das System wird mit Kooperationspartnern aus der Industrie pilotiert, um mit Hilfe von Statistiken die Leistungsfähigkeit in einer realen Umgebung zu bewerten.

Die letzte Phase der Arbeit beschreibt die Diffusion der Forschungsergebnisse. Fokus der Arbeit ist ein Beitrag zur gestaltungsorientierten Forschung klinischer Informationssysteme und zur Wissensmodellierung.

#### IV. BISHERIGE ANSÄTZE UND FORSCHUNGSLÜCKE

In diesem Abschnitt werden die Konzeptualisierung und Ergebnisse der ersten Phase [25] zusammenfassend dargestellt. Ziel der durchgeführten Literaturrecherche nach Vom Brocke [21] war die Verknüpfung der Forschungsarbeit mit bisherigen Ansätzen und die Identifikation zentraler Herausforderungen durch die Betrachtung früherer Forschungsfragen. Der Fokus lag dabei auf den Ergebnissen, den dahinterliegenden Theorien sowie auf der praktischen Anwendung der entwickelten Artefakte.

Die Systeme zur Unterstützung der computergestützten Befunddokumentation wurden in der Literatur analysiert und in Klassen differenziert. So lässt sich die Art der Dateneingabe in strukturierte und unstrukturierte Methoden sowie in manuelle und akustische Systeme unterscheiden. Die identifizierten Ansätze haben charakteristische Stärken und Schwächen. Die Dokumentation befindet sich beispielsweise in einem Spannungsfeld zwischen sprachlicher Flexibilität und Struktur [26]. Strukturierte Systeme arbeiten in der Regel mit einer hinterlegten Wissensdatenbank und können eine Komponente zur Textgenerierung besitzen. Diese Textgenerierung kann in flache, tiefe und hybride Systeme unterteilt werden [27, 28].

Aus der Literatur wurden strukturierte Systeme mit einer Komponente zur Textgenerierung abgeleitet, um die Stärken und Schwächen ähnlicher Systeme zu vergleichen. Für die Literaturrecherche wurden Schlüsselwörter und Kombinationen definiert und im Anschluss die zu durchsuchende Literatur festgelegt. Zur Übernahme der wichtigsten Journale und

Beiträge in das Ergebnis, wurden die zwölf forschungsstärksten Journale aus dem Bereich der Medizininformatik nach dem Google-h5-Ranking (2010–2014) sowie drei wissenschaftliche Datenbanken und drei Tagungsbände in die Recherche einbezogen.

In 162 Beiträgen wurden 21 Systeme zur Unterstützung der Befunddokumentation identifiziert und bewertet. Es konnte gezeigt werden, dass keines der Systeme die ermittelten Anforderungen erfüllt. Insbesondere die nötige Flexibilität und die Möglichkeit zur Individualisierung und zur offenen Wissensentwicklung wurden in bisherigen Ansätzen nicht ausreichend berücksichtigt.

Durch eine Online-Befragung konnte ein Überblick praktisch etablierter Dokumentationsarten verschiedener medizinischer Fachbereiche und praxisnahe Anforderungen gewonnen werden. Hierfür wurden Ärzte der Inneren Medizin, Pathologie und Chirurgie befragt und 118 Antworten ausgewertet. Die Zufallsstichprobe ist in der Altersstruktur repräsentativ. Von den befragten Klinikärzten sind 85 Prozent unzufrieden mit dem administrativen Aufwand. In 71 Prozent der Fälle wurde das aktuelle Dokumentationssystem als Ursache dafür genannt. Der häufigste Grund hierfür ist eine zu geringe Flexibilität bzw. die unpassenden Eingabemöglichkeiten strukturierter Systeme, gefolgt von einem zu hohen zeitlichen Aufwand bei der Berichterstellung. Die Befragung hatte gezeigt, dass die strukturierte Eingabe die bevorzugte Methode zur Erstellung von Befundberichten ist.

Die Analyse zeigt auch, dass das bisher redaktionell vorgegebene Wissen oft nicht den Vorstellungen der Anwender entspricht und 91 Prozent der Klinikärzte bereit wären sich an der Wissensentwicklung zu beteiligen. Es würden 95 Prozent der Befragten dieses Wissen mit Kollegen teilen. Als Anreiz wird hierfür in 92 Prozent der Fälle der Austausch mit fachnahen Kollegen genannt. Eine monetäre Gegenleistung wäre nur für 29 Prozent der Befragten eine Motivation.

#### V. ZIELSETZUNG UND FORSCHUNGSSTAND

Die Literaturrecherche und die quantitative Befragung haben gezeigt, dass Forschungsbedarf nach flexibleren Systemen, mit einer individuelleren Wissensbasis besteht. Daraus abgeleitet soll die leitende Forschungsfrage beantwortet werden:

*„Wie ist eine Beschreibungssprache zur Modellierung medizinische Befundvorlagen zu konzipieren, um die effiziente Berichterstellung und den offenen Austausch der entwickelten Vorlagen zu ermöglichen?“*

Als Qualitätskriterien wurden die Flexibilität, Individualität, Aktualität der Wissensbasis und die maschinelle Verarbeitung der Befundberichte für Automatisierungen und Informationsrückgewinnung identifiziert. In Interviews und Workshops werden die genauen Anforderungen an die medizinische Berichterstellung erfasst. Im Fokus der Analyse stehen Klinikärzte, medizinische Redaktionen und Hersteller klinischer Informationssysteme.



Die Beantwortung der Forschungsfrage wird in vier Teilbereiche unterteilt:

- Anforderungen an die Beschreibung medizinischer Befundvorlagen mit semantischen Metadaten.
- Ein semantisches Basismodell für die Interoperabilität unabhängiger Systeme.
- Entwurfsmuster für die Integration in klinische Informationssysteme.
- Versionierung, Autorenverwaltung, Individualisierung und Verteilung der entwickelten Vorlagen.

Es soll eine Beschreibungssprache entwickelt werden, die den Prozess der Wissenserstellung für medizinische Redaktionen vereinfacht und dem Anwender ermöglicht, Befundvorlagen direkt mitzugestalten, um bedarfsgerechte Lösungen zu entwickeln und zu teilen.

### A. Report Template Modeling Language (RTML)

In diesem Abschnitt wird die aktuelle Version der im Rahmen dieser Arbeit entwickelten Report Template Modeling Language vorgestellt. Die folgende Darstellung zeigt, wie RTML eingesetzt wird, um eine medizinische Dokumentation zu erstellen.

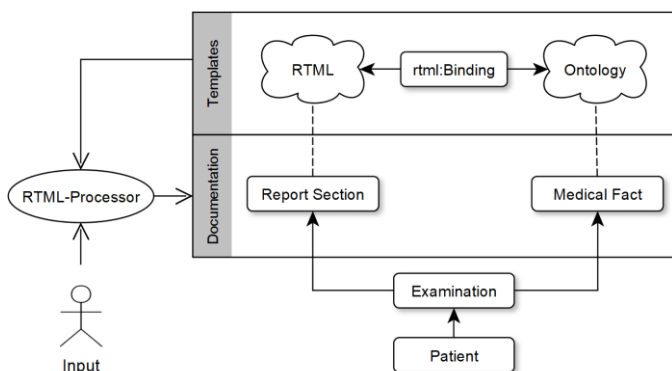


Abb. 2: Generierung der Dokumentation mit RTML

Für die Nutzung der entwickelten Vorlagen wird ein RTML-Processor benötigt, der aus Eingaben bzw. Entscheidungen die Fließtexte für den Befundbericht und ein semantisches Modell für die maschinelle Verarbeitung generiert und konsistent hält. Die Konsistenz wird dabei über eine im Modellierungsprozess konfigurierbare Bindung zwischen dem linguistischen und semantischen Modell ermöglicht. Fließtexte werden in Bereiche gruppiert, während Beobachtungen unabhängig davon direkt mit der Untersuchung verknüpft sind.

Die Anreicherung von Fließtexten mit einer formalen Semantik ist Forschungsgegenstand des Semantic Webs [29]. Um von Technologien und Erkenntnissen aus diesem Forschungsfeld zu profitieren, wird die hier entwickelte Modellierungssprache auf Basis des Resource Description Frameworks (RDF) entwickelt und kann dadurch mit Technologien auf gleicher Basis kombiniert werden. So ist beispielsweise die Verbindung zwischen Fließtexten und

semantischen Konzepten, die mit der Web Ontology Language (OWL) beschrieben wurden, möglich [30, 31].

### B. Elemente der Modellierungssprache

In der folgenden Abbildung sind die Basiselemente der Beschreibungssprache dargestellt. Diese besteht aus linguistischen Elementen, die im Befundbericht lesbar sind (*rtml:Text*) und Struktur-Elementen, die mit einer RDF-basierten Ontologie verknüpft werden können (*rtml:Element*), den Textfluss steuern (*rtml:FlowElement*) und aus variablen Bereichen, über die ein Entscheidungsbaum abgebildet wird (*rtml:Slot*).

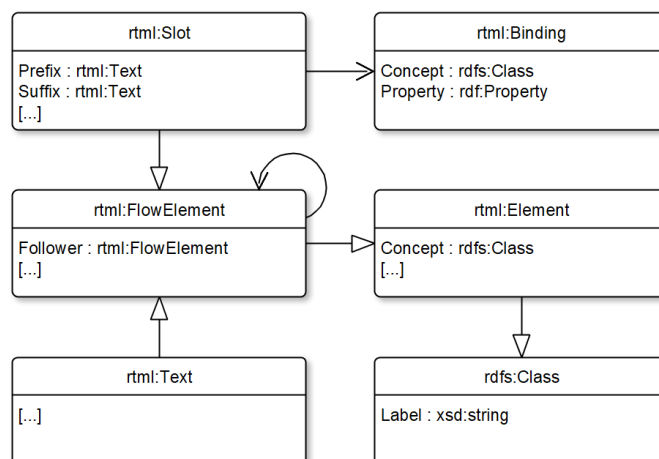


Abb. 3: Basiselemente in RTML

Ein Slot kann über Binding mit einem semantischen Modell verknüpft werden. Es gibt verschiedene Datentypen, die abgebildet werden müssen und entsprechend gibt es verschiedene Slot-Typen, die diese Anforderungen repräsentieren. Die folgende Tabelle gibt einen Überblick der wichtigsten Slot-Typen.

TABLE I. WICHTIGSTE SLOT-TYPEN

Slot	Beschreibung
rtml:Selector	Gruppierung von Elementen zu einer Auswahlliste
rtml:FlowGroup	Entscheidungsbedingte, linguistische Verbindungen
rtml:Link	Verknüpfung zu wiederverwendbaren Bausteinen
rtml:Range	Zahlenbereich, insbesondere für Messwerte relevant
rtml:UserInput	Platzhalter für Freitext
rtml:DateTime	Auswahl eines Datums und/oder einer Zeit

Slots haben allgemeine und spezielle linguistische Eigenschaften, die von einem RTML-Processor zu einem Fließtext verbunden werden. So hat jeder Slot beispielsweise einen Prefix und Suffix, der nur eingefügt wird, wenn der Slot einen Wert enthält, während es bei Auswahllisten beispielsweise die Möglichkeit gibt verschiedene Elemente auszuwählen und diese über Separatoren zu einem flüssigen Text zu verbinden.

Die Beschreibung von Beobachtungen, Pathologien, Therapien und Komplikationen und deren Verkettung kann sehr komplex werden und häufig ist hier enthaltenes Wissen wiederverwendbar. Um Redundanz im Graphen zu vermeiden wurden zwei Mechanismen eingeführt. Zum einen kann ein Element als Vertreter eines semantischen Konzepts deklariert werden und zum anderen lassen sich komplette linguistische Verbindungen in Scopes wiederverwenden. Scopes werden gruppiert (*rtml:ScopeGroup*) und können über Link-Elemente miteinander verknüpft werden.

Ein Scope enthält Text-Elemente für den Befundbericht und Slots für die Abbildung von Entscheidungen. Ein Slot kann dazu verwendet werden die Art einer Beobachtung zu definieren. Für eine Beobachtung können Eigenschaften hinterlegt sein, die sich dann wieder mit einer Slot-Text-Sequenz beschreiben lassen. So lässt sich der Detailierungsgrad immer weiter erhöhen und ein beliebig umfangreicher Entscheidungsbaum mit linguistischen Fragmenten abbilden.

Im Rahmen dieser Arbeit wird für die Modellierungssprache auch das genaue Verhalten des RTML-Prozessors und die Zusammenarbeit mit anderen Technologien des Semantic Webs spezifiziert [31–33].

### C. RTML und formale Semantik

Der RTML-Prozessor erzeugt auf Basis von Eingaben zum einen den Fließtext für den Befundbericht und zum anderen ein semantisches Abbild der dort enthaltenen Informationen für die maschinelle Verarbeitung. Hierfür sind bereits die Vorlagen mit Metainformationen versehen. Ein RTML-Element kann mit einer RDF basierten Klasse verknüpft werden. Dieselbe Instanz kann mit einer beliebigen Anzahl von RTML-Elementen verschiedener Scopes verbunden sein. So ist es beispielsweise möglich eine ausführliche Beschreibung einer Pathologie und gleichzeitig eine Zusammenfassung dieser Beschreibung mit derselben semantischen Entität zu verbinden und auch automatisch zu generieren.

Slots, die mit einem semantischen Modell verbunden sind werden über den RTML-Prozessor konsistent gehalten. Das semantische Mapping des Slots ist abhängig von den Datentypen des linguistischen und des semantischen Modells. Neben der formalen Abbildung ist auch die Beschreibung von implizitem Wissen möglich. So könnte zu einem Messwert die Einheit oder auch die physikalische Größe hinterlegt werden. Das semantische Abbild ermöglicht in Kombination mit einem Regelwerk die automatische Kodierung (z.B. ICD-10) des Befundberichts. Das Modell ist auch eine wichtige Grundlage für die Automatisierung im Dokumentationsprozess, die domänenspezifische Repräsentation und Eingabe, die semantische Suche und die Auswertung der Informationen für den Anwender.

### D. Notation der Modellierungssprache

Bisher werden Vorlagen für medizinische Berichtsschreibung meist von medizinischen Redaktionen entwickelt und textuell oder tabellarisch beschrieben. Während einfache Modelle so dargestellt werden können, wird für komplexe Beschreibungen mit semantischen Metadaten eine effizientere Notation benötigt,

um alle Zusammenhänge erfassen zu können. Die folgende grafische Notation soll die Komplexität der Wissensmodellierung für den Anwender reduzieren. In diesem Abschnitt sollen die grundsätzlichen Entwurfsprinzipien dieser Notation vorgestellt werden.

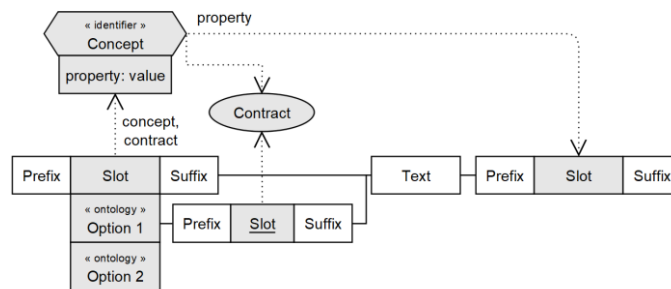


Abb. 4: Kurzfassung der RTML-Notation

Um die Wechselwirkung zwischen Elementen darzustellen werden Scopes visuell getrennt. Innerhalb eines Scopes lassen sich drei Bereiche unterscheiden:

- Linguistische Fragmente werden von Strukturelementen farblich unterschieden. Grundsätzlich werden beide Elemente im Textfluss ausgerichtet. Diese Eigenschaft ermöglicht eine Vorschau der Satzbildung.
- Die Darstellung der Slots hängt von den dort repräsentierten Datentypen ab, erfolgt aber für alle Slots kompakt. So wird beispielsweise bei Prefix und Suffix oder bei der Darstellung einer Aufzählung auf Verbindungslinien verzichtet.
- Semantische Metadaten sind über den linguistischen und strukturellen Elementen angeordnet und direkt mit diesen verknüpft.

Verbindungslinien werden zur Abgrenzung unabhängiger Elemente, zur Darstellung der Satzstruktur und zur Verknüpfung der Metadaten verwendet. Die Interviews mit Experten der Wissensmodellierung haben gezeigt, dass die grafische Notation gegenüber einer textuellen Beschreibung von einer deutlichen Mehrheit bevorzugt wird und die Erschließung aller Zusammenhänge der Vorlagenentwicklung für viele nur dadurch ermöglicht wird. Für die Modellierung der Ontologien gibt es eine breite Auswahl an Werkzeugen. Im Rahmen dieser Arbeit wird für die Demonstration und Evaluation die RTML-Notation mit einer visuellen Notation zur Entwicklung von OWL-Ontologien kombiniert [34–36].

## VI. INTEGRATION UND INTEROPERABILITÄT

Für die herstellerunabhängige Integration und den unabhängigen Austausch der Befundvorlagen ist neben der technischen Interoperabilität auch ein semantisches Modell erforderlich. Das im Rahmen dieser Arbeit entwickelte Basismodell ist ein erweiterbares Bezugssystem für klinische Informationssysteme und soll die Beziehung zwischen typischen Entitäten normieren. Es werden Entwurfsmuster vorgestellt, die dabei helfen die Modellierungssprache in ein Informationssystem zu integrieren und Lösungsvorschläge erarbeitet, durch die eine maschinelle Weiterverarbeitung, Automatisierung und Abfrage der Eingaben ermöglicht wird.

Häufig werden relationale Datenbanksysteme bei der Entwicklung klinischer Informationssysteme eingesetzt [37] und zur Gewährleistung der Informationssicherheit der Patientendaten auch von klinischen IT-Abteilungen gefordert. Um die klassischen Informationssysteme mit Technologien des Semantic Webs zu verbinden, werden in dieser Arbeit einige Lösungsvorschläge dargestellt und im Rahmen der Prototypenentwicklung demonstriert. Es wird gezeigt, wie mithilfe einer Mapping-Sprache eine Semantic-View auf die relationalen Daten (wie beispielsweise Stammdaten, Falldaten, Messwerte) des konventionellen Informationssystems generiert werden kann [38] und wie diese View mit einem SQL basierten Triplestore [39] für patientenbezogene Daten und einer effizienteren NoSQL-Implementierung [40] für die teilweise sehr komplexen, medizinischen Befundvorlagen kombinierbar ist.

Für die Entwicklung klinischer Informationssysteme mit relationalen Datenbanken werden häufig objektrationale Abbildungen genutzt, um innerhalb der objektorientierten Programmierung mit virtuellen Objekten zu arbeiten. Im Bereich des Semantic Webs werden ähnliche Technologien eingesetzt [41–44]. Im Rahmen dieser Arbeit wird ein Semantic Layer entwickelt, der die objektorientierte Abbildung der RDF-Statements zur Laufzeit ermöglicht. Der Semantic Layer wird als Basis für die prototypische Umsetzung des RTML-Prozessors genutzt.

Alle Daten werden über einen zentralen Hub (Publish-subscribe) mit semantischer Filterung verteilt. Durch dieses Muster erhalten die angeschlossenen Ziel-Komponenten vorgefilterte Entitäten, die kontextbezogen verarbeitet werden können. Der Semantic Layer ermöglicht hierfür die kontextspezifische Erweiterung der semantischen Entitäten, um beispielsweise bedarfsgerechte Repräsentationen zu realisieren. Zielsetzung der Entwurfsmuster sind Handlungsempfehlungen für die effiziente Integration der Modellierungssprache in klinische Informationssysteme.

## VII. KOLLABORATION UND VERTEILUNG

Die auf Basis der Beschreibungssprache realisierten Systeme sollen den offenen Austausch und die Mitarbeit vieler Experten bei der Entwicklung der Befundvorlagen unterstützen [45]. Entsprechend müssen administrative Metadaten hinterlegt und ein System zur Versionierung und zum technischen Austausch der Vorlagen bereitgestellt werden. Vorlagen werden hierfür in einer Arbeitsgruppe zusammengefasst, die von verschiedenen Autoren verwaltet werden kann. Es werden zusätzliche Beschreibungsdaten, wie der Kontext und das Ziel der enthaltenen Vorlagen, die Liste der beteiligten Autoren oder auch rechtliche Hinweise für die Nutzung der Vorlagen verknüpft.

Im Rahmen dieser Arbeit wird die Versionsverwaltung vom Datengraphen entkoppelt, um eine effiziente Implementierung des RTML-Prozessors und den technologieunabhängigen Austausch der Vorlagen zu ermöglichen. Änderungen werden mit Zeitstempel, Kommentar und Autor protokolliert [46]. Zum einen dient die Versionierung damit der Transparenz im Prozess der Wissensentwicklung, zum anderen aber auch um Individualisierungen der Anwender mit den Neuerungen der

Autoren einer Arbeitsgruppe zu verbinden. Um Derivate zu vermeiden, können Elemente individuell konfiguriert werden. So lassen sich beispielsweise Merkmale auf die benötigte Menge reduzieren ohne dabei die eigentliche Struktur zu verändern. Für den technischen Austausch der Versionen wird ein Format entwickelt, das die Anforderungen an eine dezentrale Verwaltung, Kompression und Transparenz erfüllt.

## VIII. DEMONSTRATION UND EVALUATION

Ziel der Demonstration ist es die praktische Umsetzung der Modellierungssprache, die Integrationsfähigkeit in ein Informationssystem und die Möglichkeiten zur kollaborativen Entwicklung zu evaluieren. Für diese Phase wird zum einen ein Wissenseditor für Autoren entwickelt, um RTML-Vorlagen und Ontologien mit computergestützten Werkzeugen zu erstellen. Zum anderen wird ein RTML-Prozessor und darauf aufbauend ein Editor für die Berichterstellung entwickelt.

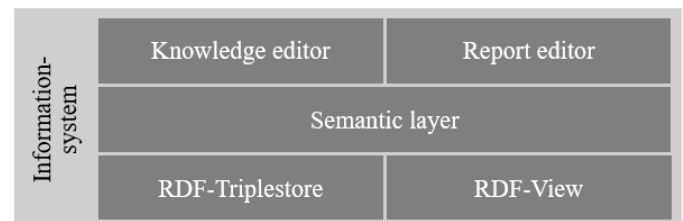


Abb. 5: Prototypen der Demonstration

In der ersten Phase der Evaluation werden Interviews und Workshops mit Autoren und Ärzten durchgeführt. Die Ergebnisse der Evaluation fließen wieder zurück in die Konzeption, um das Modell den neuen Erkenntnissen anzupassen. Wurden alle Anforderungen der ersten Phase in das Modell überführt, dann werden in der zweiten Phase, dem Feldexperiment, die Prototypen in einigen Kliniken eingesetzt. Hier wird die praktische Leistungsfähigkeit der Modellierungssprache evaluiert. So ist zum Ende der Arbeit eine umfangreiche Abdeckung der allgemeinen Anforderungen an die Wissensentwicklung im Kontext der Befundberichterstellung zu erwarten.

## IX. AUSBLICK

In der aktuellen Phase werden die Modellierungssprache und die Anforderungen an die kollaborative Entwicklung von Befundvorlagen mit Experten analysiert. Die entwickelten Prototypen ersetzen die in früheren Interviews und Workshops eingesetzten Mockups. Im anschließenden Feldversuch wird das Artefakt auch unter realen Bedingungen evaluiert. Bisher wurden Anforderungen primär aus dem Fachbereich der Gastroenterologie abgedeckt. In einem weiteren Schritt soll das Modell in weiteren Fachbereichen evaluiert werden.

Die entwickelte Beschreibungssprache bietet weiteres Forschungspotential, zum Beispiel in der automatischen Generierung von Vorschlägen für medizinische Befundvorlagen in Form von RTML auf Basis vorhandener Befundberichte oder in der Entwicklung einer klinikübergreifenden Analyseplattform für die medizinische Forschung.

## X. REFERENCES

- [1] V. Meinhardt and E. Schulz, "Kostenexplosion im Gesundheitswesen?," (ger), *DIW-Wochenbericht : Wirtschaft, Politik, Wissenschaft*, vol. 70, no. 7, pp. 105–109, 2003.
- [2] H. Püschmann, G. Haferkamp, K.-D. Scheppokat, H. Vinz, and M. Wegner, "Vollständigkeit und Qualität der ärztlichen Dokumentation in Krankenakten - Untersuchung zu Krankenunterlagen aus Chirurgie, Orthopädie, Innerer Medizin und Neurologie," *Deutsches Ärzteblatt*, vol. 103, no. 3, pp. 121–126, 2006.
- [3] F. Leiner *et al.*, *Medizinische Dokumentation. Grundlagen einer qualitätsgesicherten integrierten Krankenversorgung*, 6th ed. Stuttgart: Schattauer, 2011.
- [4] K. Blum and U. Müller, "Dokumentationsaufwand im Ärztlichen Dienst - Repräsentativerhebung des Deutschen Krankenhausinstituts," *Das Krankenhaus*, pp. 544–548, 2003.
- [5] HIMSS, *Auf den Spuren der Zeitdiebe im Krankenhaus: Die wahre Belastung durch Dokumentation an deutschen Akutkrankenhäusern wird unterschätzt*.
- [6] J. J. Mamlin and D. H. Baker, "Combined Time-Motion and Work Sampling Study in a General Medicine Clinic," *Medical Care*, vol. 11, no. 5, pp. 449–456, 1973.
- [7] W. A. Nowlan *et al.*, "PEN&PAD: A Doctors' Workstation with Intelligent Data Entry and Summaries," *Symposium on Computer Applications in Medical Care. American Medical Informatics Association*, pp. 941–942, 1990.
- [8] E. Ammenwerth and H.-P. Spötl, "The Time Needed for Clinical Documentation versus Direct Patient Care - A Work-sampling Analysis of Physicians' Activities," *Methods of Information in Medicine*, vol. 48, no. 1, pp. 84–91, 2009.
- [9] A. P. McAfee, "Enterprise 2.0: The Dawn of Emergent Collaboration," *MIT Sloan Management*, vol. 47, no. 3, 2006.
- [10] P. Haas, *Medizinische Informationssysteme und Elektronische Krankenakten*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, 2005.
- [11] V. E. Amelung, Ed., *Vernetzung im Gesundheitswesen: Wettbewerb und Kooperation*. Stuttgart: Kohlhammer, 2009.
- [12] A. W. Goldschmidt and B. Busse, Eds., *Praxishandbuch Medizincontrolling*. Heidelberg: Economica, 2005.
- [13] K. O. Bergmann and C. Wever, *Die Arzthaftung: Ein Leitfaden für Ärzte und Juristen*, 4th ed. Berlin: Springer, 2014.
- [14] T. M. Lehmann, Ed., *Handbuch der medizinischen Informatik*, 2nd ed. München, Wien: Carl Hanser Verlag, 2004.
- [15] M. Ivanović and Z. Budimac, "An overview of ontologies and data resources in medical domains," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5158–5166, <http://www.sciencedirect.com/science/article/pii/S0957417414001213>, 2014.
- [16] J. Ingenerf, "Terminologien oder Klassifikationen," *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, vol. 50, no. 8, pp. 1070–1083, 2007.
- [17] A. Meining *et al.*, "Neufassung der Standard-Terminologie in der gastroenterologischen Endoskopie – Ergebnis eines Konsensusprojekts der Deutschen Gesellschaft für Gastroenterologie, Verdauungs- und Stoffwechselkrankheiten," (ger), *Zeitschrift für Gastroenterologie*, vol. 56, no. 1, e1, 2018.
- [18] R. H. von Alan, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS quarterly*, vol. 28, no. 1, pp. 75–105, 2004.
- [19] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, 2007.
- [20] T. Wilde and T. Hess, "Forschungsmethoden der Wirtschaftsinformatik: Eine empirische Untersuchung," *Wirtschaftsinformatik*, vol. 49, no. 4, pp. 280–287, 2007.
- [21] J. Vom Brocke, A. Simons, B. Niehaves, and K. Reimer, "Reconstructing the giant: On the importance of rigour in documenting the literature search process," *ECIS 2009 Proceedings*, no. Paper 161, 2009.
- [22] M. Meuser and U. Nagel, "Das Experteninterview — konzeptionelle Grundlagen und methodische Anlage," in *Methoden der vergleichenden Politik- und Sozialwissenschaft: Neue Entwicklungen und Anwendungen*, S. Pickel, G. Pickel, H.-J. Lauth, and D. Jahn, Eds., Wiesbaden: VS Verlag für Sozialwissenschaften, 2009, pp. 465–479.
- [23] W. Mayerhofer, "Das Fokusgruppeninterview," in *Qualitative Marktforschung: Konzepte — Methoden — Analysen*, R. Buber and H. H. Holzmüller, Eds., Wiesbaden: Gabler, 2007, pp. 477–490.
- [24] T. Bucher, C. Riege, and J. Saat, "Evaluation in der gestaltungsorientierten Wirtschaftsinformatik - Systematisierung nach Erkenntnisziel und Gestaltungsziel," in *Wissenschaftstheorie und gestaltungsorientierte Wirtschaftsinformatik*, 2008, pp. 69–86.
- [25] M. Siekmann, "E-Collaboration to establish a knowledge database for the structured documentation of findings," *13. Internationale Tagung Wirtschaftsinformatik (Doctoral Consortium)*, 2017.
- [26] S. T. Rosenbloom *et al.*, "Data from clinical notes: a perspective on the tension between structure and flexible documentation," *Journal of the American Medical Informatics Association*, vol. 18, no. 2, pp. 181–186, 2011.
- [27] M. Klarner, "Hybride, pragmatisch eingebettete Realisierung mittels Bottom-Up-Generierung in einem natürlichsprachlichen Dialogsystem," Dissertation, Universität Erlangen-Nürnberg, Erlangen, 2005.
- [28] G. Görz, C.-R. Rollinger, and J. Schneeberger, "Generierung natürlichsprachlicher Texte," in *Handbuch*

- der künstlichen Intelligenz, G. Görtz, C.-R. Rollinger, and J. Schneeberger, Eds., 3rd ed.: Oldenbourg, 2000.
- [29] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Sci Am*, vol. 284, no. 5, pp. 34–43, 2001.
- [30] O. Lassila and R. R. Swick, *Resource Description Framework (RDF) Model and Syntax Specification*. [Online] Available: <http://www.w3.org/TR/PR-rdf-syntax>. Accessed on: Sep. 23 2018.
- [31] M. Dean and G. schreiber, *OWL Web Ontology Language Reference*. [Online] Available: <https://www.w3.org/TR/owl-ref/>. Accessed on: Sep. 23 2018.
- [32] H. Knublauch and D. Kontokostas, *Shapes Constraint Language (SHACL)*. [Online] Available: <https://www.w3.org/TR/shacl/>. Accessed on: Sep. 23 2018.
- [33] E. Prud'hommeaux and A. Seaborne, *SPARQL Query Language for RDF*. [Online] Available: <https://www.w3.org/TR/rdf-sparql-query/>. Accessed on: Sep. 23 2018.
- [34] J. Bārzdiņš, G. Barzdins, K. Cerans, R. Liepins, and A. Sproģis, "OWLGrEd: a UML Style Graphical Notation and Editor for OWL 2," *International Conference on Business Informatics Research*.
- [35] R. Falco, A. Gangemi, S. Peroni, D. Shotton, and F. Vitali, "Modelling OWL Ontologies with Graffoo," *European Semantic Web Conference*.
- [36] OMG, *Ontology Definition Metamodel (ODM)*. [Online] Available: <https://www.omg.org/spec/ODM>. Accessed on: Sep. 23 2018.
- [37] M. Bageshwari, P. Adurkar, and A. Chandrakar, "Clinical Database: RDBMS V/S Newer Technologies (NoSQL And Xml Database); Why Look Beyond RDBMS and Consider the Newer," *International Journal of Computer Engineering and Technology (IJCET)*.
- [38] M. Arenas, A. Bertails, E. Prud'hommeaux, and J. Sequeda, *A Direct Mapping of Relational Data to RDF*. [Online] Available: <https://www.w3.org/TR/rdb-direct-mapping/>. Accessed on: Sep. 23 2018.
- [39] R. Meersman et al., Eds., *An Evaluation of Triple-Store Technologies for Large Data Stores: On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops*: Springer Berlin Heidelberg, 2007.
- [40] C. Weiss, P. Karras, and A. Bernstein, "Hexastore: Sextuple Indexing for Semantic Web Data Management," *Proc. VLDB Endow.*, vol. 1, no. 1, pp. 1008–1019, <http://dx.doi.org/10.14778/1453856.1453965>, 2008.
- [41] T. J. Armstrong, "Enhancing Object-Oriented Technologies with Semantic Web Technologies," in *Proceedings of Semantic Web Information Management on Semantic Web Information Management*, Snowbird, UT, USA: ACM, 2014, pp. 1–8.
- [42] K. Wenzel, "KOMMA: An Application Framework for Ontology-based Software Systems," *Semantic Web – Interoperability, Usability, Applicability*, 2010.
- [43] M. Völkel and Y. Sure, Eds., *RDFReactor – From Ontologies to Programmatic Data Access*, 2005.
- [44] P. Kremen, *Building Ontology-Based Information Systems*. Prague, 2012.
- [45] J. Surowiecki, *The wisdom of crowds*: Anchor, 2005.
- [46] A. Natanael, P. Naumann, and E. Marx, "Exploring the Evolution and Provenance of Git Versioned RDF Data," *14th European Semantic Web Conference (ESWC 2017)*, vol. 2017.

# Development of a Decision Support System for Online Surgery Scheduling

Norman Spangenberg  
Information Systems Institute  
Leipzig University  
Leipzig, Germany  
spangenberg@wifa.uni-leipzig.de

**Abstract**—This work has its focus on decision support in operational business situations and especially on the very short-term decisions in Online Surgery Scheduling. This so far barely considered problem needs research for decision support, since few approaches are available that relieve the OR manager through tool support and reduce the informational, communicational and cognitive workloads needed to ensure efficient and frictionless operations. With the strong growth of generated data and the digitization of business processes that make previously unobtrusive business elements become more visible, and their combination with large-scale data processing technologies and intelligent methods of the fields of AI or Analytics, new opportunities for data-driven real-time Decision Support Systems become evident.

The objective of this research is the development of an approach that supports the operational decision processes in Operating Room Management and Online Surgery Scheduling by facilitating the information collection and reducing the cognitive effort for decision-making by providing predictive information or alternative actions. Consequently, a novel approach for a data-driven Decision Support System for Online Surgery Scheduling is presented.

**Keywords**—Decision Support Systems; Online Surgery Scheduling; Surgical Phase Recognition; Machine Learning;

## I. INTRODUCTION

Decisions and decision-making processes on the operational level of organizations are characterized by their short-term planning horizons, unexpected events, immediate decisions or frequent disturbances, and are found in various domains like maintenance, manufacturing and production management, logistics coordination and emergency handling [8, 2, 20, 23, 17]. Another highly relevant use case for short-term decisions is the operational Operating Room Management (ORM) which has the goal of operating efficient and structured processes in the operating room (OR) area at minimal costs. The latter aspect of costs and efficiency means a key problem in the daily operations of the OR area, since it is one of the most expensive hospital facilities and thus efficient business operations management is imperative.

As a part of ORM, Online Surgery Scheduling (OSS) delineates all intra-day decisions needed to ensure the execution of all planned and unplanned surgeries of the surgery schedule, with all of the concomitant uncertainties like delays, cancellations and emergency patients. It has a very short-term planning horizon of just a few hours and is faced with unexpected cancellations or add-on cases, uncertain surgery

durations and other dynamics in the environment making it hard for the OR manager to gather and process all necessary information to guarantee orderly operations. The OR manager is responsible for operational planning in the OR area, in particular for the supervision of all surgery-related resources and for complying with a created surgery schedule according to diverse performance indicators. So, for OR managers there is an ongoing pressure to optimize operational processes, e.g. increase resource utilization or reduce OR downtime. These so far barely considered problems reveal new research opportunities for decision support, since few approaches are available that relieve the OR manager through tool support and reduce the informational, communicational and cognitive workloads needed to ensure efficient and frictionless operations.

Due to the strong growth of generated data and the digitization of business processes due to connected or smart devices and the Internet of Things, plenty of business activities and previously unobtrusive business elements become more visible. Hence, the business-critical tasks of utilizing data need goal-oriented and appropriate practices to achieve the best results for supporting decision makers. But to avoid data and information overload, it is necessary to extract and enrich valuable information deduced from vast amounts of heterogeneous data and provide it to the strategic, tactical and operational management level. This can be achieved for example by the use of large-scale data processing technologies – which are prerequisites for the utilization of this data – in combination with intelligent methods of the fields of AI, Analytics and Data Science to enrich information. Hence, novel opportunities for Intelligent Systems and IS research in operational decision support become evident. Consequently, data-driven approaches for real-time Decision Support Systems (DSS) in operational decision processes form the basis of this work.

In this work, the analytics components of a DSS for OSS is presented that has the goal to utilize the data and turn it into information value to assist decision-making activities by detecting and handling relevant business events.

The remainder of this work is structured as follows. The upcoming section describes the research questions and the used methodology for developing the artefacts. In section 3, the OSS problem is described in detail while section 4 provides the solutions for these problems. Finally, the results of the solution approaches' evaluation is presented.

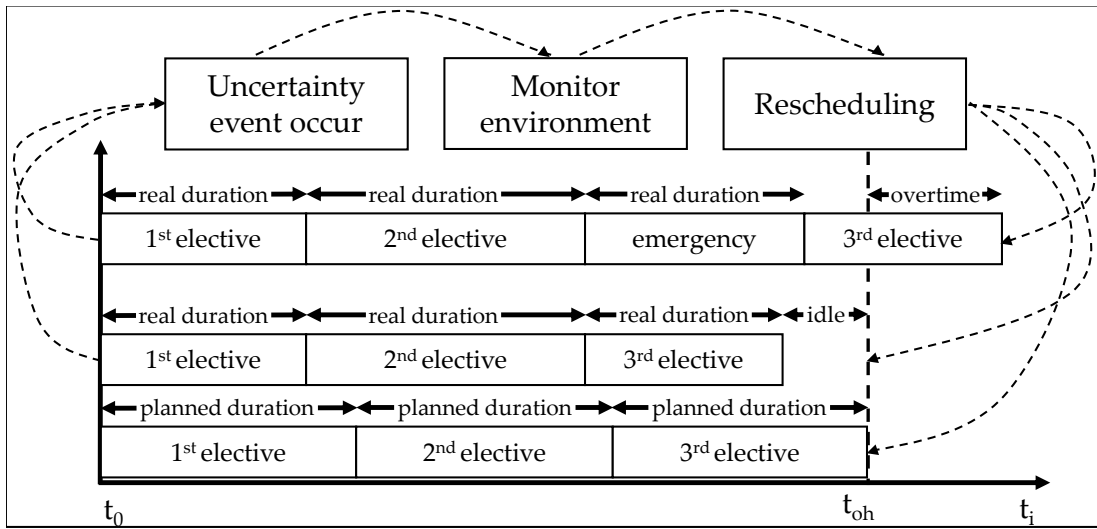


Fig. 1. OSS problem for a single OR including the challenges due to changing intervention durations and introducing emergent cases [36]

## II. RESEARCH QUESTIONS AND METHODOLOGY

Resulting from the recent findings, the hypothesis of this research is based on the assumption that the provision of adequate tools, that support decision-makers simplifies and optimizes the corresponding decision processes in OSS. As a consequence, the intention of this work is to develop an approach for a DSS that utilizes available data of varying sources for the processing, integration and enhancement of information, in order to create additional information values for decision makers in operational ORM respectively OSS. Based on intelligent methods from DSS and Business Analytics this approach should provide close to real-time information of the OR area and for optimized surgery rescheduling. Hence, the research objective is the development of an approach for supporting operational decision processes in Operating Room Management, that facilitates the information gathering and processing, as well as reduces the cognitive efforts for decision-making. Consequently, the derived overriding research question refers to this objective: *How can operational decision processes in Operating Room Management be provided with relevant and up-to-date information for decision-making?* It is detailed in three research subquestions:

- **SQ1:** Which solution approaches already exist for the support of operational decision processes?
- **SQ2:** Which requirements impose on an approach for supporting operational decision processes?
- **SQ3:** How should a Decision Support System be designed to support operational decision processes in Online Surgery Scheduling?

To ensure the systematic, structured and replicable treatment of the research questions the research methodology of design-oriented IS research described by Österle, Winter, and Brenner [29] is utilized.

- 1) **Analysis:** During the analysis phase, the research problem was identified and formulated, also a catalogue of

requirements for a corresponding DSS was specified. Systematic literature reviews, cross-sectional studies and the analysis of expert interviews were conducted to substantiate the relevance of the formulated problem as well as to enhance the requirements analysis with data from a use case perspective.

- 2) **Design:** The conception of the particular artefacts happened in the design phase. Design-oriented research methods were engaged for the development of the necessary elements to answer the achieving research goals. Modelling methods are used for the mathematical representation of aspects of the reality, while inductive approaches like reference modelling could be used for the abstract description of an optimized reality [41]. Closely linked is the prototyping, which is utilized to enable a *proof-of-concept* of the models as well as the DSS approach.
- 3) **Evaluation:** For the verification of rigour, the developed artefacts are subject of a proper evaluation. This was ensured by using methods for the benchmarking of certain artefacts as well as implementing the prototype, to show their effectiveness and efficiency for the fulfilment of specific requirements or aspects of the research goal [29].
- 4) **Diffusion:** The diffusion – meaning the dissemination of the research results in the space of the research community – constitutes the final step of the research cycle and was achieved by the presentation of research papers at several international conferences. The selection of the publications and conferences was conducted in respect to the diverging poles of the technical perspective of IS research and the use case-driven perspective in the IS research community as well as in the healthcare research community.

In addition, during the entire research process, the principles of the design-oriented IS research approach abstraction, orig-

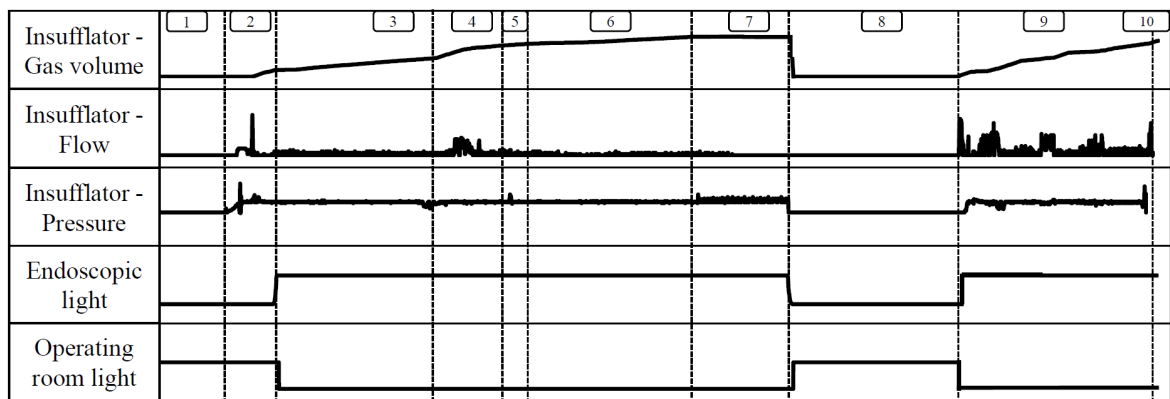


Fig. 2. Parameter values of monitored surgical devices

inality, justification and benefit are considered.

### III. PROBLEM FORMULATION

Similar to other planning and control processes, OR management is hierarchically separated into a strategic, tactical and operational level, concerned with different planning horizons. Strategic OR management deals with sizing core resources like staff and room, over a planning horizon of a year or more. Tactical OR management has a midterm planning horizon for the rough planning of a Master Surgery Schedule. Finally, the operational level of OR management is divided into offline and online OR management, considering the substantiation of tactical OR plans for a one-week planning horizon before schedule execution (offline) and the control and rescheduling of an intra-day planning horizon during schedule execution (online). The main task of operational OR management is Online Surgery Scheduling (OSS), ensuring the execution of the surgery schedule, completing all planned and unplanned surgeries this day [6]. From now on, all unexpected changes and events require immediate decisions, making OSS the most challenging and time-sensitive task in OR management. Accurate surgery scheduling is a necessity for efficient and frictionless operations in the OR area. Is this not the case, ambiguities and delays occur in the operations that further lead to interruptions, unused resources and dissatisfied staff or patients. Several subtasks can be deduced from these characteristics, which all are accompanied with challenges for gathering necessary information.

#### A. Monitoring

This implies the collection of necessary information to be aware of the current situation or unforeseen events in the OR area [15]. This information might come from computerized information systems, like the hospital information system or by the interaction with surgical staff. However, there is still much communicational efforts with surgeons needed to collect timely information and real-time updates of current status and surgical phases, delays or critical events [27]. In addition, several uncertainties and frequent changes characterize the monitoring of the OR area, that arise from varying surgery

durations and resource availability as well as stochastically occurring patient no-shows or arrivals of emergency patients [31]. Therefore, OR monitoring often is associated with a lack of knowledge that has to be completed with the experience of the OR manager, that make delays and cancellations often seem intrinsic for processes in the OR area [16].

#### B. Estimating remaining surgery durations:

Based on the observation of the current situation and conditions, the prediction of upcoming events and outcomes of surgery-related processes is another important task in OSS. The most significant process parameter for predicting future outcomes is the remaining time to the end of the surgery, meaning the time elapsed between the moment the patient is introduced to the OR and leaves the OR [32].

Reliable predictions of surgery durations are needed to make optimal decisions in OR scheduling, e.g. for calling the next elective patient to minimize patient waiting times or reduce OR idle times, if a surgery was shorter than predicted. Further, when surgeries are longer than predicted, subsequent ones need to be postponed or canceled [7]. Both cases lead to inefficiencies in the use of the ORs, recurring communications between staff, patients and OR managers, and an overload of the preoperative holding area [14]. Remaining intervention times are difficult to predict due to several factors. Often the scope of an intervention is not known in advance and becomes apparent not before the surgery is already in progress [33]. For running cases, the current status and recent events during the intervention are influencing parameters. Since current conditions, recent progress and states are not easily recognizable, OR managers need substantial experience knowledge and external information to make reliable predictions. Hence, personnel requests and inspections in the OR are necessary to complete the comprehension of the current situation in the OR area as well as obtain real-time information, e.g. for contemporaneous assignment and scheduling of emergent procedures. The absence of objective and reliable predictions adds complexity to OSS, since the OR manager is forced to interpret various situations to adapt the surgery schedule [14].



### C. Rescheduling of surgeries:

Initial surgery schedules created in the offline planning phase are often obsolete right after the beginning of the surgery day [27], due to unexpected events like delays, cancellations or urgent add-on cases require the modification of the current surgery schedule. The rescheduling of a surgery is challenging in terms of necessary experience knowledge and cognitive efforts for weighing planning alternatives, due to the uncertainties and dynamics in OR operations. The possible actions for OSS include the assignment and scheduling of emergency procedures, the cancellation of cases, as well as the reassignment of ORs and start times for already planned interventions [27]. Then, a reevaluation and rescheduling of the current surgery schedule is performed and, depending on its value of improvement, the plan is modified or remains unchanged. The main objectives and measurements for determining the value of a plan are for example resource overtime, hospitalization costs, intervention costs, OR utilization, waiting times, and patient or personnel preferences [31]. Compared to other scheduling problems in manufacturing or production environments, surgical interventions cannot be suspended and later continued at the same point. Thus, the OSS problem is treated as a predictive-reactive scheduling problem where upcoming events have to be predicted and unforeseen events have reacted to. This means, OSS is represented as a two-stage process, consisting of an initial predictive schedule that models the desired behavior of the OR area. Then, in the reactive part, this schedule is modified in response to certain events during execution [40, 1, 27].

## IV. SOLUTION APPROACH

According to the problem formulation, a predictive-reactive rescheduling strategy is utilized and supported with software tools to generate and partially update the current schedule based on incoming events with planning-relevant information. In this section an architectural approach is proposed that realizes several software components which are needed to collect and enhance the necessary information. The segmentation of the solution approach into three parts is because of separation of concerns. Nevertheless they build on top of another each subsystem uses different type of data and information.

### A. Situation Detection Subsystem

This component supports the information gathering tasks of the OR manager and automatizes it to ease and advance this process. Based on low-level real-time data of e.g. cameras, surgical devices, OR equipment or other connected devices information about intra-surgical phases in running interventions can be gathered (cf. figure 2). Besides processing the incoming data streams, SDS realizes methods for the phase detection. Lots of research exists for surgical phase detection methods. Some are image- or video-based, e.g. [5]. Others relying on electronic signals of surgical devices are described for example by [30]. All of these methods have their pros and cons, e.g. some detect minimal invasive surgical phases better than others and vice versa. In this work, Complex Event Processing (CEP)

is used for modelling surgical phases based on surgical device data and operating room equipment e.g. OR lights. According to the taxonomy proposed in [21] this component classes into the data-to-model analysis methods. The data streams are partitioned into time-windows and monitored for patterns, thresholds and average values of parameters. Detected surgical phases are shown inside the data graphs to visualize the temporal relations as well as they appear in a list to show up recent surgical phases. In addition, all ORs are presented on an aggregated level to reduce the amount of information and prevent data overload of the decision-maker CEP means the conceptual foundation for modelling, for implementing the resulting rules and patterns the Esper CEP engine [9] was used due to its strong expressiveness and good scalability.

### B. Prediction Subsystem

The projection of environmental future states is based on the understanding of a situation and needs lots of domain knowledge and experience. Only these abilities allow quick responses on events, as well as proactive decisions for deterministic behavior in certain, already known situations. In our approach these challenges are addressed by realizing a subsystem for remaining intervention time estimation. The PS utilizes the concept of the Lambda architecture (cf. fig. 3), which combines batch and stream data processing [25]. Applications using this concept benefit of the in-depth analytics possible with batch layer and low-latency of the streaming layer. The batch layer is used to process historic data of the recognized surgical phases and to train a regression model that computes a value for the remaining intervention time.

The underlying model is build with a Random Forest Regression algorithm based on five features provided by the SDS and additional data from resource databases like the Hospital Information System:

- **Event Name:** The name of the recognized current surgical phase. Some phases are unique to a specific surgery type, while others can be found among multiple surgeries. Each surgery begins with a phase called `START` and ends with a phase called `END`.
- **OPS Code:** The OPS code is the official German classification of medical procedures [12]. It holds different granularity of information about the given surgery and is used extract three features. E.g. the OPS Code of a sigmoidectomy is transformed to the OPS category 5-45, subcategory 5-455 and procedure (5-455.75).
- **Time stamp:** The time stamp gives information about the starting time of the current phase (or the end of the surgery in case of phase `END`). It is used to determine the features `TIME_PASSED`, the time since start of the surgery and `TIME_UNTIL_END` time until the end of surgery. The latter is the target variable that is aimed to predict from the other features.
- **Operating Room:** The room in which the surgery takes place. Master et al.[26] claim that the location of a surgery contains basic information about its complexity.

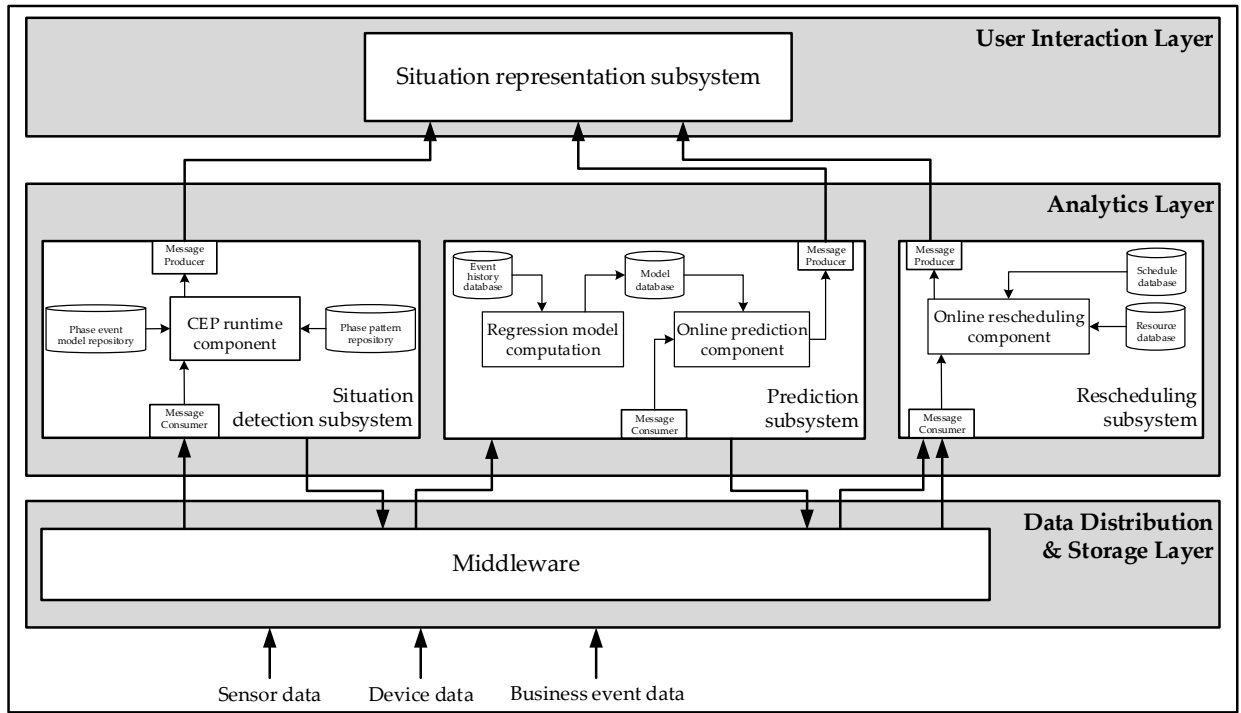


Fig. 3. Architecture model of the subsystems and relations.

Additionally different equipped OR rooms may have an influence on the procedure.

- **Phase History:** A vector containing information about recent surgical phases. Accurate predictions need the comprehension and involvement of the surgical workflow. Since phases in some cases appear more than once in a surgery the previous phase can give important indications for the current progress.

The current phase as well as an event history based on previously detected phases are factored into the model. As well, the OR and the OPS code have huge influence on the model. These are the input for the regression model to calculate an estimated value for the respective surgery based on the described features. The resulting models are stored in a model database and read by the online prediction component, which represents the speed layer of the Lambda architecture. The online prediction component again is interrelated with the SDS, meaning that a new recognized surgical phase triggers the recomputation of the remaining time. It loads the model and aligns detected phases in running surgeries with the model to update the estimated intervention time. Lastly, this starts triggering the upcoming rescheduling subsystem.

### C. Rescheduling Subsystem

In this work, it is assumed that a valid initial schedule exists and focus on the reactive task of OSS. This is based on a mixed integer linear programming (MILP) approach to model the OSS problem. Since it is a dynamic scheduling problem, it implies updating the schedule defined the previous day in reaction to external effects like incoming emergencies or

internal changes like deviations (see fig. 1). According to the rescheduling framework of [40], the OSS problem described in this work can be seen as a dynamic scheduling problem with variable arrivals of patients. The OSS problem consists of a number of characteristics, assumptions, resources and constraints. The model includes a set of indices for capacities and resource requirements of interventions, surgeons, ORs and their available time slots. Further, several parameters describe properties of the resources and entities like estimated durations, states about availability, urgency and modifiability. The model has two planning variables to optimize the schedule which are the assigned OR and the starting time of an intervention.

The number of surgeries to be scheduled on the tagged day is not known in advance, since it is likely that emergencies occur. A surgical intervention  $i$  is characterized by its surgeon  $s_i$ , the estimated duration  $d_i$  before or during the intervention and its urgency  $u_i$  according to the scale elective, urgent and emergent. Further, a modification parameter is introduced to block interventions in a specific OR at a specific time manually or after start. For each surgeon indexed  $s$ ,  $I_s$  denotes the set of jobs that are performed by that surgeon.

The objective is to find an assignment  $\sigma$  of interventions to available OR time slots and surgeons depending on the intervention duration, considering several optimization criteria and goals. The most important criteria for the OR manager – besides treatment quality – is maximizing OR utilization of each OR  $\omega_u(o_\sigma)$ , according to the method of Hans & Verbeekel [18], including the duration of all interventions per

day and the available OR working hours.

$$\text{Max } \omega_u(o_\sigma) = \sum_{i=1}^I \frac{d_i}{l} \quad (1)$$

The objective of minimizing waiting time  $\omega_w(\sigma)$  should lead to the fast scheduling of non-electives by introducing the urgency factor  $u$  and the penalty costs for the waiting time of an intervention  $c^w$ .

$$\text{Min } \omega_w(\sigma) = \sum_{i=1}^I u c^w \quad (2)$$

Further, all types of surgeries are assigned as early as possible, thus the solver minimizes overtimes  $\omega_o(\sigma)$ . Adding penalty costs  $\beta_u$  for each cancelled or reassigned intervention should lead to the effect that valid schedules with fewer reassignments/cancellations are preferred (3).

$$\text{Min } \omega_o(\sigma) = \sum_{i=1}^I \beta_u i_{canc} \quad (3)$$

A metaheuristic approach is used for solving the optimization problem of the rescheduling task. Metaheuristics don't guarantee finding an optimal solution for the optimization problem, but find an appropriate solution in a given amount of time, which is necessary for our goal to give real-time decision support. The search space is defined by two vectors: One for the OR assignments of each surgery and second a vector for non-overlapping time-slots including surgeons, surgical team and ORs. Hence, the planning variables are operating room and the combination of starting time slot and the intervention duration.

The cost function incorporates all cost factors of the constraints. Violations of the hard constraints, e.g. two surgeries at the same time in the same OR, are not allowed. The quality of a valid schedule is determined by the minimization of the soft constraints.

The metaheuristic consists of several computational steps (see Spangenberg et al. 2018 [35]), based on the principle of local search. The used algorithm is Simulated Annealing, since, it has been successfully used in dynamic scheduling domain before it is as well scalable and finds near optimal solutions [4].

Rescheduling is triggered by several factors, for example changes in remaining durations of running interventions based on the ML model. Further, the adding of emergent or urgent patients to the set of interventions leads to the execution of the rescheduling procedure. After collecting information of the PS, the RS starts to adapt the current schedule to events and changes in surgeries. The RS is responsible for the generation of valid surgery schedules based on the resources and constraints.

## V. EVALUATION

1) *Situation Detection Subsystem*: To evaluate the suitability of the SDS, the derived and implemented recognition rules were evaluated with data sets of surgical interventions recorded

TABLE I  
EVALUATION OF SURGICAL PHASE RECOGNITION FOR LAPAROSCOPIC SURGERIES (N=15).

Surgical phase	sigmaresection	proctocolectomy
Placing trocars	100.0%	88.8%
Preparation	66.7%	44.4%
Venting Fume	66.7%	88.8%
Dividing rectum	66.7%	–
Extracting rectosigmoid	83.3%	–
Preparing extra-abdominally	83.3%	77.7%
Preparing intra-abdominally	83.3%	77.7%
Retracting trocars	100.0%	100.0%
Overall	83.3%	76.4%

in the Department of Surgery at Heidelberg university hospital. The overall accuracy of our approach for the recognition of surgical phases in sigmaresection interventions is about 83%. Table I shows the effectiveness of the SDS. The phases with the highest prediction rates are the placing trocars phase, the begin and end of the non-minimally invasive tasks and the retracting of the trocars. This is because of very expressive parameter pattern in the data of the medical devices. For placing and retracting trocars these are the insufflator parameters of pressure, flow and gas volume which depict the drop and rise of the parameters at the creation of the pneumoperitoneum and the insertion of the trocars in the abdominal cavity. The non-minimally invasive tasks have high prediction rates as well, due to the opposite usage of the OR lights and the endoscope lights.

2) *Prediction Subsystem*: A leave-one-out cross validation was performed on the level of surgeries to evaluate the predictions of remaining intervention times. 14 surgeries were used to train the model and one was taken to validate the resulting model. This process was repeated for each surgery. To find the best possible prediction approach the performance was evaluated with several algorithms (for more details see [34]). It is important to consider the different mean surgery lengths (MSD)<sup>1</sup> and standard deviations in the data set. The surgery data used in our evaluation has a mean length of 183.8 minutes (std 50 min). Since the accuracy of remaining time predictions is most important for the timing of the scheduling and preparation of the next surgery, it is not relevant at the very beginning of the procedure. Hence, an increased accuracy at the end of a surgery is crucial for the usability of the system. Some of the discussed related papers also present prediction errors in minutes that are comparable to our results. They are shown along with the used surgical procedure and their mean time in table II. The results are comparable and in some cases better than the current approaches but it has to be taken into account that all systems are evaluated on different data and different kinds of surgeries which are of different length and complexity. Therefore, the comparison can not be used to determine one of the prediction engines clearly outperforming the others. This could only be achieved by a benchmark that uses unified input data.

<sup>1</sup>MSD – Mean Surgery Duration, MAE – Mean Average Error

TABLE II

COMPARISON OF APPROACHES FOR REMAINING INTERVENTION TIME PREDICTION. THE MAE/MSD RATIO DESCRIBES THE PROPORTION OF THE PREDICTION DEVIATION AND THE SURGERY LENGTH. LONG SURGERIES ARE PRESUMABLY MORE DIFFICULT TO PREDICT, THUS HAVING A HIGHER MAE.

Authors	Procedure type	MSD	MAE	$\frac{MAE}{MSD}$
Franke et al. 2013 [11]	neurosurgical discectomy	75	13.4	0.18
Franke et al. 2013 [11]	brain tumor removal	241	29.3	0.12
Guedon et al. 2016 [14]	laparoscopic cholecystectomy	-	14	-
Li et al. 2017 [22]	trauma resuscitation	46.6	6.5	0.14
Maktabi et al. 2017 [24]	lumbar discectomy	54	21.45	0.39
Twinanda et al. 2018 [38]	laparoscopic cholecystectomy	38.1	4.2	0.11
Spangenberg 2018	lap. sigmarsection & proctocolectomy	183.3	16.44	0.089

3) *Rescheduling Subsystem*: The RS was evaluated with a simulated environment of an operating room area representing 10 operating rooms each with 10 hours of operation/day and 4 starting time slots/hour. A data set of 15 surgeries simulate a surgical day. The detected intra-surgical phases trigger the calculation of remaining intervention times and use this information afterwards to start rescheduling. In this stage the interventions can have five different states:

- **Planned**: Are introduced to the system, but OR or time slot are not assigned yet.
- **Scheduled**: OR or time slots are assigned, but intervention did not start already.
- **In progress**: Intervention is running and changes in running intervention time are likely but OR isnt moveable.
- **Reassigned**: Scheduled intervention is reassigned to other OR or time slot.
- **Cancelled**: Delayed with higher priority for next day.

In addition, a formative and artificial evaluation approach according to the evaluation framework FEDS was conducted [39] by implementing a prototype of a DSS for the OSS problem. The formative aspects of the evaluation focus on the outcome of the artefact for the described problem since the goal is to improve the results of the decision processes in OR area. Namely, the Technical Risk & Efficacy evaluation strategy was conducted due to the high costs and risks for evaluation in the OR environment. By using a set of real-world data streams of surgeries, including medical and surgical device data as well as historical data of surgical phase events and the possibility to add emergent patients, a synthetic, but realistic scenario of a day in OR area was created.

## VI. CONCLUSION

In this research project, a novel solution approach for supporting the OSS problem by a real-time DSS for the OR manager was presented. The proposed approach denotes an innovative solution since most of the current approaches operate on the tactical and strategical planning and scheduling with longer time horizons. Real-time data of surgical and OR devices were utilized to monitor current states of each surgery. Predictive Analytics methods and an online optimization technique were employed to provide the decision-maker with additional information about remaining intervention times

and suggestions for rescheduling in case of uncertainties and emergencies. In this work, the involvement and modelling of domain knowledge and their linkage with above mentioned technologies are key factors for the recommendation of alternative actions in decision-making and the progressive automation of this processes. The composition of these models and the needed data processing components in an DSS architecture means a key challenge of this work. Since DSS development and research is highly domain-driven, the use case of operational OR management and OSS are an important consideration for these models and the entire work. The characteristics of operational decisions constitute one of the main challenge of today's Business and Decision Analytics research, since the progressively complex and dynamic nature of this type of decisions needs the involvement of more and up-to-date information [10, 28]. As part of the Business Information Systems Engineering research, this work classifies into the field of business and decision analytics – *"the research and application area comprising theory, methodology, and practice to analyse and support important managerial decisions"* with DSS – which tries to transform data into insights for better decision-making [37]. Decision Analytics makes use of the ideas and concepts of decision theory and DSS, but as well, methods and concepts of AI, Statistics and Computer Science form the basis of this research field [3].

The scientific contributions as well relate to the field of Business and Decision Analytics. The main challenge of this research area is the increasing complexity and dynamics of todays business decisions [3]. This work provides a novel DSS approach, innovative models and concepts which consider exactly these problems with regards to the characteristics of OSS. According to the Design Science guideline for contributions in IS research by Hevner et al. [19], the BISE knowledge base can be extended by applying existing knowledge in new and innovative ways. Hence, the contribution of this research are the innovative artefacts itself, since the mentioned problems have not been fully solved so far and the previously scarcely considered OSS use-case provides new impetus [29]. The novel approach and its artefacts presented in this research push the body of knowledge beyond prior domain constraints of OSS, making the field of research receive more testing and support, as well as problem and solution domain are

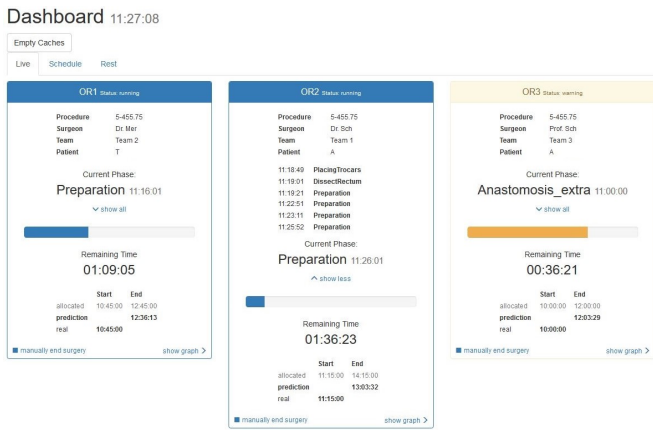


Fig. 4. Prototype - Dashboard view: ORs including raw data, surgical phases and predictions.

considered more mature. As a result of this process, knowledge becomes more abstract, more general and more comprehensive with clearer delimitation of boundaries [13].

In future work, several aspects has to be considered. The SDS for example should integrate data sources that are available in other surgical methods, e.g. using image- and video-data of minimally-invasive surgeries, identify instrument usage in open surgical procedures or location sensor data. In addition, new self-learning methods should be used to detect surgical phases, e.g. with the use of Time Series Analysis or Artificial Neural Networks. The accuracy of the Prediction Subsystem could be improved by the integration of patient-specific information like age, gender, medical history, among others. The Rescheduling Subsystem so far uses information of the current surgery schedule about necessary staff and resources. Modelling more resources and constraints would lead to a more realistic model of the OR environment. The evaluation of the DSS architecture was done with the prototypical implementation and means a proof-of-concept. Upcoming research should focus on the evaluation of the solution approach and its implementation in a real-world setting of the OR environment.

#### REFERENCES

- [1] Haldun Aytug et al. "Executing production schedules in the face of uncertainties: A review and some future directions". In: *European Journal of Operational Research* 161.1 (2005), pp. 86–110. ISSN: 0377-2217. DOI: 10.1016/j.ejor.2003.08.027.
- [2] Mehmet Başdere and Ümit Bilge. "Operational aircraft maintenance routing problem with remaining time consideration". In: *European Journal of Operational Research* 235.1 (2014), pp. 315–328. ISSN: 0377-2217.
- [3] Martin Bichler et al. "Theories in Business and Information Systems Engineering". In: *Business & Information Systems Engineering* 58.4 (2016), pp. 291–319.

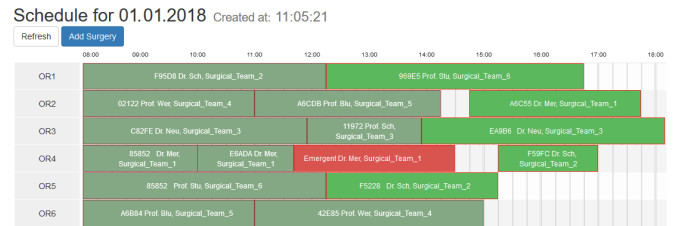
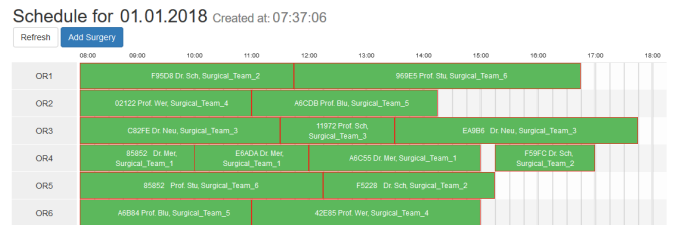


Fig. 5. Prototype - Surgery scheduling view: initial schedule and after rescheduling due to emergency (of a simulated OR plan).

- [4] Sara Ceschia and Andrea Schaerf. "Dynamic patient admission scheduling with operating room constraints, flexible horizons, and patient delays". In: *Journal of Scheduling* 19.4 (2016), pp. 377–389.
- [5] Olga Dergachyova et al. "Automatic data-driven real-time segmentation and recognition of surgical workflow". In: *International Journal of Computer Assisted Radiology and Surgery* 11.6 (2016), pp. 1081–1089. ISSN: 1861-6429.
- [6] Franklin Dexter et al. "Making Management Decisions on the Day of Surgery Based on Operating Room Efficiency and Patient Waiting Times". In: *Anesthesiology* 101.6 (2004), pp. 1444–1453. ISSN: 0003-3022.
- [7] Marinus J. C. Eijkemans et al. "Predicting the Unpredictable: A New Prediction Model for Operating Room Times Using Individual Characteristics and the Surgeon's Estimate". In: *The Journal of the American Society of Anesthesiologists* 112.1 (2010), pp. 41–49.
- [8] Ergin Erdem, Xiuli Qu, and Jing Shi. "Rescheduling of elective patients upon the arrival of emergency patients". In: *Decision support systems* 54.1 (2012), pp. 551–563. ISSN: 0167-9236. DOI: 10.1016/j.dss.2012.08.002.
- [9] EsperTech Inc. *Esper*. 2018. URL: <http://www.espertech.com/esper/>.
- [10] Andreas Fink et al. "Model-Based Decision Support in Manufacturing and Service Networks". In: *Business & Information Systems Engineering* 6.1 (2014), pp. 17–24.
- [11] Stefan Franke, Jürgen Meixensberger, and Thomas Neumuth. "Intervention time prediction from surgical low-level tasks". In: *Journal of Bio. Inf.* 46.1 (2013), pp. 152–159.
- [12] Bernd Graubner. *OPS Systematisches Verzeichnis 2014: Operationen-und Prozedurenschlüssel-Internationale Klassifikation der Prozeduren in der Medizin Version 2014*. Deutscher Ärzteverlag, 2013. ISBN: 3769135393.

- [13] Shirley Gregor and Alan R. Hevner. "Positioning and presenting design science research for maximum impact". In: *Mis Quarterly* 37.2 (2013). ISSN: 0276-7783.
- [14] Annetje Guédon et al. "'It is Time to Prepare the Next patient' Real-Time Prediction of Procedure Duration in Laparoscopic Cholecystectomies". In: *Journal of Medical Systems* 40.12 (2016), pp. 271–277. ISSN: 1573-689X.
- [15] Francesca Guerriero and Rosita Guido. "Operational research in the management of the operating theatre: a survey". In: *Health Care Management Science* 14.1 (2011), pp. 89–114. ISSN: 13869620.
- [16] Rosita Guido and Domenico Conforti. "A hybrid genetic approach for solving an integrated multi-objective operating room planning and scheduling problem". In: *Computers & Operations Research* 87 (2017), pp. 270–282. ISSN: 0305-0548. DOI: 10.1016/j.cor.2016.11.009.
- [17] Z. X. Guo et al. "An RFID-based intelligent decision support system architecture for production monitoring and scheduling in a distributed manufacturing environment". In: *International Journal of Production Economics* 159 (2015), pp. 16–28. ISSN: 0925-5273. DOI: 10.1016/j.ijpe.2014.09.004.
- [18] Erwin W. Hans and Peter T. Vanberkel. "Operating theatre planning and scheduling". In: *Handbook of Healthcare System Scheduling*. Ed. by Randolph Hall. International Series in Operations Research & Management Science. Springer US, 2012, pp. 105–130. ISBN: 978-1-4614-1733-0.
- [19] Alan R. Hevner et al. "Design science in information systems research". In: *Mis Quarterly* 28.1 (2004), pp. 75–105. ISSN: 0276-7783.
- [20] Zhi-Hua Hu and Zhao-Han Sheng. "A decision support system for public logistics information service management and optimization". In: *Decision support systems* 59 (2014), pp. 219–229. ISSN: 0167-9236.
- [21] Florent Lalys and Pierre Jannin. "Surgical process modelling: a review". In: *International Journal of Computer Assisted Radiology and Surgery* 9.3 (2014), pp. 495–511. ISSN: 1861-6429. DOI: 10.1007/s11548-013-0940-5.
- [22] Xinyu Li et al. "Progress Estimation and Phase Detection for Sequential Processes". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1.3 (2017), pp. 1–20. DOI: 10.1145/3130936.
- [23] Hao Luo, Ji Fang, and George Q. Huang. "Real-time scheduling for hybrid flowshop in ubiquitous manufacturing environment". In: *Intelligent Enterprise Systems* 84 (2015), pp. 12–23. ISSN: 0360-8352. DOI: 10.1016/j.cie.2014.09.019.
- [24] M. Maktabi and T. Neumuth. "Online time and resource management based on surgical workflow time series analysis". In: *International Journal of Computer Assisted Radiology and Surgery* 12.2 (2017), pp. 325–338. ISSN: 1861-6429.
- [25] Nathan Marz and James Warren. *Big Data: Principles and best practices of scalable realtime data systems*. Manning Publications Co., 2015.
- [26] Neal Master, David Scheinker, and Nicholas Bambos. "Predicting pediatric surgical durations". In: *arXiv preprint arXiv:1605.04574* (2016).
- [27] Jerrold H. May et al. "The Surgical Scheduling Problem: Current Research and Future Opportunities". In: *Production & Operations Management* 20.3 (2011), pp. 392–405. ISSN: 10591478.
- [28] Peter Mertens and Dina Barbian. "Beherrschung systemischer Risiken in weltweiten Netzen". In: *Informatik-Spektrum* 38.4 (2015), pp. 283–289. ISSN: 0170-6012.
- [29] Hubert Österle, Robert Winter, and Walter Brenner. *Gestaltungsorientierte Wirtschaftsinformatik: Ein Plädoyer für Rigor und Relevanz*. Infowerk, 2010. ISBN: 3000303103.
- [30] Nicolas Padoy et al. "Statistical modeling and recognition of surgical workflow". In: *Medical image analysis* 16.3 (2012), pp. 632–641. ISSN: 1361-8415.
- [31] Atle Riise, Carlo Mannino, and Edmund K. Burke. "Modelling and solving generalised operational surgery scheduling problems". In: *Computers & Operations Research* 66 (2016), pp. 1–11. ISSN: 0305-0548.
- [32] Michael Samudra et al. "Due time driven surgery scheduling". In: *Health Care Management Science* 20.3 (2017), pp. 326–352. ISSN: 13869620. DOI: 10.1007/s10729-016-9356-4.
- [33] Michael Samudra et al. "Scheduling operating rooms: Achievements, challenges and pitfalls". In: *Journal of Scheduling* 19.5 (2016), pp. 493–525.
- [34] Norman Spangenberg, Moritz Wilke, and Bogdan Franczyk. "A Big Data architecture for intra-surgical remaining time predictions". In: *Procedia Computer Science* 113 (2017), pp. 310–317. ISSN: 1877-0509.
- [35] Norman Spangenberg et al. "Implementation of a Situation Aware and Real-Time Approach for Decision Support in Online Surgery Scheduling". In: *31st IEEE International Symposium on Computer-Based Medical Systems, CBMS 2018, Karlstad, Sweden, June 18-21, 2018*, 2018, pp. 417–421. DOI: 10.1109/CBMS.2018.00079.
- [36] Norman Spangenberg et al. "Online Surgery Rescheduling - A Data-driven Approach for Real-time Decision Support". In: *Proceedings of the 20th International Conference on Enterprise Information Systems, ICEIS 2018, Funchal, Madeira, Portugal, March 21-24, 2018, Volume 1*, 2018, pp. 336–343. DOI: 10.5220/0006805103360343.
- [37] Leena Suhl and Stefan Voß. "An Introduction to the Special Focus Issue "Decision Analytics"". In: *Business & Information Systems Engineering* 6.3 (2014), pp. 129–130. ISSN: 2363-7005. DOI: 10.1007/s12599-014-0324-6.

- [38] Andru P. Twinanda et al. “RSDNet: Learning to Predict Remaining Surgery Duration from Laparoscopic Videos Without Manual Annotations”. In: *arXiv preprint arXiv:1802.03243* (2018).
- [39] John Venable, Jan Pries-Heje, and Richard Baskerville. “FEDS: A framework for evaluation in design science research”. In: *European Journal of Information Systems* 25.1 (2016), pp. 77–89. ISSN: 0960-085X.
- [40] Guilherme Vieira, Jeffrey Herrmann, and Edward Lin. “Rescheduling Manufacturing Systems: A Framework of Strategies, Policies, and Methods”. In: *Journal of Scheduling* 6 (2003), pp. 39–62.
- [41] Thomas Wilde and Thomas Hess. “Forschungsmethoden der Wirtschaftsinformatik”. In: *WIRTSCHAFTSINFORMATIK* 49.4 (2007), pp. 280–287. ISSN: 0937-6429. DOI: 10.1007/s11576-007-0064-z.

# Potentiale künstlicher neuronaler Netze zur Identifikation von Ironie im Rahmen der Sentiment-Analyse

Tetzner, Anja  
Lehrstuhl Wirtschaftsinformatik II – Systementwicklung/ Anwendungssysteme  
Technische Universität Chemnitz  
Chemnitz, Deutschland  
anja.tetzner@wirtschaft.tu-chemnitz.de

**Abstract**—Änderungen im Kommunikationsverhalten sowie dem Umgang mit Daten und Informationen führen dazu, dass der Anteil unstrukturierter Daten im Bereich der Datenanalyse steigt. Zur Auswertung unstrukturierter Daten sind spezifische Methoden erforderlich. Hierzu zählt auch die Sentiment-Analyse, welche sich als ein Teilbereich des Text Mining mit der Analyse subjektiver Meinungsäußerungen in Texten befasst. Gerade diese Inhalte bedienen sich häufig komplexer sprachlicher Mittel, die die Datenanalyse vor besondere Herausforderungen stellt. Eine dieser Herausforderungen ist die Identifikation von Ironie. Klassische Verfahren aus dem Bereich des maschinellen Lernens wie bspw. Naïve-Bayes-Klassifikatoren, Maximum-Entropy-Klassifikatoren und Support Vector Machines erzielen durchaus gute Ergebnisse im Bereich der Erkennung von Ironie, sind zu einem großen Teil aber mit einem hohen manuellen Aufwand verbunden. Verfahren aus dem Bereich künstlicher neuronaler Netze und insbesondere Deep Learning bieten eine vielversprechende Alternative, in dem sie den manuellen Aufwand bei der Durchführung der Datenanalyse verringern.

**Keywords**—Künstliche neuronale Netze, Deep Learning, Ironie, Sentiment-Analyse

## I. EINLEITUNG

Der Anteil unstrukturierter Daten im Bereich der Datenanalyse steigt, nicht zuletzt durch die Digitalisierung, die Verbreitung mobiler Endgeräte und die zunehmende Nutzung von Social Media. Insbesondere letzteres trägt durch die starke Beeinflussung der täglichen Kommunikation sowie dem Teilen von persönlichen Daten zu dem Wachstum unstrukturierter Daten bei. Diese Daten besitzen ein weitreichendes Potenzial für bspw. Unternehmen und staatliche Organisationen. Informationen, die vor der Verbreitung von Social Media nur durch direkte Ansprache oder bspw. über Studien, Meinungsumfragen oder Fokusgruppen ermittelt werden konnten, kommunizieren die Nutzer heute eigenmotiviert öffentlich. Unternehmen und staatliche Organisationen können diese Informationen nutzen, um ein besseres Verständnis der Kunden bzw. Bürger sowie eine Reflektion der öffentlichen Wahrnehmung zu erhalten.

Die effektive und effiziente Nutzung der aus Social Media gewonnen Informationen ist mit einigen Herausforderungen verbunden. Neben dem Umfang [1], [2] stellt vor allem die Art der Daten [3] eine Herausforderung dar. Das Text Mining als Forschungsbereich zur Analyse unstrukturierter Textdaten bedient sich hierfür einer Reihe von Methoden und Algorithmen aus verschiedenen Disziplinen, darunter dem maschinellen Lernen (engl.

Machine Learning), Information Retrieval, Natural Language Processing und der Statistik [4]. Statistische Analysen ermitteln unter anderem die Häufigkeit von Wörtern in einem Textdokument. Die Aussagekraft dieser einfachen Kennzahlen ist begrenzt. Sie sagen bspw. nichts über Zusammenhänge zwischen den einzelnen Inhalten eines Textes sowie der Meinung und Emotionen des oder der Autoren aus. Gerade diese Informationen besitzen aber einen großen Informationsgehalt für Unternehmen und staatliche Organisationen. Sie enthalten bspw. wesentliche Aspekte zur Einschätzung von Produkten oder Dienstleistungen durch Konsumenten, welche für das Treffen strategischer Marktentscheidungen von besonderer Bedeutung sind [5]. Der Umstand, dass Individuen im eigenen Entscheidungsprozess aktiv nach Meinung anderer suchen [6], verdeutlicht die Notwendigkeit der Analyse öffentlich verbreiteter Meinungsäußerungen. Eine wesentliche Bedeutung im Text Mining kommt daher dem Erkennen und Analysieren von Zusammenhängen in Texten zu, insbesondere in Bezug auf Meinungen. Speziell diesem Bereich des Text Mining widmet sich die Sentiment-Analyse, die definiert ist als computergestützte Untersuchung von Subjektivität und damit einhergehender Meinungen und Stimmungen in Texten [7], [8].

Aufgrund der Vielfältigkeit der natürlichen Sprache steht die Sentiment-Analyse vor einer Reihe an Problemstellungen, darunter die Differenzierung zwischen objektiven und subjektiven Texten, Domänenabhängigkeit im Verständnis von Begriffen etc. [6], [7]. Eine besondere Herausforderung stellt das Erkennen von Ironie dar. Ironie ist ein häufig genutztes sprachliches Mittel in der vorwiegend persönlichen und informellen Kommunikation [9], [10]. Das Erkennen von Ironie setzt komplexe kognitive Fähigkeiten voraus, da Ironie nicht durch eine formale Beschreibung oder klare Regeln zu definieren ist [10], was eine automatisierte, computergestützte Identifikation erschwert. Zur Lösung dieser Problemstellung bedient sich die Sentiment-Analyse verschiedener Methoden. Insbesondere Methoden aus dem maschinellen Lernen bieten großes Potential [8]. Häufige Anwendung finden vor allem Naïve-Bayes-Klassifikatoren, Maximum-Entropy-Klassifikatoren und insbesondere Support Vector Machines [5], [11]. Künstliche neuronale Netze hingegen finden bei der Identifikation von Ironie nur wenig Einsatz. In anderen Teilbereichen der Sentiment-Analyse, bspw. der Sentiment-Klassifikation, der Aspekt-Extraktion und der Erstellung von Sentiment-Lexika, konnten durch künstliche neuronale Netze und insbesondere



Deep Learning<sup>1</sup> bereits signifikant bessere Ergebnisse erzielt werden [5], [11] als mit klassischen Verfahren. Zusätzlich bietet diese Ansätze den Vorteil einer Reduktion des manuellen Aufwands bei der Datenanalyse [12]. Dieser ist darin begründet, dass insbesondere im Bereich Deep Learning die Netze direkt aus den Rohdaten Repräsentationen (Features) lernen. Nichtlineare Transformationen abstrahieren diese Repräsentationen auf höhere Level. Mit ausreichend hoher Abstraktion werden komplexe Funktionen gelernt, die für die ursprüngliche Aufgabenstellung wesentliche Aspekte aus den Daten herausfiltern [12]. Zur Verbesserung der State-of-the-Art-Ergebnisse im Bereich der Identifikation von Ironie sowie der Verringerung des aktuell teilweise hohen manuellen Aufwandes, insbesondere bei der Feature-Selektion, ist es daher als sinnvoll anzusehen, künstliche neuronale Netze zur Identifikation von Ironie zu entwickeln.

Der vorliegende Beitrag befasst sich mit der Untersuchung der Potentiale und des aktuellen Standes hinsichtlich des Einsatzes künstlicher neuronaler Netze und insbesondere Deep Learning zur Identifikation von Ironie im Rahmen der Sentiment-Analyse. Hierfür gliedert sich der Beitrag wie folgt: Kapitel zwei beschreibt die Sentiment-Analyse und die Problematik der Identifikation von Ironie in Texten näher. Kapitel drei stellt bestehende Forschungsarbeiten zum Einsatz von künstlichen neuronalen Netzen und insbesondere Deep Learning im Bereich der Identifikation von Ironie vor. Kapitel vier beinhaltet Implikationen hinsichtlich der Weiterentwicklung von künstlichen neuronalen Netzen zur Identifikation von Ironie basierend auf den in Kapitel drei beschriebenen Forschungsarbeiten. Kapitel fünf schließt den Beitrag mit einer Zusammenfassung ab.

## II. SENTIMENT-ANALYSE UND IRONIE

Liu [6] definiert die Sentiment-Analyse als die Untersuchung von Meinungen<sup>2</sup> hinsichtlich explizit oder implizit vorhandener positiver oder negativer Stimmungen, auch als Polarität bezeichnet. Eine Meinung ist eine Äußerung einer Person gegenüber einer Entität [8] und kann als ein Tupel bestehend aus fünf Komponenten erfasst werden: einer Entität, einem Aspekt dieser Entität, einer Stimmung gegenüber dem Aspekt der Entität, einem Autor sowie einem Zeitpunkt, zu dem die Meinung geäußert wurde [7]. Die Entität steht stellvertretend für eine Person, ein Gegenstand, eine Organisation, ein Ereignis oder ähnliches. Auf Grund dessen, dass die Ergebnisse der Sentiment-Analyse in Klassen eingeteilt werden, ist als Synonym für die Sentiment-Analyse auch der Begriff Sentiment-Klassifikation gebräuchlich.

Der Prozess der Sentiment-Analyse besteht aus sechs Teilschritten (Abb. 1).

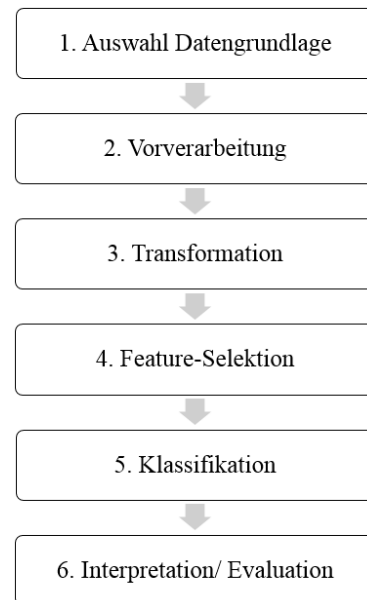


Abb. 1: Prozess der Sentiment-Analyse nach [5].

Im ersten Schritt ist die Datenquelle zu bestimmen, welche zur Informationsgewinnung relevanten Daten enthält. Die Forschung zur Sentiment-Analyse nutzt in diesem Zusammenhang vor allem Produktbewertungen [13], [14], Zeitungsartikel [13], [15], Filmkritiken [13] und Social Media (insbesondere Twitter) [9], [10], [14], [16], [17].

Die Analyse von unstrukturierten Daten zeichnet sich allgemein durch eine hohe Komplexität und Dimensionalität aus [18]. Die Schritte zwei bis vier des in Abb. 1 beschriebenen Prozesses dienen daher vor allem der Reduktion der Komplexität der Daten. Im zweiten Schritt finden Verfahren wie Tokenisierung, N-Gramm-Zerlegung, Stoppwortreduktion, Normalisierung, Wortstammreduktion, Lemmatisierung und Part-of-Speech-Kennzeichnung (POS) Anwendung. Tokenisierung und N-Gramm-Zerlegung ermitteln anhand definierter Regeln aus den extrahierten Rohdaten des ersten Prozessschrittes einzelne Terme<sup>3</sup> [19], [20]. Die Stoppwortreduktion filtert häufig vorkommende Worte, die eine geringe Aussagekraft für die Textanalyse besitzen [19]. Hierzu zählen vor allem Präpositionen und Artikel [5]. Die Normalisierung vereinheitlicht unterschiedliche Schreibweisen eines Wortes [19]. Die Wortstammreduktion reduziert Worte auf ihren Stamm [5], die Lemmatisierung reduziert Worte auf ihre Grundform [19]. Normalisierung, Wortstammreduktion und Lemmatisierung verringern den Umfang der zu analysierenden Daten, in dem sie Worte auf eine einheitliche Form zurückführen und damit Variationen eliminieren. Die POS-Kennzeichnung bestimmt die Wortfamilie jedes Wortes [19]. Bestimmten Wortfamilien, wie z.B. Adjektiven, wird im Rahmen der Sentiment-Analyse eine höhere Bedeutung zur Bestimmung der Polarität eines Textes zugemessen [7], [8]. Schritt drei generiert aus den Daten eine numerische Repräsentation. Eine Möglichkeit ist eine binäre Repräsentation, die angibt ob ein Term generell in einem Text enthalten ist (1) oder nicht (0). Alternativen arbeiten mit der Vorkommenshäufigkeit eines Terms in einem Text [5]. Die Feature-Selektion, Schritt vier, besitzt durch die

<sup>1</sup> Deep Learning als Teilbereich künstlicher neuronaler Netze zeichnet sich neben der automatisierten Extraktion von Features durch komplexe Architekturen mit einer Vielzahl von Neuronen und komplexen Verbindungen zwischen diesen Neuronen aus. Diese komplexen Strukturen ermöglichen es Deep-Learning-Netzen komplizierte Aufgabenstellungen zu lösen, bspw. im Bereich der Bilderkennung, Sprach- und Textverarbeitung [42].

<sup>2</sup> Es werden auch die Begriffe Emotionen, Gefühle, Bewertungen, Einstellung u.ä. in diesem Zusammenhang verwendet.

<sup>3</sup> Der Begriff Term fast in diesem Zusammenhang Wörter, Gruppen von Wörtern definierter Länge (N-Gramme) und Phrasen zusammen.

Festlegung der im nachfolgenden Schritt fünf zu analysierenden Features einen wesentlichen Einfluss auf die Ergebnisqualität [5], [21] und bedarf daher besonderer Sorgfalt. Grundsätzlich bestimmt der Schritt der Feature-Selektion den Informationsgehalt jedes Terms im analysierten Dokument. Die Terme mit dem höchsten Wert bzw. einem Wert über einem definierten Schwellenwert fließen in die weitere Analyse ein [21]. Chi-Quadrat-Test, Document Frequency Threshold, Information Gain, Mutual Information und Term Strength zählen zu dem am häufigsten eingesetzten Methoden im Rahmen der Feature-Selektion der Sentiment-Analyse [5].

Die eigentliche Analyse der Daten findet im fünften Schritt des Prozesses der Sentiment-Analyse statt. Medhat, Hassan und Korashy [8] unterteilen die im Rahmen der Sentiment-Analyse zum Einsatz kommenden Analysemethoden in zwei grundsätzliche Ansätze: die lexikonbasierten Ansätze und die Ansätze des maschinellen Lernens. Im Rahmen der lexikonbasierten Ansätze erfolgt auf Basis eines Textkorpus bzw. einer Ausgangswortliste, die durch Synonyme und Antonyme erweitert wird, die Erstellung eines Sentiment-Lexikons. Das Sentiment-Lexikon ist eine Auflistung von Wörtern, Wortgruppen und Phrasen und deren Polarität (positiv, negativ, neutral bzw. Abstufungen dieser), welche als Referenz zur Klassifikation von Texten dient [7], [8]. Die Ansätze des maschinellen Lernens betrachten die Sentiment-Analyse als klassisches Klassifikationsproblem mit der Besonderheit textueller Features [8].

Im letzten Schritt des Prozesses, der Interpretation bzw. Evaluierung, gilt es, die Güte der Analyseergebnisse aus dem fünften Schritt zu bewerten. Zum Einsatz kommen dabei vor allem die Metriken Accuracy (Richtigkeit), Recall (Vollständigkeit), Precision (Genauigkeit) und F1 [5].

Das vordergründige Ziel des beschriebenen Prozesses ist es, eine Aussage darüber treffen zu können, ob ein Text eine positive, negative oder ggf. auch neutrale Meinung aufweist, um dadurch Rückschlüsse über die betrachtete Entität und ihren beschriebenen Aspekt ziehen zu können. Stilistische Mittel sorgen dafür, dass die Ermittlung der Polarität eines Textes eine komplexe und anspruchsvolle Aufgabe ist. Ein stilistisches Mittel, dass die Ermittlung der Polarität eines Textes auf Grund seiner Komplexität besonders erschwert, ist Ironie. Gleichzeitig stellt die Erkennung von Ironie eine wichtige Aufgabe im Bereich der Sentiment-Analyse dar, deren Missachtung zu einer Verfälschung der Analyseergebnisse führen kann [22]–[24].

Ironie ist der figurativen Sprache zuzuordnen. Diese ist wesentlich schwieriger auszuwerten als literale Sprache, da sie ausgeprägte kognitive Fähigkeiten erfordert, die es ermöglichen, die teilweise abstrakte Aussage hinter der verbal formulierten Aussage zu verstehen [13]. Um eine ironische Aussage korrekt einordnen zu können, muss der Rezipient der Aussage über Kontextwissen verfügen, das es ihm ermöglicht, Ironie von einer faktisch wahren Aussage zu unterscheiden. Der Autor der ironischen Äußerung muss wiederum wissen, dass der Rezipient über das nötige Kontextwissen verfügt, um seine Aussage als ironisch einordnen zu können. Eine wichtige Ressource zur Erlangung des notwendigen Kontextwissens stellt der Gesprächsverlauf dar. Im Vorfeld tauschen die Gesprächspartner das zur Erkennung einer ironischen Aussage notwendige Wissen durch die Weitergabe von

Informationen, Erfahrungen etc. aus. Die ironische Aussage selbst nimmt im Gesprächsverlauf die Aufgabe einer zusammenfassenden Bewertung ein. Dieser Bewertung haftet immer ein positiver oder negativer Wert an. Die Polarität der bewertenden ironischen Aussage muss dabei nicht zwingend in der Aussage selbst erkennbar sein, sondern kann sich durch den vorangegangenen Gesprächsverlauf ergeben. Die ironische Äußerung besitzt in diesem Fall selbst keinerlei sprachlichen Bewertungsausdruck. Auf Grund der Differenz zwischen der formulierten und der tatsächlich intendierten Bewertung einer ironischen Äußerung sowie der binären Distinktivität der Bewertung, kehrt sich in den meisten Fällen, wenn auch nicht in allen, die Bewertung in ihrer Polarität um. Empirische Beobachtungen ergaben, dass ironische Äußerungen fast ausschließlich negative Bewertungen wiedergeben. Diese wiederum werden im Zuge der Differenz zwischen formulierter und tatsächlicher Aussage sprachlich positiv formuliert [25].

### III. KÜNSTLICHE NEURONALE NETZE ZUR IDENTIFIKATION VON IRONIE - AKTUELLER FORSCHUNGSSTAND

Klassische Ansätze zur Identifikation von Ironie (bspw. [9], [10], [14], [15], [18], [26]) basieren vordergründig auf lexikonbasierten Ansätzen und Methoden des maschinellen Lernens wie Naïve-Bayes-Klassifikatoren, Maximum-Entropy-Klassifikatoren und insbesondere Support Vector Machines. Diese Ansätze besitzen den Nachteil, dass sie einen erheblichen manuellen Aufwand insbesondere im Schritt der Vorverarbeitung und der Feature-Selektion (vgl. Abb. 1) verursachen [24], [27], [28]. Außerdem bedient sich ein Großteil dieser Ansätze dem sogenannten Bag-of-Word-Vorgehen, welches anhand von n-Grammen Features extrahiert. Dabei ignoriert dieses Vorgehen weitgehend die Wortreihenfolge sowie semantische Informationen [27]. Eine Identifikation von Ironie anhand einzelner Signalwörter ist allerdings nicht möglich [25], weshalb die Erfassung von Kontextinformationen eine besondere Rolle einnimmt (vgl. Kapitel II). Auf künstlichen neuronalen Netzen und insbesondere Deep Learning basierende Ansätze bieten eine Lösung dieser Problemstellung. Sie verzichten auf eine umfangreiche Vorverarbeitung und lernen die Features, sogenannte Embeddings<sup>4</sup>, direkt und automatisiert aus den Rohdaten [12], [27], [29]. Die Embeddings erfassen neben syntaktischen auch semantische Informationen [30] und ermöglichen damit, im Gegensatz zum Bag-of-Word-Vorgehen, das Erfassen von Kontextinformationen.

Einen aktuellen Überblick über die Forschung zur Identifikation von Ironie mittels computergestützter Methoden liefern die Beiträge zum International Workshop on Semantic Evaluation 2018 (SemEval 2018)<sup>5</sup>. Erstmals wurde mit dem Task 3 „Irony Detection in English Tweets“ explizit die Identifikation von Ironie als eigenständige Aufgabe in den bereits seit 1998 bestehenden Workshop aufgenommen. Die Aufgabe bestand aus zwei Teilaufgaben. Teilaufgabe A beinhaltete eine binäre Einteilung von Tweets in die Klassen ironisch und nicht-ironisch.

<sup>4</sup> Embeddings leiten aus der Struktur der Inputdaten automatisiert niedrig-dimensionale verteilten Feature-Repräsentationen ab [44].

<sup>5</sup> <http://alt.qcri.org/semeval2018/>

TABELLE I.

BEITRÄGE SAMEVAL 2018 MIT BEZUG ZU DEEP LEARNING

Beitrag	Basis Klassifikator	Teilaufgabe	Accuracy <sup>a</sup>	Precision	Recall	F1
Baziotis, Athanasiou, Papalampidi, Kolovou, Paraskevopoulos, Ellinas und Potamianos [22]	BiLSTM	A	73,21	<b>65,35</b>	69,13	67,19
		B	65,18	49,59	51,24	49,59
Cho, Kang, und Kim [31]	BiLSTM, CNN	A	64,41	54,26	65,59	59,39
		B	59,82	41,17	40,96	39,25
Dragoni [32]	RNN	A	50,38	40,85	55,95	47,22
		B	44,13	21,33	23,14	21,89
Ghosh und Veale [33]	LSTM	A	68,24	64,49	44,37	52,57
		B	<b>73,21</b>	<b>57,86</b>	50,44	<b>50,74</b>
González, Hurtado und Pla [34]	CNN in Kombination mit LSTM	A	61,10	50,59	<b>83,28</b>	62,94
		B	63,27	41,23	44,04	42,11
Marrese-Taylor, Ilic, Balazs, Matsuo und Prendinger [28]	BiLSTM	A	35,20	25,68	33,44	29,05
Nikhil und Srivastava [35]	BiLSTM	A	66,59	55,28	64,71	59,62
Peng, Wang und Zhang [36]	LSTM, BiLSTM, Attention BiLSTM, GRU, BiGRU, Attention BiGRU	A	50,89	39,12	42,77	40,86
		B	53,32	43,79	26,70	26,05
Rajalakshmi, Angel, Milton und Mirmalinee [37]	MLP	B	57,27	34,84	36,09	33,37
Rangwani, Kulshreshtha und Kumar Singh [38]	CNN (lediglich Feature-Selektion)	A	66,07	55,06	78,78	64,81
		B	60,33	46,60	50,58	47,43
San [23]	LSTM	A	61,73	51,35	67,20	58,22
		B	63,27	48,68	43,88	43,52
Vu, Nguyen, Vu, Nguyen, Catt und Trenell [39]	MLP	A	70,15	60,91	69,13	64,76
		B	65,94	54,46	44,75	44,37
Wu, Wu, Wu, Liu, Yuan und Huang [24]	BiLSTM	A	<b>73,47</b>	63,04	80,06	<b>70,54</b>
		B	60,46	48,60	<b>54,14</b>	49,47
Yin, Wang, Lan und Wang [40]	BiLSTM	A	59,57	49,36	74,28	59,31
		B	30,36	25,46	24,93	23,26

<sup>a</sup> Fettgedruckte Zahlen geben das beste Ergebnis für Teilaufgabe A an, fett- und kursivgedruckte Zahlen geben das beste Ergebnis für Teilaufgabe B an.

Teilaufgabe B beinhaltet eine feingranularere Unterteilung der Klassen in: verbale Ironie mit Umkehr der Polarität des Textes, situative Ironie, andere verbale Ironie und keine Ironie [41]<sup>6</sup>.

Neben der Umsetzung mittels klassischer Verfahren wurden 14 Beiträge<sup>7</sup> eingereicht, welche die Aufgabe der Identifikation von Ironie mittels künstlicher neuronaler Netze bzw. Deep Learning umsetzen. Dabei befassen sich 13 der 14 Beiträge mit der Teilaufgabe A und 12 Beiträge mit der Teilaufgabe B. Eine Übersicht der Beiträge inklusive der erreichten Testergebnisse ist in Tabelle I zu finden.

Die Untersuchung der Beiträge zeigt, dass sich die Autoren hauptsächlich zwei grundsätzlicher Architekturen von Deep-Learning-Netzen bedienen, Recurrent Neural Networks (RNN) und Convolutional Neural Networks (CNN). RNN dienen der Verarbeitung sequenzieller Daten. Sie verarbeiten in jedem Durchlauf jeweils einen Teil der Inputdatensequenz. Der Output des aktuellen Durchlaufes fließt durch rekursive Verbindungen in den versteckten Schichten des Netzes in den nachfolgenden Durchlauf ein, wodurch der zurückliegende Teil der Inputsequenz zu jedem Zeitpunkt Einfluss auf den aktuellen Durchlauf nimmt [42]. Auf diese Weise werden kontextuelle Informationen eines Textes erfasst [27]. RNN neigen bei einer langen Inputsequenz allerdings dazu, dass der Gradient verschwindet bzw. explosionsartig ansteigt [43]. Diesem Problem wirken Long Short-Term Memory (LSTM) bzw.

Gated Recurrent Units (GRU) als Varianten von RNN entgegen. LSTM ergänzen jedes Neuron, welches in diesem Zusammenhang auch als Memory-Zelle bezeichnet wird, um drei Gates, einem Input Gate, einem Forget Gate und einem Output Gate [33]. Die Gates sorgen dafür, dass das LSTM zwischen nützlichen und unwichtigen Informationen unterscheiden kann [36]. Lediglich nützliche Information können die Gates passieren und beeinflussen den Output der Memory-Zelle [42]. Bidirektionale LSTM (BiLSTM), als eine Sonderform der LSTM, berücksichtigen ausgehend von einem Punkt in der Inputsequenz sowohl die zurückliegenden Informationen als auch die folgenden [22]. GRU ist eine vereinfachte Variation von LSTM, welche lediglich zwei Gates nutzt, das Input Gate, auch als Update Gate bezeichnet, und das Forget bzw. Reset Gate [36]. CNN zeichnen sich durch zwei unterschiedliche Arten versteckter Schichten aus, den Convolutional-Schichten und den Pooling-Schichten. Die Convolutional-Schicht bestimmt aus einer definierten Region des Input über eine Gewichtsmatrix (Kernelfunktion) die Aktivierung und damit den Output eines Neurons. Die Pooling-Schicht aggregiert den Output der Convolutional-Schicht anhand vordefinierter Metriken um dem Umfang der Daten einzugrenzen und einem Overfitting entgegen zu wirken. Mehrere Paare an Convolutional- und Pooling-Schichten können in einem CNN aufeinander folgen, wodurch das Lernen komplexer Feature möglich ist [42]. Der Output der letzten Pooling-Schicht dient anschließend einer Schicht mit vollständig vernetzten Neuronen als Feature zur Klassifikation [44].

Baziotis et al. [22] kombinieren über zwei BiLSTM Wort- und Zeichen-Embeddings. Über einen Attention Layer wird der Beitrag eines jedes Wortes bzw. Zeichens gewichtet und die Ergebnisse anschließend zur Bestimmung eines Klassenlabes in einer Outputschicht kombiniert. Um einer fehlerhaften Klassifikation der unterrepräsentierten Klassen

<sup>6</sup> Die Trainings- und Testdaten umfassen einen Korpus von 4.792 englischsprachigen Tweets (2.396 ironisch, 2.396 nicht-ironisch). Die ironischen Tweets wurden mittels Twitter-Suche nach den Hashtags #irony, #sarcasm und #not gesammelt, die nicht-ironischen Tweets stammen aus dem selben Nutzerkreis. Zur Kontrolle wurden alle ironischen Tweets zusätzlich manuell annotiert.

<sup>7</sup> Untersucht wurden in diesem Zusammenhang die offiziellen Beiträge aus den Proceedings of International Workshop on Semantic Evaluation. Diese enthalten nicht zu jeder eingereichten Lösung einen Beitrag.

in Teilaufgabe B vorzubeugen, wurde mit Klassengewichten in der Fehlerfunktion gearbeitet.

Cho, Kang und Kim [31] vergleichen die Performance von BiLSTM und CNN. Besonderer Fokus liegt dabei auf dem Einfluss von Hashtags, die in den Embeddings gesondert berücksichtigt werden. Das BiLSTM liefert im Vergleich zum CNN das bessere Ergebnis, was nach Aussage des Autors zeigt, dass die Identifikation von Ironie durch spezifische Terme oder Wordsequenzen der Tweets beeinflusst wird.

Dragoni [32] nutzt ein RNN mit vortrainierten Wort-Embeddings basierend auf einem subjektiven, meinungsbehafteten Korpus anstatt eines allgemeinen, objektiven, wie es in den anderen Beiträgen der Fall ist.

Ghosh und Veale [33] entwickelten ein siamesisches neuronales Netz<sup>8</sup>, das aus zwei Subnetzen besteht. Jedes Subnetz umfasst eine Embedding-Schicht, initialisiert durch vortrainierte Embeddings, und ein LSTM. Der Input jedes Subnetzes besteht aus einem Teilstück eines Satzes. Über den Output der Subnetze erfolgt in einer Subtraktionsschicht ein Vergleich der beiden Satzteile. Sind beiden Satzteile semantisch inkongruent, wird der Satz als ironisch klassifiziert, andernfalls als nicht-ironisch.

González, Hurtado und Pla [34] verfolgen einen Ansatz bestehend aus der Kombination eines zweischichtigen CNN und eines anschließenden LSTM.

Marrese-Taylor, Ilic, Balazs, Matsuo und Prendinger [28] nutzen vier parallele BiLSTM mit unterschiedlicher, zufälliger Initialisierung, deren Ergebnisse sie kombinieren. Neben Wort-Embeddings setzen die Autoren zusätzliche Features auf Satz- und Token-Ebene ein, die sie mittels einfacher linguistischer Muster erzeugen. Durch entsprechende Tests zeigen die Autoren, dass die zusätzlichen Features auf Satzebene positiv zum Ergebnis beitragen, während die Features auf Token-Ebenen eher einen negativen Einfluss auf das Ergebnis ausüben.

Nikhil und Srivastava [22] zerlegen die Tweets mittels Abhängigkeitsparser (Dependency Parser) zunächst in einzelne Abschnitte, welche anschließend an eine Embedding-Schicht übergeben werden. Zwei BiLSTM verarbeiten die Embeddings weiter und übergeben ihre Ergebnisse an einen Attention Layer. Eine Outputschicht ermittelt final aus den Ergebnissen des Attention Layer die entsprechenden Klassenlabels.

Peng, Wang und Zhang [36] kombinieren zur optimalen Bestimmung eines Klassenlabels die Ergebnisse verschiedener Deep-Learning-Netze mittels Voting-Mechanismus. Insgesamt trainierten die Autoren sechs individuelle Netze, ein LSTM, ein BiLSTM, ein GRU, ein bidirektionales GRU (BiGRU), ein attention-based BiLSTM und ein attention-based BiGRU. Ein Voting-Mechanismus bestimmt das Klassenleben unter Berücksichtigung bestimmter Regeln durch die Selektion des Klassenlabels, welches von den meisten Netzen prognostiziert wird. Einzeln betrachtet erzielen das BiLSTM und das GRU die besten Ergebnisse.

<sup>8</sup> Ein siamesisches neuronales Netz besteht aus zwei Subnetzen, die mit unterschiedlichen Inputdaten trainiert werden um den Unterschied zwischen diesen Inputdaten hervorzuheben.

Rajalakshmi, Angel, Milton und Mirnalinee [37] kombinieren einen klassischen, regelbasierten Ansatz zur Feature-Selektion mit einem Multilayer Perzeptron (MLP) als Klassifikator. MLP sind einfache künstliche neuronale Netze, die aus einer Schicht Eingabeneuronen, einer oder mehreren Schichten versteckter Neuronen und einer Schicht Ausgabeneuronen bestehen. Jedes Neuron einer Schicht ist vollständig mit den Neuronen der nachfolgenden Schicht verbunden. Das MLP der Autoren besteht aus einer Schicht Eingabeneuronen, zwei Schichten versteckter Neuronen und einer Schicht Ausgabeneuronen.

Rangwani, Kulshreshtha und Kumar Singh [38] verknüpfen manuell erfasste linguistische Features mit automatisch erzeugten Embeddings eines CNN. Ähnlich wie Rajalakshmi et al. verbinden die Autoren damit klassische Ansätze mit Ansätzen des Deep Learning. Untersuchungen der Autoren haben gezeigt, dass durch die Kombination der manuell erstellten Features mit den automatisch erzeugten Embeddings bessere Ergebnisse erzielt werden können, als mit dem separaten Einsatz der beiden Ansätze. Die Klassifikation selbst erfolgte über den Klassifikator XGBoost.

San [23] nutzt zwei Arten von Embeddings, Wort-Embeddings und Emoji-Embeddings, in Verbindung mit Sentiment Scores als Input für ein LSTM. Die Ergebnisse des LSTM nutzt eine Outputschicht zur Bestimmung der Klassenlabels.

Vu, Nguyen, Vu, Nguyen, Catt und Trenell [39] übergeben manuell erstellte Features, darunter lexikalische, syntaktische, semantische und Polaritäts-Features, als Input ein einfaches MLP. Ein Vergleich der Autoren zwischen dem entwickelten MLP mit einem LSTM zeigte, dass unter Einbezug ironischer Hashtags das LSTM ein besseres Ergebnis erzielte, ohne Einbezug der ironischen Hashtags erzielt allerdings das MLP ein besseres Ergebnis.

Wu, Wu, Wu, Liu, Yuan und Huang [24] nutzen ein kompaktes BiLSTM-Modell mit mehreren BiLSTM-Schichten. Jede BiLSTM-Schicht übernimmt den Output aller vorherigen Schichten als Input. Dadurch ist ein Lernen verschiedener Level kontextueller Informationen zur gleichen Zeit möglich. Neben Wort-Embeddings werden POS-Tags an die erste BiLSTM-Schicht übergeben. Der Output der letzten BiLSTM-Schicht wird mit Sentiment-Features und Satz-Embeddings kombiniert und an eine Outputschicht übergeben, die die Klassenlabels entsprechend der Teilaufgaben bestimmt. Um die Performance ihres Modells zu verbessern, kombinierten die Autoren die Ergebnisse von zehn Modellen gleicher Architektur aber unterschiedlicher Dropout-Raten.

Yin, Wang, Lan und Wang [40] vergleichen klassische Methoden des maschinellen Lernens mit dem Ergebnis eines BiLSTM sowie einer Kombination beider Varianten. Das Ergebnis zeigt, dass eine Kombination klassischer Methoden und des BiLSTM die besten Ergebnisse liefert.

#### IV. IMPLIKATIONEN

Die eingereichten Beiträge wurden entsprechend ihres F1-Ergebnisses in eine Rangliste eingeordnet. Für die untersuchten 14 Beiträge ist die Ranglistenplatzierung in Tabelle II zu finden, die vollständige Rangliste enthält [41].

TABELLE II.                    UNTERSUCHUNG DER ARCHITEKTUR DER SEMEVAL-BEITRÄGE

Beitrag	Ranking SamEval 2018 (A/B)		Embeddings für	Zusätzliche manuell Features	Zusätzliche externe Ressourcen <sup>a</sup>	Kombination von Netzen bzw. deren Ergebnissen
	Teilaufg. A	Teilaufg. B				
Baziotis, Athanasiou, Papalampidi, Kolovou, Paraskevopoulos, Ellinas und Potamianos [22]	2	2	Zeichen, Worte (vortrainiert)	nein	nein	ja
Cho, Kang, und Kim [31]	19	12	Worte (vortrainiert) mit besonderer Behandlung von Hashtags	nein	nein	nein
Dragoni [32]	33	30	Worte (vortrainiert)	nein	nein	nein
Ghosh und Veale [33]	30	1	Worte (vortrainiert)	nein	nein	nein
González, Hurtado und Pla [34]	7	7	Worte (vortrainiert)	ja (Polaritäten/ Emotionen)	ja	ja
Marrese-Taylor, Ilic, Balazs, Matsuo und Prendinger [28]	41	-	Worte (vortrainiert)	ja (binäre Features auf Satz- und Token-Ebene)	nein	ja
Nikhil und Srivastava [35]	17	-	Phrasen	nein	nein	ja
Peng, Wang und Zhang [36]	39	27	Worte (vortrainiert)	nein	nein	ja
Rajalakshmi, Angel, Milton und Mirnalinee [37]	-	19	-	ja (regelbasierte Features)	nein	nein
Rangwani, Kulshreshtha und Kumar Singh [38]	4	4	Worte mit besonderer Berücksichtigung von Emojis	ja (Feature basierend auf linguistischen Strukturen und Nutzerverhalten)	ja	nein
San [23]	24	6	Worte (vortrainiert), Emoji (vortrainiert)	ja (Polaritäten)	ja	nein
Vu, Nguyen, Vu, Nguyen, Catt und Trenell [39]	5	5	Worte (vortrainiert)	ja (lexikalische und syntactische Features, Polaritäten)	ja	nein
Wu, Wu, Wu, Liu, Yuan und Huang [24]	1	3	Worte (vortrainiert), Sätze (vortrainiert)	ja (POS-Features)	ja	ja
Yin, Wang, Lan und Wang [40]	20	29	Worte (vortrainiert)	ja (linguistische Features)	ja	ja (KNN mit klassischen Methoden)

<sup>a</sup>Hierbei werden die für das Vortrainieren der Embeddings benötigten Ressourcen nicht gesondert berücksichtigt. Es handelt sich bei diesen Ressourcen bspw. um Sentiment-Lexika, oder Listen für die POS-Kennzeichnung.

Für beide Teilaufgaben belegen Beiträge, die künstliche neuronale Netze bzw. Deep-Learning-Ansätze zur Lösung der Aufgabenstellung nutzen, den ersten Platz. Dies verdeutlicht die generelle Überlegenheit dieser Ansätze im Vergleich zu klassischen Ansätzen zur Identifikation von Ironie.

Der Vergleich der Ergebnisse für Teilaufgabe A und B zeigt deutlich, dass die Lösungen für die Teilaufgabe A durchgängig bessere Ergebnisse erzielen. Die binäre Klassifikation der Daten in ironisch bzw. nicht-ironisch gestaltet sich demnach einfacher als eine feingranuläre Klassifikation der Daten. Weite Untersuchungen, vgl. [41], ergaben, dass bei der Identifikation der Klassen nicht-ironisch und verbale Ironie mit Umkehr der Polarität der Teilaufgabe B bessere Ergebnisse erzielt wurden, als für die Klassen situative Ironie und andere verbale Ironie. Dies ist zum Großteil in der Unterrepräsentation der zuletzt genannten beiden Klassen im Datensatz begründet. Für die Identifikation von Ironie und damit verbunden der Sentiment-Analyse ergeben sich aus diesen Ergebnissen zwei Erkenntnisse. Bei der Auswahl der Trainingsdaten ist auf eine ausgeglichene Verteilung der Daten je Klasse zu achten. Außerdem ist eine grobgranuläre, bspw. binäre,

Klasseneinteilung im Sinne korrekter Ergebnisse vorzuziehen. Andere Untersuchungen im Bereich der Sentiment-Analyse, die einen Vergleich zwischen einer groben und einer detaillierteren Klasseneinteilung bzw. Einteilung der Polarität untersuchen, unterstützen dies [30], [45], [46].

Die Betrachtung der Arten der verwendeten Netze (vgl. Tabelle I) zeigt, dass RNN-basierte Netze die mit Abstand am häufigsten genutzten künstlichen neuronalen Netze in den untersuchten Beiträgen darstellen. In elf der 14 Beiträge (rund 78,57%) nutzen die Autoren RNN bzw. Varianten davon. Die am häufigsten genutzte Variante ist das BiLSTM in sechs von elf Beiträgen (rund 54,55%). Generell besitzen RNN durch ihren Aufbau, der die Zusammenhänge und damit den Kontext eines Textes erfasst, eine grundlegend gute Eignung zur Identifikation von Ironie, bei deren Erfassung der Kontext eine essentielle Rolle spielt. Durch die Einseitigkeit der genutzten Netze ist allerdings keine generelle Aussage darüber möglich, ob RNN und speziell BiLSTM im Vergleich zu anderen Arten künstlicher neuronaler Netze prinzipiell im Vorteil sind. Beiträge aus dem Bereich der Sentiment-Analyse ohne gesonderte Berücksichtigung von Ironie zeigen, dass durch andere Arten

von künstlichen neuronalen Netzen, bspw. CNN [29], [30], [43], Recurrent CNN [27], semi-überwachte rekursive Autoencoder [47], Recursive Neural Tensor Networks [46], bessere Ergebnisse hinsichtlich Accuracy und/oder F1 erzielt werden konnten.

Bei Betrachtung der Embeddings ist auffällig, dass vor allem Wort-Embeddings Verwendung finden. Die Beiträgen [22]–[24], [31], [38], die neben Wort-Embeddings weitere Embeddings nutzen bzw. die reinen Word-Embeddings ergänzen, erzielen im Ranking grundlegend eine bessere Platzierung. Im Fall von Teilaufgabe A erreicht der Beitrag von [24] sogar den ersten Platz. Dies lässt den Schluss zu, dass die parallele Nutzung verschiedener Embeddings die Performance verbessert. Um dies genauer zu untersuchen, ist ein Vergleich verschiedener Embeddings notwendig, vgl. [22], [24], [28], [31], [32].

Eine Ergänzung durch zusätzliche manuell erstellte Features zeigt eine leichte Tendenz zur positiven Beeinflussung der Ergebnisse. Bei Betrachtung der fünf besten Beiträge für Teilaufgabe A nutzen vier der fünf Beiträge zusätzliche manuell erstellte Features, lediglich zwei der fünf schlechtesten Beiträge ergänzten die Embeddings um weitere Features. In Bezug auf Teilaufgabe B nutzten drei der fünf besten Beiträge zusätzliche manuell erstellte Features, und nur zwei der schlechtesten. Da der Einsatz zusätzlicher manuell erstellter Features lediglich eine leicht positive Tendenz hinsichtlich der Beeinflussung der Ergebnisse zeigt, ist eine Weiterverfolgung dieses Ansatzes fraglich. Bezüglich einer möglichst hohen Automatisierung des Analyseprozesses ist sogar von der Einbeziehung zusätzlicher manuell erstellter Features abzusehen, stattdessen ist eine stärkere Verfolgung der Optimierung der Embeddings zu empfehlen.

Ähnlich zur Ergänzung um zusätzliche manuell erstellte Features ist auch durch den Einsatz zusätzlicher externer Ressourcen eine leicht positive Tendenz zu erkennen. Bei Betrachtung der fünf besten Beiträge für Teilaufgabe A nutzen vier der fünf Beiträge externe Ressourcen, bei den fünf schlechtesten Beiträgen nutzt lediglich ein Beitrag zusätzliche externe Ressourcen. In Bezug auf Teilaufgabe B nutzten drei der fünf besten Beiträge zusätzliche externe Ressourcen und nur einer der schlechtesten. Eine Nutzung externer Ressourcen bedingt allerdings eine Abhängigkeit von diesen und ist daher möglichst gering zu halten. Gegebenenfalls bieten sich Alternativen an, bspw. ein Ersetzen externer Sentiment-Lexika durch eine Erweiterung des Netzes um die Erkennung des Sentiments.

Die Kombination von Netzen zeigt, betrachtet über alle Beiträge, keine generelle Performanceverbesserung oder -verschlechterung. Die Autoren, die einen solchen Ansatz verfolgten, haben innerhalb ihrer Beiträge allerdings nachweisen können, dass eine Kombination von Netzen bzw. deren Ergebnis durch Voting-Mechanismen bessere Ergebnisse erzielen. Eine generelle Ergebnisverbesserung durch die Kombination von künstlichen neuronalen Netzen zur Identifikation von Ironie ist demnach durch weitere Untersuchungen zu prüfen.

Die vorgestellten Beiträge nutzen alle den gleichen durch die Aufgabenstellung des SemEval-Workshops vorgegebenen Datensatz. Dies hat den Vorteil, dass die Ergebnisse der Beiträge miteinander vergleichbar sind.

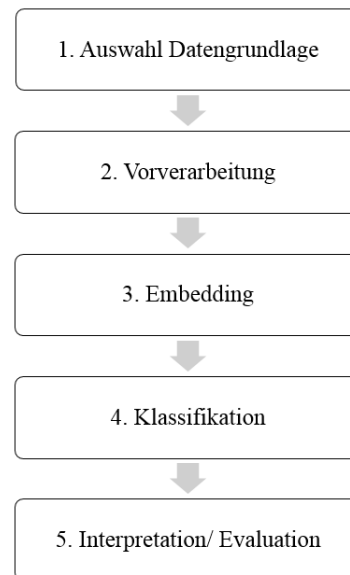


Abb. 2: Prozess der Sentiment-Analyse unter Einsatz künstlicher neuronaler Netze.

Gleichzeitig ist dadurch keine Aussage über die Übertragbarkeit auf andere Datenquellen, bspw. Zeitungsartikel oder Produktbewertungen, möglich. Twitter-Daten zeichnen sich durch ihre Kürze sowie bestimmter Informationen wie Hashtags, URLs etc. aus. Diese Daten liegen in anderen Datenquellen ggf. nicht vor. Auch hat die Länge des Textes Einfluss auf die Identifikation von Ironie. Beschränkt ein Autor seine Äußerung auf einen bis wenige Sätze, kann weniger Kontextwissen vermittelt werden und die ironische Äußerung ist konzentrierter. Bei einem längeren Text sind unter Umständen wesentlich mehr Kontextinformationen zu verarbeiten, um eine einzelne ironische Äußerung erkennen zu können. Eine Überprüfung der Ergebnisse hinsichtlich der Generalisierbarkeit bzw. Übertragbarkeit auf andere Datenquellen bzw. Datensätze ist daher notwendig.

Basierend auf den in den untersuchten Beiträgen beschriebenen Vorgehen ist eine Änderung im generellen Ablauf des Prozesses der Sentiment-Analyse zu erkennen, siehe Abb. 2. Automatisierte Embeddings ersetzen die Schritte der Transformation und insbesondere die aufwändige manuelle Feature-Selektion (vgl. Abb. 1). Die Features werden nicht mehr in zwei separaten Schritten durch manuelle Verfeinerung der Daten ermittelt, sondern durch das automatisierte Lernen von Embeddings. Die Embeddings dienen anschließend künstlichen neuronalen Netzen als Input zur Klassifikation. In Bezug auf die Vorverarbeitung der Daten zeigen die untersuchten Beiträge eine unterschiedliche Ausprägung. Mit Ausnahme von Dragoni [32] sowie Nikhil und Srivastava [35] geben die verbleibenden Autoren eine unterschiedliche Anzahl an Vorverarbeitungsschritten an, darunter die Tokenisierung, Normalisierung, Wortstammreduzierung, Lemmatisierung, Rechtschreibkorrektur, POS-Kennzeichnung sowie die Entfernung von Links, Nutzerreferenzierungen und Hashtag. Eine Aussage über den Einfluss dieser Vorverarbeitungsschritte ist anhand der Beiträge daher nicht direkt möglich, hierfür sind weitere Untersuchungen über den Einfluss auf die erzielten Ergebnisse notwendig.

## V. ZUSAMMENFASSUNG

Der vorliegende Beitrag zeigt den Stand der Forschung zur Identifikation von Ironie mittels künstlicher neuronaler Netze und insbesondere Deep Learning anhand aktueller wissenschaftlicher Beiträge zum International Workshop on Semantic Evaluation 2018 (SemEval 2018). Insgesamt 14 Beiträge wurden näher untersucht, auf deren Basis anschließend Erkenntnisse zur Nutzung künstlicher neuronaler Netze und Deep Learning zur Identifikation von Ironie sowie Verbesserungspotentiale abgeleitet wurden.

In der Gesamtheit der Beiträgen zum International Workshop on Semantic Evaluation 2018 konnten sich die Lösungsansätze, welche mit künstlichen neuronalen Netzen bzw. Deep Learning arbeiteten, gegenüber Beiträgen mit einer Fokussierung auf klassische lexikonbasierte Verfahren bzw. klassische Verfahren des maschinellen Lernens durchsetzen und damit die grundsätzliche Überlegenheit dieser Ansätze beweisen. Insgesamt bieten die erreichten Ergebnisse allerdings Verbesserungspotentiale. Zum Vergleich, im Bereich der Sentiment-Klassifikation mittels künstlicher neuronaler Netze bzw. Deep Learning konnten bereits Ergebnisse mit einer Accuracy von 95,2 % und einem F1 von 96,49% erreicht werden [27]. Eine Verbesserung und damit zuverlässigere Erkennung von Ironie im Rahmen der Sentiment-Analyse ist Ziel der angestrebten Dissertation, welches sich in der Gestaltung einer optimalen Netzarchitektur manifestiert. Gleichzeitig wird eine möglichst hohe Automatisierung, Generalisierung hinsichtlich der Datenquellen sowie Domänenunabhängigkeit angestrebt. Ansatzpunkte für die Optimierung ergeben sich aus den Implikationen aus Kapitel IV sowie den Prozessschritten aus Abb. 2. Insbesondere die Untersuchung weiterer Arten künstlicher neuronaler Netze und die Verbesserung der Embeddings werden fokussiert. In Bezug auf die Gestaltung der Embeddings ist eine stärkere Orientierung an den Charakteristiken von Ironie angestrebt, um erfolgversprechende Kombinationen von Embeddings zu identifizieren. Der Fokus liegt dabei vor allem auf dem Merkmal der Kontextabhängigkeit. Zur Erreichung der Generalisierung hinsichtlich der Datenquellen sowie einer möglichst hohen Domänenunabhängigkeit werden sowohl für das Training als auch das Testen der Netze Daten aus unterschiedlichen Datenquellen herangezogen, die ein möglichst weites Feld an unterschiedlichen Themen umfassen.

## REFERENZEN

- [1] J. James, „Data Never Sleeps 5.0“, 2017. [Online]. Verfügbar unter: <https://www.domo.com/learn/data-never-sleeps-5>. [Zugegriffen: 29-Nov-2017].
- [2] A. Montoyo, P. Martínez-Barco, und A. Balahur, „Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments“, *Decis. Support Syst.*, Bd. 53, Nr. 4, S. 675–679, 2012.
- [3] S. Stieglitz, L. Dang-Xuan, A. Bruns, und C. Neuberger, „Social Media Analytics“, *Wirtschaftsinformatik*, Bd. 56, Nr. 2, S. 101–109, Feb. 2014.
- [4] H. Hippner und R. Rentzmann, „Text Mining“, *Informatik-Spektrum*, Bd. 29, Nr. 4, S. 287–290, 2006.
- [5] R. Moraes, J. F. Valiati, und W. P. Gavião Neto, „Document-level sentiment classification: An empirical comparison between SVM and ANN“, *Expert Syst. Appl.*, Bd. 40, Nr. 2, S. 621–633, 2013.
- [6] B. Pang und L. Lee, „Opinion Mining and Sentiment Analysis“, *Found. Trends® Inf. Retr.*, Bd. 2, Nr. 1–2, S. 1–135, 2008.
- [7] B. Liu, *Sentiment Analysis and Opinion Mining*. San Rafael: Morgan & Claypool Publishers, 2012.
- [8] W. Medhat, A. Hassan, und H. Korashy, „Sentiment analysis algorithms and applications: A survey“, *Ain Shams Eng. J.*, Bd. 5, Nr. 4, S. 1093–1113, 2014.
- [9] C. Bosco, V. Patti, und A. Bolioli, „Developing Corpora for Sentiment Analysis: The Case of Irony and Senti-TUT“, *IEEE INTELLIGENT Syst.*, Bd. 28, Nr. 2, S. 55–63, 2013.
- [10] A. Reyes, P. Rosso, und T. Veale, „A multidimensional approach for detecting irony in Twitter“, *Lang. Resour. Eval.*, Bd. 47, Nr. 1, S. 239–268, 2013.
- [11] S. Sun, C. Luo, und J. Chen, „A review of natural language processing techniques for opinion mining systems“, *Inf. Fusion*, Bd. 36, S. 10–25, 2017.
- [12] Y. Lecun, Y. Bengio, und G. Hinton, „Deep learning“, *Nature*, Bd. 521, Nr. 7553, S. 436–444, 2015.
- [13] A. Reyes und P. Rosso, „On the difficulty of automatically detecting irony: beyond a simple case of negation“, *Knowl. Inf. Syst.*, Bd. 40, Nr. 3, S. 595–614, 2014.
- [14] B. C. Wallace, „Computational irony: A survey and new perspectives“, *Artif. Intell. Rev.*, Bd. 43, Nr. 4, S. 467–483, 2015.
- [15] P. Carvalho, L. Sarmento, M. J. Silva, und E. de Oliveira, „Clues for detecting irony in user-generated contents“, *Proceeding 1st Int. CIKM Work. Top. Anal. mass Opin. - TSA '09*, S. 53–56, 2009.
- [16] A. Hassan, A. Abbasi, und D. Zeng, „Twitter sentiment analysis: A bootstrap ensemble framework“, *Proc. - Soc. 2013*, S. 357–364, 2013.
- [17] A. Gianti, C. Bosco, und V. Patti, „Annotating irony in a novel italian corpus for sentiment analysis“, *Proc. 4th Int. Work. Corpora Res. Emot. Sentim. Soc. Signals*, S. 1–7, 2012.
- [18] M. Ghiassi, J. Skinner, und D. Zimbra, „Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network“, *Expert Syst. Appl.*, Bd. 40, Nr. 16, S. 6266–6282, 2013.
- [19] C. D. Manning, P. Raghavan, und H. Schütze, *Introduction to Information Retrieval*. Cambridge: Cambridge University Press, 2008.
- [20] Y. Mejova und P. Srinivasan, „Exploring Feature Definition and Selection for Sentiment Classifiers“, *Fifth Int. AAAI Conf. Weblogs Soc. Media*, S. 546–549, 2011.
- [21] Y. Yang und J. O. Pedersen, „A Comparative Study of Feature Selection in Text Categorization“, *Icml*, Bd. 97, S. 412–420, 1997.
- [22] C. Baziotis u. a., „NTUA-SLP at SemEval-2018 Task 3: Tracking Ironic Tweets using Ensembles of Word and Character Level Attentive RNNs“, in *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, 2018.
- [23] A. San, „Random Decision Syntax Trees at SemEval-2018 Task 3: LSTMs and Sentiment Scores for Irony Detection“, in *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, 2018, S. 560–564.
- [24] C. Wu, F. Wu, S. Wu, J. Liu, Z. Yuan, und Y. Huang, „THU\_NGN at SemEval-2018 Task 3: Tweet Irony Detection with Densely connected LSTM and Multi-task Learning“, in *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, 2018, S. 51–56.
- [25] M. Hartung, *Ironie in der Alltagssprache*. Radolfzell: Verlag für Gesprächsforschung, 2002.
- [26] D. Nozza, E. Fersini, und E. Messina, „A Multi-View Sentiment Corpus“, *Proc. 15th Conf. Eur. Chapter Assoc. Comput.*

*Linguist.*, Bd. 1, S. 273–280, 2017.

- [27] S. Lai, L. Xu, K. Liu, und J. Zhao, „Recurrent Convolutional Neural Networks for Text Classification“, in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI'15)*, 2015, S. 2267–2273.
- [28] E. Marrese-Taylor, S. Ilic, J. A. Balazs, Y. Matsuo, und H. Prendinger, „IIDIYT at SemEval-2018 Task 3: Irony detection in English tweets“, in *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, 2018, S. 2015–2018.
- [29] Y. Kim, „Convolutional Neural Networks for Sentence Classification“, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, S. 1746–1751.
- [30] C. N. dos Santos und M. Gatti, „Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts“, *Proc. COLING 2014, 25th Int. Conf. Comput. Linguist. Tech. Pap.*, S. 69–78, 2014.
- [31] W. I. Cho, W. H. Kang, und N. S. Kim, „HashCount at SemEval-2018 Task 3: Concatenative Featurization of Tweet and Hashtags for Irony Detection“, in *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, 2018, S. 546–552.
- [32] M. Dragoni, „NEUROSENT-PDI at SemEval-2018 Task 3: Understanding Irony in Social Networks Through a Multi-Domain Sentiment Model“, in *Proceedings of The 12th International Workshop on Semantic Evaluation*, 2018, S. 512–519.
- [33] A. Ghosh und T. Veale, „IronyMagnet at SemEval-2018 Task 3: A Siamese network for Irony detection in Social media“, in *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, 2018, S. 570–575.
- [34] J.-Á. González, L.-F. Hurtado, und F. Pla, „ELiRF-UPV at SemEval-2018 Tasks 1 and 3: Affect and Irony Detection in Tweets“, in *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, 2018, S. 565–569.
- [35] N. Nikhil und M. M. Srivastava, „Binarizer at SemEval-2018 Task 3: Parsing dependency and deep learning for irony detection“, in *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, 2018.
- [36] B. Peng, J. Wang, und X. Zhang, „YNU-HPCC at SemEval-2018 Task 3: Ensemble Neural Network Models for Irony Detection on Twitter“, in *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, 2018, S. 622–627.
- [37] S. Rajalakshmi, D. Angel, R. Milton, und T. Mimalinee, „SSN MLRG1 at SemEval-2018 Task 3: Irony Detection in English Tweets Using MultiLayer Perceptron“, in *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, 2018, S. 633–637.
- [38] H. Rangwani, D. Kulshreshtha, und A. Kumar Singh, „NLPRL-IITBHU at SemEval-2018 Task 3: Combining Linguistic Features and Emoji pre-trained CNN for Irony Detection in Tweets“, in *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, 2018, S. 638–642.
- [39] T. Vu, D. Q. Nguyen, X.-S. Vu, D. Q. Nguyen, M. Catt, und M. Trenell, „NIHRIO at SemEval-2018 Task 3: A Simple and Accurate Neural Network Model for Irony Detection in Twitter“, in *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, 2018.
- [40] Z. Yin, F. Wang, M. Lan, und W. Wang, „ECNU at SemEval-2018 Task 3: Exploration on Irony Detection from Tweets via Machine Learning and Deep Learning Methods“, in *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, 2018, S. 600–606.
- [41] C. Van Hee, E. Lefever, und V. Hoste, „Semeval-2018 task 3: Irony detection in english tweets“, in *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, 2018, S. 39–50.
- [42] J. Patterson und A. Gibson, *Deep Learning: A Practitioner's Approach*. O'Reilly Media, Inc., 2017.
- [43] D. Tang, B. Qin, und T. Liu, „Document Modeling with Gated Recurrent Neural Network for Sentiment Classification“, in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2015, S. 1422–1432.
- [44] R. Johnson und T. Zhang, „Semi-supervised Convolutional Neural Networks for Text Categorization via Region Embedding“, in *Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS 2015)*, 2015, S. 919–927.
- [45] A. Kumar u. a., „Ask Me Anything : Dynamic Memory Networks for Natural Language Processing“, in *Proceedings of the 33rd International Conference on Machine Learning*, 2016, Bd. 48, S. 1378–1387.
- [46] R. Socher u. a., „Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank“, in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 2013, S. 1631–1642.
- [47] R. Socher, J. Pennington, E. H. Huang, A. Y. Ng, und C. D. Manning, „Semi-Supervised Recursive Autoencoders for Predicting Sentiment Distributions“, in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 2011, S. 151–161.



## DISKUSSIONSBEITRÄGE ZU WIRTSCHAFTSINFORMATIK UND OPERATIONS RESEARCH

- Nr. 1 **R. Rogge**  
Internationales Kolloquium Wirtschaftsinformatik und Operations Research vom 09. bis 10. Oktober 2002 - Vorträge Teil 1. September 2003
- Nr. 2 **R. Rogge**  
Internationales Kolloquium Wirtschaftsinformatik und Operations Research vom 09. bis 10. Oktober 2002 - Vorträge Teil 2. September 2003
- Nr. 3 **R. Maier / T. Hädrich**  
Szenarien für Wissensmanagement und E-Learning in Forschung und Lehre. Juni 2004
- Nr. 4 **R. Maier / M. Trögl**  
Tagungsband des fünften interuniversitären Doktorandenseminars Wirtschaftsinformatik der Universitäten Halle(Saale), Jena und Leipzig. Dezember 2004
- Nr. 5 **R. Maier / M. Trögl**  
Wissensmanagementsysteme für Wissenskoooperationen. Mai 2005
- Nr. 6 **F. Kretzschmar / I. Reitzenstein / M. Schramm**  
Ajax-Technologien in Model2-Webanwendungen. November 2006
- Nr. 7 **F. Kretzschmar**  
The economic perspective on Service-Oriented Architectures. November 2006
- Nr. 8 **I. Reitzenstein**  
Entwurf einer Multi-Agenten-Simulation zur Analyse der Robustheit von Reputationssystemen. November 2006
- Nr. 9 **R. Maier / M. Trögl**  
Aktive Dokumente. November 2006
- Nr. 10 **R. Rogge / M. Steglich**  
Betriebswirtschaftliche Entscheidungsmodelle zur Verfahrenswahl sowie Auflagen- und Lagerpolitiken. Januar 2007
- Nr. 11 **A. Linke / R. Peters / S. Schäfer**  
PrintIT2006: E-Business in der Druck- und Medienindustrie - Eine Studie zum Status quo. 2007
- Nr. 12 **R. Peters**  
Agentenbasierte Erweiterungen einfacher Auktionen. 2007
- Nr.13 **F. Kretzschmar / R. Peters**  
Koordination von SOA-basierten Wertschöpfungsketten - Ein vergleichendes Simulationsexperiment. Juni 2007

- Nr. 14 **M. Backhaus / R. Rogge**  
Agieren statt Reagieren - Wege zu einem aktiven Management medialer Nutzungsrechte im Rundfunk. September 2007
- Nr. 15 **R. Peters**  
Tagungsband des achten interuniversitären Doktorandenseminars Wirtschaftsinformatik der Universitäten Halle, Jena und Leipzig. Dezember 2007
- Nr. 16 **I. Reitzenstein / R. Peters**  
Robuste Reputationssysteme für Elektronische Märkte - ein evolutionärer Bewertungsansatz. Dezember 2007
- Nr. 17 **I. Reitzenstein / R. Peters**  
Vertrauen im E-Commerce - Ein Überblick zu Reputationssystemen. Februar 2008
- Nr. 18 **M. Steglich / T. Mellouli / R. Rogge**  
Produktionsprogrammplanung unter Einbeziehung von Preis-Absatz-Funktionen und Lernkurveneffekten. Juni 2008
- Nr. 19 **T. Mellouli**  
Scheduling and Routing Processes in Public Transport Systems – Modeling, Optimization, and Decision Support
- Nr. 20 **R. Peters / I. Reitzenstein**  
Untersuchung der Wechselwirkung zwischen Manipulationen und Lösungskonzepten für Reputationssysteme. August 2008
- Nr. 21 **M. Steglich / T. Schleiff / R. Rogge / T. Mellouli**  
Eine Fallstudie zur Produktionsprogrammplanung unter Einbeziehung von Preis-Absatz-Funktionen und Lernkurveneffekten unter Verwendung der Modellierungssprache CMPL. Juli 2009
- Nr. 22 **T. Mellouli**  
Tagungsband des elften interuniversitären Doktorandenseminars Wirtschaftsinformatik der Universitäten Halle, Leipzig, Jena, Dresden und Freiberg. Dezember 2009
- Nr. 23 **T. Mellouli / M. Steglich**  
Beiträge zum Festkolloquium anlässlich des 65. Geburtstages von Prof. Dr. Rolf Rogge – Teil 1: Anwendungen der mathematischen Optimierung und Entscheidungsunterstützung. Oktober 2010
- Nr. 24 **T. Wöhner / S. Köhler / R. Peters**  
Metrics for Automatic Reputation Assessment in Wikipedia - A Comparative Evaluation. 2011
- Nr. 25 **T. Wöhner / S. Köhler / R. Peters**  
Automatic Editing Rights Management in Wikipedia. 2011
- Nr. 26 **K. Kittel / S. Sackmann**  
Gaining Flexibility and Compliance in Rescue Processes with BPM. 2011

- Nr. 27 **C. Schmeißer / R. Peters**  
Shilling Identification in Online Auctions. 2011
- Nr. 28 **R. Peters / C. Schmeißer**  
Sniping in Online Auctions. 2011
- Nr. 29 **T. Mellouli / M. Steglich**  
Beiträge zum Festkolloquium "Angewandte Optimierung" anlässlich des 65. Geburtstages von Prof. Dr. Rolf Rogge
- Nr. 30 **S. Sackmann / M. Hofmann**  
Beiträge zum 16. Interuniversitären Doktorandenseminar Wirtschaftsinformatik (IDS 2012)
- Nr. 31 **U. Brettschneider / R. Peters**  
Konzeption eines Business Intelligence Systems zur Stimmungsanalyse in Web 2.0 Applikationen
- Nr. 32 **M. Römer**  
Airline Crew Planning Processes: A Hierarchical Planning Perspective
- Nr. 33 **M. Römer**  
Optimization in Airline Crew Planning and Scheduling: State of the Art and Generic Network Flow-Based Approaches
- Nr. 34 **M. Römer**  
Case Studies in Medium- and Short-Term Airline Crew Planning: Establishing Consistency in Optimization Approaches

## WIRTSCHAFTSWISSENSCHAFTLICHE DISKUSSIONSBEITRÄGE

- Nr. 1 **R. K. von Weizsäcker**  
Staatsverschuldung und Demokratie. April 1992
- Nr. 2 **M. Becker**  
Theoretische und ideengeschichtliche Ansätze zur Begründung der Personalwirtschaftslehre als betriebswirtschaftliche (Teil)Disziplin. Februar 1995
- Nr. 3 **M. Nerlove / A. Razin / E. Sadka / R.K. von Weizsäcker**  
Comprehensive Income Taxation, Investments in Human and Physical Capital, and Productivity. Mai 1992
- Nr. 4 **R. K. von Weizsäcker**  
Bildung und Theorie der Lebenseinkommensverteilung. Mai 1992
- Nr. 5 **V. Meier**  
Long-Run Migration Incentives and Migration Effects: The Case of Different Fertility Rates. Juni 1993

- Nr. 6 **R. K. von Weizsäcker**  
Public Pension Reform, Demographics, and Inequality. Dezember 1993
- Nr. 7 **G. Steinmann**  
Zusammenhang zwischen Alterungsprozeß und Einwanderung. Dezember 1993
- Nr. 8 **G. Schmitt-Rink**  
Opimal Rates of Migration or The Goldenest Golden Rule of Economic Growth. Dezember 1993
- Nr. 9 **V. Meier**  
On the Demand for Long-Term Care Insurence. Juli 1994
- Nr. 10 **V. Meier**  
Altruism and the Demand for Long-Term Care Insurance. Juli 1994
- Nr. 11 **M. Becker**  
Wertende oder werturteilsfreie Betriebswirtschaftslehre? Aktuelle Anmerkungen zu einer alten Streitfrage. September 1994
- Nr. 12 **B. Wigger**  
Human Capital and International Patterns of Economic Growth. Oktober 1994
- Nr. 13 **B. Wigger**  
Bevölkerungswandel, Alterssicherung und individuelle Wohlfahrt. Dezember 1994
- Nr. 14 **A. Brüggemann / M. Klein**  
Privatization and Foreign Direct Investment in Transition Economics. Februar 1995
- Nr. 15 **V. Meier**  
Long-Term Care Insurance and Moral Hazard. Februar 1995
- Nr. 16 **V. Meier**  
Conflicts on care for the elderly, long-term care insurance, and bewuests. Mai 1995
- Nr. 17 **U. Eiteljörge**  
Das Abkommen über den internationalen Handel mit Dienstleistungen - Eine erste Bewertung. Juni 1995
- Nr. 18 **U. Eiteljörge / M. Klein**  
Size, Flexibility and Power in an Integrated World Economy. August 1995
- Nr. 19 **A. Brüggemann / M. Klein**  
Ist die Treuhand ein Modell für die Transformationsländer? Einige skeptische Anmerkungen. Januar 1996
- Nr. 20 **M. Klein**  
WTO, IMF, IBRD usw.: Koordinationsprobleme zwischen internationalen Wirtschaftsorganisationen. Dezember 1995

- Nr. 21 **V. Meier**  
Long-Term Care Insurance and Savings. Februar 1996
- Nr. 22 **U. Eiteljörge / M.Klein**  
Trade in Know-how, Direct Investment and Welfare: The Duopoly Case. Mai 1996
- Nr. 23 **A. Brüggemann**  
Social safety nets in Central and Eastern Europe: Are they efficiency enhancing? Mai 1996
- Nr. 24 **V. Meier**  
The Demand for preventive dental care. Juni 1996
- Nr. 25 **G. Steinmann / A. Prskawetz / G. Feichtinger**  
A Model on the Escape from the Malthusian Trap. August 1996

## BETRIEBSWIRTSCHAFTLICHE DISKUSSIONSBEITRÄGE

- Nr. 1 **J. Merker / G. Wäscher**  
Two New Heuristic Algorithms for the Maximal Planar Layout Problem. März 1996
- Nr. 2 **J. Kloock / S. Dierkes**  
Prozeßkostenkontrolle. März 1996
- Nr. 3 **H. Foerster / G. Wäscher**  
Simulated Annealing for the Order Spread Minimization Problem in Sequencing Cutting Patterns. Mai 1996
- Nr. 4 **J. Kloock / U. Schiller**  
Cost Budgeting and Cost Control with Marginal Costing. August 1996
- Nr. 5 **G. H. Lawson / H. P. Möller**  
The cash flow effect of retained earnings measured on a conservative basis. Oktober 1996
- Nr. 6 **M. Becker**  
Geändertes Karriereverständnis: Personalentwicklung im Zeichen von Führungs-, Fach- und Projektkarrieren. November 1996
- Nr. 7 **P. Heise / G. Wäscher**  
The Bin-Packing Problem: A Problem Generator and Some Numerical Experiments with FFD Packing and MTP. November 1996
- Nr. 8 **D. Gramlich**  
Cross Risks im liquiditätsmäßig-finanziellen Bereich von Kreditinstituten. Dezember 1996
- Nr. 9 **D. Möhlenbruch / C. Meier**  
Stand und Entwicklungsmöglichkeiten eines integrierten Controllingsystems für den Einzelhandel. Dezember 1996

- Nr. 10 **M. Becker**  
Personalentwicklung und Organisationsentwicklung als Führungsaufgabe - eine Einführung in die Thematik - (2.Auflage). Februar 1997
- Nr. 11 **M. Becker**  
Krankheitsbedingte Fehlzeiten in ostdeutschen und westdeutschen Unternehmen - Eine Untersuchung zur Ursachenanalyse und Maßnahmenplanung. Februar 1997
- Nr. 12 **V. Schwarz / G. Rother**  
Arbeitszeitgestaltung in den neuen und alten Bundesländern unter mikropolitischer Betrachtung - Eine Analyse von Betriebsvereinbarungen. Februar 1997
- Nr. 13 **S. Dierkes**  
Konzeption und Aufbau differenziert-mehrstufiger Fixkostendeckungsrechnungen. März 1997
- Nr. 14 **G. Wäscher / J. Heuer / V. Reschke / P. Schwerin**  
Local Search-Verfahren für ein Traveling Salesman- Problem in der Produktionssteuerung. April 1997
- Nr. 15 **K. Jank / G. Wäscher**  
Mischungsoptimierung in der Wurstwarenherstellung - Eine Fallstudie. Mai 1997
- Nr. 16 **M. Becker**  
Vom Objektbezug zur Subjektorientierung in der betrieblichen Weiterbildung. Mai 1997
- Nr. 17 **M. Becker**  
Unternehmensinformation und Personalentwicklung. Oktober 1997
- Nr. 18 **D. Gramlich**  
Modellgestützte Analyse bankbetrieblicher Cross Risks - Ein Beitrag zur Theorie des Risikoverbundes bei Finanzmärkten. November 1997
- Nr. 19 **B. Siebenhüner / R. Antes / H.-U. Zabel**  
Prevention Through Process Policy and Product Policy - A Study of European and German Environmental Policy. Dezember 1997
- Nr. 20 **S. Tamm**  
Ranking joint stock companies: an analysis of different multi-criteria methods. Dezember 1997
- Nr. 21 **R. M. Ebeling / H. Kuschel**  
Auswirkungen der dritten Stufe der Europäischen Währungsunion auf die "Währungsumrechnung" im Konzernabschluß. März 1998
- Nr. 22 **M. Becker / G. Rother**  
Kompetenzentwicklung. April 1998
- Nr. 23 **M. Becker**  
Flow statt Frust. Mai 1998

- Nr. 24 **M. Becker**  
Gruppenarbeit - Theoretische Grundlagen und Evaluierung an ausgewählten Praxisbeispielen. Mai 1998
- Nr. 25 **A. Sapusek**  
Fundamental Performance of Initial Public Offerings: Empirical Evidence from Germany. Juni 1998
- Nr. 26 **P. Schwerin / G. Wäscher**  
A New Lower Bound for then Bin- Packing Probleme and its Integration into MTP. Juni 1998
- Nr. 27 **F. Schneider**  
Bankbetriebliches Preismanagement - Ergebnisse einer Umfrage bei Filial- und Direktbanken. August 1998
- Nr. 28 **R. M. Ebeling / K. F. Baumann**  
Konsolidierung mehrstufiger Konzerne nach der Methode der Integrierten Konsolidierungstechnik. Januar 1999
- Nr. 29 **H. Foerster / G. Wäscher**  
Pattern Reduction in One-dimensional Cutting Stock Problems. April 1999
- Nr. 31 **H.-U. Zabel**  
Verhaltensmodell einer Ökologischen Ökonomik. Dezember 1999
- Nr. 32 **S. Dierkes**  
Erlöscontrolling im Monopol und Oligopol. Dezember 1999
- Nr. 33 **K. Renger**  
Effektivzinsberechnung nach den neuen EU-Regelungen. Dezember 1999
- Nr. 34 **N. Ehrentreich / R. Schmidt**  
The German Corporate Governance System with Special Respect to Innovation. Dezember 1999
- Nr. 35 **S. Dierkes / S. Hanrath**  
Ermittlung und Entscheidungsrelevanz kalkulatorischer Zinskosten für das Anlagevermögen bei nutzungsabhängigem Verschleiß - ein investitionstheoretischer Lösungsansatz. Dezember 1999
- Nr. 36 **R. Schmidt**  
Cross-Border Mergers and Corporate Governance - An Empirical Analysis from 1988 to 1999. Dezember 1999
- Nr. 37 **A. Sapusek**  
Comovement and Divergence of European Growth Markets: The Case of EURO.NM Markets. Dezember 1999
- Nr. 38 **D. Möhlenbruch / B. Claus / U.-M. Schmieder**  
Corporate Identity, Corporate Image und Integrierte Kommunikation als Problembereich des Marketing. Juli 2000

- Nr. 39 **R. Antes**  
Die Flußkostenrechnung als Instrument betrieblicher Nachhaltigkeit? August 2000
- Nr. 40 **V. Reschke / G. Wäscher**  
Local Search-Verfahren für die Lagerplatzvergabe in Mann-zur-Ware-Kommissioniersystemen. Oktober 2000
- Nr. 41 **M. Becker**  
Sind Wohltäter bessere Menschen? Oktober 2000
- Nr. 42 **B. Maczulaitis / R. Schmidt**  
Corporate Governance and Product Innovation at Scherting AG: A Case Study. Dezember 2001
- Nr. 43 **N. Ehrentreich / R. Schmidt**  
Situation and Trends in the German Corporate Governance System. Mai 2002
- Nr. 44 **L. Schiefner**  
Risk-Minimizing Hedging of General Cash Flows in Discrete Time. September 2002
- Nr. 45 **N. Ehrentreich**  
The Santa Fe Artificial Stock Market Re- Examined - Suggested Corretions. September 2002
- Nr. 46 **D. Schneider**  
Junger Wein in alten Schläuchen: der frühen Universität Halles. Beitrag zum Gestaltwandel des ökonomischen Denkens. September 2002
- Nr. 47 **L. Schiefner / R. Schmidt**  
Shareholder Value at Risk: Concept for Company Valuation, Implementation, and Simulation Example. April 2003
- Nr. 48 **E. Altay**  
The Effect of Macroeconomic Factors on Asset Returns: A Comparative Analysis of the German and the Turkish Stock Markets in an APT Framework. Juni 2003
- Nr. 49 **E. Altay**  
Cross-Autocorrelation between Small and Large Cap Portfolios in the German and Turkish Stock Markets. August 2003
- Nr. 50 **M. Ecke**  
Empirische Untersuchungen zu Kursaussetzungen am deutschen Aktienmarkt. August 2003
- Nr. 51 **A. Sapusek**  
Floor versus Screen Trading: Evidence from Neuer Markt. Oktober 2003
- Nr. 52 **M. Becker**  
Statt Ausbildung wieder Lehrgeld zahlen? Nur eine radikale Änderung der Denkhaltung bringt mehr Ausbildungsplätze. Oktober 2003



- Nr. 53 **M. Becker**  
Von der Taube auf dem Dach und dem Spatzen in der Hand - Anmerkungen zu Akzeptanz, Gestaltung und Motivationswirkung variabler Entlohnung. Dezember 2003
- Nr. 54 **R. Krengel / G. Kraft**  
Mindestbesteuerung - Eine systematische und ökonomische Analyse. Juli 2004
- Nr. 55 **M. Becker**  
Strategisch orientierte Auswahl und Qualifizierung von Nachwuchskräften. September 2004
- Nr. 56 **M. Becker**  
Die Organisation der Personalentwicklung als Shared Service Center. November 2004
- Nr. 57 **M. Becker**  
Die neue Rolle der Personalentwicklung. Dezember 2004
- Nr. 58 **M. Becker**  
Flow statt Frust - Der Weg zum motivierten und leistungsstarken Mitarbeiter. April 2006
- Nr. 59 **C. Bierwirth / F. Meisel**  
A fast Heuristic for Quay Crane Scheduling with Interference Constraints. November 2006
- Nr. 60 **J. Bron**  
Das G-REIT-Gesetz - eine Analyse auf Basis des Gesetzentwurfes der Bundesregierung. Dezember 2006
- Nr. 61 **M. Becker**  
Das allgemeine Gleichbehandlungsgesetz (AGG) erzwingt Diversity Management. Eine Analyse des AGG und personalwirtschaftliche Gestaltungshinweise. Januar 2007
- Nr. 62 **H.-Ulrich Zabel**  
Umweltmanagement auf dem Weg zur Nachhaltigkeit - 10 Jahre "Betriebliches Umweltmanagement" an der Martin-Luther-Universität Halle-Wittenberg. Januar 2007
- Nr. 63 **S. Berger / C. Bierwirth**  
The Collaborative Carrier Vehicle Routing Problem: Formulation and Solution Concepts. März 2007
- Nr. 64 **L. Balsler / M. Wappler / C. Weiser**  
Controlling von Shared Services Center. Juli 2007
- Nr. 65 **H.-U. Zabel**  
Egoismusannahme auf dem Prüfstand - ein Thesenpapier. Dezember 2007
- Nr. 66 **D. Möhlenbruch / M. Arnold / A. Wolf**  
Die Führung von Premiumhandelsmarken im Lebensmitteleinzelhandel - Eine informationsökonomische Analyse. Januar 2008
- Nr. 67 **D. Möhlenbruch / S. Möller / F. Ritschel**  
Bundling auf Grundlage der adaptiven Conjoint-Analyse - Bedeutung und Vorgehensweise im Bereich der Autovermietung. Januar 2008

- Nr. 68 **M. Becker / C. Köllen / S. Maihöfner**  
Wenn der Wirt das Getränk nicht mehr selbst bezahlt! Analyse und Aspekte einer nachfragerfinanzierten Hochschulbildung in Deutschland. Februar 2008
- Nr. 69 **D. Möhlenbruch**  
Kundenzufriedenheit und Beziehungsqualität im Einzelhandel - Inhalte und Messung. März 2008
- Nr. 70 **M. Becker**  
Wenn Forschung auf Realität trifft! Forschen, Erkennen und Gestalten als grundlegende Aspekte wissenschaftlichen Arbeitens. April 2008
- Nr. 71 **M. Becker / I. Labucay**  
Nach Maß geschneidert oder von der Stange gekauft? Entwicklungstendenzen und Gestaltungsempfehlungen zur Personalarbeit als "Fabrik" und "Manufaktur". April 2008
- Nr. 72 **M. Becker**  
No slack - bad slack - good slack. April 2008
- Nr. 73 **M. W. Fütting / T. B. Müller / M. Wappler / C. Weiser**  
Der Realloptionsansatz als Instrument zur Steuerung und Bewertung von F&E-Projekten. Mai 2008
- Nr. 74 **D. Möhlenbruch / G. von Wichert**  
Inhalte und Bedeutung des Defender-Modells zur Bestimmung der Defensivstrategie bei Markteintritten von Produktinnovationen. Juni 2008
- Nr. 75 **M. Becker / C. Rieger**  
Ein Fingerhut Auswahl ist besser als ein Scheffel Training. Dezember 2008
- Nr. 76 **H.-U. Zabel**  
Technologiewahl für Kreislaufwirtschaft - ein Kostenmodell
- Nr. 77 **Hans-Ulrich Zabel**  
Nachhaltige Nutzung der Elbregion - Konturen einer Gesamtkonzeption
- Nr. 78 **D. Möhlenbruch / F. Ritschel / F. Sinning**  
Einsatzmöglichkeiten des Customer Profiling im Kundenrückgewinnungsmanagement des E-Retailing
- Nr. 79 **F. Meisel**  
Operational Planning and Scheduling in Seaport Container Terminals
- Nr. 81 **F. Meisel / C. Bierwirth**  
Characterization and Generation of Benchmark Instances for Quay Crane Scheduling
- Nr. 82 **M. Becker / M. Kirchner / I. Labucay**  
Altersgerechte Weiterbildung in Unternehmen "Wer aufhört, fällt zurück"
- Nr. 83 **M. Becker / B. Niggeloh**  
Leistungsfähige Entlohnung auf der Grundlage von Stellenbündeln

- Nr. 84 **D. Möhlenbruch / D. E. Blunck / F. Ritschel**  
Einsatzmöglichkeiten von Web 2.0 Instrumenten im Multi-Channel- Marketing des Einzelhandels
- Nr. 85 **M. Becker / C. Kownatka**  
DIM-PRAX Deutschland 2010 Blitzumfrage - Erhebung der Diversity Management Praxis
- Nr. 86 **F. Meisel / C. Bierwirth**  
Integrated Berth Allocation and Crane Operations Planning in Seaport Container Terminals
- Nr. 87 **D. Möhlenbruch / S. Dölling / F. Huhn**  
Einflusspotentiale nonverbaler Kommunikation im stationären Einzelhandel - Darstellung und Kritik
- Nr. 88 **C. Bierwirth / A. Goel**  
Lenk- und Ruhezeitenregelungen im Straßengütertransport
- Nr. 89 **M. Becker / M. Kirchner**  
Dynamixity. Dynamik - Komplexität – Unsicherheit
- Nr. 90 **F. Meisel**  
Ansätze zur Gestaltung und zum Betrieb von Produktions-, Distributions- und Dienstleistungsnetzwerken
- Nr. 91 **D. Möhlenbruch / S. Dölling / N. Rotsch**  
Das Verkaufspersonal als Determinante der Einkaufsstättenwahl unter Berücksichtigung des Gendermarketing
- Nr. 92 **D. Möhlenbruch / A. Kohlmann / A. Schönborn**  
Das Internet als Vertriebskanal des Lebensmitteleinzelhandels - Gestaltungsmöglichkeiten und kritische Analyse
- Nr. 93 **J. K. Zimbelmann / C. Bierwirth**  
A Literature Survey and Classification of Humanitarian Logistics Research

#### VOLKSWIRTSCHAFTLICHE DISKUSSIONSBEITRÄGE

- Nr. 1 **H. Tietmeyer**  
Finanzmärkte und Beschaffung. Juli 1996
- Nr. 2 **V. Meier**  
On the Choice Between Capital and Labor Mobility: The Small Country Case. September 1996
- Nr. 3 **V. Meier**  
Long-Term Care Insurance and Life Insurance Demand. Oktober 1996

- Nr. 4 **M. Jäger / G. Steinmann**  
Morality and Economic Growth. Januar 1997
- Nr. 5 **V. Meier**  
Health Insurance and Preventive Behavior. Januar 1997
- Nr. 6 **A. Brüggemann**  
Enterprise Arrears in the Transformation - A Comparison of the Polish and East German Experiences. Januar 1997
- Nr. 7 **A. Brüggemann**  
Soziale Absicherung und Privatisierung durch Direktinvestitionen. Januar 1997
- Nr. 8 **Chr. Lippert / H. Ahrens / M. Rittershofer**  
The Significance of Institutions for the Design and Formation of Agro-Environmental Policy. März 1997
- Nr. 9 **U. Neyer**  
Unterinvestition als Folge asymmetrisch verteilter Information unter Berücksichtigung risikoaversen Verhaltens der Kreditinstitute - Implikationen für Ostdeutschland. März 1997
- Nr. 10 **D. Gläser / G. Schmidt / E. Harz / S. Knauth / Chr. Schäfer**  
Quo vadis ? Auswertung der Ergebnisse der Befragung der Absolventen der Wirtschaftswissenschaftlichen Fakultät im Beobachtungszeitraum Wintersemester 1993/94 bis Wintersemester 1995/96. März 1997
- Nr. 11 **G. Steinmann / M. Jäger**  
How many Immigrants can a Society integrate? August 1997
- Nr. 12 **V. Meier**  
Why the Young Do Not Buy Long-Term Care Insurance. Dezember 1997
- Nr. 13 **I. Meyne**  
International Environment Agreements as Two-Level Games. Dezember 1997
- Nr. 14 **O. Fabel / V. Meier**  
Optimal Parole Decisions. Februar 1998
- Nr. 15 **V. Meier**  
Time preference, international migration, and social security. April 1998
- Nr. 16 **U. Eiteljörge / M. Klein**  
International Know-how Trade and Foreign Direct Investment. Mai 1998
- Nr. 17 **A. Bäcker**  
Eine politökonomische Sicht der Schuldenkapazität von Emerging Economies. Dezember 1998

- Nr. 18 **D. Gläser / G. Schmidt / E. Harz / T. Keitzl / U. Kramer / D. Meißner**  
Cum grano salis. Evaluation der Lehre: Befragung der AbsolventInnen der  
Wirtschaftswissenschaftlichen Fakultät der Martin-Luther-Universität Halle-Wittenberg.  
Januar 1999
- Nr. 19 **V. Meier**  
Economic Consequences of the Posted Workers Directive. April 1999
- Nr. 20 **I. Meyne / A. Wagener**  
Which Game Do We Play? On The Role of the Strategic Variable in Environmental  
Negotiations. Mai 1999
- Nr. 21 **V. Meier**  
Prison, Teherapy, and Parole - An Economic Analysis. Juli 1999
- Nr. 22 **V. Meier**  
Can Higher Wages for Foreign Workers Increase the Welfare of the Natives? Juli 1999
- Nr. 23 **V. Meier**  
Economic theories of education. Januar 2000
- Nr. 24 **G. Nekkers**  
A Regional Insider-Outsider Model With Cooperation and Harassment. Januar 2000
- Nr. 25 **U. Neyer**  
Under- and Overinvestment as a Consequences of Informational Asymmetries -  
Implications for the Transition in Eastern Germany. Mai 2000
- Nr. 26 **J. Wilde**  
A Useful Generalization of Amemiya's Generalized Least Squares Estimator. Januar 2001
- Nr. 27 **F. Bal / G. Nekkers**  
Knowledge Accumulation and Processing via Study Findings: Methodological Issues.  
Januar 2001
- Nr. 28 **U. Neyer**  
Asymmetric Information in Credit Markets and Monetary Policy. Mai 2001
- Nr. 29 **U. Neyer**  
Bank's Behaviour in the Interbank Market and the Operational Framework of the  
Eurosystem. November 2002
- Nr. 30 **J. Wilde**  
Was reizt Sozialhilfeempfänger zum Ausstieg? Eine empirische Untersuchung mit dem  
Niedrigeinkommens-Panel. November 2002
- Nr. 31 **J. Wilde / A. Kubis**  
Nichtinanspruchnahme von Sozialhilfe - Eine empirische Analyse des Unerwarteten. März  
2004

- Nr. 32 **R. Klein**  
Medical Savings Accounts or Deductible. April 2004
- Nr. 33 **M. Jäger**  
The Economic Consequences of Immigration and why a Neoclassical Model will Fail.  
November 2004
- Nr. 34 **M. Jäger / G. Steinmann**  
A Vintage Model to Production and Consumption. November 2004
- Nr. 35 **G. Steinmann**  
Vorschläge für eine nachhaltige Familienpolitik. November 2004
- Nr. 36 **M. Ahlert**  
Aspiration Balancing Agreement - A New Axiomatic Approach to Bounded rationality in  
Negotiations. Dezember 2004
- Nr. 37 **S. Tagge**  
An Alternative Approach to a child-related Pension System. Dezember 2004
- Nr. 38 **M. Ahlert**  
Public and Private Choices in Organ Donation. Dezember 2004
- Nr. 39 **G. Steinmann**  
Die individuelle Rationalität: Kindermangel als Ergebnis der elterlichen Entscheidung über  
die Kinderzahl. Januar 2005
- Nr. 40 **U. Neyer**  
Flexibility of Monetary Policy in the Euro Area and Remuneration of Required Reserves.  
Februar 2005
- Nr. 41 **G. Steinmann**  
Kindermangel als Ursache ökonomischer und sozialer Probleme. Die Auswirkungen auf  
die sozialen Sicherungssysteme. April 2005
- Nr. 42 **A. Kubis**  
Sectoral Movement as an Incentive for Interregional Migration. Mai 2005
- Nr. 43 **H. Gude**  
Sortimentserweiterung als Reaktion auf den Markteintritt eines Billig-Anbieters.  
Mai 2005.
- Nr. 44 **H. Gude**  
Vorteile durch Produkt-Differenzierung: Zahlungsbereitschaft und Präferenzen der  
Konsumenten als Marktzutrittsbarriere. September 2005
- Nr. 45 **H. Gude**  
Über die Vorteilhaftigkeit einer temporären Duldung von Wettbewerbern.  
September 2005

- Nr. 46 **F. Bal / J. M. Vleugel**  
Opening the Black Box of Ceteris Paribus Clause. November 2005
- Nr. 47 **H. Gude**  
Warum Unternehmen von hoher Wettbewerbsintensität profitieren können: Der Fall differenzierter Produkte. Januar 2006
- Nr. 48 **M. Ahlert**  
A New Approach to Procedural Freedom in Game Forms. February 2006
- Nr. 49 **M. Ahlert**  
Guarantees in Game Forms. Februar 2006
- Nr. 50 **M. Ahlert**  
Discrete Allocation of a Divisible Good – Allocation of Chances. März 2006
- Nr. 51 **M. Ahlert / H. Kliemt**  
Necessary and Sufficient Conditions to Make Numbers Count. März 2006
- Nr. 52 **F. Heyde / U. Neyer**  
The Stability of the Banking Sector and Credit Default Swaps. February 2007
- Nr. 53 **M. Ahlert**  
If not only Numbers Count – Allocation of Equal Chances. April 2007
- Nr. 54 **M. Klein**  
Ein Input-Output-Ansatz für die Außenhandelstheorie. August 2007
- Nr. 55 **M. Klein / M. Henkel / L. Möller**  
Produktdifferenzierung und Markteintritt an Elektrizitätsmärkten: Ein Realloptionsansatz. September 2007
- Nr. 56 **A. Saalbach**  
Leistungen und Probleme der Volksrepublik China bei der Herstellung von WTO-Konformität. November 2007
- Nr. 57 **D. M. Ngui / C. Becker / W. Thomi**  
Firm efficiency differences during the structural reform period: Empirical Evidence from the Kenyan Manufacturing firms. December 2007
- Nr. 58 **D. M. Ngui / C. Becker / W. Thomi**  
Corruption, Insecurity and technical Efficiency: Evidence from Kenyan Manufacturing firms. January 2008
- Nr. 59 **L. Schwettnann**  
The Acceptance of Truncated Efficiency. November 2008
- Nr. 60 **R. Bertenrath / M. Demary / M. Jäger / M. Schattenberg**  
Aggregierter Konsum und Kreditrationierung in Deutschland. November 2008.

- Nr. 61 **E.-O. von Ledebur / J. Schmitz**  
Corn Price Behavior – Price Transmission during the Boom on Futures Markets. May 2009.
- Nr. 62 **T. Weirowski**  
Der Transmissionsmechanismus der Geldpolitik in regionaler Perspektive - Eine Analyse für West- und Ostdeutschland. Juni 2009
- Nr. 63 **P. Bönisch / S. Tagge**  
Estimating a Cost Function of German Child Care Centers. October 2009
- Nr. 64 **M. Klein**  
Financial Programming aus diskurstheoretischer Perspektive: Eine Ideenskizze zur wirtschaftspolitischen Arbeit des IWF. Dezember 2009
- Nr. 65 **T. Faxin**  
Sind Chinas Flitterwochen mit dem Dollar vorbei? Dezember 2009
- Nr. 66 **P. Bönisch / L. Schneider**  
Institutional transition, social capital mix and local ties – Does Communist legacy explain low labour mobility? February 2010
- Nr. 67 **P. Bönisch / P. Gaffert / J. Wilde**  
The Impact of Skills on Remigration Flows. March 2010
- Nr. 68 **A. Illy / M. Schwartz / C. Hornyh / M. T. W. Rosenfeld**  
Specialization diversity competition and their impact on local economic growth in Germany. March 2010
- Nr. 69 **F. Bal / J. M. Vleugel**  
On composite valuation techniques for man-made natural areas. February 2011
- Nr. 70 **M. Klein / T. Weirowski**  
Trade and Unemployment in Germany: An Empirical Exploration and Some Theory. August 2011
- Nr. 71 **M. Ahlert / L. Schwettmann**  
Ergebnisse einer Befragung zur Studienzufriedenheit im Fachbereich Wirtschaftswissenschaften. Dezember 2011
- Nr. 72 **M. Ahlert / I. Lajtos**  
60 Years after Nash's Bargaining Solution: Trends in Bargaining Theory. Dezember 2011
- Nr. 73 **W. H. Grieben / F. Sener**  
North-South Trade, Unemployment and Growth: What's the Role of Labor Unions? Dezember 2013
- Nr. 74 **S. Deole**  
The Rise of the Commons: The Post-Independence Socio-Economic Transition of Indian Society. Dezember 2013



- Nr. 75 **M. Ahlert**  
A Conceptual Model of Desire-Based Choice. November 2014
- Nr. 76 **E. Dinopoulos / W.-H. Grieben / F. Sener**  
The Conundrum of Recovery Policies: Growth or Jobs? Dezember 2014
- Nr. 77 **M. Ahlert / L. Schwettmann**  
Equity of What When Allocating Health Care Resources? Februar 2015
- Nr. 78 **L. Schwettmann**  
The (Difficult) Interdependence between Empirical and Normative Research: Empirical Social Choice and the Fair Distribution of Health Care Resources. Juni 2015
- Nr. 79 **W. Gaertner / L. Schwettmann**  
Burden Sharing in Deficit Countries: A Questionnaire-Experimental Investigation.  
September 2015
- Nr. 80 **H. Albers / C. Gornott / S. Hüttel**  
How Do Inputs and Weather Drive Wheat Yield Volatility? The Example of Germany. April 2016





