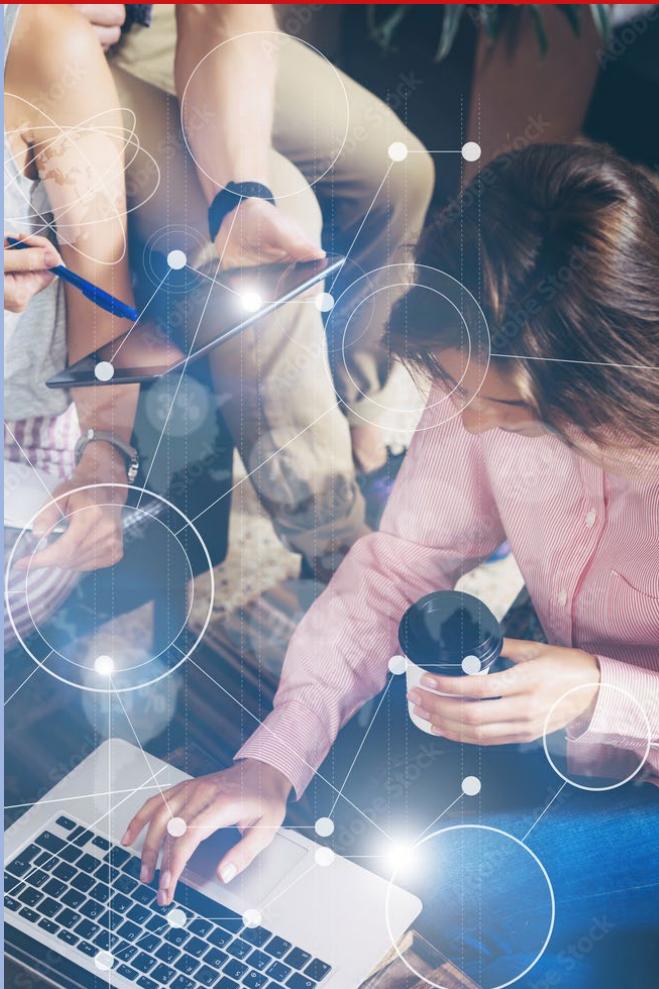


Angewandte Forschung in der Wirtschaftsinformatik 2022

Herausgeber*innen

Sandy Eggert
Claudia Lemke
Verena Majunke
Birte Malzahn
Vera Meister
Katharina Simbeck
Christian Czarnecki
Martin Wolf



Angewandte Forschung in der Wirtschaftsinformatik 2022

Tagungsband zur 35. Jahrestagung
des Arbeitskreises Wirtschaftsinformatik an Hochschulen für
Angewandte Wissenschaften im deutschsprachigen Raum (AKWI)
vom 11.09. bis 13.09.2022, ausgerichtet von
der Hochschule für Technik und Wirtschaft Berlin (HTW Berlin) und
der Hochschule für Wirtschaft und Recht Berlin (HWR Berlin)

Herausgeber*innen

Sandy Eggert
Claudia Lemke
Verena Majuntke
Birte Malzahn
Vera Meister
Katharina Simbeck
Christian Czarnecki
Martin Wolf

Angewandte Forschung in der Wirtschaftsinformatik 2022

Tagungsband zur 35. Jahrestagung des Arbeitskreises Wirtschaftsinformatik an Hochschulen für Angewandte Wissenschaften im deutschsprachigen Raum (AKWI) vom 11.09. bis 13.09.2022, ausgerichtet von der Hochschule für Technik und Wirtschaft Berlin (HTW Berlin) und der Hochschule für Wirtschaft und Recht Berlin (HWR Berlin)

Herausgeber*innen

Prof. Dr. Sandy Eggert, HWR Berlin, sandy.eggert@hwr-berlin.de

Prof. Dr. Claudia Lemke, HWR Berlin, claudia.lemke@hwr-berlin.de

Prof. Dr. Verena Majuntke, HTW Berlin, verena.majuntke@htw-berlin.de

Prof. Dr. Birte Malzahn, HTW Berlin, birte.malzahn@htw-berlin.de

Prof. Dr. Vera G. Meister, TH Brandenburg, vera.meister@fh-brandenburg.de

Prof. Dr. Katharina Simbeck, HTW Berlin, katharina.simbeck@htw-berlin.de

Prof. Dr. Christian Czarnecki, FH Aachen, czarnecki@fh-aachen.de

Prof. Dr. Martin Wolf, FH Aachen, m.wolf@fh-aachen.de

Mitglieder des Programmkomitees:

Prof. Dr. Wolfgang Alm (HS Aschaffenburg)

Prof. Dr. Gunnar Auth (HS Meißen)

Prof. Dr. Thomas Barton (HS Worms)

Prof. Dr. Christian Czarnecki (FH Aachen)

Prof. Dr. Christian Drumm (FH Aachen)

Prof. Dr. Sandy Eggert (HWR Berlin)

Prof. Dr. Ingo Elsen (FH Aachen)

Prof. Dr. Heinrich Faßbender (FH Aachen)

Prof. Dr. Dorothee Feldmüller (HS Bochum)

Prof. Dr. Andreas Heberle (HS Karlsruhe)

Prof. Dr. Frank Herrmann (OTH Regensburg)

Prof. Dr. Stephan Jacobs (FH Aachen)

Prof. Dr. Jürgen Karla (HS Niederrhein)

Prof. Ute Klotz (HS Luzern)

Prof. Dr. Bodo Kraft (FH Aachen)

Prof. Dr. Claudia Lemke (HWR Berlin)

Prof. Dr. Olga Levina (TH Brandenburg)

Prof. Dr. Verena Majuntke (HTW Berlin)

Prof. Dr. Birte Malzahn (HTW Berlin)

Prof. Dr. Vera G. Meister (TH Brandenburg)

Prof. Dr. Frank Morelli (HS Pforzheim)

Prof. Dr. Christof Mosler (HFT Stuttgart)

Prof. Dr. Jörg Puchan (HS München)

Prof. Dr. Regina Polster (HS Schmalkalden)

Prof. Dr. Kerstin Prechel (Duale Hochschule Schleswig-Holstein, Kiel)

Prof. Dr. Margit Scholl (TH Wildau)

Prof. Dr. Thomas Specht (HS Mannheim)

Prof. Dr. Juliane Siegeris (HTW Berlin)

Prof. Dr. Katharina Simbeck (HTW Berlin)

Prof. Dr. Jessica Slamka (HS München)

Prof. Dr. Martin Spott (HTW Berlin)

Prof. Dr. Ingo Stengel (HS Karlsruhe)

Prof. Dr. Jan Stöß (HS Karlsruhe)

Prof. Dr. Arif Wider (HTW Berlin)

Prof. Dr. Olga Willner (HTW Berlin)

Prof. Dr. Martin Wolf (FH Aachen)

Redaktionsschluss: 08.08.2022

Erscheinungstermin: 11.09.2022

ISBN 978-3-95545-409-8

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Veröffentlicht im GITO Verlag 2022

Titelbild: Adobe Stock 119945440 @ SFIO CRACHO

Covergestaltung: Nicole Kappe, nicole@qstall-design.de

© GITO mbH Verlag Berlin 2022

GITO mbH Verlag

Kaiserdamm 23, 14057 Berlin

Tel.: +49.(0)30.41 93 83 64, Fax: +49.(0)30.41 93 83 67

E-Mail: service@gito.de, Internet: www.gito.de

https://doi.org/10.30844/AKWI_2022





Consultant IT-Strategie, IT-Transformation und IT-Architektur CIO Advisory (w/m/d)

Jobcode:
C-100237

Deine Aufgaben

- Anspruchsvolle IT-Beratung** – Du wirst Teil eines hochmotivierten Beraterteams und hilfst unseren Kunden in Deutschland und Europa von der IT-Strategie bis zu deren Umsetzung ihre Leistungsfähigkeit in und mit der IT zu steigern.
- Innovative Lösungskonzepte** – In einem interdisziplinären Team erarbeitest du Lösungen für die IT Transformation, die Optimierung des IT-Alignments oder unterstützt die Entwicklung neuer digitaler Geschäftsmodelle unserer Kunden.
- Effiziente Optimierungsmaßnahmen** – Du verwendest Best-Practice-Ansätze eines weltweit führenden Beratungsunternehmens, um Zielorganisationen zu gestalten, Abläufe zu optimieren und Abbildungen in IT-Systemen zu unterstützen.
- Spannende Aufgaben** – Dein Aufgabengebiet ist anspruchsvoll und abwechslungsreich. Du arbeitest im Team mit Kolleginnen und Kollegen unterschiedlichster Skill Sets, um gemeinsam Projekte erfolgreich umzusetzen.
- Gezielte Förderung** – Zusammen mit deiner Mentorin oder deinem Mentor arbeitest du an deiner Karriereentwicklung. Dazu gehört u.a. ein umfassendes Weiterbildungsangebot der PwC Academy.

Dein Profil

- Dein Studium der Wirtschaftsinformatik, Wirtschaftswissenschaften, Wirtschaftsingenieurwesen oder eines vergleichbaren Studiengangs hast du erfolgreich abgeschlossen oder stehst kurz davor.
- Durch Praktika konntest du erste Branchenerfahrungen, z. B. bei Industrieunternehmen, im Handel, im Banken- oder Versicherungsumfeld oder in der Energiebranche sammeln.
- Du verfügst über grundlegende Kenntnisse in einem oder mehreren der folgenden Bereiche: IT-Strategie, IT-Sourcing, Digitalisierung, IT-Architektur oder IT-Alignment, Organisation und Transformation von IT-Einheiten sowie IT-Controlling/-Benchmarking.
- Du arbeitest gerne im Team und kannst deine analytischen und kommunikativen Fähigkeiten verbunden mit Eigeninitiative in einem Projektteam erfolgreich einsetzen.
- Sehr gute Deutsch- sowie gute Englischkenntnisse runden dein Profil ab.

PwC

Du hast Talent. Wir sind ein internationales Netzwerk aus 284.000 Mitarbeiter:innen. Du suchst einen spannenden Job. Wir bieten dir als weltweit führende Wirtschaftsprüfungs- und Beratungsgesellschaft vielfältige Möglichkeiten für deine Karriere. Du möchtest einen Arbeitgeber mit passenden Werten. Unsere Arbeitsweise ist geprägt von Vertrauen, Respekt, Weitsicht und Wertschätzung.

Standorte

Berlin, Düsseldorf, Essen,
Frankfurt a. M., Hamburg, Köln,
München

Kontakt

Claudia Schmitz
Tel.: +49 69 9585-2222

Inhaltsverzeichnis

Geleitwort der Sprecherin des Arbeitskreises Wirtschaftsinformatik an Hochschulen für Angewandte Wissenschaften im deutschsprachigen Raum <i>Prof. Dr. Vera G. Meister</i>	9
Vorwort der Herausgeber*innen <i>Prof. Dr. Sandy Eggert, Prof. Dr. Claudia Lemke, Prof. Dr. Verena Majuntke, Prof. Dr. Birte Malzahn, Prof. Dr. Vera G. Meister, Prof. Dr. Katharina Simbeck, Prof. Dr. Christian Czarnecki, Prof. Dr. Martin Wolf</i>	11
Wissenschaftliche Beiträge	
IT-Akzeptanzkriterien bei der Arbeit im Ehrenamt: Eine qualitative Studie <i>Simone Dogu, Ingo Stengel und Stefanie Regier</i>	14
Content and Ad Formats in Performance Marketing Campaigns: Brand Usage Analysis of Advertising Options on Facebook <i>Maria Margarita Cabrera Frias und Peter Konhäusner</i>	27
Anwendbarkeit von Enterprise Security Assessments sowie Enterprise Architecture Tools für KMU <i>Kendime Ismailji, Christof Mosler und Silvia Knittl</i>	43
Acquiring Android App Development Skills in a Virtual Learning Environment: Extending the Virtual Programming Lab for Moodle towards Self-Assessed App Development Tasks employing Android and Gradle <i>Frank Neumann, Juan Carlos Rodríguez-del-Pino und Sebastian Homer</i>	60
Requirements for Competence Developing Games in the Environment of SE Competence Development <i>Philipp M. Zäh, Marcel Biewendt, Martin R. Wolf und Mathias Eggert</i>	73
Asynchrone, digitale Lehrpfade in der Kostenrechnung als Open Educational Ressource <i>Katharina Simbeck und Katrin Dziergwa</i>	89
Design Science Research als wissenschaftliche Herangehensweise für Abschlussarbeiten mit Gestaltungsauftrag in anwendungsorientierten Studiengängen <i>Irene Weber</i>	101

Digital Transformation in Educational Institutions: Scrutinizing the Debate and Highlighting Success Factors <i>Linda Mai, Lynn Schmodde und Marius C. Wehner</i>	117
Process Mining Model Quality in Software Development Case Studies: An Analysis <i>Marianne Schmolke</i>	132
Die Rolle der Künstlichen Intelligenz bei kleinen und mittleren Unternehmen des Dienstleistungssektors <i>Giuseppe Sortino und Sandy Eggert</i>	148
AI-ware als sozio-technische Systeme: Entwurf eines Gestaltungsrahmens für den Einsatz von KI in Unternehmen <i>Claudia Lemke und Dagmar Monett</i>	168
Probabilistic Programming – ein neuer Baustein für die KI <i>Thomas Wengerek</i>	185
Visualisierung hochdimensionaler Daten: State-of-the-Art von nichtlinearen Methoden zur Dimensionsreduktion <i>Tobias Schmieg und Helmut Beckmann</i>	202
Kernkonzepte der Taxonomiesprache GenDifS <i>Johannes Busse</i>	214
Konzeptionierung einer Softwareanwendung zur Verdachtsdiagnostik und Triage in der Pferdemedizin <i>Laura Haase und Lucas Rahn</i>	232
Extraktion und Analyse von Schlüsselwörtern in einer Literaturrecherche zu Quantum-Computing <i>Mazlum Copurkuyu und Thomas Barton</i>	245
Konzeption eines Machine-Learnings-Verfahrens zum Lösen von Green Vehicle Routing Problemen <i>Pablo Stockhausen, Andreas Johannsen und Robert Maurer</i>	261
Einsatz von Machine Learning im Innovationsmanagement: Eine Analyse des State-of-the-Art <i>Daniel Szafarski und Helmut Beckmann</i>	273
OpenPredict - An Open Research Dataset and Evaluation Protocol for Fine-grained Predictive Testing <i>David Brodmann and Erik Rodner</i>	287

Prototypen

BPMN-CREATOR - Ein innovatives Tool zur vollautomatischen Digitalisierung zuvor manuell erstellter Geschäftsprozessmodelle <i>Selcan Ipek-Ugay, Tabea Herrmann und Eric Siegeris</i>	305
Data Synthesis for Fairness Audits of Learning Analytics Systems <i>Linda Fernsel und Katharina Simbeck</i>	316
Instrumente zur aktiven und sicheren Verbraucherteilhabe an öffentlichen Online-Diensten <i>Esther Ruiz Ben, Anne Jellinghaus, Britta Hesse, Margit Scholl und Sebastian Möller</i>	321
Serious Games als Lernmethode zur Steigerung der Informationssicherheit <i>Frauke Prott, Ulrike Küchler, Regina Schuktomow und Margit Scholl</i>	325
Userorientierter Prototyp einer Open-Source App für die Paketzustellung <i>Daniel Quiter, Birte Malzahn, Maximilian Engelhardt und Stephan Seeck</i>	335
Das Process-Simulation.Center: Prototyp eines Integrierten Managementsystems für Lehre, Forschung und Praxis <i>Carlo Simon, Lara Zakfeld und Stefan Haag</i>	344

Studentische Beiträge

Evaluierung des Dense Passage Retrievals zur Dokumentensuche in Behörden im Vergleich zu BM25 <i>Anket Mehra, Malte Prieß und Tom Vincent Peters</i>	353
Strategische Technologiefrüherkennung für die Intralogistik 4.0 in der Nutzfahrzeugindustrie <i>Christina Westermaier und Jörg Puchan</i>	363
Untersuchung des Gender Bias im deutschsprachigen Wikipedia <i>Sarah Krause</i>	372
Entwicklung einer mHealth Anwendung zur psychosozialen Unterstützung von pflegenden Angehörigen <i>Janina Krassa, Pia Christina Fisahn und David Koschnick</i>	385
Verzeichnis der Autor*innen	394

Geleitwort der Sprecherin des Arbeitskreises Wirtschaftsinformatik an Hochschulen für Angewandte Wissenschaften im deutschsprachigen Raum

Liebe Teilnehmer*innen,

zweieinhalb Jahre besonderer Herausforderungen liegen hinter uns, die uns Lehrenden und Forschenden an Hochschulen für Angewandte Wissenschaften (HAW) und unseren Partner*innen in Wirtschaft und Verwaltung enorm viel abverlangt haben. Wir sind gewachsen in dieser Zeit. Als Wirtschaftsinformatiker*innen und damit Expert*innen auf dem Feld der Digitalisierung sind wir damit proaktiv umgegangen, haben überall auch die Entwicklungschancen aufgespürt, innovative Projekte initiiert und neue Praktiken der digitalen Kommunikation und Kollaboration in Forschung und Lehre gelebt. Dennoch – oder vielleicht gerade deshalb – sind wir glücklich und froh, nach dieser langen Zeit des rein digitalen Austauschs wieder in Präsenz tagen zu können.

Die Freude und Erleichterung darüber haben sich auch in der Vielzahl der hochwertigen Einreichungen gezeigt. In einem wettbewerblichen Peer-Review-Verfahren konnten sich 19 wissenschaftliche Beiträge, sechs Prototypen und vier Student Paper durchsetzen, die im vorliegenden Tagungsband veröffentlicht werden. Weitere studentische Arbeiten werden als Poster auf der Konferenz vorgestellt. Diese Leistung wurde von 36 Autorinnen und 34 Autoren erbracht. Damit konnten wir ein wichtiges Ziel der Tagung erreichen: den Beitrag von Frauen in der angewandten Forschung der Wirtschaftsinformatik sichtbar zu machen. Auf dieses Ziel zahlt auch der Student Best Paper Award ein, der in diesem Jahr für das beste studentische Paper mit weiblicher Beteiligung verliehen wird.

Die diesjährige Tagung bietet eine Reihe weiterer Neuerungen. Erstmals seit Beginn der jährlichen Fachtagungen des AKWI im Jahr 2007 wird sie von zwei Hochschulen kooperativ ausgerichtet: der HTW Berlin und der HWR Berlin. Die Organisation wurde von einem Team aus fünf Professorinnen dieser Hochschulen gestemmt. Die Federführung lag in den Händen von Professorin Birte Malzahn. Auf Vorschlag des stellvertretenden Sprechers des AKWI, Professor Christian Czarnecki, erhalten die besten Paper der Tagung die Möglichkeit, ihre Beiträge in erweiterter Form auf dem Workshop „Wirtschaftsinformatik an HAWs“ im Rahmen der Jahrestagung INFORMATIK 2022 der Gesellschaft für Informatik (GI e. V.) vorzustellen. Damit einher geht eine erneute Publikationsmöglichkeit in der GI-Edition Lecture Notes in Informatics (LNI). Im vorliegenden Tagungsband haben wir uns am LNI-Template orientiert. Schließlich ist neu, dass in diesem Jahr nicht nur der Tagungsband als Ganzes, sondern alle einzelnen Beiträge unter einer eigenständigen DOI im Open Access publiziert werden.

Auch im vergangenen Jahr ist es dem AKWI gelungen, alle bereits eingeführten Aktivitäten in guter Qualität fortzuführen. Wir engagieren uns in der Weiterentwicklung des Systems der Qualitätssicherung in Studium und Lehre. Viele Kolleg*innen arbeiten als

Gutachter*innen in der Akkreditierung von Wirtschaftsinformatik-Studiengängen im In- und Ausland. Darüber hinaus nehmen wir Einfluss auf Prozesse und Richtlinien in diesem Handlungsfeld, wie z. B. aktuell die Neufassung der Fragebögen für das CHE-Ranking. So wurde auf unser Betreiben hin der Indikator Forschungsorientierung in der Master-Studierendenbefragung neu gefasst und auf HAWs ausgedehnt. Das von der GI initiierte Konsortium NFDIxCS – an dem auch der AKWI beteiligt ist, wurde vom NFDI-Expertengremium zur Förderung empfohlen. Im Rahmen der Jahrestagung Modellierung 2022 wurde ein Grundsatzpapier für die Behandlung von Modellen als Forschungsdaten-Artefakte initiiert.

Ein wichtiges Handlungsfeld des AKWI ist traditionell die fachbezogene Publikationstätigkeit. In der Buchreihe des Springer Verlags „Angewandte Wirtschaftsinformatik“ wurde der sechste Band mit dem Titel „Data Science anwenden – Einführung, Anwendungen und Projekte“ veröffentlicht. Ein siebenter, besonderer Band ist in Vorbereitung. Er entsteht als fachlich editierte, maschinelle Übersetzung des 2018 erschienenen dritten Bandes durch den Dienst DeepL.com. Herausgeber der Buchreihe sind die Professoren Thomas Barton und Christian Müller. Das 2013 gegründete E-Journal des AKWI hat zwei weitere Ausgaben vorgelegt. Wie schon in den bisherigen Ausgaben werden viele charakteristische Bereiche der Wirtschaftsinformatik behandelt: Geschäftsprozesse im weiteren Sinne, Systeme wie SAP, der Betrieb von Systemen und neuere Ansätze aus dem Bereich der künstlichen Intelligenz. Für diese engagierte Arbeit danken wir den herausgebenden Professoren: Christian Müller, Konrad Marfurt, Norbert Ketterer und Frank Herrmann. Viele Kolleg*innen engagieren sich als Autor*innen in beiden Publikationsorganen.

Ich freue mich auf das persönliche Treffen in Berlin und einen spannenden kollegialen Austausch.



Prof. Dr. Vera G. Meister, Sprecherin des AKWI

Vorwort der Herausgeber*innen

Liebe Leser*innen,

dieser Tagungsband zur 34. Jahrestagung des „Arbeitskreises Wirtschaftsinformatik an Hochschulen für Angewandte Wissenschaften im deutschsprachigen Raum“ (AKWI) der Gesellschaft für Informatik e.V. ist etwas Besonderes. Er enthält alle Beiträge, die nach zwei Jahren pandemiebedingter Virtualisierung unserer Jahrestagung erstmals wieder in Präsenz vorgestellt und diskutiert worden sind. Zum anderen ist dieser Tagungsband das Ergebnis einer Überlegung, initiiert von einigen Kolleg*innen unter der treibenden Kraft unserer Sprecherin Prof. Dr. Vera G. Meister, mit dieser Jahrestagung erstmals vor allem Frauen in der Wirtschaftsinformatik anzusprechen. Das zeigte sich in einem ausschließlich weiblichen Organisationskomitee aus Professorinnen der beiden ausrichtenden Hochschulen dieser Tagung in Berlin: der Hochschule für Technik und Wirtschaft Berlin (HTW Berlin) und der Hochschule für Wirtschaft und Recht Berlin (HWR Berlin). Unprätentiös wurde diese Tagung am Standort Berlin für die Ausrichtung an beiden Hochschulen vorbereitet. Mit der Aufforderung zur Einreichung von Beiträgen wurden erstmalig direkt Wissenschaftlerinnen in der Wirtschaftsinformatik aufgefordert, ihre Forschungsarbeiten im Rahmen der AKWI-Tagung zu präsentieren. Auch Studierende wurden stärker adressiert, Beiträge unter weiblicher Beteiligung einzureichen. Damit zeigt dieser Tagungsband nicht nur in gewohnter Tradition die gesamte Bandbreite der aktuellen Entwicklungen in Forschung und Industrie zum aktuellen Zeitpunkt. Dieser Band repräsentiert zudem einen Ausschnitt an anwendungsorientierten Forschungsarbeiten unter der Führung von Wissenschaftlerinnen und Studentinnen. Wir hoffen, damit zeigen zu können, dass die deutschsprachige Wirtschaftsinformatik eine diverse Disziplin ist, in der auch Frauen ihren festen Platz haben.

Wie auch schon im letzten Jahr ist das Thema Künstliche Intelligenz mit den vielen Teilgebieten wieder ein Schwerpunkt der Tagung. Aber auch eher klassische Themen zum Geschäftsprozessmanagement, zum IT-Management und zu Fragen der IT-Sicherheit sind vertreten. Zudem sehen wir den allgemein beobachtbaren Trend der stärkeren didaktischen Beschäftigung in der Vermittlung allgemeiner digitaler Kompetenzen auch bei unseren Beiträgen. Besonders freuen wir uns, vier studentische Beiträge vorstellen zu können. Weitere studentische Arbeiten wurden auf der Tagung als Poster vorgestellt, sind aber nicht in diesem Tagungsband enthalten. Sechs Beiträge über Prototypen enthalten auch studentische Arbeiten und spiegeln damit in besonderer Weise das Engagement der Studierenden in der Wirtschaftsinformatik und der verwandten Disziplinen wider, das wir als Hochschullehrende sicher auch im alltäglichen Lehrbetrieb immer wieder beobachten, und weisen auf die hohe Praxisrelevanz der Wirtschaftsinformatik hin.

Wir bedanken uns bei allen Autor*innen für die Einreichungen. Aufgrund der hohen Resonanz bei den Einreichungen in diesem Jahr musste das Organisationskomitee eine Auswahl auf der Grundlage der Ergebnisse des qualitativ hochwertigen „Double-Blind-Peer-Review“-Prozesses treffen. Die Erstellung des Bandes wurde durch die gute Kooperation, das

Vorwort der Herausgeber*innen

pünktliche Abliefern der Beiträge sowie das Beachten der Hinweise der Gutachter*innen und des Organisationskomitees vereinfacht.

Das Organisationskomitee bedankt sich damit bei allen Gutachter*innen für die konstruktiven Anmerkungen und für das zügige Erstellen der Gutachten. Zudem danken wir dem Gito-Verlag, insbesondere Frau Martina Braun, für ihre Unterstützung während der Vorbereitung des Tagungsbands und für die Übernahme der Erstellung des Dokuments, das Ihnen nun als Online-Version zur Verfügung steht.

Zum Ende noch eine persönliche Anmerkung des Organisationskomitees: Wir danken allen Kolleginnen im Team für die sehr kollegiale und fast schon freundschaftliche Zusammenarbeit in Vorbereitung dieser Tagung in Berlin. Insbesondere Frau Prof. Dr. Birte Malzahn als unserer Projektleiterin möchten wir hier unseren Dank aussprechen.

Berlin, September 2022

Prof. Dr. Sandy Eggert (HWR Berlin)

Prof. Dr. Claudia Lemke (HWR Berlin)

Prof. Dr. Verena Majuntke (HTW Berlin)

Prof. Dr. Birte Malzahn (HTW Berlin)

Prof. Dr. Vera G. Meister (TH Brandenburg)

Prof. Dr. Katharina Simbeck (HTW Berlin)

Prof. Dr. Christian Czarnecki (FH Aachen)

Prof. Dr. Martin Wolf (FH Aachen)

Wissenschaftliche Beiträge

IT-Akzeptanzkriterien bei der Arbeit im Ehrenamt

Eine qualitative Studie

Simone Dogu¹, Ingo Stengel² und Stefanie Regier³

Abstract: Die Digitalisierung ist ein zentraler gesellschaftlicher Trend des 21. Jahrhunderts. Digitale Informations- und Kommunikationstechnologien gewinnen in nahezu allen öffentlichen und privaten Lebensbereichen an Bedeutung. Auch im Bereich des freiwilligen Engagements werden zunehmend Möglichkeiten der Digitalisierung genutzt, wie z.B. die Terminkommunikation einer Übungsleiterin im Sportverein per E-Mail. Ob und wie Engagierte digitale Techniken für ihre freiwillige Tätigkeit überhaupt nutzen können, hängt u.a. davon ab, ob sie Zugang zum Internet haben und ihnen entsprechende Infrastruktur zur Verfügung steht. Neben den technischen Voraussetzungen spielt aber auch der Wille, IT im Ehrenamt zu nutzen, eine wesentliche Rolle. Diese Studie soll einen Beitrag dazu leisten, mehr über mögliche Akzeptanzkriterien bei der Nutzung digitaler Tools im Ehrenamt zu erfahren.

Keywords: Akzeptanzkriterien, IT, Ehrenamt, Experteninterviews, Digitalisierung

1 Einleitung

Freiwilliges Engagement ist eine wichtige Säule unserer modernen Gesellschaft. Im Jahr 2019 engagierten sich 28,8 Millionen (39,7 Prozent) Menschen in Deutschland ab 14 Jahren ehrenamtlich. Zwischen 1999 und 2019 ist der Anteil freiwillig Engagierter somit gestiegen. Ob Übernahme von ehrenamtlichen Positionen in Gemeinderäten, Mitwirkung in Bürgerinitiativen, Unterstützung im örtlichen Sportverein oder beim Umweltschutz: die Aufgaben und Tätigkeiten sind dabei sehr vielfältig [DB20]. Die Besonderheit bei ehrenamtlicher Arbeit ist, dass diese auf Freiwilligkeit beruht und so z. B. Unterschiede zu hauptamtlichen Tätigkeiten aufweist. In den letzten zwanzig Jahren hat sich das Ehrenamt gewandelt. Zum einen sind die Anteile an engagierten Personen gesunken, die sich mit hohem Zeitaufwand engagieren und die Leitungsfunktionen im Engagement übernehmen.

¹ Hochschule Karlsruhe, Fakultät für Informatik und Wirtschaftsinformatik, Moltkestr. 30, 76133 Karlsruhe, simone.dogu@h-ka.de

² Hochschule Karlsruhe, Fakultät für Informatik und Wirtschaftsinformatik, Moltkestr. 30, 76133 Karlsruhe, ingo.stengel@h-ka.de

³ Hochschule Karlsruhe, Fakultät für Informatik und Wirtschaftsinformatik, Moltkestr. 30, 76133 Karlsruhe, stefanie.regier@h-ka.de

Vielmehr üben immer mehr Engagierte ihre Tätigkeit in einem informell organisierten Rahmen aus, der in der Regel mit flacheren hierarchischen Strukturen einhergeht und weniger Leitungs- und Vorstandspositionen bedarf [HS17]. Damit einhergehend ist das freiwillige Engagement aktuell zunehmend stärker durch das Thema Digitalisierung geprägt, das auch in diesem Bereich neue Möglichkeiten eröffnet. Digitale Informations- und Kommunikationstechnologien gewinnen in nahezu allen öffentlichen und privaten Lebensbereichen an Bedeutung. Auch im Bereich des freiwilligen Engagements werden die Möglichkeiten der Digitalisierung genutzt. Vielfach geht es bei der Digitalisierung des Engagements um die Unterstützung von freiwilligen Tätigkeiten, die weiterhin ‚analog‘ stattfinden, wie beispielsweise die Organisation eines lokalen Stammtisches per E-Mail oder die kurzfristige Verschiebung des Trainings im Sportverein über die WhatsApp-Gruppe. Ebenso können unter die Digitalisierung des Engagements auch neue Formen des freiwilligen Engagements gefasst werden, wie z. B. die Veranstaltung eines Online-Stammtisches [DB20] [HBS19] [Hi14]. Inzwischen nutzt mehr als die Hälfte der freiwillig Engagierten das Internet im Rahmen ihrer freiwilligen Tätigkeit. Der Anteil der ehrenamtlichen Engagierten, die im Rahmen ihrer freiwilligen Tätigkeit das Internet verwenden, ist lediglich in den Jahren von 2004 bis 2009 angestiegen, seitdem aber recht stabil geblieben. Aktuell wird das Internet nur von etwas mehr als der Hälfte genutzt. Dies bedeutet auch, dass ein großer Teil der Engagierten (mehr als 40 Prozent) das Internet für ihre Tätigkeit nicht nutzt, weil es z. B. für die Ausführung bestimmter Aufgaben nicht relevant ist. Digitale Technologien spielen damit für viele Engagierte eine große Rolle in ihrer freiwilligen Tätigkeit, aber dies trifft keineswegs auf alle Menschen zu, die ehrenamtlich aktiv sind [TRH21].

Um die zentrale Annahme der vorliegenden Studie, dass Einstellungs- und Verhaltensdimensionen bei der Entscheidungsfindung der Ehrenamtlichen eine entscheidende Rolle für die Nutzung digitaler Tools spielen, zu bestätigen, fehlt es bei diesem Forschungsthema derzeit noch an detaillierter Forschung.

Die gesamte Forschungsdesign möchte Erkenntnisse zu möglichen IT-Akzeptanzkriterien liefern und ist auf einen Mixed-Methods-Ansatz ausgelegt. Sie beinhaltet eine qualitative sowie quantitative Studie. Gegenstand dieses Artikels sind der Aufbau, die Durchführung und die Ergebnisse der qualitativen Studie. Ziel dieser ist es, auf Basis der vorhandenen Forschung, mehr über die Akzeptanzkriterien von Ehrenamtlichen im Rahmen ihrer Tätigkeit zu erfahren und um eine ausreichend fundierte und aussagekräftige theoretische Grundlage für die nachfolgende quantitative Studie zu schaffen.

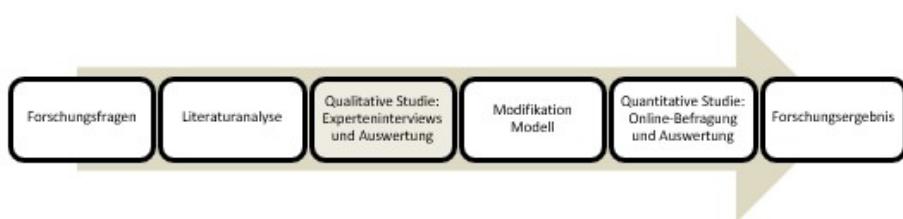


Abb. 1: Einordnung der qualitativen Studie in das gesamte Studiendesign (eigene Darstellung)

2 Forschungsstand

Im Folgenden soll einen Überblick über den relevanten Forschungsstand im Bereich Einstellungs- und Verhaltensdimensionen bei der Nutzung digitaler Tools von Ehrenamtlichen gegeben werden.

Im Bereich der Informationssysteme (IS) war die Forschung zur Akzeptanz in den letzten zwei Jahrzehnten ein dominierendes Thema im Bereich des Nutzerverhaltens. Ein bedeutender Teil der theoretischen Forschung hat sich auf das Medien-, Technologie- und Dienstleistungsakzeptanzverhalten konzentriert und dabei die Theory of Reasoned Action (TRA) [AF80], die Theory of Planned Behaviour (TPB) [Aj91] und das Technologieakzeptanzmodell (TAM) [Da89] angewendet. Diese Theorien haben zwar einen wichtigen Beitrag zur Untersuchung der Technologieübernahme durch Einzelpersonen geleistet, doch liefern sie kein ganzheitliches Bild, sondern fokussieren sich meist auf wenige Faktoren, die die Entscheidung beeinflussen.

Die Forschung von Einstellungs- und Verhaltensdimensionen im Ehrenamt fokussiert sich auf Motive und Motivationen von ehrenamtlichen Mitarbeitenden und sind eher allgemeiner Natur. Der Deutsche Freiwilligensurvey (FWS) ist eine repräsentative, seit 1999 alle fünf Jahre durchgeführte Befragung zum freiwilligen Engagement in Deutschland, die sich an Personen ab 14 Jahren richtet. Er ist die umfassendste und detaillierteste quantitative Erhebung zum bürgerschaftlichen Engagement in Deutschland. Freiwillige Tätigkeiten und die Bereitschaft zum Engagement werden in telefonischen Interviews erhoben und können nach Bevölkerungsgruppen und Landesteilen dargestellt werden (FWS 2019: N=27.759). Darüber hinaus können Veränderungen in den Formen und Kontexten des freiwilligen Engagements nachgezeichnet werden. Außerdem können die Engagierten und Personen, die sich nicht bzw. nicht mehr engagieren, beschrieben werden. Neben Motiven und Motivationen sowie Beendigungs- und Hinderungsgründe der ehrenamtlichen Tätigkeit wie z.B. fehlende Zeit [AK21] werden im Rahmen dieser Erhebung zudem Erkenntnisse zu der Internetnutzung von Ehrenamtlichen gewonnen [TRH21]. Zudem gibt es Untersuchungen zu der Bedeutung des Ehrenamts in Deutschland, auch vor dem Hintergrund des gesellschaftlichen Wandels [Ka21], sowie zu Unterschieden hinsichtlich Geschlechter, Altersgruppen, Schulbildung und Erwerbstätigkeit [Si21]. Auch die gesellschaftlichen Bereiche des freiwilligen Engagements wurden untersucht [KH21], so findet die meisten ehrenamtlichen Tätigkeiten im Bereich Sport und Bewegung, Kultur und Musik sowie im sozialen Bereich statt.

Robinson et al. [RKS21] erarbeiten in ihrer Review-Studie eine Typologie von Projekten sowie Erkenntnisse über die Nutzung und Erwartungshaltung entsprechender Projekttools für den Gebrauch in Projekten der Zivilgesellschaft. Lau et al. [La19] untersuchten die Motivationen Ehrenamtlicher, sich in soziale Projekte einzubringen und diese zu unterstützen.

Bereits 2007 untersuchten Zhang & Gutierrez [ZG07] unter Anwendung der Theory of Planned Behavior (TPB) verschiedene Einflussfaktoren der IT-Akzeptanz (z.B. Perceived Personal Usefulness, Peer Influence, Self-Efficacy, User Intention to Use IT) im Umfeld von Nonprofit-Organisationen.

Um der steigenden Komplexität gerecht zu werden, wurde das TAM um ergänzende Faktoren erweitert. Saura et al. [SPV20] untersuchten die Nutzung digitaler Tools, um neue Ehrenamtliche zu rekrutieren unter Verwendung eines erweiterten TAMs. Die Ergebnisse dieser Untersuchung zeigen, dass die Kommunikationsstrategien für NGOs und Plattformen, die auf Ehrenamtliche angewiesen sind, auf digitalen Kanälen von größter Bedeutung sind. So konnten Einflussfaktoren wie Perceived Usefulness, Perceived easy of Use, Trust und deren Bedeutung nachgewiesen werden.

Zudem gibt es eine Vielzahl an Untersuchungen, die sich einzelnen Einflussfaktoren der IT-Akzeptanz widmen, wie z.B. Personal Innovativeness (u.a. [TAS16]). Hierfür möchte diese qualitative Studie weitere Erkenntnisse liefern, um im nächsten Schritt aus den gewonnenen Einblicken ein Modell erstellen zu können.

Eine Übersicht relevanter Forschung befindet sich in nachfolgender Tabelle:

Autor(en)/Datum	Titel	Forschungsmethodik, N, Modelle	Ergebnisse
Arriagada, C. & Karnick, N., 2022	Motive für freiwilliges Engagement, Beendigungs-gründe, Hinderungs-gründe und Engagementbereitschaft	Basierend auf der Erhebung des Deutschen Freiwilligensurvey 2019 (N=27.759, Interviews)	Allgemeine Motive, Hinderungs- und Beendigungsgründe von Ehrenamtlichen
Kausmann, C., Kelle, N., Simonson, J., Tesch-Römer, C., 2022	Freiwilliges Engagement – Bedeutung für Gesellschaft und Politik	Basierend auf der Erhebung des Deutschen Freiwilligensurvey 2019 (N=27.759, Interviews)	Unterschiede und Ungleichheiten im ehrenamtlichen Engagement, Entwicklung freiwilligen Engagements vor dem Hintergrund des gesellschaftlichen Wandels
Kausmann, C. & Hagen, C., 2022	Gesellschaftliche Bereiche des freiwilligen Engagements	Basierend auf der Erhebung des Deutschen Freiwilligensurvey 2019 (N=27.759, Interviews)	Untersuchung der Bereiche des Ehrenamts (Top 3): Sport und Bewegung, Kultur und Musik, Sozialer Bereich
Lau et al., 2019	Volunteer motivation, social problem solving, self-efficacy, and mental health: a structural equation model approach	Befragung, N=1.530	Untersuchung von Motivationen, ein soziales Projekt zu unterstützen

Robinson, J.A. et al., 2021	Meeting volunteer expectations — a review of volunteer motivations in citizen science and best practices for their retention through implementation of functional features in CS tools	Review Article	Erstellung und Zusammenfassung einer Typologie von Projekten sowie Erkenntnisse über die Nutzung und Erwartungshaltung entsprechender Projekttools
Saura, JR. et al., 2020	What Drives Volunteers to Accept a Digital Platform That Supports NGO Projects?	245 Befragte, Fragebogen	TAM, Digitale Plattformen ziehen motivierte potenzielle junge Freiwillige an. IT-Akzeptanzkriterien wie Wahrgenommene Nützlichkeit, wahrgenommene Benutzerfreundlichkeit, Vertrauen
Simonson, J. et al., 2022	Unterschiede und Ungleichheiten im freiwilligen Engagement	Basierend auf der Erhebung des Deutschen Freiwilligensurvey 2019 (N=27.759, Interviews)	Untersuchung u.a. der Unterschiede hinsichtlich Geschlechter, Altersgruppen, Schulbildung, Erwerbstätigkeit sowie Vergleich Ost- und Westdeutschland
Tesch-Römer, C. & Huxhold, O., 2021	Nutzung des Internets für die freiwillige Tätigkeit	Basierend auf der Erhebung des Deutschen Freiwilligensurvey 2019 (N=27.759, Interviews)	Erkenntnisse zur Internetnutzung im Ehrenamt
Zhang, W. & Gutierrez, O., 2007	Information Technology Acceptance in the Social Services Sector Context: An Exploration.	61 Befragte, Online-Fragebogen	TPB, IT-Akzeptanzkriterien wie Wahrgenommene persönliche Nützlichkeit, Einfluss von Peers, Selbstwirksamkeit, Absicht der Nutzer, IT zu nutzen
Zhang, et al., 2009	Impact of perceived technical protection on security behaviors.	176 Befragte, Online-Fragebogen	TPB, Verhalten bei der Informations-sicherheit: Wahrgenommener Sicherheits-schutz-Mechanismus

Tab. 1: Übersicht relevanter Studien

Zusammenfassend lässt sich sagen, dass sich die aktuelle Literatur vor allem auf soziodemografische Daten wie Alter, Geschlecht und Bildung fokussiert sowie Motive der ehrenamtlichen Arbeit in Deutschland. Hinsichtlich möglicher IT-Akzeptanzkriterien bei der Nutzung digitaler Tools gibt es erste Untersuchungen, jedoch fehlt es derzeit noch an umfassender fundierter Forschung. Daher ist das Ziel dieser qualitativen Studie die Identifikation relevanter Einflussfaktoren, um die Forschungslücke in diesem Bereich zu schließen.

3 Methode

Um mehr über die Einstellungs- und Verhaltensdimensionen zu IT-Akzeptanzkriterien von Ehrenamtlichen bei Nutzung digitaler Tools zu erfahren, es derzeit jedoch empirisch-induktiver Forschung bei diesem Thema fehlt, wurde für die Gesamtstudie ein Mixed-Methods-Ansatz gewählt. Da in diesem Paper die qualitative Studie vorgestellt wird, liegt der Fokus auf der Auswahl der qualitativen Datenerhebung und Auswertung. Nach Mack et al. [Ma05] sind die gängigsten qualitativen Methoden die teilnehmende Beobachtung, Tieffeneinterviews und Fokusgruppen. Da die vorliegende Studie darauf abzielt, Einstellungsfaktoren zu erheben, sind direkte Beobachtungen möglicherweise nicht durchführbar. Außerdem haben Fokusgruppen den grundsätzlichen Nachteil, dass ihre Ergebnisse von der Gruppendynamik abhängen, was möglicherweise zu verzerrten Ergebnissen führt [CI10]. Vertiefende Interviews ermöglichen es hingegen, den „komplexen Wissensbestand“ [Fl07] von Personen über das untersuchte Thema zu erheben. Da persönliche Einzelinterviews sehr gut geeignet sind, um hervorstechende Verhaltens-, Norm- und Wirksamkeitsüberzeugungen zu eruieren [Mo08], passt diese qualitative Forschungsmethode zu den genannten Zielen dieser Arbeit.

Eine besondere Form des halbstrukturierten Interviews ist das Experteninterview, das sich weder auf sehr offene noch auf starr strukturierte Frage-Antwort-Schemata bezieht [Fl07]. Charakteristisch für das Experteninterview ist die spezifische Fokussierung auf den inhaltlichen Kontext. So rücken Rekonstruktionen über individuelle oder kollektive Lebensphasen, wie zum Beispiel persönliche Lebensverläufe in der Biografieforschung, in den Hintergrund und bilden damit nicht die Gesamtperson als Untersuchungsgegenstand ab. Viel mehr orientieren sich Experteninterviews am organisatorischen oder institutionellen Bezug als einen spezifischen Aspekt der Gesamtperson, die sich im Rahmen des Interviews in Form des Expertenwissens äußert [MN02]. Im Hinblick auf die thematische Ausrichtung werden die Interviews durch einen im Vorfeld der Untersuchung erstellten Interviewleitfaden strukturiert, der im Wesentlichen den zugrunde liegenden konzeptionellen Rahmen widerspiegelt. Dies gibt dem Interviewer die Möglichkeit, Fragen zu variieren oder einzuschlieben, so dass ausführlichere Antworten erreicht werden. Diese können möglicherweise zur Offenlegung von implizitem Wissen oder verborgenen Facetten des menschlichen Verhaltens führen [QD11]. Andererseits werden die Forschungsziele durch den Interviewplan konsequent vermittelt, der die Umsetzung systematisch steuert [LP14]. Die Zielgruppe der Experteninterviews umfasst Personen, die über ein spezifisches Praxiswissen in einem beruflichen Tätigkeitsfeld verfügen [Sc14] [Fl07]. In Anbetracht der Tatsache, dass qua-

litative Forschung eher auf Übertragbarkeit als auf Verallgemeinerbarkeit der Ergebnisse abzielt, ist eine angemessene Stichprobengröße „eine, die die Forschungsfrage adäquat beantwortet“ [Ma96]. Die Auswahl von Experten richtet sich nach Gläser & Laudel [GL10] daher folglich

- nach der Auswahl bestimmter Organisationen,
- nach Reputation und Position von relevanten Akteuren bzw. Personen,
- nach Einflussmöglichkeiten bzw. Beteiligung von Personen an relevanten Entscheidungen bzw. Handlungen.

Des Weiteren spielen bei der Auswahl der Experten auch forschungspraktische Faktoren, wie z.B. die Erreichbarkeit und Bereitschaft der Interviewpartner, eine entscheidende Rolle [GL10]. Die Auswertung erfolgt mittels einer qualitative Inhaltsanalyse. Sie zielt darauf ab, gesammeltes Textmaterial in für die vorliegende Studie aussagekräftige Erkenntnisse umzuwandeln [Pa90]. Diese lassen sich nach Hsieh & Shannon [HS05] grob in drei Methoden einteilen: konventionelle, gerichtete und zusammenfassende Analyse. Die zusammenfassende Inhaltsanalyse zielt darauf ab, die zugrunde liegenden Bedeutungen von Ausdrücken zu entdecken, indem das Auftreten eines bestimmten Wortes oder Inhalts in den erhobenen Daten analysiert wird. In Anbetracht der Ziele der vorgelagerten explorativen Forschungsphase wird in dieser Arbeit eine zusammenfassende Inhaltsanalyse durchgeführt, deren Ergebnis gebildete Kategorien sind, die im weiteren Studienverlauf bei der Modellentwicklung berücksichtigt werden.

4 Auswertung

Insgesamt wurden im Rahmen der qualitativen Studie sechs ehrenamtliche Mitarbeitende, die in verschiedenen Funktionen tätig sind, aus verschiedenen Organisationen ausgewählt und befragt. Alle Interviewpartner verfügen über langjährige Erfahrung in verschiedenen ehrenamtlichen Rollen und Organisationen und sind in Themenfelder wie Digitalisierung sowie die Ausführung verschiedener Tätigkeiten involviert.

Die Ergebnisse der qualitativen Studie liefern vielfältige Erkenntnisse über die Akzeptanzkriterien der Nutzung digitaler Tools im Ehrenamt. Insgesamt wurden aus der qualitativen Untersuchung mehrere theoretische Konstrukte abgeleitet, die die persönlichen Erfahrungen und Einstellungen der ausgewählten Befragten gegenüber der Nutzung digitaler Tools in ihrer ehrenamtlichen Tätigkeit beschreiben. Diese Konzepte dienen gemeinsam mit der Literaturanalyse als Grundlage für die theoriegestützte Ableitung des Modells, indem sie wichtige Aspekte und Determinanten des Akzeptanzverhaltens liefern. Im Nachfolgenden werden die verschiedenen Schwerpunkte aus den Interviews herausgearbeitet und dargestellt.

Fünf der Befragten gingen im Rahmen der durchgeführten Interviews auf die Bedeutung der Digitalisierung und den damit verbundenen Veränderungen des Alltags ein. Vier der Befragten gaben an, dass Ängste gegenüber Innovationen bzw. Neuerungen Einfluss auf die Nutzung digitaler Tools in der ehrenamtlichen Tätigkeit haben und diese somit verhindern können. Die Tatsache, dass Early Adopters/junge Menschen es hierbei einfacher haben, erwähnten drei der Befragten. Dass das Alter der Ehrenamtlichen bei der Nutzung digitaler Tools im Rahmen der Tätigkeit entscheidend sei, wurde explizit von vier der Befragten erwähnt.

Die Offenheit für Neues hat nach Aussage aller sechs Interviewten eine hohe Relevanz in Bezug auf ehrenamtliche Tätigkeit und Digitalisierung. Die Offenheit, neue Tools im Rahmen der ehrenamtlichen Tätigkeit auszuprobieren, findet sich in Aussagen wieder, wie „Die Offenheit ist teilweise vorhanden, aber teilweise gibt es auch die Haltung: auf die alte Art und Weise, die nicht digitale Weise, hat es ja bisher auch gut funktioniert. Also so ein bisschen das Neue als Feind.“ (I1) und „Manchmal muss man einfach ausprobieren.“ (I4)

Auch das Thema IT Security spielte im Rahmen der durchgeführten Experteninterviews eine wichtige Rolle. So gingen fünf der sechs Befragten auf das Thema „Datenschutz“ ein. Es wurden jedoch nicht nur Bedenken und Vorsicht geäußert. Neben Aussagen zu einem eher vorsichtigen Verhalten im Umgang mit digitalen Tools gab es jedoch auch Aussagen wie „Ich persönlich habe da eine recht unbedarfe Haltung dazu“ (I1).

Themen wie „Systemzuverlässigkeit“, „Rechtssicherheit“ und „Jugendschutz“ wurden hingegen nur von jeweils einem der Befragten erwähnt, unterstreichen jedoch erneut die Wichtigkeit des Sicherheitsaspekt bei der Nutzung von digitalen Tools im Ehrenamt im Allgemeinen.

Wie wichtig die Themen Nutzerfreundlichkeit und Barrierefreiheit für die ehrenamtliche Nutzung sind, zeigen Aussagen wie „Mangelnde Bedienbarkeit ist für mich ein Grund für die Nichtnutzung.“ (I1) oder „Barrierefreiheit sehe ich auch als wichtig an.“ (I5). Fünf der Befragten gingen auf diese Themen im Rahmen der Interviews ein, was darauf hindeutet, dass dieser Aspekt auch für ein Akzeptanzmodell von Relevanz sein kann.

Die empfundene Nützlichkeit wurde von vier der Interviewten angesprochen. Einen Mehrwert sehen sie vor allem in Efficiency Enhancement (4), besserer Kommunikation (4) und zentraler Datenspeicherung (4). Wird die Nutzung als Mehrwert wahrgenommen, steigt die Akzeptanz bei Nutzung von digitalen Tools. Dies wird deutlich durch Aussagen wie z.B. „Das Kontaktthalten und die Kommunikation untereinander ist wichtig, also nicht nur das Termine ausmachen, sondern auch spontaner Austausch, z. B. bei Problemen oder wenn etwas zu organisieren ist, finde ich wichtig.“ (I6)

Von ebenfalls großer Bedeutung hat sich das Thema Support/Unterstützung der Ehrenamtlichen erwiesen, d.h. wären entsprechende Unterstützungsangebote vorhanden, würden sie eher digitale Tools im Rahmen ihrer ehrenamtlichen Tätigkeit nutzen. Hierbei gibt es zwei Unterkategorien: „Technischer Support“ und „Hilfe bei der Nutzung“. Die Unterkategorie

„Technischer Support“ bezieht sich hierbei auf die Unterstützung bei rein technischen Problemen, was zwei der Befragten in ihren Interviews angegeben haben. Die Unterkategorie „Hilfe bei der Nutzung“ hingegen beschreibt die Unterstützung bzw. Hilfestellung bei der Nutzung von digitalen Tools, wie z.B. bei der Durchführung einer Videokonferenz. Dies erwähnten fünf der Befragten. Gestützt wird dies durch Aussagen wie „Und vielleicht weil es niemand gibt, der es uns professionell beibringt. Wenn ich schon mit meinem geringen IT-Wissen im Vorstand der IT-erfahrenste bin, ist das halt eine Aussage.“ (I1) und „Schulungen sind auf jeden Fall wichtig.“ (I4).

Als weiterer wichtiger Punkt hat sich die Ausstattung gezeigt. Alle sechs Interviewten erwähnten in ihren Interviews die Wichtigkeit von geeigneter Software für die Erfüllung ihrer Tätigkeiten. Zwei der sechs Interviewten gingen auf die Wichtigkeit von entsprechender Hardware ein („Ich brauche halt einen Rechner mit entsprechender Voraussetzung.“ (I3))

5 Fazit/Ausblick

Die Digitalisierung ist ein zentraler gesellschaftlicher Trend des 21. Jahrhunderts. Digitale Informations- und Kommunikationstechnologien gewinnen in nahezu allen öffentlichen und privaten Lebensbereichen an Bedeutung, so auch im ehrenamtlichen Engagement. Um die Aufgaben und Tätigkeiten erledigen zu können, kommen u.a. das Internet sowie verschiedene digitale Tools zum Einsatz. Um mehr über die Einstellungs- und Verhaltensdimensionen zu erfahren, wurde für die Studie ein Mixed-Methods-Ansatz gewählt. Basierend auf umfangreicher Literaturrecherche wurde für den qualitativen Teil ein Interviewleitfaden entwickelt und sechs halbstrukturierte Interviews mit Experten durchgeführt. Die Experten verfügen durch langjährige ehrenamtliche Tätigkeit(en) über umfangreiches Wissen und Erfahrungen in ihren jeweiligen Organisationen.

Die durch die Interviews gewonnenen Ergebnisse liefern wichtige Erkenntnisse über Akzeptanzfaktoren bei der Nutzung digitaler Tools im ehrenamtlichen Engagement. So wurden durch eine qualitative Inhaltsanalyse mehrere Kategorien als relevant für die ehrenamtliche Tätigkeit identifiziert. Zu den von mind. der Hälfte der befragten Experten genannten Themenschwerpunkten gehören:

- Offenheit für Neues (u.a. [Sv11])
- Datenschutzbedenken (u.a. [LLM20])
- geeignete Software
- Unterstützung/Hilfe bei der Anwendung
- Ängste gegenüber Neuerungen/Innovationen (u.a. [Ng08] [NdC16])
- Alter

Diese gilt es im weiteren Untersuchungsverlauf theoretisch zu fundieren und bereits bestehende Modelle in der Akzeptanzforschung zu modifizieren, so dass im Rahmen einer quantitativen Untersuchung (vgl. Abb. 1) weitere relevante Erkenntnisse zu Einstellungs- und Verhaltensdimensionen gewonnen werden können. Diese können zukünftig dabei helfen, entsprechende Angebote für Ehrenamtliche zu schaffen, die diese bei der Ausführung ihre Tätigkeiten und Aufgaben unterstützen

Literaturverzeichnis

- [Aj91] Ajzen, I.: The theory of planned behavior. In: *Organizational Behavior and Human Decision Processes*, 50, S. 179–211, 1991.
- [AF80] Ajzen, I.; Fishbein, M.: *Understanding Attitudes and Predicting Social Behavior*. Englewood Cliffs, NJ, Prentice-Hall, 1980.
- [AK21] Arriagada, C., Karnick, N.: Motive für freiwilliges Engagement, Beendigungsgründe, Hindernisgründe und Engagementbereitschaft. In: J. Simonson, N. Kelle, C. Kausmann & C. Tesch-Römer (Hrsg.): *Freiwilliges Engagement in Deutschland – Der Deutsche Freiwilligensurvey 2019*. Deutsches Zentrum für Altersfragen, Berlin 2021, S. 125 – 144, 2021.
- [CI10] Churchill, G. A., Iacobucci, D.: *Marketing research: methodological foundations*. 10. Auflage, SouthWestern Cengage Learning, 2010.
- [Da89] Davis, F. D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* 13, 319–340, 1989.
- [De83] Deshpande, R.: Paradigms Lost: on theory and method in research in marketing. *The Journal of Marketing*. Vol. 47, 4, 1983.
- [DB20] Deutscher Bundestag: *Dritter Engagementbericht. Zukunft Zivilgesellschaft: Junges Engagement im digitalen Zeitalter und Stellungnahme der Bundesregierung* (Drucksache 19/19320). Berlin, Deutscher Bundestag, 2020.
- [Fl07] Flick, U.: *Qualitative Sozialforschung. Eine Einführung*. 6. Auflage. Reinbek bei Hamburg, Rowohlt, 2007.
- [GL10] Gläser, J., Laudel, G.: *Experteninterviews und qualitative Inhaltsanalyse. Als Instrument rekonstruierender Untersuchungen*. 4. Auflage. Verlag für Sozialwissenschaften, Wiesbaden, 2010.
- [HS17] Hagen, C., Simonson, J.: Inhaltliche Ausgestaltung und Leitungsfunktionen im freiwilligen Engagement. In: J. Simonson, C. Vogel & C. Tesch-Römer (Hrsg.) *Freiwilliges Engagement in Deutschland – Der Deutsche Freiwilligensurvey 2014*, S. 299–331, Wiesbaden, Springer VS, 2017.
- [HBS19] Heinze, R. G., Beckmann, F., Schönauer, A.-L.: Die Digitalisierung des Engagements: Zwischen Hype und disruptivem Wandel. In: R. G. Heinze, S. Kurtenbach & J. Überbacher (Hrsg.) *Digitalisierung und Nachbarschaft. Erosion des Zusammenlebens oder neue Vergemeinschaftung?*, S. 61–90, Baden-Baden, Nomos Verlag, 2019.
- [Hi14] Hinz, U., Wegener, N., Weber, M., From, J.: *Digitales Bürgerschaftliches Engagement*. Berlin: FOKUS. Online: [https://www.oeffentliche-it.de/documents/10181/14412/Digitales+B%C3%BCrgerschaftliches+Engagement\(zuletzt abgerufen 11. Februar 2021\)](https://www.oeffentliche-it.de/documents/10181/14412/Digitales+B%C3%BCrgerschaftliches+Engagement(zuletzt abgerufen 11. Februar 2021)), 2014.
- [HS05] Hsieh, H. F., Shannon, S. E.: Three approaches to qualitative content analysis. Qualitative health research. Vol. 15, 9, 2005.
- [Hy00] Hyde, K. F.: Recognising deductive processes in qualitative research. *Qualitative Market Research: An International Journal*, Vol. 3, 2, 2000.
- [KH21] Kausmann, C., Hagen, C.: Gesellschaftliche Bereiche des freiwilligen Engagement. In: J. Simonson, N. Kelle, C. Kausmann & C. Tesch-Römer (Hrsg.): *Freiwilliges Engagement in Deutschland – Der Deutsche Freiwilligensurvey 2019*. Deutsches Zentrum für Altersfragen, Berlin, S. 95 – 122, 2021.

- [Ka21] Kausmann, C., Kelle, N., Simonson, J., Tesch-Römer, C.: Freiwilliges Engagement – Bedeutung für Gesellschaft und Politik. In: J. Simonson, N. Kelle, C. Kausmann & C. Tesch-Römer (Hrsg.): Freiwilliges Engagement in Deutschland – Der Deutsche Freiwilligensurvey 2019. Deutsches Zentrum für Altersfragen, Berlin, S. 319–324, 2021.
- [Ku18] Kuckartz, U.: Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. Weinheim; Basel: Beltz Juventa, 2018.
- [La19] Lau, Y., Fang, L., Cheng, L. J., Kwong, H. K. D.: Volunteer motivation, social problem solving, self-efficacy, and mental health: a structural equation model approach. *Educ. Psychol.* 39, 112–132, 2019.
- [LLM20] Li, X., Liu, X., Motiwalla, L.: Valuing Personal Data with Privacy Consideration. *Decision Sciences*, 2020.
- [LP14] Lienhard, P., Preuss, H.: Legacy, Sustainability and CSR at Mega Sport Events: An Analysis of the UEFA EURO 2008 in Switzerland. s.l.: Springer, Wiesbaden, 2014.
- [LL10] Liu, F., Lee, H. J.: Use of social network information to enhance collaborative filtering performance. *Expert Syst. Appl.* 37, S. 4772–4778, 2010.
- [Ma05] Mack, N., et al.: Qualitative research methods: a data collectors field guide. North Carolina: Family Health International, 2005.
- [Ma96] Marshall, M. N.: Sampling for qualitative research. *Family practice*. Vol. 13, 6, 1996.
- [MN02] Meuser, M., Nagel, U.: ExpertInneninterviews - vielfach erprobt, wenig bedacht: ein Beitrag zur qualitativen Methodendiskussion. In: D. Garz, & K. Kraimer(Hrsg.), Qualitativ-empirische Sozialforschung: Konzepte, Methoden, Analysen. Westdt. Verlag, Opladen. S. 441-471, 2002.
- [Mo91] Morse, J. M.: Approaches to qualitative-quantitative methodological triangulation. *Nursing research*. Vol. 40, 2, 1991.
- [Mo08] Montano, D. E., et al.: Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. *Health Behavior: Theory, Research, and Practice*. s.l.: John Wiley & Sons, 2008.
- [NdC16] Nader, L., de Campos, J.: Five Reasons to Fear Innovation. *SUR* 23 - v.13 n.23, S. 211 - 221, 2016.
- [Ng08] Nguyen, A.: FACING “THE FABULOUS MONSTER”, *Journalism Studies*, 9:1, S.91-104, 2008.
- [Pa90] Patton, M. Q.: Qualitative Evaluation and Research Methods. Newbury Park, Sage, 1990.
- [QD11] Qu, S. Q., Dumay, J.: The qualitative research interview. *Qualitative Research in Accounting & Management*. Vol. 8, 3, 2011.
- [Ra08] Raithel, J.: Quantitative Forschung. Ein Praxiskurs. Springer, Wiesbaden, 2008.
- [RKS21] Robinson, J.A., Kocman, D., Speyer, O., Gerasopoulos, E.: Meeting volunteer expectations - a review of volunteer motivations in citizen science and best practices for their retention through implementation of functional features in CS tools, In: *Journal of Environmental Planning and Management*. 2021.

- [SPV20] Saura, J. R., Palos-Sanchez, P., Velicia-Martin, F.: What drives volunteers to accept a digital platform that supports NGO projects? In: *Frontiers in Psychology*, 2020.
- [Sc14] Schmitt, J.: Social Innovation for Business Success: Shared Value in the Apparel Industry. s.l.: Springer Science & Business Media, 2014.
- [Si21] Simonson, J., Kelle, N., Kausmann, C., Tesch-Römer, C.: Unterschiede und Ungleichheiten im freiwilligen Engagement. In: J. Simonson, N. Kelle, C. Kausmann & C. Tesch-Römer (Hrsg.): *Freiwilliges Engagement in Deutschland – Der Deutsche Freiwilligensurvey 2019*. Deutsches Zentrum für Altersfragen, Berlin, S. 67-94, 2021.
- [Sv11] Svendsen, G. B., Johnsen, J. A. K., Almås-Sørensen, L., Vittersø, J.: Personality and technology acceptance: the influence of personality factors on the core constructs of the technology acceptance model. In: *Behavior & Information Technology*, 99(1), 2011.
- [TRH21] Tesch-Römer, C., Huxhold, O.: Nutzung des Internets für die freiwillige Tätigkeit. In: J. Simonson, N. Kelle, C. Kausmann & C. Tesch-Römer (Hrsg.): *Freiwilliges Engagement in Deutschland – Der Deutsche Freiwilligensurvey 2019*. Deutsches Zentrum für Altersfragen, Berlin, S. 194–213, 2021.
- [TAS16] Thakur, R., Angriawan, A., Summey, J.H.: Technological opinion leadership: The role of personal innovativeness, gadget love, and technological innovativeness, In: *Journal of Business Research*, Volume 69, Issue 8, S. 2764-2773, 2016.
- [VST17] Vogel, C., Simonson, J., Tesch-Römer, C.: Freiwilliges Engagement und informelle Unterstützungsleistungen von Personen mit Migrationshintergrund. In: J. Simonson, C. Vogel & C. Tesch-Römer (Hrsg.) *Freiwilliges Engagement in Deutschland – Der Deutsche Freiwilligensurvey 2014*, S. 601–634, Wiesbaden, Springer VS, 2017.
- [ZG07] Zhang, W., Gutierrez, O.: Information Technology Acceptance in the Social Services Sector Context: An Exploration, *Social Work*, Volume 52, Issue 3, July 2007, S. 221–231, 2007.
- [ZRL09] Zhang, J., Reithel, B. J., Li, H.: Impact of perceived technical protection on security behaviors. *Information Management & Computer Security*. Vol. 17, 4, 2009.

Content and Ad Formats in Performance Marketing Campaigns

Brand Usage Analysis of Advertising Options on Facebook

Maria Margarita Cabrera Frias¹ and Peter Konhäusner²

Abstract: The digital marketing industry is on the rise, with Facebook being one of the main advertising platforms for businesses to reach potential customers and drive results such as purchases, app downloads and registrations. In order to run a successful campaign, the appropriate kind of content is necessary to trigger the customer - still, the question for a suitable mix of content and ad format remains. Although research on online marketing has increased, there is a gap in the research of content use in performance marketing in relation to the business objectives being pursuit. This paper uses an inductive empirical method based on grounded theory with data gathered from 10 of the largest advertisers on Facebook ads in Germany, to analyze and categorize the different content and format strategies used by companies to drive performance objectives. As a result, a framework regarding content and ad format in Facebook ads for practical usage is derived.

Keywords: digital marketing, online marketing, performance marketing, paid social, social media advertising, Facebook ads, content, ad format

1 Introduction

With technological development, online marketing has become a legit way of businesses to leverage the internet in order to reach their audiences and attract potential customers [Am21]. As a consequence, in 2020 the internet was regarded as the most important channel for advertisers, reporting more than half of the total media ad spend, with ad spending on mobile doubling in the United States between 2019 and 2020 [St21] and the digital advertising spending worldwide reaching 378 billion U.S. dollars in 2020. The high spend of companies on digital marketing has been driven by its ability to reach users globally, the lower costs in comparison with traditional marketing communication, and its traceability allowing marketers to have a clear view on the effectiveness of their campaigns. This has also led to a raise in the paid social segment, also known as social media advertising, which

¹ bbw Hochschule, Leibnizstraße 11-13, 10625 Berlin, Germany, margarita.cabrera023@gmail.com, <https://orcid.org/0000-0002-2446-5430>

² GISMA Hochschule, Digital Entrepreneurship, Konrad-Zuse-Ring 11, 14469 Potsdam, peter.konhaeusner@gisoma.com, <https://orcid.org/0000-0001-6717-1304>

has grown to represent close to 35% of the value of the entire digital marketing industry [St21, St22b], aiming to achieve performance targets such as user acquisition, by using precise targeting and segmentation [PM16].

Despite the vast amount of ad spend being invested in online advertising, paid social and Facebook ads, research lacks a clear overview over the advertising methods regarding content and format which companies are using across industries. Therefore, to close this apparent research gap, the research question addressed in this paper focuses on analysing the usage of different advertising options on Facebook, and the ways that companies are making use of them, creating a theoretical framework as a result, suggesting a best practice regarding ad content and format in relation to business objectives. To address this matter, the inductive empirical method of grounded theory [GS67] is being used. Data sets from ten different companies, which use Facebook advertising, including Amazon, Spotify and Deutsche Telekom, were collected and analysed to categorize in different strategies of content and ad format usage, sourcing the data from the Facebook Ad Library. After the data was compiled and structured, clusters were created by grouping companies together showing similar characteristics, resulting in a theoretical approach of categorization regarding typical characteristics of ad format strategies using different content structures to reach their audience.

This article commences with theoretical research regarding online marketing, performance marketing, paid social, the Facebook ads algorithm and ad formats. Following background theory is an explanation on the research question and methodology used during this research. Finally, the data is presented, together with the results and a conclusion on the topic.

2 Background and Literature Review

2.1 Online Marketing

The American Marketing Association defines marketing as “the activity, set of institutions, and processes for creating, communicating, delivering, and exchanging offerings that have value for customers, clients, partners, and society at large” [Am17]. On the other side, Chaffey, Ellis-Chadwick, Mayer and Johnson [Ch06] present online marketing as “achieving marketing objectives through applying digital technologies”. As explained by Piñeiro-Otero and Martínez-Rolán [PM16], digital marketing has become its own phenomenon, bringing together customization and mass distribution, developing new channels, formats and strategies, and primarily being user-centred, more measurable and interactive. Kotler, Kartajaya and Setiawan [KKS17] describe how traditional marketing plays a major role in building awareness, but as the customers demand a closer relationship with companies, digital marketing becomes more important than ever, managing not only to reach customers, but also to drive results while developing advocacy.

As online marketing becomes more relevant for companies, the investment going into this form of marketing has also significantly increased year over year. “It was calculated that the digital advertising spending worldwide amounted to 378 billion U.S. dollars in 2020, increasing and not contracting versus the 2019 figure in spite of the economic impact of the Covid-19 pandemic” [St21]. The Covid-19 pandemic also led people to stay at home, causing a significant increase in online traffic, giving the opportunity for companies to target their audience through digital marketing initiatives and shifting traditional marketing budgets towards digital marketing [Gl21a]. It is expected that online marketing will continue to grow at an estimate COGR of 15% in the years to come, reaching 455 billion U.S dollars in 2022 and 645 billion U.S dollars in 2024. The online marketing sector has become so relevant that in 2020, “the internet was considered the most important medium for advertisers, accounting for 51 percent of total media ad spend”, with mobile ad spending doubling in the United States between 2019 and 2020 [St21].

2.2 Performance Marketing

Performance marketing represents an approach in the sphere of digital marketing where businesses pay when a specific result has been reached, such as a new lead or sale [Be21]. Berry explains how “the main thing that separates performance marketing from other types of marketing is the way businesses pay for their campaigns. Rather than paying up front or by month, advertisers pay when a specific action takes place”. Performance marketing therefore focuses on maximizing outcomes, mainly lead generation, purchases, or application downloads. In this light it is possible to also understand that “performance marketing campaign goals are typically shorter-term and focus on easy-to-measure metrics like leads generated rather than goals such as increased brand awareness” [Be21]. This measurability and ability to drive specific results are very appreciated by managers and advertisers, which is why 90% of all media spend in digital marketing is currently directed to this type of campaigns [Ed10]. Performance marketing focuses on quantifiability and a performance-oriented approach, “measuring the impact of online advertising, adapting it as needed, and making it more effective”, being ideal for user acquisition and retention [IO19]. The measurability of performance marketing can be tracked through the entire online user journey, from the moment someone watches an advertisement, clicks on a banner or purchases a product.

2.3 Paid Social

Paid social, also known as Social Media Advertising, refers to paid communication on social media platforms, such as Facebook or Twitter, using detailed targeting and segmentation, and optimizing towards goals such as website traffic, app downloads or purchases [PM16]. Paid social uses mainly videos and images that link to a website or app to transmit their messages and trigger specific results. On the other side, social media marketing (not

advertising) refers to organic (unpaid) content from companies such as pages, posts and comments, with the aim of engaging with users, helping to create brand awareness, and sharing content that could be interesting to users. Social media advertising (paid social), uses the platform to target specific user audiences, makes use of content that appears to be part of the platform, and drives particular performance goals such as lead generation, purchases or app downloads [Ma19]. According to Durante [Du21] this type of advertising has three big benefits, hyper-targeting customers, being able to build audiences and tracking ROI (return on investment). Hyper-targeting refers to the data availability on social media platforms, where marketers can target users based on age, gender, location, interests, behaviours, website visitors, email lists and more. This allows advertisers to create highly effective marketing campaigns, reaching the best fitting audiences. “Showing ads to the right individuals is a critical part of the marketing process and must be taken seriously to avoid wasted ad spend” [Du21]. On top of being able to target specific audiences, through paid social advertisers can save and mix audiences, creating new audience groups, but also being able to retarget them in the future automatically. Finally, one of the main tools available through social media advertising is the highly powerful reporting, being capable to track ”budget’s performance and analyse return on ad spend (ROAS), cost per website purchase, cost per lead and much more at a quick glance” [Du21].

Facebook has been a pioneer in the paid social industry, being one of the first ones to develop high level targeting and segmentation options, together with introducing a new targeting approach called lookalike matching, consisting on using deep third party data to algorithmically identify behaviours and find potential customers that exhibit the same “desirable” behaviours [SLR21]. On the social network, advertisers can also choose the objective of their campaign, choosing among eleven targets in the high, mid and low funnel, helping Facebook know which kind of people should be targeted, and supporting companies on getting the best results. With Facebook Ads companies can also use very granular segmentation, choosing their location, gender and other demographics, but also interests, connections to other users or past behaviours [Fa22a]. Finally, advertisers can choose a variety of formats to run, including links, images, videos, and dynamic advertisements, with which users can directly interact with a company’s inventory without even the displayed ad post. All of these types of ads can be delivered not only on the Facebook app, but also on Instagram and the Network Audience, which are apps and sites that have advertising connections with Facebook [Fa22a].

2.4 Advertising Content

No advertising campaign can be launched without content. “Content can play a vital role in driving awareness, traffic, sales, and reputation throughout the customer journey to create a comprehensive experience” [Ca17]. Images and videos are the main kinds of content utilized in commercials and online advertising. They both aim to share a message, connecting and engaging with an audience. “Images communicate thought and emotion whilst communicating a narrative in a way which words on their own can sometimes lack” [Sm20].

There are several benefits of using images in online advertising. In the first place, images can be eye catching, grabbing people's attention when using social media or the internet [Gl21b]. Images can also help to transmit a message almost immediately, without the need of lengthy explanations. Finally, images are easier, faster and cheaper to produce, allowing designers and creatives to generate the content without too much time, shooting or editing [Gl21b]. Images have also been seen by Facebook as the best content type if companies want to drive traffic to their website [Fa14]. Videos on the other hand can transmit more information than an image can, being able to explain a functionality or how to acquire a product. Videos can also be more engaging than images, keeping people's attention and getting more clicks. On the downside, videos need significantly more resources both in terms of time and money, which can be a limitation for small businesses [Gl21b]. It is important to note that in online advertising there is no need to choose one over the other, as Facebook research shows the benefits of combining content, as when an audience first sees an image, followed by a video, online conversions can increase around 30% [Fa14].

There are several formats that advertisers can use on Facebook in order to reach their target audience, having four big categories, images, videos, carousels and collections. The last two include images or videos, together with other ad components that can support companies meet their business goals [Fa22b]. On the first hand there are image ads, which are best used to "drive people to destination websites or apps through high-quality and engaging visuals" [Fa22b]. Businesses can use images to show their products or share their brand personality. Video ads are one of the preferred formats, as they allow marketers to show product features and display the brand closer. By using videos, advertisers can include sound and movement to catch people's attention, show the product's characteristics or tell their brand story. Carousels are an interactive format that can showcase up to ten videos or images in a single ad. Marketers can use this to share multiple features of a product, or even to show multiple products and link to each of them individually, within the same ad. The images or videos in the carousel can also come from the company's data base. Thanks to an integration of Facebook with companies' data bases, which allows them to connect to the product feed, showing specific products available on stock, this is known as a dynamic ad [Fa22b]. Finally, there are collection ads, which are one of the newest ad formats in the Facebook repertoire. Collection ads can include a video or image as the main component, and when the user taps on it, they can discover an immersive experience by scrolling through multiple products as if they were in the online shop, thanks to the dynamic ads benefits. These different formats are available in a similar way on the whole Meta family of platforms, including Instagram, Messenger and WhatsApp, together with partner apps and websites of the Audience Network [Fa22b].

3 Research Method

This paper aims to understand and research a best practice how companies are using content and ad format in paid social and Facebook ads in order to achieve their performance goals. As this perspective is a new and innovative approach, a method which examines phenomena and tries to extrapolate them to a bigger scope, is needed. This research uses the inductive empirical method of grounded theory [GS67] to empirically uncover the strategies used by the selected companies at the moment of running digital advertisements, under the assumption that these companies are acting logically, are optimizing their approaches and are trying to be as efficient as possible.

The present paper uses a qualitative research method, aiming to create a categorization of the strategies used by the chosen businesses regarding ad content. In the first step, an intentional selection of ten companies that use Facebook ads as a channel for their advertising efforts is done. The selection of brands to take part on the study was based on a research conducted on top Facebook advertisers in Germany based on their level of investment, with the highest investors being included in this paper [We21]. Not all the organizations mentioned in the study could take part in this research as they were no longer running advertisements on Facebook at the time of this research, wherefore two additional businesses were selected on convenience sampling basis (shown here with an asterisk). The focus on larger companies for this research is intentional, as the higher budgets can derive on faster learnings, more resources for analysis and sophisticated structured campaigns and use of best practices, which could be of use also for smaller organizations in terms of content.

#	Name	#	Name
1	Wix	6	Sky
2	Penny	7	Vodafone
3	Amazon	8	Deutsche Telekom
4	Spotify	9	Lidl*
5	Aldi	10	Zalando*

Tab. 1: Companies selected for the research [Source: Own creation]

After the organizations have been selected, the data is collected via online research with the use of the Facebook Ad Library, in order to record the ads that are currently published. The Facebook Ad Library is an open to the public tool created by Facebook in their attempt to make their platform and the advertisements published on it more transparent [Fa22c]. According to Leerssen et al. [LDHV21], the Facebook Ad library documents a selection of ads that appear on Facebook, and although it has been criticized for data not being granular

enough, it does constitute a step towards advertising transparency. For this paper, a total of one hundred ads - ten ads per every of the ten companies - were collected using the Facebook Ads library for the month of February 2022 in Germany, creating a selection of companies that were running several advertisements. With the content of each company recorded, a description of them is then noted down with keywords and qualities that are later compared and consolidated into a table. Additionally, there will be a description of the overall content published by every company and linked to the different Facebook ad objectives and advertising products aiming at providing readers a better understanding of the advertisements being published by each company. At the end, the categorization supports the interpretation of the strategy for which the content has been used.

Thanks to the content classification this study is able to analyse the formats being used by companies. Using inductive analysis to build conclusions regarding the objectives towards which businesses are using paid social advertising in Facebook, together with identifying the best kind of content for the goals being pursued, managerial implications can be derived. This qualitative study aims to reveal the media status quo for Facebook ads, to systematically understand different paid social strategies, and provide a basis for further performance marketing research.

It is important to note that for this study purposive sampling has been used, and although this is in line with the system of the inductive empirical method of grounded theory, consequences could include that the selected companies to be studied might not be the best to answer the research question, be biased, or be not diverse enough to allow for general conclusions to be drawn. The method of grounded theory has received criticism in the past by some academics, based on being too empirical, allowing a high level of freedom to researchers, or claiming that it draws over-generalized conclusions. Nevertheless, in this paper the method is considered to be appropriate for qualitative research, providing ecological validity and being a standard in social studies.

This paper uses a purposive selection of ten companies in Germany for the period of February 2022, which is not statistically significant, and although it is not a prerequisite for grounded theory, it calls for readers' and researchers' discretion regarding the conclusions drawn, as they might not apply for other companies, industries, countries, or timeframes. It is also relevant to consider that for the companies running several advertisements, a selection of ten was made for the purpose of this paper.

4 Data

In order to understand the content and format used by companies in performance marketing regarding Facebook Ads, the following companies have been observed using the Facebook Ads Library tool. Ten of their ads which had been published by February 2022 were recorded, and the creatives were organized in a table, together with keywords regarding the content type and format being used. A summary of this can be found in Table 2, which includes not only the name of the studied company, but also the industry to which it belongs, the number of ads running at the moment, and keywords regarding format and objective.

#	Name	Industry	# Ads running	Content+Format	Objective
1	Wix	Technology	23	Images Links and Story	Sign Ups
2	Penny	Supermarket	380	Images and Videos Links, story, carousel and collection	App Downloads Branding Recruiting Store Visits
3	Amazon	Retail	1700	Images Links, carousel and collection	Online Purchases
4	Spotify	Entertainment	95	images and videos Links and Story	Download Sign Up
5	Aldi	Supermarket	980	Images and Videos Links, story, carousel and collection	Online Purchases Recruiting Store Visits
6	Sky	Entertainment	160	Images and Videos Links and Story	Sign Ups Content Views
7	Vodafone	Communication	270	images and videos Links and Story	Download Sign Up
8	Deutsche Telekom	Communication	1500	images and videos Links and Story	Sign Up Recruiting
9	Lidl	Supermarket	1100	Images and Videos Links, story, carousel and collection	App Downloads Branding Recruiting Store Visits
10	Zalando	Fashion Retail	480	Images and Videos Links, story, carousel and collection	Online Purchases App Downloads Recruiting

Tab. 2: Overview of Ad Research [Source: Own creation]

5 Results and Analysis

After the advertisements were recorded, they were organized in a table including ten of the advertisements that the studied companies were running in Germany at the time of February 2022. With the aggregated information, a table sorts the types of content, format and objective being used by each of the studied companies by percentage of how much each of the factors have been utilized.

Regarding content type, the most common were images with 56%, followed by videos with 29% and finally dynamic content with 16%. Wix, Vodafone and Deutsche Telekom were the companies using the highest ratio of images with 100%, 90% and 80% respectively. For video, the businesses in which advertising context it was seen more broadly were Spotify and Sky with 60% each, and Penny and Lidl with 40%. Finally, dynamic content was the least used content type with only 16% of the 100 recorded ads using it. Amazon was the company relying mostly on dynamic ads with 90% of their ads leveraging their data base to show the products available to customers and segmenting them in categories such as technology or clothes. Other businesses taking advantage of this technology are Aldi and Zalando. 30 percent of their ads comprise dynamic content.

With this information it is possible to interrelate certain types of content use and the industries to which the companies using them belong. The use of images is mainly seen in technology and communication organizations, as they can explain a concept in a single snapshot. Videos were often observed in entertainment businesses such as Spotify and Sky, as they are able to share a preview of the content they are trying to advertise. Finally, dynamic content is leveraged mostly by e-commerce companies such as Amazon and Zalando, as they are the businesses having most complete and comprehensive catalogues of products from which customers can directly purchase. Supermarkets form the industry which makes a balanced use of all the different content types, with both Aldi and Lidl having a very balanced portfolio of 40% images, 30% videos and 30% dynamic content.

When revising format, the most used type is links with 40% of the recorded ads using an image or video that links to an external site or app store and being placed on the feed of Facebook and Instagram. Deutsche Telekom, Wix, Lidl, Vodafone and Spotify are the companies using link ads the most, linking both to their website and to an app store in the case of promoting an app. The second most common format are stories with 23% of the measured ads, which are vertical images and videos placed mainly on Instagram stories. These are popular amongst young people and mostly link to a website or app store. Stories were mainly used by Sky with 50% of the recorded ads having vertical format, followed by Spotify and Vodafone with 40% each. Carousels and collections share a third place with 16% and 14% respectively. This kind of placements allows advertisers to have an interactive ad in which users can scroll back and forth to choose interesting content using both images and videos. Carousels and collections are also very useful linking to specific products and they can even use dynamic content by being connected to a business' catalogue. Carousels were mainly used by Amazon with 50% of their ads being carousels, followed by Zalando and Aldi with 30% each. Collection ads can only be placed using dynamic content

with an image or video at the top. These were highly implemented by Amazon with 50% of the recorded ads being collections, together with Zalando and Aldi with 35% and 30% respectively. The least used type of format were images and videos that don't link to any external site, as it doesn't allow users to take any particular action. Finally, Sky and Penny were the only companies using an image or video without linking to anywhere else.

Relating this information to other aspects, it is possible to find that link ads, which are the most broadly used can be employed by companies in any industry, linking both to websites and app stores and leveraging mostly images. Story ads are being mainly used by companies in the communications and entertainment industry and use mostly videos as they are more immersive and make the ad feel more native. Carousel and collection ads are common with e-commerce and supermarket companies such as Amazon, Zalando and Aldi, as it allows them to show variety and link to specific products as well as leveraging dynamic content. Finally, there is a small minority of ads not using any kind of links and being used mostly for non-performance campaigns that search to raise awareness and don't require an immediate reaction of users, such as Sky sharing that there is a new series or Penny encouraging people to get the Covid-19 vaccine.

Finally, the third category was ad objective, for which most of the companies were using sign ups to different (mostly paid) services as the main goal of their Facebook campaigns in 35% of the recorded cases. For Wix this was the only objective used, followed by Vodafone and Deutsche Telekom with 80% and Sky with 70% of the cases. Online purchases were the second most popular objective with 27% of the businesses looking to sell something from their online shop with Amazon having this as its sole objective, followed closely by Zalando with 80% together with Lidl and Aldi with 40% each. App download was the third most used objective being highly leveraged by Spotify with 90% of its ads being redirecting people to an app store to download the Spotify app, followed by Penny with 40% and Vodafone with 20%. Store visits represented 9% of the ads in the research being used by Aldi, Penny and Lidl. Finally, other objectives used in rare cases were recruiting, branding and content view, with less than 6% each, which in this case are considered non-performance objectives being in line with the performance definition of Thomas [Th00]

Most of the companies using sign up ads were in the technology and communication industries, as they have services where customers must register to use them. Most of the sign-up ads use image link ads that explain the service offered and link to a site with more information and where people can enrol.

Online-purchase ads were popular for e-commerce businesses such as Amazon and Zalando, leveraging carousel and collection ads, in which they use a high amount of dynamic content with the products that they are promoting.

App-download ads were mainly used by technology-related companies where the app is the main product. Other businesses using app-download ads are supermarkets and communication companies, which are diversifying offering apps as a way to connect with their customers and generate new revenue streams. For this type of ads, both images and videos

are highly used, mainly in link and story ads, linking to the app store. Store visits are a popular objective for supermarkets which not only persuade people to purchase online but to visit their physical stores. For these ads, images and videos are widely used with or without a link, or even in story format.

Content view and branding are other objectives that don't seem to require to redirect users to a new site, while recruiting ads link users to the corporate page of the company.

With this information and following the path of inductive empirical method of grounded theory, a second round of analysis of the sets of concepts, codes and keywords gathered is possible, allowing to develop a theory that helps answer the research question. The previous categorization of the recorded ads and organization with the use of keywords supports the understanding of how brands are using different types of content and ad formats in performance marketing, particularly on Facebook. This enables now the creation of a theory table that suggests the best practices regarding content and format in relation to an objective that a company might pursue with the use of Facebook ads. These findings can be found in Table 3, and represent the ultimate theoretical framework derived from this research. As stated, this is based on the assumption that due to the high budgets of these companies and the optimization capabilities of Facebook Ads, the ads being run by these companies are the best performing ones.

Objective	Format	Content
Sign Ups	Link Story	Images
Online Purchases	Carousel Collection	Dynamic
App Download	Link Story	Images Video
Store Visits	Link Image/Video Story	Images Videos
Non-Performance	Image/Video Link	Image Videos

Tab. 3: Findings Best Practices Facebook Ads [Source: Own creation]

This theoretical framework which can be derived out of the research reveals that the kind of objective that a company tries to reach with the use of ads highly impacts the type of format and content it uses. The reasoning could be linked to the kind and amount of information that needs to be shared with one ad, and where this ad should take the user.

In this sense, sign-up ads benefit from using link or story ads with an image where they can place only the most relevant information and allow them to redirect users to their website where they would find more information about the service they can register to.

Online-purchase ads focus on the product that is being sold. Connecting the business' webshop or even data warehouse to Facebook is possibly the best way of advertising only

those items which are available leveraging the latest ad technology with dynamic content. Showing these products in a carousel or collection format also supports the idea of variety and allows customers to find their item of choice.

App-download ads need to link to an app store for which they require link or story ads. Diversifying the type of content with images and videos helps to show different aspects of the app. Store visits can use different kinds of content and format depending on what companies are using to attract customers to the store, such as certain products, a promotion or their unique selling proposition, allowing them to have more flexibility at the time of ad design.

Finally, non-performance based ads also enjoy more options regarding format and content, probably due to having more focus on the message being shared on the ad than what happens after, not having to redirect users to a website or an app store.

This framework suggests that using certain types of content and format can be beneficial to reach specific performance objectives. The use of the formats and content types recommended by this framework show in the next step a reduction of time, effort and investment into the design of creatives, allowing companies to focus only on the best performing for the objective at hand. Therefore, the whole process is made more efficient. This approach, moreover, would improve the performance of campaigns by allowing the algorithm to already start distributing the best creatives, instead of running images or videos that are not optimized for the objective, delivering the message more efficiently, decreasing campaign learning times and most importantly, reducing transaction costs.

6 Discussion and Conclusion

In this paper, the relation between ad content, format and objective in paid social marketing has been discussed, winding up in a theoretical framework of best practices on content and format types in the context of business objectives, based on the current actions of the biggest advertisers on Facebook ads in Germany. The research uncovers key concepts on performance marketing, social media advertising and - more specifically - Facebook advertising. The value of this research becomes more relevant as the overall investment in online marketing and Facebook ads continues to grow [St21]. As ad spend grows, so does the importance to understand the best way to utilize the channel in order to reach potential users, create a relationship with them and drive results [PM16].

This study does a thorough revision and explanation of online marketing [Bo14], performance marketing [TWC10, CA01] and paid social [DL15, SLR21] topics in order to clarify concepts and understand differences and strategies within the sector, together with industry size research [St21, St22b] to bring awareness to the magnitude and relevance of the sector in today's economy. This research also analyses advertising platforms [Fa22d, Ti22, Tw22] to understand the scope and advertising capabilities of each of these deeper,

compiled with industry information from marketing organizations within the sector [IO19, PY22]. The topic of content is also discussed, taking the use of images and videos in digital advertising into consideration [Ca17, Sm20, GI21a] by going over the Facebook ads content and format possibilities within the platform [Fa22b].

This research studies the advertisements launched by ten of the largest advertisers in Facebook in Germany during February 2022, in the aim of understanding their approach and using inductive analysis to develop a theoretical framework of the best practices regarding ad content and format respective to business objectives in Facebook ads. In order to reach this purpose, the study follows an analysis approach in line with the grounded theory method, starting with purposed sampling based on the largest advertisers in Facebook ads in 2020 in Germany [We21], followed by a collection and organization of the data, in this case the ten ads for each of the ten selected companies. With the collected data there is a round of coding with keywords based on defined characteristics: content, format and objective, aimed at providing a full picture. This broad perspective allows for a second analysis in which correlations and patterns are identified, with which a theoretical framework can be inducted based on the observed data and answering the research question, in this case the best practice framework of content and format in Facebook ads.

The findings of this research give orientation to marketing managers regarding the best type of ad content and format for their campaigns depending on their business goal, decreasing uncertainty and the misuse of resources in time and investment on the design and creation of said content, which is an important limitation that advertisers in the performance marketing industry face [Sm20, GI21a].

The research presented in the thesis is limited to the highest spending companies in Germany for February 2022, which means that smaller companies, different industries, countries, and timeframes could reveal other outcomes. Future research could evaluate more companies from different industries and also discuss if the proposed framework is valid for different contexts, as well as deepening on the reasons for why said best practices have the influence they seem to have.

Bibliography

- [Am17] American Marketing Association (AMA): What is Marketing? — The Definition of Marketing — AMA. <https://www.ama.org/the-definition-of-marketing-what-is-marketing/>, Retrieved: 03/01/2022.
- [Am21] American Marketing Association (AMA): What is Digital Marketing? <https://www.ama.org/pages/what-is-digital-marketing/>, Retrieved: 25/01/2022.
- [Be21] Berry, S.: What Is Performance Marketing and Why Does It Matter? <https://www.webfx.com/blog/marketing/what-is-performance-marketing/>. WebFX, Retrieved: 10/01/2022.
- [Bo14] Bostanshirin, S.: Online Marketing Challenges and Opportunities, 2014.
- [CA01] Clark, B.; Ambler, T.: Marketing Performance Measurement: Evolution of Research and Practice. International Journal of Business Performance Management 3, 2001.
- [Ca17] Carr, D. J.: What do we talk about when we talk about Content (and media)? <https://djc1805.medium.com/what-do-we-talk-about-when-we-talk-about-content-and-media-8f5be0a909b6>. Medium, Retrieved: 10/01/2022.
- [Ch06] Chaffey, D. et al.: Internet marketing. Strategy, implementation and practice. Financial Times Prentice Hall, Harlow, 2006.
- [DL15] Dempster, C.; Lee, J.: The Rise of the Platform Marketer. Performance Marketing with Google, Facebook and Twitter, Plus the Latest High-Growth Digital Advertising Platforms. Wiley, Hoboken, New Jersey, 2015.
- [Du21] Durante, J.: Three Reasons You Should Be Advertising On Social Media In 2021. Forbes, 2021.
- [Ed10] Edelman, D.: Branding in the Digital Age: You're Spending Your Money in All the Wrong Places. <https://hbr.org/2010/12/branding-in-the-digital-age-youre-spending-your-money-in-all-the-wrong-places>, Retrieved: 1/9/2022.
- [Fa14] Facebook: Creative Combinations That Work. <https://www.facebook.com/business/news/insights/creative-combinations-that-work>, Retrieved: 24/01/2022.
- [Fa22a] Facebook: Advertising on Facebook. <https://www.facebook.com/business/ads>, Retrieved: 17/01/2022.
- [Fa22b] Facebook: Facebook Ads guide: Ad format specs & recommendations. <https://www.facebook.com/business/ads-guide>, Retrieved: 24/01/2022.
- [Fa22c] Facebook: Facebook Ad Library. https://www.facebook.com/ads/library/?active_status=all&ad_type=political_and_issue_ads&country=DE&media_type=all, Retrieved: 30/01/2022.
- [Fa22d] Facebook: About ad auctions | Facebook Business Help Centre. https://www.facebook.com/business/help/430291176997542?id=561906377587030&ref=fbb_ads, Retrieved: 18/01/2022.
- [Gl21a] Glatzhofer, S.: Video vs. image Ads: Which is better? <https://www.splento.com/blog/photography/video-vs-image-ads-which-is-better/>, Retrieved: 22/01/2022.
- [Gl21b] Global Industry Analysts Inc: Digital Advertising and Marketing - Market Study by Global Industry Analysts, Inc. <https://www.strategyr.com/market-report-digital-advertising-and-marketing-forecasts-global-industry-analysts-inc.asp>, Retrieved: 04/01/2022.

- [GS67] Glaser, B. G.; Strauss, A. L.: The discovery of grounded theory. Strategies for qualitative research. Aldine de Gruyter, New York, 1967.
- [IO19] IONOS: Performance marketing: measuring advertising success. <https://www.ionos.com/digitalguide/online-marketing/online-sales/performance-marketing/>, Retrieved: 14/01/2022.
- [KKS17] Kotler, P.; Kartajaya, H.; Setiawan, I.: Marketing 4.0. Moving from Traditional to Digital, Wiley, 2017.
- [LDHV21] Leerssen, P.; Dobber, T.; Helberger, N.; de Vreese, C.: News from the ad archive: how journalists use the Facebook Ad Library to hold online advertising accountable, Information, Communication & Society. DOI: 10.1080/1369118X.2021.2009002
- [Ma19] Marcee: Social Media Marketing vs. Social Media Advertising. <https://evenbound.com/blog/social-media-marketing-vs-social-media-advertising>, Retrieved: 15/01/2022.
- [OA07] O'Sullivan, D.; Abela, A. V.: Marketing Performance Measurement Ability and Firm Performance. Journal of Marketing 2/71, S. 79–93, 2007.
- [PM16] Piñeiro-Otero, T.; Martínez-Rolán, X.: Understanding Digital Marketing—Basics and Actions, in: Machado, C., Davim, J. P.: MBA: Theory and Application of Business and Management Principles, pp. 37–74, 2016.
- [PY22] PYMNTS: TikTok Sets 2022 Target of \$12B in Sales. <https://www.pymnts.com/news/social-commerce/2022/tiktok-sets-2022-target-of-12b-in-sales/>. PYMNTS.com, Retrieved: 15/01/2022.
- [SG15] Schwarzl, S.; Grabowska, M.: Online marketing strategies: the future is here. Journal of International Studies 8(2), pp. 187-196, 2015. DOI: 10.14254/2071-8330.2015/8-2/16
- [SLR21] Sudhir, K.; Lee, S. Y.; Roy, S.: Lookalike Targeting on Others' Journeys: Brand Versus Performance Marketing, Cowles Foundation Discussion Paper No. 2302, New Haven, Connecticut 06520-8281, 2021.
- [Sm20] Smith, S.: The Importance of Images in Marketing - Solve. <https://solve.co.uk/seo-tips/importance-of-images-in-marketing/>, Retrieved: 22/01/2022.
- [St21] Statista: Global digital advertising market 2024 | Statista. <https://www.statista.com/statistics/237974/online-advertising-spending-worldwide/>, Retrieved: 04/01/2022.
- [St22a] Statista: Social Media Advertising - Worldwide | Statista Market Forecast. <https://www.statista.com/outlook/dmo/digital-advertising/social-media-advertising/worldwide#key-players>, Retrieved: 17/01/2022.
- [St22b] Statista: CPM advertising cost for cable TV in the U.S. 2020 | Statista. <https://www.statista.com/statistics/826039/cable-tv-cpm/>, Retrieved: 07/01/2022.
- [Th00] Thomas, M. J.: Marketing performance measurement: Directions for development. Journal of Targeting, Measurement and Analysis for Marketing 9, 2000.
- [Ti22] TikTok: TikTok For Business. <https://www.tiktok.com/business/en/apps/tiktok>, Retrieved: 17/01/2022.
- [Tw22] Twitter: Twitter Advertising. <https://business.twitter.com/en/advertising.html>, Retrieved: 17/01/2022.

- [TWC10] Tonkin, S.; Whitmore, C.; Cutroni, J.: Performance Marketing with Google™ Analytics. Strategies and Techniques for Maximizing Online ROI. Wiley, 2010.
- [We21] Wells, R.: Top Facebook Advertiser Report: Germany. <https://www.pathmatics.com/blog/top-facebook-advertiser-report-germany>, Retrieved: 20/02/2022.

Anwendbarkeit von Enterprise Security Assessments sowie Enterprise Architecture Tools für KMU

Kendime Ismailji¹, Christof Mosler², Silvia Knittl³

Abstract: Die zunehmende Cyber-Kriminalität bedarf eines präventiven Sicherheitsansatzes der Enterprise Security Architecture (ESA). Besonders kleine und mittelständische Unternehmen (KMU) tun sich oft schwer damit, Maßnahmen zur Widerstandsfähigkeit gegenüber Cyber-Angriffen umzusetzen. In diesem Artikel stellen wir unseren ESA-Ansatz für KMU vor. Dafür wurden Anforderungen hinsichtlich der Informationssicherheit an KMU identifiziert, priorisiert und mit einem bestehenden ESA-Rahmenwerk verglichen. Das Ergebnis bildet ein für KMU angepasstes Framework. Das bereits bestehende Framework wurde von 340 Fähigkeiten auf 100 Fähigkeiten gekürzt. Der Einsatz des neuen Frameworks hat bereits gezeigt, dass ESA für KMU anwendbar und vorteilhaft ist. ESA-Assessments benötigen ein geeignetes Werkzeug, welches deren Durchführung und Verwaltung unterstützt und die Ergebnisvisualisierung automatisiert. Dazu stellen wir die Anforderungsanalyse vor. Die Umsetzung der Anforderungen in einem ausgewählten Tool hat gezeigt, dass dadurch der ESA-Prozess optimiert wird. Der Werkzeugeinsatz reduziert den Pflegeaufwand und ermöglicht dadurch eine effizientere Arbeitsweise.

Keywords: Enterprise Security Architecture; ESA; KMU; EA-Tool; ISMS.

1 Präventive Sicherheit durch Enterprise Security Architecture (ESA)

Informationssicherheit hat heutzutage eine besondere Bedeutung, denn durch die zunehmende Digitalisierung verschaffen sich Cyberkriminelle neue Einfallstore. Um sich vor Cyberangriffen zu schützen, sind besondere Maßnahmen hinsichtlich der Sicherheit von Unternehmenswerten zu beachten. Enterprise Security Architecture (ESA) bietet einen präventiven Sicherheitsansatz, indem Cyber-Fähigkeiten (engl. Cyber Capabilities) zur Bewertung des Sicherheitsstatus eines Unternehmens festgelegt werden, welche die Organisationsstruktur, deren Prozesse und Technologien gesamtheitlich betrachten. Dadurch können Lücken in der Sicherheitsarchitektur identifiziert und Handlungsmaßnahmen abgeleitet werden. Insbesondere für kleine und mittelständische Unternehmen (KMU) sind Rahmenbedingungen hinsichtlich Informationssicherheit und Datenschutz schwer einzuh

1 PwC, Cyber Privacy, Friedrichstr. 14, 70174 Stuttgart, kendime.ismailji@pwc.com

2 HFT Stuttgart, Wirtschaftsinformatik, Schellingstr. 24, 70174 Stuttgart, christof.mosler@hft-stuttgart.de

3 PwC, Cyber Privacy, Bernhard-Wicki-Str. 8, 80636 München, silvia.knittl@pwc.com,

<https://orcid.org/0000-0001-9507-8713>

halten, denn bekannte Rahmenwerke, wie bspw. die ISO 27001 [ISO13] oder die Datenschutzgrundverordnung (DSGVO), sind in der Wahrnehmung häufig für große Unternehmen und Konzerne ausgelegt. Cyberkriminelle unterscheiden jedoch nicht zwischen den Unternehmensgrößen. Daher ist es umso wichtiger auch für KMU einen geeigneten Sicherheitsansatz festzulegen.

1.1 Ziele der Arbeit und Vorgehensweise

Das Ziel der Arbeit ist zum einen den Anpassungsbedarf eines bestehenden, im praktischen Einsatz befindlichen ESA-Frameworks für KMU zu bewerten. Zum anderen sollen durch einen geeigneten Tool-Einsatz die derzeitige Assessment-Dokumentation und ihre Verwaltung vereinfacht und Ergebnisse visualisiert werden. Der methodische Aufbau zum Erreichen dieser Ziele orientiert sich am Design Science-Ansatz (vgl. Hevner in [BHM20]). Dieser Ansatz umfasst gem. Hevner drei Zyklen: den der praktischen Relevanz (Relevance Cycle), der den aktuellen Stand der Technik einbezieht, den des Designs (Design Cycle), in dem aus der Verknüpfung von Theorie und Praxis ein Artefakt abgeleitet wird, und den der Evaluation (Rigor Cycle).

Entsprechend dieses Ansatzes beschreiben wir die Relevanz anhand eines aus unserer Beratungspraxis entwickelten ESA-Rahmenwerks. Zum besseren Verständnis stellen wir dieses in Abschnitt 1.2 vor. Die Herausforderungen bei der praktischen Anwendung bei KMU stellen wir in Abschnitt 1.3 dar. Verwandte Arbeiten, die sich mit ESA für KMU befassen und die einen Beitrag für unser Design liefern, stellen wir in Kapitel 2 vor. In Kapitel 3 entwickeln wir daraus ein Konzept für die Anpassung unseres ESA-Rahmenwerks für KMU. Dafür identifizieren wir die Cyber-Security-Anforderungen für KMU in Abschnitt 3.1. und konzipieren damit ein KMU-taugliches Assessment-Vorgehen in Abschnitt 3.2.

Ein weiteres Ziel ist, die bisherige Assessment-Dokumentation und -Verwaltung zu vereinfachen und die Ergebnisse zu visualisieren. Hier analysieren wir das vorgegebene Tool LeanIX, welches bereits im Kontext Enterprise Architecture genutzt wird und nun auf seine Anwendbarkeit für ESA geprüft werden soll. Damit soll der bisherige ESA-Prozess und -Assessment durchgeführt und ausgewertet werden. Zudem sollen sämtliche Ergebnisse im Tool abgebildet werden können. Darunter zählen z. B. Modellierungen, Diagramme, Roadmaps und Reifegradbewertungen. In Abschnitt 3.3 stellen wir die Anforderungsanalyse an das Tool und die Erstellung des Assessments für KMU vor. Dieses Konzept prüfen wir mit qualitativen Methoden auf Anwendbarkeit in Abschnitt 3.4. Wir diskutieren die Ergebnisse in Kapitel 4, fassen diese zusammen und geben einen Ausblick über mögliche folgende Entwicklungsschritte.

1.2 Grundlagen: Enterprise Security Architecture

Enterprise Security Architecture (ESA) [Mc20] ist ein Top-Down-Ansatz, der mit der Identifikation der Strategie und der Unternehmensziele startet. Der ESA-Ansatz ist methodisch abgeleitet vom Enterprise Architecture (EA) Management-Framework TOGAF [Th09]. Im Mittelpunkt stehen die zu untersuchenden Fähigkeiten und Kompetenzen (engl. Capabilities) des Unternehmens. Durch die Orientierung an den Methoden der EA wird ein ganzheitlicher Schutz gewährleistet, bei dem Mitarbeiter, Gebäude, Anlagen, IT-Systeme und Informationen einer Organisation berücksichtigt werden.

1.3 Herausforderungen bei der Umsetzung von ESA

Im Mittelpunkt aller Aktivitäten steht ein ESA-Framework, welches bei PwC entwickelt wurde. Mithilfe dieses aus der Praxis entstandenen Rahmenwerks wird die Unternehmenssicherheitsarchitektur aufgebaut und bewertet. Das Framework selbst kann bei Bedarf angepasst und weiterentwickelt werden und besteht aus 11 Domänen, von Strategie, Führung und Governance, bis hin zur Physischen Sicherheit, wie sie in Abbildung 1 dargestellt sind. Die Domänen sind wiederum in weitere Sub-Domänen untergliedert. Jede Sub-Domäne besteht wiederum aus verschiedenen Fähigkeiten, die die Sub-Domänen unterstützen. Insgesamt werden 340 Fähigkeiten beschrieben. Sie werden jeweils für die drei Ressourcen Menschen, d.h. Ablauf- und Aufbau-Organisation (People), Prozesse (Processes) und Technologien (Products) analysiert. Dabei definiert man vier Architektsichten (von Konzeption bis Technische Architektur), die jeweils andere Blickwinkel auf die Domänen bieten und in ihrem Zusammenspiel jedoch kohärent und konsistent sein sollten. Im Rahmen von ESA-Assessments wird das Rahmenwerk als Referenz verwendet. Die jeweilige Ist-Ausprägung in den aufgezeigten Dimensionen und Fähigkeiten wird bei den Unternehmen analysiert. Die Analyse-Ergebnisse werden in verschiedenen Berichtsformaten manuell aufbereitet und zusammengestellt. Herausforderungen bei der Umsetzung von ESA

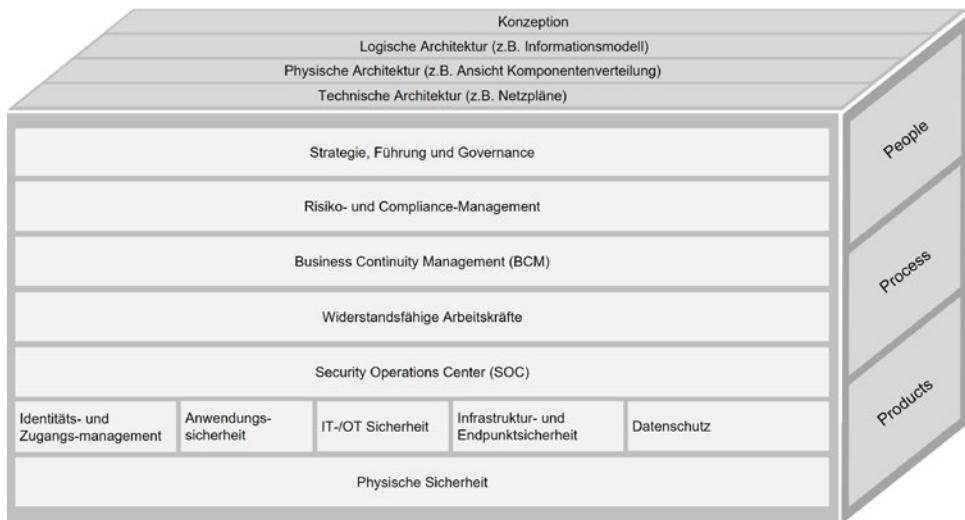


Abb. 1 Domänen und Dimensionen des ESA-Frameworks von PwC

Bisher wurden ESA-Assessments in großen Organisationen durchgeführt. Nun soll der Ansatz auch für KMU durchgeführt werden. Das dafür eingesetzte Framework ist umfangreich und detailliert, weshalb es insbesondere für Konzerne ausgelegt ist. Kleinere Unternehmen verfügen in der Regel nicht über die notwendigen Kapazitäten und das Know-how, um alle Aspekte dieser umfangreichen Assessments zu berücksichtigen. Hinzu kommt, dass inhaltlich viele der untersuchten Bereiche bei KMU vermeintlich gar keine Anwendung finden, da hier andere technische und juristische Anforderungen existieren. Daher ist es notwendig, das Assessment bzw. das ESA-Framework auf seine Anwendbarkeit für KMU zu bewerten und anzupassen.

Die ESA-Assessments unter Verwendung des ESA-Frameworks werden derzeit mittels Excel durchgeführt und dokumentiert. Das ESA-Framework besteht aus mehreren Domänen, die jeweils Sub-Domänen enthalten. Für jeden Kunden wird beim Scoping-Schritt ein individuelles Excel-Dokument erstellt, indem die Domänen des ESA-Frameworks entsprechend ausgewählt und das Framework individuell werden. Die Excel-Tabellen werden schnell unübersichtlich, Unwissende haben häufig Verständnisprobleme und verlieren den Überblick. Die Assessment-Auswertung wird ebenfalls in der Excel-Tabelle ergänzt und dokumentiert. Für den Kunden werden die ausgewerteten Beobachtungen und Empfehlungen mithilfe verschiedener anderer Tools (z.B. einer PowerPoint-Präsentation) visualisiert und berichtet. Ein Problem sind daher die Medienbrüche zwischen den einzelnen Phasen des Assessments. Werden Änderungen vorgenommen, so müssen diese manuell abgeändert und nachgezogen werden.

2 Verwandte Arbeiten

Die Vorteile von Ansätzen, die auf der Enterprise Architecture (EA) basieren, wurden in der Literatur bereits erläutert [Mc20]. Man erreicht damit eine ganzheitliche Sicht auf die Sicherheitsaspekte und involviert alle relevanten Unternehmensteile. Gleichzeitig werden auch strategische Ziele berücksichtigt. In dem erwähnten Artikel werden insgesamt 25 Security Frameworks untersucht, wobei sechs von ihnen auf EA basieren und jeweils komplett Organisationen betrachten: [Ho02], [Sc06], [RA08], [SLR09], [SCL09], [Je16].

Aus Sicht der Autoren der oben erwähnten Studie erfüllt jedoch keines dieser Ansätze die erforderlichen Anforderungen im Hinblick auf die Ganzheitlichkeit, wie sie EA zu liefern verspricht. Deshalb schlagen sie mit dem „Security Architecture Framework for Enterprises (SAFE)“ eine eigene Lösung vor, die genau wie unser Framework nach dem Design Science Research-Ansatz entwickelt wurde. Die Autoren beschreiben ihr Framework allerdings als sehr komplex und die bei der Entwicklung interviewten Teilnehmer stammen meistens aus großen Unternehmen. Der Einsatz dieses Frameworks im KMU-Kontext war nicht das Ziel der Forschung und wurde nicht ausreichend untersucht. Auch wurde bisher noch keine Fallstudie durchgeführt.

Andere Ansätze, die zwar nicht im Bereich der ESA anzusiedeln sind, praktisch jedoch ähnlich vorgehen (mit der Betrachtung von Schwerpunktbereichen und Fähigkeiten) basieren auf Reifegradmodellen. Einige davon fokussieren auf KMU und sind für uns deshalb besonders interessant: In 2014 wurde in [SR14] das „Information Security Focus Area Maturity Model (ISFAM)“ vorgestellt, dessen Assessments auf lediglich 161 Ja/Nein-Fragen basieren. Dieses Modell bietet dennoch eine ganzheitliche Sicht, bestehend aus 13 Schwerpunktbereichen (engl. Focus Areas), 51 Fähigkeiten (engl. Capabilities) und zwölf Reifegraden. Ähnlich wie in unserem ESA-Ansatz wurden diese verschiedenen Kategorien aus bekannten Security Frameworks abgeleitet (v.a. ISO 27000, CISSP [IS22] und aus dem IBM Security Framework [Bu13]). Zusätzlich wurden anhand der Abhängigkeiten von Fähigkeiten die passenden Reifegrade definiert.

In späteren Veröffentlichungen wurde das ISFAM-Modell aufgegriffen und erweitert. [Fr16] stellt mit dem Ansatz „Characterizing Organizations‘ Information Security for SMEs (CHOISS)“ eine Möglichkeit vor, das ISFAM-Modell unternehmensspezifisch anzupassen. Bei den Assessments soll jeweils ein Schwerpunkt auf die elf organisatorischen Merkmale (engl. Organizational Characteristics, kurz OCs) des jeweiligen Unternehmens gesetzt werden. Dadurch wird die Analyse deutlich zielorientierter. In [Yi20] wurde diese Idee weiterentwickelt, indem vorab definierte OCs für Cluster von ähnlichen Organisationen bei der Reifegradbetrachtung eingesetzt werden. In der vorgestellten Studie wurde beispielsweise eine Gruppe von lokalen Unternehmen der Transport-, Logistik- und Verpackungsbranche untersucht.

Die Idee, dass man vor der Assessment-Durchführung die Analyse auf das jeweilige Unternehmen anpassen muss, liegt auch unserem ESA-Ansatz zugrunde. Denn als ersten Schritt

bei der Anwendung des Modells wird immer das Scoping durchgeführt (siehe Abschnitt 1.3), bei dem man die Assessment-Domains für die Kunden individuell anpasst.

Es gibt außerdem weitere Ansätze, die helfen sollen, IT-Security in KMU zu etablieren: Eine Übersicht der vielen relevanten Standards liefert z. B. die Studie der European Union Agency for Network and Information Security (ENISA) [Ma15]. Basierend auf dieser Übersicht wird auch eine Maßnahmenliste speziell für KMU abgeleitet. Weitere solche Listen von Empfehlungen für KMU finden sich in zahlreichen Leitlinien und Handbüchern, wie z.B. [Gl20] und [Sa18]. Die Autoren dieser Publikationen weisen jedoch explizit auf die Vorteile und die Notwendigkeit von systematischen und ganzheitlichen Ansätzen hin. In [THM07] findet sich ein solcher einfacher Ansatz, mit dem Unternehmen ausgehend von ihren Sicherheitszielen in einem vierstufigen Prozess entsprechende Aktionen identifizieren, implementieren und überwachen können. Und in [ET21] findet sich eine weitere Studie zu Security-Frameworks für KMU, in der 17 einheitliche Kontrollkategorien für vier Typen von KMU-Organisationen identifiziert werden. Davon ausgehend wird ein Prozess zur Definition von Maßnahmen definiert.

Allen diesen Vorschlägen fehlt jedoch die ganzheitliche Sicht wie sie bei ESA oder den oben erwähnten ISFAM-Ansätzen zu finden ist. Insbesondere bleiben größtenteils die strategischen Unternehmensziele sowie die Organisationsarchitektur unberücksichtigt.

3 Konzeption ESA-Assessment für KMU mit Werkzeugunterstützung

Im folgenden Kapitel beschreiben wir unser Vorgehen zur Lösung der in Abschnitt 1.3 dargelegten Problemstellung.

3.1 ESA-Konzept für KMU

Es gibt verschiedene Normen für Sicherheitskontrollen, wie z.B. die ISO/IEC 27001. Die Umsetzung der Maßnahmen aus dieser Norm stellt für KMU jedoch oft eine Herausforderung dar. Daher gibt es auch speziell für KMU entwickelte Rahmenwerke. Dazu zählen z.B. CISIS12 [IT22] und VdS 10000 [Vd22]. Um das ESA-Framework für KMU anzupassen und ein entsprechendes Assessment durchführen zu können, war es notwendig, relevante Anforderungen hinsichtlich der Informationssicherheit zu identifizieren. Dafür wurden die bereits genannten Sicherheitsstandards ISO 27001, CISIS12 und VdS 10000 näher betrachtet, um herauszufinden, welche Anforderungen auf KMU zutreffen. Aus diesem Vergleich wurden die in den Rahmenbedingungen enthaltenen, identifizierten Anforderungen in neu zusammen gesetzten Domänen aggregiert. Um anschließend den Abdeckungsgrad der unterschiedlichen Rahmenbedingungen mit dem aktuellen ESA-Framework im

nächsten Schritt zu identifizieren und die entsprechende Anpassung vornehmen zu können, wurden die Anforderungen entsprechend der drei Ebenen des ESA-Frameworks als strategisch, taktisch/betrieblich und technisch klassifiziert. Aus dem Vergleich haben sich insgesamt 28 Anforderungen ergeben, die gleich zu gewichten und als Muss-Anforderungen zu betrachten waren. Ein Auszug der zusammengestellten Anforderungen ist in der Tabelle 1 zu sehen. Die Tabelle zeigt außerdem die Gegenüberstellung der Sicherheitsstandards und die Abschnitte, in denen die jeweiligen Anforderungen zu finden sind.

ID	ESA-Sub-Domäne	ISO 27001	CISIS12	VdS 10000
R11	Kontinuierliche Verbesserung	Kapitel 10	N/A	N/A
R12	Business Continuity Management	A.17	N/A	Kapitel 17
R13	Steuerung von Dienstleistern/ Lieferanten	A.15	N/A	Kapitel 4
R14	Umgang mit personenbezogenen Daten	A.18	N/A	N/A
R15	Identifikation von kritischen Anwendun-gen	A.8	Schritt 6	Kapitel 9 Kapitel 10
R16	Softwaresicherheit - Erkennung von Schadsoftware	A.12	N/A	Kapitel 10
R17	Umgang mit Informationssicherheits- vorfällen	A.16	Schritt 5	Kapitel 18
R18	Schwachstellenmanagement	A.12	N/A	Kapitel 10

Tab. 1 Übersicht: Mapping der Anforderungen der ISMS-Rahmenwerke (Auszug)

Aus dem Vergleich der Frameworks geht hervor, dass die Anforderungen der einzelnen Frameworks sich in den anderen widerspiegeln. Da CISIS12 mehr einen Prozess zum Aufbau eines ISMS darstellt, ist die Gegenüberstellung mit den einzelnen Rahmenbedingungen nicht vollständig. Anforderungen bezüglich des Datenschutzes (zum Schutz personenbezogener Daten) werden ausschließlich in der ISO 27001 beschrieben. Diese wurden dennoch den Muss-Anforderungen zugeordnet.

Die Anpassung des ESA-Frameworks erfolgte auf Grundlage der Gegenüberstellung der Rahmenbedingungen und der daraus aufgestellten Anforderungen. Für die Anpassung wurden die identifizierten Anforderungen aus der ISO 27001, CISIS12 und VdS 10000 mit dem vorhandenen ESA-Framework verglichen. Zur Schaffung eines Überblicks des Abdeckungsgrades der Anforderungen aus der ISO 27001, CISIS12 und der VdS 10000 diente ein Mapping der einzelnen Frameworks auf das ESA-Framework. Das ESA-Framework ist in einer detaillierten Excel-Datei dokumentiert. Strukturiert ist das Framework in Form einer Tabelle, in der die einzelnen Domänen, Subdomänen und Capabilities die einzelnen Spalten bilden. Die enthaltenen Domänen sind in der Abbildung 1 aufgeführt. Diese Domänen werden wiederum unterteilt in Sub-Domänen, die von mehreren Fähigkeiten unterstützt werden. Das Mapping erfolgte somit in der Excel-Tabelle des ESA-Framework. Für jedes gegenübergestellte Framework wurde in der Excel-Tabelle eine neue Spalte eingefügt, in der die passenden Controls den ESA-Fähigkeiten gegenübergestellt

und so gemappt wurden. Dabei wurden die einzelnen Controls der Frameworks (sowohl für die ISO 27001 als auch für die VdS 10000) näher betrachtet und so die Inhalte mit den ESA-Fähigkeiten verglichen. Damit erfolgte die Prüfung der Konformität der einzelnen Frameworks mit dem ESA-Framework. Das Mapping bildete die Grundlage zur Anpassung des ESA-Frameworks für KMU. Anhand des Mappings und der ausgearbeiteten Anforderungen wurden die Fähigkeiten des ESA-Frameworks mit den Schlüsselwörtern Muss, Soll und Kann wie folgt priorisiert. Muss: Diese Anforderungen sind Mindestanforderungen und somit als verpflichtend anzusehen und müssen zwingend erfüllt werden. Die Anforderungen sind in mindestens zwei der verglichenen Rahmenbedingungen enthalten. Soll: Die Umsetzung dieser Anforderungen wird empfohlen, ist jedoch nicht zwingend notwendig. Sie dienen der Erweiterung der Mindestanforderungen und sind in mindestens einem der verglichenen Rahmenbedingungen enthalten. Kann: Optional umzusetzende Maßnahmen werden mit dem Schlüsselwort „kann“ gekennzeichnet. Diese Anforderungen sind wünschenswert, jedoch von untergeordneter Bedeutung und in der Detailtiefe zu spezifisch für KMU. Diese Anforderungen sind in maximal einem der verglichenen Rahmenbedingungen aufgeführt.

Die Priorisierung erfolgte entlang der einzelnen Fähigkeiten. Die in Tabelle 1 aufgeführten Anforderungen wurden dabei als Muss-Anforderungen aufgenommen. Zur Priorisierung der Sub-Domänen wurden die Summen der Muss-Fähigkeiten gebildet. Aus der Summe der Muss-Fähigkeiten und der Sub-Domänen ergibt sich in Tabelle 2 die aufgezeigte Priorisierung der Domänen.

Damit wurde das ESA-Framework von 340 auf 100 Fähigkeiten gekürzt. Anhand der Priorisierung und des Mappings entfallen die Domänen Security Orchestration und IT/OT-Security für das KMU-Framework. Der Fokus sollte auf die zuvor priorisierten Muss-Domänen gelegt werden. Diese Domänen beinhalten grundlegende Fähigkeiten, wie den Aufbau einer Organisation, die Definition einer Sicherheitsstrategie, die Festlegung von Zielen und den Aufbau eines Informationssicherheits-Teams. Zudem sollte auch Wert auf Awareness-Maßnahmen und die Sensibilisierung von Mitarbeitenden gelegt werden. Auch die Sub-Domänen Asset Management, Business Continuity Management (BCM), Schwachstellenmanagement, Sicherheitsvorfälle und Physische Sicherheit sind wesentliche Bestandteile des ESA-Frameworks, die auch in den verglichenen Frameworks verankert sind. Sie gelten somit als Mindestanforderungen und haben auch für KMU hohe Relevanz in der Gewährleistung der Unternehmenssicherheit.

Nr.	Domäne	Priorität	Sub-Domäne (muss)	Capabilities (muss)
1.1	Security Strategy & Leadership	muss	3	6
1.2	Risk & Compliance Management	muss	4	21
1.3	Security Resilient Architecture	muss	1	7
1.4	Resilient Workforce	muss	2	5
2.1	Cyber Defence	muss	3	17

2.2	Security Orchestration	kann	0	0
3.1	Identity & Access Management	muss	4	13
3.2	Infrastructure & Endpoint Security	muss	3	8
3.3	Application Security	soll	1	2
3.4	Data Protection & Privacy	muss	3	12
3.5	IT-OT Security	kann	0	0
3.6	Physical Security	muss	2	9
		Summe:	26	100

Tab. 2 Priorisierung der Anforderungen

Aus dem Mapping der Standards mit dem ESA-Framework geht hervor, dass alle Bereiche der ISO 27001 und der VdS 10000 weitestgehend umfassend vom ESA-Framework abgedeckt werden. Da CISIS12 eher einen Prozess als Anforderungen beschreibt, sind nicht alle Schritte zuordenbar. Die DSGVO beschreibt umfassend, wie mit personenbezogenen Daten umzugehen ist, wodurch auch nur ein kleiner Teil der Verordnung im Framework vorkommt.

3.2 Bewertung des ESA-Konzeptes für KMU

Zur Bewertung der ausgearbeiteten Anforderungen hinsichtlich der IT-Sicherheit in KMU und des angepassten ESA-Frameworks wurden zwei unterschiedliche praktische Anwendungsfälle betrachtet. Im ersten Fall wurde ein ESA-Assessment auf Grundlage des angepassten ESA-Frameworks mit einem mittelständischen Unternehmen durchgeführt. Zur Identifizierung des Ist-Zustands und somit der aktuellen Reifegrade der Unternehmensfähigkeiten im Bereich Informationssicherheit wurden dabei zwei Interviews mit dem Kunden durchgeführt. Im ersten Interview lag der Fokus auf der Analyse und der Bewertung der internen Prozesse. Dabei wurden Domänen wie Informationssicherheitsorganisation, Dienstleistersteuerung, Umsetzung von Awareness-Maßnahmen, Datensicherheit, Physische Sicherheit, Gebäude- und Endgerätesicherheit und Outsourcing betrachtet. Das zweite Interview wurde mit einem der IT-Dienstleister des Unternehmens durchgeführt und fokussierte sich somit auf die IT-Prozesse, darunter Domänen wie Fernzugriff, Wartung und Monitoring. Auf Basis der Beobachtungen in den beiden durchgeföhrten Interviews wurde eine Reifegradbewertung erstellt und visualisiert (vgl. Abbildung 2). Diese bildete die Grundlage für die Definition entsprechender Handlungsempfehlungen je Bereich, in denen das Unternehmen Optimierungsbedarf aufwies. Diese dienen generell dazu, die Reifegrade zu erhöhen und die Zielsicherheitsarchitektur aufzubauen.

Erläuterung	Strategie, Führung und Governance	Sicherheitsstrategie			
		Asset Management	Informationssicherheitsmanagement System	Dienstleistersteuerung	
Nicht betrachtet/relevant	Risiko- und Compliance-Management	Krisenmanagement	Testzentrade & Simulationen	Sicherheitskultur	Sicherheitstraining und Schulung
Teilweise	Sicherheitsresiliente Architektur & Krisenmanagement	Sicherheitsüberwachung	Log Management	Vorfälle & Abhörfestnahmen	Krisenbewältigung und Organisation
Definiert	Cyber-Verteidigung	Schwachstellenmanagement	Penetration Testing	Angriffsmitnahmen	Patch Management
Optimal umgesetzt	Identitäts- und Zugangsmanagement	Identitäts & Account Management	Authentifizierungsmanagement	Starke Authentifizierungsmethoden	Zugangsverwaltung
	Infrastruktur-, Endpunkt- und Anwendungssicherheit	Netzwerkmanagement	Remote Access Control	Software-Lebenszyklus	Application Monitoring
	Datenschutz:	Datenschutz	Data Backup	Datenarchivierung	
	IT-OT-Sicherheit	IT/OT Integration	IT/OT Middleware	IT/OT-Netzkonvergenz	OT Applications Security
	Physische Sicherheit	Sicherheit im Rechenzentrum	Physische Perimeter	Zugangskontrolle	Sicherheit in Büro / Home Office

Abb. 2 Übersicht der ermittelten Reifegrade im Anwendungsbeispiel

Im zweiten Anwendungsfall wurde ein Interview mit einem kleinen Unternehmen über die aktuelle Lage der IT-Sicherheit des Unternehmens durchgeführt, um so den Bedarf an Informationssicherheitsmaßnahmen zu identifizieren. Bei den Interviews wurde schnell klar, dass das mittelständische Unternehmen einen Basisschutz bereits eingerichtet hatte. Zu den Domänen Asset- und Krisen-Management (BCM), Security Training & Education, Identity & Account Management, Incident Management und Physical Security hatte das Unternehmen bereits Maßnahmen umgesetzt. Auch interne Awareness-Maßnahmen werden regelmäßig durchgeführt. Maßnahmen in der Domäne Schwachstellen-Management waren jedoch nicht umfassend umgesetzt. Regelmäßiges Monitoring, das Testen bzw. Durchführen von Übungen zu den umgesetzten Maßnahmen werden insbesondere im Krisen- und Schwachstellenmanagement vernachlässigt. Die stichprobenartige Kontrolle von Dokumenten und Leistungen findet nicht statt.

Im Vergleich zum ersten Anwendungsfall handelt es sich beim zweiten Fall um ein wesentlich kleineres Unternehmen mit weniger Ressourcen. Das kleine Unternehmen hat sich mit dem Thema IT-Sicherheit nicht umfassend beschäftigt und somit Maßnahmen bezüglich IT-Sicherheit und Datenschutz nicht bedacht und vernachlässigt. Das Thema IT-Sicherheit hat bisher keine große Rolle gespielt und wurde als nicht notwendig empfunden. Ein Basisschutz war daher nicht vorhanden. Den Mitarbeitenden war jedoch bewusst, dass Maßnahmen zur Gewährleistung der Unternehmenswerte, insbesondere durch die steigenden Cyber-Angriffe, notwendig sind.

Bei der Betrachtung der zwei Anwendungsfälle hat sich ergeben, dass die Unternehmen generell viel auf Vertrauen setzen, das die Basis für die interne Kommunikation und Zusammenarbeit bildet. Die Mitarbeitenden verbindet eine langjährige Zusammenarbeit, wodurch ein gegenseitiges Vertrauen aufgebaut wurde. Da die Unternehmen mit unterschiedlichen Dienstleistern zusammenarbeiten, war die Domäne Third Party Management in den

Interviews besonders wichtig. Die Unternehmen legen Wert darauf, dass ihre Dienstleister gewisse Sicherheitsmaßnahmen umsetzen und vertrauen ihnen in dieser Hinsicht durch die jahrelange Zusammenarbeit. Dadurch lässt sich schlussfolgern, dass das Sicherheitsniveau der Dienstleister vorausgesetzt wird. Es wird davon ausgegangen, dass diese professionell mit den Unternehmensdaten umgehen. Entsprechende Verträge und Geheimhaltungsvereinbarungen wurden dennoch mit den Dienstleistern abgeschlossen.

Da in KMU oft ein familiäres Zusammenarbeiten vorzufinden ist, werden Kontrollen bzw. interne Audits und das Monitoring vernachlässigt. Sowohl zwischen den Mitarbeitenden untereinander als auch zu den langjährigen Dienstleistern wird auf Vertrauen gesetzt, und Monitoring, Stichproben, Tests und regelmäßige interne Audits als nicht notwendig empfunden.

3.3 Evaluation eines Werkzeugeinsatzes für ESA

Der ESA-Prozess stellt eine praxiserprobte Vorgehensweise zur Bewertung der Sicherheitsarchitektur eines Unternehmens dar. Durch die Analyse des Prozesses und die Erfahrungen aus der Praxis wurde deutlich, dass eine werkzeugunterstützende Prozessoptimierung mithilfe des Einsatzes eines Enterprise Architecture-Tools hilfreich wäre. Eine geeignete Werkzeugunterstützung würde Vorteile bieten, vor allem die Verbesserung der Daten- und Dokumentenverwaltung, Reduzierung des Pflegeaufwands sowie Effizienzsteigerung. Zur Optimierung des ESA-Prozesses soll daher ein Werkzeug zur Abbildung der Referenzarchitekturen eingesetzt werden. Dafür wurde eine Anforderungsanalyse durchgeführt. Die Basis zur Ermittlung der Anforderungen an das Tool bildete der ESA-Prozess. Die einzelnen Prozessschritte von ESA sollten im Tool abgebildet werden und damit das aktuelle Problem der vielen Medienbrüche beheben. Zudem haben sich beim praktischen Einsatz und durch den internen Austausch weitere Anforderungen ergeben. Insgesamt wurden 34 Anforderungen aufgestellt. Die Anforderungen wurden hinsichtlich ihrer Umsetzung analog dem Schablonenprinzip aus [Kl16] anhand der Schlüsselwörter Muss, Soll und Wird priorisiert. Muss-Anforderungen: Diese Anforderungen sind verpflichtend umzusetzen und sind zwingend zu erfüllen. Sie beschreiben die Mindestanforderungen an das Tool. Soll-Anforderungen: Die Soll-Anforderungen sind nicht zwingend notwendig, stellen jedoch einen Wunsch dar. Die Umsetzung wäre daher vorteilhaft und würde den Basisumfang erweitern. Wird-Anforderungen: Diese Anforderungen sind optional und haben eine untergeordnete Bedeutung. Die Realisierung dieser Anforderungen war nicht Bestandteil dieser Studie. Vielmehr dienen die Anforderungen zur zukünftigen Weiterentwicklung des Prozesses und somit auch des Tools.

Die Anforderungen wurden zudem in funktionale und nicht-funktionale Anforderungen (NF) unterteilt. Die funktionalen Anforderungen wurden untergliedert in allgemeine Anforderungen (A), Anforderungen hinsichtlich der Visualisierung der Ergebnisse (V) und zur Assessment-Durchführung (AS). Zu den allgemeinen Anforderungen zählen z. B. die Excel-Import- und Export-Funktion, die Taskverwaltung oder die Verknüpfung mit ande-

ren Tools. Das Tool sollte unterschiedliche Visualisierungsmöglichkeiten bieten, um die Ergebnisse entsprechend darstellen zu können. Die Assessment-Durchführung umfasst Anforderungen wie das Domain Scoping, Erstellen und Einpflegen von Kontrollfragen, die Durchführung von Interviews und das Dokumentieren von Handlungsempfehlungen. Ein Auszug der aufgestellten Anforderungen ist in Tabelle 3 dargestellt.

ID	Anforderung	Priorität	Erfüllungsgrad
..
A06	Export- und Importfunktion von Excel-Tabellen	Muss	vollständig
V11	Visualisierung der verschiedenen Ebenen	Muss	ausreichend
V14	Transformations-Roadmap	Muss	teilweise
V15	Dashboard (Benchmarking)	Soll, Wird	ausreichend
AS16	Domain Scoping	Muss	teilweise
NF25	Pflegeaufwand	Muss	ausreichend
NF30	Bedienbarkeit	Muss	vollständig

Tab. 3 Anforderungen an den Tooleinsatz (Auszug)

3.4 Umsetzung des ESA-Werkzeugkonzepts

Es gibt verschiedene Tools, die zur Abbildung von Referenzarchitekturen genutzt werden können. Unter anderem sind Microsoft-Produkte wie Excel, PowerPoint und Visio zur Darstellung und Dokumentation von Daten im Einsatz. Es hat sich jedoch herausgestellt, dass die Nutzung dieser Tools einen hohen Pflegeaufwand aufweist und Medienbrüche unvermeidbar sind, um alle notwendigen ESA-Prozessschritte zu durchlaufen und abzubilden. Daher sollten die in Tabelle 3 ausgearbeiteten Anforderungen in einem einzigen Tool abgebildet werden. Dafür wurde das Tool LeanIX [Le22] ausgewählt, da es bereits für EA bei PwC eingesetzt wird. Die Evaluation der Einsatzfähigkeit anderer Tools war in diesem ersten Schritt nicht Teil der Studie. LeanIX ist ein cloudbasiertes Enterprise Architecture Tool zum Planen, Analysieren und Überwachen des IT-Portfolios. Eins der Hauptprodukte von LeanIX ist das Enterprise Architecture Management. Damit lassen sich Architekturen abbilden und eine Übersicht der Applikationen, IT-Komponenten und Business Capabilities schaffen.

Um den Tool-Einsatz für ESA-Assessments zu bewerten, wurden die Prozessschritte im Tool abgebildet und Möglichkeiten gesucht, wie die einzelnen Schritte bzw. die Anforderungen damit abgedeckt werden könnten. Durch den Einsatz von LeanIX soll die Verwaltung und die Durchführung der Assessments vereinfacht werden, indem Medienbrüche verhindert und die ESA-Prozessschritte abgebildet werden. Insbesondere die Berichterstattung und damit das Generieren von Reports und anderen Darstellungen soll damit automatisiert erfolgen. Die ausgearbeiteten Anforderungen wurden in LeanIX abgebildet und

deren Umsetzung anhand der Kriterien: vollständig erfüllt, ausreichend erfüllt, teilweise erfüllt, nicht ausreichend erfüllt und nicht erfüllt bewertet.

Um das Tool in der Praxis zu testen, wurde das KMU-Assessment in LeanIX abgebildet und die Umsetzung der Anforderungen an der Durchführung eines Assessments analysiert. Dafür wurden die LeanIX-Standardfunktionen betrachtet. Ein entsprechender Workspace wurde für die Studie zur Verfügung gestellt. Das angepasste Framework bzw. die relevanten Fähigkeiten wurden mithilfe des Excel-Imports in LeanIX eingefügt und in Beziehung gesetzt. Für die Kontrollfragen, die aus den Fähigkeiten bei der Assessment-Durchführung abgeleitet werden, wurde die Survey-Funktion verwendet.

Es ist festzuhalten, dass die Funktionalitäten, die LeanIX anbietet, für die Darstellung der Ergebnisse und somit zum Generieren von Reports gut geeignet sind. Für die Durchführung der Assessments bietet sich die Survey-Funktion gut an. Die eingetragenen Antworten können jederzeit abgeändert und ergänzt werden. Wenn mehrere Interviews mit unterschiedlichen Stakeholdern geführt werden, können mehrere Survey Runs die Antworten voneinander trennen. Von den 34 aufgestellten Anforderungen wurden nur vier Muss-Anforderungen nur teilweise erfüllt und jeweils eine Soll-Anforderung nicht ausreichend und nicht erfüllt.

Die teilweise erfüllten Anforderungen könnten bei der Weiterentwicklung von LeanIX berücksichtigt werden. Es werden mit Unterstützung des Herstellers aktuell Möglichkeiten gesucht, diese Anforderungen bestmöglich abzubilden. Die Umsetzung der Anforderungen bezüglich der Transformation Roadmap kann so belassen werden. In der Roadmap wird die häufig verwendete 3-Phasen-Planung nicht abgebildet. Die Darstellung der Transformation Items in den vier bestehenden Kategorien reicht jedoch aus. Für das Domain Scoping bedarf es einer Datenbank, die aktuell nicht integriert ist. Dafür kann jedoch eine alternative Möglichkeit bei der Erweiterung des Tools gefunden werden. Der Aufwand für das Scoping in der Excel-Tabelle würde sich auch im Tool widerspiegeln. Das Problem bildet hierbei die zusätzliche Verwaltung der Excel-Dateien in einem vorgesehenen Ablageort. Die aktuelle Umsetzung ist insbesondere für Assessments mit KMU jedoch ausreichend. Die Erfüllung der Anforderung bezüglich der Übersichtlichkeit der Daten ist in LeanIX durch die verschiedenen Fact Sheets im Inventory zunächst etwas gewöhnungsbedürftig und bedarf einer kurzen Einarbeitung. Die Filterfunktion erleichtert jedoch das Verwalten der Fact Sheets und wird somit akzeptiert. Die nicht oder nicht ausreichend erfüllten Anforderungen gehören zu den Soll-Anforderungen. Somit hat die Umsetzung dieser Anforderung keine hohe Priorität. Zukünftig sollte ggf. auch für diese Anforderungen eine angemessene Lösung gefunden werden. Auch die Umsetzung bzw. Integration einer Datenbank sollte angedacht werden.

Schlussfolgernd ist festzuhalten, dass LeanIX für ESA und für die Assessments sowohl bei KMU als auch domänen spezifisch einsatzfähig ist. Das Tool beinhaltet essenzielle Funktionalitäten, die den ESA-Prozess optimieren. Ein geringer Restpflegeaufwand bleibt durch die eingeschränkte Durchführung des Domain Scopings und durch das Importieren des ESA-Frameworks bestehen.

4 Evaluation, Zusammenfassung und Ausblick

In diesem Artikel wurde die Anwendbarkeit von ESA für KMU betrachtet. Ausgehend von bekannten Informationssicherheits-Standards, wie etwa der Norm ISO/IEC 27001, der VdS-Richtlinie 10000 und CISIS12, wurde ein angepasstes ESA-Framework für die Durchführung von Assessments für KMU erstellt. Dazu wurden diese Standards miteinander verglichen und Anforderungen für KMU zusammengestellt sowie mit dem vorhandenen ESA-Framework von PwC gemappt. Damit wurde das ESA Framework von 340 Fähigkeiten auf 100 Fähigkeiten für KMU angepasst.

Auch wenn das neue Framework erst in zwei Kundenprojekten zum Einsatz kam, belegen die gewonnenen Erkenntnisse die grundsätzliche Einsetzbarkeit des angepassten ESA-Frameworks und somit die generelle Anwendbarkeit von ESA für KMU. Das angepasste Framework wird bei PwC auch zukünftig die Basis für Assessments mit KMU bilden und nach neuen Erkenntnissen weiter evaluiert und optimiert werden.

Schlussfolgernd ist festzuhalten, dass das angepasste Framework ein guter Leitfaden für die Assessments mit KMU darstellt. Im Rahmen weiterer Projekte muss der Fokus generell auf den priorisierten Sub-Domänen liegen. Die Domänen wie BCM, Schwachstellen- und Vorfalls-Management sollten näher betrachtet werden, insbesondere das Durchführen von Testdurchläufen und Simulationen. Dafür sind die entsprechenden Fähigkeiten der zugehörigen Sub-Domänen priorisiert worden. Das Framework beinhaltet die notwendigen Domänen, die auch in der Norm ISO/IEC 27001, in der VdS-Richtlinie 10000 und auch in CISIS12 aufgezählt werden, und deckt somit die notwendigen Domänen ab, ohne KMU mit der Ausführlichkeit und Detailtiefe des bisherigen ESA-Frameworks zu überfordern. Nach den bisherigen Erfahrungen sind die nun angepasste Tiefe der Sub-Domänen sowie die Anzahl der Fähigkeiten ausreichend, um KMU im Bereich Informationssicherheit zu bewerten.

Der bisherige ESA-Prozess erforderte mehrere verschiedene Tools und war aufwändig aufgrund von Medienbrüchen. Daher wurden Anforderungen für die Prozessoptimierung und den Werkzeugeinsatz für ESA-Assessments aufgestellt und in einem Tool exemplarisch umgesetzt und evaluiert. Für die Evaluation wurden die Anforderungen im Tool LeanIX abgebildet und mittels Praxisbeispielen bewertet. Die Umsetzung der Anforderungen im Tool LeanIX hat gezeigt, dass das Tool für ESA gut geeignet ist. Insbesondere für das Durchführen von Assessments mit KMU sind die aktuell vorhandenen Funktionalitäten ausreichend. Es hat sich herausgestellt, dass das Tool die Assessment-Durchführung unterstützen kann, die Dokumentation und Verwaltung sowohl vereinfacht als auch visualisiert und damit Medienbrüche reduziert.

In diesem Zusammenhang hat sich zudem herausgestellt, dass das vorhandene ESA-Framework (für große Unternehmen) Optimierungspotenzial aufweist. Die Anforderungen der Norm ISO/IEC 27001 und der VdS-Richtlinie 10000 werden in Zukunft mit aufgenommen, um eine Überleitung auf die einzelnen Standards vollständig zu ermöglichen. Zudem kann in zukünftigen Studien die Konformität der beiden ESA-Frameworks mit

weiteren Standards geprüft werden. Hierfür könnten z.B. das BSI IT-Grundschutz- und das NIST-Framework herangezogen werden.

Generell wurde bei der Anwendung erkannt, dass es außerdem empfehlenswert wäre, vorab ein Self-Assessment für Kunden zu erstellen und zugänglich zu machen. Dadurch könnten Kunden eine Selbsteinschätzung durchführen. So könnten Vorabgespräche vereinfacht werden und man erhielte die notwendigen Informationen, um basierend darauf ein kundenspezifisches Assessment zu erstellen und durchzuführen. Bezuglich des Tool-Einsatzes könnte die Umsetzung der identifizierten Soll- und Kann-Anforderungen in Betracht gezogen und das Tool ausgehend von den Anforderungen erweitert werden. Künftige Studien könnten sich daher damit befassen, wie die nicht oder nur teilweise erfüllten Anforderungen umgesetzt werden könnten.

Literaturverzeichnis

- [BHM20] Brocke, Jan vom; Hevner, Alan; Maedche, Alexander: Introduction to Design Science Research. In: Design Science Research. Cases. Springer International Publishing, S. 1–13, 09 2020.
- [Bu13] Buecker, Axel et. al: Using the IBM Security Framework and IBM Security Blueprint to Realize Business-Driven Security. IBM Redbooks, 2013.
- [ET21] ETSI: CYBER; Cybersecurity for SMEs; Part 1: Cybersecurity Standardization Essentials. Bericht ETSI TR 103 787-1, ETSI, 2021.
- [Fr16] Frederik Mijnhardt et. al: Organizational Characteristics Influencing SME Information Security Maturity. Journal of Computer Information Systems, 56(2):106–115, 2016.
- [Gl20] Global Cyber Alliance: GCA Cybersecurity Toolkit for Small Business Handbook. <https://gcato toolkit.org/wp-content/uploads/2021/06/GCA-Toolkit-Handbook.pdf>, 2020.
- [Ho02] Ho, Latifa: Security Management Framework: A New Approach based on John Zachman’s Framework for Enterprise Architecture. Bericht, GIAC Certifications, 2002.
- [IS22] ISC2: CISSP website. <https://www.isc2.org/Certifications/CISSP#>, Stand: 3.5.2022.
- [ISO13] ISO: INFORMATION SECURITY MANAGEMENT. Standard, International Organization for Standardization, Geneva, CH, 2013.
- [IT22] IT-Sicherheitscluster e.V.: CISIS12. <https://cisis12.de/en/>, Stand 23.2.2022.
- [Je16] Jeganathan, Seetharaman: Enterprise Security Architecure. ISSA Journal, 14(12):14–21, 2016.
- [Kl16] Kluge, Roland: Schablonen für alle Fälle. https://www.sophist.de/fileadmin/user_upload/Bilder_zu_Seiten/Publikationen/Wissen_for_free/MASTeR_Broschuere_3-Auflage_interaktiv.pdf, 2016. Stand: 1.3.2022.
- [Le22] LeanIX: Enterprise Architecture Management (EAM). <https://www.leanix.net/en/products/enterprise-architecture-management>, 2022. Stand 1.4.2022.
- [Ma15] Manso, Clara Galan; Rekleitis, Evangelos; Papazafeiopoulos, Fotis; Maritsas, Vasilios: Information security and privacy standards for SMEs, 2015.
- [Mc20] McClintock, Michelle et. al: Enterprise Security Architecture: Mythology or Methodology? In: Proceedings of the 22nd Inter- national Conference on Enterprise Information Systems - Volume 2: ICEIS, INSTICC, SciTePress, S. 679–689, 2020.
- [RA08] Rachamadugu, Vijaykumar; Anderson, John A.: Managing Security and Privacy Integration across Enterprise Business Process and Infrastructure. 2008 IEEE International Conference on Services Computing, 2:351–358, 2008.
- [Sa18] Sage, Ola: Very Small Business Should Use the NIST Cybersecurity Framework. Bericht, CyberRx, 2018.
- [Sc06] Scholtz, Tom: Structure and Content of an Enterprise Information Security Architecture. Bericht, Gartner, Inc., 2006.
- [SCL09] Sherwood, John; Clark, Andrew; Lynas, David: Enterprise security architecture. Bericht, SABSA White Paper, 2009.

- [SLR09] Shen, Yi; Lin, Frank; Rohm, C. E. Tapie: A Framework for Enterprise Security Architecture and Its Application in Information Security Incident Management. *Communications of the IIMA*, 9:2, 2009.
- [SR14] Spruit, Marco René; Röling, M.W.M.: Isfam: the Information Security Focus Area Maturity Model. In: ECIS. 2014.
- [Th09] The Open Group: TOGAF 9 - The Open Group Architecture Framework Version 9, 2009.
- [THM07] Tawileh, Anas; Hilton, Jeremy; McIntosh, Stephen: Managing Information Security in Small and Medium Sized Enterprises: A Holistic Approach. In: ISSE/SECURE 2007 Securing Electronic Business Processes: Highlights of the Information Security Solutions Europe/SECURE 2007 Conference. Vieweg, Wiesbaden, S. 331–339, 2007.
- [Vd22] VdS Schadenverhütung GmbH: VdS 10000 - Informations-Sicherheit für KMU. <https://vds.de/kompetenzen/cyber-security/vds-richtlinien/anforderungsrichtlinien/-leitfaeden/vds-10000-informations-sicherheit-fuer-kmu>, 2022. Stand: 23.2.2022.
- [Yi20] Yigit Ozkan, B. et. al: Modelling adaptive information security for SMEs in a cluster. *Journal of Intellectual Capital*, 21(2):235–256, 2020.

Acquiring Android App Development Skills in a Virtual Learning Environment

Extending the Virtual Programming Lab for Moodle towards Self-Assessed App Development Tasks employing Android and Gradle

Frank Neumann¹, Juan Carlos Rodríguez-del-Pino², Sebastian Homer³

Abstract: The presented project provides a virtual learning environment for university courses teaching app development using Android and Kotlin. It features both self-guided lessons and associated tests as well as practical app development tasks and associated self-assessed programming exercises. The well-known Moodle plug-in *Virtual Programming Lab* (VPL) serves as a starting point for the latter part. VPL gets extended by means for the usage of the Gradle build tool, the Android SDK tools and AndroidX tests for unit tests at the level of Android activities and fragments. In particular, the rather excessive resource requirements of the Gradle tool created numerous problems in the sand-boxes used by VPL. These issues were addressed by configuring suitable VPL settings. In addition, problems associated with the proper usage of these settings had to be addressed in VPL Altogether the proposed virtual learning environment provided a suitable means for learning in the described field as determined by questionnaires send to students.

Keywords: virtual programming lab, VPL, automated grading, automated assessment, automated evaluation, Kotlin, Android, Gradle, Android X tests

1 Introduction

Learning to develop native Android apps is an exciting and challenging endeavor for undergraduate students in computer science. It involves obtaining manifold skills ranging from the used programming language, over various aspects of Android apps and associated frameworks, to high-level architecture and programming concepts needed for such app development.

¹ HTW – Hochschule für Technik und Wirtschaft Berlin, Fachbereich 2: Ingenieurwissenschaften - Technik und Leben, Wilhelminenhofstr. 75a, 12459 Berlin, Germany, frank.neumann@htw-berlin.de, <https://orcid.org/0000-0002-8530-3283>

² ULPGC - University of Las Palmas de Gran Canaria, Department of Informatics and Systems, Tafira Campus, 35017, Spain, jc.rodriguezelpino@ulpgc.es

³ HTW – Hochschule für Technik und Wirtschaft Berlin, Hochschulrechenzentrum, Ostendstraße 25, 12459 Berlin, Germany, sebastian.homer@htw-berlin.de

In the present case, computer science undergraduate students (fifth semester) enroll for the course with a strong background in object-oriented programming, software engineering and some foundations in formal aspects of computer science as well as UML. The learning module starts from this competence level of students and aims at providing them with the above-mentioned theoretical foundations and practical skills for app development up to an intermediate level. The course's goal consists in the development of a medium size Android app by project groups of two students. For mastering the project, the students must take into account modern paradigms of Android development as advocated by the Android team with regards to architecture, design and navigation.

Altogether, the module consists of 12 units that each take three hours (1.5 hours lecture, 1.5 hours of programming exercises and homework) and three additional sessions offering advice and consultation time for the final app project.

The course is split into four learning sections as shown in Table 1.

#	Section name	Number of units	Content
1	Learning Kotlin	5	<ul style="list-style-type: none"> • Type system • Function • Classes and interfaces • Collections • Lambdas • Coroutines
2	Android basics	2	<ul style="list-style-type: none"> • Activities and layouts • Views and view groups • Intents • Manifest
3	Android app architecture	4	<ul style="list-style-type: none"> • Fragments • Navigation • Data binding • MVVM • Connectivity • Asynchronous programming
4	UI aspects	1	<ul style="list-style-type: none"> • Styles • Themes • Material design • Menus • App bars

Table 1: Content of the Android app development module

The first section focuses on providing insights into the Kotlin programming language and its concepts. In particular, the students learn the type system, functions, classes, interfaces, collection classes, lambdas and coroutines. The subsequent block on Android basics focuses on the various components used in an Android app (activities, intents, manifest, etc.)

and layouts. The following section aims at providing architectural guidance for modern Android app development as recommended by Google's architecture guides [GO21]. The last unit focuses on UI aspects such as themes, styles and menus.

2 Related Work

The following sections analyze what kind of approaches and solutions exist for a virtual learning environment for Android app development using Kotlin. The first section studies the topic of self-guided learning courses and their associated tests for self-assessment. The second section focuses on practical app development tasks and associated self-assessed programming exercises.

The analysis will take into account video guided self-learning courses (e.g. *Udacity*, *Udemy*, *Stepik*), self-learning tutorials (e.g. *Google codelabs*), and some tools aimed at creating self-assessed programming exercises (e.g. *CodeRunner*, *VPL*, *EduTools* plugin). Special attention will be paid to the integration into Learning Management Systems (LMS) used in the university context.

2.1 Self-guided Learning Courses – Conceptual Level

Over the last decade massive open online courses (MOOC) emerged and gained attraction both in academia (e.g. *OpenHPI*⁴, *Stanford Online*⁵) and on commercial eLearning platforms (e.g. *Coursera*⁶, *Linkedin Learning*⁷, *Stepik*⁸, *Udacity*⁹, *Udemy*¹⁰). MOOCs typically combine video lectures with subsequent assessment of competences. Although many free courses are available, no free MOOCs for the discussed topic of Android app development could be identified. Many of the above-mentioned MOOCs and eLearning platforms provide the necessary tooling to author new learning courses.

Overall, MOOCs are standalone solutions that in most cases are not integrated in the LMS of universities. In the opposite direction, typical LMS (e.g. *Blackboard*, *Canvas*, *ILIAS*,

LAMS, *Moodle*) used by universities may easily serve as platforms to setup self-guided learning courses combined with automated assessments. Courses in LMS may contain a blend of videos and other learning materials as well as test activities used to assess the learning progress.

⁴ <https://open.hpi.de>

⁵ <https://online.stanford.edu/>

⁶ <https://www.coursera.org/>

⁷ <https://www.linkedin.com/learning/topics/linkedin-learning>

⁸ <https://stepik.org>

⁹ <https://www.udacity.com/>

¹⁰ <https://www.udemy.com>

*Google Developers Codelabs*¹¹ provide self-learning tutorials through a step-by-step experience to setup an app or to use a particular feature in an existing app. They typically give precise coding instructions along with the necessary conceptual knowledge. On the subject of Android app development with Kotlin *Codelabs* offers excellent courses. Unfortunately, *Codelabs* do not provide means for assessing the learning progress, as their main target consists in guiding students through the stepwise process of writing code for the app.

2.2 Self-learning Exercises – Programming Level

A couple of platforms developed over the last decade that support self-learning programming exercises combined with automated assessment. *Codecademy*¹² is a commercial platform offering coding classes in currently 12 programming languages at different levels of proficiency. Users of the platform write their coding solutions, which can then be evaluated. *JetBrains Academy*¹³ falls also in this category but offers significant fewer subjects. This platform uses the *EduTools* plugin¹⁴ for Android Studio, IntelliJ and other IDEs. In addition, *EduTools* may also be connected to other eLearning platforms like Coursera. Unfortunately, no interface connecting *EduTools* to any LMS could be identified. *freeCodeCamp*¹⁵ focuses on providing free self-learning courses initially for web-technologies (HTML, CSS and JavaScript) but extended its scope towards subjects as Python and data science. Unfortunately, these eLearning providers do not integrate in any of the LMS used by universities.

On the university side, various add-ins may be employed to build self-learning programming exercises within LMS. For Moodle two noteworthy plug-ins were identified. The *Virtual Programming Lab* (VPL) [RRH12] [VP22] [Wa15] [Th15] [RRH11a] [RRH11b] [RRH11c] is a plug-in to the Moodle LMS, which supports more than 30 programming languages out of the box. It can be easily extended for additional programming languages by supplying adapted bash scripts. It allows teachers to author programming exercises alongside with the expected results as testcases. Students can edit, run and evaluate their solution in the commonly used Moodle environment. If desired, teachers can accept the grades determined by self-evaluation performed by the students. In addition, *VPL* provides teachers with capabilities to detect plagiarism among the handed-in solutions. *CodeRunner*

is another plug-in to Moodle, which supports about 10 programming languages but does not offer any check for plagiarism [CO22].

11 <https://codelabs.developers.google.com/>

12 <https://www.codecademy.com/>

13 <https://www.jetbrains.com/academy/>

14 <https://github.com/JetBrains/educational-plugin>

15 <https://www.freecodecamp.org/>

3 Problem Analysis

In the authors perception, the ideal virtual learning environment for university courses teaching app development using Android and Kotlin would allow for the authoring and usage of:

- self-guided lessons including videos and other multimedia learning materials focusing on the teaching of concepts
- associated tests assessing the reached competences of the students
- self-learning programming exercises
- automated assessment of the students' solutions handed-in for the programming-exercises

In addition, the learning environment would allow the teacher to check for plagiarism among the coding solutions. Ideally, the students would work in their usual IDE for Android app development i.e. Android Studio. In order to keep all the data for the course in one place, the results of the tests and programming exercises should be stored within the LMS. Table 2 summarizes the functional and technical requirements identified for the virtual learning environment.

ID	Group	Description
F-1	Authoring content, tests and exercises by teachers	
F-1.1		Authoring of lessons including videos and other multimedia learning materials focusing on the teaching of concepts.
F-1.2		Authoring of tests assessing the competence level of the students in the area of concepts.
F-2	Exploring content by students	
F-2.1		Explore and use course material.
F-2.2		Assessment of acquired competence level based on tests.
F-3	Authoring of programming exercises by teachers	
F-3.1		Authoring of programming exercise description.
F-3.2		Authoring of sample solution for programming exercise.
F-3.3		Coding of test cases used for the assessment of the programming exercise.

F-4	Editing, running and evaluation of coding solutions by students	
F-4.1		Editing of coding solution for programming exercise.
F-4.2		Upload of coding solution for programming exercise.
F-4.3		Running of coding solution for programming exercise.
F-4.4		Evaluation of coding solution for programming exercise.
F-5	Check for plagiarism by the teacher	Identification of similar coding solutions for programming exercises.
T-1	LMS	Moodle should be used as LMS for storing the content and the results of tests and programming exercises.

Table 2: Identified functional and technical requirements for the virtual learning environment

4 Proposed Solution

Starting from the gathered requirements and the technology landscape laid out in the previous sections, suitable technologies will be identified in a first step. Based on the selected technology stack, the desired solution will be further developed and described.

4.1 Selection of the Technology Stack

Chapter 2 described currently available technologies, platforms and tools for the realization of a virtual learning environment geared towards app development for Android and Kotlin. Those technologies will be filtered by the requirements collected in chapter 3 in order to select the best fitting technologies.

Due to the technical requirement *T-1* demanding an LMS integration, the many eLearning platforms unable to integrate with Moodle will be rejected. Moodle itself sufficiently covers the requirements groups *F-1* and *F-2*. Only *CodeRunner* and *VPL* as Moodle plug-ins will be further considered for the coverage of the requirements clusters *F-3*, *F-4* and *F-5*. Neither of both plug-ins supports developing Android apps out of the box. However, *VPL* supports the Kotlin programming language and offers plagiarism checks. In contrast, *CodeRunner* does not support Kotlin and does not provide any kind of plagiarism check. Consequently, the combination of Moodle and *VPL* covers more requirements than *CodeRunner* combined with Moodle. Neither of the two solutions supports Android app development out of the box. Finally, Moodle in combination with the *VPL* plug-in is selected as the technology stack to implement the virtual learning environment geared towards app development for Android and Kotlin.

4.2 Configuration of VPL Jail Server

Based on the chosen technology stack of Moodle and *VPL*, the first stage of implementing the desired solution will be detailed. The *VPL* jail server needs to be properly configured for building and running Android apps. In the *VPL* architecture, the jail server acts as an execution sandbox that effectively isolates code run by different users from each other. The development of Android apps requires the following parts of the Android SDK:

- *sdkmanager*: Allows for the installation of the different SDK components.
- *build-tools*: SDK component required for building Android apps.
- *platform-tools*: Contains platform tools like adb.
- *platforms*: The actual SDK packages

On the jail server, the “*/usr/lib/android-sdk*” directory is used for installing the SDK. The SDK manager¹⁶ needs to be downloaded to this directory first. With its help the other three components will be installed. The following code lines depict the installation of the above-mentioned SDK components for the Android SDK version 30.0.3 using the *sdkmanager* tool:

```
sdkmanager "build-tools;30.0.3"
sdkmanager "platform-tools" "platforms;android-30"
```

Subsequently, experiments were conducted on how to use the Gradle build environment within a sandbox on the jail server. The Gradle tool has rather excessive resource requirements that create numerous problems in the sandboxes used by *VPL*. These issues were addressed by identifying and configuring suitable *VPL* settings. It turned out that both within the *VPL* plug-in and on the jail server the maximum used memory and the maximum file size of the execution file had to be increased to 2 GB. Here the required settings for the jail server’s configuration file:

```
MAXMEMORY=2Gb
MAXFILESIZE=2Gb
```

When using those limits, a jail server of version 2.7.2 or higher is required due to issues in the XMLRPC protocol.

¹⁶ <https://developer.android.com/studio/command-line/sdkmanager>

Configuration of VPL Plug-in

Accordingly, the following limits have to be permitted in the configuration of the VPL plug-in:

Maximum memory used: 2 GB

Maximum execution file size: 2 GB

Maximum number of processes: 200

4.3 Authoring of VPL Activities

When creating a new VPL activity in Moodle, those settings have to be selected under *Advanced Settings -> Maximum execution resources limits*.

A general issue is the high number of files being used when developing an Android app in Android Studio. The layout of a minimal app consists of various files in a very specific structure depicted in Figure 1.

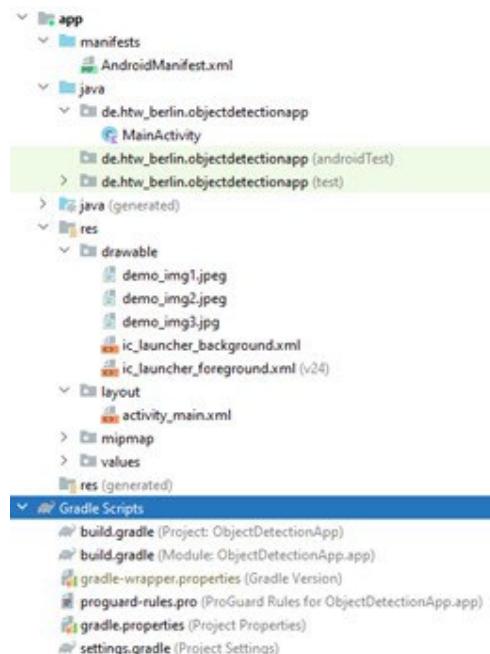


Figure 1: Structure and content of an Android app project

Besides the actual source files, it contains resource files as well as configuration files for Gradle. Drawing from previous experiences, students often fail to hand in a higher number of files. Therefore, the actual submission for the above-mentioned case will consist of only three files:

- MainActivity.kt
- activity_main.xml
- AndroidManifest.xml

Consequently, the other required files have to be provided by the setup of the *VPL activity* as shown in Figure 2. They should be added under *Advanced Settings -> Execution files*.

```

@app
.gitignore
build.gradle
proguard-rules.pro
@src
@main
@res
@drawable-v24
    ic_launcher_foreground.xml
@drawable
    ic_launcher_background.xml
@mipmap-anydpi-v26
    ic_launcher.xml
    ic_launcher_round.xml
@mipmap-hdpi
    ic_launcher.png
    ic_launcher_round.png
@mipmap-mdpi
    ic_launcher.png
    ic_launcher_round.png
@mipmap-xhdpi
    ic_launcher.png
    ic_launcher_round.png
@mipmap-xxhdpi
    ic_launcher.png
    ic_launcher_round.png
@mipmap-xxxhdpi
    ic_launcher.png
    ic_launcher_round.png
@values-night
    themes.xml
@values
    colors.xml
    strings.xml
    themes.xml
build.gradle
gradle.properties
gradlew
gradlew.bat
@gradle
@wrapper
    gradle-wrapper.jar
    gradle-wrapper.properties
settings.gradle

```

Figure 2: Android app files contained as VPL execution files

The Gradle version specified in *gradle-wrapper.properties* should be set to 7.3, as previous versions had memory issues in the jail server's sandbox environment.

Finally, the *vpl_run.sh* and the *vpl_evaluate.sh* files need to be adapted for running and evaluating an Android app. Here, the JDK version 15.0.2 is required for running *RoboElectric* tests at the level of activities and fragments. The difference between the run and the evaluate scripts consists only of the different build targets used for Gradle:

- Run: *compileReleaseSources*
- Evaluate: *testDebugUnitTest*

```
. common_script.sh

check_program java

check_program kotlinc

if [ "$1" == "version" ] ; then

    echo "#!/bin/bash" > vpl_execution

    echo "kotlinc -version &> .kotlinc_version" >> vpl_execution

    echo "cat .kotlinc_version | sed 's/.*/kotlin/kotlin/'" >> vpl_execution

    chmod +x vpl_execution

    exit

fi

JUNIT4=/usr/share/java/junit4.jar

if [ -f $JUNIT4 ] ; then

    CLASSPATH=$CLASSPATH:$JUNIT4

fi

export CLASSPATH

# search Android SDK

export ANDROID_SDK_ROOT=/usr/lib/android-sdk

# Use JDK 15.0.2

export JAVA_HOME=/usr/lib/jvm/jdk-15.0.2
```

```
# build and test using Gradle wrapper

chmod +x ./gradlew

cat common_script.sh > vpl_execution

echo "export ANDROID_SDK_ROOT=$ANDROID_SDK_ROOT" >> vpl_execution

echo "export JAVA_HOME=$JAVA_HOME" >> vpl_execution

echo "./gradlew compileReleaseSources" >> vpl_execution

chmod +x vpl_execution
```

5 Experiences and Discussion

We designed a questionnaire containing questions about the course in general and about the design, handling and the benefits of the exercises. There were 25 students enrolled in the course. 14 of them received grades in the exercises and thus can be regarded as active participants. 10 of the students filled out the questionnaire.

For each question, we used a Likert scale with a range that goes from 1 for the lowest score (strongly disagree) to 5 for the highest (strongly agree). Table 3 contains the questions used, the mean and two distribution plots of the student's responses. The left box plot shows the locality and spread of the answers' numerical values. The median is plotted as a solid vertical line and outliers as dots. The left and right hinges represent the 1. and 3. quartile. The left whisker reaches from the hinge to the lowest value at most $1.5 * \text{inter-quartile range}$ [GG22]. The right histogram shows the absolute number of each answer value in the data set.

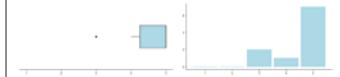
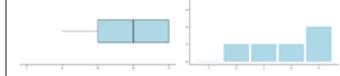
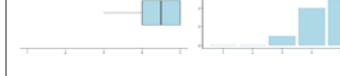
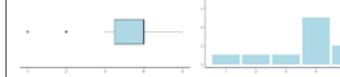
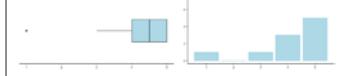
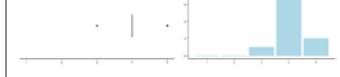
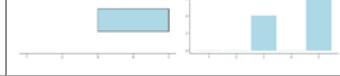
Question	Mean	Distribution plots
The programming exercises were an important supplement to the lectures.	4.5	
The goal of the tasks provided by the professor was always clear.	3.8	
I had no technical difficulties handing in and evaluating my solutions.	2.9	
The provided exercises made me think about related problems of programming.	3.8	
I was able/asked to create my own solutions for the given problem.	3.4	
I used other sources to solve the problem.	2.3	
I discussed my approaches with other students in my class.	2.8	
The programming exercises helped me to better understand the concepts of Android apps such as activities, fragments, navigation and data binding.	4.4	
I developed skills and techniques.	3.6	
The lecture and exercises increased my interest in developing Android app.	4.1	
I feel empowered to solve typical problems/issues in developing Android applications.	4.1	
The combination of lectures and practical exercises was a good preparation for the final project	4.2	

Table 3: Evaluation questions and statistics

A pleasant result of the questionnaire – the didactic goal of the course was achieved: Only one student neither agreed nor disagreed with the statement, that the exercises helped to understand the basic concepts of programming Android apps, the others agreed (4) or strongly agreed (5). 7 out of 10 students strongly agreed, that the programming exercises were an important supplement of the lectures. The majority reports an increased interest in developing Android Apps and feel empowered to solve problems or issues in that field.

3 students were faced with technical problems when handing in or evaluating their solutions, only 1 student strongly agreed, that no problems occurred. None of the students asked the technical helpdesk for assistance or failed uploading a solution.

The answers show that students did not use other sources than the provided broadly. We asked them for a list of used sources, they mentioned Google Codelabs, StackOverflow and video tutorials hosted by YouTube.

6 Conclusions

Altogether the proposed virtual learning environment provides a suitable means for learning in the described field. This statement was confirmed by the questionnaires handed in by the students. However, certain aspects of the current setup can still be improved. The answers to the question on technical difficulties when handing in and evaluating solutions suggest that especially the reliability of this aspect can be significantly increased. Another area of improvement is the long build time required by the Gradle tool. When using Gradle from within Android Studio, the long build time appears also as an issue when building a project for the first time. Afterwards, incremental builds are performed by Gradle that require significantly less time. However, in the sandbox environment of the jail server, Gradle always executes a full build. Improvements addressing this issue are under discussion at the level of *VPL* and the jail server. In addition, the *VPL* developers are studying how to incorporate the Gradle build tool, the Android SDK tools as well as the AndroidX tests so can be used in *VPL* out of the box.

On the level of the requirements gathered in chapter 3, all of them were completely covered by the selected technology stack and the implemented solution.

Requirements for Competence Developing Games in the Environment of SE Competence Development

Philipp M. Zähl¹, Marcel Biewendt², Martin R. Wolf³, Mathias Eggert⁴,

Abstract: Many of today's factors make software development more and more complex, such as time pressure, new technologies, IT security risks, et cetera. Thus, a good preparation of current as well as future software developers in terms of a good software engineering education becomes progressively important. As current research shows, Competence Developing Games (CDGs) and Serious Games can offer a potential solution.

This paper identifies the necessary requirements for CDGs to be conducive in principle, but especially in software engineering (SE) education. For this purpose, the current state of research was summarized in the context of a literature review. Afterwards, some of the identified requirements as well as some additional requirements were evaluated by a survey in terms of subjective relevance.

Keywords: software engineering, requirements, competence developing games, systematic literature review

1 Introduction

With the start of the covid pandemic in 2019, many companies were forced to invest more in digitization [FRB21]. This has ensured strong growth in the German IT industry, but at the same time the current shortage of skilled workers is slowing down the covering of the associated needs for qualified employees [BPW22; Ge21]. As a result, demand exceeds supply, which means that the prices for software and IT service providers continue to rise [BPW22]. In order to remain competitive and to be prepared for future challenges and technologies, current and future software developers must be adequately trained [Sh00].

¹ Lab for IT Organization and Management, Aachen University of Applied Sciences, Eupener Str. 70, 52066 Aachen, Germany; zaehl@fh-aachen.de; <https://orcid.org/0000-0003-3302-4415>

² Institute for Digitization, Aachen University of Applied Sciences, Eupener Str. 70, 52066 Aachen, Germany; biewendt@fh-aachen.de; <https://orcid.org/0000-0002-3410-7995>

³ Institute for Digitization, Aachen University of Applied Sciences, Eupener Str. 70, 52066 Aachen, Germany; m.wolf@fh-aachen.de

⁴ Aachen University of Applied Sciences, Eupener Str. 70, 52066 Aachen, Germany; eggert@fh-aachen.de; <https://orcid.org/0000-0002-3340-7873>

Especially with individual software, which may have to cover very special applications, good preparation for engineering-oriented work in the sense of good software engineering (SE) training is required. As current research shows, Competence Developing Games (CDGs) and Serious Games can offer a potential partial solution [WK17]. However, what are the necessary requirements for such a CDG?

1.1 The CODEGA Project

This paper is based on a thesis, which was developed in the research project CODEGA at the University of Applied Sciences Aachen [Zä22]. CODEGA (*Quality assurance of digital innovations through Competence Developing Games*) was launched in 2021 at the Institute for Digitization Aachen (IDA), in collaboration with the Lab for IT Organization & Management (ITOM) of the Department of Electrical & Information Engineering. The central goal of CODEGA was to investigate whether software quality can be improved by CDGs. This was based on findings from previous research, which demonstrated increased motivation even in more *boring* activities, such as software testing. The project was implemented through the development of an own CDG, on which necessary investigations were subsequently carried out. As a result, the ZeroDebt (ZD) game was created.

ZeroDebt is an online multiplayer deduction game developed as part of the CODEGA project. The gameplay and game setting are strongly based on the game Among Us by *Innersloth*, in which players can move freely in a spaceship. At the beginning of the game, the two roles Crewmate and Impostor are assigned. The task of the Crewmates, who make up the bulk of the group, is to solve the tasks distributed on the map or to identify all Impostors and eliminate them by majority vote. The unknown Impostors, in turn, must eliminate the Crewmates.

During the game, players cannot communicate with each other. Instead, they can discuss their observations or guesses about who the Impostor might be after a body is found or during called emergency meetings. If a majority can agree on a potential Impostor, that Impostor is thrown out of the spaceship. If all Impostors have been eliminated before they could kill enough Crewmates, the Crewmates have won. The Crewmates tasks are simple click or drag'n'drop tasks that can be completed within a few seconds. Even dead Crewmates can continue to complete them as ghosts, putting permanent pressure on the Impostor.

ZeroDebt (ZD) transforms the game in that the Crewmates' tasks are to be solved through software development. For this, a real software project of the players is read in and examined for problems/risks by a SonarQube instance. These identified problems are then introduced into the game environment as Crewmate tasks. Thus, while playing, a player can reduce the technical debt of his project and pave the way for further progress. While the Imposter in Among Us has a countdown after each kill in which he can't kill another player for the time being, in ZD he also has software problems to solve that unlock his ability.

Each completed problem lets him kill one person.

ZeroDebt represents a kind of platform on which many more investigations will be possible. Besides the original goal of improving code quality, the development towards a software engineering CDG is also promising. For this reason, a requirements elicitation is essential for the further development steps.

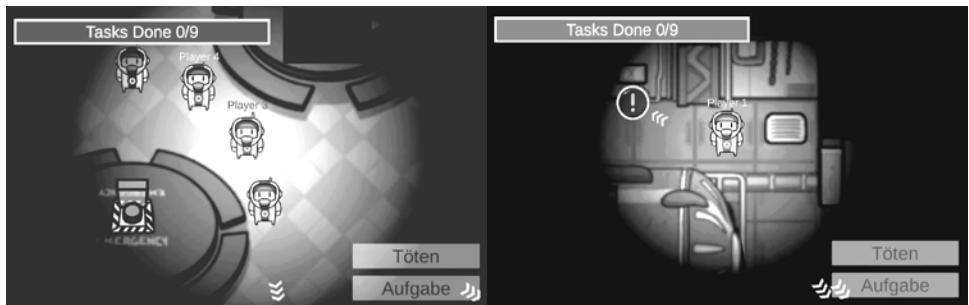


Fig. 1: Screenshots from ZeroDebt

2 Research Method

In order to determine the requirements of a CDG for competence development in the SE environment, a literature research is to be carried out first. Afterwards, some of the identified requirements as well as some additional requirements will be evaluated according to their subjective relevance within the scope of a survey.

3 Systematic Literature Review

3.1 Methodology

The methodological procedure of this research is divided into the three phases *Planning the review, Execution and Analysis* [Ga19; Mc13]. In order to select the literature appropriately and to prevent arbitrariness in the selection of literature, search criteria should be established and the search sources used should be defined [Br15]. The resulting search process is shown in Figure 2. The literature search was conducted in January 2022.

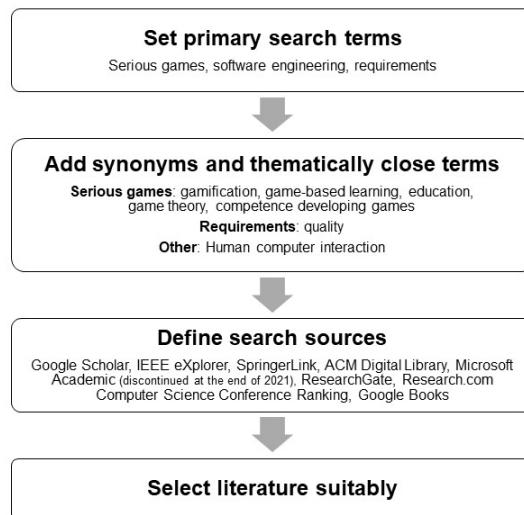


Fig. 2: Search process of the literature review

Studies with collected primary data as well as publications consisting of summaries of secondary literature should be included in the literature search. Thus, it should be ensured that continued research approaches and discovered correlations are not neglected. Furthermore, the following exclusion criteria are established: (1) no full text available; (2) non-English, Dutch or German language studies; (3) publications that only present a self-created learning game without evaluating it with primary data or justifying its success based on similarities to studies from secondary literature.

3.2 Results

The literature review included 19 publications, which include different perspectives, such as psychological or game theoretical perspectives, and different game definitions of Competence Developing Games. Thereby, the literature comes from different years of the last two decades (see Figure 3).

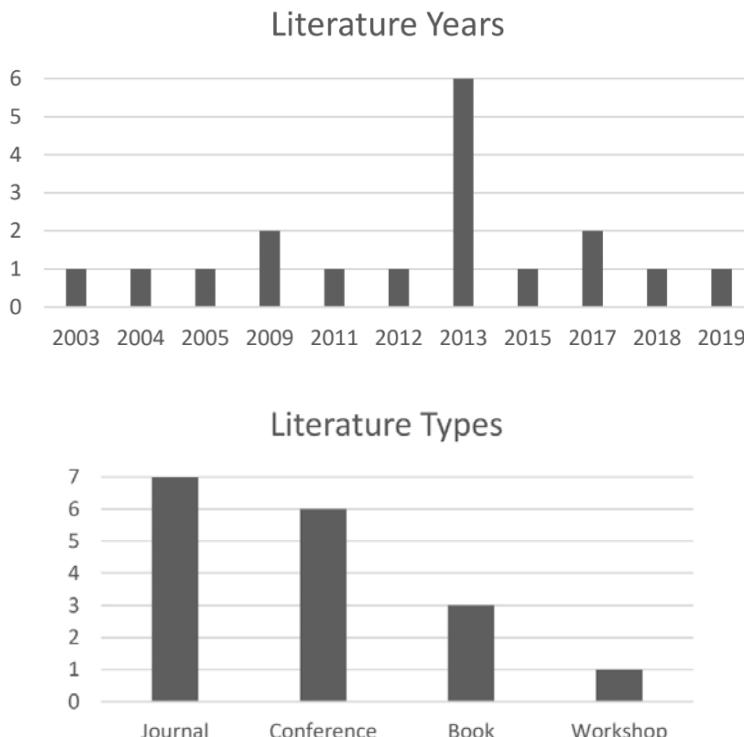


Fig. 3: Statistics on the selected literature

For the sake of clarity, the requirements identified are grouped by category.

Theories of Game Design and common aspects

Games have goals defined according to the SMART principle, or clear, step-by-step, and concise goals [AM18; Ga19; THE09; We13] that appropriately relate learning and business objectives to user interests [We13]. To describe these user interests or needs, different models exist, such as Mark LeBlanc's Taxonomy of Game Pleasures [LHZ04]. According to this, different experience factors or game forms are defined that can be enjoyable for players:

- Sensation (aesthetics, good sound, or tactile pleasure)
- Fantasy (immersion in the fictional world and suspension of disbelief)

- Narrative (suspense or a sequence of events)
- Challenge (by requiring effort through tasks or puzzles)
- Fellowship (feelings of joy, friendship, or community)
- Discovery (exploration, uncovering what is hidden, variety, or strategy)
- Expression (self-expression, creativity, and change)
- Submission (game structures, surrender, and forgetting reality)

It should be noted that players prefer certain game forms [Te16]. Therefore, to reach a broad target group, several of these should be considered in the game design.

Furthermore, players should have the possibilities to achieve a *flow*. Flow refers to a state in which the player is completely immersed in the game and what they are doing. This would include “the feeling of energized concentration, full participation, and success in the process of an activity”. However, this requires the game to have a well-chosen level of difficulty and perceived challenges that neither over- nor under-challenge the player. [BT13; Cs01]

Resources

CDGs require a large financial investment. In addition, teachers need “computers with *gaming specification* (usually high-end devices), technical support, familiarity with game-based learning (GBL) content, adequate preparation time, and a suitable group of participants.” [THE09] According to [SWM15], time-related factors must also be considered in CDGs. Furthermore, a dilemma exists between the formal framework and the minimum time to generate motivation. The limited time has to be used efficiently accordingly, which requires a suitable integration of the curriculum. [THE09] A game duration of about two hours with a game frequency of once a week is recommended - thus the game motivation is the highest [SC13]. Repetition of the game is essential, since playing once is hardly more effective than conventional learning - only through repetition higher learning successes can be achieved [Wo13].

Learning and environment

The introduction of new teaching methods is often difficult because a strong culture of formal, traditional learning has been established [BT13]. Thus, in the transition phase towards learning by Competence Developing Game, it must be taken into account that old patterns

and other influences can strongly influence learning behavior.

A Competence Developing Game has to be introduced into the curriculum step by step in order to accustom the learners slowly to the new learning environment. A too fast change leads to a “cognitive overload”, which can manifest itself in confusion or frustration - thus it is no longer guaranteed that the participants can fully understand and enjoy the game [AM18; BT13]. Furthermore, the game should not be mixed with other teaching methods or forms of teaching, as this reduces motivation and complicates the learning process. “The students would be distracted by the new rules, elements, and learning methods,” which would “blurring the learning experience as a whole” [AM18; Wo13]. Thus, it can also prevent participants from seeing the game “as an unnecessary obstacle” [BT13]. At the same time, they should have a possibility knowledge outside of the game, so that they can fill gaps in their knowledge if necessary.

Within the game, skill building should be achieved through situational learning, exploratory learning or learning through failure should be made possible [GS09]. At the same time, motivation, as the most important factor for learning, must always be maintained [Ge03; THE09]: “When motivation dies, learning dies and playing stops. Since good games are highly motivating to a great many people, we can learn from them how motivation is created and sustained” [Ge03]. In the school or university learning environment, after the introduction of a CDGs, participants should not be given complete autonomy over their learning, as this has a negative impact on learning success [BT13].

Game Art and Aesthetics

Participants learn significantly more when CDGs are played in a group [Wo13]. The thus possible competitive or collaborative game concepts as well as the “certain degree of social comparison” have a positive effect on learning behavior [Pi12; We13]. Furthermore, through collaborative multiplayer game concepts or debriefing, experiences can be discussed and solutions debated, e.g., whether there would have been a better way to master the task, thus “completing the learning experience” [Ga19; GS09; THE09]. Such concepts have already been proven in large commercial online games, where players operate in inhomogeneous teams, sharing knowledge and making the most of different skills. “In the process, they create distributed and dispersed knowledge within a community in ways that would please any contemporary high-tech, cross-functional-team-centered workplace.” [Ge03]

A good CDG should correspond in its appearance to a professionally developed application or, in this case, a commercial game [BT13]. After all, participants have the same expectations when it comes to the design of CDGs as they do for usual games. If the design is not well chosen or sufficiently mature, this leads accordingly to “disappointment, which is reflected in less motivation and learning” [Wo13]. However, this is contrasted with the fact that common games are primarily designed for entertainment and thus may also show vio-

lent or sexual content, which is unsuitable for a CDG in an instructional setting [THE09]. Basically, CDGs should be designed following game theory and rules of game design to be seen as an appealing and motivating application [GS09]. The visuals of the game world/interface also influence learning behavior: While photorealistic, unrealistic, and cartoon-like representations do not produce a significant difference in motivation but have a negative impact on learning success, a CDG with schematic visuals, on the other hand, can achieve higher motivation without affecting learning [Wo13].

The game genre has received little attention in the selected literature with regard to motivation and learning success. However, it could be observed that “girls and women [prefer] other gaming platforms than men [...], such as also mobile and social games as well as predominantly casual games, simulation and role-playing games. Women often have different game motives than men, i.e. joy and fun, creativity, building up, playing together are in the foreground and less the competition or competitive elements. Men and women also differ in the preferred context of play: Women often play alone, while men often play in virtual communities and teams.” Thus, the gender of the intended target audience should be considered when choosing the game genre before developing a CDG. Game intensity is identical for both genders. [Ha17]

In the case of adventure and role-playing games, it must be noted that the game story has a very high influence on engagement and motivation [Wo13]. Furthermore it can be used to impart knowledge [THE09]. In the game world, participants should have the freedom to interact through a set of defined actions to interact in the game world [THE09]. These actions must be obvious and the mechanisms/tools needed to perform them must be understandable and clearly identifiable [Ga19; We13]. To avoid participants receiving relevant information without context or too early, information should be delivered just-in-time. Good games bring information into the game worlds that players move through in such a way that there is always a reference and the meaning is clear. [Ge03]

Game Design, mechanics and concepts

First of all, participants in CDGs should not be only consumers, as in schools, but also producers of their knowledge and curriculum [Ge03]. Thus, CDGs require bringing producers and consumers together to build a dynamic ecosystem [PvC16].

CDGs should include an appropriate game intro. The game intro is tasked with “confronting players in the early stages of the game with problems [that] are specifically designed to allow them to form good generalizations about what will work well later when they are confronted with more complex problems”. In addition, the game intro is intended to be a hidden tutorial, that is, a game-like instruction manual where essential game mechanics are explained and tried out. [Ge03]

The protagonist of the game has a separate role in CDGs, as personal identification with

the game character is strongly related to the motivation to play: “The more a player can manipulate a game character and make decisions that impact on the character, the more the player invests in the character and the game at a deep level. This investment appears to be the deepest foundation of a players motivation in sticking with and eventually mastering a game.” - [Ge03] The identification with the character is even more important for women than for men [Ha17].

A CDG should provide a lot of feedback to the participants [GS09; THE09; We13]. In addition, reward systems are necessary, such as badges or a hall of fame where the best players are presented. The prerequisites/performance requirements for such rewards must be known and progress should be visible. Furthermore, in school and university contexts, it is a good idea to link such rewards to grade bonuses or comparable benefits [So17].

Game content specifically for software projects

Basically, the best possible representation of reality must be created. On the one hand, tools from the real world (such as development environments or technologies) can be used in the CDG for this purpose; in addition, simulation modeling should be implemented hybirdly by means of Discrete-Event Simulation (DES) and System Dynamics (SD) [Pi12]. Furthermore, norms, industry standards, and other environmental factors should be considered in the CDG [Pi12; THE09].

Tasks

The decisive attribute of a task or its challenge is its difficulty and its further progression through the game progress. Dealing with tasks or challenges can be represented by an expertise cycle, where players are repeatedly confronted with similar problems at the beginning “until the players reach a routine, self-evident mastery of certain skills. Then the game confronts the players with a new problem, such as a new type of enemy or a boss, which forces the players to forcing them to rethink the mastery they have come to take for granted, and to integrate their old skills with new ones. This cycle repeats itself throughout the entire game” [Ge03]. A CDG thus has to find a middle ground where the tasks are challenging enough for the participants, but easy enough that the tasks are manageable within a foreseeable period of time. Tasks that are too difficult and almost impossible to master will result in that fewer people want to play the game - the right level of difficulty, in turn leads to increased motivation: “Since games are often challenging, but doable, they are often also pleasantly frustrating, which is a very motivating state for people. To achieve this, good games allow players to adapt the game to their own abilities and learning styles” [Ge03]. It has been found helpful to consider the background, prior knowledge, and previously gained skills of the participants [Ga19; SWM15; THE09]. In addition, the tasks must become more difficult as the game progresses, so as to accommodate the improved skills of

the players [Ga19; Ge03; We13]. Thus, CDGs can consistently “operate at the extreme and growing edge of a player’s competence of a player” and “remain challenging but doable, while schools often must operate at the lowest common denominator” [Ge03]. As a result participants have “a sense of immersion” and can “engage with maximum motivation to actively concentrate”. This moment is also referred to as the flow state [BT13; GLE14], which also has a comparable meaning in colloquial language (cf. meaning “I am in the flow”).

Essential for a well-balanced level of difficulty are the explanations or tasks associated with the task. In addition to good comprehensibility [SWM15], tasks with active as well as passive instructions should be equally given [Wo13]. In general, care should be taken to ensure that participants receive sufficient guidance and instruction [GS09] and that help or hints are available when needed. In addition to in-game explanations, additional resources should be provided, as presented earlier [Wo13]. Participants can fall back on these if the in-game explanations are not sufficient for them. Especially in Competence Developing Games, tasks should be able to be implemented by clear, standardized solutions, with which a sample solution can be provided after completion of a task [Ga19]. Thereby, sample solution as well as the achieved score should be comprehensible or supplemented by explanations [GS09]. Last, tasks should cover different perspectives and levels of abstraction [Pi12]. Furthermore, the duration of the task should be indicated or at least recognizable [Ga19].

Technical aspects

The game should be highly scalable to “allow for the simultaneous training of a large number of learners” [THE09]. A modular/component-based architecture through web technologies would not only satisfy scalability, but is platform-independent, so that participants can ideally play the CDG on their smartphones [Pi12]. In addition, the game must be adaptable to accommodate future changes or different usage scenarios [SWM15]. The performance of the player and the game must be measurable and the possible interactions must be intentional or planned [We13].

4 Primary Data: Survey

It has been shown which requirements a good Competence Developing Game should fulfill according to the secondary literature. For this purpose, selection criteria were first defined, after which then literature was searched and selected. By means of a uniform examination scheme the requirements were extracted and grouped. To complement the research, a survey was conducted. The scientific findings of this paper are not based on the results of the survey, yet they reaffirm the relevance of the topic.

A total of 130 people from the IT environment (Among others: trainees, students and employees) participated, with an average age of 25. Just under 100 of them are studying a subject related to IT. Of the IT employees, about 58 % work in software development. Using Likert scales, the participants were able to evaluate already mentioned, but also further requirements.

As figure 4 shows, the factors *Good design*, *Intuitive operation* as well as *Persistent game progress* received the highest approval. On the other hand, *character creation and customization*, the *choice of when to solve which task* and *comparison with others* also received approval, but also many dissenting votes.

Using free text fields, the participants could name game components that are perceived as particularly demotivating:

- Paid content
- punishments for supposedly wasted time
- time pressure
- daily “login compulsion”
- punishing respawn mechanics e.g. by long ways to the previous location/progress.

In the last section of the survey, participants were presented with a possible concept whereby the ZeroDebt CDG could function as a software engineering CDG. 72 % of the participants were in favor of such a concept. At the same time, however, it was noted that not all SE processes might be suitable or that the prerequisites of these processes, such as communication, might be limited.

A correlation analysis was used to test whether a person's frequency of play affects the requirements queried. A two-sided T-test was performed to test significance, with a significance level of 5 %. Transitive dependencies were tested for significance without findings.

Avg. (1 highest agreement, 5 lowest)

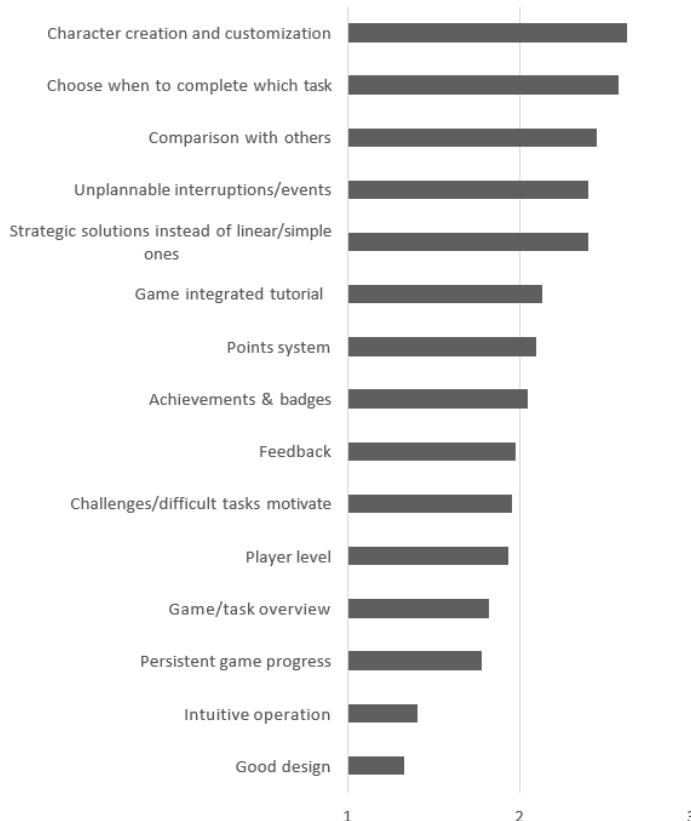


Fig. 4: Survey results [Zä22]

According to this, frequent gamers have the following variations of the requirements (cf. figure 5): The more often the players play, the...

- ...more they want to determine the task order themselves.
- ...more they want to be challenged by more difficult tasks.
- ...more they prefer a persistent game progress.
- ...less they need an integrated tutorial.
- ...shorter a task should take.

5 Discussion

When comparing the requirements determinations in the literature and primary data collection, contradictions can be found:

- *Leaderboard/comparison* with others is an important factor for CDGs according to the literature. However, the conducted survey revealed that this game component is not popular among all players - especially few players stated that the comparison is rather demotivating. This may be due to the fact that few players achieve poorer game results due to their lack of experience. This is because a leaderboard shows them, from their perspective during and after the game, that they did poorly compared to others. This negative rating could have a negative impact on game motivation.
- *Game Optics*: The literature review found that schematic game designs facilitate the highest learning success without having a significant negative impact on game motivation. However, in the survey, respondents indicated that schematic game designs were the least likely to be played. This can possibly be attributed to either an insufficient supply of such games or a lack of interest in playing such games. Thus, the inhibition to play a CDG in schematic design might be higher than for conventional games.

The listed contradictions should be verified by future investigations. Overall, however, it could be observed that the requirements from the literature predominantly coincide with the requirements of our own primary data collection.

Last, it was found that the existing literature for CDGs in the software engineering environment has so far neither formulated explicit requirements for the domain-specific content nor discovered any limitations.

6 Conclusion & Future Work

The intention of this paper was to identify requirements for Competence Developing Games in the software engineering environment. For this purpose, a literature research was conducted.

With regard to the problem statement, it can be stated that CDGs can certainly be used for the further education of students and employees, since previous studies as well as this work have reported positive results. However, it must be noted that such games have to fulfill a high number of requirements, which means that a great deal of effort must go into planning and conception. Otherwise, there is a risk that the game will not find sufficient acceptance or that the learning objectives will not be achieved.

As indicated in the previous section, some contradictions exist between the theoretically optimal as well as the desired requirements. These should be further evaluated in future studies. Furthermore, it is not yet known whether a CDG has to fulfill all requirements in order to be successful. A comparison with conventional computer games would be a promising approach for this.

Last, no domain-specific requirements for software engineering could be identified. Here, it would be useful to further differentiate through primary data.

Acknowledgment

We would like to thank the student team for their excellent development work.

References

- [AM18] Alhammad, M. M.; Moreno, A. M.: Gamification in software engineering education: a systematic mapping. *Journal of Systems and Software* 141/, pp. 131–150, 2018.
- [BPW22] Bayer, F.; Pauly, B.; Wohlrabe, K.: Branchen im Fokus: ITK-Branche./75, pp. 56–59, 2022.
- [Br15] Brocke, J. v.; Simons, A.; Riemer, K.; Niehaves, B.; Plattfaut, R.; Cleven, A.: Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research. *Communications of the Association for Information Systems* 37, Aug. 2015.
- [BT13] Berkling, K.; Thomas, C.: Gamification of a Software Engineering course and a detailed analysis of the factors that lead to its failure. In: 2013 International Conference on Interactive Collaborative Learning, ICL 2013. Pp. 525–530, 2013.
- [Cs01] Csikszentmihalyi, M.; Nakamura, J.; Snyder, C. R.; Lopez, S. J.: *Handbook of Positive Psychology*. Oxford University Press, 2001.
- [FRB21] Faraj, S.; Renno, W.; Bhardwaj, A.: Unto the breach: What the COVID-19 pandemic exposes about digitalization. *Information and Organization* 31/1, p. 100337, 2021, issn: 14717727.
- [Ga19] Gasiba, T.; Beckers, K.; Suppan, S.; Rezabek, F.: On the Requirements for Serious Games Geared Towards Software Developers in the Industry. In: 2019 IEEE 27th International Requirements Engineering Conference (RE). Pp. 286–296, 2019.
- [Ge03] Gee, J. P.: What Video Games Have to Teach Us About Learning and Literacy. *Computers in Entertainment* Vol. 1/, p. 20, 2003.
- [Ge21] Gerd Zika; Christian Schneemann; Markus Hummel; Florian Bernardt; Michael Kalinowski; Tobias Maier; Anke Mönnig; Stefanie Steeg; Marc Ingo Wolter: Die langfristigen Folgen von Covid-19, Demografie und Strukturwandel für die Bundesländer./, 2021.
- [GLE14] Granic, I.; Lobel, A.; Engels, R. C.: The benefits of playing video games. *American psychologist* 69/1, p. 66, 2014.
- [GS09] Gresse von Wangenheim, C.; Shull, F.: To Game or Not to Game? *IEEE Software* 26/, pp. 92–94, 2009.
- [Ha17] Hahn, S.: *Gender und Gaming: Frauen im Fokus der Games-Industrie*. transcript Verlag, 2017.
- [LHZ04] LeBlanc, M.; Hunicke, R.; Zubek, R.: MDA: A formal approach to game design and game research. In: Game Design and Tuning Workshop at the Game Developers Conference. Vol. 4, pp. 1–5, 2004.
- [Mc13] McHugh, M.; Cawley, O.; McCaffery, F.; Richardson, I.; Wang, X.: An agile V-model for medical device software development to overcome the challenges with plan-driven software development lifecycles. In: 2013 5th International Workshop on Software Engineering in Health Care (SEHC). Pp. 12–19, 2013.
- [Pi12] Pieper, J.: Learning software engineering processes through playing games. In: 2012 Second International Workshop on Games and Software Engineering: Realizing User Engagement with Game Engineering Techniques (GAS). Pp. 1–4, 2012.

- [PvC16] Parker, G. G.; van Alstyne, M. W.; Choudary, S. P.: Platform revolution: How networked markets are transforming the economy and how to make them work for you. WW Norton & Company, 2016.
- [SC13] Su, C.-H.; Cheng, C.-H.: 3D game-based learning system for improving learning achievement in software engineering curriculum. *Turkish Online Journal of Educational Technology* 12/, pp. 1–12, 2013.
- [Sh00] Shaw, M.: Software engineering education: A roadmap. In: *Proceedings of the Conference on the Future of Software Engineering*. Pp. 371–380, 2000.
- [So17] Souza, Mauricio Ronny de Almeida; Constantino, K. F.; Veado, L. F.; Figueiredo, E. M. L.: Gamification in Software Engineering Education: An Empirical Study. In: *2017 IEEE 30th Conference on Software Engineering Education and Training (CSEE T)*. Pp. 276–284, 2017.
- [SWM15] Studt, R.; Winterfeldt, G.; Mottok, J.: Measuring software engineering competencies. In: *2015 IEEE Global Engineering Education Conference (EDUCON)*. Pp. 908–914, 2015.
- [Te16] Tekofsky, S.; Miller, P.; Spronck, P.; Slavin, K.: The effect of gender, native English speaking, and age on game genre preference and gaming motivations. In: *International Conference on Intelligent Technologies for Interactive Entertainment*. Pp. 178–183, 2016.
- [THE09] Tang, S.; Hanneghan, M.; El Rhalibi, A.: Introduction to games-based learning. Games-based learning advancements for multi-sensory human computer interfaces: Techniques and effective practices/, pp. 1–17, 2009.
- [We13] Webb, E. N.: Gamification: When It Works, When It Doesn't. In: *Proceedings of the Second international conference on Design, User Experience and Usability: Health, Learning, Playing, Cultural, and Cross-Cultural User Experience*. Pp. 608–614, 2013.
- [WK17] Wolf, M. R.; König, J. A.: Competence Developing Games - Ein Überblick. In (Eibl, M.; Gaedke, M., eds.): *INFORMATIK 2017*. Gesellschaft für Informatik, Bonn, pp. 385–391, 2017.
- [Wo13] Wouters, P.; Nimwegen, C.; Oostendorp, H.; Spek, E.: A Meta-Analysis of the Cognitive and Motivational Effects of Serious Games. *Journal of Educational Psychology* 105/, 2013.
- [Zä22] Zähl, P. M.: Förderung von Software Engineering Kompetenzen durch Competence Developing Games am Beispiel ZeroDebt, Bachelorarbeit, FH Aachen, 2022.

Asynchrone, digitale Lehrpfade in der Kostenrechnung als Open Educational Ressource

Katharina Simbeck¹, Katrin Dziergwa²

Abstract: Auch in den betriebswirtschaftlichen Grundlagenlehrveranstaltungen verfügen Studierende über sehr unterschiedliche Vorkenntnisse. Asynchrone, digitale Lehrpfade helfen Studierenden mit Vorkenntnissen, das Wissen selbstständig zu überprüfen und aufzufrischen. Studierenden ohne Vorkenntnisse oder mit Deutsch als Zweitsprache können im eigenen Tempo Lehrveranstaltungen interaktiv vor- und nachbereiten. Das im Folgenden vorgestellte digitale Lernmaterial nutzt die Aktivität „Test“ des weit verbreiteten Lernmanagementsystems Moodle für Lehrpfade zum Thema Kostenrechnung mit ansteigender Schwierigkeit. Das Material steht als Open Educational Ressource für die weit verbreitete Lernplattform Moodle zum Einsatz in eigenen Lehrveranstaltungen zur Verfügung

Keywords: Open Educational Ressource, Moodle, Kostenrechnung, asynchrone Lehre, digitale Lehre

1 Einleitung

Die Studierenden an deutschen Hochschulen verfügen über sehr unterschiedliche Vorkenntnisse. Die fachspezifische Lernkompetenz unterscheidet sich zum einen nach der Herkunft der Studierenden (Schultyp, Bundesland, nichtdeutsche Hochschulzugangsberechtigung). Viele Studierende beginnen ihr Studium nicht direkt nach dem Abitur sondern arbeiten zunächst oder machen eine Ausbildung. Der Anteil der Studienberechtigten, die im Jahr des Erwerbs der Hochschulzugangsberechtigung das Studium aufnehmen, sinkt seit Jahren und lag im Jahr 2018 nur noch bei 43% [AW20]. 65% der Studienberechtigten haben die allgemeine Hochschulreife an allgemeinbildenden Schulen (z.B. Gymnasien, Gesamtschulen) erworben; 35% haben die Studienzugangsberechtigung an beruflichen Schulen erworben [AW20]. Darüber hinaus arbeiten zwei Drittel der Studierenden neben dem Studium oder betreut Kinder (6%) und andere Angehörige [Mi17]. Daraus ergibt sich die Notwendigkeit der Flexibilisierung und Digitalisierung der Lehre ergänzend zu klassischen Vorlesungen, auch nach dem Ende der Pandemiesituation.

In diesem Beitrag stellen wir einen Ansatz zur digitalen Ergänzung der Lehre im Fach Kostenrechnung durch asynchrone, digitale Lehrpfade vor, die von Studierenden für Kompetenzerwerb und –überprüfung ergänzend oder an Stelle der Lehrveranstaltung genutzt

¹ HTW Berlin, Treskowallee 8, 10318 Berlin, simbeck@htw-berlin.de, <https://orcid.org/0000-0001-6792-461X>

² HTW Berlin, Treskowallee 8, 10318 Berlin, dziergwa@htw-berlin.de

werden können. Die Lehrpfade wurden für das weit verbreitete Lernmanagementsystem Moodle entworfen und können durch Lernende und Lehrende getestet und wiederverwendet werden.

2 Hintergrund

Der Einsatz von digitalen Medien in der Lehre (auch als eLearning bezeichnet) kann der Bereitstellung von Lehrmaterial dienen, selbstgesteuerte Lernphasen in Ergänzung zu traditioneller Lehre ermöglichen oder sogar neue, alternative Lernformate etablieren (z.B. virtuelle Labore, interaktive Übungen) [Br07]. Die Vorteile von digitaler Lehre sind ihre Flexibilität, Skaleneffekte und Ressourceneffizienz in Bezug auf Lehrkräfte und Räume, individualisierte Inhalte und individualisierte Zeiteinteilung, die ubiquitäre Zugänglichkeit und die Möglichkeit zum Austausch von Lernenden untereinander, z.B. in Diskussionsforen [AA15]. Andererseits ist der Einsatz digitaler Lernmethoden nicht für die Vermittlung aller Arten von Kompetenzen geeignet und erfordert von den Lernenden eine höhere Selbstmotivation und bessere Selbstorganisation [AA15].

Open Educational Ressources (OER) sind im Internet erhältliche Lernmaterialien, die ohne technische, finanzielle oder rechtliche Barrieren für andere nutzbar gemacht werden [Hy21]. Die Veröffentlichung als OER wird vor allem dann gefordert, wenn Lehrmaterial öffentlich finanziert erstellt wird [Hy21].

Zahlreiche Distanz-Lehrmethoden können asynchron eingesetzt werden, d.h. sie setzen nicht voraus, dass die Lernenden gleichzeitig online sind, z.B. Arbeitsblätter, Videos [MRB11]. Die asynchronen Lehrmaterialien können durch asynchrone Lehrformen z.B. persönliches Feedback (in Diskussionsforen, per Email) [MRB11] oder synchrone Lehre in Präsenz- oder Distanzform ergänzt werden. Asynchrone Lehre ermöglicht die größtmögliche Flexibilität sowohl für Lernende, als auch für Lehrende, setzt jedoch die Fähigkeit zum selbstgesteuerten Lernen voraus [KLK20]. Ein Nachteil asynchroner Lehre ist häufig die fehlende Interaktion der Studierenden untereinander und fehlende informelle Kommunikationsmöglichkeiten mit der Lehrperson [MRB11]. In Folge der Covid19 Pandemie wurden asynchrone Lehrkonzepte in den letzten Semestern in vielen Disziplinen entwickelt und eingesetzt [KLK20, Gu20, Da20, DM20, BP20]. Auch nach der pandemiebedingten Einschränkung der Präsenzlehre können die einmal entwickelten Materialien zur Flexibilisierung und Individualisierung der Lehre genutzt werden.

Moodle ist ein weltweit eingesetztes Open Source Lernmanagementsystem (LMS), das an Hochschulen, Schulen und anderen Organisationen eingesetzt wird. Laut eigenen Angaben gibt es weltweit über 300 Millionen Nutzer*innen [Mo22]. Moodle bietet Lehrenden die Möglichkeit selbstständig Materialien und Aktivitäten zu erstellen und den Lernenden bereitzustellen. Fragen in der Fragensammlung können dabei importiert werden, was einen Austausch unter Lehrenden auch über Organisationsgrenzen hinweg erleichtert. Aus den Fragen werden Tests erstellt, die in Kursen sowohl zum Lernen/Lehren als auch

zum Prüfen eingesetzt werden können. Die Einstellungen bei Tests ermöglichen es ein allgemeines oder je nach Aufgabentyp auch antwortspezifisches Feedback anzuzeigen. Ebenso kann die Verfügbarkeit eines Tests von den Ergebnissen eines inhaltlich vorgelagerten Tests oder vom Abschluss einer anderen Aktivität abhängig gemacht werden. Dadurch kann man die Aktivitäten den Lernenden strukturiert z.B. als Lernpfad zur Verfügung stellen. Moodle Plugins wie z.B. Level up! bieten darüber hinaus die Möglichkeit die Kurse um Gamification Elemente zu erweitern.

3 Lerntests zur Kostenrechnung

Im Folgenden werden die als OER verfügbaren Lerntests vorgestellt. Diese können im Moodlekurs³ ausprobiert werden und für den eigenen Einsatz als Moodle XML-Datei angefordert werden. Zur Verwendung der Lerntests im eigenen Moodlekurs müssen sie dort importiert werden. Die als OER zur Verfügung stehenden Lerntests wurden in jedem Semester durch Folien, selbsterstellte Lehrvideos, Aufgabenblätter und Verweise zu digitalen Lehrbüchern ergänzt.

3.1 Aufbau und Prinzipien

Im Moment sind grundlegende Themen der beiden Teilgebiete der Kostenrechnung (Vollkostenrechnung und Teilkostenrechnung) abgedeckt und nach Themen gegliedert, so dass eine teilweise Bearbeitung möglich ist. Studierende können die Lerntests beliebig oft und bei freier Zeiteinteilung (Asynchronität) wiederholen, sie werden nicht bewertet. Die Lerntests eignen sich auch nicht zur Bewertung individueller Leistungen, da sie keine Zufallselemente enthalten. Die abgedeckten Themen sind auch Inhalt von kaufmännischen Ausbildungsberufen, die Lerntests können deshalb auch an Berufsschulen eingesetzt werden. Die Entwicklung der Lerntests erfolge nach den folgenden drei Prinzipien:

- Selbsterklärender Inhalt: Durch vorgegebene Antwortmöglichkeiten und textuelle Erläuterungen sind die Lernenden in der Regel in der Lage, auf die richtigen Antworten zu schließen, auch wenn der Lerninhalt für sie neu ist.
- Ansteigende Schwierigkeit.
- Automatisiertes und sofortiges Feedback.

In dem Beispiel zur Deckungsbeitragsrechnung in Abb. 1 ist die Umsetzung der Prinzipien des selbsterklärenden Inhalts und der ansteigenden Schwierigkeit erkennbar. Im ersten

³ <https://moodle.htw-berlin.de/course/view.php?id=36158>

(Benutzer: kostenrechnung-test, Passwort: Kostenrechnung_2022)

Schritt wird ein typischer (einfacher) Aufgabentext zur Deckungsbeitragsrechnung gegeben. Die weitere Lösung der Aufgabenstellung wird vereinfacht, indem die Lernenden angehalten sind, im ersten Schritt die gegebenen Angaben aus dem Text zu entnehmen. Danach sind die Formeln zur Berechnung der gesuchten Werte für Stückdeckungsbeitrag und Gesamtdeckungsbeitrag angegeben, die in Verbindung mit den vorher ermittelten gegebenen Variablen die Lösung erlauben. Die Berechnung des Betriebsergebnisses im nächsten Schritt wird dann textuell erläutert.

Wir wollen ein einfaches Beispiel zur Deckungsbeitragsrechnung betrachten.

Eine Tischlerei stellt nur ein einziges Produkt her: den Tisch "Classic", welcher zum Preis von 230 Euro verkauft wird. Für den Tisch fallen Materialkosten in Höhe von 60 Euro je Stück an. Alle anderen Kosten der Tischlerei sind fix (Löhne, Miete) und betragen 3.000 Euro im Monat.

In diesem Monat wurden 30 Tische gefertigt und verkauft.

Entnehmen Sie dem Text die gegebenen Informationen und tragen Sie sie in diese Tabelle ein:

Stückpreis	e	<input type="text"/>	Euro
variable Kosten	kv	<input type="text"/>	Euro
fixe Kosten	FK	<input type="text"/>	Euro
Menge	x	<input type="text"/>	Stück

Ermitteln Sie Stückdeckungsbeitrag ($db = e - kv$) und Gesamtdeckungsbeitrag ($DB = db * x$)!

$$db = \boxed{} \text{ Euro}$$

$$DB = \boxed{} \text{ Euro}$$

Das Betriebsergebnis ergibt sich jetzt aus der Differenz zwischen Gesamtdeckungsbeitrag und Fixkosten:

$$\text{Betriebsergebnis} = \boxed{} \text{ Euro}$$

Abb. 1: Beispiel aus dem Lerntest zur Deckungsbeitragsrechnung

In den Lerntests werden verschiedene der Standard-Aufgabentypen für Moodletests verwendet, bspw. „Drag and Drop auf Text“ (Abb. 2 linker Teil, Abb. 5), Lückentext (Abb. 1, Abb. 3, Abb. 4) oder „Mehrfach Wahr/Falsch Frage“ (Abb. 2 rechter Teil).

3.2 Vortest

Der Vortest ist vom Aufbau und Inhalt her sehr einfach gehalten (Abb. 2). Die Studierenden lernen das Format der Lerntests kennen, erleben die Wiederholbarkeit und das sofortige Feedback. Der Vortest dient darüber hinaus dem Erlernen der richtigen Verwendung der Fachbegriffe. Die Erfahrung zeigt, dass Studierende häufig Absatz, Umsatz, Erlös wechseln und Schwierigkeiten mit dem Zusammenhang zwischen Stückpreis und Erlös haben.

<p>Wie berechnen wir den Erlös?</p> <div style="display: flex; justify-content: space-around; align-items: center;"> <input type="text"/> * Menge = <input type="text"/> <div style="margin-left: 10px;"> <input type="button" value="Stückpreis"/> <input type="button" value="Erlös"/> </div> </div>	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left;">Wahr</th> <th style="text-align: left;">Falsch</th> </tr> </thead> <tbody> <tr><td><input type="radio"/></td><td><input type="radio"/> Der Erlös ist der Gewinn.</td></tr> <tr><td><input type="radio"/></td><td><input type="radio"/> Der Stückpreis kann auch als Stückerlös bezeichnet werden.</td></tr> <tr><td><input type="radio"/></td><td><input type="radio"/> Die Absatzmenge entspricht dem Umsatz.</td></tr> <tr><td><input type="radio"/></td><td><input type="radio"/> Die Absatzmenge kann auch als Absatz bezeichnet werden.</td></tr> <tr><td><input type="radio"/></td><td><input type="radio"/> Der Stückpreis entspricht meist den Stückkosten.</td></tr> <tr><td><input type="radio"/></td><td><input type="radio"/> Die Verkaufsmenge kann auch als Absatz bezeichnet werden.</td></tr> <tr><td><input type="radio"/></td><td><input type="radio"/> Der Verkaufspreis kann auch als Stückerlös bezeichnet werden.</td></tr> </tbody> </table>	Wahr	Falsch	<input type="radio"/>	<input type="radio"/> Der Erlös ist der Gewinn.	<input type="radio"/>	<input type="radio"/> Der Stückpreis kann auch als Stückerlös bezeichnet werden.	<input type="radio"/>	<input type="radio"/> Die Absatzmenge entspricht dem Umsatz.	<input type="radio"/>	<input type="radio"/> Die Absatzmenge kann auch als Absatz bezeichnet werden.	<input type="radio"/>	<input type="radio"/> Der Stückpreis entspricht meist den Stückkosten.	<input type="radio"/>	<input type="radio"/> Die Verkaufsmenge kann auch als Absatz bezeichnet werden.	<input type="radio"/>	<input type="radio"/> Der Verkaufspreis kann auch als Stückerlös bezeichnet werden.
Wahr	Falsch																
<input type="radio"/>	<input type="radio"/> Der Erlös ist der Gewinn.																
<input type="radio"/>	<input type="radio"/> Der Stückpreis kann auch als Stückerlös bezeichnet werden.																
<input type="radio"/>	<input type="radio"/> Die Absatzmenge entspricht dem Umsatz.																
<input type="radio"/>	<input type="radio"/> Die Absatzmenge kann auch als Absatz bezeichnet werden.																
<input type="radio"/>	<input type="radio"/> Der Stückpreis entspricht meist den Stückkosten.																
<input type="radio"/>	<input type="radio"/> Die Verkaufsmenge kann auch als Absatz bezeichnet werden.																
<input type="radio"/>	<input type="radio"/> Der Verkaufspreis kann auch als Stückerlös bezeichnet werden.																
<p>Wie ermittelt man, ganz allgemein, den Gewinn?</p> <div style="display: flex; justify-content: space-around; align-items: center;"> <input type="text"/> - <input type="text"/> = <input type="text"/> <div style="margin-left: 10px;"> <input type="button" value="Kosten"/> <input type="button" value="Gewinn"/> <input type="button" value="Erlös"/> </div> </div>																	

Abb. 2: Aufgaben aus dem Vortest

3.3 Vollkostenrechnung

Für die Vollkostenrechnung stehen neun Lerntests zur Verfügung. Das Instrument des Kalkulationsschemas der Vollkostenrechnung wird dabei auf zwei Tests verteilt gelehrt. Der erste Lerntest beinhaltet die Erläuterung der Begriffe Kostenträger, Einzel-/Gemeinkosten sowie die Kalkulation der Selbstkosten, der zweite Lerntest die Preiskalkulation. Zu Beginn des zweiten Lerntests müssen die Studierenden die bekannten Elemente des Kalkulationsschemas bis zu den Selbstkosten in die richtige Reihenfolge bringen und die Reihenfolge der neuen Elemente aus einer verbalen Erläuterung entnehmen. Die Kalkulationsbeispiele werden ausgehend von sehr einfachen Beispielen schrittweise ergänzt. Eine häufige Fehlerquelle ist die Kalkulation geplanter Rabatte und Skonti. Diese wird deshalb ausführlich erläutert und mit Verständnisfragen ergänzt. Im dritten, kürzeren, Lerntest lernen die Studierenden, dass das Kalkulationsschema sowohl zur Vor- als auch zur Nachkalkulation verwendet werden kann und berechnen an einem Beispiel Kostenüber- bzw. Kostenunterdeckung.

Die Gemeinkosten berechnen wir als prozentualen Zuschlag auf die Einzelkosten:

$$\text{Gemeinkosten} = \text{Gemeinkostenzuschlagssatz} * \text{Einzelkosten}$$

Um die Herstellungskosten zu kalkulieren, gehen wir nach folgendem Schema vor (Reihenfolge beachten!):

	in Euro	Gemeinkostenzuschlagssatz
Materialeinzelkosten	500	
Materialgemeinkosten		20%
Fertigungseinzelkosten	200	
Fertigungsgemeinkosten		300%
Herstellungskosten		

Berechnen sie die Gemeinkosten für Material und Fertigung.

Ermitteln Sie dann die Herstellungskosten (Summe von Material und Fertigungskosten).

Geben Sie alle Beträge als einfache Zahl ohne Eurozeichen, Punkt oder Komma ein.

Abb. 3: Beispielaufgabe aus dem Lerntest zur Kalkulation der Selbstkosten

Für die Darstellung des Instruments des Betriebsabrechnungsbogens (BAB) gibt es vier Lerntests. Im ersten Lerntest werden Aufbau, Begriffe und Ziel des BAB eingeführt. Drei Umlageverfahren (Blockumlage, Stufenleiter-, Gleichungsverfahren) werden jeweils in einem weiteren Lerntest beispielhaft vorgestellt. Die Lerntests zum BAB sind für Studierende, die sich erstmalig und ernsthaft mit dem Thema beschäftigen deutlich zeitaufwendiger als die Lerntests zum Kalkulationsschema.

Im Lerntest zur Kostenträgerrechnung wird die einstufige und mehrstufige Divisionskalkulation gezeigt. Der letzte Lerntest zur Vollkostenrechnung stellt die Maschinenstundensatzrechnung an einem Beispiel dar.

3.4 Teilkostenrechnung

Für das Themengebiet der Teilkostenrechnung sind bisher nur zwei Themen abgedeckt: Deckungsbeitragsrechnung (Abb. 4) und Eigen- oder Fremdfertigung (Abb. 5). Weitere Themen sind in Vorbereitung.

Die Tischlerei möchte den Tisch "Classic" jetzt nicht nur in Birkenholz (variable Kosten 60 Euro/Stück) sondern zusätzlich auch in Eiche (variable Kosten 80 Euro/Stück) anbieten. Für das Modell Eiche wird ein höherer Preis verlangt (400 Euro).

Vervollständigen Sie die Deckungsbeitragsrechnung!

	Birke		Eiche	
	je Stück	Gesamt	je Stück	Gesamt
Menge	20		10	
Erlös	230 Euro	Euro	Euro	Euro
variable Kosten	Euro	Euro	Euro	Euro
Deckungsbeitrag	Euro	Euro	Euro	Euro
db in % vom Erlös (gerundet auf ganze Zahlen)	%		%	
Fixkosten	3000 Euro			
Betriebsergebnis		Euro		

Das Betriebsergebnis ermitteln wir hier, in dem wir von der Summe der Gesamtdeckungsbeiträge die Fixkosten subtrahieren.

Überlegen Sie kurz: Welches Produkt ist am profitabelsten? Die Antwort diskutieren wir auf der nächsten Seite.

Abb. 4: Beispielaufgabe aus dem Lerntest zur Deckungsbeitragsrechnung

Wir haben bereits gelernt, dass der Deckungsbeitrag beschreibt, inwiefern ein Produkt zur Deckung der beiträgt.

In diesem Lerntest beschäftigen wir uns mit einem Anwendungsgebiet der Deckungsbeitragsrechnung. Unternehmen stehen häufig vor der Entscheidung, ein Zwischenprodukt dazu zu kaufen oder es selbst zu fertigen.

Diese Entscheidung (Eigen- oder Fremdfertigung) bezeichnen wir auch als Entscheidung.

Überlegen Sie doch einmal kurz, welche Vor- und Nachteile Sie jeweils in Bezug auf Eigenfertigung oder Fremdbezug sehen!

Fixkosten

Make or Buy

Abb. 5: Beispielaufgabe aus dem Lernteste zum Aufgabentyp Eigen- oder Fremdfertigung

4 Einsatzmöglichkeiten und Erfahrungen

Die Materialien wurden seit Pandemiebeginn mehrfach in verschiedener Weise eingesetzt, getestet und weiterentwickelt. Eine Übersicht der Lerntests und der Anzahl der Versuche (mehrere Versuche möglich) je Semester ergibt sich aus Tab. 1.

	WiSe 20/21	SoSe 21	WiSe 21/22
Teilnehmer_innen im Kurs	95	74	108
Vortest (Begriffe)	275	235	253
1 Kalkulation der Selbstkosten	139	100	166
2.1 Preiskalkulation	343	247	259
2.2 Vor- und Nachkalkulation	284	157	195
3.1 BAB Einführung	168	136	162
3.2 BAB Blockumlageverfahren	140	114	123
3.3 BAB Stufenleiterverfahren	120	88	110
3.4 BAB Gleichungsverfahren	131	107	113
4 Kostenträgerrechnung	135	96	132
5 Maschinenstundensatzrechnung	135	82	117
6a Deckungsbeitragsrechnung	124	86	81
6b Eigen- oder Fremdfertigung	99	68	69

Tab. 1: Anzahl der Versuche je Lerntest je Semester

4.1 Rein asynchrone Lehre in Verbindung mit Gamification

Für das Wintersemester 2020/21 erfolgte mit großem Aufwand die Erstellung und der erstmalige Einsatz des Materials. Die Lehre wurde auf Grund der Pandemiesituation komplett in Distanzform durchgeführt. Der Kurs wurde vollständig asynchron durchgeführt, d.h. die Studierenden konnten das Material unabhängig von Ihrem Stundenplan bearbeiten. Das Material der Folgewoche, also des Folgethemas, wurde erst sichtbar, wenn das jeweilige Thema abgeschlossen war. Die Lerntests wurden ergänzt durch selbst erstellte Videos, Verweise auf ein digitales Lehrbuch und Übungsaufgaben. Die Übungsaufgaben wurden wöchentlich korrigiert (individuelles Feedback und Bewertung). Durch die asynchrone Gestaltung des Kurses konnten die Studierenden mehrere Wochen vorarbeiten und so bereits zu Semesterbeginn bewertungsrelevante Punkte sammeln. Andererseits konnten sie in der Bearbeitung nicht mehrere Wochen zurückfallen, das sie sonst die Abgabetermine für die semesterbegleitenden Leistungen verpasst hätten.

Der Kurs integrierte mehrere Gamification-Elemente. Die Studierenden starteten nach dem Einstellungstest (entspricht dem Vortest) in der Rolle Financial Analysis Intern. Mit fortschreitender Bearbeitung der Lerntests erhielten sie nacheinander die Rollen Junior Financial Analyst, Senior Financial Analyst, Group Manager Finance, Associate Director Finance, Finance Director und CFO (Chief Financial Officer). Die englischsprachigen Bezeichnungen der Rollen ermöglichen eine einfache genderneutrale Darstellung und entsprechen in internationalen Unternehmen der Realität. Die Rollenprofile wurden als farblich ansprechend formatierte Textfelder mit kleinem Bild in Abhängigkeit von in Moodle erfüllten Voraussetzungen (erfolgreiche Lerntests) ein- und ausgeblendet. Zusätzlich wurde das Moodle Gamification Plugin Level up! sowie die Moodle Funktionalität Fortschrittsbalken verwendet. So konnten sowohl Studierende als auch Lehrende den Fortschritt der Studierenden vergleichen und beobachten. Wöchentliche Emails motivierten die Studierenden, am Ball zu bleiben und kontinuierlich weiter zu lernen.

Bei einer freiwilligen, anonymen Befragung der Kursteilnehmenden erhielt die Kursform sehr gutes Feedback. Die Studierenden gaben an, dass ihnen der Kurs aus den folgenden Gründen gefiel:

- „Die eigenständige Bearbeitung, weil ich dadurch so flexibel wie möglich arbeiten konnte.“
- „Am meisten hat mir gefallen, dass sie den Kurs spielerisch für uns gestaltet haben (mit den Beförderungen).“
- „asynchron, freie Zeiteinteilung; Game-Konzept, weil es anspröt“
- „Das asynchrone Lernen, weil man seine Zeit so einteilen kann, wie es für einen passt.“

Die Studierenden wünschten sich schnellere Antwortzeiten zu Fragen im Forum und die Möglichkeit, das ganze Semester vorzuarbeiten (nicht alle Inhalte waren zu Semesterbeginn schon fertig).

4.2 Hybridform – asynchrone und synchrone Elemente

Im Sommersemester 2021 wurden die erstellten asynchronen Lehrmaterialien erneut eingesetzt, aber durch synchrone Lernphasen in Form von Videovorlesungen ergänzt. Hierbei fand die Vorlesung (seminaristischer Unterricht) immer abwechselnd synchron und asynchron statt, die Übungen (Bearbeitung von Rechenaufgaben zur Kostenrechnung) fanden komplett synchron statt. Das schrittweise Absolvieren der Lerntests (es mussten 90% der Punkte erreicht werden) war wiederum Voraussetzung für die Anzeige der jeweils folgenden Lehrmaterialien und den Zugriff auf die Abgaben. Die Lerntests konnten beliebig oft wiederholt werden. Der Karrierefortschritt vom Financial Analysis Intern zum CFO wurde über die im Moodle Gamification Plugin „Level up!“ erzielten Punkte dargestellt.

In diesem Semester nutzte nur eine Person, die Möglichkeit, anonymes Feedback zum Kurs zu geben: „Ich konnte den Inhalt in meinem Tempo abarbeiten. Die aufgenommenen Lernvideos fand ich ebenfalls sehr hilfreich.“ Kritisch dagegen wurde die Aufforderung gesehen, Lehrbuchkapitel ergänzend zu lesen: „Die Verweise auf das Lehrbuch (ganze Kapitel) waren teilweise etwas too much.“

4.3 Synchrone Lehre mit asynchronen Ergänzungen für das Selbststudium

Im Wintersemester 2021/22 wurde die Lehrveranstaltung in erster Linie in Präsenzform durchgeführt. In der zweiten Semesterhälfte wurden die Veranstaltungen dann hybrid synchron gestreamt. Die Lerntests und Selbstlernmaterialien wurden ergänzend eingesetzt. Die Studierenden konnten jederzeit darauf zugreifen und sie zur Klausurvorbereitung nutzen. Studierende, die die Lerntests vor Weihnachten, also vor Beginn der Klausurlernphase, vollständig absolvierten, konnten 5 Zusatzpunkte bekommen. Hiermit sollte das kontinuierliche Lernen gefördert werden. 50 der 104 im Kurs eingeschriebenen Studierenden machten hiervon Gebrauch.

5 Zusammenfassung

Es steht außer Frage, dass die Interaktionen zwischen Lernenden und Lehrenden sowie zwischen den Lernenden untereinander einen großen Beitrag zu Motivation, Verständnis und der Entwicklung eines akademischen Habitus beiträgt [PH81]. Digitale, asynchrone Elemente können die Lehre auch nach den pandemiebedingten Präsenzeinschränkungen sinnvoll ergänzen, da sie zeitliche und örtliche Flexibilität erlauben, sofortiges Feedback ermöglichen und dazu beitragen, individuelle Lernlücken zu schließen. Die LMS Aktivität „Test“ wird üblicherweise zur Kompetenzmessung oder zur Übung von Aufgabentypen eingesetzt. Im hier vorgestellten Material werden Tests dagegen zur Vermittlung von Material verwendet. Die Lernenden werden durch das ständig leicht steigende Niveau und die kontinuierlichen Rückfragen dazu angeregt, sich mit dem Inhalt aktiv zu beschäftigen, statt ihn nur zu konsumieren. Die übliche Trennung von Vermittlung – Einübung – Überprüfung wird damit aufgehoben.

Mit diesem Beitrag veröffentlichen wir eine Reihe von zwölf Moodle-basierten Lerntests zur Voll- und Teilkostenrechnung als Open Educational Ressource, die zur Vermittlung der betriebswirtschaftlichen Grundlagen in Bachelorstudiengängen oder kaufmännischen Ausbildungsberufen eingesetzt werden können.

Literaturverzeichnis

- [AA15] Arkorful, V.; Abaidoo, N.: The role of e-learning, advantages and disadvantages of its adoption in higher education. International Journal of Instructional Technology and Distance Learning 1/12, S. 29–42, 2015.
- [AW20] Bildung in Deutschland 2020. Ein indikatorengestützter Bericht mit einer Analyse zu Bildung in einer digitalisierten Welt. wbv, Bielefeld, 2020.
- [BP20] Brady, A. K.; Pradhan, D.: Learning without Borders: Asynchronous and Distance Learning in the Age of COVID-19 and Beyond. ATS scholar 3/1, S. 233–242, 2020.
- [Br07] Bremer, C.: Qualität in der Lehre durch eLearning – Qualität im eLearning. In (Horst, C. a. d.; Ehler, H. Hrsg.): eLearning nach Bologna. Prozesse, Projekte, Perspektiven. Grupello-Verl., Düsseldorf, 2007.
- [Da20] Danjou, P.-E.: Distance Teaching of Organic Chemistry Tutorials During the COVID-19 Pandemic: Focus on the Use of Videos and Social Media. Journal of Chemical Education 9/97, S. 3168–3171, 2020.
- [DM20] Dutton, Y.; Mohapatra, S.: COVID-19 and Law Teaching: Guidance on Developing An Asynchronous Online Course for Law Students. SSRN Electronic Journal 447, S. 130, 2020.
- [Gu20] Guo, S.: Synchronous versus asynchronous online teaching of physics during the CO-VID-19 pandemic. Physics Education 6/55, S. 65007, 2020.
- [Hy21] Hylén, J.: Open educational resources: Opportunities and challenges, 2021.
- [KLK20] van der Keylen, P. et al.: Asynchronous, digital teaching in times of COVID-19: a teaching example from general practice. GMS journal for medical education 7/37, Doc98, 2020.
- [Mi17] Middendorff, E., Apolinarski, B., Becker, K., Bornkessel, P.: Die wirtschaftliche und soziale Lage der Studierenden in Deutschland 2016. 21. Sozialerhebung des Deutschen Studentenwerks – durchgeführt vom Deutschen Zentrum für Hochschul- und Wissenschaftsforschung. https://www.dzhw.eu/pdf/sozialerhebung/21/Soz21_hauptbericht_barrierefrei.pdf, Stand: 5.4.2022.
- [Mo22] Moodle: Moodle Statistics. <https://stats.moodle.org/>, Stand: 16.3.22.
- [MRB11] Murphy, E.; Rodríguez-Manzanares, M. A.; Barbour, M.: Asynchronous and synchronous online teaching: Perspectives of Canadian high school distance education teachers. British Journal of Educational Technology 4/42, S. 583–591, 2011.
- [PH81] Portele, G.; Huber, L.: Entwicklung des akademischen Habitus: zum Problem der Konzeptbildung in der Hochschulsozialisationsforschung: Identität und Hochschule.

Design Science Research als wissenschaftliche Herangehensweise für Abschlussarbeiten mit Gestaltungsauftrag in anwendungsorientierten Studiengängen

Irene Weber¹

Abstract: Der Forschungsansatz Design Science Research eignet sich sehr gut für Abschlussarbeiten mit Gestaltungsauftrag in anwendungsorientierten Studiengängen, hat sich jedoch auch als schwierig umsetzbar und erklärbungsbedürftig gezeigt. Hier wird ein einfach umsetzbares Vorgehensmodell nach Design Science Research für Abschlussarbeiten mit Gestaltungsauftrag in der Industrie vorgestellt, zusammen mit einer Gliederungsvorlage und erklärenden Visualisierungen zu Design Science Research. Die Gliederungsvorlage wird an einem Beispiel demonstriert. Erste Anwendungen haben gezeigt, dass das Vorgehensmodell den Studierenden Hilfestellung bietet und das Schreiben gut strukturierter, aussagekräftiger Abschlussarbeiten mit Gestaltungsauftrag erleichtern kann.

Keywords: Design Science Research, Thesis, Academic Writing, Scientific Writing, Information Systems, Industrial Engineering, Business Engineering, Wirtschaftsinformatik, Wirtschaftsingenieurwesen, Abschlussarbeit, Wissenschaftliches Arbeiten.

1 Einleitung

1.1 Problemstellung und Zielsetzung

Abschlussarbeiten in anwendungsorientierten Studiengängen finden oft nicht an der Hochschule, sondern in Unternehmen statt, und bearbeiten dort reale, praktische Aufgabenstellungen. Solche Abschlussarbeiten, im Folgenden als Industrie-Abschlussarbeiten bezeichnet, tragen akademisches Knowhow in die Praxis und bieten den Studierenden wertvolle Praxiserfahrung und Zugang zu potenziellen Arbeitgebern [BKR21, S. 281, Hu19, RDH15, S. 30]. Häufig beinhaltet die Aufgabenstellung, eine Lösung für ein bestehendes Problem zu entwickeln, also einen Gestaltungsauftrag. Eine gestaltende Industrie-Abschlussarbeit (GIAA) ist seitens des Unternehmens darauf ausgerichtet, praktischen Nutzen zu schaffen, während die Hochschule Wissenschaftlichkeit fordert [BKR21, S. 281, Hu19, Kn21]. Den Vorteilen einer Abschlussarbeit in der Industrie steht hier ein möglicher Zielkonflikt gegenüber [BKR21, S. 281].

¹ Hochschule für Angewandte Wissenschaften Kempten, Fakultät für Maschinenbau, Bahnhofstraße 61, 87435 Kempten, irene.weber@hs-kempten.de, <https://orcid.org/0000-0003-2743-1698>

Design Science Research (DSR) ist ein Forschungsansatz, der Wissenschaftlichkeit und praktische Nützlichkeit vereinbart. DSR entwickelt Lösungen für Problemsituationen in der realen Welt, evaluiert sie und gewinnt daraus wissenschaftliche Erkenntnisse [Ba18, S. 358]. Damit eignet sich DSR als Forschungsdesign für Abschlussarbeiten mit Gestaltungsauftrag, um die Anforderungen der Praxis wie der Hochschule zu erfüllen.

DSR ist in vielen Bereichen einsetzbar, so in Technik [HE96], Betriebsführung/Operations Management [ACH16, DLC19], Software Engineering [En20, Kn21], Wirtschaftsingenieurwesen/Industrial Engineering [Sc13, Go21], Interaktionsdesign [Mü21] und Erziehungswissenschaften, dort bezeichnet als Design-Based Research (DBR) [OSO22, Re17]. Auch in Wirtschaftsinformatik/Information Systems (IS) ist DSR etabliert [Wi08, S. 470] und es existieren Leitlinien für DSR in IS. Viele IS-DSR-Publikationen nennen die Leitlinien von [He04] oder [Pe07] als ihre methodische Grundlage [ELE19]. Bekannt ist auch die Empfehlung zur wirkungsvollen Darstellung von DSR-Projekten in Publikationen von [GH13]². Jedoch ist die Methodik der DSR in IS immer noch in der Diskussion [HSS19, Jä22, 29ff, We19, S. 5706]. Kritikpunkte sind, dass die Leitlinien widersprüchlich und interpretationsbedürftig sind, die Terminologie uneinheitlich ist und dass es kein allgemein akzeptiertes Prozessmodell gibt [CTR19, S. 1856f, HSS19, S.6282]. Viele DSR-Veröffentlichungen setzen die Vorgaben der zugrunde gelegten Leitlinien nicht vollständig um [DPS14]. Vor allem Forschungsneulingen fällt es schwer, die Logik der DSR zu verstehen und die Leitlinien richtig anzuwenden [CTR19, S.1844, MGS19, S. 163].

Ein weiterer Interessenkonflikt ergibt sich bei Industrie-Abschlussarbeiten aus Geheimhaltungsansprüchen. Die Aufgabenstellungen und auch die bearbeitenden Studierenden sind oft stark in die Unternehmen integriert. Diese verlangen in der Regel einen Sperrvermerk, der die Publikation der Abschlussarbeit verbietet [BKR21, S. 284, Hu19, S. 54, Kr05, S. 8, Kr15]. Während an Hochschulen und Universitäten intern durchgeführte Abschlussarbeiten in manchen Fächern bereits in wissenschaftliche Publikationen münden können [Kn21, Kr05, S. 8, Sc09, S. 12], bleibt dies Studierenden und betreuenden Hochschulen bei Industrie-Abschlussarbeiten durch die Sperrvermerke verwehrt. Auch können die Studierenden ihre Abschlussarbeit nicht nutzen, etwa für Bewerbungen oder um sie weiterzuführen.

Das Schreiben einer Abschlussarbeit bedeutet für viele Studierenden eine große Herausforderung. Dies zeigt unter anderem das große Angebot an diesbezüglicher Ratgeberliteratur [He21, S. 27f]. Um eine große Leserschaft anzusprechen, sind solche Ratgeber typischerweise allgemein gehalten und geben entsprechend unspezifischen Rat [Br21]. Industrie-Abschlussarbeiten fordern aber meist besonders großen zeitlichen Aufwand [BKR21, S. 281]. Eine leicht verständliche und direkt umsetzbare Anleitung, etwa in Form eines Beispiel-Forschungsprozesses, kann hier zum Erfolg beitragen [WD21, S. 300f].

Diese Arbeit stellt ein DSR-Vorgehensmodell sowie eine Gliederungsvorlage für Abschlussarbeiten mit Gestaltungsauftrag vor. Die Ziele sind: Das Vorgehensmodell soll leicht umsetzbar sein (Ziel 1), das Prinzip von DSR verdeutlichen und dazu anleiten, neben

² Laut Google Scholar bisher über 3000 Mal zitiert

der Entwicklung einer praktisch nützlichen Problemlösung auch wissenschaftlich zu arbeiten, um wissenschaftlichen Erkenntnisgewinn zu ermöglichen (Ziel 2). Ein Vorschlag für die Gliederung der schriftlichen Ausarbeitung ergänzt das Vorgehensmodell. Beide sollen zusätzlich darauf hinwirken, die Arbeit möglichst allgemein und unabhängig von einem Industriepartner zu beschreiben (Ziel 3), um den Bedarf an Geheimhaltung zu verringern und es zu erleichtern, allgemein interessante Ergebnisse zu veröffentlichen.

1.2 Methodik und Überblick

Das Vorgehensmodell wurde selbst nach der DSR entwickelt. Zunächst wurde eine erste Version einer Gliederungsvorlage für Abschlussarbeiten mit Gestaltungsauftrag auf Basis von [GH13, Pe07] erstellt und drei publizierte Abschlussarbeiten mit Gestaltungsauftrag [Gu18, Ri19, Ho22], die nicht als DSR-Projekte gefasst sind, skizzenhaft auf die Gliederungsvorlage abgebildet. Dies ergab eine zweite Version der Gliederungsvorlage. Diese wurde drei Bacheloranden und einem Masteranden, die eine GIAA durchführen, vorgestellt und mündlich erläutert. Dabei zeigte sich, dass es vorteilhaft ist, zusätzlich zur Gliederungsvorlage ein handlungsleitendes Vorgehensmodell zu erstellen, da die Gliederung nicht der Ablauflogik des Vorgehens entspricht (wie auch bei [Kn21]). Gliederungsvorlage und Vorgehensmodell wurden weiter verfeinert und zwei Diagramme entwickelt, um das Prinzip der DSR zu verdeutlichen.

Der folgende Abschnitt nennt verwandte Arbeiten. Anschließend werden das Vorgehensmodell und die Gliederungsvorlage vorgestellt. Die Gliederungsvorlage wird demonstriert und das Vorgehensmodell evaluiert. Die Arbeit schließt mit Diskussion, Fazit und Ausblick.

1.3 Verwandte Arbeiten

[GKM15, MGS17, MGS19] befassen sich mit der Anwendung von DSR für PhD-Theesen in IS. [MGS19] haben dafür eigene Leitlinien entwickelt. [Kn21] stellt Leitlinien und eine Gliederungsvorlage für Industrie-Masterarbeiten mit Gestaltungsauftrag im Software Engineering vor. [St17] stellen ebenfalls eine vereinfachende Visualisierung der DSR vor, jedoch mit anderer Zielsetzung. Anregungen zu DSR in Abschlussarbeiten geben [BCS19, BKS20]. Beide präsentieren bekannte DSR-Ansätze, ohne eigene spezifische Anleitungen zu entwickeln. [BKS20] bietet ergänzend eine Checkliste zur nachträglichen Überprüfung des DSR-Vorgehens.

2 Vorgehensmodell und Gliederungsvorlage

2.1 Verständnis von Design Science Research

Das DSR-Verständnis dieser Arbeit folgt [ACH16]. Im Zentrum steht ein Projekt, das eine Lösung für eine Problemsituation entwickelt. Dies symbolisiert der als „EP“ (Entwicklungsprojekt) bezeichnete Kasten in Abb. 1. Ein „Design Science Research Projekt“ (DSRP) beinhaltet ein EP und dient zugleich wissenschaftlichen Zielen.

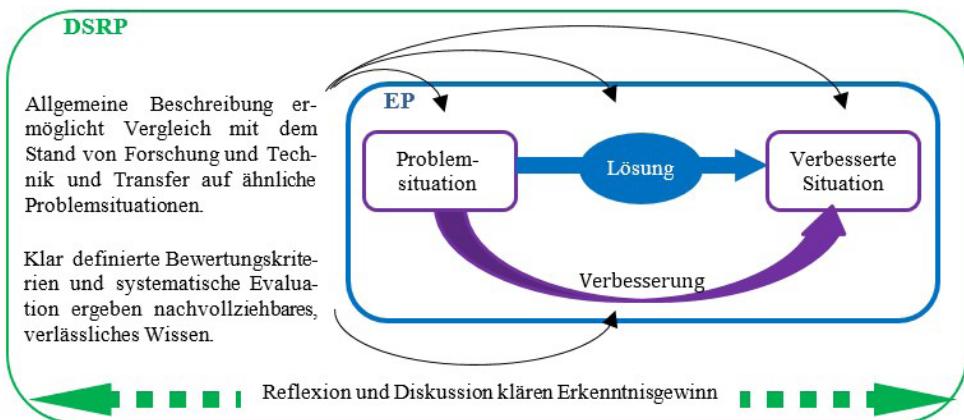


Abb. 1 Prinzip eines Entwicklungsprojekts (EP) und Design Science Research-Projekts (DSRP)

Ein EP kann wissenschaftliches Wissen darüber beitragen, wie sich eine bestimmte Klasse von Problemsituationen verbessern lässt. Dafür ist es erforderlich, diese Klasse von Problemsituationen möglichst allgemein zu beschreiben und abzugrenzen. Die wissenschaftliche Validität eines DSRPs ergibt sich aus der Evaluation, die die erzielte Verbesserung bewertet. Dazu muss in Form von prüfbaren Kriterien klar dokumentiert werden, welche Art von Verbesserungen angestrebt werden, und eine objektive und überprüfbare Auswertung der Kriterien muss zeigen, ob und wie weit die Lösung diese Verbesserung herbeiführt. Eine kritische Reflexion und Diskussion soll herausarbeiten, worin der erzielte Wissensfortschritt besteht und ob sich die Erkenntnisse des DSRPs über das konkrete EP hinaus auf andere Problemsituationen derselben Klasse übertragen lassen [ACH16]. Der folgende Abschnitt stellt ein Vorgehensmodell vor, das ein DSRP in einzelne Aktivitäten aufschlüsselt.

2.2 DSR-V-Modell als Vorgehensmodell

Dieser Abschnitt beschreibt das Vorgehensmodell in knapper Form, aber möglichst verständlich formuliert, so wie es Studierenden zur Verfügung gestellt werden soll. Abb. 2 zeigt die Aktivitäten des Vorgehensmodells im Überblick. Ein DSR-Projekts läuft auf mehreren Ebenen ab: auf wissenschaftlicher Ebene, wirtschaftlicher Business-Ebene und technischer Ebene. Vertikal sind die Aktivitäten in Abb. 2 in diesen Ebenen angeordnet. Das Vorgehensmodell bildet ein V-Modell. Aktivitäten im linken Zweig definieren Ziele und Anforderungen, Aktivitäten im rechten Zweig evaluieren, wie weit diese erfüllt werden (symbolisiert durch Pfeile). Die horizontale Anordnung entspricht einem idealisierten zeitlichen Ablauf, der die Logik des Vorgehens widerspiegelt, analog zu [Pe07]. In der tatsächlichen Durchführung laufen Aktivitäten auch in anderer Reihenfolge oder parallel ab. Rücksprünge, Überarbeitungen und Ergänzungen, während das Projekt voranschreitet, sind wahrscheinlich, etwa zur Verfeinerung der Ziele oder durch weitere Literaturrecherchen. Nur die grau unterlegten Aktivitäten sollten sich direkt mit der Anwendungssituation des Industriepartners befassen und diese etwa als Fallbeispiel für Demonstration und Evaluation nutzen. Die anderen Aktivitäten sollen möglichst allgemein bleiben.

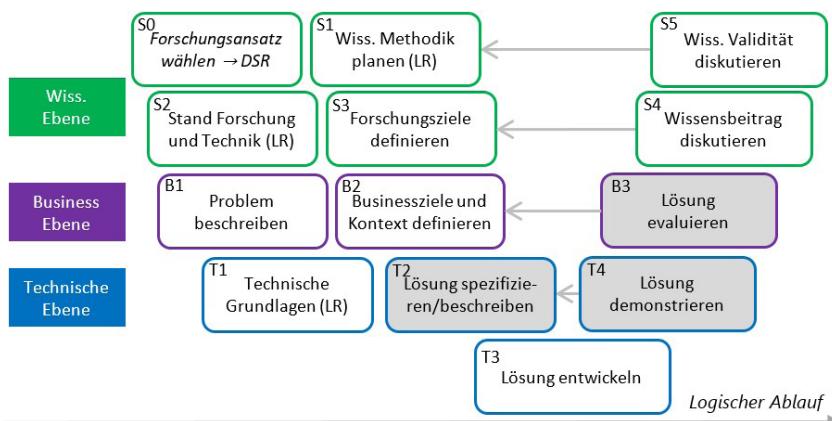


Abb. 2 DSR-V-Modell als Vorgehensmodell für gestaltende Industrie-Abschlussarbeiten (T2, T4, B3 mit Informationen des Industriepartners, LR: Literaturrecherche, wiss.: wissenschaftlich)

2.3 Aktivitäten des DSR-V-Modells

Die Erläuterung der einzelnen Aktivitäten folgt mit Ausnahme der ersten Aktivität S0 den Ebenen des Vorgehensmodells, beginnend mit der Business-Ebene.

S0: Den Anstoß für das DSR-Projekt einer GIAA gibt in der Regel ein praktisches Problem. Damit kann sich die Arbeit für DSR als Forschungsansatz entscheiden und das DSR-V-Modell anwenden.

B1: Eine der ersten Aktivitäten des DSRP ist, das Problem mit allgemein geläufigen Fachbegriffen (statt etwa mit firmeninternen Begriffen) und abstrakt zu beschreiben, also losgelöst von der konkreten Problemsituation beim Projektauftraggeber. Eine abstrakte, allgemeine Problembeschreibung lässt sich mit dem Stand der Forschung und Technik in Beziehung setzen und vergleichen. Möglicherweise benötigt diese Aktivität Input aus der Literatur, um aktuell geläufige Fachbegriffe und Darstellungsweisen herauszufinden.

B2: Hier wird festgelegt, welcher Art die entwickelte Lösung sein soll, zum Beispiel Entwurf, Prototyp, produktiv nutzbares Softwaretool, Berechnungsmethode, organisatorischer Prozess, Workflow oder Kombinationen daraus. Geklärt werden auch Eigenschaften und Besonderheiten des Anwendungskontexts, auf den die Lösung zugeschnitten wird.

Auf die Problembeschreibung aufbauend werden die Business-Ziele definiert. Es wird konkret aufgeschlüsselt, welche quantitativen und qualitativen Verbesserungen das Projekt bewirken soll. Quantitative Verbesserungen sollten durch messbare Kennzahlen ausdrückt werden (eingesparte Kosten, Material, Zeit, weniger Fehler, genauere Vorhersagen, bessere Kommunikation, höherer Durchsatz, ...). Qualitative Verbesserungen können zum Beispiel darin bestehen, dass sie neue Möglichkeiten eröffnen wie neue Services, Geschäftschancen, Auswertungen, Transparenz etc.

B3: Die Evaluation aus Business-Sicht untersucht, ob und wie gut die entwickelte Lösung die Business-Ziele erfüllt. Hier sind Vergleiche mit der ursprünglichen Problemsituation (Abb. 1) und mit dem Stand von Forschung und Technik interessant. Die Evaluation ist ausschlaggebend für die Wissenschaftlichkeit des DSRPs [ACH16, Pe12]. Sie muss objektiv (unvoreingenommen) und systematisch erfolgen und so dokumentiert werden, dass ihre Ergebnisse nachvollziehbar sind. Dazu sollte sie wissenschaftliche Methoden nutzen.

S1: Diese Aktivität wählt die wissenschaftlichen Methoden für die Evaluation aus und plant ihre Anwendung. Besonders eignen sich Fallstudien, Simulationen, Kriterienbasierte Evaluation, Teilnehmendenbeobachtung oder -befragung, Expertenbefragungen (Fokusgruppen) und Informiertes Argument [JP14, Pe12, Kr16].

S2: Die Idee von DSR ist, durch ein EP eine praktische Problemlösung und zugleich wissenschaftliches Wissen zu generieren. Um technisch und wissenschaftlich auf dem aktuellen Stand zu arbeiten und darauf aufzubauen zu können, wird dieser in Fach- und wissenschaftlicher Literatur recherchiert und erfasst.

S3: Diese Aktivität befasst sich damit, Forschungsfragen zu formulieren. Jedes DSRP kann Fragen wie die folgenden bearbeiten [ACH16, TDA19]: Wie ist eine (allgemein verwendbare) Lösung für das Problem beschaffen? Wie kann eine allgemein verwendbare Lösung auf die konkrete Problemstellung angepasst werden? Ist eine (bekannte) Lösung geeignet, um auch im vorliegenden Anwendungsumfeld die definierten Ziele zu erreichen? Wie kann eine Lösung für das Problem entwickelt werden? Wo und wie kann die Lösung praktisch eingesetzt werden?

In manchen Fällen lassen sich auch wissenschaftliche Fragen bearbeiten, die sich nicht direkt auf die Lösungsentwicklung richten. Diese ergeben sich vielleicht im Verlauf des Projekts oder am Ende, wenn die Diskussion das Projekt reflektiert und allgemein interessante, wissenschaftlichen Erkenntnisbeiträge herausarbeitet.

S4: Die Diskussion muss explizit zu den Forschungsfragen Stellung nehmen und sie beantworten. Sie kann auch weitere interessante Aspekte des Projekts reflektieren, etwa

- Vorteile und Nachteile der entwickelten Lösung, auch im Vergleich zu alternativen Lösungsansätzen für ähnliche Problemstellungen aus der Literatur oder zu eigenen verworfenen Lösungsversuchen, Erläuterung und Begründung von Design-Entscheidungen, ...
- Vorteile und Nachteile der eingesetzten Methoden und des Vorgehens beim Entwickeln: welche Methoden haben gut funktioniert, welche waren weniger zweckmäßig, worauf sollten ähnliche zukünftige Projekte achten, was hat sich im Rückblick als ungünstig oder erfolglos erwiesen (sogenannte Lessons Learned), wo sind unerwartete Schwierigkeiten aufgetreten, was war einfacher als erwartet? ...
- neue Einsichten oder ein geändertes Verständnis der bearbeiteten Problemstellung, eine neue Formalisierung der Problemstellung, ein allgemeines anpassbares Lösungsschema, einen Ordnungsrahmen, um das Problem und verwandte Problemstellungen zu charakterisieren, ...
- innovatives Vorgehen bei Entwicklung oder Evaluation, neue Kennzahlen oder Bewertungsschemata, ...
- Möglichkeiten, das DSRP fortzuführen.

S5: Zur Diskussion gehört auch eine Reflexion darüber, wie valide die gewonnenen Erkenntnisse sind. Welche Schwächen und Einschränkungen hat das Vorgehen bei der Evaluation? Wurden die erhobenen Informationen korrekt erfasst? Bilden die erhobenen Informationen die Business-Ziele gut ab? Lassen sich die Ergebnisse auf andere, ähnliche Problemstellungen übertragen? Diese Fragen berühren die Konstrukt-Validität, interne Validität und externe Validität der Evaluation auf Business-Ebene, siehe etwa [GRW08].

T1: Die technische Ebene behandelt die eigentliche Lösungsentwicklung. Es ist zu klären, welche fachlich-technischen Grundlagen und Methoden bei der Entwicklung der Lösung zum Einsatz kommen.

T2: Die Lösung, ihre Architektur, Komponenten, Funktionen usw. wird beschrieben.

T3: Die Lösung wird entwickelt mit fachlichen und technischen Kenntnissen, Methoden und Werkzeugen.

T4: Es wird demonstriert, dass die entwickelte Lösung wie beschrieben oder spezifiziert funktioniert, das heißt, dass sie die technischen, funktionalen etc. Anforderungen erfüllt, zum Beispiel durch reale oder simulierte Anwendung auf einen Beispielfall.

2.4 Verwendung von Literatur

Ein DSRP kann Information aus verschiedenen Arten von Quellen heranziehen: (1) Literatur zum Stand der Forschung, wie Beiträge in wissenschaftlichen Zeitschriften und Konferenzbänden, Dissertationen und Abschlussarbeiten, akademische Arbeitsberichte und Vorabdrucke (Preprints); (2) Fachliteratur zum Stand der Technik in der Praxis wie Fachzeitschriften, kommerzielle Studien, Whitepapers, Projektberichte und ähnliches; (3) fachliche und technische Grundlagen aus Fach-, Lehr-, Handbüchern, Software- und Gerätedokumentationen, etc.; (4) Lehr- und Handbücher zum wissenschaftlichen Arbeiten und zu wissenschaftlicher Methodik.

2.5 Gliederungsvorlage

Abb. 3 zeigt eine Gliederungsvorlage (DSR-Thesis, DSRT) für gestaltende Abschlussarbeiten, die als DSRP durchgeführt werden.

<u>1 Einleitung</u>	<u>4 Grundlagen</u>
1.1 Motivation	<u>5 Beschreibung der Lösung</u>
1.2 Problembeschreibung	5.1 Architektur/Überblick
1.3 Anwendungskontext und Zielsetzung	5.2 Komponente 1
1.4 Forschungsfragen	5.3 Komponente 2
1.5 Aufbau der Arbeit	...
<u>2 Stand von Forschung und Technik</u>	<u>6 Demonstration</u>
2.1 Stand der Forschung	<u>7 Evaluation</u>
2.2 Stand der Technik	<u>8 Vorgehen bei der Entwicklung</u>
<u>3 Methodik</u>	<u>9 Diskussion</u>
3.1 Forschungsdesign	10 Fazit und Ausblick
3.2 Wissenschaftliche Methoden	
3.3 Fachliche Methoden	

Abb. 3 DSRT-Gliederungsvorlage für gestaltende Abschlussarbeiten

Die DSRT ist als Maximalvorlage zu verstehen. Im zeitlichen Rahmen einer Abschlussarbeit ist es nicht immer möglich, alle Abschnitte voll auszuarbeiten. Der Maximalvorschlag soll Möglichkeiten aufzeigen, die Abschlussarbeit bewusst zu gestalten: die Bearbeitenden können fachliche Grundlagen im Grundlagenteil aufbereiten und wiedergeben, oder unter „Stand der Technik“ oder „Fachliche Methoden“ nur mit Literaturangaben darauf verweisen. Das Vorgehen bei der Entwicklung kann unter „Entwicklung“ ausführlich dargestellt oder unter „Fachliche Methoden“ kurz erläutert werden. Ebenso kann der Abschnitt „Evaluation“ das Vorgehen bei der Evaluation ausführlich darlegen oder lediglich Ergebnisse nennen, während der Abschnitt „Methoden“ das Vorgehen mit Verweis auf die Literatur kurz skizziert.

3 Demonstration und Evaluation

Um die Anwendung der Gliederungsvorlage zu demonstrieren, wird eine gestaltende Industrie-Diplomarbeit [Gu18], die im Original nicht als DSR konzipiert ist, als DSRT strukturiert, siehe Abb. 4. Die Demonstration möchte nicht nachträglich diese Diplomarbeit bewerten oder Schwächen aufdecken, sondern die Unterschiede aufzeigen, die die Konzeption als DSRT mit sich bringen und auf die die DSRT hinführen soll.

Die kursiv gedruckten Stellen zeigen, welche Teile der DSRT die Diplomarbeit wenig oder gar nicht ausarbeitet und womit diese etwa zu füllen wären. Die Diplomarbeit enthält darüber hinaus Teile, welche die Gliederungsvorlage nicht vorsieht und Abb. 4 daher nicht zeigt: ein Unternehmensporträt und eine Beschreibung des Fertigungsverfahrens im Unternehmen, die beide für das Thema der Diplomarbeit nicht wesentlich sind. Diese Demonstration am Beispiel zeigt, dass die DSRT sich eignet, um eine GIAA schlüssig zu präsentieren, und dass sie Bearbeitenden verdeutlichen kann, wie die Ziele 1 bis 3 umzusetzen sind.

Vorversionen des Vorgehensmodells und der DSRT wurden Bearbeitern einer GIAA vorgeschlagen (1 x Masterarbeit, 3 x Bachelorarbeiten). In allen Fällen haben die Studierenden die Idee der DSRT leicht verstanden, befanden eine Gliederungsvorlage als hilfreich und waren gern bereit, ihre Arbeit als DSRT zu schreiben. Ein Bachelorand äußerte, dass die Bachelorarbeit so „sicher wissenschaftlicher wird“. Der Masterand erklärte, dass ihm die Vorgaben „Unsicherheit nehmen“. Zwei der Bachelorarbeiten sind inzwischen fast oder ganz abgeschlossen. Die Studierenden haben ihre Arbeit als DSRT strukturiert, Forschungsfragen formuliert und ihre Lösungen demonstriert sowie evaluiert.

4 Diskussion

Die bisherigen Erfahrungen sind, dass das DSR-V-Modell und die DSRT gut verständlich und anwendbar sind (Ziel 1). Die Evaluation der entwickelten Lösungen in Bezug auf die Business-Ziele ist bei einer GIAA offenbar leicht möglich, da mit dem Industriepartner

ein geeignetes Umfeld bereitsteht. Dazu angeleitet setzen die Abschlussarbeiten die Evaluation gut um. Der Hinweis, den aktuellen Stand von Forschung und Technik zu erheben und unterschiedliche Arten von Literatur zu verwenden, wird ebenfalls gut umgesetzt (Ziel 2). Eine Anleitung zur wissenschaftlichen Validierung (S5 des DSR-V-Modells) haben die bisher genutzten Versionen des Vorgehensmodells noch nicht beinhaltet. Wie eine solche Validierung in einer GIAA ausschen kann, zeigt das Beispiel [Bo19, S. 53]. Vorgehensmodell und Gliederungsvorlage sensibilisieren dafür und leiten dazu an, die Abschlussarbeit allgemeiner und in vielen Teilen vom Industriepartner unabhängig zu formulieren. Die bisher mit dem Vorgehensmodell unterstützten Abschlussarbeiten konnten dies noch nicht ganz realisieren (Ziel 3).

Titel: Konzept zur Einführung eines Werkzeugmanagements (WM)	
1 Einleitung	4 Grundlagen
1.1 Motivation	Identifikationsverfahren Markierungsmethoden
Produktion erfordert Werkzeuge, die der Werkzeugbau eines Unternehmens passend zum Auftrag fertigt. Wiederverwendung vorhandener Werkzeuge spart Kosten und Zeit...	
1.2 Problembeschreibung	5 Beschreibung
Ein besonderes Problem des WM ist der Verschleiß. Durch den Gebrauch ändern sich Maße der Werkzeuge. Ein Werkzeug ist also nicht mehrfach für Aufträge mit gleicher Spezifikation verwendbar...	Sollkonzept des neuen Arbeitsablaufs Bedienoberfläche und nötige Funktionen Datenbank Technik für Markierung und Identifikation, ...
1.3 Anwendungskontext und Zielsetzung	6 Demonstration
Konzept für ein WM-System bei einem KMU mit Ziehmaschinen, das folgende Probleme verbessert: (1) Werkzeuge prozesssicher identifizieren (2) Aktuelle Maße von Werkzeugen einfach erfassen und abrufen (3) Passendes Werkzeug leicht finden	Beschreibung eines Beispiel-Arbeitsablaufs beim Industriepartner Beispieldaten in der entworfenen Datenbank <i>genauere Beschreibung des Arbeitsablaufs entsprechend einer Simulation</i>
1.4 Forschungsfragen	7 Evaluation
Wie kann eine Lösung gestaltet werden?	(1) Experimente zur Auswahl von Identifikationsverfahren und Markierungsmethoden (2) Aufwandsabschätzungen (Simulation) (3) Simulation mit Mitarbeitern, Befragung
1.5 Aufbau der Arbeit	8 Vorgehen bei der Entwicklung
...	Experimente zur Auswahl von Identifikationsverfahren und Markierungsmethoden Experimente mit Technik für Markierung und Identifikation
2 Stand von Forschung und Technik	9 Diskussion
2.1 Stand der Forschung	Die Ziele wurden im Wesentlichen erreicht (als „Zusammenfassung“) <i>Besonderheiten, Stärken, Schwächen</i> <i>Anwendbarkeit in anderen Bereichen</i> <i>Vergleich mit anderen WM-Tools</i>
Industrie 4.0, Maschinelles Lernen für Verschleißvorhersage?	
2.2 Stand der Technik	
Fachbücher, Abschlussarbeiten, Dokumentationen anderer Tools für WM	
3 Methodik	10 Fazit und Ausblick
3.1 Forschungsdesign	Implementierung als nächster Schritt
DSR	
3.2 Wissenschaftliche Methoden	
Mitarbeiterbefragung in einem realen Betrieb (für Ist-Analyse), <i>Simulation, Experiment</i>	
3.3 Fachliche Methoden	
Entity Relationship-Modellierung, Nutzwertanalyse, Quality Function Deployment-Analyse	

Abb. 4 Demonstration der Gliederungsvorlage am Beispiel von [Gu18]

Ein gebräuchliches Grob-Gliederungsschema für gestaltende Abschlussarbeiten lässt sich auf „Einführung – Grundlagen – Ist-Analyse – Entwicklung/Sollkonzept – Schluss“ kondensieren, zum Beispiel [Gu18, Ri19], Dieses Schema entspricht der Logik eines EP in Abb. 1. Demgegenüber nennt die DSRT zusätzlich explizit Teile, die das EP zum wissenschaftlichen DSRP ergänzen. Von einer GIAA kann nicht zwingend ein umfänglicher wissenschaftlicher Beitrag im Sinne neuen Wissens gefordert werden. Der praktische Nutzen sollte im Konfliktfall vor dem Ziel eines wissenschaftlichen Erkenntnisgewinns Priorität bekommen, um die Interessen der Studierenden zu wahren, die ihre GIAA bei einem potenziellen Arbeitgeber durchführen. Der wissenschaftliche DSRT-Ansatz bringt jedoch auch für die Unternehmen Vorteile. Insbesondere gewinnen sie eine fundierte Evaluation und den Abgleich der Unternehmenssituation mit dem Stand von Forschung und Technik. Besonders wenn eine GIAA einen Entwurf, ein Konzept oder eine Handlungsempfehlung ausarbeitet (im Unterschied zu einer implementierten, praktisch nutzbaren Lösung), bildet eine simulierte Demonstration (T3 in Abb. 2) einen Funktionscheck und kann zu einer detaillierteren und schlüssigeren Lösung hinführen.

Die DSRT ordnet den Abschnitt über das Vorgehen bei der Entwicklung nach der Beschreibung ein, so wie auch [Kn21] education of software engineers must include strong and applied education in empirical research methods. For most students, the master's thesis is the last, but also most applied form of this education in their studies. Problem: Especially thesis work in collaboration with industry requires that concerns of stakeholders from academia and practice are carefully balanced. It is possible, yet difficult to do high-impact empirical work within the timeframe of a typical thesis. In particular, if this research aims to provide practical value to industry, academic quality can suffer. Even though constructive research methods such as Design Science Research (DSR). Die Beschreibung des Entwicklungsprozesses fällt leichter und wird klarer, wenn sie sich auf die fertige Lösung beziehen kann. Im Entwicklungsprozess kommt es oft zu Rücksprüngen und nachträglichen Änderungen. Die rückblickende Perspektive macht es oft einfacher, den Prozess strukturiert darzustellen und das Vorgehen und Entwurfsentscheidungen zu begründen.

Die Leitlinien für DSR-Masterarbeiten von [Kn21] sind komplementär zum hier vorgestellten Vorgehensmodell. [Kn21] richtet sich an Universitätsstudierende mit guten Kenntnissen im wissenschaftlichen Arbeiten, aber wenig Praxiserfahrung in einem betrieblichen Umfeld. Das hier präsentierte Vorgehensmodell soll vor allem das wissenschaftliche Vorgehen und Verfassen der Abschlussarbeiten unterstützen. Es richtet sich an Studierende von HAWen, die in der Regel Praxiserfahrung besitzen und sich nach bisheriger Erfahrung gut in ein betriebliches Umfeld integrieren können. Die technisch-fachliche Entwicklung einer Lösung bildet meist die Hauptaktivität der GIAA. Das Vorgehensmodell lässt diese offen, denn sie kann je nach Aufgabenstellung sehr unterschiedlich ausfallen und die Studierenden sollten die nötigen fachlichen Kenntnisse aus ihrem Studium mitbringen.

Diese Arbeit legt ein einfaches, pragmatisches Verständnis von DSR zugrunde, um das Vorgehensmodell leicht verständlich und anwendbar zu halten. Hinsichtlich der DSR als Wissenschaftstheorie bietet diese Arbeit mit dem DSR-V-Modell eine neuartige Darstellung und Visualisierung bekannter Ansätze [Pe07, GH13, ACH16], um damit, wie etwa auch [St17], zum Verständnis und zur Verbreitung der DSR beizutragen.

Demonstration, Evaluation und Diskussion dieser Arbeit stützen sich auf wenige publizierten GIAAs sowie auf die nicht publizierten GIAAs, die die Autorin betreut hat, da Recherchen nur wenige publizierte GIAAs erbracht haben. Diese sind in hier relevanten Aspekten wie der Grobgliederung („Einführung – Grundlagen – Ist-Analyse – Entwicklung/Sollkonzept – Schluss“), den bearbeiteten Gestaltungsaufträgen, der Ausarbeitung einer Diskussion ähnlich zu den der Autorin bekannten unveröffentlichten GIAAs³. Dies spricht dafür, dass sich die hier präsentierten Überlegungen und Ergebnisse verallgemeinern lassen.

5 Fazit und Ausblick

Hier wird ein einfach umsetzbares Vorgehensmodell nach Design Science Research für Abschlussarbeiten mit Gestaltungsauftrag vorgestellt, zusammen mit einer Gliederungsvorlage und erklärenden Visualisierungen zur DSR. Erste Anwendungen haben gezeigt, dass sie den Studierenden Hilfestellung bei ihren Abschlussarbeiten bieten, das Schreiben gut strukturierter, aussagekräftiger Abschlussarbeiten erleichtern und zur wissenschaftlichen und praktischen Qualität der Abschlussarbeiten beitragen können.

Zu den Zielen des Ansatzes gehört auch, auf eine möglichst unternehmensneutrale Darstellung von Abschlussarbeiten, die in Unternehmen stattfinden, hinzuwirken. Dies soll es erleichtern, Abschlussarbeiten oder Teile davon zu veröffentlichen und dabei die Geheimhaltungserfordernisse der Unternehmen zu wahren. Ein breiterer Austausch von Ergebnissen aus Abschlussarbeiten würde die Weiterentwicklung der Unternehmen wie auch die wissenschaftliche Arbeit der beteiligten Hochschulen fördern.

Eine erweiterte Version des Vorgehensmodells sollte wissenschaftliche Methoden und systematische Literaturauswertung genauer beschreiben oder zugängliche Literatur zum Nachlesen vorschlagen, um den Studierenden ihre Anwendung zu erleichtern. Studierende orientieren sich beim Schreiben ihrer Abschlussarbeiten erfahrungsgemäß auch an Abschlussarbeiten, die sie im Industriepartnerunternehmen oder von Mitstudierenden erhalten. Da dieser DSRT-Ansatz neu ist, sind noch keine danach erstellten Abschlussarbeiten öffentlich verfügbar. Um den DSRT-Ansatz noch anwendbarer zu machen, wären musterhafte DSRT-Arbeiten als Beispiele sehr hilfreich.

Dank. Dank an Susanne Ertel, Constantin Weber und Antonia Weber für wertvolle Diskussionen und an Stephan Eberle, Kilian Mayrhans, Marcel M. Alici und Leonard Blosch zudem für die Bereitschaft, diesen Ansatz zu erproben.

3 <https://www.hs-kempten.de/fakultaet-maschinenbau/labore/informationssysteme-im-maschinenbau/labor>

Literaturverzeichnis

- [ACH16] van Aken, J.; Chandrasekaran, A.; Halman, J.: Conducting and publishing design science research: Inaugural essay of the design science department of the Journal of Operations Management. *Journal of Operations Management* 1/47-48, S. 1–8, 2016.
- [Ba18] Baskerville, R. et al.: Design Science Research Contributions: Finding a Balance between Artifact and Theory. *Journal of the Association for Information Systems* 5/19, S. 358–376, 2018.
- [BCS19] Bergener, K.; Clever, N. C.; Stein, A.: *Wissenschaftliches Arbeiten im Wirtschaftsinformatik-Studium: Leitfaden für die erfolgreiche Abschlussarbeit*. Springer Gabler, Berlin, 2019.
- [BKR21] Benner-Wickner, M.; Kneuper, R.; Riccardi, L.: Betreuung unternehmensspezifischer Abschlussarbeiten—Zahlen, Erfahrungen und Handlungsempfehlungen aus dem Bereich IT und Technik. In (Hattula, C.; Hilgers-Sekowsky, J.; Schuster, G. Hrsg.): *Praxisorientierte Hochschullehre*. Springer Fachmedien, Wiesbaden, S. 279–291, 2021.
- [BKS20] Benner-Wickner, M.; Kneuper, R.; Schlömer, I.: Leitfaden für die Nutzung von Design Science Research in Abschlussarbeiten. *IUBH Discussion Papers - IT & Engineering*, Erfurt, 2020.
- [Bo19] Botter, A.: Die Entwicklung eines Prozessmodells zur strukturierten Weiterentwicklung von IT-Mitarbeitern. HS Mittweida. Bachelorarbeit, 2019.
- [Br21] Brommer, S.: Schreibberatung in Buchform. Ein kritischer Blick in ausgewählte Ratgeber zum wissenschaftlichen Schreiben. *Exposé-Zeitschrift für wissenschaftliches Schreiben und Publizieren* 2/2, S. 17–18, 2021.
- [CTR19] Cater-Steel, A.; Toleman, M.; Rajaeian, M. M.: Design science research in doctoral projects: An analysis of Australian theses. *Journal of the Association for Information Systems* 12/20, S. 1844–1869, 2019.
- [DLC19] Dresch, A.; Lacerda, D. P.; Cauchick-Miguel, P. A.: Design science in operations management: conceptual foundations and literature analysis. *Brazilian Journal of Operations & Production Management* 2/16, S. 333–346, 2019.
- [DPS14] Dwivedi, N.; Purao, S.; Straub, D. W.: Knowledge Contributions in Design Science Research: A Meta-Analysis. In (Tremblay, M. C. et al. Hrsg.): *Advancing the Impact of Design Science: Moving from Theory to Practice: 9th International Conference. DESRIST 2014*. Springer International Publishing, Cham, S. 115–131, 2014.
- [ELE19] Engel, C. T.; Leicht, N.; Ebel, P.: The Imprint of Design Science in Information Systems Research: An Empirical Analysis of the AIS Senior Scholars? Basket: 40th International Conference on Information Systems. *ICIS 2019*, S. 1–17, 2019.
- [En20] Engström, E. et al.: How software engineering research aligns with design science: a review. *Empirical Software Engineering* 25, S. 2630–2660, 2020.
- [GH13] Gregor, S.; Hevner, A. R.: Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly* 2/37, S. 337–355, 2013.
- [GKM15] Gerber, A.; Kotze, P.; van der Merwe, A.: Design Science Research as Research Approach in Doctoral Studies: Twenty-first Americas Conference on Information Systems, 2015.

- [Go21] Goecks, L. S. et al.: Design Science Research in practice: review of applications in Industrial Engineering. *Gestão & Produção* 28, 2021.
- [GRW08] Gibbert, M.; Ruigrok, W.; Wicki, B.: What passes as a rigorous case study? *Strategic management journal* 13/29, S. 1465–1474, 2008.
- [Gu18] Gusner, L.: Konzept zur Einführung eines Werkzeugmanagementsystems. TU Wien. Diplomarbeit, 2018.
- [HE96] Hubka, V.; Eder, W. E.: *Design Science Introduction to the Needs, Scope and Organization of Engineering Design Knowledge*, 1996.
- [He04] Hevner, A. R. et al.: Design Science in Information Systems Research. *MIS Quarterly* 1/28, S. 75–105, 2004.
- [He21] Heine, C.: Schreibwissenschaft und Schreibratgeber zum wissenschaftlichen Schreiben. *trans-kom* 1/14, S. 23–42, 2021.
- [Ho22] Hoskova, K.: Entwicklung und Visualisierung einer Berechnungslogik für Datenqualität in der Instandhaltung. TU Wien. Diploma Thesis, 2022.
- [HSS19] Holtkamp, P.; Soliman, W.; Siponen, M.: Reconsidering the Role of Research Method Guidelines for Qualitative, Mixed-methods, and Design Science Research: Proceedings of the 52nd Hawaii International Conference on System Sciences. *HICSS 2019*, S. 6280–6289, 2019.
- [Hu19] Hubert, F.: Praxisorientierte Abschlussarbeiten im Unternehmen. *WiSt-Wirtschaftswissenschaftliches Studium* 4/48, S. 53–56, 2019.
- [Jä22] Järvinen, P.: Improving guidelines and developing a taxonomy of methodologies for research in information systems. University of Jyväskylä. Dissertation, 2022:43:28.
- [JP14] Johannesson, P.; Perjons, E.: *An Introduction to Design Science*. Springer International Publishing, Cham, 2014.
- [Kn21] Knauss, E.: Constructive Master's Thesis Work in Industry: Guidelines for Applying Design Science Research: 2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering Education and Training. *ICSE-SEET*, S. 110–121, 2021.
- [Kr05] Kramer, J. W.: Zur Forschungsaktivität von Professoren an Fachhochschulen, untersucht am Beispiel der Hochschule Wismar. *Wismarer Diskussionspapiere* 05, 2005.
- [Kr15] Kratzke, N.: Sperrvermerke bei Abschlussarbeiten: Ein Kommentar. Weshalb Sperrvermerke kritisch hinterfragt werden sollten. *Informatik-Spektrum* 5/38, S. 409–413, 2015.
- [Kr16] Kriglstein, S. et al.: Evaluation Methods in Process-Aware Information Systems Research with a Perspective on Human Orientation. *Business & Information Systems Engineering* 6/58, S. 397–414, 2016.
- [MGS17] van der Merwe, A.; Gerber, A.; Smuts, H.: Mapping a Design Science Research Cycle to the Postgraduate Research Report. In (Liebenberg, J.; Gruner, S. Hrsg.): *ICT Education*. Springer International Publishing, Cham, S. 293–308, 2017.
- [MGS19] van der Merwe, A.; Gerber, A.; Smuts, H.: Guidelines for conducting design science research in information systems: 40th International Conference on Information Systems. *ICIS 2019*, S. 163–178, 2019.

- [OSO22] Oppl, S.; Stary, C.; Oppl, S.: On the Ambiguous Nature of Theory in Educational Design-based Research—Reflecting and Structuring from an IS Perspective. EDeR. Educational Design Research 1/6, 2022.
- [Pe07] Peffers, K. et al.: A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems 3/24, S. 45–77, 2007.
- [Pe12] Peffers, K. et al.: Design Science Research Evaluation. In (Peffers, K.; Rothenberger, M.; Kuechler, B. Hrsg.): Design Science Research in Information Systems. Advances in Theory and Practice. DESRIST 2012. Springer, Berlin u.a., S. 398–410, 2012.
- [RDH15] Roessler, I.; Duong, S.; Hachmeister, C.-D.: Welche Missionen haben Hochschulen? Third Mission als Leitung der Fachhochschulen für die und mit der Gesellschaft. Centrum für Hochschulentwicklung Nr. 182, 2015.
- [Re17] Reinmann, G.: Design-Based Research. In (Dorothea, S.; Hermann, N. Hrsg.): Gestaltungsorientierte Forschung—Basis für soziale Innovationen: Erprobte Ansätze im Zusammenwirken von Wissenschaft und Praxis. Berichte zur Beruflichen Bildung. W. Bertelsmann Verlag, Bielefeld, S. 49–64, 2017.
- [Ri19] Richter, L.: Optimierung des Materialbereitstellungsprozesses der Hilfs- und Betriebsstoffe. Hochschule Mittweida. Bachelor Thesis, 2019.
- [Sc09] Schuh, C.: Publikationsverhalten im Überblick—eine Zusammenfassung der einzelnen Diskussionsbeiträge. In (Alexander von Humboldt Stiftung Hrsg.): Publikationsverhalten in unterschiedlichen wissenschaftlichen Disziplinen. Beiträge zur Beurteilung von Forschungsleistungen. Alexander von Humboldt-Stiftung, S. 6–13, 2009.
- [Sc13] Schuh, G. et al.: Potenziale einer Forschungsdisziplin Wirtschaftsingenieurwesen. Herbert Utz Verlag, München, 2013.
- [St17] Storey, M.-A. et al.: Using a visual abstract as a lens for communicating and promoting design science research in software engineering: 2017 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement. ESEM 2017. IEEE, S. 181–186, 2017.
- [TDA19] Thuan, N. H.; Drechsler, A.; Antunes, P.: Construction of Design Science Research Questions. Communications of the Association for Information Systems 1/44, S. 332–363, 2019.
- [WD21] Wegener, L.; Dietrich, T.: Betreuung einer praxisorientierten Bachelorarbeit am Beispiel der Prozessoptimierung der Urlaubsverschiebung des fliegenden Personals im Flugbetrieb. In (Hattula, C.; Hilgers-Sekowsky, J.; Schuster, G. Hrsg.): Praxisorientierte Hochschullehre. Springer Fachmedien, Wiesbaden, S. 293–304, 2021.
- [We19] Werner, M.: Exploring Design Science Research Project Characteristics—An Initial Empirical Investigation: Proceedings of the 52nd Hawaii International Conference on System Sciences. HICSS 2019, 2019.
- [Wi08] Winter, R.: Design science research in Europe. European Journal of Information Systems 5/17, S. 470–475, 2008.

Digital Transformation in Educational Institutions: Scrutinizing the Debate and Highlighting Success Factors

A qualitative study on the current use of learning management systems and learning analytics in Germany

Linda Mai¹, Lynn Schmodde², Marius C. Wehner³

Abstract: In German education, digitalization is evolving, and the COVID-19 pandemic has accelerated this process due to closed schools and universities. While these external conditions were almost the same for all institutions, they have dealt with remote teaching in different ways and developed various new learning concepts. We conducted 27 interviews with headmasters, e-learning managers, and teachers from Germany to gain insights into varying attitudes towards learning management systems, as well as the motivations and concerns about future teaching with digital support. Using an inductive method, we identify different obstacles, including lack of digital competencies, necessity of technical equipment, concerns about replacing traditional lessons, but also enthusiasm for new opportunities. We argue that a meaningful use of learning management systems and learning analytics depends on an open and impartial debate on data protection and equal opportunities. Overall, this paper contributes to current research with proposition development about privacy rights and inequalities.

Keywords: Data protection, learning analytics, digital transformation, personalized learning, emergency remote teaching

1 Introduction

Learning Management Systems (LMS) played a vital role in Germany while schools and universities were closed in spring 2020 and winter 2020/2021 due to the COVID-19 pandemic. Unlike in other countries, home schooling is not common in Germany, hence, teachers and instructors were forced to explore new possibilities of teaching, which is now also

¹ Heinrich Heine University Düsseldorf, Junior Professorship for Business Administration, esp. Corporate Management, Universitätsstraße 1, 40225 Düsseldorf, linda.mai@hhu.de, <https://orcid.org/0000-0003-4847-4152>

² Heinrich Heine University Düsseldorf, Junior Professorship for Business Administration, esp. Corporate Management, Universitätsstraße 1, 40225 Düsseldorf, lynn.schmodde@hhu.de

³ Heinrich Heine University Düsseldorf, Junior Professorship for Business Administration, esp. Corporate Management, Universitätsstraße 1, 40225 Düsseldorf, marius.wehner@hhu.de

known as emergency remote teaching (ERT) [Ho20]. Virtually overnight, also students had to adapt to new types of learning, and they had to deal with the required level of self-management and self-organization as well as the technical setup at home. During this time, the so far underutilized LMS suddenly played an increased role for institutions, as they provide an inclusive learning environment for academic progress [Br21]. For example, teachers and instructors provided video clips, tasks, and materials, or got in contact with the students through videocalls. Hence, LMS offered the opportunity to bridge the distance between instructors and students. But this forced transformation was not without difficulties, which was not only the case in German institutions. A study from Sheffield Hallam and Glasgow Universities shows that a lot of students had difficulties with their emotional well-being, managing the workload as well as technological requirements [Ma22]. In Japan, ERT was dissatisfying for first year-students due to the lack of personal contacts, but overall, the LMS was considered as a satisfactory tool [La21]. Furthermore, there is evidence from Thailand that perceived usefulness, perceived ease-of-use, and behavioral intention to use are important factors for LMS usage [KC21]. Pre-pandemic research on LMS and the use of data with Learning Analytics (LA) already indicated that both LMS and LA offer many opportunities in a more digitalized education, especially about individualized teaching, and learning. LMS can provide documents and teachers can manage their courses. LMS also offer teachers information about the learning progress of learners. As LMS store information about activities in the virtual learning environment, this data can be used for an analysis of learners' individual learning behavior and prediction of success. LA can hence encourage the learning process through the identification of learners who may not pass a class [GSR17], support students to reflect on learning [Ag18]. LMS and LA are crucial tools for individualized teaching and learning, as it facilitates instructors to avoid a too general approach of teaching [Ag18].

However, especially in Germany, institutions currently do not exhaust all possibilities of LMS and LA due to different reasons, that include technical hurdles and reservations about new technologies. To understand the complex situation of digital transformation, there is a need for qualitative research to create new insights [GCH13]. We take on a stakeholder-oriented approach, as especially the development of LA needs the views of various critical stakeholders to ensure optimal deployment [BD16], [GDS15], and we explore which factors are conducive or obstructive to future use. By conducting 27 interviews, we investigate the needs and experiences of schools and universities with remote teaching, LMS and LA during the pandemic, and we derive important enablers for digitalization and identify obstacles for the process. In doing so, we map the ongoing debate in Germany about the implementation and usage of LMS and LA.

Hence, the purpose of this qualitative study is to examine the different promoting and inhibiting conditions for the use of LA in schools and universities. We focus on identifying underlying factors that influence the acceptance of LA. Although LA is not a permanent element in education yet in most institutions, it is important to figure out the differences between the institutions to identify the enabling environment for a conscious use. Therefore, we want to answer the following research questions:

- What are drivers and obstacles for digitalization and the use of LA?
- Which factors can make the transformation successful?

2 Conceptual background

Digital transformation will be an important responsibility for schools and universities in the next years. The process includes addressing various issues like learning from the COVID-19 crisis, developing guidance and support, fostering digital skills of educators, and evaluating practical guidelines [Eu20]. Several stakeholders are involved in this transformation process, especially when institutions want to implement new technologies based on LA [GD12]. Hence, focusing on the different stakeholders will reveal the different opinions and attitudes toward digital technologies. This chapter introduces the current debate about concerns and opportunities that emerge from transformations and recent technological developments. LA offers a variety of applications for schools and universities, for instance “monitoring and analysis, prediction and intervention, tutoring and mentoring, evaluation and feedback, adaptation, personalization, and recommendation, as well as reflection, concerning both the administrative area of the institutions as well as the teaching and learning process that takes place within the classroom” [OGA21]. This broad definition shows the different possibilities of implementing LA, which are based on personal data. Personal data is commonly unstructured, while analytics algorithms, in turn, offer useable insights and knowledge [Gr15]. At the same time, dashboards or warnings of dropping out can only be beneficial and effective with high data quality. Educational applications of LA can prevent students’ course failing by an “early warning system” that can motivate students to use support systems, such as individual treatments [SCB20]. All these analyses can be conducted in the LMS. The LA system, in turn, connects the learners, data, metrics, and intervention, building a closed feedback loop in the form of a cycle [Cl12].

At the same time, the implementation of machine learning methods in analytics applications can be challenging [SCB20]. For example, the use of personal data for success prediction might be problematic if the underlying algorithm design and configuration is influenced by human subjectivity [Mi16]. This underlying human subjectivity could create unfairness in case it is not giving all students the same opportunities [Mi16]. For institutions, offering additional support systems means making additional resources available, such as additional staff. Depending on the funding and the type of institutions (e.g., universities vs. schools), educational institutions often have to deal with limited resources, which can be problematic, as LA projects need well trained staff with knowledge, and there is a lack of staff available for such projects [If17]. This combination of potentially “unfair” prediction systems and limited resources could result in a problematic situation for at-risk students, as they would not receive the same support from their institutions.

However, these challenges are not only of technical nature, but also include instructors’ worries about privacy and ethical use of data: there is empirical evidence from 22 Austra-

lian universities that many institutional leaders and academic staff are worried about privacy and ethical use of data using applications like performance management and profiling of students and reported a lack of transparency about how and why LA are being used and data security [WHH16]. A few years ago, the daily use of an LMS was still at the beginning and staff was interested in using LA, but they also reported a lack of clarity, access, training, and time as barriers for using LA [We15].

Regarding students as equally important stakeholders, a study showed that students are generally interested in the use of LA systems but are also worried about the use of personal data, and they expect transparency about data being collected, the analysis and storage of data, as well as sensitive and responsible data management [IS19]. These concerns show the students' awareness of possible problems. Therefore, they should participate in the implementation process of LA in institutions [IS19]. In sum, both students as well as teachers and instructors as different stakeholders are important drivers of development, albeit with potential different interests in using different LA applications. Next to instructors' and students' expectations, the requirements for LA systems also include political aspects. There is a need for a political framework, but in Europe, “national policies for learners’ data or guidelines that govern the ethical usage of data in research or education” are still missing [No19]. With our study, we focus on the requirements for digital transformation in schools and universities with a particular focus on learning analytics. However, the identified issues are also relevant for other change and transformation processes in educational institutions. We focus on the transformation from ERT to well-structured online courses and LA applications in German schools. Specifically, we want to identify both concerns and opportunities.

The current literature provides little knowledge on how applications like predictive modelling for at-risk students serve the usefulness and problem centered approaches [WKO21]. Furthermore, there is need for work that focuses on problems with LA and data and keeps in mind that even interest of people, correct interpretation of data, use as intended and generalization of analytics can not be the solution to all problems [WSB21]. All in all, LA needs to take the complexity of the learning context into account by using models that both integrate data sources, instructor knowledge of the context and personalized learning support actions [Pa18]. Collaborative research and development with practitioners can help to understand the requirements and influence changes [RL20].

3 Method

3.1 Sample selection and data collection

To explore important drivers and obstacles for the ongoing process of digitalization in educational institutions, we use an inductive approach, as this procedure helps to identify

hidden factors for ongoing processes. Using the Gioia methodology, we map the complex dynamics in educational organizations, which helps to create and validate important constructs [GCH13]. Beginning in June 2021, we approached 183 schools and 35 universities via e-mail and telephone in Germany. Our sampling criteria included that the institutions already used a LMS for teaching or learning purposes. Until October 2021, we conducted 27 telephone interviews with persons who were familiar with the system in the institution or even responsible for the development of the LMS. We interviewed 14 persons from German schools (i.e., schools' headmasters and responsible teachers) as well as 13 persons from German universities (i.e., coordinators for e-learning, didactics, or media). There is a difference between the two groups: Universities have dedicated positions for the support of learning platforms. In schools, learning platforms are rather supported by regular teachers (e.g., for informatics), besides their actual jobs and in consultation with the school management. Therefore, the persons in our sample have different backgrounds and the use, support, and development of the platform plays different roles in their jobs. The schools' interviewees were 47 years old on the average ($SD=9,3$ years) and had between 5 and 30 years of working experience in schools. 11 of the interviewees were male and the institutions have used platforms for a period between 1,5 and 15 years. Some of the schools had no experience with LMS before the lockdown in March 2020. The universities' interviewees were on average 45 years old ($SD=9,8$ years) and had between 2 and 25 years of experience with working at a university. 11 persons of them were male and the universities have used LMS for a period between 7 and 20 years. In general, the universities started earlier with the use of LMS and all of them had experience with LMS when the institutions were closed during the pandemic.

Our conducted interviews were semi-structured, meaning that we asked the questions as openly as possible so that the respondents had the opportunity to answer very broadly. Starting with general questions about the current use of LMS and LA in their institutions, we also specifically asked questions regarding their evaluations of chances and risks regarding the use of LMS and LA. The interview ended with questions regarding the interviewees' personal outlook on the future as well as demographic questions. Throughout the interviews, we used the neutral terms "learning platforms" as well as "learning software", but avoided the expression "learning analytics", because this term could potentially be misunderstood and evoke negative attitudes. For instance, in conversations with schools, teachers immediately thought of unwanted surveillance.

3.2 Coding and data analysis

To analyze our interview material, we used the MAXQDA software. The first two authors worked independently for the 1st order terms, which represented a lot of categories close to the informants' terms. Afterwards, we compared all identified codes, and then generated the 2nd order themes. In this second step, the identified themes are reflected on a more theoretical level [GCH13]. Our 2nd order themes include future-oriented applications from the instructors' perspectives, wishes and ideas, as well as requirements for the future

application in terms of individual learning opportunities. Moreover, the 2nd order themes distinguish between the perceived risks of different users, experienced problems in recent years regarding both the platform and institutionally related barriers. After this step, we elaborated the aggregating dimensions of the 2nd order themes. They represent the overarching constructs and offer a summary of existing advantages and chances, restrictions and barriers, and future perspectives on LA as well as requirements for possible applications. To put the different identified themes in context with each other, we visualized them in a comprehensive framework (see Figure 1).

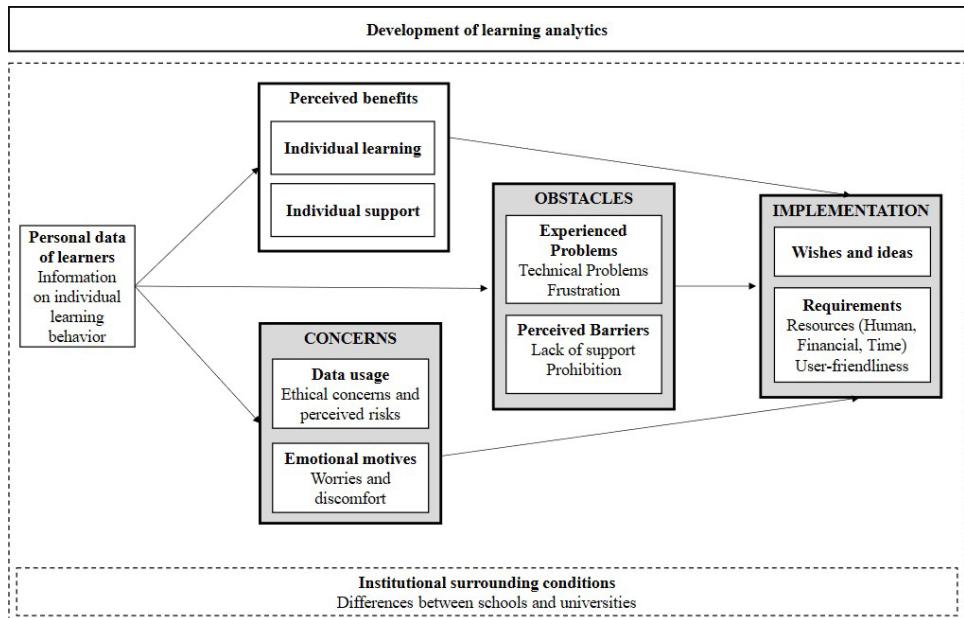


Fig. 1. Influences of LMS and LA usage in institutions

4 Results

Our analysis maps the current issues concerning the digital transformation in educational institutions that we derive from our qualitative interviews. We name the 1st order codes, which represent the informant-centric terms, and the 2nd order themes, which represent the researcher-centric concepts [GCH13]. The 1st order codes, directly derived from the interviews, demonstrate different aspects regarding the debate of the implementation of LA and LMS, which results from the different stages of experience and institutional factors, which we will discuss in the following. The derived 2nd order themes and aggregate dimensions, in turn, form the basis for further discussion and research, and cover a range of influencing factors for using LA in institutions. Figure 1 shows the different issues. The white boxes

show the 2nd order concepts with some essential examples. The grey boxes show the aggregate dimensions.

Research question 1

The concepts provide important clues about drivers and obstacles and can therefore be used to answer the first research question. As indicated in the graphic, it is important to work through these as an institution so that the next steps for implementation can then be taken.

The collection of personal data leads to emotional and rational motives that are also interrelated. For our analysis, we specifically distinguish between rational concerns, like ethical problems, and emotionally driven concerns. This helps us to understand what users really need or how an LA system can look like that appeals to all users. The benefits and opportunities, in turn, are related to the type of implementation. In the following, we will illustrate the different aspects along the identified themes around concerns about data usage, emotional motives, experienced problems, perceived barriers, wishes and ideas, and requirements.

Concerns about data usage

The ethical concerns include the problems with surveillance risk, data privacy and security gaps. Interviewees argued that using personal data violates the GDPR. Data outflows might lead to the problem that users do not trust the system anymore. The amount of the coded 1st order terms reflects the importance of data related concerns. For some interviewees it was also problematic that the context of the data is missing. The participants' statements could be covered by the following: *no use of personal data, purpose of data use is important, no access to individual data, personal data traces, use of LA is unclear, no use of LA in general, surveillance risk, data privacy, risk of data leaking to companies, frequency of video views, no log-data, activity progress, examination results provide key figures on study success, cohort comparison for teachers, context of the data is missing, security gaps, manipulation, no use of data to build up inhibitions*. These terms show a range of different attitudes toward the issue of date usage and reflect both strict negations, but also potential ideas for a meaningful data usage that includes clearly defined purposes, for example:

“And this creates data traces from the students that are already very personal”
(Interview_12).

Emotional motives

Our interviewees also expressed concerns that stem from emotional perceptions regarding the use of LA and possible risks. Some interviewees expressed their fear that traditional teaching or cultural competencies get lost by using LMS. They were also unsure if a system can support students better than a teacher, and they were worried about losing creativity and the opportunity for individual feedback and criticism. Specifically, digital teaching is perceived as a threat for the traditional way that schools and universities worked. At the

same time, some instructors compare themselves and their own proficiency with the digital system and do not perceive digital teaching as a valuable addition to their teaching. In contrast to the possibilities LA might have for supporting learning, a lot of teachers do not see the chances for improving teaching, but they rather felt they were in a “competition” with the systems. Some instructors also fear that students could lose social and learning competencies when using digital technologies. The amount of 1st order terms shows the different reasons for rejections and the weight of this issue: *restriction of creativity and criticism, automated tests are perceived as negative, differences intensify through automation, bad experiences with learning pathways, no added value through LA, complaints about evaluation of the activity, cultural techniques are lost, formalities must be clarified, real experiences are lost through digital teaching, writing competence suffers, support of the students needed, negative development of competences, intransparency in the event of problems, skepticism towards algorithms*. These terms reflect different worries and discomfort with the use of personal data, that may result from previous negative experiences, but also from indistinct worries like:

“The same thing - since I’ve been in the teaching profession for a while now, I can say - I would now say subjectively that the ability to concentrate and the ability to grasp things has decreased on average over a longer period of time” (Interview_4).

“These are massive disadvantages because simply these values or, how do I put it, these cultural techniques that you know so well, the classical ones. They are simply neglected. Whether it’s writing or this self-organization” (Interview_9).

“Depending on how you then implement it, it can be very streamlined and limit creativity and critical processes” (Interview_12).

Experienced problems

The first issue of the identified surrounding conditions is the aspect of stakeholders’ previous experience and resulting new opportunities. The recent years have been an intriguing time for education due to the COVID-19 pandemic. Therefore, a lot of changes had to be done to ensure ERT and it is possible that these changes created an overload for both instructors and students. Using new technologies can cause technical problems and frustration and at the same time a need to deal with issues like self-management individually. Issues that were mentioned by our interviewees were mostly related to students’ problems with LMS and LA in general, rather than pandemic-specific problems. The 1st order terms show the frustration and problems from a perspective outside teaching and learning: *technical hurdles, hacker attacks, processes are blocked, technical problems, time is a critical factor*. Therefore, using technologies to ensure better support in teaching could be a door opener for more applications like:

“So, on the one hand, having the technology under control, but on the other hand, also taking enough responsibility for yourself that you can also shape it. And many of our students can’t do that very well” (Interview_15).

Perceived barriers

The second issue of the surrounding factors is the problem of perceived barriers which include general prohibition and a lack of support. Data protection plays a dual role here. On the one hand, like described above, data security is associated with negative emotions like concerns regarding own safety. On the other hand, a very strict interpretation of data protection leads to prohibition and constitutes an obstacle for further development. Stakeholders can perceive data protection as problematic because it hinders any trials and tests and therefore also a user-friendly development. The 1st order themes were: *lack of support from the federal state, lack of collaboration between schools, prohibition by the university*. Although there are just three identified themes, these terms reflect barriers that can not be bypassed. At the same time, potential solutions to these problems must be found on a regional or federal level, for instance regarding financial support of institutions as well as the initiation of policies regarding data and privacy rights. Therefore, not only the institutions themselves but also the German government play a crucial role for the further development of LMS and the implementation of LA like:

“Yes, that not quite so much is blocked by data protection” (Interview_7).

Wishes and ideas

On the other hand, a lot of interviewees also highlighted the useful aspects of LMS and LA, which contrasts with the existing worries and negative attitudes. For instance, some interviewees mentioned the individualization of learning processes, which is a very important aspect of learning in schools and universities. Although there was a range of difficulties and concerns, the 1st order terms shows a more positive picture of future learning and a need for digital competencies: *fun while learning, preparation for professional future, promote problem solving skills, digital communication, digital competencies are starting advantages for students, students build up competencies, support of learning processes/human learning, access and added value through digital media, merging digital and face-to-face teaching, improved communication with students*. These applications and competencies show the motivation to improve systems and the wish for a stronger implementation of LMS and LA, for example:

“(…) where he has problems and I offer him solutions or exercises in this area. Individualized. I think these are things that will be extremely interesting and exciting next year” (Interview_8).

“Because the children are all different speeds and that cannot be standardized at all” (Interview_6).

Requirements

A further identified theme follows from the two previous issues and covers factors, such as resources, that are required for a successful implementation in institutions. Above all,

personal support from well-trained people is an important factor that ensures a meaningful use of LMS and LA, and hence can contribute to the acceptance of new technologies. In addition, time and money are also resources that are needed for a transition to LMS and LA, which, hence, an institution must provide to ensure a seamless implementation. The 1st order terms show the requirements on different levels with different responsibilities: *adapt teaching to the digital world, change of exams, clear communication about log-in data, further development by school administration, LMS must suit the students, intuitive application, rooms in school (for working), LMS needs people who are on fire for it, necessity of enthusiasm, communication between colleagues important for processes, learning-friendly LMS, equal opportunities must be guaranteed, new didactics, resources are needed (personnel, training), time-consuming training for LMS, time-intensive digital concepts*. These themes show how many processes in institutions should be taken into consideration when developing a global approach and that the cooperation of many different stakeholders (e.g., instructors, school administration, technicians) are crucial for a useful implementation of LMS and LA, for example:

“It is not enough to provide platforms, but the staff must be trained, I need appropriate contact persons, I need didactics who are familiar with the e-tools and know how to transform classroom teaching into digital teaching” (Interview_13).

Research question 2

The concepts provide initial indications of the challenges institutions face in taking advantage of the opportunities offered by technological support. To answer the second question, focusing on the connections between the concepts is important. Thus, it became apparent that perceived advantages and disadvantages by no means cancel each other out. Rather, it is the case that concerns, and obstacles must first be dealt with in order to make the advantages clear to users and to make use possible in the first place in the case of absolute prohibitions. Here it is important to find clear rules and clear communication with all stakeholders involved. We therefore see critical factors in this. In a second step, the wishes, requirements and needs of the users can be included in the further development.

5 Discussion

5.1 Implications

The results from our 27 interviews provide insights into the current debate among schools and universities regarding the implementation of LMS and LA, and offer multiple implications. Our findings also reveal research questions that are still unclear and where further research is needed to understand the users' attitude. First, we highlight the ethical discussion

around the implementation of these systems. Second, we shed light on the relevant factor of communication among stakeholders as well as legally protected guidelines. Finally, we comment on the need of resources (i.e., personnel, financial) and adequate training of stakeholders for a meaningful development of LMS and LA. As a general note, all results should be interpreted in the light of the different contexts of schools and universities, which should be considered when developing guidelines and recommendations.

First, for a conscious communication with all stakeholders, it can be helpful to address possible problems directly and show how the implemented system(s) meet these concerns. As we show, even unspecified concerns of stakeholder can cause feelings of uncomfortableness. Therefore, it is important for all responsible persons in institutions to minimize concerns among their employees by using straightforward instructions, using legally protected guidelines, and avoid increasing ambiguity in their communication. As an example, communicating “predictive analytics” is a technological expression which can be perceived as unpersonal, but offering additional support through “proactively identifying struggling students” can be perceived as helpful [WKO21]. For a better understanding of the underlying mechanisms of communication with different stakeholders, further research is needed on risks and chances of using distinctive terms and language.

Second, all users (i.e., instructors, students, responsible persons in educational institutions) need to have time and training with the implemented systems to create positive experiences. This process requires well-trained contact persons for the users’ special questions and further training in general for all users. Specifically educational staff plays a key role when it comes to the success of digital transformation as well as a conscious and reasonable communication. The involvement of all stakeholders forms the basis for building up trust and acceptance among all users, which increases the chances of a meaningful implementation of new systems. Therefore, the emotional factor should be considered.

Third as our results show, the success of LMS depends on how different issues are addressed. The identified emotional and rational motives of the interviewees are closely related. Existing rational issues can lead to unclear worries and can cause negative attitudes toward technologies. To better categorize the ethical issues with new technologies, it is helpful to apply a framework [Mi16].

Mittelstadt et al. (2016) show how algorithms create evidence from data for given outcome, which motivates actions and can increase the responsibility for actions driven by algorithms. However, algorithms rely on correlations rather than causal knowledge which can lead to unjustified actions [Mi16]. For the educational context, this can mean that students do not get the support they need, because the personal reasons for a specific learning behavior are not considered in automatic decisions. At this point, positive experiences can help users better than technical explanations. This is exacerbated by the fact that algorithmic decisions can be biased like every other human decision due to social, technical, and emergent biases, which means that algorithms cannot compensate biases in human decision-making [Mi16]. This includes, for example, the human bias of instructors, which reduces their judgment accuracy [UW21]. Additionally, informational privacy can be challenging

[Mi16]. This can have a twofold effect for education. Personal data often is necessary for useful analysis, but the European data protection law (General Data Protection Law, 2016/679) is strict and unnecessary personal identification is problematic. Nevertheless, opportunities of the implementation of LMS and LA should be taken into consideration.

5.2 Limitations

The Gioia method comes with some limitations, that should be taken into consideration [GCH13]. First, as a qualitative method, statements about relative proportions are difficult. Additionally, this method does not analyze every aspect of the interviews, because the aim is the identification of more general topics for further research. Therefore, further studies that focus on specific facets of the themes that we discuss here could yield more detailed insights. Second, the timing of the interviews could be a constraint, as many institutions found themselves in the situation between pandemic ERT and a more “conscious” development of new teaching and learning concepts. Hence, it could be interesting to repeat the interviews after a few years to ascertain whether the interviewees changed their perspectives on LMS and LA. Third, personal attitudes, for instance towards privacy, rights can change over time [PS14].

6 Future research and practical implications

Our qualitative research implies three main research avenues. First, we would like to shed light on the important role of teachers and instructors in the digitalization progress. As our interviews show, teachers notice a change of the traditional understanding of their profession. This, in turn, provides a novel—and hence, uncertain—experience for them. At the same time, their evaluation of new systems is a crucial driver for the implementation of new types of teaching and learning. Hence, the overall success of LA strongly depends on the question of which factors will help in supporting teachers’ and instructors’ positive evaluations of those novel technologies and systems. Future research should focus on investigating variables that will foster the enthusiasm of teachers and instructors and diminishing the perceived obstacles that dominate the debate at the moment.

Second, our research gives insights into the role of students in the ongoing change in the educational system and provides avenues for further research on student-oriented teaching. The COVID-19 pandemic provided a stressful situation for young people, which was reinforced due to the switch to distance learning, and the dramatic change of formerly well-known concepts of learning, studying and instruction. The results of our qualitative research show that many solutions during this time were implemented as emergency solutions, hence it is difficult to disentangle the influence of the overall stressful situation on the individual perceptions of LMSs. For instance, as our interviewees mentioned, some students were not only concerned about their learning achievements, but also about the

lack of social contacts. This, in turn, provides a fruitful path for future research, as it has to be worked out which digital learning concepts will really hold in the future and can be further developed for a useful application under more “normal” circumstances with re-opened schools and universities.

The third aspect that we would like to highlight focuses on the topics of data safety and protection, which will be a decisive point for further development. The results of our qualitative research indicate that the debate on LMSs and LA is strongly marked by negative attitudes towards these technologies due to data safety and privacy concerns. Therefore, the first step in approaching these obstacles is the (political) agreement on data directives in Germany which specifically focus on the usage of LMSs and LA. While the interviewees admitted that LMS in general are used more than before the pandemic, this did not lead to a heightened employment of LA. At the same time, we believe that the debate on LMSs and LA and its upsides will be strongly inhibited as long as stakeholders mainly focus on data and privacy concerns, which is why we call for further research to understand the existing negative attitudes, and to search for ways to improve communication of the benefits of new technologies to support the development.

To address these challenges in practice, we recommend the collaboration of different stakeholders, including experts from IT, didactics, and pedagogics. Additionally, it is important to reflect the habits of different institutions, as they are in demand of different strategies. As we already mentioned, time and staff are critical factors for a successful digital change. Due to their size, universities often have special centers for e-learning, but smaller schools are even more dependent on external financial support and well-trained staff to facilitate the implementation of LMSs and LA, as well as to monitor and audit data regulations. For universities, it is challenging to implement new technologies as a “one-size-fits-all” solution due to the habits of different faculties. Therefore, political stakeholders should rethink the problem of sufficient resources in schools and universities.

This implicates that there is a need for clear government positions on legal concerns. This would provide planning security for institutions to manage the implementation and development of the systems within the legal framework, while also providing benefits for students and instructors. Only if institutions will be provided with both resources and legal safety, they will be able to access digital opportunities.

References

- [Ag18] Aguilar, S. J.: Learning Analytics: at the Nexus of Big Data, Digital Innovation, and Social Justice in Education. *TechTrends* 1/62, S. 37–45, 2018.
- [BD16] Buckingham Shum, S.; Deakin Crick, R.: Learning Analytics for 21st Century Competencies. *Journal of Learning Analytics* 2/3, S. 6–21, 2016.
- [Br21] Bradley, V. M.: Learning Management System (LMS) Use with Online Instruction. *International Journal of Technology in Education* 1/4, S. 68, 2021.
- [Cl12] Clow, D.: The learning analytics cycle. In (Dawson, S.; Haythornthwaite, C. Hrsg.): *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*. ACM Press, New York, New York, USA, S. 134, 2012.
- [Eu20] European Commission: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. *Digital Education Action Plan 2021-2027. Resetting education and training for the digital age*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0624&from=EN>.
- [GCH13] Gioia, D. A.; Corley, K. G.; Hamilton, A. L.: Seeking Qualitative Rigor in Inductive Research. *Organizational Research Methods* 1/16, S. 15–31, 2013.
- [GD12] Greller, W.; Drachsler, H.: Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Journal of Educational Technology & Society* 3/15, S. 42–57, 2012.
- [GDS15] Gašević, D.; Dawson, S.; Siemens, G.: Let's not forget: Learning analytics are about learning. *TechTrends* 1/59, S. 64–71, 2015.
- [Gr15] Grindrod, P.: Mathematical underpinnings of analytics. Theory and applications. Oxford University Press, Oxford, New York, 2015.
- [GSR17] Gašević, D.; Siemens, G.; Rose, C. P.: Guest Editorial: Special Section on Learning Analytics. *IEEE Transactions on Learning Technologies* 1/10, S. 3–5, 2017.
- [Ho20] Hodges, C.; Moore, S.; Lockee, B.; Trust, T.; Bond, A.: The Difference Between Emergency Remote Teaching and Online Learning. <https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning>, Stand: 12.04.2022.
- [If17] Ifenthaler, D.: Are Higher Education Institutions Prepared for Learning Analytics? *TechTrends* 4/61, S. 366–371, 2017.
- [IS19] Ifenthaler, D.; Schumacher, C.: Releasing Personal Information Within Learning Analytics Systems. In (Sampson, D. et al. Hrsg.): *Learning Technologies for Transforming Large-Scale Teaching, Learning, and Assessment*. Springer International Publishing, Cham, S. 3–18, 2019.
- [KC21] Kaewsaiha, P.; Chanchalor, S.: Factors affecting the usage of learning management systems in higher education. *Education and Information Technologies* 3/26, S. 2919–2939, 2021.
- [La21] Lafleur, L.: JALT Postconference Publication - Issue 2020.1; August 2021. *JALT Post-conference Publication* 1/2020, S. 125, 2021.

- [Ma22] Marshall, E. M. et al.: The impact of remote teaching on statistics learning and anxiety. *MSOR Connections* 1/20, S. 90–101, 2022.
- [Mi16] Mittelstadt, B. D. et al.: The ethics of algorithms: Mapping the debate. *Big Data & Society* 2/3, 205395171667967, 2016.
- [No19] Nouri, J. et al.: Efforts in Europe for Data-Driven Improvement of Education – A Review of Learning Analytics Research in Seven Countries. *International Journal of Learning Analytics and Artificial Intelligence for Education (iJAI)* 1/1, S. 8, 2019.
- [OGA21] Oliva-Cordova, L. M.; Garcia-Cabot, A.; Amado-Salvatierra, H. R.: Learning Analytics to Support Teaching Skills: A Systematic Literature Review. *IEEE Access* 9, S. 58351–58363, 2021.
- [Pa18] Pardo, A. et al.: OnTask: Delivering Data-Informed, Personalized Learning Support Actions. *Journal of Learning Analytics* 3/5, 2018.
- [PS14] Pardo, A.; Siemens, G.: Ethical and privacy principles for learning analytics. *British Journal of Educational Technology* 3/45, S. 438–450, 2014.
- [RL20] Reeves, T. C.; Lin, L.: The research we have is not the research we need. *Educational Technology Research and Development*, S. 1–11, 2020.
- [SCB20] Smith, B. I.; Chimedza, C.; Bührmann, J. H.: Global and Individual Treatment Effects Using Machine Learning Methods. *International Journal of Artificial Intelligence in Education* 3/30, S. 431–458, 2020.
- [UW21] Urhahne, D.; Wijnia, L.: A review on the accuracy of teacher judgments. *Educational Research Review* 32, S. 100374, 2021.
- [We15] West, D. et al.: Higher Education Teachers' Experiences with Learning Analytics in Relation to Student Retention. *Australasian Journal of Educational Technology*, 2015.
- [WHH16] West, D.; Huijser, H.; Heath, D.: Putting an ethical lens on learning analytics. *Educational Technology Research and Development* 5/64, S. 903–922, 2016.
- [WKO21] Wise, A. F.; Knight, S.; Ochoa, X.: What Makes Learning Analytics Research Matter. *Journal of Learning Analytics* 3/8, S. 1–9, 2021.
- [WSB21] Wise, A. F.; Sarmiento, J. P.; Boothe Jr., M.: Subversive Learning Analytics. *LAK21: 11th International Learning Analytics and Knowledge Conference*, S. 639–645, 2021.

Process Mining Model Quality in Software Development Case Studies

An Analysis

Marianne Schmolke¹

Abstract: With increasing digitalization speed, the need for knowledge about how Software Development can be improved, increases as well. Frequently encountered challenges are skill deficits, problematic behavior or activity sequences, and deviations from common and expected paths. Process Mining provides promising opportunities to remedy these issues. Process Mining applied to Software Development activities for example is an analytics method that can provide insights before a process has been executed as there is a huge amount of log files to be potentially analyzed. As promising case studies in this field have been conducted, there is a need to find quality measures to effectively learn from the results found. Therefore, a systematic analysis has been applied to the relevant sources using the relevant criteria of the Comprehensive Process Model Quality Framework (CPMQF).

Keywords: Process Analytics, CPMQF, Software Engineering, Model Quality Measures

1 Introduction

Practical informatics is a field which is expected to produce solutions to real-word problems. Software development in practical informatics therefore aims at improving certain use case scenarios and classically at producing viable software products. As this is a primary focus, the quality of case study process models seems to have less importance. As technologies have come up to ease the analysis of large data amounts, e.g. log data in a significant manner, this priority does not seem applicable anymore.

This paper aims at providing guidelines for good Process Mining Case Study Models in order to enrich their fit-for-use-level. In case of a high fit-for-use level models can be reused on a strong foundation and explanations based on these models increase their impact for the scenarios in question.

¹ Hochschule Wismar, Abteilung Wirtschaftsinformatik, Philipp-Müller-Str. 14, 23966 Wismar, Marianne.Schmolke@hs-wismar.de

A powerful instrument of automatically constructing process models from log data is Process Mining. Generally, Process Mining aims at building a process model in order to describe the behavior contained in event logs of information systems ([Va05]). Event logs are produced by process-aware information systems (e.g. Workflow Management Systems). Usually, these event logs contain information about the start and completion of process steps together with context data, e.g. who executes a task and what is necessary to execute the task.

[Ca20] found that only 1 study (3.12%) in their Literature Review on Software Development Analytics in Practice used Process Mining methods and tools.

That means that there hasn't been a sufficient amount of studies indicating empirical quality as result of collective experience in the Process Mining Analytics domain.

2 Related work

2.1 Overview of Secondary Studies

A set of literature analyses has been conducted in order to structure the field:

[Ru07] focus on technical possibilities of mining software process information. They work with abstracted and algorithmically processed event logs. Their output is the characterization of developers in the software engineering process as well as insights about problematic behavior or activity sequences.

[No16] analyze classical Process Mining applications. They extract hidden information about the Software Development process and analyze event logs of the types: problem report, change request, feature request and information request. They then identify typical application fields, case study categories and the challenges: Process knowledge, data quality, tool integration, result evaluation and usability.

[Ca16] conduct a meta-analysis on Process Mining in Software Development. They identify high or low detail and accurateness of Software Development processes and investigate process and social network patterns of relevance for generating improvement suggestions.

[Do17] conduct a systematic mapping study on Process Mining Software Processes. They analyze 40 papers dealing with the relevant research topics which can be classified into Micro and Macro Process papers. They aim at identifying the structure of and connections between core Process Mining Software Process elements.

[Ur21] have the goal to identify the perspectives of Process Mining, its definitions, the

application of Process Mining within the Software Engineering and successful practical implementations in the Software Development process. They analyze 12 papers, most importantly complying to the search criteria: (“Process Mining” OR “Process Mining Perspective” OR “Software Engineering”) AND (“Control-flow perspective” OR “Organizational perspective” OR “Case perspective” OR “Time perspective”). As a core result they extract several types of definitions and degrees of practical application.

An analysis which has not been conducted is the application of a process model framework particularly to Process Mining Software Development process models. To close this gap is the overall goal of this paper.

2.2 Detailed findings from primary studies

The body of literature encompasses a set of findings regarding the well-founded application and interpretation of Process Mining models in Software Development:

The approach was mostly evaluated as beneficial to Process Management System (PMS) implementation conditions. [Ru07] identified the possibility of a ‘faster PMS creation’. [Ru14b] suggested to embed the model into the development lifecycle due to effectiveness, [Ca19] found that their model accuracy was sufficient. The other works did at least not falsify this assumption and presented theoretical or practical benefits e.g. as optimization examples.

General challenges and advantages of Process Mining Software Development were inter alia:

- [Ma18] e.g. found that it is hard to evaluate if a team is self-organized based on the information contained in an event log. There is still the need for more heuristics to identify typical log characteristics of self-organized teams.
- [Ma21] e.g. found that the longest path of executive deletion commits had to be done manually. They as well found similar sequences of commits. Additionally, most and least used activities and traces easy to find.

The optimization of developing environments can inter alia be supported by these findings: [Po11] recommend the separation between the preprocessing and the analysis step. The reuse of analysis techniques is recommended.

- [Se14] found that the identification of important user roles, ‘check requirement’, ‘define specification’, ‘define solution’, ‘check impact’, ‘define testing strategy’ need to be mandatory. Additionally, the context and details need to be understood.

Insights regarding the developer learner models are:

- [Ca19] found that the more proficient a team the less complex their models.
- [Ca21] found these developer fingerprints: high performers: cautious coder, cautious coder/test skipper, insecure/testers, insecure/debuggers, balanced coders/confident; low performers: general coding limitations (2), limited python/algorithms skills (2).

Specifically, teams using Scrum could profit from the following workflow optimizations. [Ma18] found that loops deserve special attention and identified the possibility to:

- check whether all issues are prioritized and assigned to some resources;
- identify Scrum roles, even though they were not recorded in the case-handling system by analyzing the organizational perspective;
- assess cross-functionality of the team based on patterns in a dotted chart (they were evaluated with the project manager).

2.3 Increasing Process Mining Model Result Quality

Regarding the more technical perspective these works have to be considered, as they provide practical guidelines regarding (partly specific) thresholds: [Ze18] find that there is no specific threshold most suitable for sequence encoding, as this greatly depends on the event log. They observe that using lower threshold values, e.g. 0–0.4, leads to less complex models. They recommend to use a lower threshold value first, which also reduces computation time due to a smaller constraint body size, and based on the obtained result increase or decrease the threshold value if necessary. Recommendations for choosing the right dependency measure, frequency and best dependency measure applying the particular approach of the Heuristic Miner can be found in [We06].²

² A dependency measure above the dependency threshold value [WE06]; frequency higher than the positive observations threshold value (*ibid.*), and a dependency measure for which the difference regarding the best dependency measure is below the relative to best threshold value (*ibid.*).

An example of applying a framework to assess Process Mining result quality is [Fa22]. They apply the SEQUAL framework to analyze whether syntactic, semantic, or pragmatic issues occur including deadlocks, lack of synchronization, superfluous activity, labeling issues, line crossing, and crooked alignment. The approach used here which is elaborated in the next chapter is based on the model extracted from previous works in [Me18].

3 Approach

3.1 Applied Framework

The focus lies on papers where Process Mining and Software Development are an essential component of the article – in the sense that the terms appear in the document title.

The criteria of the Comprehensive Process Model Quality Framework are assessed regarding their applicability to the Process Mining domain and then applied to the models found according to these search criteria:

Reviewed (full paper) English articles containing at least one Process Model including the terms “Process Mining” AND “Software Development” OR “Software Engineering” in Title from or key words from 2007 on from the following databases: ACM Digital Library (dl.acm.org), IEEE Xplore (ieeexplore.ieee.org), Proquest (proquest.com), ScienceDirect (sciencedirect.com), Wiley Online Library (onlinelibrary.wiley.com), Springer (link.springer.com) and Google Scholar (scholar.google.com).

3.2 Model analysis

The model developed in [Me18] encompasses the state-of-the-art body of knowledge regarding process model quality.³ The aim is to extract the criteria applicable to Process Mining models generated on a scientific basis.

	Title	Publication	Year	Source	Ref
a)	Using Process Mining in Software Development	Journal	'11	IEEE	[Le11]
b)	Conformance analysis on Software Development: an experience with process mining	Journal	'11	Int. J. of Business Process Integration and Management	[Cr11]
c)	Process Mining Multiple Repositories for Software Defect Resolution from Control and Organizational Perspective	Conference	14	MSR 2014	[Gu14]

Tab. 1: Papers (P) I

³ High resolution PDF of the model: <https://www.janclaes.info/downloads/CPMQF.pdf>

	Title	Publication	Year	Source	Ref
d)	Process Mining support for Capability Maturity Model Integration-based software process assessment, in principle and in practice	Journal	14	J. Softw.: Evol. and Proc.	[Sa14]
e)	Applied process mining in Software Development	Conference	14	SACI	[Se14]
f)	The Discovery of the Implemented Software Engineering Process Using Process Mining Techniques	Journal	16	IJACSA	[Za16]
g)	An Exploratory Study on Usage of Process / Mining in Agile Software Development	Conference	17	SPICE	[Er17]
h)	A new perspective on Process-oriented Software Engineering based on BPMN Process Mining	Conference	17	AITEM	[Ca17]
i)	Assessing Agile Software Development Processes with Process Mining: A Case Study	Conference	18	IEEE Business Informatics	[Ma18]
j)	Evaluating coding behavior in Software Development processes: A process mining approach	Conference	19	ICSSP	[Ar19]
k)	Assessing Software Development Teams' Efficiency using Process Mining	Conference	19	ICPM	[Ca19]
l)	Profiling Software Developers with Process Mining and N-Gram Language Models	Journal	21	J. o. Systems and Software	[Ca21]
m)	Using process mining for Git log analysis of projects in a Software Development course	Journal	21	Educational and Information Technologies	[Ma21]

Tab. 2: Papers (P) II

Title	Publication	Year	Source	Ref	Title
n)	Application of Constructive Modelling and Process Mining Approaches to the Study of Source Code Development in Software Engineering Courses	Journal	21	J. o. Communications Software and Systems	[Sh21]
o)	Business process analysis of programmer job role in Software Development using process mining	Conference	22	ISICO	[Fa22]

Tab. 3: Papers (P) III

Regarding the papers not presented in the selection (Tab. 1 - Tab. 3): [Ru14_a] illustrates how Process Mining models can look like in agile development and [Ru14_b] in Software Development in general without focusing on the specific outcome of a model.

In [Sa09] and [Ba18] no process model has been generated. [Po11] does not stress the software engineering/ development perspective. These terms do not appear in title or key words in contrast e.g. as in [Gu14].

These criteria were not assessed: Proactive modeling is not tested in detail – as the recommendation to use Process Mining is being met by all analyzed studies. The syntactical rules of Process Mining methods are applied, i.e. they are not questioned regarding their quality. The optional guidelines are left out in order to focus on the mandatory guidelines.

The following criteria (C) were assessed from which a selection was kept for the analysis:

10 Process Modeling Guidelines – Error probability and understandability

1. Use no more than 31 nodes -> applicable
2. Use no more than 3 in-or outputs per connector -> not applicable
3. Use no more than 2 start and end event -> fulfilled by all models
4. Model as structured as possible -> mechanism of automatic model generation
5. Use design patterns to avoid mismatch -> Process Mining can discover design patterns, e.g. [Ba03] but the models are not adjusted by using design patters. That means, this criterion does not apply.
6. Avoid OR-joins and OR-splits -> not applicable
7. Minimize the heterogeneity of connectors -> not applicable
8. Minimize the level of concurrency -> not applicable
9. Use verb-object activity labels -> applicable
10. Decompose a model with more than 31 nodes -> not applicable

These criteria were extracted:

1. Use no more than 31 nodes
9. Use verb-object activity labels

C P	a)	b)	c)	d)	e)	f)	g)	h)	i)	j)	k)	l)	m)	n)	o)
1	y	x	y	y	y	y	x	y	y	x	y	y	y	y	y
9	0 ⁴	y	x	x	x	y	y	x	x	x	y	x	y	y	y

Tab. 4: Applicable Criteria; legend: x=not fulfilled, 0=not applicable y=fulfilled

These criteria sets were not assessed: The Gateway Complexity Rules cannot be influenced in a Process Mining model. The gateways that are mined have no alternative in the representation.

Regarding the patterns used, the explanation given regarding criterion 5 from the “Process Modeling Guidelines – Error probability and understandability” applies.

“Reaching GoM – Basic guidelines -> Syntactic correctness” does not apply as assessment criterion to Process Mining. The models have their own validation mechanisms.

1. Document the meta-model. The overall meta-model for Process Mining can be found in [Po11] modified form [Do11].
2. Use a tool for automatic syntactic checks -> Does not apply to Process Mining where the model already is automatically generated.

Reaching GoM – Basic guidelines -> Semantic correctness

1. Use clearly defined domains and scopes -> applicable
2. Use internal and external feedback loops -> applicable
3. Try to capture the ‘soft’ process-related issues -> applicable

C P	a)	b)	c)	d)	e)	f)	g)	h)	i)	j)	k)	l)	m)	n)	o)
1	y	x	y	0	y	y	y	x	y	y	y	y	y	y	y
2	x	x	y	y	y	y	y	y	y	y	y	y	y	y	y
3	x	x	y	y	y	y	y	y	y	y	y	y	y	y	x

Tab. 5: Applicable Criteria
x=not fulfilled, 0=not applicable y=fulfilled

4 Markov-Chain notation

In the following, the models found in the papers listed in (Tab. 1 - Tab. 3) are assessed.

[Le11] (a) explore the domain of a Software Development Process in a large-sized Brazilian Software Company. The scope is to exclude the performance perspective due to lack of duration data of the logs. As this is not the focus of the study, it does not cause issues. Regarding the internal and external feedback loops (C2) [Le11] already reflect the gap: “A set of activities being executed in a process instance are only a subset of the activities defined in the formal process. So, a deeper conformance analysis is still required.” Therefore, this criterion is not fulfilled. ‘Soft’ process-related issues can be interpreted as less quantifiable, i.e. perceptual and social aspects, mostly addressed by change management initiatives in software projects. [Le11] relate to rather cost-related issues. The category 3 could have been considered more thoroughly.

[Cr11] (b) clearly state that they mainly want to obtain an understanding of the challenges involved in the process. They refer to a specific Software Development project but do not abstract the investigated domain or scope. The authors thoroughly reflect their results which confronted them with certain interpretation challenges. They state that further investigation is needed. Due to the lack of evidence in the model, conclusions on ‘soft’ process-related issues were not possible.

[Gu14] (c) discover a runtime process model for a bug resolution process spanning three repositories using the process mining tool, Disco, and conduct process performance and efficiency analysis. They identify bottlenecks, and detect basic and composite anti-patterns. In addition to control flow analysis, they mine event logs to perform an organizational analysis and discover metrics such as the handover of work, subcontracting, joint cases and joint activities. Regarding C2 they use a broad selection of metrics to test the validity of the model which can be seen as an equivalent to feedback cycles. ‘Soft’ process-related issues are addressed e.g. by differentiating generalists and specialists and observing that social performers contribute more actively.

[Sa14] (d) deliberately choose a broad scope using ‘complexity’ of the selected processes as one criterion. This approach results in a more general model than if the scope definition was more restricted. Therefore, this criterion is not applicable here. Regarding the feedback, several other analyses for additional projects were conducted in order to validate their model. They closely observe the corresponding real world process. They used their insights for actively enabling the development team considering the ‘soft’ process-related issues.

[Se14] (e) analyze the case of a software product release involving an external company. They explain their scope in detail and define all relevant elements clearly. Their real-world model is statistically analyzed in a way that the internal feedback criterion is fulfilled. They address ‘soft’ process-related issues by e.g. identifying training necessities and discovering that the support team did not ensure if the project team has understood the context or the details of the requirements.

[Za16] (f) focus on a subset of the project life cycle by the MS Visual Studio Scrum 3.0, i.e. the work items: feature, product backlog item, bug, and task. They apply quality dimensions to the model (Pareto Front model). The log is well described according to accepted thresholds whereas much more behavior than recorded can occur. The ‘soft’ process-related issues are discovered by interpreting correlations. It is assumed that the scrum team is not adhering to the process definition. Whereas specific work items have to be analyzed individually.

[Er17] (g) present a model of an agile Scrum team executing Product Backlog Items, i.e. user stories or specifications inter alia. Scope and definition are clear.

The relevant patterns are analyzed by the team as a feedback mechanism and ‘soft’ process-related issues are considered, e.g. the analysis is used as an approach to reflect one’s own work as an agile team member.

[Ca17] (h) use the domain of a manufacturing company to conduct a conformance analysis regarding an initially BPMN-mined process. They do not name the scope of the initially mined model. Still, their conformance testing is conducted in a mainly quantitative manner. They conclude that nonconformities can be difficult to detect if only predefined KPIs are used to control the process. Nevertheless, the ‘soft’ process-related issues, such as, discussing the warehouse’s personnel oversight (‘can generate a big delay’+ ‘probably represents a local problem’) are considered.

[Ma18] (i) analyze two projects in an IT organization providing software products and services whereas Jira Software and Scrum are used. They apply a set of mining methods to test the models. They list several conclusions which also include ‘soft’ process-related issues.

[Ar19] (j) create high- and low-performer models of Software Development students of an advanced Java course. They use grading as a measurement to classify the students with a certain mined process profile. They consider ‘soft’ process-related issues by e.g. pointing out graphical feedback possibilities.

[Ca19] (k) create a model of practitioner activities executed in an IDE, including its artifacts and additional details on the ecosystem of components supporting the process. They use the internal and external validity criterion to assess the quality of their results and find clear threats to validity, i.e. idle time could not be interpreted and due to study organizational reasons the students might not have been the adequate surrogates. They clearly find ‘soft’ process-related issues, e.g.: the less complexity in the models, the more proficient students were in the task.

[Ca21] (l) create developer profiles according their development process.

They use N-gram language models and text mining in addition to the Process Mining models. The developer classifications include ‘soft’ process-related issues.

[Ma21] (m) create models of the project development process from the perspective of Git log attributes. Their feedback method was a manual analysis of the discovered model. They find ‘soft’ process-related issues such as inadequate distribution of work as a reason for deadline issues.

[Sh21] (n) construct a model about the task to write a program to calculate the minimum number of coins needed to give change to a customer. They use a formalized approach including code history and a direct experiment with students. They consider ‘soft’ process-related issues, e.g. factors complicating the program development process- which are leaving certain decisions to the teacher in the end (i.e.: grading).

[Fa22] (o) work on finding out the flow carried out by programmers in Software Development courses. They apply the SEQUAL framework to assess the quality of the process model. E.g. one of their main result components (issues of business process quality assurance) does not consider ‘soft’ process-related issues – as it is not in their scope.

Reaching GoM (Guidelines of Modeling) – Basic guidelines -> Relevance

1. Have sufficient constructs in the meta-model to represent the elements of the real world -> Applies to Process Mining automatically as the real-world processes are the center of the mining activity.
2. Use continuous feedback loops -> The constant repetition of mining activities can be considered as feedback loops.

Reaching GoM – Basic guidelines -> Economic efficiency

1. Use process scope definitions and clear objectives and targets -> similar to C1 of “Reaching GoM – Basic guidelines -> Semantic correctness”. The objective and target are similar throughout all papers as they are all related to process mining in software engineering – aiming at finding optimization scenarios.
2. Use business frameworks -> Frameworks such as Porter’s 5 Forces Model or BCG Growth-Share Matrix inter alia are clearly related to classical business goals. Optimizing with Process Mining is not only limited to these. Optimization goals in Software Engineering cannot be clearly categorized into traditional business frameworks. E.g. software architectures can be assessed in terms of technical debt ([Li19]), indicating the long-term vision on its profitability. Clearly quality can often be valued opposed to a short-term cost-benefit ratio.
3. Select only relevant users to participate in feedback loops->This criterion does not apply because the feedback loops are rather automatized by the mining approach.
4. Re-use models -> This is a core element of Process Mining.
5. Use reference models -> This is a core element of Process Mining.
6. Use state of the art modeling tools -> Different contexts require different mining algorithms. They are chosen for optimization reasons which comply to the state-of-the-art requirement.

Summary of applicable criteria:

1. ***Use no more than 31 nodes***
9. ***Use verb-object activity labels***
1. Use clearly defined domains and scopes
2. Use internal and external feedback loops
3. Try to capture the ‘soft’ process-related issues

C P	a)	b)	c)	d)	e)	f)	g)	h)	i)	j)	k)	l)	m)	n)	o)
I	y	x	y	y	y	y	x	y	y	x	y	y	y	y	y
9	0 ⁵	y	x	x	x	y	y	x	x	x	y	x	y	y	y
1	y	x	y	0	y	y	y	x	y	y	y	y	y	y	y
2	x	x	y	y	y	y	y	y	y	y	y	y	y	y	y
3	x	x	y	y	y	y	y	y	y	y	y	y	y	y	x

Tab. 6: These criteria where assessed per case study model

The criteria automatically fulfilled by applying Process Mining have are:

- Use no more than 2 start and end event
- Model as structured as possible
- Document the meta-model
- Have sufficient constructs in the meta-model to represent the elements of the real world
- Use continuous feedback loops
- Re-use models
- Use reference models
- Use state of the art modeling tools

An applicable quality indicator is to have at least four criteria fulfilled. This applies to the majority of models. There does not seem to be a connection between time of publication and quality although more recent studies generate high quality (5 times “y”) models.

⁵ Markov-Chain notation

4 Conclusions

In general, it can be stated that the Process Mining approach itself already embodies characteristics that increase the quality of the resulting models. Regarding the criteria which can be assessed particularly per model it can be stated that they are not applicable to create a clear order of the selected studies. Nevertheless, they help to frame the applied model in order to predict their practical relevance and weigh the conclusions drawn in interpretation of the connected results. A particular conclusion can be drawn from assessing the work of [Fa22]: They work on finding out the flow carried out by programmers in Software Development courses. They apply the SEQUAL framework to assess the quality of the process

model. E.g. one of their main result components (issues of business process quality assurance) does not consider ‘soft’ process-related issues – as it is not in their scope. This finding can be interpreted as a hint to further differentiate the term ‘soft’ process-related issues, e.g. to also define ‘counter-intuitive’ process-related issues.

References

- [Ar19] Ardimento, P., Bernardi, M. L., Cimitile, M., Maggi, F. M.: Evaluating coding behavior in Software Development processes: A process mining approach. In: IEEE/ACM International Conference on Software and System Processes (ICSSP) 2019, pp. 84-93.
- [Ba18] Bala, Saimir, Mendling, J.: Monitoring the Software Development process with process mining. In: International Symposium on Business Modeling and Software Design, Cham 2018. Springer.
- [Ca16] Caldeira, J., Brito e Abreu, F.: Software development process mining: Discovery, conformance checking and enhancement. In: IEEE/10th International Conference on the Quality of Information and Communications Technology (QUATIC) 2016, pp. 254-259.
- [Ca17] Carmignani, G., Cimino, M. G. C. A., Failli, F., Santone, A., Vaglini, G.:A new perspective on Process-oriented Software Engineering based on BPMN Process Mining. In: XIII Convegno dell'Associazione Italiana di Tecnologia Meccanica AITEM 2017, pp. 1-10.
- [Ca19] Caldeira, J., Brito e Abreu, F., Reis, J., Cardoso, J.: Assessing Software Development Teams' Efficiency using Process Mining. 10.1109/ICPM.2019.00020, pp. 65-72. 2019.
- [Ca20] Caldeira, J., Cardoso, J., Ribeiro, R., Werner, C.: Software Development Analytics in Practice: A Systematic Literature. arXiv preprint arXiv:2101.06733. 2020
- [Ca21] Caldeira, J., Brito e Abreu, F., Reis, J., Cardoso: Profiling Software Developers with Process Mining and N-Gram Language Models. arXiv preprint arXiv:2101.06733. 2021.
- [Cr11] Da Cruz, J. I. B., Ruiz, D. D. Conformance Analysis on Software Development: An Experience with Process Mining. International Journal of Business Process Integration and Management, 5(2), 2011, pp. 109-120.
- [Do17] Dong, L., Liu, B., Li, Z., Wu, O., Babar, M. A., Xue, B. A Mapping Study on Mining Software Process. In: IEEE/24th Asia-Pacific Software Engineering Conference (APSEC) 2017, pp. 51-60.
- [Er17] Erdem, S., Demirors, O.: An Exploratory Study on Usage of Process Mining in Agile Software Development. 187-196. 10.1007/978-3-319-67383-7_14, 2017.
- [Fa22] Fauzi, R., Andreswari, R.: Business Process Analysis of Programmer Job Role in Software Development Using Process Mining. Procedia Computer Science, 197, pp. 701-708, 2022.
- [Gu14] Gupta, M., Sureka, A., Padmanabhu, S.: Process Mining Multiple Repositories for Software Defect Resolution from Control and Organizational Perspective. In: Proceedings of the 11th Working Conference on Mining Software Repositories 2014, pp. 122-131.
- [Le11] Lemos, A. M., Sabino, C. C., Lima, R. M., & Oliveira, C. A.: Using Process Mining in Software Development Process Management: A case study. In: 2011 IEEE International Conference on Systems, Man, and Cybernetics, pp. 1181-1186.
- [Li19] Lilienthal, Carola. Langlebige Software-Architekturen: Technische Schulden analysieren, begrenzen und abbauen. dpunkt. verlag, 2019.
- [Ma18] Marques, R., da Silva, M. M., Ferreira, D. R.: Assessing Agile Software Development Processes with Process Mining: A case study. In: IEEE 20th Conference on Business Informatics (CBI) 2018, Vol. 1, pp. 109-118.

- [Ma21] Macak, M., Kruzelova, D., Chren, S., Buhnova, B.: Using process mining for Git log analysis of projects in a Software Development course. *Education and Information Technologies*, 26(5), 5939-5969, 2021.
- [Me18] De Meyer, P., Claes, J.: An Overview of Process Model Quality Literature - The Comprehensive Process Model Quality Framework. *arXiv preprint arXiv:1808.07930*, 2018.
- [No16] Norambuena, BK, Zepeda, VV: Process Mining Applications in Software Engineering. *Risti-Revista Iberica de Sistemas e Tecnologias de Informacao* 21, 2016, 51-66.
- [Po11] Poncin, W., Serebrenik, A., Van Den Brand, M.: Process Mining software repositories. In: *IEEE/ 15th European conference on software maintenance and reengineering 2011*, pp. 5-14.
- [Ru07] Rubin, V., Günther, C., van der Aalst, W. M., Kindler, E., Dongen, B., Schäfer, W. Process Mining Framework for Software Processes. *10.1007/978-3-540-72426-1_15*, pp. 169-181, 2007.
- [Ru14a] Rubin, V. A., Mitsyuk, A. A., Lomazova, I. A., van der Aalst, W. M.: Process Mining Can be Applied to Software Too!. In *Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement 2014*, pp. 1-8.
- [Ru14b] Rubin, V. A., Mitsyuk, A. A., Lomazova, I. A., van der Aalst, W. M.: Agile development with software process mining. In *Proceedings of the 2014 International Conference on Software and System Process*, pp. 70-74.
- [Sa14] Samalikova, J., Kusters, R. J., Trienekens, J. J., & Weijters, A. J. M. M.: Process Mining support for Capability Maturity Model Integration-based software process assessment, in principle and in practice. *Journal of Software: Evolution and Process*, 26(7), pp. 714-728, 2014.
- [Se14] Sebu, M. L., Ciocarlie, H.: Applied process mining in Software Development. In *IEEE 9th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI) 2014*, pp. 55-60.[Sh21] Shynkarenko, V., Zhevaho, O.: Application of Constructive Modeling and Process Mining Approaches to the Study of Source Code Development in Software Engineering Courses. *Journal of Communications Software and Systems* 2021, 17(4), pp342-349.
- [Ur21] Urrea-Contreras, Silvia Jaqueline, et al. "Process Mining Perspectives in Software Engineering: A Systematic Literature Review." *2021 Mexican International Conference on Computer Science (ENC)*. IEEE, 2021.
- [Va05] Van der Aalst, W. M., Medeiros, A. K., Weijters, A. J.: Genetic process mining. In: *International conference on application and theory of petri nets 2005* (pp. 48-69). Springer, Berlin, Heidelberg.
- [We06] Weijters, A. J. M. M., van Der Aalst, W. M., & De Medeiros, A. A.: Process Mining with the Heuristics Miner-Algorithm. *Technische Universiteit Eindhoven, Tech. Rep. WP*, 166 (July 2017) , pp. 1-34, 2006.
- [Za16] Zayed, M. A., Farid, A. B.. The discovery of the implemented software engineering process using process mining techniques. *Int. J. Adv. Comput. Sci. Appl*, 1(7), pp. 279-286, 2016.
- [Ze18] van Zelst, S. J., van Dongen, B. F., van der Aalst, W. M., Verbeek, H. M. W.: Discovering Workflow Nets Using Integer Linear Programming. *Computing*, 100(5), pp. 529-556.

Die Rolle der Künstlichen Intelligenz bei kleinen und mittleren Unternehmen des Dienstleistungssektors

Giuseppe Sortino¹, Sandy Eggert²

Abstract: Dieser Beitrag zeigt die Ergebnisse einer Untersuchung zur aktuellen und zukünftigen Rolle der Künstlichen Intelligenz bei kleinen und mittleren Unternehmen des Dienstleistungssektors im Zuge des digitalen Wandels auf. Entlang einer Literaturanalyse sowie mehreren Experteninterviews werden neben den aktuellen Herausforderungen, Einsatzbereiche und Möglichkeiten der KI, notwendige Voraussetzungen zur Einführung der KI sowie ethische Aspekte herausgestellt.

Keywords: Künstlichen Intelligenz, Einsatzbereiche der KI, Dienstleistungssektor, kleine und mittlere Unternehmen

1 Einleitung

Kleine und mittlere Unternehmen machen in Deutschland über 99% der Gesamtwirtschaft aus [Vo21]. Das Digitalzeitalter und die damit einhergehende Transformation der Unternehmen ist im vollen Gange und sowohl medial als auch gesellschaftlich eine omnipräsente Thematik. Dabei gilt es, sich den neuen Gegebenheiten anzupassen, um auch weiterhin konkurrenzfähig zu bleiben [BA20]. Der Dienstleistungssektor ist mit einem Marktanteil von 76% in Deutschland von besonderer Bedeutung [SG21]. Ein Indikator dafür könnte die Outsourcingstrategie vieler Unternehmen sein. Bereiche wie die IT-Wartung oder Personalrekrutierung werden oftmals in den Mittelstand ausgelagert [SG21]. Die Covid-19 Pandemie der letzten zwei Jahre hat diesen Trend weiterhin verstärkt und die Vorteile von digitalen Tools verdeutlicht [Di21]. Dabei nimmt der Einsatz von Künstlicher Intelligenz im Zuge der Digitalisierung eine relevante und zukunftsweisende Rolle ein. KI birgt das Potenzial, deutliche Verbesserungen entlang der gesamten Wertschöpfungskette vorzunehmen und Bereiche der Unternehmen erfolgreich zu übernehmen [LS19]. Die gelungene Einführung solcher KI-Anwendungen und die generelle Meisterung des digitalen Wandels stellen jedoch häufig eine große Herausforderung für KMU Deutschlands dar. Derartige Entwicklungen geben Anlass zur Sorge und sind Gegenstand vieler Überlegungen und Debatten bezüglich der Digitalisierung [LM19].

¹ HWR Berlin, Badensche Str. 52, 10825 Berlin, giuseppe.sortino@outlook.com

² HWR Berlin, Badensche Str. 52, 10825 Berlin, sandy.eggert@hwr-berlin.de

2 Zielstellung und Vorgehen

Im Fokus steht die Untersuchung der aktuellen und zukünftigen Rolle der Künstlichen Intelligenz bei kleinen und mittleren Unternehmen des Dienstleistungssektors. Damit

einhergeht zunächst die Ermittlung des Status quo sowie die Identifizierung von Potenzialen, die mit dem Einsatz von KI verbunden sind. Weiterhin sollen aktuelle Hemmnisse des KI-Einsatzes bei KMU im Dienstleistungssektor aufgedeckt werden.

Im ersten Schritt wurden mit Hilfe einer Literaturanalyse basierend auf Standardwerken die zentralen theoretischen Aspekte des Themenbereichs KI aufgezeigt. Weiterhin wurden Internetquellen (Studien, Befragungen) herangezogen, um aktuelle Entwicklungen und Potenziale sowie Hemmnisse im KMU-Bereich zu erfassen. Um die ermittelten Aspekte mit einer praxisbezogenen Sichtweise zu stützen und die Einordnung zu spezifizieren, wurden ergänzend mit semistrukturierten Leitfäden Expert*inneninterviews durchgeführt, welche anschließend qualitativ-empirisch angelehnt an Mayring ausgewertet wurden. Zunächst wurden hierbei induktive Kategorien entlang Mayring gebildet, indem die Fragen zusammengefasst wurden [MA15]. Die qualitative Inhaltsanalyse erfolgte anschließend in den Kategorien: Einsatzbereiche und Möglichkeiten der KI, Voraussetzungen und Einführung der KI sowie ethische Aspekte.

3 Digitale Transformation bei KMU

Grundlage der Identifikation von Potenzialen und Hemmnissen ist zunächst die Ermittlung des IST-Zustandes des KI-Einsatzes im KMU-Bereich grundsätzlich und im Speziellen im Dienstleistungssektor.

3.1 Digitale Transformation 4.0 – Kleine und mittlere Unternehmen

Der Begriff Digitalisierung, welcher in unserer Gesellschaft omnipräsent ist, gilt als Megatrend [WS18]. Von Digitalisierung wird allgemein gesprochen, wenn analoge Zustände durch digitale und computerhandbare Modelle ganz oder teilweise ersetzt werden [WS18]. Hierbei werden Geschäftsprozesse, Wertschöpfungsprozesse, menschliche Ressourcen, Produkte, Organisationen und Unternehmen einem Wandel ausgesetzt [WS18]. Die Ursachen für die Digitalisierung sind vielfältig, lassen sich aber in Anlehnung an Bialas und Alan grob in drei Kategorien aufteilen [BA20]:

Gesellschaftlicher Wandel: Ein Treiber der Digitalisierung ist die steigende Erwartungshaltung seitens der Konsument*innen. Die Konsument*innen nehmen stärker Einfluss auf Produkte und Produktentwicklung, so dass zunehmend personifizierte Produkte und Kundenerlebnisse im Vordergrund stehen [BA20].

Ökonomischer Wandel: Durch die Digitalisierung und die damit einhergehende Optimierung aller Geschäftsfelder, waren viele Unternehmen einem erhöhten Wettbewerbsdruck ausgesetzt. Unternehmen, die eine gute Positionierung auf den Märkten besaßen, wurden in diversen Branchen z. B. von Plattformbetreiber*innen unter Druck gesetzt. Die neuartigen Plattformbetreiber*innen konnten sich signifikante Anteile des Kund*innenkontakts sichern [BA20].

Technologischer Wandel: Die Digitalisierung ist nur möglich, wenn die technologischen Mittel verfügbar sind. Das mooresche Gesetzes besagt bspw., dass sich die Rechenleistung eines Computerchips alle 12-24 Monate verdoppelt [Ro16]. Die steigenden technologischen Möglichkeiten haben somit einen direkten Einfluss auf die Realisierbarkeit neuer digitaler Konzepte [BA20].

In Deutschland befinden sich kleine und mittlere Unternehmen derzeit am Anfang der Digitalisierung. Nur ein Fünftel der KMU sind so weit digitalisiert, dass sie Datenmodelle und Algorithmen für ihre Prozesse und Produkte nutzen können [LSGS18]. 80% der KMU befinden sich noch in dem Stadium der Computerisierung, welches sich lediglich durch eine unterstützende Funktion oder digitalen Modellierung ihrer Geschäftsmodelle ausdrückt. Diese Unternehmen nutzen ihre Computer und das Internet, um digital zu „sehen“ und digital zu „agieren“, können aber nicht digital „abbilden“ (Daten) bzw. digital „entscheiden“ (KI) [LSGS18]. Ein signifikanter Anstieg des Digitalisierungsgrads wird mit der Größe eines Unternehmens in Verbindung gebracht. Die kleinen Unternehmen sind häufig weniger digitalisiert als mittlere oder größere Unternehmen. Der digitale Index der IW Consult, DataLovers und beDirect, welches ein Punktesystem von 0-100 nutzt und alle deutschen Unternehmen nach äußereren Indikatoren bewertet, fand heraus, dass die kleinen Unternehmen einen Indexwert von 3,9 erreichten [LSGS18].

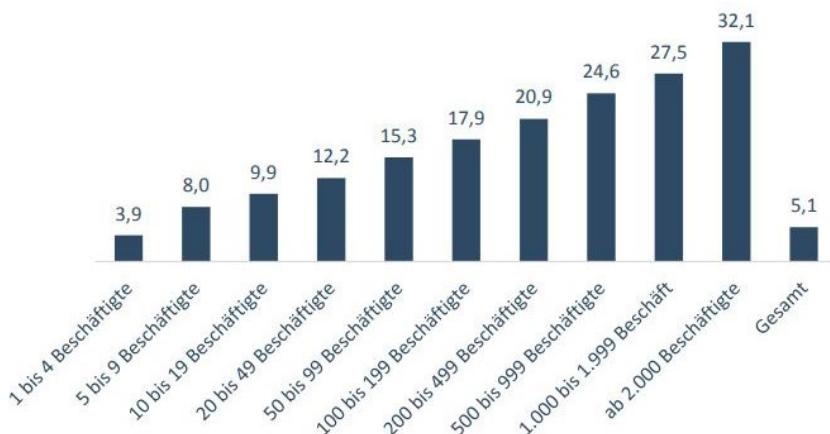


Abb. 1: Digitalindex nach Unternehmensgrößenklasse [LSGS18]

Kleine und mittlere Unternehmen mit 1-499 Mitarbeitenden, erreichten durchschnittlich einen Index von 12,58. Große Unternehmen, die mindestens 500 Mitarbeiter*innen beschäftigen, erreichten einen Durchschnittswert von 28,06. Diese Studie verdeutlicht somit die großen Unterschiede der Unternehmen und die im Verhältnis schwächere Digitalisierung von KMU [LSGS18]. Trotz der jährlich steigenden positiven Entwicklung der Digitalisierung, kann ein Grund für den Rückstand der KMU das traditionelle Verhalten und die miteingehende Vorliebe, altbewährte Methoden weiterhin zu nutzen sein [Li19]. Der Studie „Digital Business Readiness“ zufolge, gaben 75% der befragten KMU an, dass die digitale Transformation ein wichtiges Thema sei. Gleichzeitig haben jedoch über 50% der befragten KMU immer noch keine digitale Strategie [KA18].

3.2 Reifegradmodelle

Bei der Ermittlung von Digitalisierungsbestrebungen von KMU können Reifegradmodelle sinnvoll eingesetzt werden. Im Zuge der Digitalisierung gilt es eine dynamische und aktive Strategie zu entwickeln, um sich an die heutigen Gegebenheiten adaptieren zu können. Der Reifegrad ermittelt den digitalen Stand des Unternehmens, welcher konkлюдierend einen Überblick über die Fähigkeiten und Defizite verschafft [BA20]. Dieser Status Quo schafft ein besseres Bewusstsein für den IST-Zustand hinsichtlich der eigenen Digitalisierung und lässt individuelle Strategien für das Unternehmen zu [FCRJS16]. Reifegradmodelle bestehen hierbei aus klar definierten Merkmalen und Ausprägungen, die in Entwicklungsstufen abstrakt abgebildet werden [HPKGL19]. Hierbei werden verschiedene Methoden für die Beurteilung herangezogen. Neben strukturierter Datenerhebung und Prozessanalysen kann mit Hilfe von Interviews des Managements und Stakeholdern auch ein „Digital Maturity Assessment“ Fragebogen für die Mitarbeiter*innen erstellt werden [BA20].

Digital Maturity Assessment: Ergebnisse der Bewertung der digitalen Reife



Betrachtete Unternehmens-Dimensionen	Blind spot (Start)	Schwach (Rebuild)	Medium (Leverage)	Perfekt (Keep)
 Strategie, Innovation & Wachstum				
 Kundenerlebnis				
 Lieferkette & operativer Betrieb				
 Finanzen, Recht & Steuern				
 Informationstechnologie (IT Grundstruktur)				
 Risiko & Cyber-Sicherheit				
 Personal & Organisation (Führung)				

Abb. 2: Beispiel für ein Digital Maturity Assesment [BA20]

Nicht alle existierenden Modelle sind passend für kleine und mittlere Unternehmen. Eine Literaturanalyse der Universität Potsdam bezüglich geeigneter Reifegradmodelle für KMU identifizierte 94 Modelle, die den digitalen Wandel berücksichtigen. Allerdings wurden nur sechs Modelle als geeignet für KMU interpretiert [HPKGL19]. Bestehende Modelle können häufig nicht angewendet werden, weil sie nur Teile des Unternehmens, wie einzelne Prozesse bewerten oder für große Unternehmen entworfen wurden [HPKGL19]. KMU unterscheiden sich in der Regel stark von großen Unternehmen. Sie bieten spezifische Dienstleistungsangebote an und besitzen häufig keine breite Produktpalette. Unternehmer*innen, welche teilweise gleichzeitig Eigentümer*innen sind, üben einen direkten Einfluss auf die Unternehmensstrategie aus. Dies muss in großen Unternehmen nicht zwingend der Fall sein. Die Führungspersönlichkeit der KMU ist folglich ein ausschlaggebender Faktor [HPKGL19]. Die Erhebung der Reifegrade erfolgt über definierte Dimensionen, welchen Fragestellungen zugeordnet werden. Gängige Dimensionen für Reifegradmodelle der digitalen Transformation für KMU sind Strategie, Kund*innen, Produkte und Services, Prozesse, Organisation und Technologie [HPKGL19]. Die Ergebnisse der Reifegradermittlung werden ebenfalls in diesen Dimensionen wiedergegeben und zum Teil auch grafisch visualisiert. Eine geringe Bewertung lässt eine niedrige Ausprägung der Umsetzung der Digitalisierung entlang der Dimensionen schlussfolgern. Infolgedessen sind Defizite erkennbar und darauf aufbauend kann eine individuelle Strategie entwickelt werden.

3.3 Transformationsstrategien für KMU

Die Strategiebestimmung für eine digitale Transformation ist ein elementarer Bestandteil, um auch in Zukunft konkurrenzfähig zu bleiben [KA18]. Wichtig ist hierbei, dass sich KMU häufig in ihrer Strategie und Ausgangslage signifikant voneinander unterscheiden. Ein geeigneter Ausgangspunkt kann dabei der individuell ermittelte Reifegrad sein, welcher zuvor erwähnt wurde.

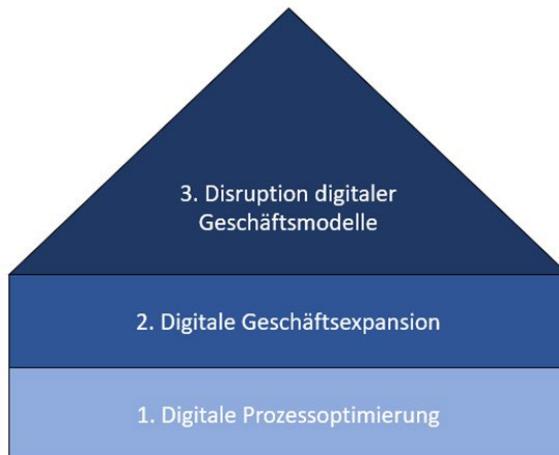


Abb. 3: Dimensionen der Digitalisierung, in Anlehnung an [Ji20]

Nach Fischer lassen sich jedoch drei Kern-Dimensionen zur Umsetzung der Digitalisierungsstrategie identifizieren [Ji20]. Die digitale Prozessoptimierung bezieht sich dabei auf alle Geschäftsprozesse, die den operativen Geschäftsablauf betreffen. Im Ergebnis liegen effizientere und schnellere Prozesse vor, die u.a. mit reduzierten Durchlaufkosten messbar sind [KA18]. Im Rahmen der digitalen Geschäftsexpansion sollen Unternehmen ihre bestehende Marktschließung inkrementell verändern und ausbauen. Im Dienstleistungssektor bezieht sich dies vor allem auf die Gebiete entlang des Absatz- und Kund*innenmarktes. Marketing oder Vertrieb nehmen hierbei eine zentrale Rolle ein [KA18]. Beispielsweise könnte der Einstieg eines Handelsunternehmens in das E-Commerce-Geschäft sein. Hierbei würde das Unternehmen seine Dienstleistung/Ware nicht nur im Geschäft vor Ort anbieten, sondern auch auf Plattformen im Internet. Unternehmen die fortgeschritten sind und beispielsweise damit anfangen, ihre Kund*innendaten auszuwerten, können zudem ihr Geschäft digital expandieren, indem sie ihre Produkte basierend auf die neu gewonnenen Erkenntnisse erweitern und sich den Kund*innen anpassen [Ji20]. Die Disruption digitaler Geschäftsmodelle bezieht sich hingegen auf die Erschließung komplett neuer Geschäftsmodelle. Diese sollen das Potenzial haben, die alten und existierenden Modelle abzulösen. Disruption bedeutet häufig eine gänzliche Umgestaltung/Auflösung des Modells und nicht die alleinige Innovation der bestehenden Geschäftsmodelle. Durch die von Schnellebigkeit geprägte digitale Wirtschaft ist eine agile Adaption häufig vorteilhaft [Ji20].

4 Künstliche Intelligenz

Für Künstliche Intelligenz existieren eine Vielzahl von Definitionen, die je nach Sichtweise variieren. Die Uneinigkeit der Definition entsteht vorwiegend aufgrund der Größe des Themas und des mangelnden Konsenses über die Begrifflichkeit „Intelligenz“ [BS19]. Übereinstimmung herrscht lediglich darin, dass die KI einen Bereich der Informatik abdeckt und ein menschenähnliches Verhalten in Bezug auf die Intelligenz aufweist [Ka17]. Folglich gilt es charakteristische Merkmale und Teilgebiete der Künstlichen Intelligenz herauszustellen, um Einsatzbereiche kleiner und mittlerer Unternehmen des Dienstleistungssektors identifizieren zu können.

4.1 Starke und schwache KI

Allgemein betrachtet lässt sich KI in zwei Oberkategorien aufteilen. Die sogenannte „starke KI“ befasst sich mit dem Versuch menschliche Vorgänge zu imitieren oder abzubilden. Kennzeichnend für die starke KI ist die Fähigkeit, Empathie zu empfinden und ein Bewusstsein zu entwickeln. Obwohl die Forschung diesbezüglich noch keinen Durchbruch verzeichnen konnte, stößt man häufig auf diese Annahme und Voraussetzung [BS19]. Ein weiterer Aspekt, der in Betracht gezogen werden muss, ist die Kompetenz sich kontinuierlich weiterentwickeln zu können, um Limitationen innerhalb der Fähigkeiten zu vermeiden. Die Maschine sollte adaptiv auf ihre Umwelt wirken und ihre Handlungen sollten sich mit den Handlungen eines Menschen vergleichen können bzw. sollten sich nicht von den eines Menschen unterscheiden [Ma21]. Hingegen wurde die schwache KI in der Regel konzipiert, um eine exakte, bestimmte Aufgabe zu bewältigen. Die Entwickler*innen müssen den Lösungsansatz kennen und dieses in die Kodierung miteinfließen lassen. Die schwache KI besitzt, wie die starke KI ein menschenähnliches Verhalten in Bezug auf „Intelligenz“, beschränkt sich jedoch auf die vorbestimmte Aufgabe, die ihr zugeteilt wurde [Ma21].

4.2 Machine learning

Das maschinelle Lernen ist ein Teilgebiet der KI und befasst sich mit dem Erkennen von Zusammenhängen. Dabei agiert das System „lernend“ und wendet im Zuge der Lernprozesse Methoden und Algorithmen an. Das Wissen bzw. die Erfahrung generiert die Maschine oder Software über bestehende Datensätze, die zur Verfügung gestellt wurden. Die Programmierer*innen müssen hierbei nicht mehr alles expliziert codieren, denn die Maschine lernt eigenständig. Beispielsweise werden bei Sprachsteuerungen Algorithmen angewendet, um das Gesprochene in Form von Daten zu analysieren und Befehle zu identifizieren [BS19]. Derzeit existieren im Wesentlichen drei verschiedene Methoden für das Maschinelle lernen [BS19]:

- Supervised Learning - überwachtes Lernen
- Unsupervised Learning - unüberwachtes Lernen
- Reinforcement Learning - verstärkendes Lernen

Das Supervised Learning (überwachtes Lernen) wird häufig für sogenannte Klassifizierungen genutzt, wobei die Lösung bereits bekannt ist. Beispielhaft für dieses Konzept ist die Kategorisierung von Bildern, wobei die Aufgabe dieser Methode darin besteht, Abbildungen in zuvor definierten Kategorien einzuordnen [KLR19]. Beim Unsupervised Learning sucht die Methode eigenständig nach Mustern und Kategorien. Die korrekte Lösung ist hierbei nicht vorgegeben [M21]. Anders als beim Supervised Learning existieren keine vorbestimmten Kategorien und die Methode würde die Daten eigenständig segmentieren [KLR19]. Die Reinforcement Learning Methode hingegen lernt Strategien, um Problemstellungen zu lösen [KMSKB19]. Als Google mit AlphaGo aus dem DeepMind Projekt, den internationalen Champion Lee Sedol im Brettspiel „Go“ schlagen konnte, wurden Reinforcement Learning Methoden angewendet [Th22]. Dieses Verfahren könnte zum Beispiel auch dafür genutzt werden, um den kürzesten Weg aus einem Labyrinth zu finden [KMSKB19].

4.3 Deep learning

Deep Learning ist ein Teilbereich des Machine Learnings, welches künstliche neuronale Netze (KNN) anwendet. KNN wurden anhand der menschlichen biologischen neuronalen Netze im Gehirn entwickelt und sollen den Menschen während des Lernprozesses nachahmen. Die KNN bestehen hauptsächlich aus Knoten und Kanten, die jeweils für die Neuronen und Synapsen des Menschen stehen [BS19]. Die Deep Learning Algorithmen bestehen aus tiefen neuronalen Ebenen, die durch lineare und nichtlineare Einheiten miteinander verbunden sind [MR20]. Hierbei existiert eine Eingangs- und Ausgangsschicht, die von den versteckten Schichten getrennt sind. Aus den versteckten Schichten des KNN leitet sich zudem der Begriff „Deep Learning“ ab. Die Anzahl bzw. die „Tiefe“ der Schichten wirkt sich direkt auf die Fähigkeit Probleme zu lösen aus. Komplexe Modelle können mehrere hundert Schichten besitzen und dementsprechend komplexere Probleme lösen [FMS21]. Auch wenn sich diese Technologie vielseitig einsetzen lässt, kann diese Art von Algorithmus bspw. für Sprach- und Gesichtserkennung optimal genutzt werden. Sie lernen kontinuierlich und verbessern sich selbst, ohne dass ein Mensch eingreifen muss. Hierbei können die Deep-Learning-Algorithmen Millionen von Bildern analysieren und sich selbst trainieren [Ti20]. Die einzelnen Ebenen im KNN gewinnen Erkenntnisse, um sie der nächsten neuronalen Ebene weiterzugeben. Dabei beginnt man in der Regel mit binären Grundoperationen wie das Erkennen von Pixeln und ihre farblichen Ausprägungen. Das finale Ergebnis solcher Lernprozesse ist z. B. die Erkenntnis, um welches Gesicht es sich auf einem Bild handelt [Ti20].

4.4 KI-Potenziale für KMU des Dienstleistungssektors

Das wissenschaftliche Institut (WIK) sieht Potenziale im KI-Einsatz und regt an, dass sich kleine und mittlere Unternehmen zunehmend um die Einführung von KI-Anwendung bemühen sollten. Gerade durch die internationale Konkurrenz könnte der deutsche Mittelstand abgehängt werden und wirtschaftlichen Schaden erleiden. Diese Meinung teilten 70% der befragten Expert*innen im Rahmen der Umfrage der WIK. Das Optimierungspotenzial durch die KI im Mittelstand ist laut vieler Expert*innen evident [LS19].

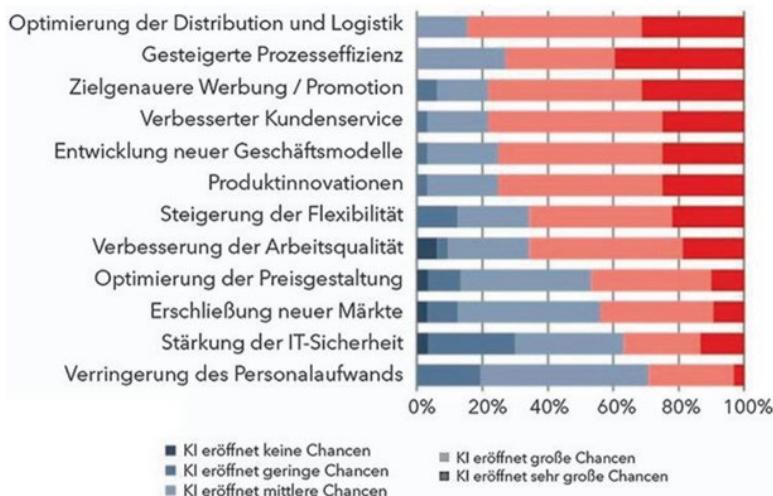


Abb. 4: Einschätzung der Chancen der KI im Mittelstand [LS19]

Entsprechend der Umfrage der WIK könnte der Einsatz von KI entlang der gesamten Wert schöpfungskette Mehrwerte erzielen. Die Bereiche „Prozesseffizienz“, „Optimierung der Distribution und Logistik“ und „Zielgenauere Werbung / Promotion“ werden als besonders relevant für die KI gesehen. Auch wenn häufig über eine Substitutionsgefahr im Zuge der KI gesprochen wird, sieht man in der „Verringerung des Personalaufwands“ wenig Potenzial [LS19]. Der ehemalige Schachweltmeister Garri Kasparov teilte eine ähnliche Meinung, demnach würde die KI den Menschen fördern bzw. ergänzen und nicht vollständig ersetzen [LS19]. Grundsätzlich hat die KI laut dem „McKinsey Global Institute“ mehr Potenzial als die Dampfmaschine während der Industrialisierung. Es wird prognostiziert, dass die globale Wirtschaft im Jahr 2030 um 1,2% jährlich steigen wird. Im Vergleich führte die Dampfmaschine mit ihrer Einführung zu einem Wachstum von 0,3%. Für Deutschland wird eine Wachstumsrate von 1,3% erwartet. Dieser Wert liegt um 0,1% über dem weltweiten Durchschnitt. Das Wachstumspotenzial wird sich auf etwa 50 Mrd. EUR belaufen [LM19]. Laut der Studie „Künstliche Intelligenz im Mittelstand – Relevanz, Anwendungen, Transfer“ waren 84% der befragten Experten der Meinung, dass die Optimierung der Distribution und Logistik für den Mittelstand durch die KI eine besondere Möglichkeit darstellt [LM19].

2030: Künstliche Intelligenz wird für die Logistik unverzichtbar

Wie verbreitet werden die folgenden Szenarien in zehn Jahren sein?



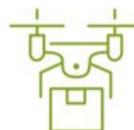
71%

Künstliche Intelligenz übernimmt viele Aufgaben in der Logistik, etwa die Planung von Routen oder das Bestellen von Waren



59%

Waren werden mit autonomen Fahrzeugen zwischen Unternehmen trasportiert



42%

Waren werden mit autonomen Drohnen bis zum Endkunden transportiert

Abb. 5: Einschätzung deutscher Unternehmen über den Stellenwert der KI für die Logistik
[MFWTHC21]

Entlang einer Bitkom-Studie stellt die KI eine unverzichtbare Technologie für die Logistikbranche dar. Dabei prognostizieren ca. 70 % der Teilnehmer*innen sogar, dass der Einsatz von künstlicher Intelligenz in den nächsten zehn Jahren eine Notwendigkeit darstellen wird. Das Optimierungspotenzial der intelligenten Routenplanung und Echtzeitauswertung der Daten und Bestellungen würde einen signifikanten Mehrwert für die Unternehmen schaffen. Der hohe gesellschaftliche und wirtschaftliche Stellenwert und die Tatsache, dass die Prozesse relativ einfach zu verstehen sind und dementsprechend leicht für eine KI umzusetzen sind, trägt zur Relevanz innerhalb der Logistikbranche bei [MFWTHC21]. Ein weiterer Aspekt ist der herrschende Personalmangel in diesem Sektor. Die vierteljährige Umfrage des Münchener Ifo Instituts fand heraus, dass zweidrittel der Unternehmen in diesem Sektor vom Fachkräftemangel betroffen sind und folglich eine Beeinträchtigung von 43% der Tätigkeiten hinnehmen müssen [SLB21]. Die KI könnte dieser Beeinträchtigung entgegenwirken, indem sie autonome Lieferungen von Fahrzeugen und Drohnen ermöglicht [MFWTHC21]. Des Weiteren waren 78% der befragten Expert*innen der WIK davon überzeugt, dass der Kund*innenservice der Unternehmen mithilfe der KI verbessert werden kann. Besonders im Dienstleistungssektor lassen sich viele Erkenntnisse gewinnen. Solche Unternehmen, die sich auf den Kund*innenservice konzentriert haben, können durch eine datengetriebene KI umfassendes Kund*innenwissen erlangen, um konkludierend bessere Kund*innenbeziehungen zu erlangen [BH21]. KI gesteuerte Chatbots stellen z.B. eine Möglichkeit dar, den Kund*innen personalisierte Erfahrung, Komfort und Schnelligkeit zu geben [BS19].

4.5 Hemmnisse für KI in KMU

Gegenwärtig ist der Einsatz von KI in deutschen KMU nicht weit verbreitet. Es besteht jedoch ein positiver Trend, denn der Einsatz von KI hat sich innerhalb eines Jahres (2020-2021) fast verdoppelt. Zudem sind laut einer Umfrage aus dem Jahre 2021 68,2% der Befragten der Meinung, dass KI-Technologien einen Mehrwert für die Volkswirtschaft in Deutschland darstellen, was die überwiegende Akzeptanz der KI widerspiegelt [DG21].

Einsatz von Künstlicher Intelligenz in Unternehmen

in Prozent der befragten Unternehmen aus Industrie und industrienahen Dienstleistungen

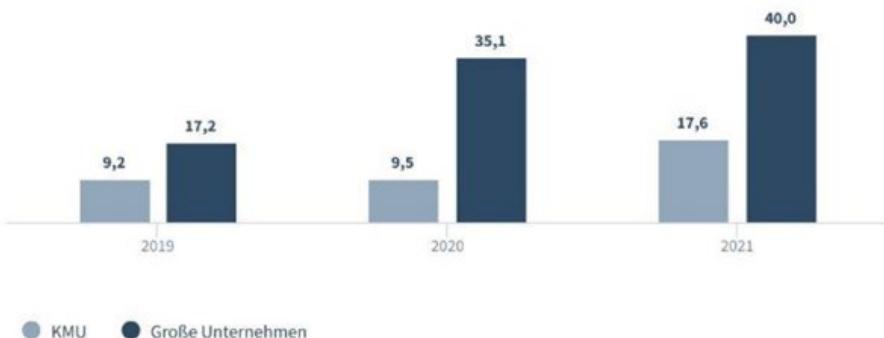


Abb. 6: Einsatz von Künstlicher Intelligenz in Unternehmen [DG21]

Trotz der positiven Entwicklung der letzten drei Jahre in beiden Größenordnungen, verdeutlicht die Abbildung 6 den Rückstand der KMU im Vergleich zu den größeren Unternehmen. Die Kategorisierung der KMU erfolgte hier entsprechend der Definition der EU-Kommission. Gründe für den Rückstand können die häufig zu hoch beurteilte Einschätzung des Aufwandes im Zuge einer Implementierung von KI-Technologien und ein unterschätztes Empfinden des Nutzens einer KI-Anwendung im Unternehmen sein [LG21]. Basierend auf der Umfrage der „Mittelstand-Digital-KI“, ist das fehlende Knowhow bzw. sind die fehlenden Fachkräfte in KMU ein Hauptgrund für den Rückstand des KI-Einsatzes [LM19].



Abb. 7: Inwieweit hemmen folgende Gegebenheiten den Einsatz von KI im Mittelstand? [LG21]

Exemplarisch hierfür sind die häufig vakanten Arbeitsstellen für IT-Fachkräfte, denn KMU können oft nicht mit den Gehältern von Großunternehmen mithalten [LM19]. Im Jahr 2019 blieben in Deutschland grundsätzlich 43% der KI- und Data Science Stellen unbesetzt [ECK21]. Auf Platz 2 befindet sich laut den befragten Expert*innen der mangelnde digitale Reifegrad der KMU. Auch wenn nicht alle KI-Lösungen ein sehr hohes Datenvolumen benötigen, haben Unternehmen, die sich in einer digitalen „Unreife“ befinden, Schwierigkeiten mit der Sammlung von Daten. Machine Learning Methoden brauchen beispielsweise Daten, um trainiert zu werden. Dies ist, wie bereits in vorherigen Kapiteln erwähnt, eine fundamentale Bedingung. Wie häufig angenommen, sind die begrenzten finanziellen Möglichkeiten der KMU laut dieser Umfrage nicht das Hauptproblem. Dennoch ist zu berücksichtigen, dass viele KI-Projekte mit hohen Anschaffungskosten einhergehen. Eine skeptische Haltung gegenüber neuartigen Technologien wie die KI, erschwert zusätzlich das nötige Vertrauen, um langfristige Investitionen wie diese zu ermöglichen [ECK21].

5 Expert*inneninterviews

Um die aus Studien und Umfragen ermittelten Erkenntnisse zu Potenzialen und Hemmnissen zu überprüfen, wurden drei Expert*inneninterviews durchgeführt. Bei der Auswahl der Expert*innen wurde die aktuelle Position im Unternehmen bzw. der Organisation geprüft sowie ob sich die bisherigen Erfahrungen in den im Rahmen dieser Untersuchung betrachteten Bereich einordnen lassen. Insgesamt konnten zwei Experten aus dem Dienstleistungssektor und eine Expertin aus dem Fachgebiet der Digitalisierung im Bereich KMU gewonnen werden (siehe Tabelle 1).

Expert*innen-interviews	Position im Unternehmen	Bisherige Erfahrungen
E 1	Branchen Manager	IT-Dienstleistungsberatung für KMU
E 2	Solution Manager	IT-Dienstleistungsberatung für KMU Mehrjährige Erfahrung mit KI-Projekten
E 3	Dozentin und Teamleiterin	Dozentin für den Fachbereich Wirtschaftsinformatik, Publikationen zur Digitalisierung im KMU-Bereich, Berufserfahrung: Stabstellen im Vorstand

Tab. 1: Angaben Experteninterviews

Die Experten E1 und E2 bieten einerseits Dienstleistungen für KMU an und weiterhin ist es selbst ein KMU, welches sich mit KI-Themen auseinandersetzt.

5.1 Einsatzbereiche und Möglichkeiten der KI

Das Unternehmen der Experten E1 und E2 bietet unterschiedliche KI-Produkte bzw. Vorgehensweisen für ihre Kund*innen an. Die Cognitive Secretary, Cognitive Data Scientist und das Individuelle KI-Projekt. Hierbei handelt es sich um Werkzeuge, die verwendet werden, um Anwendungen für ihre Kund*innen zu modellieren. Das Produkt „Cognitive Secretary“ wendet eine kognitive Verarbeitung von Daten oder beispielsweise Dokumenten an. Ein Praxisfall für diese Produkt, ist bspw. ein Projekt, in welches sich E1 und E2 zum Zeitpunkt des Interviews befanden. Hierbei haben sie Prozesse eines medizin-technisches Sanitätshaus mithilfe der KI automatisiert. Täglich sind ca. 5000 Aufträge per Fax eingetroffen, in welchen Bestellungen enthalten waren. Zu diesen Bestellungen gehören spezielle Strümpfe und dazugehörige Informationen wie Körperteil, Maße, Farbe und Sachen wie Materialstärke. Hierbei handelte es sich um handschriftlich beschriebene Dokumente, die anschließend von 40 Mitarbeitenden erfasst wurden. Die KI „Cognitive Secretary“ sollte nach der Einführung die Dokumente erhalten und sie anschließend aufbereiten, um sogenannte Artefakte zu eliminieren. Nachdem dies geschehen ist, wurden die Aufträge mithilfe bestehender Formular-Typen kategorisiert und aufgenommen. Die KI konnte anschließend 80% der Aufträge automatisch verarbeiten und aufnehmen. Des Weiteren konnte an dieser Stelle Personaleinsatz eingespart werden. Das „Cognitive Data Scientist“ Produkt behandelt Themen entlang der Stammdaten Harmonisierung. E2 befand sich in einem Projekt, in welchem 150.000 Artikel in Form von Daten aufbereitet werden mussten. Hierbei war der Datensatz mit Dubletten versehen und befand sich in einer grundsätzlich ungenügenden Datenqualität, weil die genaue Differenzierung von Produkten nicht möglich war. Die KI konnte hierbei die Daten eigenständig harmonisieren und im Sinne der weiteren Verarbeitung zur Verfügung stellen. Das individuelle KI-Produkt bedeutet hingegen, dass es auch möglich ist, maßgeschneiderte und individuelle Produkte für ihre Kunden umzusetzen. In der Regel tritt das ein, falls „Cognitive Secretary“ oder

„Cognitive Data Scientist“ nicht benötigt wird. Trotz der erzielten Erfolge im Bereich KI im Zuge von Projekten ihrer Kund*innen, wendet das Unternehmen der beiden Experten keine KI-Anwendungen für eigene Zwecke an. E1 erklärte, dass es aktuell keinen konkreten Anwendungsfall gäbe, welcher den Einsatz solcher Technologien rechtfertigen würde. Prinzipiell sei das Unternehmen gegenüber solchen Themen jedoch offen, weil sie sich selbst einer regelmäßigen Evaluierung aussetzen. E2 zur Folge ist der Fachkräftemangel, der im Unternehmen herrscht ein Grund, weshalb manche Projekte nicht vollzogen werden können.

5.2 Voraussetzungen und Einführung der KI

Um eine erfolgreiche Implementierung von KI-Anwendungen in Unternehmen zu ermöglichen, müssten laut aller befragten Expert*innen, die benötigten Voraussetzungen geschaffen werden. E1 geht zuerst davon aus, dass ein konkreter Anwendungsfall identifiziert werden muss, der mithilfe von klassischen Digitalisierungsmethoden nicht zu lösen ist. Dies würde den eigentlichen Prozess initiieren. Laut E3 müssen neben den Use Cases vor allem die Daten im Unternehmen identifiziert werden, mit denen die KI im späteren Verlauf arbeiten soll. Dabei liegt der Fokus auf der Aufbereitung der Daten, um sie in einen auswertbaren Zustand zu versetzen. Für KI-Anwendungen sind korrekte Daten schlussendlich eine fundamentale Bedingung. KMU sollten sich diesbezüglich ebenfalls um die technischen Voraussetzungen kümmern und nicht nur mithilfe von Excel arbeiten. KI-Anwendungen, die neuronale Netze anwenden, benötigen beispielsweise geeignete Tools, mit denen sie arbeiten können. KMU, die keine Daten zur Verfügung haben, sollten zunächst eine ganzheitliche IT-Strategie entwickeln, um sich ausreichend zu digitalisieren. Dabei würde man mit einzelnen Bereichen eines Unternehmens beginnen, um sogenannte „quick wins“ oder „low hanging fruits“ zu erzielen. Ein Indikator für die digitale Bestimmung des Unternehmens sind laut E3 sogenannte Reifegradmodelle. Der Einsatz dieser Modelle ermöglicht das Erkennen der Voraussetzungen für IT-Anwendungen. Vergleichend ist E1 ähnlicher Meinung. Laut E1 sind ca. 80% der ihm bekannten KMU nicht genügend für den Einsatz der KI vorbereitet. Derweil befinden sich viele KMU noch in einem anfänglichen Zustand der Digitalisierung, weshalb noch viele Grundlagen fehlen. Im Zuge seiner Tätigkeit als Consultant mit der Spezialisierung auf ERP-Systeme, sind Problemstellungen, basierend auf einer mangelnden Digitalisierung häufige Themen. Er selbst nutzte aber keine Reifegradmodelle, um seine Kund*innen einzustufen. Häufig erfolgen zunächst Prozessanalysen, nachdem die Kund*innen sich für eines ihrer Produkte entschieden haben. Prinzipiell würden diese Modelle jedoch die Arbeit seiner Meinung entsprechend erleichtern. E2 nutzt ebenfalls keine Reifegradmodelle, weil ein modellhafter Ansatz in der Praxis oft nicht umsetzbar ist. Das primäre Ziel ist für ihn nicht die Ermittlung eines Reifegrades, sondern ein explorativer Ansatz mit dem Kund*innen. Hierbei schätzt E2 die benötigten Voraussetzungen im Gespräch mit den Kund*innen ab. Aus seiner Perspektive ist ein digitaler Reifegrad höchstens ein Indikator für mehr Verkauf, weil er den Kund*innen andere IT-Anwendungen anbieten kann, um eine Grundlage zu bilden. Des Weiteren sind Know-how und IT-Fachkräfte eine weitere Voraussetzung für IT- und KI-Anwendungen. Laut E2

mangelt es vor allem bei KMU an Geld und Ressourcen. Diverse Studien belegen hierbei die Tendenz, dass kleine Unternehmen immer mehr von den Größeren abhängt werden. E3 erkennt prinzipiell auch den herrschenden Fachkräftemangel in allen Bereichen und die reduzierten Ressourcen, bemängelt hingegen besonders die unzureichende Kreativität der KMU in ihren Recruitingprozessen. Häufig sind Unternehmen in der Kundengewinnung sehr kreativ und innovativ, haben aber Schwierigkeiten dasselbe für die Mitarbeitergewinnung umzusetzen. E3 ist der Meinung, dass Unternehmen, die nicht mit den Gehältern von Großunternehmen mithalten können, im Austausch andere Leistungen anbieten sollten, für die Mitarbeitende Abstriche im Gehalt hinnehmen würden.

5.3 Ethische Aspekte

Die drei befragten Expert*innen teilten im Zuge der Interviews ihre Sorgen bezüglich der KI und die Einhaltung von ethischen und moralischen Normen. E1 sieht die Gefahr prinzipiell in allen Bereichen, die KI-Bewertungsmuster auf Menschen anwenden. Ausgangspunkt seiner Sorgen ist der Modus Operandi der Technologie. Hierbei ist eines ihrer Hauptaufgaben die Kategorisierung und Auswertung von Daten, die wiederum zukünftige Prognosen zulassen. Die KI arbeitet hierbei kommentarlos und unbeachtet. Beispielhaft hierfür ist die Auswertung von Gesundheitsdaten der Menschen. Unternehmen könnten mithilfe von KI-Auswertungen die Beiträge personifiziert errechnen und darüber hinaus im wirtschaftlichen Sinne entscheiden, ob sich die Annahme der Patient*innen lohnt. Die KI ist laut E1 geeignet, komplexe Zusammenhänge zu erkennen. Sie eignet sich jedoch nicht für Themen, bei denen Menschen einer Evaluation von Maschinen ausgesetzt sind, und wüsste zum jetzigen Zeitpunkt nicht, wie man die KI effektiv steuern könnte. E2 nimmt eine ähnliche Stellung ein und sieht die Lernfähigkeit der KI entlang von gesellschaftlichen Themen als Gefahr. Der Einsatz von KI für die Mitarbeiter*innengewinnung müsste beispielsweise ethisch überprüft und von Menschen schlussendlich genehmigt werden. Die vollständige Automatisierung solcher Prozesse darf hierbei nicht eintreffen und sollte in Bezug auf ihre Wissensgenerierung limitiert werden. E3 ist der Meinung, dass Jobs im Niedriglohnsektor wahrscheinlich einer Substitution ausgesetzt sein werden. Allerdings würde das eher Arbeitsplätze betreffen, welche eine hohe Fluktuation des Personals aufweisen. Hierbei würde die KI vielmehr die Fachkräfte entlasten, weil sie sich mit repetitiven Aufgaben, wie beispielsweise die Ticketerstellung nicht mehr befassen müssten. Aktuell werden vor allem Nebentätigkeiten automatisiert, in denen ein Mangel von Mitarbeiter*innen herrscht. Problematisch sieht er eher den möglichen Kontrollverlust von KI und die zunehmende Abhängigkeit solcher Technologien. Derartige Entwicklungen sollten den Menschen sensibilisieren und gegebenenfalls motivieren, präventive Maßnahmen zu ergreifen.

6 Auswertung

Die kleinen und mittleren Unternehmen sind das Rückgrat der deutschen Wirtschaft und machen über 99% der Unternehmen aus [Vo21]. Folglich nehmen die KMU eine zentrale Rolle im Zuge der Digitalisierung ein. Nicht verwunderlich scheinen diesbezüglich die gesellschaftlichen Debatten über die digitale Transformation und wie diese am besten gestaltet wird. Aus den Ergebnissen lässt sich schließen, dass sich KMU in Deutschland derzeit am Anfang der Digitalisierung befinden. Hierbei sind laut einer Studie 80% der KMU im Stadium der Computerisierung und können noch nicht digital in Form von Daten „abbilden“ bzw. digital mithilfe von KI „entscheiden“. Zusätzlich herrschen große Unterschiede besonders zwischen den kleinen und größeren Unternehmen in Bezug auf die Digitalisierung. Positiv zu beurteilen ist jedoch der überwiegende Konsens der KMU über die strategische Bedeutung der Digitalisierung, um auch in Zukunft konkurrenzfähig zu bleiben. Der Eintritt in die digitale Transformation kann hierbei mithilfe geeigneter Reifegradmodellen ermöglicht werden, die eine facettenreiche und für die KMU gerechte Bewertung berücksichtigen. Wichtig ist hierbei zu erkennen, dass sich KMU häufig in ihrer Strategie und Ausgangslage signifikant voneinander unterscheiden und der individuell ermittelte Reifegrad ein Ausgangspunkt für weitere Vorhaben sein könnte. Die digitale Prozessoptimierung, die digitale Geschäftsexpansion und die Disruption digitaler Geschäftsmodelle sind jedoch drei Kern-Dimensionen einer Digitalisierungsstrategie und sollten hierbei verfolgt werden.

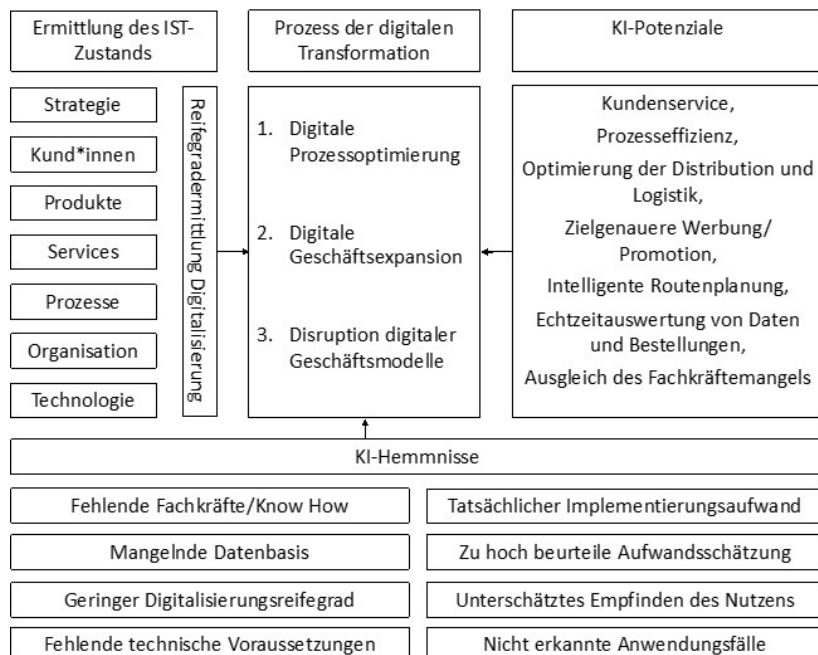


Abb. 8: Potenziale und Herausforderungen des KI-Einsatzes für KMU im Dienstleistungsbereich

Zusammenfassend zeigt Abbildung 8, welche Potenziale mit dem KI-Einsatz für KMU verbunden sind. Die Notwendigkeit des Vorantreibens von KI-Bestrebungen zeigt sich auch in der internationalen Konkurrenzfähigkeit. Einsatzbereiche der KI existieren entlang der gesamten Wertschöpfungskette. Vor allem im Dienstleistungssektor lassen sich Potenziale, in den Bereichen Prozesseffizienz, Optimierung der Distribution und Logistik sowie Ermöglichung von zielgenauerer Werbung, identifizieren. Trotz der positiven Entwicklungen der letzten Jahre in Bezug auf die Nutzung von KI, ist KI im KMU-Bereich nicht weit verbreitet. Als Hemmnis werden vor allem die fehlenden Fachkräfte und der mangelnde digitale Reifegrad der KMU gesehen.

Literaturverzeichnis

- [BA20] Biallas, S. & Alan, Y. (2020). Digitalisierung im Mittelstand: Trends, Impulse und Herausforderungen der digitalen Transformation. Springer Gabler. <https://doi.org/10.1007/978-3-658-29291-1>, S. 45-50.
- [BH21] Bruhn, M. & Hadwich, K. (2021). Künstliche Intelligenz im Dienstleistungsmanagement: Band 1: Geschäftsmodelle – Serviceinnovationen – Implementierung (Bd. 1). Springer Gabler. <https://doi.org/10.1007/978-3-658-34324-8>
- [BS19] Buxmann, P. & Schmidt, H. (2019). Künstliche Intelligenz: Mit Algorithmen zum wirtschaftlichen Erfolg (2. Aufl.). Springer Gabler. <https://doi.org/10.1007/978-3-662-61794-6>, S. 6.
- [DG21] Demary, V. & Goecke, H. (2021). Wie KMU Künstliche Intelligenz nutzen. <https://www.iwkoeln.de/studien/vera-demary-henry-goecke-wie-kmu-kuenstliche-intelligenz-nutzen.html>, S. 1-2.
- [Di21] Bundesministerium für Wirtschaft und Energie (BMWi). (2021). Digitalisierung in Deutschland – Lehren aus der Corona-Krise: Gutachten des Wissenschaftlichen Beirats beim Bundesministerium für Wirtschaft und Energie (BMWi). <https://www.bmwk.de/Redaktion/DE/Publikationen/Ministerium/Veroeffentlichung-Wissenschaftlicher-Beirat/gutachten-digitalisierung-in-deutschland.html>, S. 3.
- [ECK21] Ecker, W., Coulon, C.-H. & Kohler, M. (2021). KI in die Anwendung bringen: Eine Gemeinschaftsaufgabe für Hochschulen, Forschungseinrichtungen, Unternehmen und
- [FCRJS16] Frey, U., Cachelin, J. L., Richards, B., Jäger, J. & Sieber, P. (2016). ku Führung von Kleinunternehmen: Digitalisierung (16. Aufl.). PostFinance AG. https://www.alexandria.unisg.ch/250062/1/LOW_160152_PostFinance_ku16_mai16_DE_V26.pdf, S. 4.
- [FMS21] Feiner, M. & Schöllhorn, M. (2021). KI4Industry: KI für den Mittelstand. <https://arxiv.org/ftp/arxiv/papers/2106/2106.09455.pdf#page=30>, S. 11.
- [HPKGL19] Hölzle, K., Petzolt, S., Kullik, O., Gerhardt, F. & Lehrstuhl für Innovationsmanagement und Entrepreneurship, Universität Potsdam. (2019). Reifegradmessung zur digitalen Transformation von KMU. Gleich wie Autor/in. <https://doi.org/10.13140/RG.2.2.28402.66242>, S. 11-23.
- [Ji20] J. Fischer, F. (2020). Digitalisierung im Mittelstand: Trends, Impulse und Herausforderungen der digitalen Transformation. Springer Gabler. <https://doi.org/10.1007/978-3-658-29291-1>, S. 24-25
- [Ka17] Kaplan, J. (2017). Künstliche Intelligenz: Eine Einführung (1. Aufl.). mitp Verlags GmbH. <https://books.google.de/books?hl=en&lr=&id=mUIzDwAAQ-BAJ&oi=fnd&pg=PT2&dq=k%C3%BCnstliche+intelligenz+definition&ots=CdtY-B328aO&sig=jgoAIiYrMvdLkqFEd98p4oBi5o#v=onepage&q=k%C3%BCnstliche%20intelligenz%20definition&f=false>, S. 5-135.
- [KA18] Kugler, S. & Anrich, F. (2018). Digitale Transformation im Mittelstand mit System: Wie KMU durch eine innovative Kultur den digitalen Wandel schaffen. Springer Gabler. <https://doi.org/10.1007/978-3-658-22914-6>, S. VI-5.
- [KLR19] Kersting, K., Lampert, C. & Rothkopf, C. (2019). Wie Maschinen lernen: Künstliche Intelligenz verständlich erklärt. Springer Fachmedien. <https://doi.org/10.1007/978-3-658-26763-6>, S. 23-25.

- [KMSKB19] Kleesiek, J., M. Murray, J., Strack, C., Kaassis, G. & Braren, R. (2019). Wie funktioniert maschinelles Lernen? Springer Medizin Verlag. <https://doi.org/10.1007/s00117-019-00616-x>, S. 25.
- [LG21] Lundborg, M. & Gull, I. (2021). Künstliche Intelligenz im Mittelstand: So wird KI für kleine und mittlere Unternehmen zum Game Changer. https://www.mittelstand-digital.de/MD/Redaktion/DE/Publikationen/ki-Studie-2021.pdf?__blob=publicationFile&v=5, S. 13.
- [Li19] Lindner, D. (2019). KMU im digitalen Wandel: Ergebnisse empirischer Studien zu Arbeit, Führung und Organisation. Springer Gabler. <https://doi.org/10.1007/978-3-658-24399-9>, S. 1-5.
- [LM19] Lundborg, M. & Märkel, C. (2019). Künstliche Intelligenz im Mittelstand: Relevanz, Anwendungen, Transfer. https://www.mittelstand-digital.de/MD/Redaktion/DE/Publikationen/kuenstliche-intelligenz-im-mittelstand.pdf?__blob=publicationFile&v=5, S. 7-9.
- [LS19] Lundborg, M. & Schrade, L. (2019). Digitalisierung im Mittelstand: Trends, Impulse und Herausforderungen der digitalen Transformation. Springer Gabler. <https://doi.org/10.1007/978-3-658-29291-1>
- [LSGS18] Lichtblau, K., Schleiermacher, T., Goecke, H. & Schützdeller, P. (2018). Digitalisierung der KMU in Deutschland: Konzeption und empirische Befunde. https://www.iwconsult.de/fileadmin/user_upload/projekte/2018/Digital_Atlas/Digitalisierung_von_KMU.pdf, S. 16-17.
- [M21] Matzka, S. (2021). Künstliche Intelligenz in den Ingenieurwissenschaften: Maschinelles Lernen verstehen und bewerten. Springer Vieweg. <https://doi.org/10.1007/978-3-658-34641-6>, S. 13.
- [MA15] Mayring, P. (2015). Qualitative Inhaltsanalyse: Grundlagen und Techniken (Beltz Pädagogik) (Neuausgabe Aufl.). Beltz., S. 67.
- [Ma21] Maximilian Röser, A. (2021). Charakterisierung von schwacher und starker Künstlicher Intelligenz. MA Akademie Verlags- und Druckgesellschaft mbH. <https://www.econstor.eu/handle/10419/234520>, S. 33.
- [MFWTHC21] Murrenhoff, A., Friedrich, M., Witthaut, M., ten Hompe, M., Henke, M. & Clausen, U. (2021). KÜNSTLICHE INTELLIGENZ IN DER LOGISTIK: Future Challenges in Logistics and Supply Chain Management. Fraunhofer-Institut für Materialfluss und Logistik IML.
- [MR20] Microsoft & Ronsdorf, J. (2020, 9. April). Microsoft erklärt: Was ist Deep Learning? Definition & Funktionen von DL. News Center Microsoft Deutschland. Abgerufen am 16. Januar 2022, von <https://news.microsoft.com/de-de/microsoft-erklaert-was-ist-deep-learning-definition-funktionen-von-dl/>
- [Ro16] Roth, A. (2016). Einführung und Umsetzung von Industrie 4.0: Grundlagen, Vorgehensmodell und Use Cases aus der Praxis. Springer Gabler. <https://doi.org/10.1007/978-3-662-48505-7>, S. 24.
- [SG21] KfW Bankengruppe, Schwartz, M. & Gerstenberger, J. (2021, 1. Oktober). KfW-Mittelstandspanel 2021: Mittelstand beweist Anpassungsfähigkeit in der Corona-Krise – Fundament der Kleinen allerdings mit sichtbaren Rissen. kfw.de. <https://www.kfw.de/PDF/Download-Center/Konzernthemen/Research/PDF-Dokumente-KfW-Mittelstandspanel/KfW-Mittelstandspanel-2021.pdf>, S.23

- [SLB21] Semmann, C., Lauenroth, L. & Bosse, T. (2021, 12. August). <https://www.dvz.de/sonderseiten/daten-center/detail/news/logistik-fachkraeftemangel-verschaerft-sich-wieder.html>. DVZ. Abgerufen am 4. Februar 2022, von <https://www.dvz.de/sonderseiten/daten-center/detail/news/logistik-fachkraeftemangel-verschaerft-sich-wieder.html>
- [Th22] DeepMind. (o. D.). The challenge: Go is known as the most challenging classical game for artificial intelligence because of its complexity. Abgerufen am 16. Januar 2022, von <https://deepmind.com/research/case-studies/alphago-the-story-so-far>
- [Ti20] Tiedemann, M. (2020, 8. Mai). Deep Learning in der Praxis: 5 Anwendungsfälle für Deep-Learning-Algorithmen. alexanderthamm. Abgerufen am 19. Januar 2022, von <https://www.alexanderthamm.com/de/blog/deep-learning-in-der-praxis/>
- [Vo21] Institut für Mittelstandsforschung Bonn. (o. D.-b) Mittelstand im Überblick: Volkswirtschaftliche Bedeutung der KMU. Institut für Mittelstandsforschung Bonn (ifm Bonn). Abgerufen am 21–12–16, von <https://www.ifm-bonn.org/statistiken/mittelstand-im-ueberblick/volkswirtschaftliche-bedeutung-der-kmu/deutschland>
- [WS18] Wolf, T., Strohschen, J.-H. & Informatik-Spektrum volume 41. (2018). Digitalisierung: Definition und Reife: Quantitative Bewertung der digitalen Reife. Springer-Verlag Berlin Heidelberg 2018. <https://doi.org/10.1007/s00287-017-1084-8>, S. 58 - 59

AI-ware als sozio-technische Systeme: Entwurf eines Gestaltungsrahmens für den Einsatz von KI in Unternehmen

Claudia Lemke¹ und Dagmar Monett²

Abstract: Das aktuelle KI-Momentum spricht für den Einsatz intelligenter Algorithmen in einer Vielfalt von Anwendungsbereichen, die unser Leben schon verändern oder verändern werden. Allerdings stellt in letzter Zeit der KI-Hype andere Teilgebiete in den Schatten und verschleiert gleichzeitig die Implikationen und Risiken, die KI-basierte Systeme mit sich bringen können. Die Chancen und Herausforderungen solcher Systeme brauchen sowohl moderne, ethisch konforme Handlungsempfehlungen als auch eine Forschungsagenda, die eine richtige, an den Nutzer*innen orientierte Gestaltung ermöglichen. Wir untersuchen in diesem Beitrag die Grundlagen und Merkmale KI-basierter Systeme und schlagen einen auf der sozio-technischen Systemtheorie basierenden Gestaltungsrahmen vor, um KI in Organisationen erfolgreich umzusetzen.

Keywords: AI-ware, KI, Künstliche Intelligenz, socio-technical systems theory, sozio-technische Systeme.

1 Problemstellung und Zielsetzung

Anwendungen der Künstlichen Intelligenz (KI) beeinflussen die Art und Weise, wie Unternehmen zukünftig Entscheidungen treffen werden [FMS19]. Zudem innovieren sie durch eine stärkere Automatisierung betriebliche Abläufe, Aufgaben und Strukturen [LM20]. Der Einsatz von KI-basierten Systemen in den verschiedenen betrieblichen Unternehmensbereichen ist seit Jahren zunehmend, wie verschiedene Studien und Umfragen zeigen [Ch21], [RG22], [Ox19]. Aus Sicht des privaten Investments in KI-Ideen, -Lösungen und -Startups zeigt der jüngste AI Index Report 2022 der Stanford University [Zh22] eine Verdopplung im Vergleich zum Vorjahr auf ca. 93,5 Milliarden USD. Auch wenn die verschiedenen Studien und Umfragen je nach Design unterschiedliche Schwerpunkte und Erkenntnisse präsentieren, unbestritten scheint zu sein, dass KI insgesamt und deren Abbildung als konkrete Applikationen mittlerweile als technologische Lösung für die Gestaltung von Unter-

¹ Hochschule für Wirtschaft und Recht Berlin, Fachbereich Duales Studium, Fachrichtung Wirtschaftsinformatik, Alt-Friedrichsfelde 60, Berlin, 10315, claudia.lemke@hwr-berlin.de,
<https://orcid.org/0000-0003-0151-1489>

² Hochschule für Wirtschaft und Recht Berlin, Fachbereich Duales Studium, Fachrichtung Informatik, Alt-Friedrichsfelde 60, Berlin, 10315, dagmar.monett-diaz@hwr-berlin.de,
<https://orcid.org/0000-0001-5750-972X>

nehmen nicht mehr wegzudenken sind. Damit forciert der betriebliche Einsatz von KI-basierten Systemen zwangsläufig ähnliche Herausforderungen wie der Einsatz klassischer betrieblicher Informationssysteme in den 80er und 90er Jahren des letzten Jahrhunderts [BH00], [PM85], [Ve94].

Darüber hinaus zeigen sich aber auch neue, bislang nicht adressierte Herausforderungen. Diese fokussieren zum einen die besonderen digital-ethischen Problemstellungen beim Design der Algorithmen und des Datenmanagements (vgl. [Co20], [IE19]) sowie bei deren funktionsorientierten Nutzung, den Verwendungsmöglichkeiten und Formen der Weiterentwicklung, insbesondere auch im Bereich des maschinellen Lernens (ML). Zum anderen gestaltet sich aufgrund der besonderen Natur der Technologie Künstliche *Intelligenz* der Entwicklungsprozess solcher Systeme selbst anders, so dass neue bzw. veränderte Verfahren der Softwareentwicklung notwendig werden, die z.B. auch die Belange des Trainings von Algorithmen, wie beim maschinellen Lernen, oder den Wechsel von Laborumgebungen zu realen Unternehmenssituationen berücksichtigen müssen [FMS19]. Auf einer Metaebene indes offenbart sich die genuine Problemstellung von KI – die einer bislang nicht konsensfähigen Definition von KI oder noch konkreter der von Intelligenz [MLT20], [Wa19]. Eine nach wie vor nicht allgemeingültig akzeptierte Definition von KI bietet den Raum für vielerlei Interpretationen über computerbasierte Systeme der Daten- und Informationsverarbeitung mit der Fähigkeit einer künstlichen bzw. maschinellen Intelligenz. Bei einer tiefergehenden Prüfung halten solche Systeme teilweise nicht die Versprechungen über die Abbildung intelligenter Fähigkeiten. Zudem fördert der aktuelle Medienhyphen um KI Missverständnisse über die wirklichen Einsatzpotenziale und Grenzen von KI [MD19] und führt damit zwangsläufig zu falschen und überzogenen Erwartungen sowie übersteigerten Annahmen zu den Chancen und Risiken des Einsatzes, gerade auch im betrieblichen Kontext. Somit erscheint die Gestaltungsaufgabe für einen verantwortungsvollen und vertrauenswürdigen Einsatz von KI-basierten Systemen in Unternehmen mehrdimensional und komplex.

Eine erste allgemeingültige Betrachtung von KI-basierten Systemen kann diese ungeachtet der verschiedenen möglichen KI-Modelle [RN20] als Systeme zur automatisierten Entscheidungsfindung definieren – in diesem Beitrag mit Bezug auf die Unterstützung betrieblicher Aufgabenbereiche. Ohne die breit geführte akademische Diskussion über die Ansätze einer *moral agent vs. moral patient theory* (siehe [Mi18], [Lo19], [Co20]) aufgreifen zu wollen, werden in diesem Beitrag KI-basierte Systeme als (moralische) Ratgeber (vgl. [Mi18], [Lo19], [Co20]) betrachtet, die einer letzten, menschlichen Instanz zur Entscheidungsdurchsetzung bedürfen. Nicht zuletzt wegen der nach wie vor nicht konsensfähigen Definition von (künstlicher) Intelligenz und des praktischen KI-Effekts [Mc04], sollten KI-basierte Systeme im betrieblichen Einsatz als Entscheidungsratgeber fungieren, die eine menschliche Entscheidungsfindung unterstützen und erweitern, z.B. durch eine schnellere und umfassendere Analyse und Auswertung von bestimmten Datenmengen [LMM21].

Auf der Grundlage dieser Prämissen können KI-basierte Systeme mit Bezug auf ihren betrieblichen Einsatz als eine zur maschinellen Intelligenz fähigen Form von Informationssystemen betrachtet werden, die als sozio-technische Systeme menschliche und maschi-

nelle Komponenten umfassen. Die gestaltungsorientierte Wirtschaftsinformatik liefert als interdisziplinäre Wissenschaftsdisziplin den Erkenntnisstand zur Gestaltung betrieblicher Informationssysteme [OWB10]. Eine Übertragbarkeit der damit verbundenen Konzepte und Ansätze auf *KI als intelligente sozio-technische Systeme* sollte daher stärker in den Fokus rücken. Dieser Beitrag beginnt exakt an dieser Stelle mit der Betrachtung des Einsatzes von KI-basierten Systemen unter Nutzung der *sozio-technischen Systemtheorie* (eng.: socio-technical systems theory) [Tr81]. Dieser Beitrag liefert einen begrifflichen Überblick zu KI einschließlich zu den Entwicklungen und Einflussfaktoren dieser Technologie. Danach erfolgt eine erste theoretisch basierte Adaption der sozio-technischen Systemtheorie auf die Belange des Einsatzes KI-basierter Systeme in Organisationen mit dem Ziel, eine Handlungs- bzw. Forschungsagenda zu formulieren. Diese kann als Ausgangspunkt dienen für nachfolgende notwendige, zwangsläufig auch empirisch basierte, Untersuchungen, die nicht Gegenstand dieses Beitrags sind. Das Ergebnis dieses Beitrags ist es, den Diskurs über die *richtige* Gestaltung von KI-basierten Systemen in Unternehmen auf der Grundlage einer systemischen, kybernetischen und interdisziplinären Betrachtung zu stärken.

Das übergeordnete Ziel dieses Beitrags besteht darin, eine Theoriegrundlage zu schaffen, um die überzogenen Erwartungen an KI und die zum Teil fehlgeleiteten Projekte [FMS19] von den ernsthaften und erfolgreichen Bestrebungen des Einsatzes von KI-basierten Systemen zu trennen bzw. den bereits erfolgreich in Organisationen verankerten KI-basierten Lösungen. Somit kann mit diesen Ausführungen auch ein Beitrag geleistet werden zum Erkenntnisstand der Wirtschaftsinformatik im Bereich der Gestaltung sozio-technischer Systeme, hier in Erweiterung auf sozio-technische Systeme mit der Fähigkeit zur Abbildung intelligenter Eigenschaften.

2 Begriff und Entwicklung von KI

2.1 Die Definition von KI und das KI-Momentum

In der Vergangenheit hat es ein typisches und tiefgreifendes Missverständnis darüber gegeben, was maschinelle oder künstliche Intelligenz bedeutet oder was diese leisten kann. Dies hat die Fortschritte auf dem Gebiet der KI teilweise gebremst und ihre Glaubwürdigkeit seit ihrer Gründung mehrmals beschädigt (vgl. [Mc76], [Mi19]). Versprechungen über intelligente Fähigkeiten in Maschinen sind seit Jahrzehnten im KI-Gebiet allgegenwärtig, insbesondere während der „AI Springs“ (auch „AI Summers“ genannt) bzw. weiteren Entwicklungszeiträumen, in denen (vermeintlich) bedeutende Fortschritte erzielt wurden und werden. Der Enthusiasmus wächst unaufhaltsam, ebenso wie die Erwartungen bzw. der Anspruch an die Leistungsfähigkeit von Maschinen. Neue Entwicklungen und das Gebiet selbst werden auf einen Status gehoben, der über die tatsächlichen Leistungen hinausgeht [Sh56]. Es ist nur eine Frage der Zeit, bis dieser Hype zusammenbricht und folglich Investitionen gekürzt werden und das Interesse an dem Gebiet nachlässt (vgl. [Ma13], [VF22]).

KI ist aber nach wie vor weit davon entfernt, die allgemein gewünschte magische Kugel zu sein, die all unsere (Menschheits-)Probleme lösen kann. Die Gründe für diese Höhen und Tiefen sind vielfältig. Einer der Wichtigsten scheint zu sein, dass es keinen Konsens über eine Definition von KI gibt, auch nicht von Intelligenz als Begriff (vgl. [MLT20]). Es ist verständlich, dass es immer wieder Teilgebiete geben wird, die je nach wissenschaftlichem Fortschritt mehr oder weniger erfolgreich werden; andere entstehen neu und wieder andere verschmelzen. Aber im KI-Gebiet herrscht eine Kontroverse seit seiner Entstehung, charakteristisch für die Unfähigkeit, einen Konsens über die grundlegendsten Konzepte und Techniken der KI zu erzielen.

In den letzten Jahren hat der *Trugschluss des maschinellen Lernens* (in Englisch, *the machinelearnization fallacy*) sowohl die öffentliche Wahrnehmung von KI als auch die der KI-Insider*innen dominiert. Wir behaupten, dass es sich bei diesem Trugschluss um den Drang handelt, alle Probleme mit maschinellem Lernen lösen zu wollen und dafür die Realität auf eingeschränkte Modelle reduziert wird. Datengesteuerte, statistisch basierte Algorithmen (wie künstliche -tiefe- neuronale Netze und verwandte Techniken) sind jedoch nicht für alle Problemstellungen unserer Welt die am besten geeignetsten. Die einzelnen Auswirkungen, z.B. bei der Fehlklassifizierung von Katzenbildern, werden verallgemeinert als Auswirkungen auf die Gesellschaft und ihren Strukturen insgesamt betrachtet. Darauf hinaus kann es für die KI-Gemeinschaft und ihre Fortschritte schädlich sein, von KI zu sprechen, wenn damit nur eines ihrer Teilgebiete gemeint ist [Cr21], [La21].

Die wesentlichen Fortschritte, die zurzeit in KI gemacht werden, betreffen jedoch hauptsächlich die Fähigkeiten zur Muster- und Spracherkennung auf niedriger Ebene [Jo19] und damit, die „schwache“ KI. Diese sind bei weitem noch nicht genug fortgeschritten, um als „intelligent“ zu gelten (vgl. [MD19], [Mi19], [Sm19]). Im Bereich der Big-Data-Anwendungen für Unternehmen werden viele Projekte als KI-Projekte deklariert, die zudem dem klassischen Aufgabenbereich des IT-Projektmanagements untergeordnet werden. Das suggeriert nicht nur den Mitarbeiter*innen, sondern auch den Aktionär*innen und der Öffentlichkeit, dass KI auf moderne und innovative Weise nach den klassischen Paradigmen des IT-Managements betrieben werden kann. Doch hinter dem Geheimnis verbirgt sich die Anwendung klassischer statistischer Methoden, die mit KI-Algorithmen wie dem maschinellen Lernen realisiert werden können, aber nicht unbedingt. So entstehen Verzerrungen in der Vorstellung, dass KI-getriebene Projekte (und damit die Annahme von Software-Engineering) und Lösungen nachhaltig und sinnvoll in Unternehmen eingesetzt werden können. Diese Tendenzen zwingen dringend zu einer Erweiterung des Untersuchungsproblems. Die gestaltungsorientierte Wirtschaftsinformatik liefert mit der Untersuchung der Konstruktion betrieblicher Informationssysteme, wie es [Si10, S. 35] skizziert, einen bereits vorhandenen, interdisziplinär geprägten Erkenntnisbeitrag, der sich lohnt, auf die Untersuchung der Konstruktion KI-basierter Systeme zu übertragen.

2.2 Der Begriff AI-ware

Im Grunde genommen kann jede definitorische Betrachtung von KI auf den Terminus *Software* zurückgeführt werden. Im Sinne der Zusammenfassung von Computerprogrammen, den Algorithmen, und den dafür notwendigen Daten als ein immaterielles Abbild von Problemstellungen der realen Welt [Cr21] sind KI-basierte Systeme stets auch Softwarelösungen. Im Fall von KI indes herrscht bei einigen Vertreter*innen, nicht nur der eigenen Fachdisziplin, die Auffassung, KI sei magisch und bedürfe daher einer neuen begrifflichen Kategorie im Kanon der Computertechnik bzw. Informations- und Kommunikationstechnik (IT) (vgl. [La21], [Jo19]). So wurde z.B. zwischen KI und Software-Engineering bzw. Softwareentwicklung im Besonderen und der Informatik im Allgemeinen künstlich eine unscharfe Grenze gezogen, die mit der Zeit immer stärker wurde. Es ist nicht verwunderlich, dass sich das Software-Engineering zu der etablierten Disziplin entwickelt hat, die es heute ist, während der Bereich der KI in der algorithmischen Entwicklung stecken geblieben ist, die hauptsächlich in akademischen Umgebungen improvisiert wurde und nur wenige praktische Anwendungen in großem Maßstab hatte. Die traditionelle Softwarestruktur, die nicht nur alle Phasen der Softwareentwicklung, sondern auch das Management dieser und die Qualitätssicherung umfasst [So16], ist zwar bei einigen KI-gestützten Systemen anders, aber nicht bei allen. Sie erfordern eine stärkere algorithmische Abstimmung, Ungewissheit, Mehrdeutigkeit, einen anderen Umgang mit Daten (wenn es sich um eine kritische Ressource handelt) und eine höhere Fehlertoleranz und Instabilität der Systeme [Oz20].

AI-ware ist ein Sammelbegriff mit einer mehrfachen Bedeutung [ML21]!

Zum ersten zielt der Begriff *AI-ware* auf die Notwendigkeit ab, sich bewusst zu machen, was KI ist und was nicht und welche Teilbereiche betroffen sind sowie über eine kritische Einschätzung zum sinnvollen Einsatz von KI– *a) Definition von KI*.

Zum zweiten – und hier zeigt sich dann auch der direkte Bezug zur Betrachtung als sozio-technisches System – muss *AI-ware* in einen organisatorischen Rahmen eingebettet werden. Als alleinstehende Lösung oder Teil größerer betrieblicher Informationssysteme unterstützen KI-basierte Systeme in erster Linie die Aufgaben der Informationsverarbeitung in Unternehmen und interagieren damit zwangsläufig mit den Mitarbeiter*innen, da diese aus ethischen Aspekten stets als moralische Ratgeber zu sehen sind – *b) AI-ware als sozio-technisches System*. Die verschiedenen KI-Modelle mit ihren Algorithmen müssen zunehmend den Labor- bzw. Experimentierstatus verlassen und die Robustheit, Sicherheit und Verlässlichkeit aufweisen, um tatsächlich im direkten Einsatz für betriebliche Aufgabenbereiche nutzbar zu sein. Das erfordert die Gestaltung sowohl technischer als auch sozialer Faktoren.

Die notwendigen Prinzipien zur Gestaltung sozio-technischer Systeme [Ch76], [Cl00] können in den Entwurf KI-basierter Systeme einfließen und um die Erfüllung der Anforderungen aus den besonderen Herausforderungen im Umgang mit intelligenten Systemen erweitert werden. Damit ergibt sich die dritte Dimension von *AI-ware*: die Verbindung von

KI und Software-Engineering. Die Konstruktion von KI-basierten Systemen profitiert vor allem – aber nicht nur – von der Vielfalt der etablierten Techniken und bewährten Verfahren des Software-Engineering und erweitert diese gleichzeitig um die entsprechenden KI-Spezifika – c) *Softwareentwicklung für AI-ware*.

3 Entwurf eines Gestaltungsrahmens AI-ware

3.1 Die sozio-technische Systemtheorie als Bezugsrahmen

Unternehmen stehen seit den 50er Jahren des letzten Jahrhunderts vor der Herausforderung, die Potenziale der IT auf eine nutzbringende Art und Weise in die jeweilige Organisationsstruktur zu integrieren. Eine systemische Betrachtung ermöglicht hierbei die bewusste Gestaltung der damit verbundenen Veränderungen in den Strukturen, Aufgaben und Abläufen eines Unternehmens aus einer sozialen *und* einer technologischen Perspektive. Die gestaltungsorientierte Wirtschaftsinformatik liefert hier den erkenntnistheoretischen Rahmen zur Gestaltung einer betrieblichen Informationsverarbeitung. Als Objektbereich der Wirtschaftsinformatik steht der Mensch in Wechselwirkung mit der IT und den zu unterstützenden Aufgaben [He10, S. 7]. Systemisch gewonnene Erkenntnisse ermöglichen mit Bezug auf die Gestaltung betrieblicher Informationssysteme die notwendige gesamthaft Herangehensweise [Ap97]. Soziale und technische Perspektiven werden miteinander kombiniert und berücksichtigen die „Verhaltensaspekte bei der Entwicklung und Implementierung von Technologien“ [Mu22, S. 5]. Für die Gestaltung KI-basierter Systeme bedeutet dies zwangsläufig die Anwendung von Verfahren und Vorgehensweisen auch zur Gestaltung der Interaktion mit den menschlichen Nutzer*innen des Systems und zur Sicherstellung definierter Begleitprozesse von im Einsatz befindlichen KI-basierten Systemen. Jüngere Arbeiten in der anwendungsorientierten Forschung der Gestaltung betonen mittlerweile eine notwendige interdisziplinäre Sicht [Di19]. Die sozio-technischen Systemtheorie in den Erkenntnisbereich der gestaltungsorientierten Wirtschaftsinformatik zu verorten, ist daher sinnvoll. Es liefert die konkrete Motivation des Beitrags für eine tiefere Auseinandersetzung mit diesem Konzept und der Übertragbarkeit auf AI-ware. Das nach Sinz [Si10, S. 28] formulierte Konstruktionsproblem (i: Untersuchungsobjekt, ii: Untersuchungsgegenstand, iii: Lösungsraum) einer gestaltungsorientierten Wirtschaftsinformatik umfasst in der Übertragung auf die Herausforderung in der Konstruktion von AI-ware folgende Aspekte:

- i. KI-basierte Systeme für einen Unternehmenseinsatz können nicht nur auf eine algorithmische Betrachtung beschränkt werden. Allein schon die Definition einer sinnvollen betrieblichen Problemstellung erfordert ein Verständnis über die korrekte Anwendung der verschiedenen KI-Modelle zur Abbildung der dahinterliegenden betrieblichen Aufgaben. Zudem müssen Entscheidungen über den geeigneten Umgang mit Daten getroffen werden, unter anderem über die Herkunft der Daten, deren Nutzbarkeit und Verwendungszweck sowie zu Datenschutz und

Datensicherheit. Aber auch Entscheidungen über eine passende IT-Infrastruktur sind notwendig, ebenso wie über die Ausgestaltung der Prozesse zum Betreiben und Weiterentwickeln der KI-basierten Systeme.

- ii. Eine ganz wesentliche Herausforderung indes ist die Gestaltung der Interaktion der Systeme mit den menschlichen Anwender*innen. Dies bezieht auch die Fragestellung mit ein, inwieweit die Ergebnisse eines KI-basierten Systems in die Arbeitsaufgaben innerhalb der Organisation einfließen sollten und wie und welchem Umfang hoheitliche Entscheidungsaufgaben autonom oder unterstützend bzw. begleitend durch die *Maschine* realisiert werden.
- iii. Insgesamt werden in den eben genannten Bereichen jeweils ganz grundlegende digital-ethische Facetten berührt, die eine spezifische Antwort schon während der Gestaltung der Systeme erfordern [LMM21]. Gerade die Konstruktion vertrauenswürdiger Lösungen, wie diese durch die EU gefordert werden [Ho19], schränkt den möglichen Lösungsraum an Gestaltungsalternativen ein.

Die sozio-technische Systemtheorie [Tr81] scheint daher ein geeigneter Ansatz zu sein, diese mehrdimensionalen Herausforderungen zur Konstruktion von AI-ware umfassend zu untersuchen und Handlungsempfehlungen im Sinne einer gestaltungsorientierten Wirtschaftsinformatik [OWB10] abzuleiten. Darüber hinaus scheint dieser Ansatz auch geeignet, empirisch orientierte Forschungsarbeiten, z.B. zur Erhebung der tatsächlichen Einsatzbereiche von KI-basierten Systemen und deren Reifegrad in Unternehmen, strukturell begleiten zu können. Gerade die Betrachtung als sozio-technisches System ermöglicht zudem auch die Entwicklung eines verhaltensorientierten Forschungsdesigns, die eine menschenzentrierte Gestaltung von Systemen präferieren, wie es z.B. auch die EU-Richtlinien für eine vertrauenswürdige KI [Ho19] fordern.

Einzelne, neuere Arbeiten zeigen, dass die sozio-technische Systemtheorie bereits für verschiedene Forschungsdesigns zur Untersuchung eines betrieblichen Einsatzes von KI-basierten Systemen angewendet wird. Hier wären nur beispielhaft [YXA22] zu nennen in der Anwendung dieser Theorie für die Untersuchung von KI-basierten Systemen am Arbeitsplatz, oder von [Ch22] zur Untersuchung der Zusammenarbeit von Mitarbeiter*innen und KI-Lösungen. Aber auch bei eher komplexeren IT-Technologien, wie z.B. dem Technologiebündel Industrie 4.0, nutzen aktuelle Arbeiten diese Theorie, wie [DCS17], [Mu22] und [SN20] zeigen. Im Fokus ist stets eine ganzheitliche Untersuchung von Technologie in der Interaktion mit einem organisatorischen Rahmen, um die Wirksamkeit dieser zu betrachten.

Sozio-technische Systeme können als Gebilde verstanden werden, die aus den vielfältigen Interaktionen zwischen den Menschen einer Organisation bestehen im Zusammenspiel mit den technischen Komponenten und den verschiedenen Beziehungen zum organisatorischen Umfeld [Mu22]. Dabei fokussieren die sozialen Komponenten auf Elemente wie Menschen, Beziehungen, Organisationen, Anreize sowie die Leistungsmessung, während die technischen Komponenten Elemente beinhalten wie Technologie, Innovation, Wissen, Prozesse und Methoden [Ap97], [Mu22], [Tr81]. Diese beiden Komponenten müssen im Einklang zueinander gestaltet werden, um einen sinnvollen und nutzbringenden Einsatz

einer Technologie in ein organisatorisches Arbeitsumfeld zu garantieren. Nach dem ursprünglichen sozio-technischen Modell von [Le65] können vier Dimensionen den Untersuchungsraum für sozio-technische Systeme und den damit verbundenen organisatorischen Änderungen liefern: *people*, *task*, *structure*, und *technology*. Jede dieser Dimensionen ist miteinander verbunden und Änderungen in einer Dimension haben Auswirkungen auf die anderen und bedürfen damit der Nachjustierung des gesamten Systems. Die ersten drei Dimensionen betrachten hierbei eine soziale Perspektive, während die vierte sich technischen Aspekten widmet. Entsprechend der Definition nach Hess [He10] deckt sich dieser Dimensionenansatz mit dem Objektbereich der Wirtschaftsinformatik.

Einige aktuelle Arbeiten haben diese vier Dimensionen um zwei weitere erweitert [Da14], sowie um eine intensivere Betrachtung externer Umfeldbedingungen. Für den Entwurf eines ersten *Gestaltungsrahmens AI-ware* jedoch wird im Weiteren die Version der vier Dimensionen nach Leavitt [Le65] genutzt. Auch hier zeigen aktuelle Untersuchungen (vgl. [KL17]), dass diese ausreichend sind, um grundlegende Anforderungen an die sozio-technische Systemgestaltung zu realisieren. Grundsätzlich zeigen z.B. Arbeiten von [Cl00] und [BS11], dass dieser systemische Ansatz sehr überzeugend für das Design neuer Technologien und einem technologie-induzierten organisatorischen Wandel herangezogen werden kann [Da14]. Die Nutzung der sozio-technischen Systemtheorie betrachtet Lösungen als offene Systeme [BS11], die mit ihrer Umwelt im Austausch stehen. Die Gestaltung und Einführung KI-basierter Systeme sollte auch diesem Grundsatz unterliegen. Die Nutzung der sozio-technischen Systemtheorie ist im Einklang mit einer gestaltungsorientierten Wirtschaftsinformatik [OWB10] und kann eine Erweiterung des Erkenntnisbeitrags liefern durch eine Systematisierung von anwendungsorientierten Fragestellungen und die Definition geeigneter Forschungsfragen, auch für eine empirisch geführte Untersuchung, des konkreten Konstruktionsproblems AI-ware.

3.2 Entwurf eines Gestaltungsrahmens: Handlungsagenda AI-ware

Einer der wesentlichsten Vorteile in der Anwendung der sozio-technischen Systemtheorie besteht darin, die Gestaltung KI-basierter Systeme unter sozialen Gesichtspunkten der Interaktion in Organisationen zu betrachten. Daraus erwächst nicht nur Gestaltungsbedarf für die damit verbundenen betrieblichen Veränderungen bei der Einführung dieser neuen Form von Informationssystemen, vor allem in Bezug auf die Neugestaltung von Stellen und einzelnen Arbeitsaufgaben und -abläufen [Da14]. Ganz entscheidend trägt diese Theorieanwendung dazu bei, die traditionellen Designanforderungen an solche Systeme weg von einer bislang nur vorherrschenden Informatik-Sicht zu einer multidisziplinären zu erweitern.

Entsprechend der sozio-technischen Systemtheorie steht AI-ware im Sinne eines Synonyms für KI-basierte Anwendungen für ein sozio-technisches System, dass aus sozialen und technischen Komponenten und deren Interaktion in Relation zu den organisatorischen Rahmenbedingungen des Umfelds besteht. Abb. 1 skizziert solch einen möglichen Gestal-

tungsrahmen AI-ware. Hierbei adressiert die soziale Dimension ‚Aufgaben‘ die Gesamtheit der Arbeitsaufgaben, die zur Verrichtung des Organisationszwecks notwendig sind und in arbeitsteiligen Organisationen umfassend und vielschichtig analysiert und synthetisiert werden müssen. Die Dimension ‚Akteure‘ richtet sich an die Mitarbeiter*innen des Unternehmens als humane Ressourcen zur Erfüllung der Arbeitsaufgaben. Der Begriff Akteure³ orientiert sich hier vor allem an den mit den Aufgaben verbundenen Kompetenzen und Befugnissen und berücksichtigt für diese Erfüllung die sozialen Merkmale. ‚Struktur‘ umfasst die Gesamtheit aller formalen und informalen Arbeitsabläufe, die neben den Arbeitsaufgaben das Organisationsgebilde vervollständigt. Die technische Dimension ‚Technologie‘ meint alle Tools, Techniken und im Fall des aktuellen Untersuchungsobjekts alle mit dem Einsatz KI-basierter Systeme tangierenden Technologien. Grundsätzlich umfasst es die Gesamtheit aller Technologien, auch Maschinen im weitesten Sinne und Werkzeuge, für die Unterstützung des Organisationszwecks. Diese Dimensionen besitzen eine unscharfe organisatorische Grenze zur jeweiligen Unternehmensumwelt.

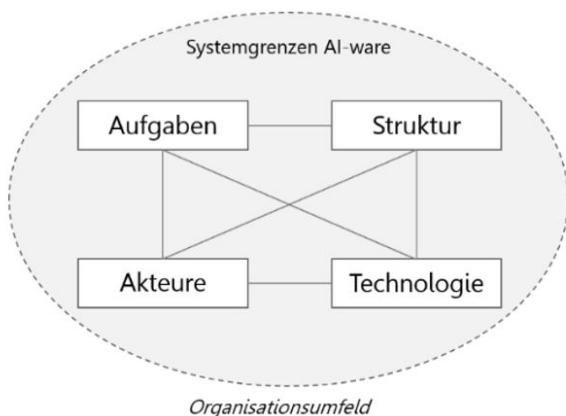


Abb. 1: Gestaltungsrahmen AI-ware nach der sozio-technischen Systemtheorie

Alle Dimensionen stehen zueinander in Relation und sind eingebettet in ein spezifisches Organisationsumfeld. Nach [Da14] kann dieses Umfeld in einer Erweiterung die Stakeholder der Organisation betrachten, Regularien und Vorgaben sowie allgemeine wirtschaftliche sowie finanzielle (und gesellschaftliche) Gegebenheiten. Die Konkretisierung dieses Gestaltungsrahmen AI-ware umfasst damit die Ausgestaltung dieser vier Dimension, die explizite Betrachtung der Interaktion zwischen diesen und sollte insbesondere die spezifischen Umfeldbedingungen berücksichtigen. Daraus kann folgende **Handlungsagenda AI-ware** abgeleitet werden, mit deren Hilfe Forschungsfragen gefunden und definiert werden können. Im Sinne einer gestaltungsorientierten Wirtschaftsinformatik ist die erste Ableitung einer Handlungsagenda normativ zu verstehen und enthält Ziel-Mittel-Aussagen [Oe10, S. 3].

³ Der Begriff wird geschlechtsneutral verwendet.

I. Aufgaben

Die Arbeitsaufgaben einer modernen Organisation unterliegen schon seit längerem der systematischen Überprüfung zur Automation. Vor allem repetitive Aufgaben zeigen von jeher ein überdurchschnittlich hohes Potenzial zur Automatisierung bzw. Übernahme durch technologische Lösungen. Hierfür sind jedoch nicht immer KI-basierte Systeme notwendig. Zunehmend zeigt sich jedoch gerade im Bereich datenbasierter Entscheidungen, dass KI-basierte Systeme hier wesentliche Vorteile bringen können [LMM21] und damit Lösungen umfassender Geschäftsanalysen bieten [Da18]. Die eigentliche Herausforderung besteht darin, das tatsächliche Problemfeld an Aufgaben für einen Einsatz von KI-basierten Systemen zu finden, zu definieren und gleichzeitig den Umfang einer zukünftigen Nutzbarkeit zu bestimmen. Das beinhaltet auch die Fragestellung, inwieweit Anwendungen mit KI-Anteilen die Robustheit und Verfügbarkeit aufweisen müssen, um dauerhaft in einem kontinuierlichen Gebrauch weg von einem prototypischen Einsatz genutzt zu werden. Folgende Teilespekte sollten unter anderem näher beleuchtet werden, wie diese z.B. in [LM20] und [Ng20] adressiert werden:

- Ermittlung der Natur von Arbeitsaufgaben vor dem Hintergrund einer KI-basierten Erbringung und damit auch die Entscheidung über den Grad einer möglichen Automation,
- Ermittlung des Umfangs einer Unterstützung und/oder Übernahme, auch vor dem Hintergrund einer weiterführenden und organisatorisch breiteren Nutzbarkeit solcher Aufgabenbestandteile,
- gezielte Neugestaltung betrieblicher Aufgaben durch KI-basierte Anteile sowie
- Aufbau eines Kennzahlensystems zur Ermittlung der Wirksamkeit für eine KI-basierte Abbildung.

Dieser Teilbereich wird auch medienwirksam stark diskutiert, beginnend mit der, mittlerweile teilweise revidierten, Studie von Frey und Osborne [Fo17] in Bezug auf die Zukunft von Arbeit und den Wegfall von Arbeitsaufgaben aufgrund der Übernahme durch KI. Auch aktuellere Studien zeigen nach wie vor ein hohes Optimierungspotential für bestimmte Aufgabenbereiche [Ch21]. Das Adoptionspotenzial für KI wird jedoch eher unspezifisch erfasst ohne konkrete Sichtbarkeit einzelner Aufgaben oder eben von Aufgabengruppen oder gar Prozessen, so dass hier nach wie vor Untersuchungsbedarf besteht über die Eignung von Arbeitsaufgaben und deren Automatisierung und Unterstützung durch KI-basierte Systeme. So werden vor allem operative Aufgaben der Serviceerbringungen als Aufgabenbereich angesehen, mit KI-basierten Systemen zu unterstützen, ebenso wie allgemeine analytische Fähigkeiten [Ch21].

II. Akteure

Diese Dimension wird bislang zu wenig betrachtet, da allgemeinhin davon ausgegangen wird, dass KI-basierte Systeme zu einer angestrebten Vollautomation führen [La21]. Damit wird auch der Ansatz eines menschenzentrierten Designs zu wenig beachtet. Daher sind in diesem Bereich vor allem zwei Untersuchungsbereiche wesentlich, die im Wesentlichen auch durch die Forderung aus der EU-Richtlinie [Ho19] zur menschenzentrierten Gestaltung von KI-basierten Systemen gestützt wird:

- Gestaltung KI-basierter Systeme beginnend mit einer menschenzentrierten Vorgehensweise des Designs, auch unter Berücksichtigung digital-ethischer Aspekte und unter Berücksichtigung von (ökonomischer und ökologischer) Nachhaltigkeit [Di19], [Le19] sowie
- Ermittlung mehrdimensionaler sozialer Faktoren in der Interaktion der Systeme mit dem Menschen (im Arbeitsumfeld und als Kund*innen der Produkte und Services des Unternehmens), auch unter der Berücksichtigung von Faktoren wie Entscheidungsgröße und weiteren Prinzipien wie Erklärbarkeit und Transparenz, Verantwortlichkeit, Sicherheit und Fairness [LMM21].

III. Struktur

Die Prozesse des Unternehmens stehen in enger Beziehung zu den einzelnen Arbeitsaufgaben. Daher gelten für diese Dimension ähnliche Herausforderungen wie für die Dimension der Aufgaben. Hier sollten jedoch weitere Untersuchungsgegenstände betrachtet werden, in erster Linie die folgenden, wie diese unter anderem auch bei der Gestaltung von klassischen Informationssystemen fokussiert werden [vgl. u.a. He10]:

- Wirksamkeit einer strategischen Einordnung der Technologie KI insgesamt in den Unternehmenskontext,
- Zielorientierung bei der Umsetzung bzw. Durchsetzung strategischer Ziele, z.B. über eine definierte Governance-Struktur,
- Unternehmenskulturelle Auswirkungen eines stärkeren Einsatzes KI-basierter Systeme im Hinblick auf die Veränderung normativer Unternehmenswerte und Normen,
- Ausmaß und Eignung von Maßnahmen zur Schulung und Weiterbildung der Mitarbeitenden im Hinblick auf die Anwendung KI-basierter Systeme sowie
- Ermittlung von Strukturen für eine geeignete skalierbare Abbildung prototypischer Szenarien für den Einsatz KI-basierter Systeme hin zu wirksamen laufenden Anwendungen.

IV. Technologie

Diese Dimension betrachtet den technischen Rahmen des Einsatzes KI-basierter Systeme. Vor allem drei Untersuchungsbereiche sind hier relevant, die ganz wesentlich die Erkenntnisträume des IT-Managements [Kr15] adressieren:

- Eignung traditioneller Modelle und Vorgehensweise betrieblicher Anwendungsentwicklung (Software-Engineering) und Erweiterung um die Herausforderungen bei der Entwicklung KI-basierter Systemlösungen,
- Aufbau einer geeigneten IT-Infrastruktur für einen dauerhaften robusten und sicheren Betrieb solcher Systeme inkl. grundlegender Integrationsfragen in die bestehende IT-Landschaft des Unternehmens sowie
- Untersuchung der Eignung klassischer Modelle aus dem IT-Management für die notwendigen Prozesse der technischen Integration und Betriebsführung (Operation) solcher Systeme.

V. Organisationsumfeld

Das regulatorische Umfeld von KI weist seit den letzten Jahren eine hohe Dynamik auf. Die EU hat mit ihren Richtlinien für eine vertrauenswürdige KI [Ho19] grundlegende Regeln im Umgang mit KI ganz allgemein geschaffen. Die Überführung dieser in gesetzliche Vorgaben, z.B. auch über die Definition von Risikolevel für den Einsatz, ist bereits in Arbeit, wie aktuelle Beiträge zeigen [Ta18]. Eine digital-ethisch konforme Gestaltung des Organisationsumfelds erfordert einen Dreiklang in Bezug auf die Einhaltung definierter ethischer Prinzipien (z.B. nach Vorgabe der EU), den vorgegebenen behördlichen Regularien und Gesetzen sowie einer formulierten Governance im Umgang mit KI im Unternehmen [Fl18]. Gesamthaft gesehen ergeben sich drei wesentliche Handlungsstränge, die in [ML21] als *AI-ware Agenda* skizziert worden sind:

- Gesellschaftliche und digital-ethische Betrachtungsschwerpunkte, die alle drei sozialen Dimensionen und das Organisationsumfeld betreffen.
- Technologische Betrachtungsschwerpunkte in Bezug auf Daten und Algorithmen sowie die Gestaltungsansätze zur Auswahl und Anwendung einzelner KI-Modelle und deren Umsetzung (Softwareentwicklung). Diese adressieren die technische Dimension.
- Wirtschaftliche, finanzielle und regulatorische Betrachtungsschwerpunkte, in denen die vier Dimensionen eingebettet sind.

Im Sinne der Konstruktion eines offenen sozio-technischen Systems wird auch der notwendige interdisziplinäre Charakter im Kontext KI berücksichtigt.

4 Zusammenfassung

Der verantwortungsvolle Umgang und die ethisch konforme Gestaltung KI-basierter Systeme erfordert die Berücksichtigung eines Rahmens, in dem nicht nur technische, sondern auch soziale sowie organisatorische Aspekte im Vordergrund stehen. Basierend auf der sozio-technischen Systemtheorie haben wir in diesem Beitrag den Gestaltungsrahmen AI-ware eingeführt und die Handlungsagenda AI-ware abgeleitet, mit deren Hilfe KI-basierte Systeme entwickelt und Forschungsfragen, sowohl anwendungsorientiert als auch empirisch, in Bezug auf den betrieblichen Einsatz KI-basierter Systeme thematisiert werden können. In den letzten Jahren ist eine zunehmende Entwicklung unterschiedlicher Ansätze, Richtlinien und Regulierungen (siehe z.B. [Di19], [Ho19], [Ie19], [Ta18]) zu beobachten, die das Thema des richtigen Einsatzes von KI in Unternehmen und deren Auswirkungen mit verschiedenen Schwerpunkten zu erklären versuchen. Allerdings fehlt an vielen Arbeiten ein übergeordneter Rahmen, um alle Aspekte der Entwicklung solcher Systeme zu betrachten. Das Ziel dieses Beitrags als Startpunkt für eine theoretische Fundierung aus der interdisziplinären Sicht der gestaltungsorientierten Wirtschaftsinformatik sollte sowohl den Erwartungen an KI als auch den tatsächlichen Entwicklungen in dem Bereich gerecht werden. Die sozio-technische Systemtheorie kann solch ein theoretisches Fundament liefern, auf dem weitere Forschungsarbeiten systematisiert erfolgen können.

Literaturverzeichnis

- [Ap97] Appelbaum, S. H.: Socio-technical systems theory: an intervention strategy for organizational development. *Management Decision*, 35(6), 452-463, 1997.
- [BH00] Brynjolfsson, E.; Hitt, L. M.: Beyond Computation: Information Technology, Organizational Transformation and Business Performance. *Journal of Economic Perspectives*, 14(4): 23-48, Fall 2000.
- [BS11] Baxter, G.; Sommerville, I.: Socio-technical systems: From design methods to systems engineering. *Interacting with Computers*, 23(1): 4-17, January 2011.
- [Ch76] Cherns, A.: The Principles of Sociotechnical Design, *Human Relations*, 29(8): 783-792, 1976.
- [Ch22] Chowdhury, S. et al.: AI-employee collaboration and business performance: Integrating knowledge-based view, socio-technical systems and organisational socialisation framework. *Journal of Business Research*, 144:31-49, 2022.
- [Ch21] Chui, M. et al.: The state of AI in 2021. McKinsey & Company, 2021. <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/global-survey-the-state-of-ai-in-2021>, Zugriff: 03.05.2022.
- [Cl00] Clegg, C. W.: Sociotechnical principles for system design. *Applied Ergonomics*, 31(5): 463-477, October 2000.
- [Co20] Coeckelbergh, M.: AI Ethics. Cambridge, Massachusetts: The MIT Press, 2020.
- [Cr21] Crawford, K.: Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press, New Haven, 2021.
- [Da18] Davenport, T. H.: From analytics to artificial intelligence. *Journal of Business Analytics* 1(2): 73-80, 2018.
- [Da14] Davis, M. et al.: Advancing socio-technical systems thinking: A call for bravery. *Applied Ergonomics*, 45(2), Part A: 171-180, March 2014.
- [DCS17] Davies, R.; Coole, T.; Smith, A.: Review of Socio-technical Considerations to Ensure Successful Implementation of Industry 4.0. *Procedia Manufacturing*, 11: 1288-1295, 2017.
- [Di19] Dignum, V.: Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way. Springer Nature Switzerland, 2019.
- [Fl18] Floridi, L.: Soft Ethics and The Governance of the Digital. *Philosophy & Technology*, 31: 1-8, 2018.
- [FMS19] Fountaine, T.; McCarthy, B.; Saleh, T.: Building the AI-Powered Organisation: Technology isn't the biggest challenge. Culture is. *Harvard Business Review*, 63–73, July-August 2019.
- [FO17] Frey, C. B.; Osborne, M.: The Future of Employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114: 254-280, January 2017.
- [He10] Hess, T.: Erkenntnisgegenstand der (gestaltungsorientierten) Wirtschaftsinformatik. In H. Österle, R. Winter, W. Brenner (Hrsg.), *Gestaltungsorientierte Wirtschaftsinformatik: Ein Plädoyer für Rigor und Relevanz*, infowerk ag, 2010. <https://www.alexandria.unisg.ch/213292/1/ATTD05CN.pdf#page=15>, Zugriff: 04.07.2022.

- [Ho19] HEG-KI: Ethik-Leitlinien für eine vertrauenswürdige KI. Hochrangige Expertengruppe für künstliche Intelligenz, Europäische Kommission, 2019. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>, Zugriff: 06.05.2022.
- [IE19] IEEE: Ethically Aligned Design. First Edition: A vision for prioritizing human well-being with Autonomous and Intelligent Systems. IEEE, 2019. <https://ethicsinaction.ieee.org/#series>, Zugriff: 06.05.2022
- [Jo19] Jordan, M.: Artificial Intelligence—The Revolution Hasn't Happened Yet. Harvard Data Science Review, 1(1), 2019.
- [KL17] Kopackova, H.; Libalova, P.: Smart city concept as socio-technical system. In Proceedings of the 2017 International Conference on Information and Digital Technologies (IDT), IEEE, July 2017.
- [Kr15] Kremer, H.: Informationsmanagement (6. Aufl.). Springer Gabler, 2015.
- [La21] Larson, E. J.: The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do. Harvard University Press, Cambridge, MA, 2021.
- [Le65] Leavitt, H. J.: Applied organizational change in industry: Structural, technological and humanistic approaches. Handbook of organizations (pp. 1144-1170), Rand McNally & Co., Chicago, Illinois, 1965.
- [LM20] Lemke, C.; Monett, D.: KI und Automatisierung: Wirkung auf die Arbeit der Zukunft. In J. Nachtwei; A. Sureth (Hrsg.), Sonderband Zukunft der Arbeit, Human Resources Consulting Review, 12: 400-403, VQP, 2020.
- [LMM21] Lemke, C.; Monett, D.; Mikoleit, M.: Digitale Ethik in datengetriebenen Organisationen und deren Anwendung am Beispiel von KI-Ethik. In Barton T., Müller C. (Hrsg.), Data Science anwenden: Einführung, Anwendungen und Projekte, S. 33-52, Angewandte Wirtschaftsinformatik, Wiesbaden: Springer Vieweg, 2021.
- [Lo19] Loh, J.: Roboterethik. Eine Einführung. Berlin, Suhrkamp Taschenbuch Wissenschaft, 2019.
- [Ma13] Marcus, G.: Hyping Artificial Intelligence yet again. The New Yorker, 2013. <https://www.newyorker.com/tech/annals-of-technology/hyping-artificial-intelligence-yet-again>, Zugriff: 04.07.2022.
- [Mc04] McCorduck, P.: Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence, A. K. Peters/CRC Press, 2004.
- [Mc76] McDermott, D.: Artificial intelligence meets natural stupidity. ACM SIGART Bulletin, 57, pp. 4-9, April 1976.
- [MD19] Marcus, G.; Davis, E.: Rebooting AI: Building Artificial Intelligence We Can Trust. Pantheon Books, New York, 2019.
- [Mi18] Misselhorn, C.: Grundfragen der Maschinenethik. Ditzingen: Reclam, 2018.
- [Mi19] Mitchell, M.: Artificial Intelligence: A Guide for Thinking Humans. Pelican Random House, UK, 2019.
- [ML21] Monett, D.; Lemke, C.: AI-ware: Bridging AI and Software Engineering for responsible and sustainable intelligent artefacts. In van Giffen, B., Kochler, J., Brenner, W., Albayrak, C. A. (eds.), Managing Artificial Intelligence (pp. 66-73), Workshop Paper Series, INFORMATIK 2021, Institute of Information Management, University of St. Gallen, 2021.

- [MLT20] Monett, D.; Lewis, C.W.P.; Thórisson, K.R. (eds.): Special Issue “On Defining Artificial Intelligence.” *Journal of Artificial General Intelligence*, 11(2): 1-100, 2020.
- [Mu22] Münch, C. et al.: Capabilities of digital servitization: Evidence from the socio-technical systems theory. *Technological Forecasting and Social Change*, 176, March 2022.
- [Ng20] Ng, A.: AI Transformation Playbook: How to lead your company into the AI era. Landing AI, 2020. https://landing.ai/wp-content/uploads/2020/05/LandingAI_Transformation_Playbook_11-19.pdf, Zugriff: 04.07.2022.
- [Oe10] Österle, H. et al.: Memorandum zur gestaltungsorientierten Wirtschaftsinformatik. In H. Österle, R. Winter, W. Brenner (Hrsg.), *Gestaltungsorientierte Wirtschaftsinformatik: Ein Plädoyer für Rigor und Relevanz*, infowerk ag, 2010. <https://www.alexandria.unisg.ch/213292/1/ATTD05CN.pdf#page=9>, Zugriff: 04.07.2022.
- [OWB10] Österle, H.; Winter, R.; Brenner, W. (Hrsg.): *Gestaltungsorientierte Wirtschaftsinformatik: Ein Plädoyer für Rigor und Relevanz*, 2010. <https://www.alexandria.unisg.ch/213292/1/ATTD05CN.pdf>, Zugriff: 05.05.2022.
- [Ox19] Oxford: Government Artificial Intelligence Readiness Index 2019. Oxford Insights and the International Development Research Centre, 2019. <https://www.oxfordinsights.com/ai-readiness2019>, Zugriff: 03.05.2021.
- [Oz20] Ozkaya, I.: What Is Really Different in Engineering AI-Enabled Systems? *IEEE Software*, 3-6, July/August 2020.
- [PM85] Porter, M. E.; Millar, V. E.: How Information Gives You Competitive Advantage. *Harvard Business Review*, July, 1985. <https://hbr.org/1985/07/how-information-gives-you-competitive-advantage>, Zugriff: 05.05.2022.
- [RG22] Rao, A.; Greenstein, B.: PwC 2022 AI Business Survey. PwC, 2022. <https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-business-survey.html>, Zugriff: 03.05.2022.
- [RN20] Russell, S.; Norvig, P.: *Artificial Intelligence: A Modern Approach*, 4. Aufl., Pearson Series in Artificial Intelligence, 2020.
- [Sh56] Shannon, C. E.: The Bandwagon. *IRE Transactions on Information Theory*, 3, 1956.
- [Si10] Sinz, E.: Konstruktionsforschung in der Wirtschaftsinformatik: Was sind die Erkenntnisziele gestaltungsorientierter Wirtschaftsinformatik-Forschung? In H. Österle, R. Winter, W. Brenner (Hrsg.), *Gestaltungsorientierte Wirtschaftsinformatik: Ein Plädoyer für Rigor und Relevanz*, infowerk ag, 2010. <https://www.alexandria.unisg.ch/213292/1/ATTD05CN.pdf#page=35>, Zugriff: 04.07.2022.
- [Sm19] Smith, B. C.: *The Promise of Artificial Intelligence: Reckoning and Judgment*. The MIT Press, Cambridge, MA, 2019.
- [SN20] Sony, M.; Naik, S.: Industry 4.0 integration with socio-technical systems theory: a systematic review and proposed theoretical model. *Technology in Society*, 61, 101248, 2020.
- [So16] Sommerville, I.: *Software Engineering* (10. Aufl.). Pearson, 2016.
- [Ta18] Taylor, S. et al.: Responsible AI – Key Themes, Concerns & Recommendations for European Research and Innovation. European AI Alliance, European Commission, 2018. <https://doi.org/10.5281/zenodo.1303252>, Zugriff: 06.05.2022.

- [Tr81] Trist, E. L.: The Evolution of Socio-technical Systems: A Conceptual Framework and an Action Research Program. *Issues in the quality of working life*, 2(2): 1-67, Ontario Ministry of Labour, Ontario Quality of Working Life Centre, 1981.
- [Ve94] Venkatraman, N.: IT-Enabled Business Transformation: From Automation to Business Scope Redefinition. *MIT Sloan Management Review*, Magazine Winter, 1994.
- [VF22] Vinsel, L.; Funk, J.: Blinded by the Hype. *Open Mind*, 2022. <https://www.openmindmag.org/articles/blinded-by-the-hype>, Zugriff: 04.07.2022.
- [YXA22] Yu, X.; Xu, S.; Ashton, M.: Antecedents and outcomes of artificial intelligence adoption and application in the workplace: the socio-technical system theory perspective. *Information Technology & People*, Vol. ahead-of-print No. ahead-of-print, 2022.
- [Zh22] Zhang, D. et al.: The AI Index 2022 Annual Report. AI Index Steering Committee, Stanford Institute for Human-Centered AI, Stanford University, March 2022.

Probabilistic Programming – ein neuer Baustein für die KI

Motto: Wenig Daten? Dann schlau Raten!

Thomas Wengerek¹

Abstract: Wenn statt riesiger Datenmengen ein generatives Modell (also ein probabilistischer Simulator) für die Erzeugung bzw. Entstehung der Daten der Anwendungsdomäne zur Verfügung steht, kann dieser Simulator durch bayesianische Inferenz „invertiert“ werden, d.h. es können diejenigen latenten Parameter des Modells bestimmt werden, die für jeweils vorgegebene Daten verantwortlich sind. Universelle Probabilistische Programmiersprachen (PPL) sind derart expressiv, dass sich mit ihnen das generative Modell, etwaige Bedingungen an die Daten und das Vorwissen des Anwenders bzgl. der latenten Parameter des Modells als ein integriertes Programm formulieren lassen. Die Ausführung dieses Programms erzeugt dann Stichproben der bayesianischen Posterior-Verteilung der gesuchten Modell-Parameter. Durch moderne Allzweck-Algorithmen zur Inferenz kann diese Verteilung exploriert und für den jeweiligen Anwendungszweck interpretiert werden. In komplexen Problemstellungen können somit in der Sprache der probabilistischen Logik formulierte Fragen durch Probabilistic Programming beantwortet werden. Dies eröffnet für viele Anwendungsgebiete, auch in der Wirtschaftsinformatik, ungeahnte Möglichkeiten. Die vorliegende Arbeit soll die Ideen und Potentiale des Probabilistic Programming einem breiteren Kreis von Anwender*innen nahebringen. Dazu werden einerseits die grundlegenden Konzepte sorgfältig dargestellt und in ihrem Zusammenspiel motiviert und andererseits werden diese neuen Ideen anhand mehrerer Anwendungsbeispiele aus dem einschlägigen Forschungsumfeld vorgestellt. Dabei wurden einige dieser Beispiele, für die Softwaresysteme im Open Source Bereich verfügbar sind, auch praktisch erprobt.

Keywords: Bayesianische Inferenz, generatives Modell, probabilistische Logik, Inferenz-Algorithmen, Modell-Invertierung

1 Generative Modelle

Akteure in der Welt müssen mit Unsicherheit umgehen können. Das Wissen über die Welt ist ausschnitthaft und verrauscht, weil die Daten immer unvollständig und fehlerbehaftet sind. In dieser Lage muss ein solcher Akteur versuchen, das vorhandene Wissen so zu verwalten, dass es 1.) akkumuliert und flexiblen Schlussfolgerungsmechanismen zugänglich gemacht werden kann, und 2.) in geeigneter Weise probabilistisch interpretiert werden

¹ Hochschule Stralsund/Fakultät Wirtschaft, Zur Schwedenschanze 15, 18435 Stralsund,
thomas.wengerek@hochschule-stralsund.de

kann. Wie lassen sich diese beiden scheinbar unversöhnbaren Anforderungen produktiv in einem gemeinsamen Rahmen vereinen? In diesem Dilemma hilft die Beobachtung, dass es zwei verschiedene Wege [NJ01] zur Modellierung gibt: den Diskriminativen und den Generativen. Die Regressions- und Klassifikationsverfahren des maschinellen Lernens wählen den diskriminativen Weg. Dabei dienen inhärente Eigenschaften der konkret vorliegenden Daten als Unterscheidungsmerkmale. Beim generativen Weg hingegen wird versucht, einen Erzeugungsprozess (generatives Modell) für die zu beobachtenden Daten zu finden, um diese einerseits reproduzieren und andererseits vorhersagen zu können. Ein solcher Erzeugungsprozess wird in der Regel stochastischer Natur sein. Und wenn die Beschreibung eines solchen generativen Modells mittels eines geeigneten, strukturreichen Formalismus erfolgt, dann kann das Modell auch zur Arbeitsgrundlage für Schlussfolgerungsmechanismen werden, um mit ihrer Hilfe „Was-wäre-wenn-Fragen“ im Kontext des Modells zu beantworten (darauf wird im nächsten Abschnitt konkreter eingegangen). Darüber hinaus könnte es sogar möglich werden, eine über die aktuellen Daten hinausgehende, abstraktere Betrachtungsebene einzunehmen und hypothetische Überlegungen über das Modell selbst anzustellen.

2 Probabilistische Programme

Das Alles ist vielversprechend und ambitioniert – bleibt nur die Frage: wie lässt es sich formalisieren?

Hierzu bietet sich die Verwendung der probabilistischen Logik [Ja03] in der Form grafischer Modelle in Bayes-Netzen [Pe88, KF09] oder noch konsequenter im Model-based Machine Learning [Bi13, Wi19] an. Ein Bayes-Netz definiert eine Wahrscheinlichkeitsverteilung, die auf einer gerichteten, azyklischen Graphenstruktur beruht, die die wechselseitigen Unabhängigkeiten der bedingten Wahrscheinlichkeiten von Teilmengen der relevanten Größen widerspiegelt. Es sei angemerkt, dass in der probabilistischen Logik mit bedingten Wahrscheinlichkeiten sogar nicht-monotone Schlussfolgerungen in natürlicher Weise vorkommen können [Gro86] – eine Eigenschaft, die in der klassischen Logik nicht ohne gravierende Erweiterungen realisierbar ist. Dieser Graph ist somit das Gerüst des zugrundeliegenden Modells. Da die Expressivität graphischer Modelle für viele Anwendungen aber nicht ausreicht, wurde vorgeschlagen, Logik und Wahrscheinlichkeit in einer neuen, formalen Struktur, dem stochastischen Lambda-Kalkül [RP02], zu vereinen und als eine universelle Probabilistische Programmiersprache (PPL) zu realisieren [Gn08], wobei dieser Sprache zwei weitere Merkmale hinzugefügt werden: (i) Variablen können expliziten Bedingungen hinsichtlich ihres erlaubten Wertebereichs unterworfen werden (sog. „conditioning“). Und (ii) um die Ergebnisse von Berechnungen dieser Sprache auch in herkömmlichen Programmiersprachen weiterverwenden zu können, wird eine spezielle Schnittstelle benötigt (sog. „query“). Berechnungen, die in dieser Sprache formuliert sind, werden fortan als „Probabilistische Programme“ (PP) bezeichnet. PP beschreiben immer eine Wahrscheinlichkeitsverteilung – allerdings in Stichproben-Semantik („sample-based semantics“). Was bedeutet das?

Jeder Lauf des PP entspricht einem Pfad („trace“) durch den Berechnungsbaum. An Knotenpunkten dieses Baums können probabilistische Funktionen aufgerufen werden, die den weiteren Weg durch den Baum nicht-deterministisch beeinflussen. Das Ergebnis eines solchen Pfades wird also durch die akkumulierten Wirkungen der Wahrscheinlichkeitsverteilungen, die an den Knoten des Berechnungsbaumes wirken, bestimmt. Insofern gehorcht das Ergebnis ebenfalls einer Gesamtwahrscheinlichkeitsverteilung – eben derjenigen des gesamten PP. Jeder Lauf des PP resultiert daher in einer Stichprobe („sample“) der Gesamtwahrscheinlichkeitsverteilung, die durch den Programmcode des PP definiert wird. Man erkaufte sich also die universelle Expressivität einer Turing-vollständigen Probabilistischen Programmiersprache damit, dass die Ergebnisse der mit ihrer Hilfe geschriebenen PP, zwar als Wahrscheinlichkeitsverteilungen vorliegen – aber eben nicht in geschlossener Form, sondern nur als Stichproben-Prozess beim wiederholten Ausführen des PP.

Halten wir fest: mit einer universellen PPL kann - qua Programmcode - jede vorstellbare Wahrscheinlichkeitsverteilung (in Stichprobensemantik) beschrieben werden. Für die Implementierung generativer Modelle steht somit ein universelles Konstruktionswerkzeug zur Verfügung.

3 Bayesianische Inferenz und moderne Inferenz-Algorithmen

Daten sollen nicht nur durch probabilistische Simulationen (d.h. generative Modelle) reproduziert werden können, sondern in Bezug auf diese Simulationen sollen auch Fragen gestellt und Schlussfolgerungen [Me18] gezogen werden können. Wie sieht ein geeigneter, probabilistischer Schlussfolgerungsmechanismus in diesem Kontext nun aus?

Das Konzept der bayesianischen Inferenz [Mc20] in der Form der elementaren Bayes-Regel hilft hier weiter:

$$P(\theta | D) = (P(D | \theta) * P(\theta)) / P(D) \quad (1)$$

Das generative Modell wird intern durch Modellparameter θ (sog. latente Variablen) kontrolliert. Je nachdem welche Werte θ annimmt, wird das generative Modell bestimmte Daten D erzeugen. Dieser Zusammenhang wird durch die Likelihood $P(D|\theta)$ dargestellt. Die bedingte Wahrscheinlichkeit $P(D|\theta)$ lässt sich auch als kausale Perspektive interpretieren, in der θ die Ursache und D die Wirkung ist. Zusätzlich sei es möglich, explizite Bedingungen („conditioning“) an die zu beobachtenden Daten zu stellen. Diese Bedingungen stellen die spezifischen Fragen dar, die im Zuge der Inferenz beantwortet werden sollen. Der Kürze der Darstellung halber sei hier angenommen, dass diese Bedingungen in der Likelihood bereits subsumiert seien. Das im Vorfeld vorhandene Wissen (oder „Un-Wissen“) über die Modellparameter wird durch die Apriori-Verteilung $P(\theta)$ ausgedrückt. Dies ist derjenige Aspekt der bayesianischen Statistik, der häufig als sog. „subjektives Wissen“ kontrovers diskutiert wird [La18]. $P(D)$ ist die Randverteilung. Sie stellt einen Normierungsfaktor dar. Die linke Seite von (1) ist die Posterior-Verteilung $P(\theta|D)$. Sie stellt das eigentliche Ziel der

Inferenz dar. Die Posterior-Verteilung beschreibt, welche Modellparameter wahrscheinlich dafür verantwortlich sind, wenn bestimmte Daten vorliegen. Man kann auch sagen: die Inferenz invertiert die kausale Richtung von Ursache und Wirkung, da von $P(D|\theta)$ auf $P(\theta|D)$ geschlossen wird:

$$P(\theta | D) \leftarrow P(D | \theta) \quad (2)$$

Diese Invertierung ist der entscheidende Aspekt der Inferenz. Zwar ist die Bayes-Regel formal gesehen symmetrisch, aber in der konkreten Anwendung gibt es nun einmal die Unterscheidung von Ursache (latente Variablen, Modellparameter) und Wirkung (Daten), und wir interessieren uns hier für die Modellparameter. Durch die Invertierung wird es somit möglich, die Ursache, also die Modellparameter, einer gegebenen Wirkung zu bestimmen. Man kann dies formal auch so ausdrücken [La18]:

$$P(\text{Ursache} | \text{Wirkung}) \Leftrightarrow P(\text{Wirkung} | \text{Ursache}) \quad (3)$$

Mit modernen Inferenzalgorithmen [WMM14] kann die in Stichprobensemantik vorliegende Posterior-Verteilung $P(\theta|D)$ näherungsweise bestimmt werden. D.h. die implizit als Programmcode des PP vorliegende Posterior-Verteilung

$$P(\theta | D) \approx \text{Probabilistisches Programm} \quad (4)$$

wird durch einen Inferenzalgorithmus explizit ausgewertet. Bei komplexen, hoch-dimensionalen Problemen kann dies approximativ geschehen oder es wird zumindest eine Maximum-Aposteriori Schätzung (MAP) [Mc20] bestimmt. Eine wichtige Klasse derartiger Algorithmen basiert auf Markov-Chain Monte-Carlo Methoden (MCMC) in der speziellen Variante von Metropolis-Hastings [La18]. Das PP wird dazu eigens in Continuation-Passing-Style (CPS) transformiert, damit der Inferenzalgorithmus optimal mit dem Code arbeiten kann [GS22]. Aber daneben werden viele weitere Verfahren erprobt. Die Verbesserung der Inferenzalgorithmen bzgl. Problemgröße und Geschwindigkeit ist Teil der aktuellen Forschung. Einen guten Überblick dazu bietet [Me18].

Die übergeordneten Forschungsbemühungen um die „Probabilistischen Programmierung“ setzen sich das ambitionierte Ziel [Me18], viele Gebiete des maschinellen Lernens auch für Nicht-Expert*innen leichter zugänglich zu machen, indem die Modellbildung auf ein Programmierproblem zurückgeführt wird, für das keine tiefgehenden Spezialkenntnisse erforderlich sind. Und die Inferenz erfolgt dann automatisch durch leistungsfähige, universell einsetzbare Verfahren. Das Potential dieser Forschungsagenda wird auch von der Anwendungsseite mit Interesse verfolgt und aktiv unterstützt [Pyr17].

4 Anglican und ein didaktisches Beispiel

Es gibt eine Vielzahl von PPL, die auf den oben vorgestellten Konzepten beruhen. Darin ist die Gruppe der universellen PPL besonders interessant, weil es bei ihnen wenige Beschränkungen der Ausdrucksmitte bei der Code-Erstellung gibt und weil auch externe Bibliotheken für die Implementierung des generativen Modells herangezogen werden dürfen. Darüber hinaus stehen für diese universellen PPL eine ganze Reihe unterschiedlicher Inferenzverfahren zur Verfügung. Eine dieser universellen, probabilistischen Programmiersprachen, die eine gewisse Vorreiterrolle einnimmt, ist Anglican [To16, Ang18]. Anglican basiert auf der modernen Lisp-Variante Clojure [Clo22]. Ein instruktives Anwendungsbeispiel - quasi das „HelloWorld“ der „Probabilistischen Programmierung“ - ist das Äquivalent zur linearen Regression (siehe Abb. 1). Dabei sollen die sechs Datenpunkte mit folgenden x-y-Werten $[0, 0.6], [1, 0.7], [2, 1.2], [3, 3.2], [4, 6.8], [5, 8.2], [6, 8.4]$ durch eine lineare Funktion f mit der Steigung s und dem Achsenabschnitt b approximiert werden. Das PP ist ein Anglican-Ausdruck folgenden Inhalts: die lineare Funktion f ist das generative Modell. Ihre Funktionsparameter, s und b , sind die latenten Variablen des Modells. Als Apriori-Verteilungen dieser latenten Variablen werden Normalverteilungen (Mittelwerte jeweils 0, Standardabweichungen 2 bzw. 6) angenommen. Das „conditioning“ (im Code mit `observe` bezeichnet) beschreibt die Anforderungen, die durch die Daten an das Modell gestellt werden. D.h. konkret: die Funktionswerte ($f x$) sollen möglichst nah bei den y-Werten der Datenpunkte liegen. Damit der Inferenz-Algorithmus stabil arbeiten kann, wird auch hier eine Normalverteilung angenommen. Das PP `linear-regression-posterior` stellt also insgesamt die Frage: welche Werte müssen die latenten Variablen des generativen Modells annehmen, damit es zu den Datenpunkten möglichst gut passt? Ganz unten im Code wird durch den Vektor $[s b]$ expliziert, welche Form die Antwort dann haben soll.

```
(defquery linear-regression-posterior []
  (let [s (sample (normal 0 2))
        b (sample (normal 0 6))
        f (fn [x] (+ (* s x) b))]
    (observe (normal (f 0) 0.5) 0.6)
    (observe (normal (f 1) 0.5) 0.7)
    (observe (normal (f 2) 0.5) 1.2)
    (observe (normal (f 3) 0.5) 3.2)
    (observe (normal (f 4) 0.5) 6.8))
```

```
(observe (normal (f 5) 0.5) 8.2)
(observe (normal (f 6) 0.5) 8.4)
[s b]))
```

Als Nächstes wird das PP mit dem Inferenz-Algorithmus Metropolis-Hastings Markov Chain Monte Carlo [La18] ausgewertet, um die gesuchte Posterior-Verteilung darzustellen. Sie ist zweidimensional, da es sich um zwei latente Variablen (s und b), eben die Geradenparameter, handelt. Um die Wirkung des conditioning zu verdeutlichen, wird in Abb.1+2 diese Posterior-Verteilung (bzw. die aus ihr resultierenden Geraden in Abb.1, wenn man die latenten Variablen als Geradenparameter interpretiert) einmal ohne (hierbei wurden alle `observe`-Anweisungen aus dem PP entfernt) und einmal mit conditioning dargestellt. In beiden Fällen ist der Inferenzalgorithmus dann jeweils für sich genommen ausgeführt worden. In Abb.2 fällt dabei sofort auf, dass die Posterior-Verteilung mit conditioning (in rot) in beiden Dimensionen signifikant schmäler ist als die ohne conditioning (in blau). Die Expressivität einer PPL erlaubt es also, gezielte (aber immer probabilistisch notierte) Bedingungen an die latenten Variablen im PP zu formulieren. Die daraus durch den Inferenzalgorithmus realisierte Posterior-Verteilung wird dann mit hoher Wahrscheinlichkeit Stichproben der latenten Variablen generieren können, die die gewünschten Zielvorgaben des conditioning möglichst gut erfüllen. Die Posterior-Verteilung ohne conditioning (in blau) hingegen reproduziert letztlich nur die Apriori-Verteilung der latenten Variablen (im PP die Normalverteilungen für s und b). Dass die Glockenkurven hierbei nur schemenhaft zu erkennen sind, beruht auf der geringen Stichprobenanzahl von 100. Diese geringe Anzahl wurde aus didaktischen Gründen gewählt, damit in Abb.1 einzelne Linien visuell noch unterschieden werden können.

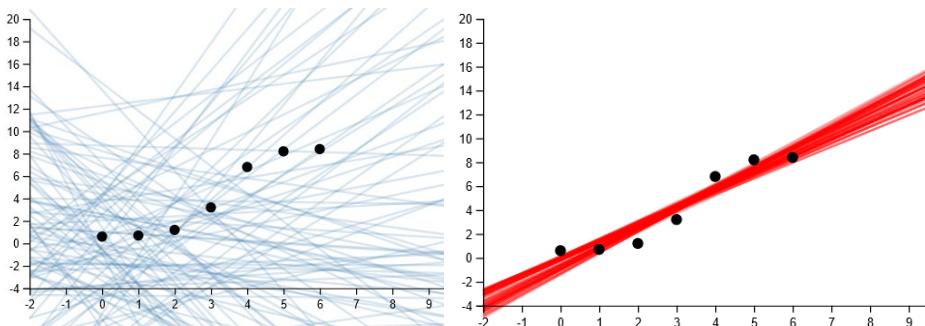


Abb.1: (links) Hier hat ein Probabilistisches Programm (ohne conditioning) 100 Paare $[s \ b]$ der latenten Variablen erzeugt. Als Geradenparameter interpretiert ergeben sie eine regellose Anordnung von Geraden (in blau). (rechts) Das gleiche Programm, nun aber mit conditioning, erzeugt wieder Geradenparameter (in rot). Die zugehörigen Geraden approximieren die Datenpunkte nun aber in nahezu optimaler Weise. Es sei betont: hier wurde keines der üblichen Regressionsverfahren, z.B. die Kleinsten-Quadrate-Methode o.ä., verwendet.

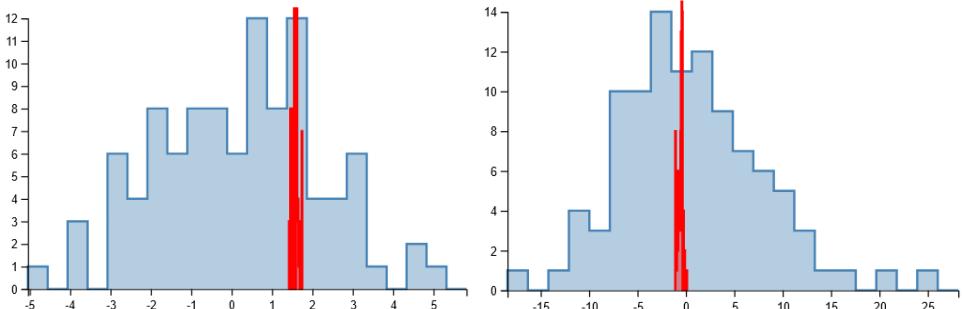


Abb.2: Diese Histogramme zeigen die durch die Inferenz bestimmten Posterior-Verteilungen der latenten Variablen: (links) für die Steigungen s und (rechts) für die Achsenabschnitte b der Geraden. Die breiten Verteilungen (in blau) gehören zum PP ohne conditioning und die extrem schmalen Verteilungen (in rot) gehören zum PP mit conditioning. Diese schmalen Verteilungen der Parameter s und b führen zur gebündelten Ausrichtung der Geraden in Abb.1 (rechts).

Dieses Beispiel verdeutlicht die Methodik und die Merkmale der „Probabilistischen Programmierung“. Folgende allgemeine Vorgehensweise lässt sich skizzieren:

- Wenn eine Situation vorliegt, in der ein Problem gelöst werden soll bzw. an die eine Frage gestellt werden soll, und nur wenige Daten vorhanden sind, es aber möglich ist, ein generatives Modell der Situation zu formulieren, welches hinreichend allgemeingültig ist und durch geeignete latente Variablen auch flexibel parametrisiert werden kann, dann besteht die Chance, „Probabilistische Programmierung“ einzusetzen.
- Ein Probabilistisches Programm (PP) beschreibt das Problem durch ein generatives Modell und eine in diesem Kontext gestellte, spezifische Frage, die durch geeignete Konditionierungen ausgedrückt wird.
- Das Inferenzverfahren arbeitet dann mit dem PP und realisiert die darin verborgene Inversion des konditionierten, generativen Modells, indem sie die zu dieser Inversion passende Verteilung der latenten Variablen des Modells bestimmt. Diese Verteilung ist die Antwort (z.B. in der Form einer MAP) auf die Fragestellung.
- Es gibt sehr viele unterschiedliche Inferenzverfahren mit jeweils eigenen Stärken und Schwächen. Manche dieser Inferenzverfahren sind „Allzweck-Werkzeuge“ mit breit gefächerten Einsatzbereichen. Der Metropolis-Hastings-Algorithmus in der obigen Regressionsaufgabe füllt z.B. die Rolle aus, die üblicherweise das Gauß-verfahren eingenommen hätte. Diese universelle Einsetzbarkeit von Inferenzverfahren ist die Basis für die immer wieder artikulierte Ambition [Go14, Me18] des Forschungsfeldes der „Probabilistischen Programmierung“, nämlich viele Bereiche des maschinellen Lernens auch Nicht-Expertinnen und Nicht-Experten einfacher zugänglich zu machen bzw. sogar zu automatisieren.

- Anwenderinnen und Anwender können agnostisch hinsichtlich der zugrunde-liegenden Techniken und Konzepte der Inferenzverfahren vorgehen und sich ganz auf die konkrete Beschreibung ihres jeweiligen Anwendungsproblems und der zu beantwortenden Fragestellung im PP konzentrieren (Abb.3). Der Programmcode des PP ist daher rein deklarativ und somit kurz.



Abb.3: Fragen an generative Modelle werden als Konditionierungen in Probabilistischen Programmen formuliert. Inferenzverfahren beantworten diese Fragen durch die Berechnung passender Parameterwerte der Modelle in der Form von Wahrscheinlichkeitsverteilungen.

5 Komplexe Anwendungsbeispiele

Vier komplexe, sehr unterschiedliche Anwendungen werden im Folgenden vorgestellt, um die Breite der Anwendungsmöglichkeiten aufzuzeigen. (5.1) Zuerst wird eine Problemstellung aus der angewandten Mechanik untersucht. (5.2) Danach wird ein System vorgestellt, das große Datensätze, die Inkonsistenzen, Fehler und Lücken besitzen, automatisch bereinigt. Für diese ersten beiden Beispiele wurden eigene Experimente durchgeführt. Die eingesetzte Hardware bestand dabei aus einer Workstation mit einem AMD Ryzen 3955WX-Prozessor (4,30 GHz, 16 Kerne) und 256GB RAM. Danach werden zwei weitere repräsentative Beispiele (5.3 + 5.4) erwähnt, die aus der Forschungsliteratur stammen.

5.1 Lösen von Mechanik-Puzzles

Dieses Beispiel wurde zuerst auf der Machine Learning Summer School 2015 (MLSS2015) in Tübingen von F. Wood und Mitarbeitern vorgestellt [Phy15] und für die hier dargestellten Experimente verwendet und für die Auswertung (siehe Abb.6) um diverse Grafikfunktionen erweitert. Die zu lösende Aufgabe besteht darin, eine Anzahl von Bällen, die von einer erhöhten Rampe herunterrollen, in einen Zielbehälter zu befördern (siehe Abb.4). Dazu kann eine bestimmte Anzahl von kurzen, waagerechten Hinderniselementen („bumper“) völlig frei in der gesamten in Abb.4 gezeigten Szene so platziert werden, dass die Bälle auf diese bumper treffen, von diesen abprallen, auf weitere bumper treffen und sich auf diese Weise dem Zielbehälter nähern und mit etwas Glück letztlich in ihm landen.

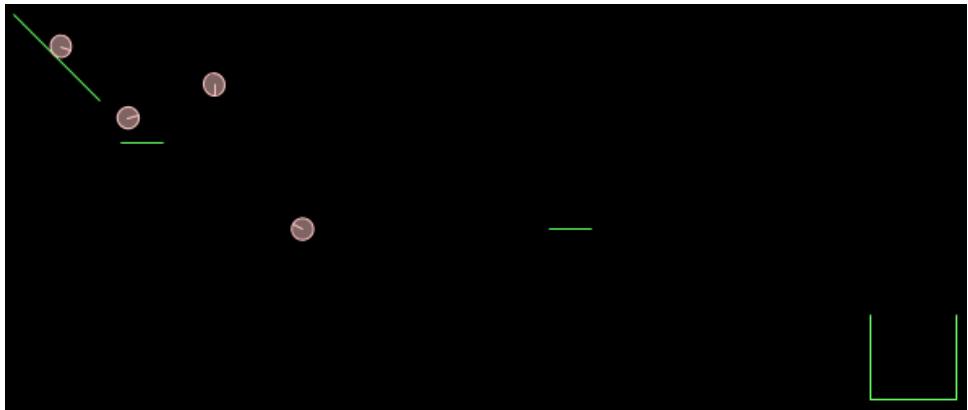


Abb.4: Das ist die Ausgangssituation. Die Bälle rollen von der Rampe oben links herunter, treffen auf den ersten zufällig platzierten bumper, prallen von diesem ab und fallen dann ins Nichts. Wie könnte man zusätzliche bumper so platzieren, dass die Bälle vom bumper zu bumper hüpfen und schließlich im Zielkorb unten rechts landen? Ein Lauf in dieser Simulation besteht immer aus 20 Bällen, danach wird gezählt wie viele von im Zielkorb landen. Dies führt allerdings zu zusätzlicher Komplexität, da die Bälle einer solchen Gruppe bei entsprechenden bumper-Platzierungen auch untereinander kollidieren können. In manchen Situationen kann das sogar hilfreich sein und für die Lösung ausgenutzt werden (siehe dazu Abb.7).

Für das PP (s.u.) ist es notwendig, eine leistungsfähige Physik-Engine [Box18] als externe Bibliothek in das generative Modell `simulate-world` zu integrieren. Dass dies möglich ist, ist eine der besonderen Stärken von Anglican. Eine Variante dieser Engine wird auch in einschlägigen Spielen (z.B. Angry Birds o.ä.) eingesetzt. Mit `create-world` werden die Positionen der Rampe, des Zielkorbs und die Apriori-Verteilung von acht bumper definiert. Dazu werden die bumper in der gesamten in Abb.4 gezeigten Szene gleichförmig (`uniform-continuous`) verteilt und in der Liste (sog. lazy sequence) `bs` gespeichert. Zusätzlich können hier aber auch noch weitere Hindernisse eingebaut werden, die die Aufgabe drastisch erschweren. Ein Lauf besteht jeweils aus 20 Bällen. Mit `balls-in-box` wird beobachtet, wie viele von ihnen dann im Zielkorb landen. Die Konditionierung `observe` besagt, dass 20 Bälle im Korb sein sollen. Eine Stichprobe, die dieses PP erzeugt, besteht aus der aktuellen Trefferzahl `num-balls` und der Liste der aktuellen bumper-Positionen `bs` als latente Variablen des Modells, für die Werte durch den Inferenz-Algorithmus gesucht werden, die die Konditionierung erfüllen sollen.

```
(defquery physics1 []
  (let [n-bumpers 8
    f (fn [] (list (sample (uniform-continuous -5 14))
                  (sample (uniform-continuous 0 10)))))
```

```

bs (repeatedly n-bumpers f)

w0 (create-world bs)

w1 (simulate-world w0)

num-balls (balls-in-box w1) ]

(observe (normal num-balls 1) 20)

(list num-balls bs)))

```

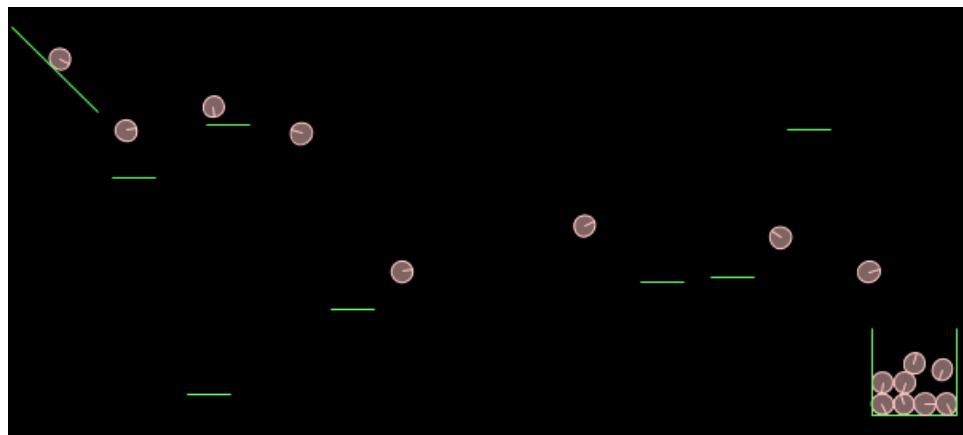


Abb.5: Nach der Inferenz: eine MAP-Stichprobe aus einer bumper-Verteilung, die zum Ziel führt.

Der hier verwendete Metropolis-Hastings Inferenzalgorithmus garantiert im Limes zwar die Konvergenz bis zur vollständigen, expliziten Darstellung der Posterior-Verteilung, aber in einem so komplexen, da hochdimensionalen Problem, kann wirkliche Konvergenz aufgrund des Rechenaufwands kaum je erreicht werden. Insofern dient die Inferenz hier als effizientes Explorationsverfahren (siehe dazu auch Abb.6), durch das relevante Bereiche der Verteilung zumindest approximiert werden können. Wenn dann derartige Bereiche explizit vorliegen, kann aus ihnen, z.B. in der Form einer MAP, eine mögliche Lösung der Aufgabenstellung leicht extrahiert werden (Abb.5 und 7).

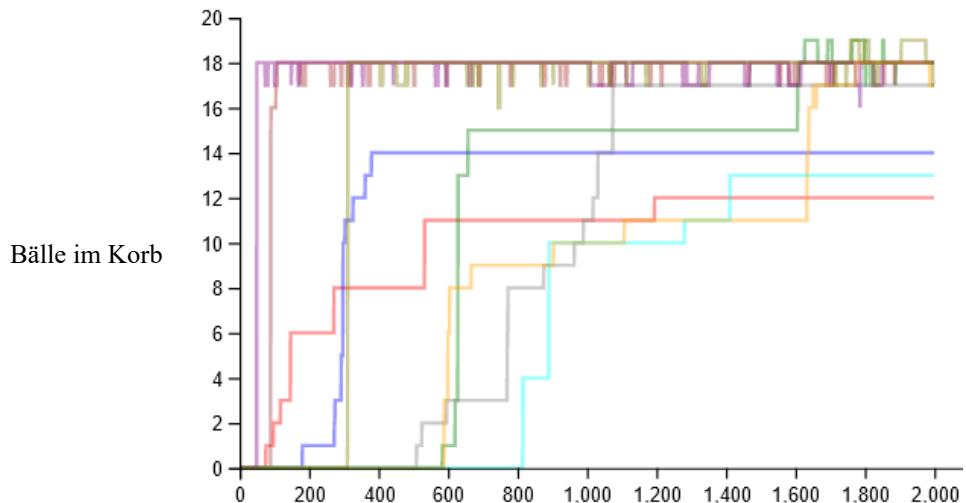


Abb.6: Der Inferenzalgorithmus bei der Arbeit: während der Burn-In-Phase wird die durch das „Probabilistische Programm“ definierte Verteilung stichprobenartig exploriert und dabei eine ihrer Modi gesucht. Ein Modus entspricht dem Teil der Posterior-Verteilung, der mit hoher Wahrscheinlichkeit die Konditionierung erfüllt. Die latenten Variablen des Modells und damit der gesuchten Posterior-Verteilung sind die x-y-Positionen der acht bumper. D.h. diese Verteilung ist eine zerklüftete Berg-und-Tal-Landschaft in einem 16-dimensionalen Raum. Wenn ein solcher Modus, d.h. ein Berg, gefunden wurde, wird er bestiegen. Dieser an einen Gradientenaufstieg erinnernde Vorgang ist typisch für Metropolis-Hastings und wurde hier zehnmal durchgeführt, da die anfängliche Stichprobe der Apriori-Verteilung der bumper-Positionen vorentscheidend dafür ist, welche Mode gefunden wird. Jedes Mal wird also eine neue Lösung gefunden. Im Diagramm wird dies jeweils mit einer bestimmten Farbe dargestellt. In jeder der zehn Durchführungen wurde die Simulation jeweils 2000-mal mit je 20 Bällen pro Simulationslauf ausgeführt. Die einzelnen Simulationsläufe entsprechen der Ordinate. Für jeden Simulationslauf wurde die Anzahl der Bälle im Zielkorb bestimmt und entsprechend auf der Abszisse eingetragen. Man erkennt anhand der stufenförmigen Linienverläufe deutlich, wie die Anzahl der Bälle im Zielkorb beim Fortschreiten der Inferenz jedes Mal sukzessive ansteigt.

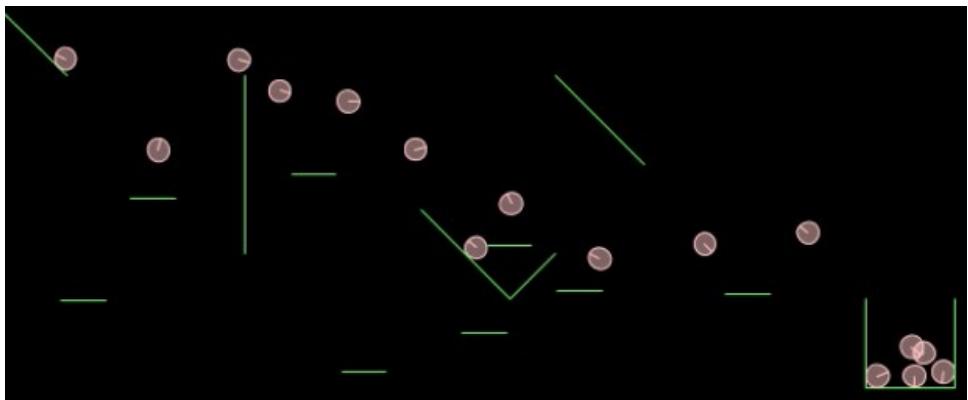


Abb.7: Diese Aufgabe wurde durch zusätzliche Hindernisse erschwert und die Komplexität der Wechselwirkung der Bälle untereinander muss hier ebenfalls berücksichtigt werden. Für die von der Inferenz gefundene Lösung wird eines der Hindernisse (senkrechte Linie im linken Bereich) von den Bällen als zusätzlicher bumper verwendet (blauer Pfeil links). Und der eingeklemmte Ball im Zentrum der Szene (blauer Pfeil in der Mitte) dient nun selbst als bumper für die anderen Bälle.

Von einer höheren Warte aus betrachtet, kann man sagen: die Inferenz invertiert das generative Modell in einem lokalen Gebiet seines Parameterraums gemäß den Vorgaben des „Probabilistischen Programms“ – d.h. sie findet qua expliziter Darstellung der Posterior-Verteilung Werte für diese Parameter, die zum conditioning des PP passen.

5.2 Datensäuberung mit PClean

Inkonsistente und fehlerhafte Daten sowie Datenlücken sind in allen Anwendungsbereichen von Wirtschaft, Verwaltung und Wissenschaft geradezu omnipräsent. Datensäuberung ist sehr aufwendig. Daher besteht hier immenser Automatisierungsbedarf. Rein regel-basierte Verfahren haben sich als untauglich erwiesen. Kommerzielle Systeme kombinieren verschiedene Verfahren des maschinellen Lernens mit adaptierbaren Regel-systemen. PClean [Le21] ist ein neues, auf Prinzipien der „Probabilistischen Programmierung“ basierendes System, das in einschlägigen Standard-Benchmarks mit gängigen, kommerziellen Systemen konkurriert kann bzw. sogar leichte Vorteile aufweist und bei Datensätzen mit anspruchsvoller Fehlerarten oder bei der Skalierung auf sehr große Datensätze (Tabellen mit 2 Mio. Zeilen) klar überlegen ist. PClean ist in Julia [Jul22] als Open Source Projekt implementiert [Pcl22]. Es definiert eine eigene DSL (domain specific language) zur Erstellung „Probabilistischer Programme“ zur Modellierung von Daten-tabellen und enthält spezielle Inferenzalgorithmen, die auf diese DSL abgestimmt sind, um die besonderen Skalierungsanforderungen zu erfüllen.

In einem mit PClean geschriebenen „Probabilistischen Programm“ kann mit Hilfe der DSL ein Vorschlag für die relevanten Klassen formuliert werden, aus denen sich der Gesamtdatensatz zusammensetzt. In diesen Klassen werden dann die Apriori-Verteilungen für die Daten angegeben und auch mögliche Fehlerquellen für diese Daten werden mit Hilfe der DSL spezifiziert, z.B. Vertauschungen von Buchstaben, falsche Einheiten bei numerischen Größen, Verwechslung von Groß- und Kleinschreibung, Falschschreibung von Namen, etc. Die Konditionierung auf die Daten geschieht durch den zu bereinigenden Gesamtdatensatz D selbst - also die große, fehlerbehaftete Tabelle, die gesäubert werden soll. Der spezielle Inferenzalgorithmus von PClean berechnet dann aus dem PP eine Posterior-Verteilung von relationalen Datenbanken, deren interne Daten nun aber gereinigt sind. Aus einer Stichprobe der Endphase dieser Inferenzberechnung kann dann wieder eine Gesamttafel D' rekonstruiert werden, die gereinigt ist. Fassen wir zusammen, was hier passiert: die Inferenz bestimmt für die Parameter des auf den verunreinigten Datensatz D konditionierten generativen Modells M eine geeignete Posterior-Verteilung über D'. Das entspricht dem Schluss von der Wirkung auf die Ursache. Wenn man das in der anderen Richtung betrachtet, also von der Ursache auf die Wirkung, folgt daraus: das generative Modell mit genau diesen Parametern D' würde somit wieder den ursprünglichen, also verunreinigten Datensatz D hervorbringen – formal ausgedrückt: $M(D') = D$. Die durch die Inferenz gefundenen Parameter des generativen Modells repräsentieren aber das, was wir suchen, nämlich den gereinigten Datensatz D' bzw. eine relationale Datenbank, in der dieser Datensatz enthalten ist. In diesem Sinne invertiert die Inferenz also das generative Modell M.

Um PClean auszuprobieren, wird nun ein synthetischer Datensatz einer Länge von 50.000 Zeilen mit Informationen zu Mietwohnungen verwendet, der viele verunreinigte Daten enthält. Dieser Datensatz wird via [Pcl22] zur Verfügung gestellt und in [Le21] detailliert erläutert. Er besteht aus vier Spalten: Zimmertyp, Mietpreis, Stadt bzw. Region, Bundesstaat. Folgende Fehler kommen vor: 1.) Namen von Städten/Regionen sind zu 2% falsch geschrieben. Viele unterschiedliche Korrekturvarianten sind i.d.R. möglich. 2.) Es fehlen 10% der Angaben zu Bundesstaaten. In den Bundesstaaten kommen häufig auch gleichlautende Städte bzw. Regionen vor, so dass die Zuordnung nicht eindeutig ist. 3.) Preisangaben sind zu 1% in einem falschen Zahlenformat bzw. in einer falschen Einheit angegeben. 4.) Der Wohnungstyp mit der Zimmeranzahl fehlt in 10% aller Fälle.

Die Datenreinigung dauerte mit der oben spezifizierten Hardware ca. 33 Sekunden, wobei ca. 27% dieser Zeit auf den Garbagekollektor und die Kompilation entfallen sind und während der Berechnung ca. 12,6 GB Speicherplatz benötigt wurde. Es steht ein korrekter Datensatz („ground truth“) zur Verfügung [Pcl22], der mit dem bereinigten Datensatz verglichen wurde, um die Leistungsfähigkeit von PClean zu messen: die Erfolgsquote (Verhältnis der Anzahl korrekter Reparaturen zur gesamten Fehlerzahl) betrug 0,69 und die Güte (Verhältnis der Anzahl korrekter Reparaturen zur Anzahl aller Reparaturversuche) betrug ebenfalls 0,69. Genau diese Werte wurden auch in [Le21] genannt. Die Autoren behaupten außerdem, dass PClean in weitergehenden Vergleichstests bzgl. dieses Datensatzes kommerziellen Systemen klar überlegen sei und belegen dies durch entsprechende Ergebnisse. Sie berichten darüber hinaus von der Bereinigung eines über 2 Mio. Zeilen

umfassenden Datensatzes aus dem medizinischen Bereich. Sie führen aus, dass ohne einen entsprechenden „ground truth“-Datensatz die Erfolgsmessung der Bereinigung aber natürlich nur stichprobenartig erfolgen könne. Hierbei hätte sich eine Erfolgsquote von 90% ergeben. Die Autoren betonen ausdrücklich, dass die Behandlung eines derart großer Datensatzes völlig außerhalb der Möglichkeiten kommerzieller Systeme läge.

5.3 Captcha Breaking

Es gibt mindestens drei Wege, eine Captcha-Hürde zu überwinden: (1) Man hat sehr viele, korrekt gekennzeichnete Captcha-Daten und trainiert damit ein Neuronales Netz. Dafür gibt es im Internet bereits eine Vielzahl von Do-it-yourself-Bastelanleitungen und Apps. Oder (2) man setzt ausgeklügelte Methoden aus der Bildverarbeitung und Muster-erkennung ein [MM03]. Oder (3) man verwendet „Probabilistische Programmierung“ wie in [Ma13]: darin wird mit Hilfe der universellen PPL Venture [Ven22] ein PP geschrieben, das einen Captcha-Generator als generatives Modell enthält. Durch ihn werden die Elemente einer Zeichenfolge zufällig angeordnet, skaliert und verrauscht. Die zugrunde-liegenden Zeichen sind die latenten Variablen des generativen Modells. Nun wird ein reales Captcha-Bild, dessen zugrundeliegenden Zeichen aber nicht gegeben sind, als Kon-ditionierung des PP verwendet. Durch Inferenz mittels Metropolis-Hastings MCMC wird das generative Modell (also der Captcha-Generator) invertiert, so dass Abschätzungen für die Werte der latenten Variablen möglich werden, d.h. die zugrundeliegenden Zeichen werden erkannt. Die Autoren [Ma13] berichten hier von einer Erkennungsrate von 70%.

5.4 Strategieplanung in dynamischen Modellen

Zur Entwicklung geeigneter Strategien zur Steuerung des dynamischen Ablaufs des Pandemie-Geschehens (COVID-19) wurden existierende epidemiologische Simulatoren in ein PP integriert [Wo22]. Als PPL wurde PyProb [Pyp22] verwendet, das auch bereits im CERN (Elementarteilchenphysik) eingesetzt wird. Dadurch dass sich Planungs-aufgaben als Inferenz darstellen lassen, konnten Maßnahmen zur Pandemiebekämpfung so aufeinander abgestimmt werden, dass sich die sozio-ökonomischen Kosten, die Belastung des Gesundheitssystems, Todesraten, etc. im Rahmen des Simulationsmodells minimieren ließen und auch Ausgangssperren konnten so ganz verhindert werden.

Zusammenfassung und Ausblick

Die zentrale Idee der „Probabilistischen Programmierung“ besteht in der auf ausgewählte Daten konditionierten Inversion eines generativen Modells des Anwendungsproblems, um geeignete Werte für die latenten Variablen dieses Modells zu bestimmen. Dafür reichen i.d.R. vergleichsweise geringe Datenmengen aus. Kurz gesagt: die Konditionierung ist eine Frage an das Modell, die dann von der Inferenz beantwortet wird. Wenn nur wenige Daten zur Verfügung stehen, aber ein generatives Modell für die Aufgabenstellung verfügbar ist, kann „Probabilistische Programmierung“ ein Weg sein, um Probleme anzu-gehen, die mit etablierten Methoden des Machine Learning nur schwer zu behandeln sind.

Das ambitionierte Fernziel des Forschungsfeldes der „Probabilistischen Programmierung“ ist die „automatisierte“ Lösung von Aufgabenstellungen des Machine Learning durch universelle Inferenzverfahren. Nicht-Experten sollen das jeweilige Problem dann nur noch deklarativ als „Probabilistisches Programm“ formulieren müssen. Den Rest übernimmt die Inferenz. Für dieses Fernziel wird die Verfügbarkeit leistungsfähigerer Inferenz-verfahren entscheidend sein. Als ein möglicher Weg wird dazu u.a. „Deep Probabilistic Programming“ [Pyr17, Pyp22, SWF16] propagiert. Besonders interessant ist auch die Möglichkeit der Integration einer externen Simulationsanwendung in ein Probabilistisches Programm (siehe z.B. 5.1). Die Simulationsanwendung übernimmt darin dann die Rolle des generativen Modells [Ang18, Pyp22].

In der Kognitionsforschung wird „Probabilistische Programmierung“ zur Analyse und Modellierung von Lern- und Entscheidungsprozessen sowie von Sozialverhalten verwendet [Te11, Go16, Ev22, Go17, St22]. Und auch die Einsatzmöglichkeiten für die Wirtschaftsinformatik i.A. sind breit gestreut: z.B. im Business Process Reengineering, in der modellbasierten Planung und Steuerung komplexer Prozesse und in der Risiko-Modellierung, in Business Analytics, in IT-Sicherheit und in Quantitative Finance.

Literaturverzeichnis

- [Ang18] Anglican, probprog.github.io/anglican, Stand: 24.3.2022
- [Bi13] Bishop C.M.: Model-based machine Learning, Phil Trans R Soc A 371:20120222, Royal Society Publ., 2013
- [Box18] Clojure wrapper für die Java Physik-Engine jbox2d, github.com/floybix/cljbox2d, 2018, Stand: 24.3.2022
- [Clo22] The Clojure Programming Language, clojure.org, Stand: 24.3.2022
- [Ev22] Evans, O.; Stuhlmüller, A.; Salvatier, J.; Filan, D.: *Modeling Agents with Probabilistic Programs*, agentmodels.org, Stand: 24.3.2022
- [Gn08] Goodman, Noah; Mansinghka, Vikash; Roy, D.M.; Bonawitz, K.; Tenenbaum, J.B.: Church: A Language for Generative Models. In: Proc. of the 24th Annual Conference on Uncertainty in Artificial Intelligence (UAI 2008), AUAI Press, 2008
- [Go16] Goodman, N.D.; Tenenbaum, J.B., Buchsbaum,D.; Hartshorne, J.; Hawkins, R.; O'Donnell, T.J.; Tessler, M.H.: *Probabilistic Models of Cognition*, probmods.org, 2016, Stand: 24.3.2022
- [Go17] Gopnik, A.: Making AI more Human. In: Scientific American, Juni 2017
- [Gro86] Grosof, B.N.: Non-Monotonicity in Probabilistic Reasoning, In: Proceedings of the Second Conference on Uncertainty in Artificial Intelligence (UAI1986), 1986
- [GS22] Goodman, N.D.; Stuhlmüller, A.: The Design and Implementation of Probabilistic Programming Languages, dippl.org, Stand: 24.3.2022
- [Ja03] Jaynes, E.T.: Probability Theory: The Logic of Science, Cambridge Univ. Press, 2003
- [Jul22] The Julia Programming Language, julialang.org, Stand: 24.3.2022
- [KF09] Koller, D.; Friedman, N.: Probabilistic Graphical Models, MIT Press, 2009
- [La18] Lambert, B: A Student's Guide to Bayesian Statistics, SAGE Publications Ltd., 2018
- [Le21] Lew, A.K.; Agrawal, M.; Sontag, D.; Mansinghka, V.K.: PClean: Bayesian Data Cleaning with Domain-Specific Probabilistic Programming. In: Proc. of the 24th Int. Conf. on AI and Statistics (AISTATS 2021), PMLR, 2021
- [Ma13] Mansinghka, V.K.; Kulkarni, T.D.; Perov, Y.N.; Tenenbaum, J.B.: Approximate Bayesian Image Interpretation using Generative Probabilistic Graphics Programs. In: Adv. in Neural Information Processing Systems 26 (NIPS 2013), Curran Ass., Inc., 2013
- [Mc20] McElreath, R.: Statistical Rethinking (2nd ed.), CRC Press – Taylor&Francis Group – Chapman & Hall Book, 2020
- [Me18] van de Meent, J.-W.; Paige, B; Yang, H; Wood, F.: An Introduction to Probabilistic Programming, arxiv.org/abs/1809.10756, Stand: 24.3.2022
- [MM03] Mori, G; Malik, Jitendra. Recognizing Objects in Adversarial Clutter: Breaking a Visual CAPTCHA. In: Proc. of the Int. Conf. on Comp. Vis. and Pattern Recog., IEEE, 2003
- [NJ01] Ng, A.; Jordan, M.I.: On Discriminative vs. Generative classifiers: A comparison of logistic regression and naïve Bayes. In: Adv. in Neural Inf. Proc. Systems 14 (NIPS 01), MIT Press, 2002

- [Pcl22] PClean: A Domain-Specific Probabilistic Programming Language for Bayesian Data Cleaning, github.com/probcomp/PClean, Stand: 24.3.2022
- [Pe88] Pearl, J.: Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, 1988
- [Phy15] Probabilistic Programming Repositories - MLSS2015 (practical materials for MLSS Tübingen) - physics, bitbucket.org/probprog/, Stand: 24.3.2022
- [Pyp22] PyProb, github.com/pyprob/pyprob, Stand: 24.3.2022
- [Pyr17] Uber AI Labs Open Sources Pyro, eng.uber.com/pyro, Stand: 24.3.2022
- [RP02] Ramsey N.; Pfeffer, A.: Stochastic lambda calculus and monads of probability distributions. In: Proc. of the 29th ACM-SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL 2002), ACM, 2002
- [St22] Stuhlmüller, A.: Forest - repo. for generative models, forestdb.org, Stand: 24.3.2022
- [SWF16] Salvatier, J.; Wiecki, T.V.; Fonnesbeck, C.: Probabilistic programming in Python using PyMC3. PeerJ Computer Science 2:e55, 2016
- [Te11] Tenenbaum, J.B.; Kemp, C.; Griffiths, T.L.; Goodman, N.D.: How to Grow a Mind: Statistics, Structure, and Abstraction. In: Science, Vol. 331, 2011
- [To16] Tolpin, D.; van de Meent, J.-W.; Yang, H.; Wood, F.: Design and Impl. of Probabilistic Programming Language Anglican, In: Proc. of the 28th Symp. on the Impl. and Appl. of Functional Progr. Lang. (IFL 2016), ACM, 2016
- [Ven22] Venture, probcomp.csail.mit.edu/software/venture, Stand: 24.3.2022
- [Wi19] Winn, J. (et al.): Model-based Machine Learning, mbmlbook.com, Stand: 24.3.2022
- [WMM14] Wood, F.; van de Meent, J.-W.; Mansinghka, Vikash: A New Approach to Probabilistic Programming Inference. In: Proc. of the 17th Int. Conf. on Artificial Intelligence and Statistics (AISTATS 2014), Vol. 33 of Proc. of PMLR, 2014
- [Wo22] Wood, F.; Warrington, A.; Naderiparizi, S.; Weilbach, C.; Masrani, V.; Harvey, W.; Scibior, B.; Grefenstette, J.; Campbell, D.; Nasseri, A.: Planning as Inference in Epidemiological Dynamic Models. In: Frontiers in Artificial Intelligence (31. 3.2022), www.frontiersin.org

Visualisierung hochdimensionaler Daten: State-of-the-Art von nichtlinearen Methoden zur Dimensionsreduktion

Tobias Schmieg, Helmut Beckmann¹

Abstract: In vielen Forschungsdisziplinen ist die Visualisierung von hochdimensionalen Daten ein wichtiger Schritt der Datenanalyse. Eine Möglichkeit solche Daten zu visualisieren ist durch nichtlineare Methoden zur Dimensionsreduktion. In dieser Arbeit wird durch eine umfangreiche Literaturanalyse der State-of-the-Art von Methoden zur nichtlinearen Dimensionsreduktion ermittelt. Durch 51 untersuchten Publikationen werden 21 verschiedene Methoden identifiziert, wobei acht dieser Methoden näher betrachtet werden. Die älteren sechs dieser acht Methoden können zwar gut die lokale Struktur von hoher Dimensionalität in zwei oder drei Dimensionen darstellen, doch nur t-SNE (t-Distributed Stochastic Neighbor Embedding) und UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction) können sowohl die globale als auch die lokale Struktur gleichzeitig beibehalten. Diese acht näher betrachteten Methoden stellen den aktuellen State-of-the-Art dar.

Keywords: nichtlineare Dimensionsreduktion, hochdimensionale Daten, State-of-the-Art, t-SNE, UMAP

1 Einleitung

Eine der vielfältigsten Techniken zur Visualisierung von Daten sind Scatterplots oder auch Streudiagramme genannt. Hierdurch können paarweise Korrelationen zwischen Dimensionen visualisiert werden. Diese Visualisierungstechnik kann, ebenfalls wie andere Techniken, höchstens dreidimensionale Daten darstellen. Durch eine paarweise Darstellung können zwar noch Korrelationen in Datensätzen mit mehr als drei Dimensionen untersucht werden. Jedoch werden dadurch auch nur die Zusammenhänge zwischen Dimensionen sichtbar, welche auch direkt in einer Visualisierung verglichen werden. Zusammenhänge über mehrere Dimensionen hinweg sind nur schwer zu erkennen [Ng20].

Jedoch ergeben sich immer mehr Anwendungsfälle, bei denen die Visualisierung hochdimensionaler Daten eine essenzielle Rolle spielt. Verschiedene Disziplinen, sowohl wissenschaftlich als auch in der Wirtschaft, nutzen Visualisierungen hochdimensionaler Daten,

¹ Hochschule Heilbronn, Studiengang Wirtschaftsinformatik, Max-Planck-Str. 39, 74081 Heilbronn,
{tschmieg@stud.; helmut.beckmann@hs-heilbronn.de}

um erste Erkenntnisse aus diesen zu ziehen. Diese sind unter anderem der medizinische Bereich, die Biowissenschaft oder auch die Analyse sozialer Netzwerke [Ha21]. So kommen Zellkerne, welche für die Brustkrebsforschung relevant sind, auf ca. 30 Variablen. Bei der Bildverarbeitung oder der Umgang mit Wort-Vektoren für Dokumentenverarbeitung kann die Anzahl der Dimensionen vierstellig oder mehr werden [vH08]. Daraus ergibt sich das Problem, dass nur eine gewisse Anzahl an unabhängigen Dimensionen gleichzeitig dargestellt werden kann [GS93].

Eine Lösung hierfür sind Methoden zur Dimensionsreduktion aus dem Fachgebiet des Machine Learning. Diese lassen sich in zwei Kategorien einordnen. Zum einen in lineare Methoden, welche versuchen, die Distanzen zwischen den Datenpunkten zu erhalten. Die wohl bekannteste Methode hierfür ist die Hauptkomponentenanalyse (Englisch: Principal Component Analysis (PCA)). Auf der anderen Seite gibt es noch die nichtlinearen Methoden. Diese versuchen durch nichtlineare Distanzen oder durch lokale Abstände die mehrdimensionalen Daten in einer niedrigeren Dimension darzustellen [Ho33]. Jedoch unterscheiden sich hier die Ansätze stark im Vorgehen und in dem genauen Ziel der Dimensionsreduktion. Die Anwendungsgebiete sind interdisziplinär und nicht auf Machine Learning beschränkt. Daraus ergibt sich folgende Forschungsfrage:

Welche Methoden sind State-of-the-Art für nichtlineare Dimensionsreduktion von hochdimensionalen Daten im Hinblick auf Datenvizualisierung?

Ziel dieser Forschungsarbeit soll es sein, den State-of-the-Art zu nichtlinearen Dimensionsreduktionsmethoden (NLDR-Methoden) darzustellen. Zur Beantwortung dieser Forschungsfrage werden in dieser Forschungsarbeit Publikationen untersucht, welche NLDR-Methoden anwenden, um das eigene Forschungsziel zu erreichen, oder Publikationen, welche NLDR-Methoden näher betrachten bzw. sogar eigene Ansätze vorstellen. Hierfür werden die im Rahmen einer umfangreichen Literaturrecherche identifizierten Publikationen auf die Nennung und Anwendung bzw. nähere Betrachtung von NLIDR-Methoden untersucht. Eine quantifizierte Darstellung der Untersuchungsergebnisse soll die Verbreitung von NLDR-Methoden sowie die Entwicklung des Forschungsgebietes über den betrachteten Zeitraum aufzeigen. Zur Erreichung dieser Zielsetzung werden nachfolgend die nötigen theoretischen Grundlagen gelegt. Aufbauend wird die Methodik sowie das Vorgehen zur Quantifizierung der Vorkommnisse genauer erläutert, welche im darauffolgenden Kapitel auf Basis des zuvor definierten Schemas analysiert wird, um Aussagen über die identifizierten NLDR-Methoden abzuleiten. Eine entsprechende Ergebnisdarstellung stellt die Erkenntnisse dieser Forschung am Ende dieser Forschungsarbeit zusammen. Hieraus resultierende Besonderheiten werden dediziert vorgestellt und bieten die Grundlage für weitergehende Forschung.

2 Dimensionsreduktion

Um einen Überblick über die theoretischen Grundlagen zu erhalten, wird der Begriff Dimensionsreduktion explizit dargestellt und erläutert. Die Nutzung von Dimensionsreduktion hat viele Anwendungsgebiete, darunter die Visualisierung hochdimensionaler Daten. Über die letzten Jahrzehnte kamen viele Techniken zur Visualisierung hochdimensionaler Daten auf. Doch viele dieser versuchten, lediglich mehr als zwei Dimensionen darzustellen. Die Interpretation wurde dem Nutzer der Visualisierung überlassen, was den eigentlichen Mehrwert in Frage stellt [vH08]. Deshalb besteht eine hohe Relevanz der Dimensionsreduktion für die Visualisierung hochdimensionaler Daten. Methoden hierfür lassen sich unter anderem in lineare und nicht-lineare Methoden einteilen [LV07]. Bei PCA als vermutlich bekanntestes Beispiel der linearen Dimensionsreduktion, wird durch die Linearkombination von Variablen eines Datensatzes versucht möglichst viel Varianz in wenigen Dimensionen zu erhalten [Hu20] [Ha21]. Nichtlineare Methoden nutzen zu diesem Zweck nichtlineare Distanzmetriken oder die lokale Struktur zwischen Datenpunkten. So versucht Locally-Linear Embedding (LLE) [RS00] die Nachbardatenpunkte des hochdimensionalen Raumes ebenfalls als Nachbarn im niedrigdimensionalen Raum darzustellen. Andere Methoden wie Stochastic Neighbor Embedding (SNE) und darauf aufbauende Methoden, nutzen Wahrscheinlichkeiten, um die Abstände zwischen Datenpunkten darzustellen [Ha21]. Auch die Methoden t-SNE (t-Distributed Stochastic Neighbor Embedding) (welche eine der auf SNE aufbauende Methoden ist) und UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction) nutzen für die Berechnung von Abständen Wahrscheinlichkeiten. Jedoch liegt bei diesen Methoden nicht nur die lokale Struktur von Datenpunkten im Fokus, sondern auch die globale Struktur des Datensatzes [vH08] [MHM18].

3 Methodik zur Literaturanalyse

Für das Vorgehen zur Beantwortung der Forschungsfrage wird eine umfangreiche Literaturanalyse gewählt. Nach Fettke [Fe06] wird eine fünfstufige Literaturanalyse durchgeführt, welche durch die Vorwärts- und Rückwärtssuche und das Erstellen einer Konzeptmatrix nach Webster und Watson [WW02] ergänzt wird.

Im Verlaufe dieser Forschungsarbeit werden die NLDR-Methoden herausgearbeitet. Dabei wird unterschieden, ob eine Methode nur genannt oder näher betrachtet wird. Betrachtet bedeutet dabei, dass die Funktionsweise oder Vor- und Nachteile der Methode beleuchtet werden oder die Methode angewendet wird. Außerdem wird bei der Untersuchung der Primärliteratur einer Methode, die Klassifizierung jener Methode in der Analyse nicht mit einbezogen, um eine dadurch entstehende Unschärfe zu vermeiden. Dies führt dazu, dass in der Konzeptmatrix (Abb. 1) pro Methode eine Klassifizierung Betrachtet häufiger vorzufinden ist, als in den folgenden Analysen. Die genaue Klassifizierung findet unter der Einschätzung der Autoren in Einbezug der oben genannten Punkte statt. Um die relevanten Methoden im Voraus zu filtern, werden Methoden, welche nur ein Vorkommnis aufweisen, nicht weiterverfolgt. Anschließend werden die Ergebnisse der Literaturanalyse in Form einer Konzeptmatrix untersucht, um die Forschungsfrage zu beantworten und den State-of-the-Art darzustellen.

4 Ergebnisse

Folgend werden die Ergebnisse der Literaturanalyse dargestellt. Mithilfe des Python-Visualisierungsframework matplotlib wird die Konzeptmatrix quantitativ aufbereitet.

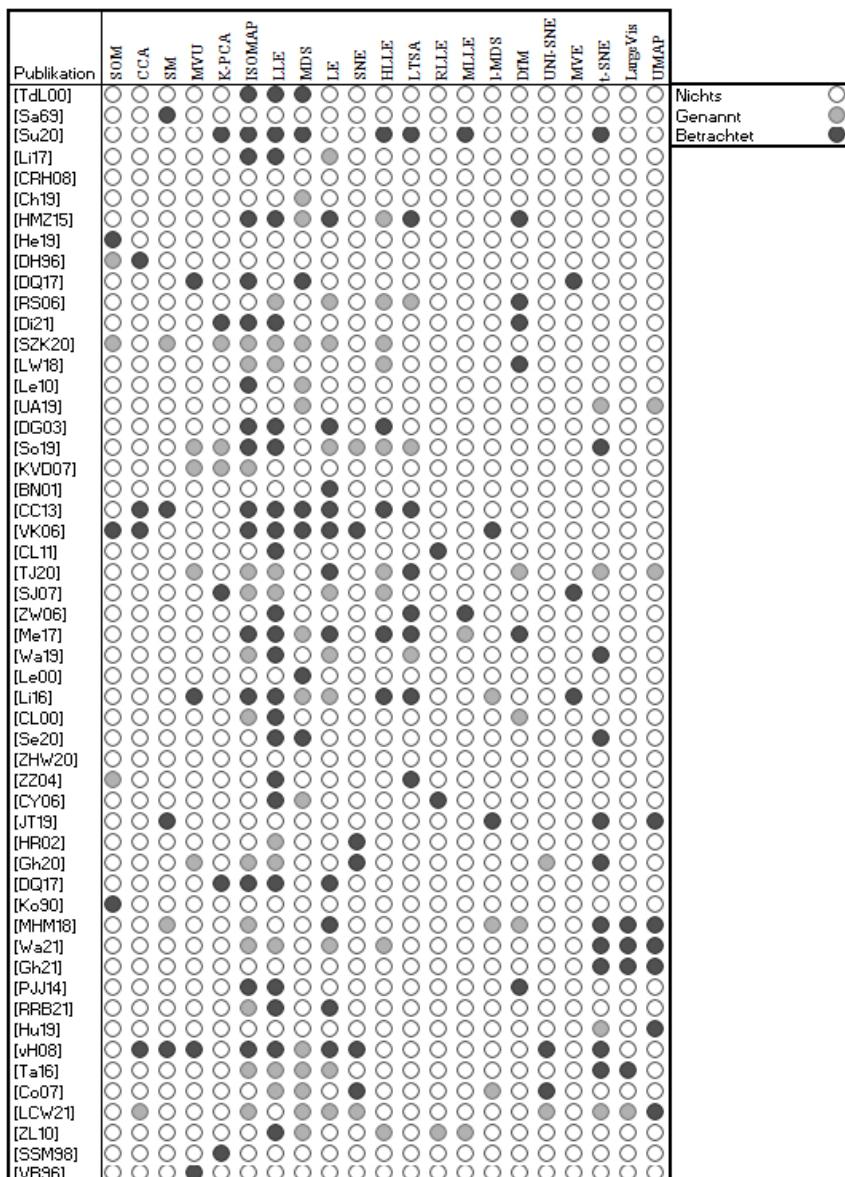


Abb. 1: Konzeptmatrix (eigene Darstellung)

Abb. 1 stellt die Konzeptmatrix dar, welche den Kern der Ergebnisse dieser Forschungsarbeit widerspiegelt. Zur Visualisierung der Vorkommnisse wird eine abgewandelte Form von Harvey-Balls verwendet. Die dreistufige Klassifizierung der Vorkommnisse innerhalb einer Publikation wird durch eine weiße Kugel für keine Nennung (Nichts), grau für Nennung (Nennung) und schwarz für Betrachtung (Betrachtet) wiedergegeben. Insgesamt werden 21 Methoden für die weitere Analyse miteinbezogen.

4.1 NLDR-Methoden

In Abbildung 2 werden die identifizierten NLDR-Methoden nach dem Veröffentlichungsjahr der dazugehörigen Primärquelle dargestellt. Im Durchschnitt hat eine NLDR-Methode ca. zehn Vorkommnisse. Auffälligkeiten zwischen den verschiedenen Methoden im Zusammenhang mit der Klassifizierung von genannt und betrachtet kann nicht festgestellt werden. Die Self-Organizing Map (SOM) [Ko90] ist mit dem Veröffentlichungsjahr 1990 die älteste und UMAP [MHM18] mit 2018 die jüngste NLDR-Methode. Trotz des Zeitraums von 28 Jahren machen die Veröffentlichungen im Zeitraum von 2000 bis 2004 (Isometric Mapping (ISOMAP) [TdL00], LLE [RS00], Multidimensional Scaling (MDS) [Le00], Laplacian Eigenmaps (LE) [BN01], SNE [HR02], Hessian LLE (HLLE) [DG03] und Local Tangent Space Alignment (LTSA) [ZZ04]) 63 % der Vorkommnisse aus.

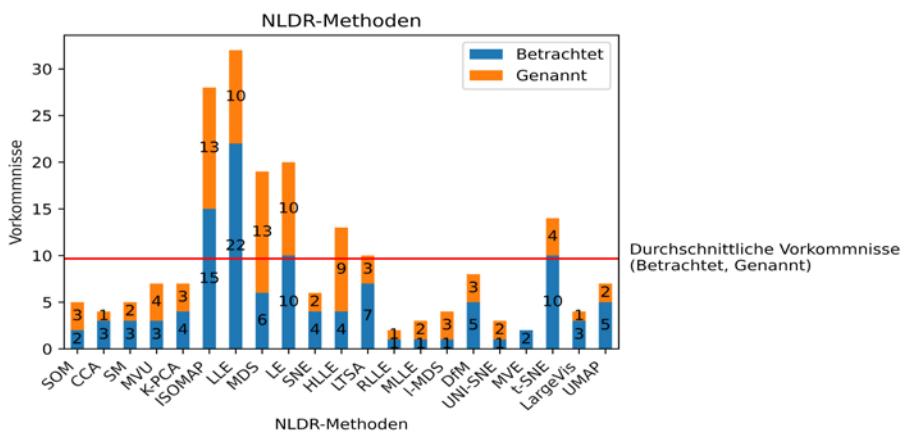


Abb. 2: Vorkommnisse der NLDR-Methoden (eigene Darstellung)

Zwar wurden einige Publikation der Datengrundlage in den Jahren direkt nach dem eben genannten Zeitraum veröffentlicht, trotzdem erschien ein Großteil ab 2017, wie in Abbildung 3 kenntlich wird.

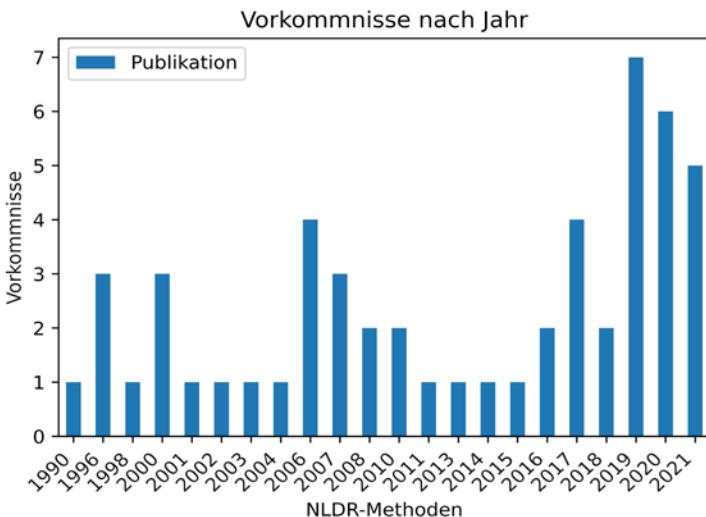


Abb. 3: Publikationen nach Veröffentlichungsjahr (eigene Darstellung)

Die nähere Betrachtung von NLDR-Methoden beschränkt sich deshalb im Folgenden auf ISOMAP, LLE, MDS, LE, HLLE, LTSA, t-SNE und UMAP. Bis auf UMAP weisen diese Methoden mindestens zehn Vorkommnisse auf. Da UMAP die jüngste Methode ist, kann für diese trotz lediglich sieben Vorkommnissen genügend Relevanz zur weiteren Betrachtung vorausgesetzt werden.

4.2 Nähere Betrachtung

Die acht NLDR-Methoden, welche für die weitere Betrachtung ausgewählt werden, machen 70 % der Vorkommnisse aus. Diese werden in Abbildung 4 dargestellt. In neueren Publikationen, wie die Primärliteratur zu t-SNE oder UMAP werden unter anderem ISOMAP und LE angewendet und entsprechend mit t-SNE und UMAP verglichen. Hier zeigt sich, dass ISOMAP, LE und weitere exemplarische Methoden nur schwer sowohl die lokale als auch die globale Struktur von Datensätzen gleichzeitig wiedergeben können. Weiterhin performen zwar alle angewendeten Methoden relativ gut auf künstlichen Datensätzen, jedoch bei realen Datensätzen wie MNIST zeigt sich die Schwäche der älteren Methoden [vH08], [MHM18]. Auch sind zu große Datensätze zum Teil ein limitierender Faktor für die älteren Methoden [MHM18]. Da sich LLE ebenfalls auf die direkten Nachbarn eines Datenpunktes fokussiert, hat diese Methode ähnliche Nachteile [RS00]. Zwar bietet HLLE Vorteile gegenüber LLE und anderen vorherigen NLDR-Methoden, trotzdem gilt es als robustere Variante der LEE Methode [DG03]. LTSA wendet Tangentenräume in der Nachbarschaft eines Datenpunktes an, um die lokale Geometrie darzustellen. Durch diese Tan-

gentenräume wird anschließend das globale Koordinatensystem konstruiert [ZZ04]. Doch auch hier liegt der Fokus auf den lokalen Nachbarschaften der Datenpunkte. Aussagen zu MDS sind schwer zu treffen, da das klassische MDS zwar eine lineare Methode ist, diese aber auch oft als Synonym für nichtlineare Abwandlungen wie beispielsweise local MDS genutzt wird.

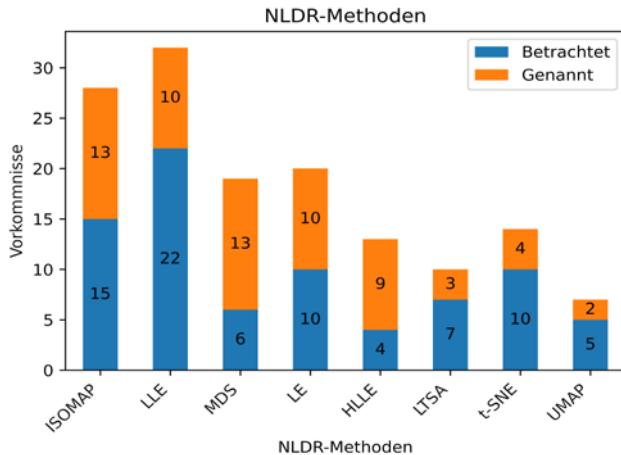


Abb. 4: Vorkommnisse der näher betrachteten NLDR-Methoden (eigene Darstellung)

In den Primärquellen zu t-SNE und UMAP wird bereits ein anderer Anspruch an die Visualisierungen, als die vorher diskutierten NLDR-Methoden deutlich. Während die globale Struktur der Datenpunkte in einer Visualisierung der Methoden ISOMAP, LLE, LE, HLLE und LTSA höchstens nebensächlich erwähnt wird, so wird dies bei t-SNE und UMAP neben der lokalen Struktur als klare Anforderung gestellt [vH08], [MHM18]. T-SNE basiert auf SNE, welches ebenfalls sechs Vorkommnisse aufzuweisen hat. Sowohl t-SNE als auch SNE geben die euklidischen Distanzen durch Wahrscheinlichkeiten wieder, um die Ähnlichkeiten zwischen Datenpunkten im niedrigdimensionalen Raum darzustellen. Jedoch nutzt t-SNE hierfür die studentische t-Verteilung, anstatt die Gauß-Verteilung. Außerdem wird für t-SNE eine effizientere Kostenfunktion angewendet, welche auch bei der Abwandlung Symmetric SNE genutzt wird. UMAP funktioniert im Kern relativ ähnlich zu t-SNE, jedoch konstruiert UMAP im hochdimensionalen Raum eine Repräsentation, welche genutzt wird, um die Repräsentation im niedrigdimensionalen Raum zu optimieren [MHM18].

4.3 Fazit

Es existiert eine Vielzahl an NLDR-Methoden. In dieser Arbeit wurden durch eine umfangreiche Literaturanalyse acht relevante Methoden identifiziert. Um Visualisierung von hochdimensionalen Daten zu erstellen, welche sowohl die lokale als auch die globale Struktur möglichst beibehalten, wird die Auswahl weiter eingeschränkt. Besteht nämlich solche Anforderung, so können von diesen acht Methoden lediglich t-SNE und UMAP diese im größeren Rahmen treffen. Zwar können t-SNE und vor allem UMAP nicht so viele Vorkommnisse in der Konzeptmatrix aufweisen, wie andere Methoden, jedoch kann hier das spätere Veröffentlichungsjahr als Hauptgrund angeführt werden.

5 Diskussion und Ausblick

Als Ergebnis dieser Forschungsarbeit können acht NLDR-Methoden genannt werden. Diese können die meisten Vorkommnisse im Untersuchungsgegenstand aufweisen. Jedoch ist hier anzumerken, dass die Vorkommnisse nicht direkt in Relevanz für den State-of-the-Art für NLDR-Methoden übertragen werden können. Weiterhin ist die Aussagekraft der Ergebnisse durch die Anzahl der untersuchten Publikationen von 51 Stück ebenfalls begrenzt. Zusätzlich haben ältere Methoden mehr Zeit gehabt Beachtung bei den Anwendern zu finden, während eventuelle besser performende Methoden nicht unbedingt schon die nötige Bekanntheit erreicht haben. Auf der anderen Seite konnte keine genauere Betrachtung der Methoden unter der Berücksichtigung verschiedener Anforderungen durchgeführt werden. So spielt es eine Rolle, ob eher die globale oder lokale Struktur für einen Anwendungsfall wichtig ist. Ebenso wurde die Performance im Hinblick auf Rechenzeit und -speicher nicht berücksichtigt. Auch die Datengröße und Anzahl von Dimensionen kann unterschiedliche Auswirkungen auf die Performance der verschiedenen Methoden haben. Zuletzt kommt noch hinzu, dass das Veröffentlichungsdatum der Primärquellen der Methoden nicht mit der Veröffentlichung der eigentlichen Methode übereinstimmen muss. Trotzdem können die Ergebnisse als Anhaltspunkt für den State-of-the-Art von NLDR-Methoden weiterverwendet werden.

Weiterführende Forschungsarbeiten könnten die acht NLDR-Methoden der Abb. 4 näher betrachten, um klare Aussagen über die Performance der Methoden treffen zu können. Außerdem könnte ein klares Kriteriumsraster oder Key-Performance-Indicator festgelegt werden, mit denen eine Evaluierung möglich ist. Weiterhin wäre interessant, inwieweit die älteren Methoden noch Relevanz aufweisen durch beispielsweise einfachere Implementierung oder Ähnliches.

Literaturverzeichnis

- [BN01] Belkin, M.; Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering: Nips, S. 585–591, 2001.
- [CC13] Chahooki, M. A. Z.; Charkari, N. M.: Learning the shape manifold to improve object recognition. Machine vision and applications 1/24, S. 33–46, 2013.
- [Ch19] Chen, N. et al.: Application of computational intelligence technologies in emergency management: a literature review. Artificial Intelligence Review 3/52, S. 2131–2168, 2019.
- [CL06] Coifman, R. R.; Lafon, S.: Diffusion maps. Applied and Computational Harmonic Analysis 1/21, S. 5–30, 2006.
- [CL11] Chen, J.; Liu, Y.: Locally linear embedding: a survey. Artificial Intelligence Review 1/36, S. 29–48, 2011.
- [Co07] Cook, J. et al.: Visualizing similarity data with a mixture of maps: Artificial Intelligence and Statistics, S. 67–74, 2007.
- [CRH08] Carter, K. M.; Raich, R.; Hero III, A. O.: An Information Geometric Framework for Dimensionality Reduction, 2008.
- [CY06] Chang, H.; Yeung, D.-Y.: Robust locally linear embedding. Pattern recognition 6/39, S. 1053–1065, 2006.
- [DG03] Donoho, D. L.; Grimes, C.: Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. Proceedings of the National Academy of Sciences 10/100, S. 5591–5596, 2003.
- [DH96] Demartines, P.; Herault, J.: CCA: “Curvilinear Component Analysis. 15° Colloque sur le traitement du signal et des images, 1995 ; p. 921-924, 1996.
- [Di21] Ding, J.-E. et al.: Dopamine Transporter SPECT Image Classification for Neurodegenerative Parkinsonism via Diffusion Maps and Machine Learning Classifiers. arXiv preprint arXiv:2104.02066, 2021.
- [DQ17] Ding, C.; Qi, H.-D.: Convex optimization learning of faithful Euclidean distance representations in nonlinear dimensionality reduction. Mathematical Programming 1/164, S. 341–381, 2017.
- [Fe06] Fettke, P.: State-of-the-Art des State-of-the-Art. Wirtschaftsinformatik 4/48, S. 257, 2006.
- [Gh20] Ghojogh, B. et al.: Stochastic neighbor embedding with Gaussian and Student-t distributions: Tutorial and survey. arXiv preprint arXiv:2009.10301, 2020.
- [Gh21] Ghojogh, B. et al.: Uniform Manifold Approximation and Projection (UMAP) and its Variants: Tutorial and Survey. arXiv preprint arXiv:2109.02508, 2021.
- [GS93] Gross, M. H.; Seibert, F.: Visualization of multidimensional image data sets using a neural network. The Visual Computer 3/10, S. 145–159, 1993.
- [HA17] Henry, R. P.; Alfred, R.: Synergy in facial recognition extraction methods and recognition algorithms: International Conference on Computational Science and Technology, S. 358–369, 2017.

- [Ha21] Haiyang Zhu et al.: Visualizing large-scale high-dimensional data via hierarchical embedding of KNN graphs. *Visual Informatics* 2/5, S. 51–59, 2021.
- [He19] Hemmati, S. et al.: Bringing Manifold Learning and Dimensionality Reduction to SED Fitters. *The Astrophysical Journal* 1/881, L14, 2019.
- [HMZ15] Hao, Z.-H.; Ma, S.-W.; Zhao, F.: Atlas compatibility transformation: A normal manifold learning algorithm. *International Journal of Automation and Computing* 4/12, S. 382–392, 2015.
- [Ho33] Hotelling, H.: Analysis of a complex of statistical variables into principal components. *Journal of educational psychology* 6/24, S. 417, 1933.
- [HR02] Hinton, G.; Roweis, S. T.: Stochastic neighbor embedding: NIPS, S. 833–840, 2002.
- [Hu20] Hude, M. von der: Dimensionsreduktion - Hauptkomponentenanalyse englisch: principal components (PCA): Predictive Analytics und Data Mining Eine Einführung mit R. Springer Fachmedien Wiesbaden, Wiesbaden, S. 83–92, 2020.
- [Hu19] Hurley, N. C. et al.: Visualization of Emergency Department Clinical Data for Interpretable Patient Phenotyping. arXiv preprint arXiv:1907.11039, 2019.
- [JT19] Johannemann, J.; Tibshirani, R.: Spectral Overlap and a Comparison of Parameter-Free, Dimensionality Reduction Quality Metrics. arXiv preprint arXiv:1907.01974, 2019.
- [Ko90] Kohonen, T.: The self-organizing map. *Proceedings of the IEEE* 9/78, S. 1464–1480, 1990.
- [KVD07] Khurd, P.; Verma, R.; D., Christos (2007): Kernel-Based Manifold Learning for Statistical Analysis of Diffusion Tensor Images. In: (Karssemeijer, N.; Lelieveldt, B. Hrsg.): *Information Processing in Medical Imaging*. Berlin, Heidelberg, 2007. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 581–593.
- [LCW21] Liang, Y.; Chaudhuri, A.; Wang, H.: Visualizing the Finer Cluster Structure of Large-Scale and High-Dimensional Data: International Conference on Knowledge Science, Engineering and Management, S. 361–372, 2021.
- [Le00] Leeuw, J. de: Multidimensional Scaling, 2000.
- [Le10] Lei, Y. et al.: Fast ISOMAP Based on Minimum Set Coverage. In (Huang, De-Shuang and Zhang, Xiang and Reyes García, Carlos Alberto and Zhang, Lei Hrsg.): *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence*. Springer Berlin Heidelberg, Berlin, Heidelberg, S. 173–179, 2010.
- [Li16] Lin, T. et al.: Nonlinear Dimensionality Reduction by Local Orthogonality Preserving Alignment. *Journal of Computer Science and Technology* 3/31, S. 512–524, 2016.
- [Li17] Lin Ma et al.: Adaptive Neighboring Selection Algorithm Based on Curvature Prediction in Manifold Learning, 2017.
- [LV07] Lee, J. A.; Verleysen, M.: Characteristics of an Analysis Method. In (Lee, J. A.; Verleysen, M. Hrsg.): *Nonlinear Dimensionality Reduction*. Springer New York, New York, NY, S. 17–45, 2007.
- [LW18] Lin, C.-Y.; Wu, H.-T.: Embeddings of Riemannian manifolds with finite eigenvector fields of connection Laplacian. *Calculus of Variations and Partial Differential Equations* 5/57, S. 1–39, 2018.

- [Me17] Mehta, K. et al.: Modified locally linear embedding with affine transformation. National Academy Science Letters 3/40, S. 189–196, 2017.
- [MHM18] McInnes, L.; Healy, J.; Melville, J.: Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426, 2018.
- [Ng20] Nguyen, Q. V. et al.: Evaluation on interactive visualization data with scatterplots. Visual Informatics 4/4, S. 1–10, 2020.
- [PJJ14] Parekh, V. S.; Jacobs, J. R.; Jacobs, M. A.: Unsupervised nonlinear dimensionality reduction machine learning methods applied to multiparametric MRI in cerebral ischemia: preliminary results: Medical Imaging 2014: Image Processing, 90342O, 2014.
- [RRB21] Ray, P.; Reddy, S. S.; Banerjee, T.: Various dimension reduction techniques for high dimensional data analysis: a review. Artificial Intelligence Review, S. 1–43, 2021.
- [RS00] Roweis, S. T.; Saul, L. K.: Nonlinear dimensionality reduction by locally linear embedding. Science 5500/290, S. 2323–2326, 2000.
- [Sa69] Sammon, J. W.: A Nonlinear Mapping for Data Structure Analysis. IEEE Transactions on Computers 5/C-18, S. 401–409, 1969.
- [Se20] Seok, H.-S.: Performance comparison of dimensionality reduction methods on RNA-Seq data from the GTEx project. Genes & genomics 2/42, S. 225–234, 2020.
- [SJ07] Shaw, B.; Jebara, T.: Minimum volume embedding: Artificial Intelligence and Statistics, S. 460–467, 2007.
- [So19] Song, W. et al.: Improved t-SNE based manifold dimensional reduction for remote sensing data processing. Multimedia Tools and Applications 4/78, S. 4311–4326, 2019.
- [Su20] Sun, Z. et al.: A Survey on Dimension Reduction Algorithms in Big Data Visualization, S. 375–395, 2020.
- [SZK20] Shamai, G.; Zibulevsky, M.; Kimmel, R.: Efficient Inter-Geodesic Distance Computation and Fast Classical Scaling. IEEE Transactions on Pattern Analysis and Machine Intelligence 1/42, S. 74–85, 2020.
- [Ta16] Tang, J. et al.: Visualizing Large-Scale and High-Dimensional Data: Proceedings of the 25th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, S. 287–297, 2016.
- [TdL00] Tenenbaum, J. B.; de Silva, V.; Langford, J. C.: A Global Geometric Framework for Non-linear Dimensionality Reduction. Science 5500/290, S. 2319–2323, 2000.
- [TJ20] Ting, D.; Jordan, M. I.: Manifold Learning via Manifold Deflation. arXiv preprint arXiv:2007.03315, 2020.
- [UA19] Urpa, L. M.; Anders, S.: Focused multidimensional scaling: interactive visualization for exploration of high-dimensional data. BMC bioinformatics 1/20, S. 1–8, 2019.
- [vH08] van der Maaten, L.; Hinton, G.: Visualizing data using t-SNE. Journal of machine learning research 11/9, 2008.
- [VK06] Venna, J.; Kaski, S.: Local multidimensional scaling. Neural Networks 6/19, S. 889–899, 2006.

- [Wa19] Wang, J. et al.: Multi-cancer samples clustering via graph regularized low-rank representation method under sparse and symmetric constraints. *BMC bioinformatics* 22/20, S. 1–15, 2019.
- [Wa21] Wang, Y. et al.: Understanding how dimension reduction tools work: an empirical approach to deciphering t-SNE, UMAP, TriMAP, and PaCMAP for data visualization. *J Mach. Learn. Res* 22, S. 1–73, 2021.
- [WW02] Webster, J.; Watson, R. T.: Analyzing the past to prepare for the future: Writing a literature review. *MIS quarterly*, S. xiii–xxiii, 2002.
- [ZHW20] Zhang, S.; Huang, W.; Wang, Z.: Plant species identification based on modified local discriminant projection. *Neural Computing and Applications* 21/32, S. 16329–16336, 2020.
- [ZL10] Zhang, S.-W.; Liu, J.: Weighted locally linear embedding for plant leaf visualization: International Conference on Intelligent Computing, S. 52–58, 2010.
- [ZW06] Zhang, Z.; Wang, J.: MLLE: Modified Locally Linear Embedding Using Multiple Weights, S. 1593–1600, 2006.
- [ZZ04] Zhang, Z.; Zha, H.: Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM journal on scientific computing* 1/26, S. 313–338, 2004.

Kernkonzepte der Taxonomiesprache GenDifS

Johannes Busse¹

Abstract: GenDifS ist eine neue Sprache zur Notation von Terminologien und Taxonomien gemäß der Formel *Genus Proximum, Differentia Specifica*. GenDifS nutzt das Opensource Mindmap Tool Freemind/Freeplane zur Notation und ist in frühen und in späten Phasen des Ontology Engineerings ohne Medienbruch nutzbar. Auf der Sachebene werden das Sprachkonzept, die Syntax und die Semantik der Kernkonzepte der Sprache erstmals vorgestellt. Auf der Metaebene trägt der Aufsatz zur Methodologie der Design Science Research bei: Wie lassen sich Artefakte von Typ Terminologie und Sprache durch Rigor entwickeln und evaluieren?

Keywords: Semantic Web, Ontologie, Taxonomie, SKOS, OWL2, Design Science Research.

1 Problemstellung

Seit Aristoteles unterscheidet man zwei Begriffe, indem man zunächst einen „besten“ gemeinsamen Oberbegriff sucht (das sogenannte Genus Proximum, die nächsthöhere Gattung) und dann in Bezug auf diesen Oberbegriff das treffendste spezifische Merkmal (die sogenannte Differentia Specifica, den Artunterschied) bestimmt.

Beispiel: In einer Lehrveranstaltung sei gewünscht, eine Taxonomie zu entwickeln, z.B. zur Domäne „Schnitzel“. In der Prüfung werde das Begriffsverständnis überprüft, indem der spezifische Unterschied zwischen verschiedenen Schnitzeln erfragt wird, Beispiele: Q: Unterschied Schweineschnitzel – Kalbschnitzel? A: Das Fleisch kommt vom Schwein / vom Kalb, leere Schnittmenge. Q: Schweineschnitzel – Kinderschnitzel? A: Herkunft des Fleisches, Zielgruppe; große Schnittmenge. Q: Kalbschnitzel – Wiener Schnitzel? A: Das ist das Gleiche.

¹ HAW Landshut, Fakultät für Informatik, Am Lurzenhof 1, 84036 Landshut, busse@haw-landshut.de

Das Beispiel zeigt, dass der charakteristische Unterschied manchmal im Wert eines Attributs, manchmal im Attribut selbst liegt, das wir zur Unterscheidung heranziehen. Offensichtlich gibt es unterschiedliche Unterschiede. Einen noch nicht begrifflich erfassten Gegenstandsbereich begrifflich zu durchdringen bedeutet, über eine logisch konsistente Systematik von Unterschieden zu verfügen.²

Terminologien entfalten in den begrifflichen Grundlagen eines Faches eine tiefgreifende Ordnungsmacht. Dies gilt erst recht für die Methoden und Sprachen, mit denen Taxonomien entwickelt und notiert werden. Entsprechend scheint es uns geboten, eine Sprache zur Taxonomie-Entwicklung im Sinne der Design Science Research (DSR) auch mit Rigor zu konstruieren und zu evaluieren. Wir stellen unsere neue Sprache GenDifS also nicht nur vor, sondern reflektieren und begründen unser Vorgehen ausführlich insbesondere in Bezug auf Hevners Darstellung der DSR ([HMPR04], im Folgenden pars pro toto mit *Hevner 2004* zitiert). Wir betreiben also auch Wissenschaftstheorie, zu der teilweise auch „philosophierende“ Argumentationsmuster hinzugehören.

1.1 Abstraktes und konkretes Business-Problem

Ein erster Schritt zur konkreten Evaluierbarkeit besteht darin, nicht lediglich abstrakte eine Sprache zu evaluieren, sondern auch im Detail anzugeben, welches konkrete Business-Problem mit der neuen Sprache angegangen werden soll.

Stark abstrahiert bearbeiten wir die Herausforderung, Terminologien und Taxonomien visuell als Mindmap formaler zu fassen. Genauer besteht unser *abstraktes Business Problem* darin, eine Sprache zu entwickeln, mit der die Begriffe eines Gegenstandsbereichs als eine Taxonomie (a) visuell in einer Mindmap leicht editiert und in einer Form dargestellt werden können, die (b) formal wohldefiniert ist und die (c) so in die Semantic-Web-Sprache OWL2 übersetzbare ist, dass (d) ein Standard OWL2-Reasoner einzelne Objekte anhand ihrer Merkmale automatisch klassifizieren kann. Zusätzlich soll (e) die Mindmap in einen SKOS-Thesaurus übersetzbare sein.

Das bekannteste Modellierungspattern für die Strukturierung von Taxonomien ist die seit mehr als 2000 Jahren bekannte Formel *Genus Proimum, Differentia Specifica*, die wir im folgenden *GPDS-Modellierungspattern* nennen. *GenDifS* wollen wir die Sprache nennen, die (a) im Kern das GPDS-Modellierungspattern realisiert sowie (b) dieses Modellierungspattern um ausgewählte komplementäre Modellierungspatterns erweitert. *GenDifS*

² Unsere Auswahl der Beispiele mag irritieren: Wäre es in einem Publikationsorgan des AKWI nicht angemessener, Begriffe aus der Wirtschaftsinformatik als Beispiele heranzuziehen, statt über Schnitzel, Milch, Pferde zu reden? In diesem Fall nicht. Denn unsere Methode weist Wörtern der Alltagssprache eine hohe Relevanz für die Theoriebildung zu, da Alltagssprache einerseits unmittelbar und ohne theoretischen Hintergrund verstanden wird, aber im Unterschied zur Fachsprache eben noch nicht begrifflich sauber verfasst ist. Um exemplarisch einen Weg von der Alltags- zur Fachsprache aufzuzeigen, können wir nicht schon mit einer Fachsprache beginnen.

verspricht, das abstrakte Business-Problem der Taxonomie-Konstruktion leichter, schneller und schöner lösbar zu machen.

Abstrakt ist dieses Problem, weil es von konkreten Rahmenbedingungen abstrahiert (lat. ab-: weg-; und lat. trahere: ziehen, schleppen). Wir verstehen Hevner 2004 so, dass es allerdings genau die konkreten Rahmenbedingungen sind, die ein Business Problem im Sinne einer Design-Science-Forschung konkret bearbeitbar und insbesondere auch konkret evaluierbar machen. Tatsächlich beschäftigt uns nicht das abstrakte Business Problem, sondern eine Reihe von ähnlichen, *konkreten* Business Problemen. Diese sind dadurch definiert, dass teils systematische, teils zufällige Rahmenbedingungen zu einem konkreten angewandten Problem zusammengewachsen sind. (lat. con-: zusammen; lat. crescere: wachsen).

In unserem *ersten konkreten Business-Problem* erarbeiten sich Studierende den Inhalt eines Lehrbuches zunächst als Mindmap, um dann Teile dieser Mindmap in eine höherwertige Wissensrepräsentation wie insbesondere eine Terminologie oder eine vollwertige Taxonomie überführen. Der Artefakt-Typ „Terminologie“ aus Hevner 2004 ist in der Didaktik im Rahmen der revidierten Lernzieltaxonomie von Anderson und Krathwohl als Lernzieltyp in der *knowledge dimension* auf der ersten und zweiten Stufe angesiedelt und damit Grundlage vieler höherer Lernprozesse: „A: Factual Knowledge. Aa: Knowledge of terminology. [...] B: Conceptual Knowledge [...] Ba: Knowledge of classifications and categories.“ ([Kra02], p.214).

In der Tat besteht eine wesentliche Aufgabe akademischer Bildung darin, eine analytisch saubere Begriffsbildung voranzutreiben. Vorbildhaft wollen wir hier als Lehrende vorangehen: In unserem *zweiten konkreten Business Problem* sind wir Lehrenden es, die als Domänen-Experten die Begrifflichkeit eines Faches als Taxonomie formalisieren und Studierenden visualisiert als Lehr- und Referenzmaterial zur Verfügung stellen. Abbildung 1: Beispiel Schnitzel zeigt die Formalisierung dieses Beispiels in GenDifS.



Abb. 1: Beispiel Schnitzel

Die Gemeinsamkeit unserer konkreten Business-Probleme besteht darin, dass einzelne Menschen im Zuge ihres Lernens und Lehrens sich eine Wissensdomäne erarbeiten und mit einem Mindmap-Tool strukturiert zusammenfassen. Typische Entwicklungsstadien der entstehenden Wissensrepräsentationen sind (a) eine gewöhnliche Mindmap als eine im wesentlichen baumartig strukturierte und oft mit reichhaltigem Layout, Icons etc. gestaltete Sammlung kleiner Informationseinheiten; (b) ein Glossar als eine Sammlung von Begriffen und Erläuterungen; falls in den Erläuterungen auf andere Glossarbegriffe Bezug genommen wird, werden diese Bezüge z.B. durch eine Verlinkung deutlich gemacht, es entsteht ein Wiki; (c) ein Begriffssystem als semiformale oder formale Strukturierung der Bezüge der Glossarbegriffe untereinander, insbesondere als Taxonomie.

Eine Analyse der intendierten konkreten Business Probleme ergibt folgende konkrete Anforderungen: Die GenDifS-Mindmap soll (a) über möglichst alle Phasen des Ontology-Engineering-Prozesses (gewöhnliche Mindmap, Glossar, Taxonomie) verwendbar sein und (b) alle Phasen nach- und nebeneinander ohne Medienbruch unterstützen. (c) Die Mindmap soll von Domain-Experten, die eine leichtgewichtige Einführung in Semantic-Web-Technologien erhalten haben, mit aktiver Sprachkompetenz in Eigenverantwortung aktiv produziert (d.h. geschrieben) werden können, und (d) insbesondere in der akademischen Lehre von Studierenden nach einer kurzen Einführung in GenDifS passiv rezipiert (d.h. gelesen) werden können.

1.2 Lösungsidee zum einleitenden Beispiel

Ursprünglich hatten wir nur einziges, etwas unspezifisches Artefakt als Entwicklungsziel vor Augen, nämlich eine domänen spezifische „Sprache“ – was auch immer man sich unter „Sprache“ vorstellen mag. Im Zuge der Überlegungen, wie man diese Sprache im Sinne von Hevner (2004) mit *Rigor* evaluieren könnte, ergaben sich viele Möglichkeiten, viele Abhängigkeiten, viele verdeckte Variablen, und auch immer mehr philosophische Fragen.

Schnell wurde klar, dass das scheinbar monolithische Artefakt „Sprache“ sich tatsächlich aus mindestens vier z.T. sehr unterschiedlichen Teil-Artefakten zusammensetzt. Im Einzelnen haben wir vor uns: a) eine *Methode*, um mit dieser Sprache Taxonomien zu entwickeln; b) eine kleine Menge von weiteren *Modellierungspatterns*, die zusätzlich zum grundlegenden GPDS-Modellierungspattern erst die angestrebte Praxisrelevanz sicherstellen; c) eine *konkrete Syntax und Semantik* für die Notation dieser Modellierungspatterns in einer Mindmap; d) eine prototypische *Implementierung* dieser Sprache in Form eines Skriptes, das eine GenDifS-Mindmap insbesondere in eine OWL-Ontologie und einen SKOS-Theaurus übersetzt.

Im vorliegenden Aufsatz stellen wir die Kernelemente, namentlich die Modellierungspatterns sowie die Syntax und Semantik, von GenDifS dar.

2 Stand von Forschung und Technik

Methoden: Eine Übersicht über Ontology-Engineering-Methoden geben z.B. [SP21] und [IMMS13] sowie [BSWZ07], [SSS04]. Es fällt auf, dass die Mehrzahl der Methoden davon ausgeht, dass man sich schon in frühen Phasen einer Modellierung für ein Tool und/oder eine Modellierungssprache entscheiden muss – typischerweise RDF(S), SKOS oder OWL, wie z.B. auch in [SGL12]. Allerdings widerspricht dies u.E. Tom Grubers Design-Kriterium des *minimal encoding bias*: „The conceptualization should be specified at the knowledge level without depending on a particular symbol-level encoding. An *encoding bias* results when representation choices are made purely for the convenience of notation or implementation. Encoding bias should be minimized, because knowledge-sharing agents may be implemented in different representation systems and styles of representation.“ [Gru95].

Die Sprache GenDifS adressiert das Designkriterium des *minimal encoding bias* im Kern, indem sie die Wahl zwischen RDF(S), OWL oder SKOS nicht schon zur Modellierungszeit erfordert, sondern OWL und SKOS gleichzeitig erzeugt, und zwar in einer Weise modularisiert, in der das OWL- und das SKOS-Modul sowohl stand-alone wie auch in beliebiger Kombination gemeinsam verwendet werden können.

Modellierungspatterns: Mit dem GPDS-Modellierungspattern zur Modellierung von Wissensstrukturen können wir uns auf einen Stand der Technik berufen, der seit mehr als 2000 Jahren besteht. In [SKOS] erkennen wir das GPDS-Modellierungspattern auch in sogenannten Collections wieder. Tatsächlich gab der Baum des Porphy in Verbindung mit den Collections des Milch-Beispiels aus dem SKOS-Primer den maßgeblichen Anstoß zur Entwicklung von GenDifS. Das Schema von SKOS würdigt die Bedeutung dieses Ordnungsprinzips, indem es dazu eine eigene Klasse `skos:Collection` mit zugehörigen Relationen wie `skos:member` bereithält. In GenDifS schlagen sich Collections in `BY SOME` als maßgebliches zentrales Modellierungspattern nieder.

Neben dem Export einer GenDifS-Mindmap in einen SKOS-Thesaurus suchen wir auch eine Lösung für einen Export in eine Taxonomie, die in der Semantic-Web-Sprache OWL2 formuliert ist und mit einem geeigneten OWL-Reasoner eine Klassierung von Objekten anhand ihrer Attribut-Wert-Paare ermöglicht. Für unser abstraktes Business Problem verfügen wir in der W3C-Publikation [Gro05] mit dem *Pattern 2: Values as subclasses partitioning a feature / using a fact about the individual* über eine OWL-Codierung, die genau der gesuchten Formalisierung entspricht. Dieser Stand der Technik löst das abstrakte Business Problem, nicht jedoch unsere zwei konkreten Business Probleme. Denn das zitierte Modellierungspattern mit einem Tool wie Protegé [Mus15] direkt in OWL zu formulieren ist selbst für Experten kompliziert, aufwändig und fehleranfällig. Für unsere Zielgruppen ist dieses Vorgehen gänzlich ungeeignet.

Der Lösungsansatz für GenDifS besteht darin, (a) das GPDS-Modellierungspattern in Gestalt der `skos:Collection` mit der existierenden formalen Lösung des W3C für *OWL value sets* konzeptuell zusammenzuführen, und (b) die erforderlichen Informationen in einer geeigneten Syntax und Semantik formal derart darzustellen, dass (c) ein Compiler eine auto-

matische Übersetzung nicht nur nach SKOS, sondern auch nach OWL2 vornehmen kann.

Mindmaps und Ontologien: Mindmap-basierte Notationen, die eine Ontologie oder eine vollwertige Taxonomie erzeugen, konnten wir nicht recherchieren. Zwar lassen sich aus einer Mindmap mit geringem Aufwand Klassenbäume erzeugen, die sich aber mangels Informationsgehalt nicht zur Klassierung von Instanzen anhand von Eigenschaften nutzen lassen. „Nackte“ Klassenhierarchien sind in der Welt der Taxonomien nur halbfertige Gebilde. Reine RDF(S)-Tools, Thesaurus- oder SKOS-Editoren sind i.A. zu ausdruckschwach, um eine Taxonomie so zu notieren, dass eine Klassierung erfolgen kann.

Am ehesten scheint UML die Modellierung einer Ontologie oder einer Taxonomie zu unterstützen. Eine genaue Durchsicht der in dem Survey [MNN20] genannten Tools zeigt allerdings, dass mit UML zwar Klassenhierarchien, nicht aber die zur Klassierung speziell benötigten, hier beschriebenen Strukturen notiert werden können. Auch unser eigenes, um 2010 entwickeltes Tool *semAuth* ([Bus14]) kann nur bedingt die für Taxonomien typische Klassierung leisten. Zwar lassen sich die für Klassierungen notwendigen und hinreichenden Bedingungen durch F-Logik ([KLW95]) sehr elegant modellieren. Relationen hatten wir jedoch damals als Domain und Range modelliert, was wir heute als Fehler interpretieren.

Unsere **prototypische Implementierung** `gendifs.py` greift möglichst weitgehend auf existierende Technologie zurück. GenDifS hat kein eigenes GUI, sondern liest die von den Mindmap-Tools *Freemind* oder *Freeplane* nativ verwendete XML-Datei `*.mm` ein. Aus den Mindmap-Knoten wird zunächst der Code im Turtle-Format erzeugt, wie er auch in diesem Aufsatz abgedruckt ist. Dieser Code wird anschließend durch `rdflib` [BOP+20] wieder eingelesen. Ivan Hermans Reasoner `owlrl` [Her14] berechnet dann mittels Forward-Inferenz den deduktiven Abschluss aller über OWL-RL und RDF(S) erreichbaren Tripel. Neu entwickelt wurde also alleine die Generierung von `ttl`-Code auf Grundlage ausgewählter Mindmap-Knoten im XML-Format.

Tatsächlich glauben wir, mit der hier vorliegenden Erstpublikation von GenDifS eine Innovation anzubieten, die es so bisher noch nicht gab.

3 Methodologie

Wir verorten die Entwicklung von GenDifS als typische Design-Science-Research (DSR) ([HMPR04]), und zwar in unserem Fall spezifisch an einer Hochschule für angewandte Wissenschaften. Unsere programmatische Anwendungsorientierung schlägt sich u.A. darin nieder, dass wir zwar ein abstraktes Business-Problem lösen wollen, dieses aber vorwiegend aus der Perspektive einer Gruppe von konkreten Business-Problemen bearbeiten.

Wann und an welchen Schnittstellen in einem DSR-Prozess Evaluation stattfinden kann und soll, wird in der DSR uneinheitlich gesehen [BHM20]. Breite und Offenheit des Evaluationsbegriffs spiegeln sich in der Vielfalt der Evaluationsmethoden wider. Wir wollen Evaluation im weitesten Sinn verstehen als den systematischen Abgleich von Ist und Soll, von Vorstellung, Modell und Wirklichkeit. Hevner 2004 beschreibt drei Cycles, in denen durch Evaluation Rigor unterstützt wird. Der Relevance- (R-) Cycle stellt sicher, dass das entwickelte Artefakt tatsächlich das richtige Problem löst. Der Knowledge-Base- (KB-) Cycle stellt den Bezug zu Theorien, Methoden und dem Stand von Wissenschaft und Technik her, und der Build-and-Evaluate (BE-) Cycle „*is the heart of any design science research project. This cycle of research activities iterates more rapidly between the construction of an artifact, its evaluation, and subsequent feedback to refine the design further*“ ([HMPR04], p90f).

Es stellt sich die Frage: Wie gestalten wir im Kontext von GenDifS die Evaluation im BE-Cycle und behalten gleichzeitig die Relevanz und das bestehende Wissen im Blick? In unserer Entwicklung greifen der R- und der BE-Cycle durch die Auswahl unserer Modellierungsbeispiele eng ineinander, während der KB-Cycle existierende Lösungsansätze beisteuert. Speziell verzahnen sich BE- und R-Cycle über die Auswahl der zu modellierenden informellen Wissensrepräsentationen, die für die spezifisch gewählten konkreten Business-Probleme typisch sind. Im BE-Cycle modellieren wir Beispiele in GenDifS und beurteilen das Modell subjektiv, um darauf aufbauend die verwendeten Sprachkonstrukte anzupassen und weiterzuentwickeln.

Es schließt sich der Test auf Wartbarkeit an: Eine weitere Informationsquelle wird hinzugenommen, das Modell dann bearbeitet, erweitert und vor allem korrigiert. Mit diesem Vorgehen testen wir auch die eingangs beschriebene Anforderung, der zufolge eine schrittweise Weiterentwicklung einer Ontologie möglich sein muss.

4 Ergebnisse

Die maßgebliche Motivation für GenDifS ist das Konzept der Klasse `skos:Collection`, das im SKOS-Primer [SKO] unter der Überschrift *4 Advanced SKOS: When KOSs are not Simple Anymore > 4.1 Collections of Concepts* in dem folgenden Milch-Beispiel veranschaulicht wird:

```
milk
<milk by source animal>
  cow milk
  goat milk
  buffalo milk
```

In GenDifS wird dieses Beispiel wie in Abb. 2 formalisiert:

ONTOLOGY Milch-Beispiel **Milch** BY Milch_hat_Herkunft SOMETier **Kuhmilch** SOME *Kuh*

Abb. 2: Beispiel Milch

Offensichtlich sind in diesem Beispiel zwei parallele Teilmengen-Beziehungen enthalten: Die erste, in Fettdruck wiedergegebene Beziehung `Kuhmilch rdfs:subClassOf :Milch` spannt insgesamt den für Mindmaps typischen Baum auf. Mit diesem Baum eng verflochten ist aber auch eine zweite, in der Abbildung in unterstrichenem Kursivdruck wiedergegebene Beziehung `:Kuh rdfs:subClassOf :Tier` enthalten. GenDifS interpretiert obige Mindmap als zwei zusammengehörende, ineinander verflochtene Bäume. Dieses Syntax-Detail ist eine Folge von explorativem, spielerischem Modellierungshandeln und entspringt der Beobachtung, dass Unterscheidungen in einem Sekundärbaum bisweilen einen Primärbaum strukturieren.

Wir skizzieren die formale Semantik der Sprache, indem wir exemplarisch die Übersetzung ausgewählter GenDifS-Idiome in das Format Turtle (*.ttl) angeben.

4.1 BY ... SOME

Kern der Sprache GenDifS ist das Modellierungspattern *Genus Proximum, Differentia Specifica (GPDS)*, also die Angabe eines Oberbegriffs plus Charakterisierung einer Teilmenge durch ein charakteristisches Merkmal (hier: ein charakteristisches Attribut-Wert-Paar). Dieses Modellierungspattern unterscheidet eine Taxonomie von einer reinen Klassenhierarchie und erlaubt die gesuchte Klassierung in Form eines Top-Down-Inferencings von der allgemeinen Klasse zur speziellen Klasse. Es wird in GenDifS durch `SOME ...` als Enkel-Knoten eines `BY ... SOME ...` implementiert.

Beispiel: *Wenn X eine Milch ist, und X als Herkunft eine Kuh hat, dann ist X eine Kuhmilch.* Der in obiger Abb. 2: Beispiel Milch umrandete Knoten `SOME Kuh` wird vom Compiler in folgenden Code übersetzt:

```
# namespaces to be aligned by the user
@prefix ex: <http://some.location/ns/ex#> . # OWL A-Box
@prefix : <http://some.location/ns/default#> . # OWL T-Box

# SOME.1, owl
:Kuh
    rdfs:subClassOf :Tier .

# SOME.2, owl-classification
:Milch_hat_Herkunft_SOME_Kuh a owl:Class ;
    owl:equivalentClass [ a owl:Restriction ;
        owl:onProperty :Milch_hat_Herkunft ;
        owl:someValuesFrom :Kuh ] ;
    rdfs:subClassOf :restrictions_BY .

# SOME.3, owl-classification
:Milch_AND_Milch_hat_Herkunft_SOME_Kuh a owl:Class ;
    rdfs:subClassOf :Kuhmilch ;
    owl:equivalentClass [ a owl:Class ;
        owl:intersectionOf ( :Milch :Milch_hat_Herkunft_SOME_Kuh ) ] .
```

In Prädikatenlogik entspricht dies der Formel $\forall X \ [\text{Kuhmilch}(X) \leftarrow \text{Milch}(X) \wedge \exists Y \text{Kuh}(Y) \wedge \text{Milch_hat_Herkunft}(X, Y)]$, die zugleich auch als Definition der formalen Semantik dieses Konstruktions dient. Die GenDifS-Website erklärt an eben diesem Milch-Beispiel im Detail das zugehörige Inferencing.

Ein in GenDifS formuliertes Modell kann in verschiedene Module exportiert werden, die per Namespace unterschieden werden und beliebig miteinander kombiniert werden können. RDF-Resources mit dem Defaultnamespace : (Doppelpunkt) definieren wie oben gezeigt die Klassen (die T-Box) aus dem *Taxonomie-Modul*. RDF-Resources mit dem Namespace `ex:` sind automatisch angelegte Instanzen (eine A-Box) der OWL-Klassen und bilden das *Example-Modul*, das eine testgetriebene Entwicklung der Ontologie unterstützt.

```
# SOME.8, owl-test
ex:Milch_ID_571743754
  a :Milch ;
  :Milch_hat_Herkunft ex:Kuh_ID_571743754 ;
  gendifs:classifyLike ex:Kuhmilch_ID_571743754 .
ex:Kuhmilch_ID_571743754 a :Kuhmilch .
ex:Kuh_ID_571743754 a :Kuh .
```

Für die Instanz `ex:Milch_ID_571743754` ist zu erwarten, dass sie aufgrund ihrer ebenfalls exemplarisch angegebenen Attribute und Werte gleich wie die Instanz `ex:Kuhmilch_ID_571743754` klassifiziert wird. Weil diese `ex`-Beispiele aus Klassennamen und einer ID zusammengesetzt sind, können als Nebeneffekt auch in der GUI von Protegé taxonomische Kategorien und exemplarische Instanzen klar unterschieden werden.

Im *SKOS-Modul* werden Konzepte eines Thesaurus als Instanzen der Klasse `skos:concept` angelegt. Diese Konzepte (die natürlich nicht selbst Elemente des SKOS-Schema sind, denn wir wollen ja nicht das SKOS-Schema selbst erweitern) halten wir im Namespace `cpt` zusammen:

```
@prefix cpt: <http://some.location/namespace/cpt#> . # concepts
# SOME.9, skos
cpt:Kuh
  rdf:type skos:Concept ;
  skos:broader cpt:Tier .
```

Das durch `BY ... SOME` realisierte Modellierungspattern bildet zwar den Kern der Sprache GenDifS, ist für sich genommen aber nicht ausdrucksstark genug, um wichtige praktische Anforderungen in der Taxonomie-Entwicklung abzudecken. Welche der vielen zusätzlich möglichen Modellierungspatterns tatsächlich in GenDifS aufgenommen werden sollen sind Design-Entscheidungen der Sprache, die sich in einem verzahnten R- und BE-Cycle bewähren müssen oder verworfen werden. Einige Sprachelemente dienen primär der Anschlussfähigkeit und Akzeptanz (z.B. REL), andere ergänzen das GPDS-Paradigma in idealer Weise (z.B. SUP). Wieder andere vermisst man, wurden aber absichtlich nicht implementiert, um die Komplexität der Sprache zu reduzieren. Wir diskutieren im Folgenden exemplarische Sprachelemente.

4.2 Superclass (SUP)

Die Bedeutung von SUP lässt sich am neuen Beispiel „Rapphengst“ gut zeigen: Pferde lassen sich u.A. nach Geschlecht, Fellfarbe unterscheiden; es gibt Heiß-, Warm- oder Kaltblüter; Mutationen des sog. Grey-Gens sorgen für das Ausschimmeln des Fells, usw. Viele dieser Kombinationen haben eigene Klassenbezeichner wie z.B. Rapphengst oder Schimmelstute. Ausschließlich baumartige strukturierte Wissensrepräsentationen kommen hier an ihre Grenze, da mit zunehmender Baumtiefe die Kombinatorik zuschlägt: männliche/weibliche Pferde, schwarze/weiße Pferde, warmblütige männliche schwarze Pferde usw.

Im oberen Teilbaum der Abb. 3: Beispiel Rapphengst wird die Modellierung in einer Mindmap schwierig: Wenn eine Klasse wie in unserem Beispiel Rapphengst an verschiedenen Stellen redundant erstellt würde, wäre dies ein Fehler in jeder auf Verständnis und Wartbarkeit ausgerichteten Wissensrepräsentation. (Das rote X-Icon schließt in GenDifS die entsprechenden Teilbäume aus der Übersetzung aus.) Da aber ein Wechsel zu einem in der Darstellung mächtigeren Graph-Tool aus verschiedenen Gründen nicht in Frage kommt und insbesondere auch nicht zu unseren konkreten Business Problemen passen würde, wollen wir an der Mindmap-Darstellung festhalten und innerhalb dieser Wissensrepräsentation eine Problemlösung suchen.



Abb. 3: Beispiel Rapphengst

Eine Lösung bietet die Idee des inversen Baums: Durch das Sprachelement `SUP` (Super-class) werden die *Kindknoten* eines Knotens (a) nicht wie sonst in Ontologie-Editoren (z.B. auch Protegé) als Unterbegriffe, sondern als *Oberbegriffe* einer Klasse interpretiert, wobei (b) ihre Schnittmenge insgesamt als eine *hinreichende Bedingung* für die aktuelle Klasse interpretiert wird. Für den in Abb. 3: Beispiel Rapphengst gekennzeichneten Knoten `SUP` erzeugt der Compiler folgenden Code:

```
# SUP.1, owl-classification
:Hengst_AND_Rappe a owl:Class ;
    rdfs:subClassOf :Rapphengst ;
    owl:equivalentClass [ a owl:Class ;
        owl:intersectionOf ( :Hengst :Rappe ) ] .

# SUP.2, owl-test
ex:Hengst_AND_Rappe_ID_693686695
    a :Hengst ;
    a :Rappe ;
    gendifs:classifyLike ex:Rapphengst_ID_693686695 .
ex:Rapphengst_ID_693686695 a :Rapphengst .
```

In Prädikatenlogik entspricht dies der Hornklausel $\forall X [\text{Rapphengst}(X) \leftarrow \text{Rappe}(X) \wedge \text{Hengst}(X)]$, die auch die formale Semantik dieses Sprachkonstrukts definiert.

`SUP` ergänzt das zentrale Modellierungspattern `BY ... SOME` in mehrfacher Weise. So hebt die „Richtungsumkehr“ der Leserichtung von Klassendefinitionen die Einschränkung von Mindmaps, nur Bäume und keine Graphen darstellen zu können, in interessanter Weise auf. Mit `SUP` lässt sich ein Graph als eine Kombination aus sich verzweigenden und zusammenführenden Wegen kreuzungsfrei als Baum repräsentieren.

`SUP` ergänzt das GPDS-Modellierungspattern aber nicht nur in der Mindmap-Visualisierung, sondern auch theoretisch. Wenn wir mit dem GPDS-Modellierungspattern eine Kategorie durch ein Genus Proximum und eine Differentia Specifica definieren, dann liegt eine sogenannte *intensionale Definition* vor, in der die Bedeutung eines Begriffs im Wesentlichen durch ein charakteristisches Merkmal definiert wird. `SUP` gesellt zu dieser Definitionsart die sogenannte *Nominaldefinition* hinzu, in der neue Begriffe extensional allein durch die Schnittmengenbildung anderer, bereits definierter Begriffe konstruiert werden. Der charakteristische Unterschied liegt im „klassifikatorischen“ Informationsgehalt: Bei intensionalen Definitionen wird ein neues charakteristisches Merkmal eingeführt, mit dem eine neue Unterscheidung getroffen werden kann. Dazu im Gegensatz sind Nominaldefinitionen „merkmalsfrei“ in dem Sinn, dass sie keine neuen Informationen hinzufügen, mit der man neue Unterscheidungen treffen könnte. Der alleinige, aber in der Praxis bedeutsame Wert von Nominaldefinitionen besteht darin, wichtige oder häufige Schnittmengen von bereits definierten Begriffen prägnant zu benennen.

4.3 REL und REV

Die eingangs erwähnte Forderung, insbesondere auch die Frühphasen semantischer Modellierungen zu unterstützen, macht es unerlässlich, an weit verbreitete Modellierungsgewohnheiten wie insbesondere auch Semantische Netze anschließen zu können. Dazu stellt auch GenDifS das Sprachelement `REL` (und das aus html bekannte inverse `REV`) bereit, das typisch ist für gerichtete Graphen mit gelabelten Kanten. Ein Semantisches Netz lässt sich geradlinig in einen SKOS-Thesaurus übersetzen. Die Übersetzung nach RDF(S) oder OWL bereitet hingegen Probleme, die auf der GenDifS-Website ausführlicher diskutiert werden.

Wir erwähnen die Probleme um das Konstrukt `REL` hier zumindest abstrakt, um eine wichtige Frage zu motivieren: Wie will man am Markt der Modellierungssprachen eine neu konstruierte Sprache zu anderen existierenden Sprachen positionieren? Wir verfolgen eine um Ergänzung bemühte Nischen-Strategie: GenDifS will mit anderen Wissensrepräsentationen nicht konkurrieren, sondern bestehendes Wissen, das in Form z.B. eines Semantischen Netzes, eines ER-Diagramms, eines BPMN-Modells etc. modelliert wurde, als relevant anerkennen, und bei Bedarf einzelne semantische Unbestimmtheiten aus dem Blickwinkel einer Taxonomie genauer klären.

4.4 Nicht realisierte Sprachelemente

Auch wenn es in Bezug auf Syntax, Semantik und Implementierung technisch möglich ist (und die Versuchung groß war), haben wir andere naheliegende Sprachkonstrukte nach reiflichem Überlegen *nicht* realisiert. Dazu gehören insbesondere alle Aspekte in Klassifikationen, die explizit oder implizit eine Negation enthalten (wie insbesondere `ONLY` oder `DISJOINT`), da diese ganz wesentlich die Komplexität erhöhen würden.

5 Evaluation

Eine der wichtigsten Eigenschaften von Design-Forschung ist eine systematische Evaluation als Beitrag zu *Rigor*, d.h. einer „sauber dokumentierte[n] wissenschaftliche[n] Herleitung anhand von anerkannten Kriterien“ ([OsterleWB10]). Wie im Abschnitt Methodologie dargestellt, siedelt Hevner 2004 Evaluation im oben eingeführten Relevance- (R-), Knowledge-Base- (KB-) und Build-and-Evaluate- (BE-) Cycle an. Wir stimmen vom Brocke ([SB12], [BHM20]) zu, für den eine Vielzahl unterschiedlicher Evaluationen den MörTEL zwischen allen Stufen im Design-Science-Research-Prozess darstellt.

Unsere eigene Evaluation siedelt sich bisher vorwiegend im hochfrequenten BE-Cycle an. Anhand vieler minimalistischer Beispiele motivierten und bewerteten wir einzelne Design-Entscheidungen insbesondere zum Teilartefakt Syntax und Semantik. Im BE-Cycle entdeckten wir immer wieder typische Problemlagen, interessante Modellierungspatterns und auch eigene blinde Flecken. (So kam etwa das Sprachkonstrukt ^{SUP} erst relativ spät zu GenDifS hinzu.) Faktisch sind wir in diesem Cycle so vorgegangen: (1) Identifiziere hoch vorstrukturierte Sachtexte oder existierende Modelle; (2) stelle diese als GenDifS-Mindmap dar; (3) beurteile die Mindmap und das erzeugte Modell. Das Evaluationsinteresse war hier: Wie geradlinig lässt sich der dargestellte Text in voll formalisierte Statements übersetzen? Wie wird die GenDifS-Methode unterstützt, d.h. lassen sich unklare Beziehungen zunächst vorläufig darstellen, jedoch so, dass der Übergang zur vollen Formalisierung leicht möglich ist? Wie könnten die existierenden Modellierungspatterns und Sprachkonstrukte von GenDifS verbessert werden?

Im Kern stellt sich die Frage, was man mit einem bestimmten Evaluations-Arrangement tatsächlich evaluiert. Denn wie in der Einleitung dargestellt, haben wir ja in unserem Fall vier Teil-Artefakte – Modellierungsmethode, Modellierungspatterns, konkrete Syntax und Semantik, Tool und Implementierung – vor uns, die eng zusammenhängen und ein komplexes System bilden. Wie evaluieren wir die Sprache nicht indirekt über die damit erzeugte Ontologie und auch nicht das Tool, und auch nicht die Kompetenz einer konkreten Nutzerin, sondern tatsächlich die Sprache? (Und wir wiederholen uns: Was immer man sich unter „Sprache“ genau vorstellen will?)

Verständlichkeit und Verwendbarkeit einer Sprache wie GenDifS in konkreten Business-Cases haben wir in Selbstanwendung ja auch schon bewertet und können uns auch ein Evaluationsdesign unter Rigor-Anforderungen vorstellen. Eine direkte Evaluation der Sprache für den abstrakten Usecase scheint uns evidenzbasiert dagegen nur schwer vorstellbar. Statt auf empirische Untersuchungen würden wir hier eher auf einen argumentativen Expertendiskurs setzen, im Sinne der Design-Science-Research also Bezüge zur Knowledge-Base herstellen. Ein wichtiges Kriterium für die *direkte Evaluation* einer Ontologiesprache haben wir mit Tom Grubers *minimal encoding bias* schon kennengelernt. Durch die leitende Idee von GenDifS, eine Taxonomie in einer abstrakten Baum-Darstellung zu modellieren, die keine frühe Entscheidung für eine Encodierung in OWL, RDF(S) oder SKOS mit der damit jeweils automatisch verbundenen Semantik erfordert, sowie die Selbstbeschränkung auf wenige Sprachkonstrukte, die die Auswahl von Rea-

sonern und OWL2-Profilen einfacher macht, kommen wir diesem Kriterium näher als andere Ontologie-Notationen.

Als anderes wichtiges Kriterium für die direkte Evaluation von GenDifS erachten wir die Kompatibilität mit anderen Modellierungssprachen. GenDifS konzentriert sich auf das Kerngeschäft Taxonomie und spielt hier seine Stärken aus. Bezuglich dem praxisrelevanten Evaluationskriterium der Anschlussfähigkeit war uns wichtig, keine Konkurrenz, sondern eine friedliche Koexistenz und Komplementarität mit Darstellungen wie ER-Schemata oder UML-Klassendiagrammen herzustellen. Mit der Mindmap-Notation schließt GenDifS nahtlos an andere etablierte Ökosystemen des nicht voll formalisierten Wissensmanagements an und erlaubt den Brückenschlag in die Welt der vollständigen logisch-semantischen Formalisierung.

6 Methodenkritik

Bei der Suche nach geeigneten Evaluationsmethoden direkt zur Sprache sind wir allerdings auf ein größeres Problem gestoßen: Je länger wir uns mit dem Thema „Evaluation einer neuen Sprache“ befasst haben, desto weniger glauben wir, dass ein Artefakt vom Typ Sprache in ähnlicher Weise wie andere Artefakt-Typen gemäß dem Design-Science-Paradigma nach Hevner 2004 mit Rigor entwickelt und evaluiert werden kann. Was ist in Bezug auf unser aus Methode, Modellierungspatterns, Syntax und Semantik, Tool und Implementierung bestehendes System gemeint mit dem Anspruch „Wir evaluieren eine Sprache“? In der DSR besteht Konsens, dass eine Artefaktentwicklung natürlich ein zunehmend gutes Verständnis der Problemstellung erzeugt:

In the early stages of a discipline or with significant changes in the environment, each new artifact created for that discipline or environment is “an experiment” that “poses a question to nature” (Newell and Simon 1976, p 114). [HMPR04]

Im besonderen Fall, in dem das Artefakt eine neue Sprache ist und eine solche Sprache genau dann gelungen ist, wenn man mit ihr bestimmte Dinge schöner, konziser, eleganter, differenzierter etc. ausdrücken kann: Dann verändert solch eine Sprache unsere Sicht auf die Welt. Konsequenterweise müsste man Artefakte vom Typ Sprache danach evaluieren, in welcher Hinsicht sie eine neue Weltsicht erzeugen. Inwiefern diese Evaluation aber mit Rigor erfolgen kann: Dazu gibt Design-Science-Research, solange sie im Sinne von Thomas Kuhn als Normalwissenschaft betrieben wird, keine Anleitung.

Ob im speziellen Fall einer neuen Ontologie-Sprache auch eine neue philosophische Ontologie folgt, können wir nicht beurteilen. Aber wenn es einen Unterschied macht, ob man ein Begriffsyste in RDF(S), in SKOS, in OWL oder eben auch in GenDifS formalisiert: Dann hat auch GenDifS das Potential, einen Unterschied zu machen – nicht die schlechteste Eigenschaft für eine Sprache, die die Unterscheidung von Unterscheidungen zu ihrer Sache macht.

Literatur

- [BOP+20] Boettiger,C; Ooms, J; Leinweber, K; Hester, J: rdflib v0.2.3. Zenodo, January 2020. doi:10.5281/ZENODO.1098478. Stand 2022 aktuell ist RDFlib 6.1.1, <https://rdflib.dev/> (2022-06-25)
- [BSWZ07] Braun, S; Schmidt, A.P.; Walter, A.; Zacharias, V.: The Ontology Maturing Approach for Collaborative and Work Integrated Ontology Development: In Chen, L. et al (ed.): Proceedings of the First International Workshop on Emergent Semantics and Ontology Evolution, CEUR-WS.org, 2007.
- [Bus14] Busse, J.: Semantische Modelle Mit Mindmaps. In Keller, S.A.; Schneider, R.; Volk, B. (Hrsg.): Wissensorganisation und -Repräsentation mit digitalen Technologien, S. 115–127. DE GRUYTER, December 2014. doi:10.1515/9783110312812.115.
- [FVO+20] Franco W.; Viktor, C.; Oliveira, A. et al: Ontology-based Question Answering Systems over Knowledge Bases: A Survey:. In Proceedings of the 22nd International Conference on Enterprise Information Systems, 532–539. Prague, Czech Republic, 2020. doi:10.5220/0009392205320539.
- [Gro05] W3C Working Group. Representing Specified Values in OWL: „Value Partitions“ and „Value Sets“. <https://www.w3.org/TR/swbp-specified-values/>, May 2005. (2022-06-25)
- [Gru95] Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing? International Journal of Human-Computer Studies, 43(5-6):907–928, November 1995. doi:10.1006/ijhc.1995.1081.
- [Her14] Herman, I.: Owl-RL, A Simple Owl2 RL Reasoner On Top Of Rdflib. Zenodo, October 2014. doi:10.5281/ZENODO.14543.
- [HMPR04] Hevner, A.R.; March, S.T.; Park, J.; Ram, S.: Design Science in Information Systems Research. MIS Quarterly, 28(1):75–105, March 2004.
- [IMMS13] Iqbal, R. et al: An Analysis of Ontology Engineering Methodologies: A Literature Review. Research Journal of Applied Sciences, Engineering and Technology, 6(16):2993–3000, September 2013. doi:10.19026/rjaset.6.3684.
- [KLW95] Kifer, M.; Lausen, G.; Wu, J.: Logical foundations of object-oriented and frame-based languages. Journal of the ACM, 42(4):741-843, July 1995. doi:10.1145/210332.210335.
- [Kra02] Krathwohl, D.R.: A Revision of Bloom’s Taxonomy: An Overview. Theory into Practice, 2002. doi:10.1207/s15430421tip4104_2.
- [MNN20] Mkhinini, M.M. et al.: Combining UML and ontology: An exploratory survey. Computer Science Review, 35:100223, February 2020. doi:10.1016/j.cosrev.2019.100223.
- [Mus15] Musen, M.A.: The protégé project: a look back and a look forward. AI Matters, 1(4): 4–12, June 2015. doi:10.1145/2757001.2757003.
- [SB12] Sonnenberg, C.; vom Brocke, J.: Evaluations in the Science of the Artificial – Reconsidering the Build-Evaluate Pattern in Design Science Research. In Hutchison, D. et al: Design Science Research in Information Systems. Advances in Theory and Practice, volume 7286, pages 381–397. Springer 2012. doi:10.1007/978-3-642-29863-9_28.

- [SGL12] Suárez-Figueroa, M.G.; Gómez-Pérez, A.; Fernández-López, M.: The NeOn Methodology for Ontology Engineering. In Suárez-Figueroa, M.G. et al. (ed): *Ontology Engineering in a Networked World*, pages 9–34. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. doi:10.1007/978-3-642-24794-1_2.
- [SKOS] SKOS Simple Knowledge Organization System Primer. W3C Working Group Note 18 August 2009. <https://www.w3.org/TR/skos-primer/>. (2022-06-25)
- [SP21] Spoladore, D.; Pessot, E.: Collaborative Ontology Engineering Methodologies for the Development of Decision Support Systems. Case Studies in the Healthcare Domain. *Electronics*, 10(9):1060, April 2021. doi:10.3390/electronics10091060.
- [SSS04] Sure, Y.; Staab, S.; Studer, R.: On-To-Knowledge Methodology (OTKM). In Staab, S. and Studer, R. (ed): *Handbook on Ontologies*, pages 117–132. Springer, 2004.

Die Online-Dokumentation zu GenDifS unter <http://jbusse.de/gendifs/> beschreibt die hier verwendete Version 0.5 vom Sommer 2022. Das Handout zum GenDifS-Talk auf der Tagung AKWI 2022 findet sich unter <http://jbusse.de/gendifs/akwi2022.html>.

Konzeptionierung einer Softwareanwendung zur Verdachtsdiagnostik und Triage in der Pferdemedizin

Entwicklung eines algorithmischen Angebots zur Entlastung von Tierärzt*innen

Laura Haase¹, Lucas Rahn²

Abstract: Veterinärmediziner*innen stehen unter einer immensen Arbeitsbelastung, die eine zunehmende Triage bezüglich der zu behandelnden Tiere erfordert. Diese kann durch entsprechend gestaltete Softwareanwendungen unterstützt werden. Am konkreten Beispiel eines Systems für die Pferdemedizin zeigt der folgende Beitrag, welche Risiken durch eine solche Anwendung entstehen können und wie diesen begegnet werden kann. Außerdem wird die Gestaltung einer möglichen Wissensbasis als Grundlage des Systems beleuchtet und ein erster Algorithmus zur Datenanalyse gezeigt. Abschließend werden weiterführende Forschungsfragen definiert.

Keywords: Expertensystem, wissensbasiertes System, regelbasiertes System, Diagnosesystem, Triage software, Medizinsoftware, Veterinärmedizin, Pferdemedizin, Telemedizin

1 Einleitung

Ende 2019 versorgten in Deutschland 1600 Tierärzt*innen über 1,2 Millionen Pferde allein in Privatbesitz (vgl. [Bu20a], [Za21]). Neben der tiermedizinischen Betreuung zu Sprechstundenzeiten ist dabei auch die Sicherstellung der Versorgung an Wochenenden, Feiertagen und nachts zu gewährleisten. Bei gleichzeitig zum Teil mangelhafter Zahlungsbereitschaft der Kund*innen wird dieser Arbeitsumfang zu einer kaum leistbaren Aufgabe. Dadurch beenden vermehrt Tierärzt*innen ihre reguläre Teilnahme am Notdienestring und Tierkliniken geben ihren Klinikstatus ab, um keine Rund-um-die-Uhr-Versorgung mehr anbieten zu müssen (vgl. [Ti18]). Vor diesem Hintergrund nimmt die Relevanz der Triage in der Veterinärmedizin kontinuierlich zu. So wird beispielsweise im pferdemedizinischen Bereich durch die betreuende Tierarztpraxis meist zunächst eine telefonische Einschätzung des Gesundheitszustands des betroffenen Tieres durchgeführt, um dieses sinnvoll in die Behandlungsreihenfolge der nächsten Stunden oder Tage einzugliedern. Diese Aufgabe ist jedoch, abhängig vom Vorwissen der anrufenden Person, für die Tierärzt*innen zum Teil sehr zeitintensiv. Eine für Endanwender*innen konzipierte Softwareanwendung, die

1 Hochschule für Wirtschaft und Recht Berlin, Fachbereich Duales Studium, Studiengang Informatik, Alt-Friedrichsfelde 60, 10315 Berlin, Laura.Haase@hwr-berlin.de

2 Hochschule für Technik und Wirtschaft Berlin, Fachbereich Informatik, Studiengang Angewandte Informatik, Wilhelminenhofstraße 75A, 12459 Berlin, Lucas.Rahn@student.htw-berlin.de

in der Lage ist, anhand einer algorithmisch angeleiteten Eingabe der auffälligen Gesundheitsindikatoren der Tiere eine Verdachtsdiagnose und eine Einschätzung der Dringlichkeit der medizinischen Betreuung zu liefern, könnte hier Abhilfe schaffen und Tierärzt*innen wie -kliniken entlasten. Eine Softwarelösung ohne Öffnungszeiten und zeitliche Abrechnungstarife könnte dabei auch diejenigen Tierhalter*innen erreichen, die Unsicherheit über die Notwendigkeit eines Tierarztkontakts verspüren, aus verschiedenen Gründen jedoch zur Abklärung nicht zur Wahrnehmung eines kostenpflichtigen Termins z.B. bei einem Online-Sprechstunden-Anbieter bereit sind (vgl. [De20], [Tr20]).

Eine repräsentative Befragung des Bitkom e.V. und der Bayerischen TelemedAllianz ergab bereits 2017 ein hohes Interesse der Bevölkerung an der Nutzung von Gesundheitsapps. Die Nutzungsbereitschaft diesen gegenüber liegt dabei deutlich vor der Nutzungsbereitschaft gegenüber anderen digitalen Gesundheitsangeboten (vgl. [Ma17]). Untersuchungen der Präferenzen speziell von Tierbesitzer*innen bestätigen dieses Interesse auch innerhalb des veterinärmedizinischen Bereichs (vgl. [KSS19]).

Im Folgenden soll die Umsetzbarkeit, sowie die notwendigen Rahmenbedingungen einer mobilen App zur Auswertung von durch Pferdehalter*innen eingegebene Gesundheitsindikatoren ihrer Tiere untersucht und beleuchtet werden. So werden mögliche Nutzungsrisiken und ein für ein solches System verwendbares Wissenskonzept mit seinen Stärken und Schwächen beschrieben und eine erste Untersuchung der Genauigkeit von Entscheidungen auf der gewählten Wissensbasis vorgenommen. Abschließend wird ein Ausblick zur Bedeutung der produzierten Ergebnisse gegeben.

2 Rechtliche Aspekte

Zunächst soll ein kurzer Überblick über die rechtliche Situation gegeben werden, innerhalb derer sich eine Softwareanwendung zur Diagnoseunterstützung von Tieren befindet. Für fachfremde Unternehmer*innen, also solche ohne veterinärmedizinischen Hintergrund, existieren derzeit keine expliziten gesetzlichen Regulierungen hinsichtlich der Bereitstellung telemedizinischer Angebote für Tierhalter*innen. Rechtliche Rahmenbedingungen aus dem Bereich der Humanmedizin finden aufgrund der gesetzlich deutlich abgesetzten Unterscheidung zwischen Mensch und Tier keine Anwendung (vgl. [Te20]). Als eine Orientierung können derzeit die Empfehlungen der Arbeitsgemeinschaft Telemedizin der Bundestierärztekammer (nachzulesen in [De20]) zur Sicherung hoher Ethikstandards und zum Schutz der Kund*innen dienen. Innerhalb derer ist beispielsweise definiert, dass ohne klinische Untersuchung lediglich die Äußerung einer Verdachtsdiagnose möglich ist. Wenn die Festlegung einer abschließenden klinischen Diagnose durch erfahrene Veterinärmediziner*innen nicht über ein digitales Tool möglich ist, sollte gleichzeitig nicht angenommen werden, dies zeitnah durch ein Softwareprodukt allein ermöglichen zu können. So sollte auch eine entsprechende Anwendung die Nutzenden in jedem Fall auf die Limitierungen der getroffenen Aussagen hinweisen (vgl. [Bu20b], [Za21], [Mö21]).

3 Nutzungsrisiken

Die Erhebung der Gefährdungen, die potenziell durch eine Anwendung zur Verdachtsdiagnostik und Triage von Pferden entstehen könnten, umfasste mehrere Aspekte. So wurden mögliche Anwendungsrisiken aufgrund der Nutzung im Stressfall (der Situation am Stall eines potenziell erkrankten Tieres), möglicher Softwarefehler und Auswirkungen potenziell mangelnder Nutzendenfreundlichkeit aufgelistet. Außerdem wurden, mittels einer Onlineumfrage, Pferde-behandelnde Veterinärmediziner*innen hinsichtlich vorstellbarer Risiken befragt. Die Umfrage umfasste, nebst Angaben zur Berufserfahrung der Tierärzt*innen, verschiedene Fragen zu Häufigkeit und Auswirkungen von Selbstdiagnosen durch Pferdebesitzer*innen in der Praxis und Kenntnisse über existierende Anwendungen mit einem ähnlichen Ziel wie vorliegend beschrieben. Außerdem wurden Angaben zu möglichen Vor- und Nachteilen einer Anwendung zur Verdachtsdiagnostik und Triage, insbesondere auch im Verhältnis zu anderen bestehenden Möglichkeiten wie einer Internetsuche, erfasst. Die Umfrage wurde per E-Mail und über Nachrichten-Apps direkt an Tierärzt*innen versendet, sowie mehrere Landestierärztekammern um eine Weiterleitung an ihre Mitglieder gebeten. Die Umfrage war für 21 Tage, vom 14. Juni bis zum 05. Juli 2021 geöffnet und wurde mittels des Onlinetools LamaPoll erstellt. Insgesamt wurden 15 Fragen gestellt, davon 7 Fragen zur freien textuellen Beantwortung und 8 Fragen mit Antwortauswahlmöglichkeiten.

An der Umfrage nahmen 27 Pferde-behandelnde Tierärzt*innen teil, die eine Berufserfahrung von mindestens zwei und zu knapp 60% (16) mehr als 10 Jahren hatten. Durchschnittlich wurden sieben zu behandelnde Pferde pro Tag angegeben, wodurch eine hohe fachliche Feldexpertise der Teilnehmenden angenommen werden kann. 26 der 27 Befragten gaben an, Selbstdiagnosen durch Pferdehalter*innen in der Praxis zu erfahren. Davon schätzen 89% (23) das Vorkommen der Selbstdiagnosen als „häufig“ oder „sehr häufig“ ein. 17 der Tierärzt*innen zeichneten anschließend ein deutlich negatives Bild der Auswirkungen und wiesen unter anderem auf Zeitverzögerungen in der tierärztlichen Konsultation und damit medizinischen Behandlung der Tiere hin, die je nach Krankheitsbild zum Teil fatale Folgen nach sich ziehen können. Auch die Erzeugung eines falschen Gefühls der Informiertheit der Pferdehalter*innen und daraus resultierend sehr kritischen Begegnung der tierärztlichen Einschätzungen wurden in diesem Kontext genannt. Ein*e Teilnehmende*r bewertete die Auswirkungen einer zuvor durchgeführten Internetrecherche zu bemerkten Gesundheitsindikatoren des Pferdes ausschließlich positiv und verwies auf eine bessere Grundlage für Gespräche im Rahmen der Untersuchung. Acht der Tierärzt*innen äußerten sich neutral oder ausgewogen zu den Auswirkungen von Selbstdiagnosen und verwiesen dabei insbesondere auf die Abhängigkeit der Auswirkungen vom individuellen Vorwissen der jeweiligen Person. Bereits auf dem Markt verfügbare Anwendungen zur zielgerichteten Unterstützung der Pferdehalter*innen in der Entscheidung hinsichtlich medizinischer Versorgung waren keinem*keiner der Teilnehmenden bekannt. Hinsichtlich möglicher Nutzungsrisiken einer Softwareanwendung für Pferdehaltende zur Verdachtsdiagnostik und Triage nannten knapp 50% (13) der Befragten die Sorge, dass die Software auf Basis der Eingaben der Nutzer*innen zu falschen Rückschlüssen kommen und daher eine falsche Empfehlung zur Dringlichkeit der tierärztlichen Konsultation geben könnte. Sechs Antworten spiegelten die Befürchtung wider, dass die Eingaben der Nutzer*innen unzurei-

chend oder falsch sein könnten, wodurch auch ein korrekt arbeitendes System keinen korrekten Output liefern könnte. Vier Tierärzt*innen äußerten außerdem die Bedenken, dass eine durch eine Softwareanwendung gestellte Verdachtsdiagnose die Wahrscheinlichkeit einer Selbsttherapie durch die Pferdehalter*innen erhöhen könnte. 63% (17) der befragten Veterinärmediziner*innen schätzen die Risiken einer spezialisierten Softwareanwendung dabei jedoch nicht höher ein, als die einer Internetrecherche. Als Hauptargument der Einschätzenden eines höheren Risikos der Anwendung wurde das durch eine spezialisierte Software erzeugte höhere Sicherheitsgefühl hinsichtlich der Verdachtsdiagnose genannt. In diesem Zuge wurde die Notwendigkeit betont, Nutzer*innen noch einmal deutlich zu machen, dass die Behandlungshoheit ausschließlich Tierärzt*innen obliegt.

Um potenziell entstehenden Risiken angemessen zu begegnen, müssen geeignete Risikobeherrschungsmaßnahmen definiert werden. Dazu gehören insbesondere für Nutzer*innen deutlich sichtbare Informationen hinsichtlich der Grenzen der Softwareanwendung und der Gefahren eigenmächtiger Behandlungen. Zusätzlich sollten verschiedene Gestaltungsoptionen für die Anwendung getestet werden, die die Qualität der Nutzer*inneneingaben verbessern können, indem die Nutzenden explizit in der Erfassung der Gesundheitsindikatoren angeleitet werden. Die im System enthaltenen Informationen sollten dabei dauerhaft durch Fachexpert*innen überwacht werden. Zusätzlich sollten Empfehlungen hinsichtlich der Triage nicht ausschließlich von der gestellten Verdachtsdiagnose, sondern insbesondere auch von dem Vorhandensein kritischer Gesundheitsindikatoren abhängig gemacht werden.

4 Wissenskonzept

4.1 Überblick

Als Grundlage einer Softwareanwendung zur medizinischen Unterstützung von Pferdehalter*innen muss zunächst eine Wissensbasis geschaffen werden. Wissen entspricht dabei einer Beschreibung von Objekten und deren Eigenschaften, in diesem Fall insbesondere der Pferdekrankheiten, die auf mit Semantik angereicherten, vernetzen Daten beruht (vgl. [Ha10]). Das hier fokussierte medizinische Wissen stellt dabei aufgrund seines immensen Umfangs und der damit einhergehenden schweren Überblick- und Dokumentierbarkeit eine besondere Wissensform dar. In der Medizin berufen sich Fachexpert*innen im Rahmen ihrer Entscheidungen häufig, neben explizit in Fachliteratur formulierbaren Informationen, auf eigenständig gesammeltes, implizites Erfahrungswissen. Dieses kann nur schwer ausdrücklich erfragt und benannt werden, sodass auch eine Abbildung dessen in durch Software verarbeitbarer Form nur begrenzt möglich ist (vgl. [SS08], [SRN17], [Pu90]).

Speziell im Bereich der Pferdemedizin existiert zusätzlich die Problematik, dass keine statistisch verlässlichen und aktuellen Aussagen zu Krankheitshäufigkeiten aufzufinden sind.

Angaben in der Literatur oder auf Websites enthalten meist nur unscharfe Begriffe wie „häufig“ oder „selten“ und konnten auf stichprobenartige Nachfrage nicht statistisch verlässlich belegt werden. Einzelne umfangreichere Untersuchungen wie die National Equine Health Survey des Royal Veterinary College und der Vereinigung britischer Tierärzte [Na18] geben lediglich einen allgemeinen Überblick über den Gesundheitsstatus der untersuchten Pferde, sind jedoch für die Nutzung innerhalb eines Expertensystems nicht ausreichend präzise aufgeschlüsselt. Zumeist werden im Rahmen existierender Erhebungen auch nur Informationen zu einer geringen Anzahl Krankheiten, einem konkreten Krankheitsbild oder sehr groben Krankheitsgruppierungen, wie „Erkrankungen des Bewegungsapparats“ abgebildet (vgl. [Je21], [Ha18], [Le21], [Na18]).

4.2 Mögliche Datenquellen

Für die vorliegend betrachtete Softwareanwendung kommen grundsätzlich mehrere verwendbare Datenarten in Frage. Es können beispielsweise in der Fachliteratur verfügbare Informationen zu Krankheiten verwendet werden, alternativ ist auch die Nutzung von Felddaten, die von Tierärzt*innen im Rahmen der Klinik- oder Praxisdokumentation erfasst werden, möglich. Auch kann mithilfe entsprechend konzipierter Befragungen versucht werden, das tierärztliche Wissen zu dokumentieren, wie beispielsweise von Qin et al. [Qi16] beschrieben. Dabei existieren je nach Möglichkeit verschiedene Vor- und Nachteile.

Der Ansatz der Dokumentation vorhandenen Wissens menschlicher Expert*innen verspricht den Vorteil, dass, zumindest theoretisch, Zugriff auf das gesamte Wissen der Tierärzt*innen möglich ist, das von diesen im Alltag eingesetzt wird. Limitiert wird dies jedoch durch die bereits beschriebene hohe Schwierigkeit und den hohen Zeitaufwand dieses implizit vorliegende Wissen in explizites, dokumentiertes Wissen zu überführen. Hierbei ist von einer hohen Fehleranfälligkeit auszugehen. Auch die Vollständigkeit der Übertragung ist in Frage zu stellen (vgl. [SS08]). Der Nutzung von Informationen aus Fachliteratur ist in dieser Hinsicht der Vorzug zu geben, da es sich dabei um bereits explizit formuliertes, vor Veröffentlichung mehrfach geprüftes Wissen handelt. Die Fachliteratur bildet zudem die Grundlage für das veterinärmedizinische Studium und somit auch die Grundlage des Wissens ausgebildeter Tierärzt*innen. Seltene Krankheitsbilder sind hierin beschrieben. Felddaten enthalten wiederum im Gegensatz dazu einen deutlich erhöhten Anteil subjektiver Eindrücke, bilden jedoch zugleich über das Grundlagenwissen hinaus auch praktische Erfahrungen mit ab. So enthalten sie beispielsweise Symptomintensitäten und Häufigkeiten des Auftretens bestimmter Krankheiten. Gleichzeitig können jedoch zu einem gewissen Anteil Fehldiagnosen oder Verfälschungen von Informationen durch persönliche Spezialgebiete und der damit einhergehenden Häufung der Behandlung bestimmter Krankheitsbilder vorhanden sein. Seltene Krankheitsbilder können zudem in Praxisdokumentationen unter Umständen nicht abgebildet sein (vgl. [Ma20], [SS08]).

4.3 Auswahl der Datenquelle(n)

Die Beleuchtung der Vor- und Nachteile der einzelnen Datenquellen zeigt, dass keiner hiervon ein eindeutiger Vorzug vor den jeweils anderen gegeben werden kann. In anderen Veröffentlichungen bezüglich diagnoseunterstützender Systeme wird häufig auf die Auswertung von Felddaten oder auf die Nutzung von Wissen aus Expert*innenbefragungen gesetzt (vgl. z.B. [AQR15], [Ca20], [FP17], [Ni17], [Si19], [Su18], [Qi16], [Ud19]). Diesem Beispiel folgend, mit dem Ziel der Ermöglichung der Nutzung impliziten Wissens, wurden ebendiese Ansätze auch für die vorliegend beschriebene Softwareanwendung ausgewählt. Um über entsprechende Datensätze zu verfügen, wurden diverse veterinärmedizinisch tätige Personen und Organisationen kontaktiert. Der auf die Kontaktaufnahme folgende Rücklauf war gering, und verneinte zum überwiegenden Teil die Möglichkeit der Unterstützung des Vorhabens. Als Gründe wurden meist Datenschutzbedenken, fehlende Personalkapazitäten zur Durchführung von Befragungen und Anonymisierung der vorhandenen Datensätze und das Vorliegen der Dokumentation in einer für die Verarbeitung unzureichenden Form genannt. Einige Einrichtungen gaben an, Vorhaben wie das angefragte nach Umstellung der eigenen IT-Infrastruktur oder bei entsprechender Unterstützung in der Datenerfassung zukünftig gegebenenfalls unterstützen zu wollen.

Da auf Basis der erhaltenen Rückmeldungen die Verwendung von Felddaten und die Durchführung von Befragungen nicht länger zielführend umsetzbar erschien, wurde die Vorgehensweise anschließend neu ausgerichtet. Für die Weiterarbeit wurde stattdessen ein Literatur-basierter Ansatz, unter Herausarbeitung von dessen Vorteilen, ausgewählt. Zu diesem Zwecke wurde das von diversen Expert*innen deutscher veterinärmedizinischer Universitäten erstellte Handbuch Pferdepraxis [Br16] herangezogen. Die in Textform vorliegenden Informationen mussten folgend in eine geeignete, durch Algorithmen möglichst fehler- und verlustfrei verarbeitbare Form überführt werden. Aus ebendem Grund der Fehler- und Verlustfreiheit wurde auf die Verwendung eines automatischen Textextraktionstools verzichtet und stattdessen eine manuelle Extraktion von für relevant befundenen Daten durchgeführt. Zu letzteren zählten neben dem Namen der Krankheit insbesondere Leit- und weitere Symptome, falls anwendbar eine minimale Symptomdauer für die Möglichkeit des Vorliegens z.B. einer chronischen Krankheit und Prädestinationen bestimmter Geschlechts-, Alters- oder Rassegruppierungen. Auf diesem Wege wurden 136 Krankheiten verschiedener Auftretenshäufigkeiten für die Weiterverarbeitung extrahiert.

4.4 Datenanalyse

Die eben beschriebenen Datensätze stellen in ihrem logischen Kontext verknüpfte WENN-DANN Regeln dar, welche häufig in Expertensystemen verwendet werden (vgl. [SS08], [SRN17]). Beispielsweise kann folgende Aussage getroffen werden: „WENN das Pferd Husten zeigt UND/ODER das Pferd Leistungsminderung zeigt UND/ODER das Pferd Atembeschwerden zeigt UND/ODER das Pferd Atemgeräusche zeigt UND/ODER ... UND die Symptome seit mindestens 6 Wochen bestehen, DANN sollte als Diagnose eine

Chronisch obstruktive Bronchitis in Betracht gezogen werden.“ Es ist zu beachten, dass es sich bei diesen Angaben naturgemäß nicht um abschließende GENAU-DANN-WENN-Aussagen handeln kann, da einerseits gleiche Symptome bei verschiedenen Krankheiten auftreten können und andererseits nicht alle Symptome bei jedem Krankheitsfall auftreten müssen. Da aufgrund dieser Einschränkungen mehrere passende Verdachtsdiagnosen existieren können, musste eine Möglichkeit gefunden werden, diese der Wahrscheinlichkeit des Zutreffens nach zu sortieren. Dazu wurden mehrere Bepunktungssysteme für die Wahrscheinlichkeit des Zutreffens einer Diagnose getestet. Der als am besten bewertete Ansatz (für das Bewertungsverfahren siehe dem folgenden Abschnitt 4.5) addiert dabei auf die Wertung jeder Diagnose pro aufgetretenem Leitsymptom 10 Punkte und pro aufgetretenem weiteren Symptom 5 Punkte. Wurde durch den*die Nutzer*in ein Gesundheitsindikator eingegeben, der keinen Zusammenhang zur Krankheit aufweist, werden hierfür jeweils 5 Punkte abgezogen. Die Zahlen von 5 und 10 Punkten wurden gewählt, um einen deutlichen Unterschied in der Gewichtung der einbezogenen Informationen schaffen zu können, gleichzeitig jedoch Spielraum für zukünftige Erweiterungen des Algorithmus um geringgradiger relevante und damit geringer bepunktete Aspekte zu schaffen.

Durch die Subtraktion von Punkten bei Nichtverknüpfung von Symptomen können differentialdiagnostisch deutlichere Unterschiede ausgemacht werden, als bei einer ausschließlichen Abbildung der Punktzahlen verknüpfter Symptome. Zusätzlich zur symptomatischen Betrachtung wurden auch weitere Informationen über das betreffende Pferd miteinbezogen. So werden beispielsweise jeweils 10 Punkte zur Wertung einer Diagnose addiert, wenn das Pferd zu einer Rasse gehört, für die dieses Krankheitsbild typisch ist. Da aufgrund diverser Rassenvermischungen in modernen Pferdezüchtungen rassespezifische Merkmale immer weiter an Bedeutung verlieren, wird mithilfe einer Addition des gleichen Wertes wie im Fall eines Leitsymptoms einer erhöhten Wahrscheinlichkeit Rechnung getragen. Liegt das Alter des Pferdes nicht in der für eine Diagnose typischen Altersspanne, werden 10 Punkte subtrahiert, um die verringerte Wahrscheinlichkeit des Krankheitsbildes abzubilden. Ebenso wird bei Vorliegen eines für die Krankheit untypischen Geschlechts verfahren. Existieren Krankheiten nur bei konkreten Geschlechtern, werden diese für andere Merkmalsausprägungen vollkommen ausgeschlossen. Gleiches gilt für die erfasste Mindestdauer des Vorliegens der Gesundheitsindikatoren (siehe Abb. 1).

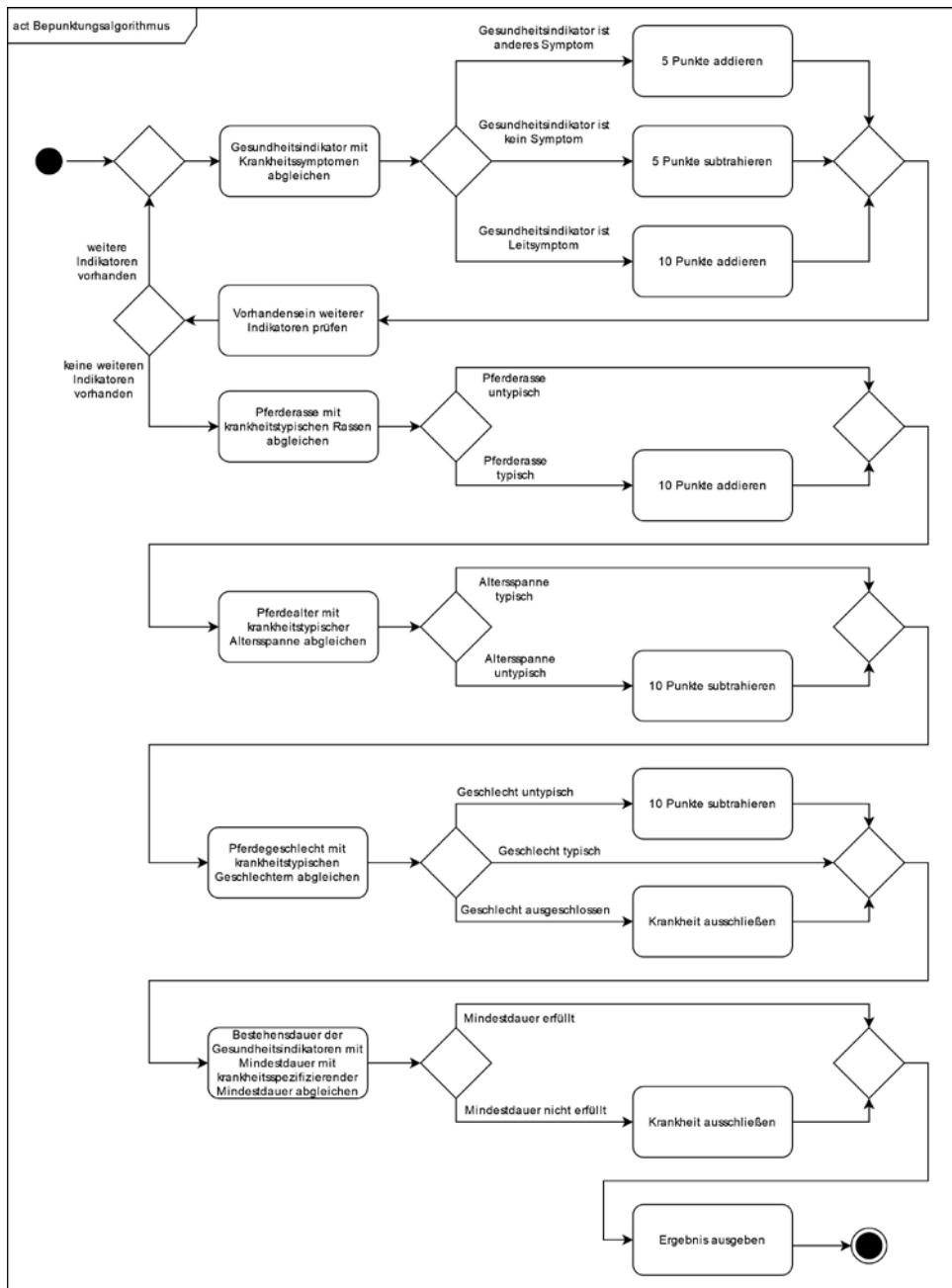


Abb. 1: Flussdiagramm des Bepunktungsalgorithmus zur Auswertung eingeggebener Gesundheitsindikatoren

4.5 Systemgenauigkeit

Für die Untersuchung und den Vergleich der getesteten Auswertungsverfahren wurde das Testverfahren mit Vignetten unter Laborbedingungen, wie beispielsweise von Gilbert et al. [Gi20] beschrieben, ausgewählt. Der Vorteil des Verfahrens besteht darin, dass reproduzierbare Testergebnisse erzeugt und verschiedene Eingabevarianten simuliert werden können. Jede der Testvignetten enthält dabei die betrachteten Pferdeinformationen, alle Symptome in unsortierter Reihenfolge und die korrekte Diagnose für den jeweiligen Fall. Es soll so überprüft werden, ob der genutzte Algorithmus aus allen aufgrund der WENN-DANN Regeln in Frage kommenden Krankheiten die korrekte Diagnose als die wahrscheinlichste ermitteln kann. Gemeinsam mit einer Fachtierärztin für Pferde wurden 11 explorative Testvignetten entwickelt, deren Diagnosen in der verwendeten Datenbank enthalten waren. Die Ergebnisse der überprüften Algorithmen wurden in drei Kategorien unterteilt:

1. Die korrekte Diagnose wurde als wahrscheinlichste Diagnose ermittelt.
2. Die korrekte Diagnose wurde als eine der fünf wahrscheinlichsten Diagnosen ermittelt.
3. Die korrekte Diagnose wurde nicht als eine der fünf wahrscheinlichsten Diagnosen ermittelt.

Bei einer Eingabe aller in der jeweiligen Testvignette enthaltenen Symptome zeigte der gewählte Auswertungsalgorithmus ein sehr gutes Ergebnis von 100% Auswertungen in der Kategorie 2. Rund 90% davon in Kategorie 1.

Es muss jedoch davon ausgegangen werden, dass reale Nutzer*innen nicht alle Krankheitsanzeichen ihres Pferdes erkennen würden. Aus diesem Grund wurde der Test erweitert und um jede Variante der Eingabe (nur eines der Symptome, zwei der Symptome, drei der Symptome, ...) ergänzt. Jede dieser Eingabevarianten wurde erneut dem Algorithmus zur Auswertung eingegeben. Unter diesen Bedingungen ermittelte der beste Algorithmus in 61% der Fälle ein Ergebnis der Kategorie 1. In 84% lag das Auswertungsergebnis in Kategorie 2. Die so ermittelte Genauigkeit kommt der einer realen Auswertung aufgrund der Unvollständigkeit vieler Eingaben deutlich näher (Auswertung durch den entwickelten Prototyp siehe Abbildung Abb. 2).



Abb. 2: Auswertung einer Nutzer*inneneingabe durch den entwickelten Prototyp

Das erzielte Ergebnis muss dabei dennoch unter Berücksichtigung verschiedener Rahmenbedingungen betrachtet werden. So handelte es sich bei dem genutzten Vignettensatz um eine sehr kleine Auswahl an Testfällen, die unter Umständen zufällig eine sehr aussagekräftige oder einzigartige Symptomatik zeigen können. Parallel dazu enthielt die Datenbank zum Zeitpunkt der Durchführung der explorativen Tests wie bereits beschrieben 136 Krankheiten. Unter Umständen waren in Frage kommende Differentialdiagnosen darin nicht abgebildet und erleichterten damit dem Algorithmus die Auswahl der korrekten Diagnose.

5 Fazit und Ausblick

Im Rahmen der Untersuchung der Anwendung zur Ermittlung von Verdachtsdiagnosen für Pferde und Unterstützung der Pferdehalter*innen bei einer Entscheidung über den Zeitpunkt des Hinzuziehens von Tierärzt*innen (Triage) zeigte sich, dass für Fachexpert*innen insbesondere die Aufklärung über die Grenzen eines solchen Softwareprodukts gegenüber den Nutzer*innen relevant ist. Risiken der Nutzung einer entsprechenden Anwendung und passende Maßnahmen zur Risikobeherrschung wurden definiert. Es ist bei der Umsetzung

der Maßnahmen zu beachten, dass sich die befragten Tierärzt*innen die ausschließlich textuell beschriebene Softwarelösung möglicherweise anders vorgestellt haben können, als intendiert war. Eine Validierung der ermittelten Risiken und Risikobeherrschungsmaßnahmen anhand einer umgesetzten Reallösung kann daher sinnvoll sein. Die Verfügbarkeit von Datensätzen aus dem Feld als Grundlage der Forschung an dem pferdemedizinischen System stellte sich während der Recherche als wenig zufriedenstellend dar. Mithilfe des in Folge dessen gewählten Literatur-basierten Datensatzes konnten in einem explorativen Testlauf dennoch gute erste Resultate der Auswertung von Krankheitssymptomen erzielt werden. Diese gilt es zukünftig auch in einem größeren Rahmen, zum einen hinsichtlich der Testfälle, zum anderen hinsichtlich des Umfangs enthaltener Krankheiten, zu verifizieren. Der verwendete rudimentäre Algorithmus sollte dahingehend weiterentwickelt werden, dass er auch den prozentualen Anteil vorliegender Symptome einer Krankheit miteinbezieht. So können auch Krankheiten mit einer geringen Anzahl charakterisierender Symptome sinnvoll abgebildet werden. Auch das Vorliegen mehrerer Krankheitsbilder parallel sollte zukünftig bei der Auswertung in Betracht gezogen werden. Es ist zu überprüfen, ob eine linear steigende Wahrscheinlichkeit für das Vorliegen einer Krankheit realistisch ist, oder ob beispielsweise ein exponentieller Ansatz gewählt werden muss. Bei einer signifikanten Erhöhung der im System gespeicherten Krankheitsbilder sollte zudem über eine sinnvolle Unterscheidung zwischen selten und häufig auftretenden Krankheiten nachgedacht und ein zielführendes differentialdiagnostisches Vorgehen festgelegt werden. Dazu sollte auch eine Erweiterung der zugrundeliegenden Datensätze von ausschließlich Literatur-basierten Daten hin zu vielfältigeren, heterogenen Datenquellen mit einer geeigneten algorithmischen Verknüpfung in Betracht gezogen werden. Ein besonderer Fokus sollte zukünftig auch auf die Verfeinerung der durch das System durchgeföhrten Triage gelegt werden. Mithilfe der oben beschriebenen Dringlichkeitsermittlung kann anhand von Notfallsymptomen und -krankheiten zunächst weitgehend sichergestellt werden, dass Fälle keine zu geringgradige Einstufung erfahren. Eine weiterführende Prozessoptimierung mit zusätzlichen von den Nutzer*innen abgefragten Informationen kann jedoch noch eine deutlich präzisere Einstufung ermöglichen.

Durch ein System wie das beschriebene könnte eine Unterstützung für Pferdehalter*innen und Tierärzt*innen bereitgestellt werden, die potenziell in der Lage ist, die Leidensdauer der Pferde bis zur Diagnosestellung und Behandlung, auch in Fällen seltener Krankheitsbilder, zu verkürzen. Neben einer emotionalen und finanziellen Entlastung für die Pferdebesitzer*innen kann gegebenenfalls auch ein Beitrag zur Entlastung der Veterinärmediziner*innen geleistet werden, indem die Patientenbetreuenden gezielte Vorinformation erfahren und in der Einzelfallpriorisierung unterstützt werden. Der entwickelte Prototyp zeigt vielversprechende Ansätze zur Umsetzung dieser Vorteile, wenngleich weitere Entwicklungsschritte und Untersuchungen notwendig sind, um eine abschließende Aussage treffen zu können.

Literaturverzeichnis

- [AQR15] Ahmad, P.; Qamar, S.; Rizvi, S. Q. A.: Techniques of Data Mining in Healthcare: A Review. International Journal of Computer Applications 120(15), S. 38–50, 2015.
- [Br16] Brehm, W.; Gehlen, H.; Ohnesorge, B.; Wehrend, A.: Handbuch Pferdepraxis, 4. Auflage, Thieme, 2016.
- [Bu20a] Bundestierärztekammer e.V.: Statistik 2019: Tierärzteschaft in der Bundesrepublik Deutschland. Deutsches Tierärzteblatt 07/20, S. 860-870, 2020.
- [Bu20b] Bundesverband praktizierender Tierärzte e.V.: Telemedizin und Künstliche Intelligenz - Chancen der Digitalisierung nutzen. Jahresbericht 2020, 1:28, 2020.
- [Ca20] Caballé, N.; Castillo-Sequera, J; Gómez-Pulido, J; Gómez-Pulido, J; Polo-Luque, M.: Machine Learning Applied to Diagnosis of Human Diseases: A Systematic Review. Applied Sciences 10(15), 5135, 2020.
- [De20] Der Einzug der Telemedizin in die Tiermedizin, <https://bundangestellttertieraerzte.de/der-einzug-der-telemedizin-in-die-tiermedizin/>, Stand 10.06.2021.
- [FP17] Fatima, M.; Pasha, M.: Survey of Machine Learning Algorithms for Disease Diagnostics. Journal of Intelligent Learning Systems and Applications 9(1), S.1–16, 2017.
- [Gi20] Gilbert, S.; Mehl, A.; Baluch, A.; Cawley, C.; Challiner, J.; Fraser, H.; Millen, E.; Montazeri, M.; Multmeier, J.; Pick, F.; Richter, C.; Türk, E.; Upadhyay, S.; Virani, V.; Vona, N.; Wicks, P.; Novorol, C.: How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. BMJ Open 10(12), e040269, 2020.
- [Ha10] Harrach, H.: Risiko-Assessments für Datenqualität, 1. Auflage, Vieweg+Teubner, 2010.
- [Ha18] Hartung, K.: Auswertung der Equiden-Sektionen im Institut für Veterinär-Pathologie der Universität Leipzig und Dresden von 1890 bis 2013. Inaugural-Dissertation zur Erlangung des Grades eines Doctor medicinae veterinariae (Dr. med. vet.) durch die Veterinär-medizinische Fakultät der Universität Leipzig, 2018.
- [Je21] Jedes vierte Pferd in Deutschland leidet an einer Atemwegserkrankung - Boehringer Ingelheim Vetmedica, <https://www.vetmedica.de/atemwege>, Stand: 03.07.2021.
- [KSS19] Kogan, L.; Schoenfeld, R.; Santi, S.: Medical Updates and Appointment Confirmations: Pet Owners' Perceptions of Current Practices and Preferences. Frontiers in Veterinary Science 6(80), 2019.
- [Le21] Lexikon :: Tierärztliche Praxis für Pferde :: Dr. med. vet. Inka Kreling, <https://www.tierarzt-inka-kreling.de/lexikon/detail.php?nr=174&rubric=Lexikon>, Stand 03.07.2021.
- [Ma17] Markt für Digital Health mit großem Wachstumspotenzial | Bitkom e.V., <https://www.bitkom.org/Presse/Presseinformation/Markt-fuer-Digital-Health-mit-grossem-Wachstumspotenzial.html>, Stand: 15.08.2021
- [Ma20] Malpede, B.; Dogan, G.; Moreland, S.; Cheheltani, R.; Fischer, B.; Lee, S.; Leavitt, N.; Doyle, O.; Rigg, J.: AI Algorithms for Disease Detection: Methodological Decisions for Development of Models Validated Through a Clinical, Analytical, and Commercial Lens. Pharmaceutical Management Science Association Journal, 08(3), 2020.
- [Mö21] Möllenbeck, S.: Telemedizin in der Tiermedizin. pan Bocholt, 1:14–15, 2021.

- [Na18] National Equine Health Survey (NEHS) 2018, <https://www.bluecross.org.uk/sites/default/files/downloads/NEHS-results-2018.pdf>, Stand 24.02.2021.
- [Ni17] Nilashi, M.; Ahmadi, H.; Shahmoradi, L.; Mardani, A.; Ibrahim, O.; Yadegaridehkordi, E.: Knowledge Discovery and Diseases Prediction: A Comparative Study of Machine Learning Techniques. *Journal of Soft Computing and Decision Support Systems* 4(5), S. 8–16, 2017.
- [Pu90] Puppe, F.: Problemlösungsmethoden in Expertensystemen, 1. Auflage, Springer, 1990.
- [Qi16] Qin, H.; Xiao, J.; Gao, X.; Wang, H.: Horse-Expert: An Aided expert system for diagnosing horse diseases. *Polish Journal of Veterinary Sciences* 19, S. 907–915, 2016.
- [Si19] Singh, A. K.: A Comparative Study on Disease Classification using Machine Learning Algorithms. In: 2nd International Conference on Advanced Computing and Software Engineering ICACSE-2019, S. 197–202, 2019.
- [SRN17] Styczynski, Z.; Rudion, K.; Naumann, A.: Einführung in Expertensysteme, 1. Auflage, Springer Vieweg, 2017.
- [SS08] Spreckelsen, C.; Spitzer, K.: Wissensbasen und Expertensysteme in der Medizin, 1. Auflage, Vieweg+Teubner, 2008.
- [Su18] Sunny, A. D.; Kulshreshtha, S; Singh, S.; Srinabh, M. B.; Sarojadevi H.: Disease Diagnosis System By Exploring Machine Learning Algorithms. *International Journal of Innovations in Engineering and Technology* 10(2), S. 14–21, 2018.
- [Te20] Telemedizin (II): Netflix-Mentalität der Tierhalter?, <https://www.wir-sind-tierarzt.de/2020/05/telemedizin-iinetflix-mentalitaet-der-tierhalter/>, Stand 10.06.2021.
- [Ti18] Tierärztlicher Notdienst: Arbeitsrechtliche Zeitbombe mit Image-Sprengstoff, <https://www.wir-sind-tierarzt.de/2018/02/tierarzt-notdienst-arbeitszeit/>, Stand 10.06.2021.
- [Tr20] Triage in der Tiermedizin: Ein Blick hinter die Kulissen, <https://www.tierarzt-rueckert.de/blog/details.php?Kunde=1489&Modul=3&ID=21087>, Stand 10.06.2021.
- [Ud19] Uddin, S.; Khan, A.; Hossain, E.; Moni, M. A.: Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making* 19, 281, 2019.
- [Za21] Zahlen und Fakten aus Pferdesport und Pferdezucht | FN, <https://www.pferd-aktuell.de/deutsche-reiterliche-vereinigung/zahlen--fakten>, Stand: 03.03.2022.

Extraktion und Analyse von Schlüsselwörtern in einer Literaturrecherche zu Quantum Computing

Mazlum Copurkuyu¹, Thomas Barton²

Abstract: Durch die große Menge an wissenschaftlichen Publikationen, die meist als unstrukturierte Daten vorliegt, nehmen Komplexität und Arbeitsaufwand eines Literature-Review Prozesses stetig zu. Auch im Forschungsgebiet Quantum Computing hat sich die Anzahl wissenschaftlicher Veröffentlichungen in den letzten Jahren stark erhöht. Dieser Beitrag gibt einen Überblick, wie man Text-Mining-Methoden zur Informationsextraktion bei der Literaturrecherche zu Quantum Computing einsetzen kann. Das zentrale Forschungsziel besteht in der Anwendung von Text-Mining zur automatischen Extraktion und Visualisierung von Schlüsselwörtern auf Basis der Abstracts von wissenschaftlichen Publikationen. Dieser Ansatz verwendet zum einen die TF-IDF-Methode und auf der anderen Seite den Word2Vec-Algorithmus, um die automatische Erfassung sowie die Verarbeitung relevanter Literatur zu ermöglichen. Anschließend wird eine visuelle Darstellung der Ergebnisse wie z.B. dynamische Word-Clouds durchgeführt. Aus der Analyse werden Erkenntnisse für den Forschungsbereich Quantum Computing abgeleitet.

Keywords: Literaturrecherche, Text-Mining, Keyword Extraction, TF-IDF, Word2Vec, Quantum Computing

1 Einführung

Eine grundlegende Aufgabe in der Forschung besteht darin, die vorhandene Literatur durch ein Literatur-Review zu identifizieren und zu verstehen, um den Kontext zu ermitteln, weiterführende Forschung zu betreiben sowie die Forschungsgemeinschaft auf dem neuesten Stand zu halten [Ta20]. Zu diesem Zweck ist es zwingend notwendig, die vorhandene Literatur zu einem Forschungsthema zu berücksichtigen. Diese Aufgabe ist jedoch bei der ständig steigenden Anzahl von Veröffentlichungen kaum noch zu bewältigen, und ihre Auswertung gestaltet sich schwierig. Um das Problem der Informationsflut zu lösen, sollte der Prozess der Literaturrecherche strukturiert sein und etablierten Rahmenwerken folgen.

Einer der bekanntesten und weit verbreiteten Vorgehensweise zur Durchführung eines (manuellen) Literatur-Reviews ist das Framework vom Brocke et al. [vo09], das in Abb. 1 dargestellt ist.

¹ Hochschule Worms, Fachbereich Informatik, Erenburgerstr.19, 67549 Worms

² Hochschule Worms, Fachbereich Informatik, Erenburgerstr.19, 67549 Worms, barton@hs-worms.de, <https://orcid.org/0000-0001-6736-7040>

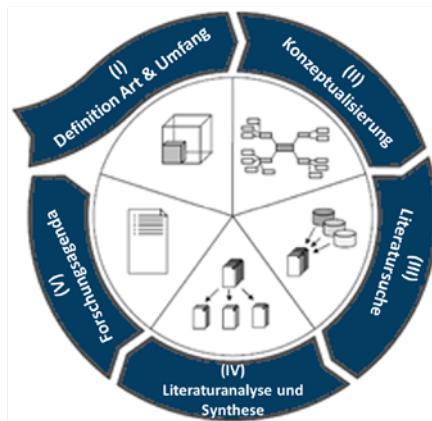


Abb. 1: Prozess zur Durchführung einer Literaturanalyse (vom Brocke et al. [vo09])

Die Durchführung eines Literaturreviews ist jedoch aufgrund der vielen manuellen Tätigkeiten wie Suchen und Herunterladen, Dokumentation des Prozesses, Textscreening usw. sehr zeitaufwändig und mühsam. Das Verfahren zur Durchführung von Literatur-Reviews zu unterstützen und zu automatisieren wäre sehr hilfreich, um mit der immer schnelleren Veröffentlichung von Artikeln Schritt halten zu können [Ta20]. Dabei gilt es zu prüfen, ob die zeitaufwendige Arbeit für die Auswertung sowie die Analyse von wissenschaftlichen Artikeln mit Hilfe von Text-Mining unterstützt werden kann.

In dieser Arbeit werden Methoden des Text-Mining eingesetzt, um Informationen aus wissenschaftlichen Publikationen zu generieren. Hierbei werden Verfahren der Keyword-Extraction und Methoden für Word-Embeddings zur Extraktion von Informationen aus den wissenschaftlichen Abstracts verwendet. Das Ziel ist es, Muster und Verbindungen zu verwandten Themen in einem Forschungsbereich zu ermitteln.

2 Grundlagen

Dieses Kapitel führt in das Quantum Computing und Text-Mining ein. Zuerst wird der Begriff Quantum Computing erläutert. Anschließend wird ein Überblick über Text-Mining gegeben und damit die für diesen Beitrag essenziellen Begriffe sowie Zusammenhänge vermittelt.

2.1 Quantum Computing

Die global vernetzte mobile Welt macht eine nachhaltige und effiziente Verarbeitung und Analyse von Informationen zu einer immer komplexeren Herausforderung. An dem Punkt, an dem klassische Prozessoren heute an ihre physikalischen Grenzen stoßen, könnten Quantencomputer in Zukunft Lösungen liefern. Diese basieren auf der Wechselwirkung quantenmechanischer Zustände [Ba20].

Ein Quantencomputer arbeitet nicht mit klassischen Bits, sondern mit Quantenbits - kurz Qubits. Sie können z.B. durch den Spin eines Elektrons gebildet werden. Die Funktion von Qubits in Quantencomputern basiert auf zwei Schlüsselprinzipien der Quantenphysik: Superposition und Verschränkung, die im Folgenden kurz erläutert werden [Ba20].

Superposition bedeutet, dass ein Qubit nicht nur den Zustand 1 oder 0 einnehmen kann, sondern beide „gleichzeitig“ sowie alle Zustände dazwischen, ähnlich wie eine sich drehende Münze [Ho18]. Solange sie sich bewegt, kann das Ergebnis sowohl Kopf oder Zahl werden. Erst wenn man sie stoppt und den Ausgang misst, wird das Ergebnis bestimmt [Ho18]. Durch die Messung wird dieser Quantenzustand, die Superposition, zerstört. Solange man den Zustand eines Qubits nicht misst, beschreibt man diesen als Superposition, also als Überlagerung aller möglichen Zustände [Ho18]. Der Zustand der Superposition wird durch die Messung und die damit verbundene Interaktion entweder zu 0 oder 1 kollabieren [Ba20].

Die Verschränkung ist der Schlüssel zur Effizienz von Quantencomputern. Verschränkung bedeutet, dass Qubits miteinander unter Wechselwirkungen stehen [Ba20]. Die Verschränkung funktioniert nur, wenn beide Qubits sich in Superposition befinden. Beeinflusst man eines, werden zeitgleich auch alle Partner-Qubits beeinflusst. Es besteht also eine Korrelation zwischen den Zuständen beider Qubits. Sobald der Zustand des einen gemessen wird, ist auch der Zustand des zweiten Qubits bekannt und so kollabiert die Superposition beider Qubits [Ba20]. Dieser Umstand gilt auch bei verschränkten Qubits, die voneinander räumlich getrennt sind. Es sei erwähnt, dass die Verschränkung nach der Messung der Qubits aufgehoben wird [Ho18].

Dadurch ist es möglich, in einem Quantencomputer parallele Berechnungen durchzuführen, aber der Zugriff auf die Ergebnisse der Berechnungen sind eingeschränkt. Der Zugriff auf die Ergebnisse ist gleichbedeutend mit einer Messung, wodurch der Quantenzustand gestört wird. Da die Messung probabilistisch ist, erhält man ein zufälliges Ergebnis [Ba20]. Zudem muss auch die Superposition der Qubits möglichst lange erhalten bleiben, also die Kohärenzzeit möglichst lang sein, denn selbst kleinste Zustandsänderungen verändern den Rechenvorgang. Weiterhin ist die Übertragung der Informationen so empfindlich, dass sie bereits durch das Rauschen der Elektronik gestört werden kann [Ba20]. Solche Fehlerquellen zu minimieren, ist eine spannende Herausforderung.

Die potenziellen Vorteile und Anwendungen des Quantum Computing für die Gesellschaft sind vielfältig. Richtig eingesetzt, sind Quantencomputer unglaublich schnell und effektiv.

Sie können in wenigen Sekunden Berechnungen durchführen, für die die heutigen klassischen Supercomputer viel länger brauchen würden. Diese Tatsache wird von Experten auch als Quantenüberlegenheit bezeichnet [Ba20]. Quantencomputer stellen einen Paradigmenwechsel in der Datenverarbeitung dar. Umso faszinierender ist es, dieses Gebiet zu verfolgen. Daher wird in dieser Arbeit dieses Forschungsgebiet mithilfe von Text-Mining-Verfahren genauer untersucht, um die Themenschwerpunkte, die bisherige Entwicklung sowie die zukünftigen Trends des Forschungsgebietes – sofern möglich – zu erfassen.

2.2 Überblick über Text-Mining

Text-Mining ist ein weit gefasster Oberbegriff, die eine Reihe von unterschiedlichen Verfahren zur Analyse und Verarbeitung von semistrukturierten und unstrukturierten Textdaten beschreiben. Das gemeinsame Ziel hinter jeder dieser Verfahren ist „Texte in Zahlen zu verwandeln“, damit leistungsstarke Algorithmen auf große Dokumenten-datenbanken angewendet werden können [Mi12]. Die Konvertierung von Text in ein strukturiertes numerisches Format und die Anwendung analytischer Algorithmen erfordert das Verständnis, wie man diese Verfahren für die Analyse von Texten sowohl nutzt als auch kombiniert [Mi12].

Text-Mining ist ein weitgehend automatisierter Prozess für die Gewinnung von neuen Erkenntnissen aus Textdokumenten [Mi12]. Die Anwendungen des Text-Mining sind so breit gefächert und ihre Ziele so vielfältig, dass es schwierig ist, seine Leistung in allgemeinen Worten zu beschreiben. Im Gegensatz zu anderen etablierten statistischen Methoden ist Text-Mining eine relativ neue und nicht standardisierte analytische Methode zur Wissensentdeckung. Folglich ist es eine Herausforderung, eine allgemeine Vorgehensweise für Text-Mining zu erstellen. Anwendungen des Text-Mining werden in erster Linie durch “trial and error“ auf der Grundlage persönlicher Erfahrungen und Vorlieben bestimmt [Mi12]. Während Data-Mining-Methoden relativ ausgereift sind, gibt es in der Literatur keine allgemein akzeptierten Methoden, die die besten Praktiken im Text-Mining in einem beliebigen Bereich widerspiegeln.

Feldmann [FS07] geht auf funktionaler Ebene davon aus, dass Text-Mining dem allgemeinen Modell einiger klassischer Data-Mining-Anwendungen folgt. Auch Hippner und Rentzmann [HR06] betonen, dass Text-Mining einen ähnlichen Aufbau wie ein klassischer Data-Mining-Prozess aufweist, sich aber in der Datenaufbereitung unterscheidet, weil beim Text Mining eine zusätzliche linguistische Datenaufbereitung erforderlich ist, damit die fehlende Datenstruktur rekonstruiert werden kann.

Die Vorgehensmodelle in der Literatur sind zu allgemein, um sie für einen konkreten Anwendungsfall einzusetzen. Die Prozessmodelle aus der Literatur können aber als Grundgerüst dienen und auf den Anwendungsfall dieser Arbeit übertragen werden (siehe folgenden Abschnitt); wobei die große Bedeutung der Datenvorverarbeitungsphase hervorzuheben ist, da sie eine wichtige Rolle in Text-Mining spielt.

3 Einsatz von Text-Mining

Die Abb. 2 zeigt das Vorgehen für die durchgeführte Studie. Es basiert auf dem Vorgehensmodell von Tauchert et al. [Ta20].



Abb. 2: Grafische Darstellung zur Vorgehensweise bei der Durchführung der Literaturrecherche und -auswertung, die mit Hilfe von Text Mining unterstützt wird [Ta20]

Zunächst werden die Phasen nachstehend kurz erläutert, anschließend werden diese im darauffolgenden Kapitel angewendet. Eine erst kürzlich erschienene Publikation hat die Entwicklung einen digitalen Plattform zum Inhalt, die Forschende dabei unterstützt, eine systematische Literaturnanalyse zu unterstützen [Am22].

3.1 Phase 1: Aufgabedefinition

Ähnlich zu Schieber et al. [SH14] und Hippner et al. [HR06] erfolgt eine Festlegung des Zwecks der Recherche und ein Festlegung der Suchbegriffe und der Datenquellen. Um ein gründliches Verständnis zu erlangen und die Ziele genau zu definieren, ist das notwendige Wissen zur Durchführung eines Literatur-Reviews zu ermitteln. Die Definition der Problem- sowie Zielstellung dieser Arbeit ergeben sich aus der Einleitung (siehe Kapitel 1).

3.2 Phase 2: Abstracts-Sammlung

Die zweite Phase besteht daraus, die wissenschaftlichen Publikationen zu selektieren und die ermittelten Abstracts herunterzuladen, die in den nachfolgenden Prozessschritten analysiert werden sollen. Dazu wird eine Recherche in der englischsprachigen ACM Digital Library [AC22] durchgeführt, und Abstracts zu wissenschaftlichen Publikationen heruntergeladen, die dem vom Benutzer definierten Suchbegriff und dem Datumsbereich entsprechen.

3.3 Phase 3: Aufbereitung der Abstracts mittels NLP-Methoden

In Text-Mining liegen die Textdaten in natürlicher Sprache, d.h. in halbstrukturierter Form, vor. Es ist schwierig, Muster aus halbstrukturierten Daten zu extrahieren, da diese Daten nicht für Text-Mining-Verfahren einsetzbar sind. Um Text-Mining durchzuführen, ist es daher notwendig, die Rohtexte einem Prozess zu unterziehen, bei dem verschiedene Methoden auf sie angewandt werden. Als Ergebnis werden bereinigte Token zurückgegeben. Token sind einzelne Wörter oder Wortgruppen, die als Merkmale für weitere Analysen sowie als Input für Text-Mining-Verfahren dienen [FS07]. Diese verfeinerten Daten sind für den Anwender besser geeignet, um sachkundige Daten zu extrahieren. Dieser Prozess wird als „Vorverarbeitung“ von Daten bezeichnet. Die Auswahl der Textvorverarbeitungsverfahren beeinflussen die Darstellung des Textdokuments und damit auch die darauffolgenden Ergebnisse erheblich [FS07]. Viele der Methoden zur Textvorbereitung und -aufbereitung haben ihre Wurzeln in Natural-Language-Processing.

Es wurden drei Vorverarbeitungsmethoden gewählt, wobei die erste die Filterung von Stopwörtern ist. Dies ist eine einfache und schnelle Methode, um Wörter herauszufiltern, die keine Bedeutung haben und nur in Verbindung mit anderen Wörtern verwendet werden [AHN19]. Das zweite gewählte Verfahren war die Lemmatisierung. Da die eingesetzten Text-Mining-Methoden zur Schlagwortextraktion auf der Häufigkeit des Auftretens eines Wortes in einem Dokument beruhen, ist es notwendig, das Lemma zu identifizieren. Dadurch wird sichergestellt, dass verschiedene morphologische Varianten eines Wortes auf ihr gemeinsames Lemma zurückgeführt und somit als Teil seiner Häufigkeit gezählt werden. In diesem Fall wurde die Lemmatisierung dem Stemming vorgezogen, denn beim Stemming wird einfach ein Wort aufgrund seines Aussehens ausgeschnitten, während die Lemmatisierung ein kalkulierter Prozess ist [AHN19]. Bei der Lemmatisierung ist außerdem zu beachten, dass es oft schwieriger ist, diese für eine neue Sprache durchzuführen als für einen Stemming-Algorithmus. Da eine Lemmatisierung viel mehr Wissen über die Struktur einer Sprache erfordert. Glücklicherweise sind alle Abstracts auf Englisch, so können vorgefertigte Pakete in Python bei der Lemmatisierung verwendet werden. Um gute Ergebnisse zu erzielen, sollten vorher jedoch Tags bezüglich der Wortart mithilfe von POS-Tagging an die Lemmatisierung übergeben werden, da sonst möglicherweise nicht alle Wörter auf die gewünschten Lemmata reduziert werden können [AHN19].

3.4 Phase 4: Keyword-Extraktion

Nachdem alle relevanten Abstracts in einem einheitlichen Format vorliegen, werden im nächsten Schritt die wichtigsten Stichwörter daraus extrahiert. Keywords sind Schlüsselbegriffe, die dabei helfen, die Dokumente zusammenzufassen und einen Überblick über den Inhalt zu vermitteln [FPW99]. Das manuelle Auffinden der relevanten Schlüsselbegriffe, die den Inhalt des Dokuments beschreiben, kann je nach Länge und Anzahl der Dokumente sehr anspruchsvoll sein. Daher wäre die Anwendung eines automatisierten Prozesses, der Schlüsselwörter aus Dokumenten extrahiert, sinnvoll. Im Bereich des Text-Mining gibt es

zahlreiche Methoden der Keyword-Extraction, die z.B. in statistische, überwachte, teil-überwachte und unüberwachte Ansätze kategorisiert werden können [SS15].

Bereits Barton und Kokoev [BK21] nutzten die unüberwachte Lernmethode RAKE, um Schlüsselwörter aus den wissenschaftlichen Arbeiten zu extrahieren. Für diesen Beitrag wird die statistisch-basierte TF-IDF-Methode ausgewählt, da sie Schlüsselwörter für ein bestimmtes Dokument in einer Dokumentensammlung gewichtet. Hierbei werden Wörter als einzigartig (Schlüsselwort) oder unwichtig (Nicht-Schlüsselwort) für das Dokument und somit für den gesamten Korpus eingestuft. Die zu Grunde liegende Idee ist, eine Dokumentensammlung als Beurteilung für die Worthäufigkeit zu verwenden, um die Wichtigkeit eines einzelnen Wortes für ein Dokument festzustellen [MRS08]. Wenn ein Term in vielen Dokumenten vorkommt, geht man davon aus, dass dessen Bedeutung gering ist. Falls ein Begriff in wenigen Artikeln sehr häufig vorkommt, scheint die Relevanz des relevanten Wortes für diese bestimmten Artikel hoch zu sein.

Zunächst berechnet man die Komponente „Termfrequenz“ (TF). Diese bewertet (gewichtet) jedes Wort entsprechend der Vorkommenshäufigkeit. Nun wird ein Blick auf die Formel geworfen. Davor werden zunächst folgende Notationen definiert:

- f ist die Gesamtanzahl der Dokumente
- d ist ein gegebenes Dokument aus dem Datenbestand
- D ist Sammlung aller Dokumente
- w ist ein bestimmtes Word in einem Dokument

TF wird nach der folgenden Formel berechnet

$$tf(w,d) = \log(1+f(w,d)),$$

wobei die reine (Vorkommens) Häufigkeit des Wortes in Dokument ist. Die Logarithmierung wurde verwendet, um die Termfrequenz der Dokumente und des Korpus zu normalisieren.

Mit der obigen Formel werden alle Terme als gleich wichtig angesehen. Daher führt man die zweite Komponente „inverse Dokumentenhäufigkeit“ ein, um das Gewicht von Begriffen zu verringern. Die inverse Dokumentenhäufigkeit beschreibt, wie viel Information ein Wort liefert, d. h. ob es in allen Dokumenten häufig oder selten vorkommt. Man definiert die inverse Dokumentenhäufigkeit () eines Terms wie folgt [MRS08]:

$$idf(w, D) = \log\left(\frac{N}{f(w, D)}\right)$$

Man kombiniert nun die Definitionen der Termhäufigkeit und der inversen Dokumenthäufigkeit. Dadurch erhält man das statistische Maß TF-IDF, welches dem Term eine Gewichtung im Dokument zuweist, das durch

$$\text{tfidf}(w,d,D) = \text{tf}(w,d) * \text{idf}(w,D)$$

gegeben ist [MRS08].

Diese Arbeit untersucht, wie sich die Themen im Bereich Quantum Computing innerhalb jedes Jahres sowie im Zeitverlauf entwickelt haben. Die Themen werden hierbei als Schlüsselwörter betrachtet und die wichtigsten Begriffe werden auf der Grundlage der Ergebnisse des TF-IDF-Algorithmus pro Jahr identifiziert. Dazu berechnet man den durchschnittlichen TF-IDF-Score aller Wörter über alle Abstracts aus einem bestimmten Jahr.

3.5 Phase 5: Wortrepräsentation mit Word2Vec

Die Methode Worteinbettung (englisch “word embedding“) wird in der Computerlinguistik verwendet, um die Bedeutung eines Wortes aus seinem Kontext zu extrahieren bzw. darzustellen. Dieses Verfahren stellt Wörter als numerische Wortvektoren dar, bei denen semantisch ähnliche Wörter auf nahe gelegene Punkte im geometrischen Raum abgebildet werden. Wörter, die in einem ähnlichen Kontext verwendet werden, werden in einem unmittelbaren Vektorraum abgebildet. Das Praktische an der Darstellung von Wörtern als Vektoren ist, dass sie sich dadurch für mathematische Operatoren eignen. Zum Beispiel kann man Vektoren miteinander addieren und subtrahieren.

Es gibt mehrere mögliche Algorithmen, die die Aufgabe der Berechnung von Worteinbettungen lösen können. Angesichts der Beliebtheit und Nützlichkeit wurde für dieses Projekt der Word2Vec-Algorithmus gewählt. Word2Vec ist ein neuronales Netz und kann in Kombination mit der Kosinus-Ähnlichkeitsformel dazu verwendet werden, ähnliche Wörter für ein vorgegebenes Wort zu finden. Das Verfahren wurde im Jahr 2013 von Mikolov et al. [Mi13] bei Google entwickelt. Es sei erwähnt, dass in diesen Beitrag die Funktionsweise sowie die mathematischen Details des word2vec-Modells nicht weiter eingegangen wird. Weiterführende Information sind bei Goldberg et al. [GL14] und Weng [We21] zu finden.

Die Vorgehensweise hier ist es, die Schlüsselwortextraktion zu erweitern. Die Basis-Schlüsselwörter können aus der vorherigen Phase 4 (siehe Abschnitt 3.4) identifiziert werden. Anschließend können weitere ähnliche Schlüsselwörter, die im gleichen Kontext wie die Basis-Schlüsselwörter vorkommen, mithilfe des Word2Vec-Verfahrens sowie dem Kosinus-Ähnlichkeitsmaß ermittelt werden. In diesem Beitrag wird Word2Vec mit demselben Datensatz trainiert, aus dem die vorherigen Schlüsselwörter erstellt wurden. Hierbei dienen die eindeutig identifizierten Schlüsselwörter aus Abschnitt 3.4 als Input, um eine neue Liste von weiteren Schlüsselwörtern zu erhalten.

3.6 Visualisierung der Ergebnisse

Der letzte fehlende Baustein für dieses Projekt ist die Visualisierung der Ergebnisse. Daher werden im Folgenden kurz Informationen über die beiden verwendeten Visualisierungsmethoden bereitgestellt.

Ein Bündel visuell dargestellter Wörter wird als Tag-Cloud, auch Word-Cloud (Wortwolke), bezeichnet [Ku07]. Dieses Bündel von Wörtern wird nach bestimmten Kriterien ausgewählt und hilft bei der explorativen Textanalyse, wichtige Wörter und kontextbezogene Themen aus einem großen Textkorpus hervorzuheben [Ku07]. Im Allgemeinen wird bei der Erstellung von Tag-Clouds die Darstellungsgröße der Wörter herangezogen. Diese ist abhängig von der Häufigkeit eines Wortes. D. h. je häufiger ein bestimmtes Wort in einem Text vorkommt, desto größer erfolgt seine Darstellung in der Tag-Cloud. In dieser Arbeit wird die Häufigkeit durch den berechneten TF-IDF-Wert der Wörter aus der Phase 4 ersetzt und dadurch wird für die Abbildung von Schlüsselwörtern eine relativ aussagekräftige Darstellung ermöglicht.

Für eine Visualisierung der mehrdimensionalen Wort-Vektoren (siehe Abschnitt 3.5) müssen diese in zweidimensionale Wort-Vektoren umgewandelt werden. Dies wird mittels t-Distributed Stochastic Neighbor Embedding (t-SNE) umgesetzt. T-SNE ist ein unüberwachtes und nicht-lineares Verfahren, das von Laurens van der Maaten und Geoffrey Hinton [vH08] im Jahr 2008 entwickelt wurde.

Die Grundidee von t-SNE besteht darin, den mehrdimensionalen Raum unter Beibehaltung des relativen paarweisen Abstands zwischen den Punkten zu reduzieren, wodurch die Abstände zwischen den Punkten identisch bleiben. D. h. der Algorithmus bildet mehrdimensionale Daten auf zwei oder mehr Dimensionen ab, wobei die Punkte, die ursprünglich weit voneinander entfernt waren, auch weit entfernt liegen, und nahe Punkte auch in nahe Punkte umgewandelt werden. Man kann sagen, dass t-SNE nach einer neuen Datendarstellung sucht, bei der die Nachbarschaftsbeziehungen erhalten bleiben.

4 Ergebnisse der Literaturanalyse mittels Text-Mining-Methoden

In diesem Kapitel werden die Ergebnisse der Text-Mining-Anwendung nach der zuvor beschriebenen Methode vorgestellt. Alle Wörter, die mit dem TF-IDF- oder Word2Vec-Verfahren extrahiert wurden, werden bewusst kursiv geschrieben.

4.1 Extraktion von wissenschaftlichen Abstracts

Für diesen Beitrag wird der Suchbegriff „Quantum Computing“ ausgewählt. Um die Suche zu erweitern, wurden die Begriffe „quantum computation“, „quantum computer“ und

„quantum computers“ hinzugefügt. Die daraus resultierende Treffermenge wurde durch die Auswahl „Research Article“ in der Kategorie „Content Type“ weiter eingegrenzt. Für eine bessere Übersicht wurden ausschließlich Artikel von 2008 bis 2021 in Betracht gezogen. Vor dem Jahr 2008 sowie – bedingt durch den Zeitpunkt der Recherche – nach dem Jahr 2021 standen wenige wissenschaftlichen Arbeiten zur Verfügung.

Das folgende Balkendiagramm in Abb. 3 zeigt die Gesamtzahl der Veröffentlichungen für jedes einzelne Jahr. Es ist zu erkennen, dass die Anzahl der Veröffentlichungen von 2008 bis 2021 im Bereich Quantum Computing deutlich gestiegen ist. Im Jahr 2008 wurden in diesem Bereich nur 29 Research-Papers veröffentlicht, während die entsprechende Zahl der Arbeiten im Jahr 2021 196 betrug. Das Wachstum nimmt über die Jahre stetig zu, vor allem ab dem Jahr 2015 ist eine stärkere Zunahme der Veröffentlichungen zu verzeichnen.

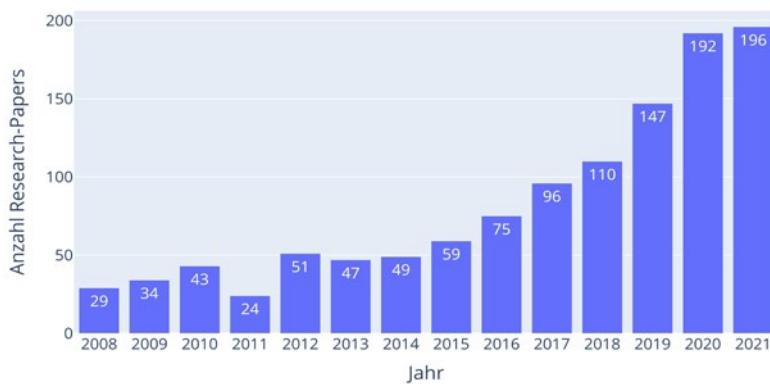


Abb. 3: Entwicklung der Anzahl von wissenschaftlichen Publikationen zu Quantum Computing zwischen 2008 und 2021

4.2 Extraktion der Schlüsselwörter sowie deren Visualisierung mithilfe von Wortwolken

Im Folgenden werden diese Schlüsselwörter ermittelt und visuell analysiert. Hierbei wird die Methode von Weiwei et al. [We10] übernommen. Die vorgeschlagene Methode verwendet die dynamische Tag-Clouds, um die zeitliche Entwicklung von Inhalten zu veranschaulichen. Diese zeitbasierten Word-Clouds stellen verschiedene statische Wortwolken zu konkreten Zeitpunkten dar. Dadurch entsteht ein aussagekräftiges Trenddiagramm, das die Veränderung der Word-Clouds im Laufe der Zeit darstellt. So erhält man einen visuellen Überblick über die unterschiedlichen thematischen Schwerpunkte.

Bei der Erstellung von Word-Clouds werden die durchschnittlichen TF-IDF-Werte herangezogen. Die Größe der Wörter spiegelt den durchschnittlichen TF-IDF-Wert für das Wort in diesem entsprechenden Jahr wider und gibt dabei die Wichtigkeit für das jeweilige Schlüsselwort – in diesem Fall sind es Bigramme – Abb. 4 zeigt die Wortwolken zu dem beschriebenen Korpus aus dem obigen Abschnitt. Es ist anzumerken, dass Begriffe wie *quantum computing*, *quantum computation*, *quantum computer* und *quantum computers* aus dem Korpus entfernt wurden, da ihre Verwendungshäufigkeit beträchtlich wäre und die anderen Begriffe überdecken würde.

Für die Visualisierung werden zunächst die Abstracts nach Jahr gruppiert und für jedes Jahr wird eine Wortwolke mit jeweils 15 Bigrammen erstellt. Die folgende Abb. 4 zeigt die Ergebnisse für die Jahre von 2019 bis 2021.

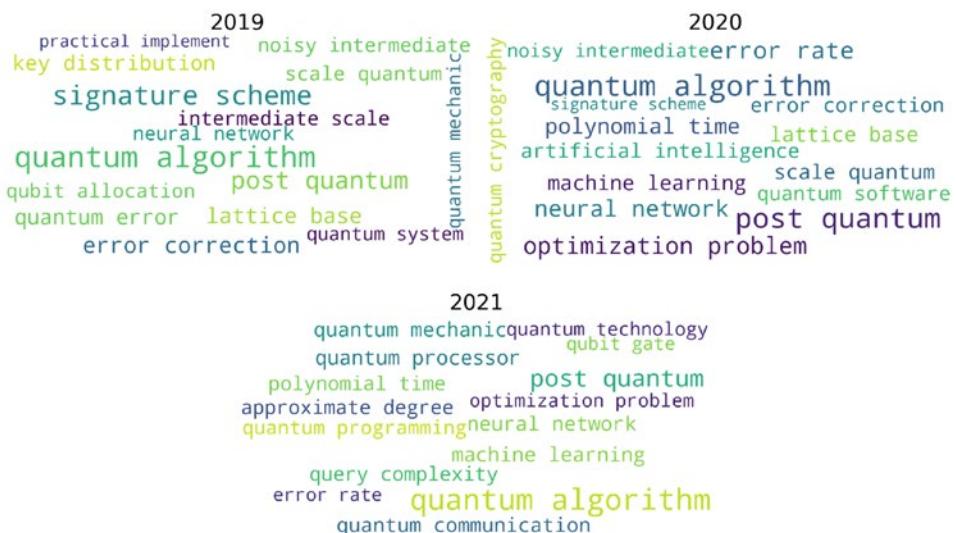


Abb. 4: Visualisierung von extrahierten Schlüsselwörtern mit Hilfe von Word-Clouds für die Jahre 2019 bis 2021

4.3 Zukunft des Quanten Computing

Auf Basis der Wortwolken konnten wichtige Themen für die Zukunft des Quanten Computing identifiziert werden, z. B. *noisy intermediate*, *intermediate scale*, *machine learning*, *artifical intellegence*, *neural network* und *error correction*. Die beiden Bigramme *noisy intermediate* sowie *intermediate scale* deuten auf das offene Kompositum „Noisy Intermediate Scale Quantum Computing“, kurz NISQ.

Im Folgenden wird die zeitliche Relevanz der Wörter untersucht. Für die zeitliche Betrach-

tung der extrahierten Schlüsselwörter wird eine Kennzahl eingeführt, um eine objektivere Interpretation zu ermöglichen sowie die Bedeutung von Forschungsthemen im zeitlichen Verlauf abzubilden. Die Kennzahl spiegelt die durchschnittliche Anzahl der Schlüsselwörter pro Artikel für jedes Jahr. Anschließend wird diese Zahl für die Erstellung der Liniendiagramme verwendet, wie die folgende Abb. 5 zeigt:



Abb. 5: Zeitliche Entwicklung von Publikationen zu ausgewählten Schlüsselwörtern

Ab dem Jahr 2015 wächst die Bedeutung von Themen wie *nisq*, *machine learning* und *artificial intelligence*. Die hohe Anzahl an wissenschaftlichen Publikationen in den letzten drei Jahren (siehe Abb. 3) sowie diese genannten Trends deuten auf eine rasche Ausweitung des Forschungsgebiets Quanten Computing hin und zeigen, dass sich mehr und mehr Wissenschaftler für das Quantum Computing interessierten.

Anhand der identifizierten Stichwörter lässt sich die Zukunft von Quanten Computing in drei Phasen einteilen:

Die erste Phase lautet NISQ. Dies ist die Ära der kleinen, fehleranfälligen oder „noisy intermediate scale quantum“ (NISQ) Maschinen, wie Preskill [Pr18] es zum ersten Mal im Jahr 2018 formuliert hat. Verrauscht deshalb, weil man nicht genügend Qubits für die Quantenfehlerkorrektur zur Verfügung hat. Und „*Intermediate Scale*“ wegen ihrer kleinen Qubit-Anzahl.

Nach der NISQ-Ära würde das Rauschen verhindern, dass Quantencomputer mit größerer Anzahl an Qubits eingesetzt werden. Quantencomputer benötigen daher eine Fehlerkorrektur. Ein Ziel wird daher sein, ein verlässliches Verfahren für die Quanten-Fehlerkorrektur in einem Quantencomputer zu implementieren, welches das Rauschen schneller korrigiert als es erzeugt wird. Daher wird solch ein Fortschritt in der Forschung als ein technologischer Meilenstein für die Entwicklung eines fehlertoleranten Quantencomputers gesehen.

Die letzte Phase betrifft vor allem das Wort *machine learning*. Für die nähere Untersuchung der letzten Phase wurde mit dem t-SNE-Algorithmus und dem Word2-Vec-Modell eine zweidimensionale Darstellung der semantisch ähnlichen Wörter im Umfeld des Bi-

gramms *machine learning* durchgeführt. Es werden zunächst die ersten 50 Datenpunkte manuell ausgewertet und anschließend diejenigen mit einer Beschriftung angezeigt, die von Interesse sein könnten. Es sei erwähnt, dass eine Messung der Distanz zwischen zwei Wortvektoren zu keiner gültigen Aussage führt. In der Abb. 6 kann der Abstand zwischen zwei Wortvektoren lediglich als nah oder weit entfernt bezeichnet werden.

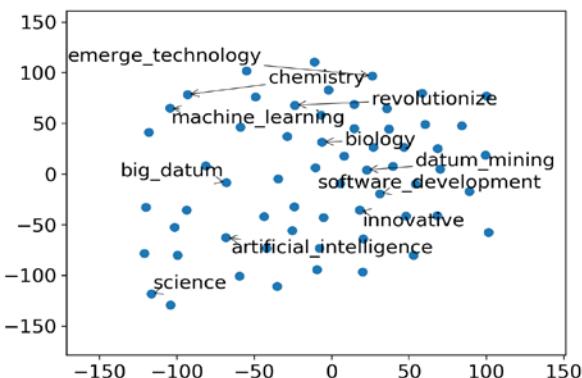


Abb. 6: Zweidimensionale Darstellung semantisch ähnlicher Wörter zu „*machine learning*“

Unter Quanten-Machine-Learning versteht man den Einsatz von Quantencomputern für das maschinelle Lernen. Die in Abb. 6 dargestellten Wörter *innovative*, *emerge technology* sowie *revolutionize* weisen darauf hin, dass Quantum-Machine Learning ein aufstrebendes Gebiet in Wissenschaft und Technologie ist. Es ist der Schnittpunkt von Quantenphysik und maschinellem Lernen. Das maschinelle Lernen könnte mithilfe von Quantencomputern zukünftig in den Bereichen *biology*, *chemistry* und *design automation* eingesetzt werden.

5 Fazit

Im Rahmen einer Literaturrecherche und -analyse zu Quanten Computing lassen sich mit Hilfe von Text-Mining-Methoden Schlüsselwörter identifizieren und auswerten, welche die zeitliche Entwicklung relevanter Themenfelder für das Forschungsgebiet Quanten Computing gut beschreiben. Auch aktuelle Themenfelder, die in einzelnen Zeiträumen vorherrschend waren, lassen sich so ableiten. Die Auswertung einer Literaturanalyse kann mit Hilfe von Text-Mining unterstützt und beschleunigt werden.

Die Anwendung der dynamischen Word-Clouds und der TF-IDF-Methode führten zu dem Ergebnis, dass sich die Forscher/innen in den letzten Jahren verstärkt für die Themen maschinelles Lernen, künstliche Intelligenz, NISQ und Fehlerkorrektur konzentrierten. Es ist zu erwarten, dass diese auch die zukünftigen Trends im Bereich des Quanten Computing

bilden werden, da viele Forschungsfragen noch offen sind. Die Einführung eines zusätzlichen Worteinbettungsverfahren erwies sich als nützlich für das weitere Verständnis der extrahierten Schlüsselwörter und für die Identifikation weiterer Themen und möglicher Trends.

Nach wie vor sind Forschende gefordert, die Literatur zu verstehen, zu verarbeiten und relevantes Wissen zu extrahieren. Insbesondere müssen die Schlüsselwörter, die mithilfe von TF-IDF oder Word2Vec extrahiert werden, genau überprüft werden, da sie nicht immer die eigenen Erwartungen widerspiegeln. So kann beispielsweise eine Word-Cloud ein Schlüsselwort beinhalten, welches zur weiteren Untersuchung weniger geeignet ist. Daher empfiehlt es sich, auf Basis der Word-Clouds oder t-SNE-Visualisierung Stichwörter manuell auszuwählen und diese in einen Wörterkatalog zu übertragen. Darüber hinaus müssen auch Forschungslücken weiterhin eigenständig von Forscher/innen identifiziert werden.

Zusammenfassend lässt sich sagen, dass die Extraktion von Schlüsselwörtern einen unterstützenden Mechanismus zur Beschleunigung des Literature-Review-Prozesses bietet und als ein Hilfsinstrument für die Teilautomatisierung eines manuellen Prozesses betrachtet werden kann.

Literaturverzeichnis

- [AC22] ACM Digital Library: ACM Digital Library. <https://dl.acm.org/>, Abruf am 01.04.2022.
- [AHN19] Anandarajan, M.; Hill, C.; Nolan, T.: Practical Text Analytics. Maximizing the Value of Text Data. Springer International Publishing, Cham, 2019.
- [Am22] Ammirato, S et al. Digitalising the Systematic Literature Review process: the MySLR platform. *Knowledge Management Research & Practice*, 2022, S. 1-18.
- [Ba20] Bauckhage, C. et al.: Quantum Machine Learning. Eine Analyse zu Kompetenz, Forschung und Anwendung, 2020. <https://www.bigdata-ai.fraunhofer.de/de/publikationen/quantum-ml.html>. Abruf am 04.04.2022.
- [BK21] Barton, T.; Kokoev, A.: Text Mining bei einer wissenschaftlichen Literaturauswertung: Extraktion von Schlüsselwörtern zur Beschreibung von Inhalten. In Barton, T.; Müller, C. (Hrsg.): *Data Science anwenden (Angewandte Wirtschaftsinformatik)*. Springer Vieweg, S. 193–200, 2021.
- [FPW99] Frank, E.; Paynter Gordan W.; Witten, I. H.: Domain-Specific Keyphrase Extraction. *International Journal of Computer Applications*, S. 668–673, 1999.
- [FS07] Feldman, R.; Sanger, J.: *The text mining handbook. Advanced approaches in analyzing unstructured data*. Cambridge University Press, Cambridge, 2007.
- [GL14] Goldberg, Y.; Levy, O.: word2vec Explained: deriving Mikolov et al.’s negative-sampling word-embedding method, 2014.
- [Ho18] Homeister, M.: *Quantum Computing verstehen*. Springer Fachmedien Wiesbaden, Wiesbaden, 2018.
- [HR06] Hippner, H.; Rentzmann, R.: Text Mining. *Informatik-Spektrum* 4/29, S. 287–290, 2006.
- [Ku07] Kuo, B. Y.-L. et al.: Tag clouds for summarizing web search results. In (Williamson, C. et al. Hrsg.): *Proceedings of the 16th international conference on World Wide Web - WWW ‘07*. ACM Press, New York, New York, USA, S. 1203, 2007.
- [Mi12] Miner, G. et al.: Practical text mining and statistical analysis for non-structured text data applications. Elsevier Academic Press, Waltham, Amsterdam, Boston, Heidelberg, London, New York, Oxford, Paris, San Diego, San Francisco, Singapore, Sydney, Tokyo, 2012.
- [Mi13] Mikolov, T. et al.: Efficient Estimation of Word Representations in Vector Space, 2013.
- [MRS08] Manning, C. D.; Raghavan, P.; Schütze, H.: *Introduction to information retrieval*. Cambridge University Press, New York, 2008.
- [Pr18] Preskill, J.: Quantum Computing in the NISQ era and beyond. *Quantum* 2, S. 79, 2018.
- [SH14] Schieber, A.; Hilbert, A.: Entwicklung eines generischen Vorgehensmodells für Text Mining. Technische Universität Dresden, Fakultät Wirtschaftswissenschaften, 2014.
- [SS15] Siddiqi, S.; Sharan, A.: Keyword and Keyphrase Extraction Techniques: A Literature Review. *International Journal of Computer Applications* 2/109, S. 18–23, 2015.
- [Ta20] Tauchert, C. et al.: Towards an Integrative Approach for Automated Literature Reviews Using Machine Learning. In (Bui, T. Hrsg.): *Proceedings of the 53rd Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences, 2020.

- [vH08] van der Maaten, L.; Hinton, G. E.: Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9, S. 2579–2605, 2008.
- [vo09] vom Brocke, J. et al.: Reconstructing the giant: On the importance of rigour in documenting the literature search process: ECIS, 2009.
- [We10] Cui, W. et al.: Context-preserving, dynamic word cloud visualization. In: 2010 IEEE Pacific Visualization Symposium (PacificVis). IEEE, 2010. S. 121-128
- [We21] Weng, L.: Learning Word Embedding. <https://lilianweng.github.io/lil-log/2017/10/15/learning-word-embedding.html>, Stand: 19.09.2021.

Konzeption eines Machine-Learnings-Verfahrens zum Lösen von Green Vehicle Routing Problemen

Pablo Stockhausen¹, Andreas Johannsen², Robert Maurer³

Abstract: Dieser Beitrag entwickelt ein Konzept zur praktischen Umsetzung eines Machine- Learning-Verfahrens zum Lösen von Vehicle Routing Problemen im Kontext einer nachhaltigen “Letzte-Meile”-Logistik, welches durch einen Prototyp umgesetzt und getestet wurde. Dabei wurden Aspekte von kombinatorischen Optimierungsalgorithmen in Form eines Ameisenalgorithmus zur Unterstützung des angewendeten Machine-Learning-Systems verwendet. Der Prototyp basiert auf einem “Reinforcement Learning”-System und verwendet als Algorithmus “REINFORCE mit Baseline”. In einer Vergleichsanalyse wird der Prototyp mit dem bekannten Vertreter für kombinatorische Optimierungsalgorithmen, Google-OR, an Hand von zwei Anwendungsszenarien verglichen. Die kombinatorischen Optimierungsalgorithmen konnten sich hinsichtlich der Lösungsqualität gegenüber dem Prototyp durchsetzen. Dafür überzeugt der Prototyp in der Laufzeit und dem Automatismus. Es wurde festgestellt, dass der verwendete Machine-Learning-Kontext für kleine bis mittelgroße Logistikdomänen nur geringe Vorteile ergibt. Eine Verwendung von lernenden Systemen für das Vehicle Routing Problem empfiehlt sich bei einem größeren Stoppvolumen und einer erweiterten IT-Infrastruktur. Letztlich bietet der Prototyp in diesem Beitrag eine Alternative gegenüber dem Outsourcing an Drittanbietern zum Lösen von Routingproblemen.

Keywords: Vehicle Routing Problem; Machine Learning; Reinforcement Learning; Ant Colony Optimization; Letzte- Meile- Logistik; Ameisenalgortyhmus.

1 Universität Potsdam, Wirtschafts- und Sozialwissenschaftlichen Fakultät, August-Bebel-Straße 89, 14482 Potsdam, stockhausen017@gmail.com

2 Technische Hochschule Brandenburg, Fachbereich Wirtschaft, Magdeburger Str. 50, 14770 Brandenburg an der Havel, johannse@th-brandenburg.de>

3 Technische Hochschule Brandenburg, Fachbereich Wirtschaft, Magdeburger Str. 50, 14770 Brandenburg an der Havel, robert.maurer@th-brandenburg.de

1 Vehicle Routing Probleme im Kontext der „Letzten-Meile“-Logistik

1.1 Motivation und Zielsetzung

Im Jahr 2020 umfasste das Sendungsvolumen in Deutschland 4,05 Mrd. Paket-, Express- und Kuriersendungen mit einer in den kommenden Jahren steigenden Tendenz, wobei der Hauptanteil an Sendungen den B2C Bereich bedient hat [BU21]. Zu dem Anstieg an Paketsendungen und den damit auftretenden Routingproblemen, ist auch das wissenschaftliche Interesse für Vehicle Routing Probleme (VRP) auf Basis von modernen IT-Lösungen gestiegen. Eine Suche nach den Begriffen „Vehicle Routing“ in der Datenbank IEEE Explore (<https://ieeexplore.ieee.org>) ergab über 6500 Treffer in den letzten zehn Jahren. Ungeachtet dessen, gibt es nur wenige wissenschaftliche Arbeiten im Bereich der Machine-Learning (ML) -Verfahren zum Lösen von VRP. Eine Suche mit Google Scholar (<https://scholar.google.com>) ergab 15 Veröffentlichungen in den letzten fünf Jahren, welche im Titel die Begriffe „Vehicle Routing“ und „Machine-Learning“ verwendet haben. Zusätzlich bewegen sich die fundamentalen Forschungserkenntnisse in einem überwiegenden theoretischen Umfeld mit Schwerpunkt auf mathematischen Modellen und Annahmen, jedoch mit wenig Praxisbezug. Die Hauptmotivation dieser Arbeit liegt darin, ein praktisches Konzept, in Form eines Prototyps, zum Lösen von VRP zu entwickeln und im Besonderen die Aspekte eines Capacitated Vehicle Routing Problems (CVRP) zu berücksichtigen. Dieser Beitrag verwendet CVRP und VRP gleichbedeutend. Weiterhin soll gezeigt werden, dass sich ML-Verfahren in einem produktiven Umfeld durchaus beweisen und eine langfristige Alternative zum Outsourcing an Drittanbieter darstellen können. Dies wird durch einen Vergleich, hinsichtlich der Effektivität unter ausgewählten Bewertungskriterien, zwischen ML-Verfahren und den klassischen Optimierungsalgorithmen zum Lösen von VRP dargestellt.

1.2 Thematische Hinführung

Das Vehicle Routing Problem (VRP) beschreibt im Gebiet der emissionsfreien „Letzte-Meile“-Großstadtlogistik ein kombinatorisches Optimierungsproblem, welches folgende Grundfrage behandelt: „Was ist der optimale Satz von Routen für einen bestimmten Fuhrpark, um eine bestimmte Anzahl von Kunden zu beliefern?“ [BRN15, S.1]. Das VRP wurde erstmalig von Dantzig & Ramser (1959) durch ihre wissenschaftliche Arbeit „The Truck Dispatching Problem“ diskutiert, wobei der Problemkontext die Lieferung von Kraftstoffen darstellte und mithilfe von algorithmischen Wegen gelöst wurde [DR59]. Zum Lösen von VRP eignen sich besonders Machine-Learning (ML)-Verfahren, da sie algorithmisch vorgehen und erfahrungsbasiert entscheiden können. Die Erfahrungen ergeben sich durch die in bestimmten Themendomänen auftretenden Abonnement-Strukturen und den daraus

resultierenden bekannten VRP-Instanzen. Der spezifische Anwendungsfall des VRP entsteht im Rahmen dieser Arbeit durch den Kontext einer „Letzten-Meile“-Großstadtlogistik. Dies ist eine moderne Form der urbanen Logistik und umfasst den letzten Logistik Schritt einer Paketzustellung. Im Besonderen wird ein emissionsfreier Ansatz angestrebt, der durch die Verwendung von Lastenrädern sowie Mikro-Depots realisiert wird und die Bezeichnung als Green-VRP begründet [RSC20, S. 6]. Der Aspekt einer „Letzten-Meile“-Großstadtlogistik spielt beim Lösen von kombinatorischen Optimierungsproblemen dahingehend eine Rolle, dass in der Regel von kurzen Routendistanzen ausgegangen werden kann und Touren durch die Verwendung von Lastenfahrrädern klaren Kapazitätsbeschränkungen unterliegen. Die Planung effizienter Touren ist somit nicht nur ein ausschlaggebender Faktor für das Ausmaß der hohen Logistikkosten einer emissionsfreien Lieferung, sondern auch nötig, um mit der Alternative von Diesel- bzw. Benzin-Kleintransportern mithalten zu können.

1.3 Herausforderungen & Wettbewerbsfähigkeit

Es bestehen Herausforderungen in der späteren Implementierung des Prototyps bezüglich des Wettbewerbs zwischen klassischen Optimierungsalgorithmen und des verwendeten ML-Verfahrens zum Lösen von VRP. Besonders im Bezugspunkt der Performance wird das gewählte ML-Verfahren des Prototyps anfänglich fordernder sein, da es mehr Prozessschritte, wie beispielsweise das Trainieren des Modells, durchlaufen muss. Der Prototyp ist dann wettbewerbsfähig, wenn: Er kurze Routen auf Basis der Tour lange bildet, die Ergebnisse in wenigen Sekunden bereitstellt, allgemein auf nicht bekannte VRP-Instanzen reagiert, auf allen VRP-Instanzen gleich gut arbeitet und kein manuelles Eingreifen benötigt wird.

2 Methodik

2.1 Auswahl des Machine-Learning -Verfahrens

Im Rahmen dieses Beitrages verwendet das ML-Modell ein Reinforcement Learning (RL)-System zum Lösen von VRP der „Letzte-Meile“-Großstadtlogistik [RA18]. Das RL unterscheidet sich maßgeblich von den Alternativen des Supervised- und Unsupervised-Learnings, da es eine andere Herangehensweise für den Aufbau eines Lernenden-Systems verwendet. Das lernende Modell beschreibt im RL einen Agenten oder auch Entscheidungsträger, der ein Environment beobachtet, Aktion darauf ausführt und deren Dynamiken, beginnend von einem unwissenden Startpunkt eigenständig lernt [RA18, S. 2]. Grundlegend steht $A_{(st)} \subseteq A$ für die Menge der Aktionen, die der Agent zu einem Zeitpunkt t ausführen kann [Lo20, S.15]. Dabei beschreibt $a_t \subseteq A$ die ausgewählte Aktion in einem Zeitpunkt t des

Agenten. Nach jeder Aktion kann der Agent zwei Arten von Belohnungen erhalten: Eine sofortige oder eine verzögerte Belohnung. Eine sofortige Belohnung findet Anwendung bei Aktionen des Agenten, die eine unmittelbare Beurteilung zu lassen. Diese wurde auch in dem entwickelten Prototyp verwendet. Zum Beispiel lässt sich das Überqueren einer roten Ampel unmittelbar bewerten, da ein negatives Verhalten direkt erkennbar ist und nicht erst durch Folgeverhalten erkenntlich wird. Unter einer verzögerten Belohnung kann beispielsweise eine Aktion in einer Schach- oder Go-Partie verstanden werden, weil die Belohnung sich an den Folgereaktionen misst. Diese wurde auch in der wissenschaftlichen Arbeit von Silver et al. (2017) verwendet [Si17]. Es wurde sich für RL-System entschieden, um ohne menschliches Vorwissen, Lösungen für komplexe Optimierungsprobleme zu finden und somit Entwicklungs- und Dispositionsaufwand einzusparen.

2.2 Trainings- und Ergebnisdaten

In Kooperation mit einem Unternehmen, welches sich auf nachhaltige Transportlösungen spezialisiert hat, konnten als Datengrundlage die anonymisierten Kundendaten der Öko-box-Anbieter in Berlin zum Trainieren und Testen des Modells verwendet werden. Die Datenobjekte verfügen über folgende Eigenschaften: Längengrad, Breitengrad, Stoppgewicht, Stoppvolumen und einem anonymisierten Identifikator. Die Datenobjekte sind für ein ausgewähltes Datum und einem bestimmten Micro-Depot spezifisch.

2.3 Prototyping

Zur Fundamentierung der Forschung und auf Basis der praktischen Motivation wurde ein qualitativ-konstruktivistischer Ansatz in Form des Prototypings ausgewählt. Der Prototyp soll in der Endfassung einen Service darstellen, der ML-Modelle trainiert, speichert und anwendet. Das ML-Modell ist gegenüber dem Versender, Micro-Depot, Lieferant, Wochentag, Fahrzeuggewicht, Fahrzeugvolumen sowie ML-Verfahren sensibel und wird auf Instanz Basis gespeichert. Das ML-Modell wird nur durch Stopps von einem Tag bis zu einer Woche trainiert, um zu vermeiden, dass Abonnement-Kunden durch doppeltes Auftreten die Strategie des Agenten verfälschen.

2.4 Vergleichsanalyse

Im Kontext dieses Beitrags wird eine Vergleichsanalyse zwischen dem entwickelten Prototyp und dem VRP-Solver von Google-OR durchgeführt. Der VRP-Solver basiert nach Google (2021) auf heuristischen Algorithmen, die als „First Solution Strategy“ kategorisiert werden und optional durch „Local-Search“-Strategien erweitert werden können [Go21]. Die Resultate aus dem Vergleich sollen den Prototypen evaluieren, Verbesserungsmöglich-

keiten aufdecken und das Potenzial einer Implementierung von Reinforcement Learning (RL)-Verfahren zum Lösen von VRP in Erwägung ziehen. Die Analyse ist in zwei Teile gegliedert. Im ersten Teil der Vergleichsanalyse liegt der Hauptfokus auf einer Effektivitätseinschätzung des entwickelten Prototyps mit der zugrundeliegenden Basisfunktionalität des ausgewählten RL-Algorithmus „REINFORCE mit Baseline“ zum Lösen von VRP. Hier nach beruht die Effektivitätseinschätzung auf den Schwerpunkten der Trainings- sowie Testlaufzeit, aber auch der Vergleichskriterien der Distanz und des zeitlichen Aufwandes. Der Distanzaufwand wird mit der Haversine-Formel belegt. Hierbei wird der Zeitaufwand durch die hinterlegten Fahrzeuggeschwindigkeiten, die Distanz und einer definierten Stoppverweildauer von fünf Minuten errechnet. Weitere Vergleichskriterien für ein aufbauendes Experiment wären die Profitabilität, Service-Qualität, Konsistenz sowie externe Faktoren [VLM19, S. 2]. Es wird ein Stoppvolumen von 430 Stopps verteilt auf vier Wochen im Mai 2021 betrachtet, wobei das ML-Modell nach jedem Tag mit 200 Iterationen trainiert wird. Es wird aufbauend trainiert, sodass jede folge Woche auf die Erfahrung der Vorangegangenen aufbaut. Im zweiten Teil der Analyse liegt das Augenmerk auf dem Verbesserungspotenzial des ML-Verfahrens und dem Ausarbeiten von Anpassungsimpulsen, die aus dem vorherigen ersten Teil abgeleitet werden. Das Ziel ist es, sich so konvergent wie möglich an den errechneten Distanz- sowie Zeitaufwand der Konkurrenz anzunähern, um dadurch mit einer besseren Laufzeit zu überzeugen. Im zweiten Teil wird derselbe Problemkontext für einen aussagekräftigeren Vergleich betrachtet. Hierbei wird analog zum ersten Teil nach jedem Tag mit angepassten 115 Iterationen trainiert. Die Iterationsanzahl wurde auf 115 herabgesetzt, um mögliches „overfitting“ des ML-Modells gegenüber den Trainingsdaten zu verhindern.

3 Ergebnisse der Forschungsarbeit

3.1 Modellierung und Konzeption des Machine-Learning-Modells

Das ML-Modell beruht grundlegend auf einem Markov Decision Process (MDP) in Verbindung mit einem Ant Colony Optimization (ACO)-Algorithmus. Dies ist ein Verfahren der kombinatorischen Optimierung, wobei die Grundsteine des Modells durch Hyeong Soo Chang et al. (2004) gelegt wurden [Ch04]. Das ACO-Verfahren wurde im Trainingsmodus des Prototyps als erste Prozessinstanz eingesetzt, um dem RL-Verfahren im anschließenden Prozessschritt eine Vorahnung der Übergangswahrscheinlichkeitsverteilung bereitzustellen und schon trainierte ML-Modelle für den betrachteten Problemkontext einzustimmen. Vergleichsweise wurde ein ähnliches Unterfahren von Yuan Sun et al. (2020) durchgeführt, welche aber als Ausgangslage ein durch ML unterstützendes ACO-Modell betrachtet haben [Su20]. Das ML-Modell wird durch ein MDP definiert. Bei Markov-Modellen sind der Folgezustand und die RL-Belohnung nur vom momentanen Zustand und der gewählten Aktion des Agenten abhängig [Lo20, S.15]. Das angepasste MDP in dem entwickelten Prototyp wurde in Anlehnung an Puterman (2014) modelliert [Pu14]. Als ML-Algorithmus

verwendet der Prototyp den „REINFORCE“-Algorithmus oder auch Monte-Carlo Policy-Gradient, um eine optimale Policy π^* zu finden. Für die Entwicklung des Prototyps und der Anwendung des „REINFORCE“-Algorithmus wurde sich an dem Buch „Reinforcement Learning – An Introduction“ von Sutton & Barto (2018) orientiert [SB18]. Dabei zeichnen sich Monte-Carlo-Methoden dadurch aus, dass kein ganzheitliches Wissen über das Environment für die Findung einer optimalen Policy nötig ist. Dies liegt daran, dass das System von den Interaktionen mit dem Environment lernt [SB18, S. 91]. Der „REINFORCE“-Algorithmus des Prototyps baut auf einer parametrisierten Policy $\pi(a|s, \theta)$ auf und verwendet zusätzlich einen Vergleich zwischen den momentanen Aktionswerten und einer willkürlichen Baseline. In abgewandelter Form ist der „REINFORCE mit Baseline“-Algorithmus nach Sutton & Barto (2018) konstruiert und wurde in Abb. 1 dargestellt [SB18, S. 330]. Eine Baseline kann dabei eine zufällige Variable sein oder wie im Rahmen dieses Prototyps eine „state-value“-Funktion, die eine Bewertung über die möglichen Belohnungen des gesamten Zustandsraumes widerspiegelt [SB18, S. 329].

Data : Iterationen k , parametrisierte Policy $\pi(a|s, \theta)$, parametrisierte „state-value“-Funktion (baseline) $V(s, w)$

Result : optimal parametrisierte Policy π^*

initialisieren der Policy-Parameter θ ;

initialisieren der Gewichte w ;

for $i \leftarrow 0$ **to** k **do**

$Eps \leftarrow$ Episode nach $\pi(\cdot|\cdot, \theta) = \{S_0, A_0, R_1, \dots, S_{T-1}, A_{T-1}, R_T\}$;

foreach step $t \in Eps$, $t = 0, 1, \dots, T - 1$ **do**

$G \leftarrow \sum_{j=t+1}^T \gamma^{j-t-1} \times R_j$;

$\delta \leftarrow G - V(S_t, w)$;

$w \leftarrow w + \alpha \times \gamma^t \times \delta \times \nabla V(S_t, w)$;

$\theta \leftarrow \theta + \alpha \times \gamma^t \times \delta \times \nabla \ln \pi(A_t | S_t, \theta)$;

end

end

Abb. 1 „REINFORCE mit Baseline“-Algorithmus in Anlehnung an Sutton und Barto (2018)
[SB18, S.330]

3.2 Prototypenentwicklung

Das Backend des Prototyps wird in der Programmiersprache Python entwickelt und das Web-Frontend, welches zum besseren Debugging nötig ist, in React. Es wurden bewusst keine vorhandenen ML-Frameworks wie beispielsweise Keras, Tensorflow oder PyTorch verwendet. Die Entscheidung liegt darin begründet, dass im Falle der Verwendung von

komplexen ML-Frameworks das Risiko einer Abhängigkeit von dem jeweiligen Framework-Support bestünde. Darüber hinaus zeigt sich ein besserer Verständnis- und Erfahrungsgewinn bei einer Eigenproduktion von komplexen ML-Verfahren gegenüber dem Einsetzen verschiedener ML-Frameworks, wobei eher dazu tendiert wird, das Framework zu lernen und nicht das darauf aufbauende ML-Verfahren. Weiterhin traten bei der Entwicklung des Prototyps spezielle Herausforderungen auf. Anführend dafür ist die Thematik des lokalen Minimums, in dieser sich der Prototyp in den Anfangsphasen der Entwicklung häufig verfangen hat. In Abb. 2 ist ein Beispieldurchlauf der kumulierten Belohnungen, welche als Gesamtlänge in km aller gebildeten Touren in einer Episode verstanden werden, über 2000 Episoden dargestellt. In diesem Fall betrachtet der Agent ein Problemkontext von 19 Stopps verteilt in Berlin. Dabei findet er zum Ende hin eine gute Lösung, wobei ihm sich die Beste Lösung nicht erschließt, obwohl er diese bereits in den ersten 250 Iterationen entdeckt hatte. Dies lag zum einen daran, dass der Lernfaktor zu niedrig eingestellt war, aber zum anderen der Erkundungsfaktor den Agenten nicht genug zum erweiterten Erkunden des Environments beeinflusst hatte. Zur besseren Erkundung wurde sich einer angepassten dynamischen Epsilon-Erkundung bedient [SB18, S.30].

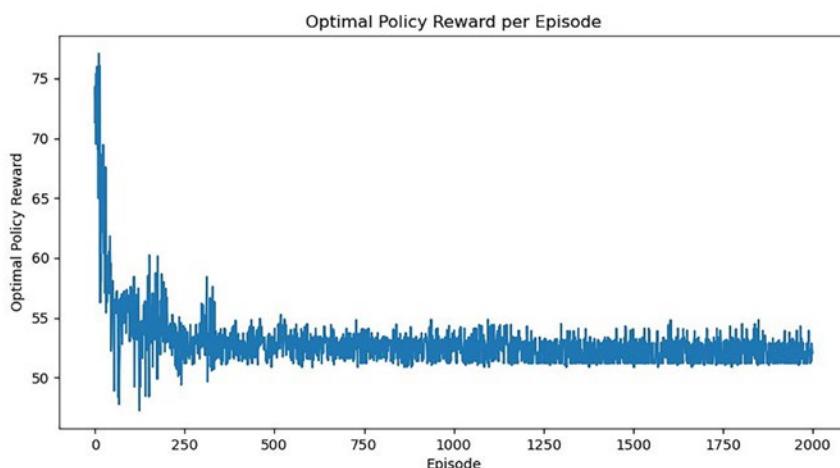


Abb. 2 Beispiel Durchlauf eines ML-Verfahren im lokalen Minimum

4 Interpretation

4.1 Evaluation der Ergebnisse

Im ersten Teil der Vergleichsanalyse und der verwendeten Basis Implementation ist zu erkennen, dass das ML-Verfahren in den ausgewählten Bewertungspunkten schlechter abgeschlossen hat als der Konkurrent. Ein Durchsetzen anhand der Laufzeit ist zwar möglich, dafür sollten aber die Touren nicht deutlich mehr Distanzaufwand in Anspruch nehmen. Die Bezahlung im Bereich der „Letzten-Meile“-Logistik erfolgt in der Regel pro ausgefahrenen Stopp. Aufgrund dessen wollen die Fahrer nicht durch schlecht optimierte Routen gebremst werden. Der Prototyp mit der Grundlagen Implementation von „REINFORCE mit Baseline“ erkennt im ersten Analyseschritt teilweise nahe gelegene Stopps nicht und wählt dadurch überflüssig längere Strecken aus. Er weicht im Durchschnitt +6,94 km und +14,13 min von der Lösung des Konkurrenten ab. Die Resultate des zweiten Analyseschritts zeigen eine Steigerung in den Bewertungspunkten gegenüber der ersten Version. Die Abweichungen der kumulierten Distanzen für die gebildeten Touren an den jeweiligen Tagen sind in Abb. 3 dargestellt. Die durchschnittlichen Abweichungen zwischen der Prototyp Lösung und OR Strategie liegt bei +6,51 km +11,01 min.

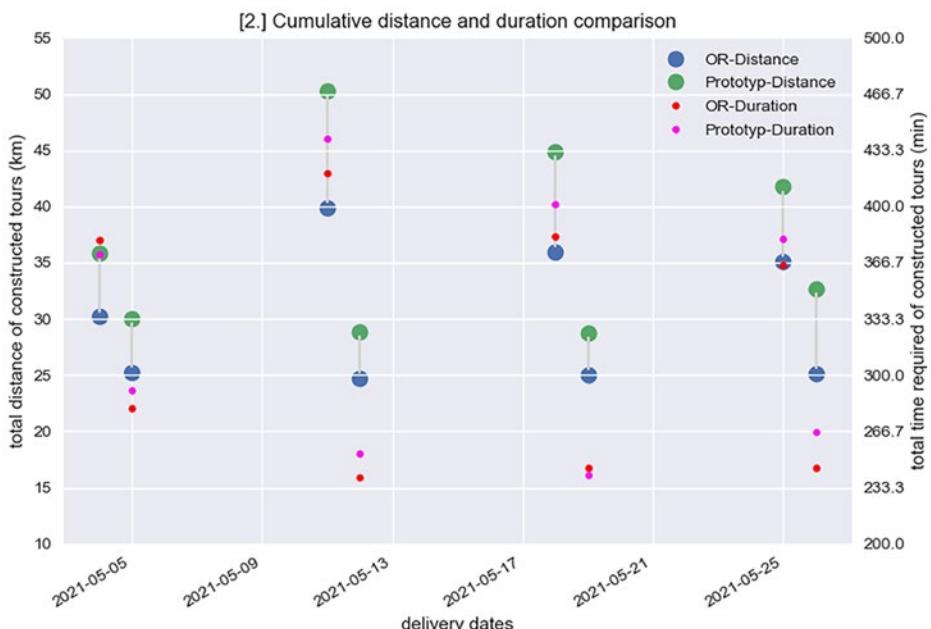


Abb. 3 Vergleich des kumulierten Distanz- und Zeitaufwandes zwischen OR Lösung und erweiterter Prototyp Lösung im zweiten Analyseschritt

Der Prototyp ist auch im zweiten Teil der Analyse bezüglich der Laufzeit zum Lösen von VRP-Instanzen deutlich schneller als der Konkurrent, weil er erfahrungsbasierte Entscheidungen treffen kann. Der nötige zeitliche Trainingsaufwand für den zweiten Analyseschritt wurde in Abb. 4 dargestellt.

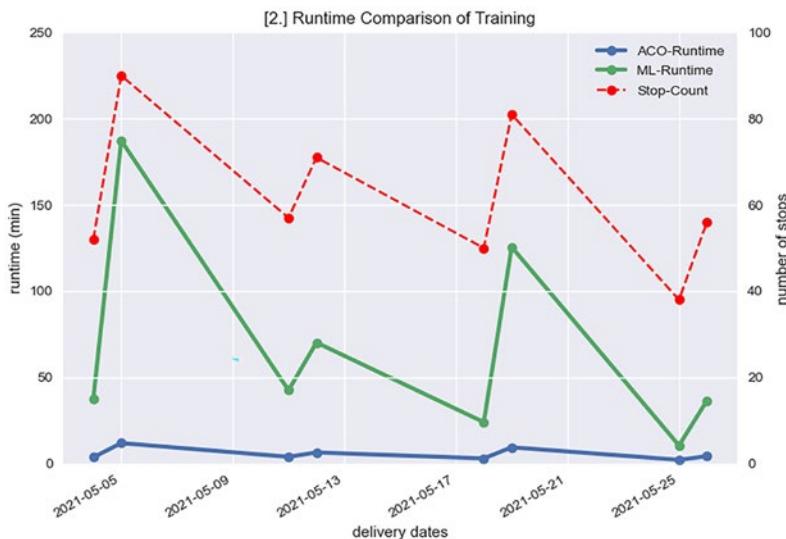


Abb. 4 Vergleich des Trainings-Laufzeitverhaltens zwischen OR und der Prototyp Lösung im zweiten Analyseschritt

Dieser ist für eine produktive Umgebung in Relation zu der Stoppanzahl noch zu aufwendig und sollte im Hinblick auf die Wettbewerbsfähigkeit optimiert werden. In Hinsicht auf die genannten Herausforderungen und damit indirekten Anforderungen an die Wettbewerbsfähigkeit des Prototyps, im Gliederungspunkt 1.3, lassen sich folgende resultierende Punkte, welche der Prototyp in der Lage ist zu erfüllen bzw. nicht zu erfüllen, ableiten:

- Nach Durchlauf des Trainingsprozesses ist der Prototyp in der Lage eine gute bis sehr gute Lösung zu finden.
- Der Prototyp kann ohne manuelles Eingreifen Lösungen erarbeiten.
- Im Testmodus wird das Ergebnis des Prototyps in Millisekunden bereitgestellt.
- Der Prototyp schneidet optimal auf hauptsächlich bekannten VRP-Instanzen ab. Sollten der überwiegende Teil der VRP-Instanz dem Prototyp nicht bekannt ist, wird der Prototyp keine annähernd optimale Lösung finden.

- Aufgrund dessen, dass der Prototyp auf nicht bekannte VRP-Instanzen schlechter reagiert, ist es auch nicht gegeben, dass er auf allen VRP-Instanzen grundlegend gleich arbeitet.

4.2 Vergleich zu kombinatorischen Optimierungsalgorithmen

Die Vergleichsanalyse ergab, dass das ML-Verfahren in seiner jetzigen Implementation nicht das gleiche Niveau der Lösungsqualität für VRP-Instanzen aufweisen kann wie Optimierungsalgorithmen. Des Weiteren zeigten die ausgewählten Anpassungsimpulse für die Basis Implementation von „REINFORCE mit Baseline“ keine signifikante Verbesserung. Die Anpassungsimpulse erreichten das der Prototyp in weniger Iterationen eine bessere Lösung erzielte. Die Verbesserung der Lösung weicht aber nur geringfügig von der Basis Implementation ab. Ungeachtet dessen weisen Vergleichsstudien von RL-Verfahren mit verschiedenen Bewertungskriterien und gleicher Laufzeit selten eine optimale Lösung gegenüber kombinatorischer Optimierungsalgorithmen auf [Ma20]. Dies wird damit begründet, dass es das grundlegende Ziel von RL-Verfahren ist, sowohl schlechte Lösungen zu vermeiden als auch eine durchschnittlich gute Lösung zu erzielen. Folglich wird somit auch das Ziel der Prototypen definiert. Im Gegensatz zu dem Prototyp betrachtet der Google-OR-Solver die Struktur des Problems ganzheitlich und erreicht, mit genug Laufzeit und Rechenleistung, asymptotisch eine optimale Lösung. Neben Google-OR gibt es auch andere Alternativen für das Lösen von kombinatorischen Optimierungsproblemen wie beispielsweise Concorde TSP Solver (<https://www.math.uwaterloo.ca/tsp/concorde.html>) oder die Dienste von openrouteservice (<https://openrouteservice.org/>).

4.3 Relevanz für die „Letzte-Meile“-Logistik

Im Bezugspunkt der Relevanz des Prototyps für die „Letzte-Meile“-Logistik und unter der Rücksprache kooperierenden Unternehmen, lassen sich folgende Resultate ableiten. Der Prototyp ist für ein geringes Stoppvolumen von weniger als 1200 Stopps pro Tag nicht effizient genug. Eine der Begründungen ist, dass der manuelle Dispositionsaufwand und das Einsetzen von kombinatorischen Optimierungsalgorithmen für weniger als 400 Stopps in einem kleinem „Letzte-Meile“ Gebiet deutlicher geringer sind als der Aufwand des Trainierens des Prototyps. Erst ab 400 Stopps schwächtelt der Google-OR-Solver in der Laufzeit mit 10 s. Unter der Retrospektive und dem Feedback des kooperierenden Unternehmens, wurde die Entscheidung des RL für die „Letzte-Meile“ überdacht und Ideen bezüglich anderer ML-Systeme ausgearbeitet. Ein Favorit ist es Supervised-Learning einzusetzen, um die Dispositionsart der Zulieferer-Disponenten zu lernen, sodass die ML-Komponente unterstützend wirkt und nicht automatisiert VRP-Instanzen löst.

5 Zusammenfassung und Ausblick

Der vorliegende Beitrag bietet ein Konzept zur Basis Implementation eines Reinforcement Learning (RL)-Verfahrens, in Form von „REINFORCE mit Baseline“ kombiniert mit einem Ant Colony Optimization Algorithm (ACO) zum Lösen eines Vehicle Routing Problems (VRP). Darüber hinaus stellt der entwickelte Prototyp auf Basis der Erweiterungsimplementierung von „REINFORCE mit Baseline“ eine Alternative gegenüber Drittanbietern wie beispielsweise Google-OR da. Des Weiteren werden Impulse zum adaptiven Lösen von VRP mittels verschiedener Optimierungsmechanismen innerhalb des Prototyps geboten. Mit Ausblick auf die weitere Forschungslandschaft von Machine-Learning (ML) und VRP haben die Untersuchungsergebnisse gezeigt, dass eine erweiterte Implementation von RL-Verfahren ein gutes bis sehr gutes Ergebnis im Kontext von einem NP-schweren Problem hinsichtlich der untersuchten Bewertungspunkte und gegenüber den klassischen Optimierungsverfahren erzielen können. Für die weitere Zukunft des Prototyps wäre ein Reduzieren der Komplexität der einstellbaren Parameter denkbar, um mögliches „overfitting“ zu vermeiden. Eine Erweiterbarkeit des Prototyps hinsichtlich anderer RL-Verfahren wäre möglich und bedeutsam, um die Resultate zu validieren. Eine Erkundung weiterer RL-Verfahren ist dahingehend wichtig, da im Gebiet von RL die geringsten Änderungen der Parameter oder des verwendeten Verfahrens starke Abweichungen im resultierenden Ergebnis auslösen. Es wäre in diesem Zusammenhang lohnend zu untersuchen, wie sich das Implementieren von Proximal Policy Optimization (PPO), welches eine neue Art von Policy-Gradienten-Methoden im RL beschreibt, auf die Lösungsqualität auswirkt. Dabei zeichnet sich PPO vor allem durch eine einfachere Implementation und einen einheitlichen Aufbau aus [Sc17].

Literaturverzeichnis

- [BRN15] Braekers, K., Ramaekers, K. & Nieuwenhuyse, I.: „The Vehicle Routing Problem: State of the Art Classification and Review“, *Computers & Industrial Engineering*, Vol. 99, 2015.
- [BU21] Bundesverband Paket und Expresslogistik e. V. (BIEK) (Hg.) (2021) KEP-Studie 2021 – Analyse des Marktes in Deutschland: Eine Untersuchung im Auftrag des Bundesverbandes Paket und Expresslogistik e. V. (BIEK), KE-CONSULT Kurte&Esser GbR [Online]. Verfügbar unter <https://www.biek.de/download.html?getfile=2897> (Stand: 17.08.2021).
- [Ch04] Chang, H. S., W. J. Gutjahr, Yang, J. & Park S.: „An ant system approach to Markov decision processes“, *Proceedings of the 2004 American Control Conference*, 3820-3825 vol.4, 2004.
- [DR56] Dantzig, G. B. & Ramser, J. H.: „The Truck Dispatching Problem“, *Management Science*, Vol. 6, No. 1, S. 80–91 [Online]. Verfügbar unter <http://www.jstor.org/stable/2627477>, 1959.
- [Go21] Google (2021) Routing Options [Online]. Verfügbar unter https://developers.google.com/optimization/routing/routing_options#local_search_options (Stand: 01.06.2021).
- [Lo20] Lorenz, U.: Reinforcement Learning, Springer Berlin Heidelberg, 2020.
- [Ma20] Mazyavkina, N., Sviridov, S., Ivanov S. & Burnaev E.: „Reinforcement Learning for Combinatorial Optimization: A Survey“, CoRR, abs/2003.03600, 2020.
- [Pu14] Puterman, M. L.: Markov Decision Processes: Discrete Stochastic Dynamic Programming [Online], Wiley. Verfügbar unter <https://books.google.de/books?id=VvBjBAAQBAJ>, 2014.
- [RA18] Richard, S. S. & Andrew, G. B.: Reinforcement Learning: An Introduction, A Bradford Book, 2018.
- [RSC20] Richter, R., Söding, M. & Christmann, G.: Logistik und Mobilität in der Stadt von morgen. Eine Expert*innenstudie über letzte Meile, Sharing-Konzepte und urbane Produktion, 2020.
- [SB18] Sutton, R. S. & Barto, A. G.: Reinforcement Learning: An Introduction [Online], The MIT Press. Verfügbar unter <http://incompleteideas.net/book/the-book-2nd.html>, 2018.
- [Sc17] Schulman, J., Wolski, P., Dhariwal, P., Radford, A., & Klimov, O.: „Proximal Policy Optimization Algorithms“, CoRR, abs/1707.06347, 2017.
- [Si17] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T. & Hassabis, D.: „Mastering the game of Go without human knowledge“, *Nature*, Vol. 550, 354, 2017.
- [Su20] Sun, Y., Wang S., Shen, Y., Li, X., Ernst, A. T., & Kirley, M.; Boosting Ant Colony Optimization via Solution Prediction and Machine Learning [Online], 2020.
- [VLM19] Vidal, T., Laporte, G. & Matl, P.: „A concise guide to existing and emerging vehicle routing problem variants“, *European Journal of Operational Research*, Vol. 286, 2019.

Einsatz von Machine Learning im Innovationsmanagement

Eine Literaturanalyse

Daniel Szafarski und Helmut Beckmann¹

Abstract: In volatilen Zeiten, in denen die verfügbare Datenmenge stetig steigt, nehmen Unternehmen das IM in den Fokus, um ihre Wettbewerbsfähigkeit zu sichern. Aus diesem Grund ist eine stetig steigende Anzahl an Publikationen zu verzeichnen, welche sich aktuell insbesondere mit der frühphasigen Ideengenerierung mithilfe natürlich sprachlicher Texte beschäftigen. Zur Analyse werden sowohl unüberwachte als auch überwachte Lernverfahren angewendet, wobei insbesondere Clustering-Algorithmen sowie Klassifikations- und Regressionsanalysen genutzt werden. Die präskriptive Analyse sowie die geschäftsbezogenen Innovationen sind in diesem Zusammenhang noch kaum untersucht, weswegen sich hieraus weitergehender Forschungsbedarf ableiten lässt.

Keywords: Innovationsmanagement, Machine Learning, Künstliche Intelligenz, Digitalisierung, Literaturanalyse

1 Einleitung

Durch eine volatile Umgebung, in der sich die Kundenbedürfnisse schnell verändern und die Innovationsgeschwindigkeit des Marktes steigt, rückt das Innovationsmanagement (IM) in den Fokus von Unternehmen, um ihre Wettbewerbsfähigkeit zu sichern [Ha18]. Aufgrund der enormen Relevanz von Innovationen für ein Unternehmen wird deren Entstehungsprozess bereits seit Jahrzehnten in der Forschung und Praxis stark thematisiert [Mü20]. Auch im Bereich des IM konnte die vielseitige Anwendbarkeit von Machine Learning (ML) in zahlreichen Case Studies demonstriert werden [Mü20]. Insbesondere vor dem Hintergrund stetig steigender Datenmengen können ML-Methoden nicht nur als Teil, sondern ebenfalls zur Optimierung des Prozesses genutzt werden [MG20, Mü20].

Wie die Boston Consulting Group in ihren jährlichen Umfragen aufzeigt, gehört im Jahr 2021 für zwei Drittel der Unternehmen das Thema „Innovation“ zu den wichtigsten drei strategischen Zielen. Gleichzeitig belegt die Studie, dass 80 % der B2C Unternehmen, also Unternehmen, die an Endkunden vertreiben, lediglich grundlegende Systeme zur Erhebung und Auswertung ihrer Kundendaten nutzen. Demgegenüber stehen fehlende Kompetenzen in der Auswertung und Nutzung von Daten [Bo21].

¹ Hochschule Heilbronn, Studiengang Wirtschaftsinformatik, Max-Planck-Str. 39, 74081 Heilbronn,
{daniel.szafarski; helmut.beckmann}@hs-heilbronn.de

Demzufolge werden aktuell keine beziehungsweise kaum ML-Methoden in den Unternehmen genutzt, obwohl diese wertvolle Erkenntnisse für das IM ermöglichen. In einem kürzlich publizierten Artikel von [Ha21] wird der Einsatz von Künstlicher Intelligenz (KI) für das IM untersucht und auf den zielführenden Einsatz in Organisationen hingewiesen. Einschränkend berücksichtigt werden muss, dass lediglich die ersten Prozessschritte des IM thematisiert werden und der dritte Schritt gänzlich ausgelassen wird. Die Literatur bietet zum aktuellen Zeitpunkt kaum einen Überblick über die Nutzung von ML-Verfahren für das IM, weswegen [Ro18] eine Forschungslücke in diesem Bereich aufzeigen. Dies wird als Motivation für die detaillierte und ganzheitliche Betrachtung des Einsatzes von ML-Methoden im Rahmen des IM durch den vorliegenden Beitrag genutzt. Hierbei soll mithilfe einer umfangreichen Literaturanalyse die folgende Forschungsfrage beantwortet werden:

Welche Machine Learning Methoden kommen im Rahmen des Innovationsmanagements zum Einsatz?

Ziel dieser Arbeit ist es, den aktuellen Forschungsstand aufzubereiten und hiermit einen Beitrag zum Überblick des Fachgebiets zu leisten. Dabei wird mithilfe einer Konzeptmatrix der Einsatz verschiedener Verfahren entlang des IM untersucht, wodurch eine Orientierungshilfe für weitergehende Forschungsvorhaben entsteht. Zur Erreichung dieses Ziels ist diese Arbeit wie folgt strukturiert. Im zweiten Abschnitt werden zunächst die theoretischen Grundlagen vermittelt, worauf aufbauend die Methodik zur systematischen Literaturanalyse dargestellt wird. In Abschnitt vier werden die Ergebnisse analysiert und veranschaulicht. Abschließend folgt im fünften Abschnitt ein Fazit der gewonnenen Erkenntnisse sowie eine Ableitung von möglichen zukünftigen Forschungsbedarfen.

2 Grundlagen und Begriffe

Der folgende Abschnitt erläutert die zwei Begriffe *Innovationsmanagement* und *Machine Learning Methoden*, um ein einheitliches Begriffsverständnis sowie den nötigen theoretischen Hintergrund zur Beantwortung der Forschungsfrage sicherzustellen.

2.1 Innovation und Innovationsmanagement

Der Innovationsbegriff wird in der Literatur sowohl für den Entstehungsprozess eines neuen Produktes als auch zeitgleich für das neue bzw. verbesserte Produkt selbst verwendet. Im Kontext dieser Arbeit wird die erstere Bedeutung verwendet, wonach allgemein betrachtet, unter dem Begriff Innovation eine neue Art und Weise verstanden werden kann, wie Tätigkeiten durchgeführt und anschließend vermarktet werden [Po90]. Im Detail bedeutet dies die Schaffung neuartiger, deutlich vom Vorgänger abgrenzbarer, Zweck-Mit-

tel-Kombinationen durch bislang unbekannte Produkte, Dienstleistungen und Prozesse bzw. deren Verknüpfungen [Ha16]. In Abhängigkeit des Betrachtungsgegenstands sowie der Auswirkung auf den Markt lassen sich dabei unterschiedliche Innovationstypologien differenzieren. In der vorliegenden Arbeit wird hierfür die Untergliederung in technische Innovationen, die sowohl Prozesse als auch Produkte beinhalten, *organisationale* sowie *geschäftsbezogene Innovationen* verwendet [ZW95].

Eine Innovation grenzt sich durch die Kommerzialisierung eines Unternehmens von einer Erfindung ab [Ro88]. Folglich lässt sich Innovation als Kernprozess eines Unternehmens definieren, um innerhalb einer volatilen Umgebung zu überleben [SJ07]. Die betrieblichen Aktivitäten zur Steuerung und Ausgestaltung der Innovationsaktivitäten eines Unternehmens werden unter dem Begriff des IM zusammengefasst. Hierunter fallen sowohl strategische Tätigkeiten zur Definition der Innovationsziele wie auch operative Aufgaben, die unter anderem die Planung, Organisation und Kontrolle des Innovationsprozesses sowie die Schaffung entsprechender Rahmenbedingungen innerhalb des Unternehmens beinhalten [PS96]. Nachfolgend werden Aspekte wie bspw. die Innovationskultur nicht weiter thematisiert und der Schwerpunkt der Analyse auf die Tätigkeiten, die im Zusammenhang mit dem Innovationsprozess stehen, gelegt.

Für den Unternehmenserfolg sowie den Erfolg des IM ist es von besonderer Importanz, eine systematische Durchführung des Innovationsprozesses sicherzustellen [GWR09]. Folglich haben sich in der Vergangenheit diverse Vorgehensmodelle etabliert, die sich aufgrund der Diversität hinsichtlich Zielsetzung und Innovationssubjekt legitimieren lassen [Ev10]. Viele Ansätze eint die grundlegende Aufgliederung des Prozesses in die Bereiche Ideen, Projekte und Produkte nach dem Konzept des Innovationstrichters, wonach aus einer Vielzahl von Ideen am Ende des Prozesses eine Handvoll marktreifer Produkte entsteht [FF11]. Im Wesentlichen wird diese Grundidee in dem Dreiphasenmodell des Innovationsprozesses nach [Th92] abgebildet. In der ersten Phase *Ideengenerierung* bestimmt das Unternehmen sein Suchfeld und beginnt mit der Findung bzw. dem Vorschlag von neuen Ideen. Diese werden im zweiten Schritt *Ideenakzeptierung* hinsichtlich ihrer Sinnhaftigkeit für eine potenzielle Umsetzung hin validiert und anschließend Realisierungspläne entwickelt, über welche am Ende der Phase entschieden wird. Fällt die Ideenbewertung positiv aus, werden diese im dritten Prozessschritt, der *Ideenrealisierung*, tatsächlich implementiert und anschließend am Markt vertrieben sowie der Erfolg gemessen. [Th92]. Dieses Prozessmodell repräsentiert mit seinen drei Phasen die wesentlichen Prozessschritte der in [Ev10] untersuchten zwölf verschiedenen Modelle. Zum selben Ergebnis kommen ebenfalls [FF11].

2.2 Machine Learning Methoden

Die steigende Menge an Daten folgert den Wunsch nach automatisierter Verarbeitung, um neue Erkenntnisse sowie Innovationen zu generieren [MG20, Mu12]. In diesem Zusammenhang hat sich innerhalb der Informatik der Teilbereich der KI herausgebildet, welches versucht, mithilfe von Computern intelligentes Verhalten zu simulieren. Es fehlt bislang allerdings an einer einheitlichen, über mehrere Disziplinen anerkannten Definition, da KI als Sammelbegriff für Datenanalyseprozesse zum Erhalt oder Anwenden von Wissen verwendet wird [Gr14]. Während viele Publikationen der untersuchten Forschungsfrage diesen Begriff nutzen, wird nachfolgend stattdessen auf den Teilbereich des ML genauer eingegangen. Dieser ermöglicht es, ein System mit menschenähnlicher Intelligenz ohne explizite Programmierung zu erzeugen. Hierdurch erhält das System die Möglichkeit, selbstständig in einer spezifischen Aufgabe aus Erfahrungen zu lernen und sich hierdurch stetig zu verbessern [JM15]. Hierbei kann in Abhängigkeit von Grad des Analysewertes die rein beschreibende (*deskriptive*), erklärende (*diagnostische*), vorausschauende (*prädiktive*) sowie optimierende (*präskriptive*) Analyse differenziert werden [Ba20].

Außerdem werden in Abhängigkeit der Aufgabe, die das System löst, sowie der Vorgehensweise innerhalb des Trainings, die nachfolgend dargestellten Typen unterschieden. Erstens wird der Begriff des *überwachten Lernens* (engl. *supervised learning*) verwendet. Hierunter wird ein Lernalgorithmus verstanden, welcher durch die Abbildung eines bzw. jedes Eingabewertes x auf eine Ausgabe y eine möglichst akkurate Vorhersage trifft. In Abhängigkeit der Ausgabevariablen $y^{(i)}$ kann die Art des Problems bestimmt werden. Ist diese kategorisch, so handelt es sich um eine Klassifizierung oder eine Mustererkennung bzw. eine Anomalieerkennung. Handelt es sich um ein reellwertiges oder ordinale Skalar, so liegt meist ein Regressionsproblem vor [JM15, MRT18, Mu12].

Zweitens wird der Bereich des *unüberwachten Lernens* (engl. *unsupervised learning*) differenziert. Im Gegensatz zu überwachtem Lernen sind die Vorhersagewerte sowie die Aufgaben unbekannt und können folglich nicht zum Überwachen des Lernalgorithmus verwendet werden. Ein typischer Anwendungsbereich ist das Clustering und die Assoziationsanalyse, wobei darüber hinaus unüberwachtes Lernen bspw. zur Dimensionsreduktion hochdimensionaler Daten, Erkennung von Graphstrukturen, Konzeptbeschreibungen oder Matrixvervollständigung genutzt werden kann [JM15, MRT18, Mu12].

Drittens wird neben den beiden zuvor genannten Verfahren das *bestärkende Lernen* (engl. *reinforcement learning*) unterschieden. Hierbei wird ein Ergebnis entweder belohnt oder bestraft, um eine entsprechende Anpassung im Training sowie nachgefolgerter Prognose zu erzielen. Im Gegensatz zum überwachten Lernen wird hierbei nicht explizit die korrekte Ein- bzw. Ausgabe präsentiert [JM15, MRT18, Mu12].

Neben den bereits beschriebenen Ansätzen werden heutzutage weitere Mischformen sowie je nach Publikation zusätzliche Verfahren differenziert [JM15]. Hierzu zählen bspw. das selbstüberwachte sowie Transferlernen. Diese sind allerdings in der Praxis aktuell noch weniger verbreitet und werden nachfolgend nicht genauer erläutert [AbPH21].

3 Methodik zur Literaturanalyse

Zur Beantwortung der vorliegenden Forschungsfrage wird die strukturierte Literaturanalyse nach [Fe06] sowie [WW02] verwendet. Gemäß dem Vorgehensmodell von [Fe06] wurde im ersten Abschnitt die Problemdefinition beschrieben, welche durch die folgenden Schritte Literatursuche und -auswertung, Analyse und Interpretation sowie Präsentation der identifizierten Literatur vervollständigt wird [Fe06]. Für die Literatursuche werden gemäß den Leitlinien geeignete Auswahlkriterien ausgewählt, welche nachfolgend dargestellt werden. Im Rahmen dieses Beitrags werden lediglich Publikationen aus den Jahren 2018 (inkludiert) bis 2022 (soweit vorhanden) betrachtet und auf einen historischen Rückblick verzichtet. Dies deckt sich ebenfalls mit der Publikation von [Ro18], welche im Jahr 2018 noch eine Forschungslücke in diesem Bereich identifiziert haben. Eine weitere Einschränkung wurde durch die Fokussierung auf deutsche sowie englischsprachige Publikationen sowie deren öffentliche Verfügbarkeit gewählt. Für *SpringerLink* wird die zusätzliche Einschränkung auf Conference Paper und Artikel vorgenommen. Als Schlagwörter werden die Begriffe *innovation management* und *machine learning* verwendet und in Form eines Suchterms in die verwendeten Literaturdatenbanken eingegeben. Diese sind: *AIS Electronic Library*, *IEEE Xplore Digital Library*, *SpringerLink* und *Science Direct*. Hierdurch konnten in Summe 398 Publikationen identifiziert werden. Nach Entfernung der Duplikate, Sichtung der Abstracts sowie einer Vorwärts- und Rückwärtssuche konnte diese Anzahl auf 38 Beiträge, die für diese Arbeit relevant sind, reduziert werden. Diese werden im anschließenden Abschnitt in Form einer Konzeptmatrix dargestellt und analysiert.

4 Forschungsergebnisse

Im nachfolgenden Abschnitt werden die Ergebnisse der Literaturanalyse dargestellt. Hierfür wird nach dem Vorgehen von [WW02] eine Konzeptmatrix erstellt, welche die wesentlichen Konzepte, die in Abschnitt zwei dargestellt sind, aufgreift. Dabei bilden die Phasen des IM-Prozesses sowie die dargestellten ML-Verfahren das Zentrum der Konzeptmatrix, um die Problemstellung dieser Forschungsarbeit zu untersuchen. Ergänzend hierzu werden der Ursprung, die Innovationstypologie, die Datenart sowie der Analysetyp und -aufgabe abgetragen. Wird eine der Ausprägungen in der untersuchten Publikation dargestellt oder lässt sich diese aufgrund der Beschreibung folgern, so wird dies durch ein *X* in der Konzeptmatrix dargestellt. Folgend werden die Besonderheiten erläutert und die Konzeptmatrix in Tabelle eins vollständig dargestellt.

4.1 Allgemein

Innerhalb der untersuchten Stichprobe lassen sich lediglich drei Publikationen einem praktischen Ursprung zuordnen. Diese weisen im Kontext der untersuchten Forschungsfrage allerdings keine Unterschiede zu den anderen betrachteten Publikationen auf, weswegen auf eine Differenzierung im weiteren Verlauf verzichtet wird. Eine weitere Betrachtungsdimension im Zusammenhang mit dem Ursprung stellt die Anwendungsdomäne der Publi-

kation dar. In diesem Punkt ist die Fokussierung auf eine spezielle Domäne in 44,7 % der Stichprobe festzustellen, wobei sich fünf Publikationen auf den Bereich der Produktion sowie zwei weitere im Spezifischen der Automobilbranche zuwenden. Setzt man die Publikationen in den zeitlichen Verlauf, so ist ein kontinuierlicher Anstieg innerhalb des Betrachtungszeitraums festzustellen. Während im Jahr 2018 vier und 2019 fünf Publikationen veröffentlicht worden sind, sind es im Jahr 2020 bereits 11 und im Jahr 2021 16 Arbeiten. Dies veranschaulicht zum einen die Aktualität dieses Themenfeldes und weist außerdem die steigende Forschungsaktivität nach.

		Ursprung	Typologie	IM-Prozess	ML-Verfahren	Datenart	Analysetyp	Analyseaufgabe
	Akademisch Praktisch	Domäne	technische Innovationen organisational Innovationen geschäftslogische Innovationen	Ideengenerierung Ideenakzeptierung Ideenreiseleitung Phasenunabhängig überwachtes Lernen	unüberwachtes Lernen Bestückende Lernen Selbstüberwachtes Lernen Transferlernung	nicht spezifiziert Natürliche Sprache Nummerische Daten Bilder / Videos	deskriptive Analyse diagnostische Analyse prädiktive Analyse präskriptive Analyse	Clusteranalyse Klassifikation Regression Anomalieerkennung Konzeptbeschreibung Assoziationsanalyse Dimensionsreduktion Visual Analytics nicht spezifiziert
Literaturverweis								
[AT18]	X	-	X	X	X	X	X	X
[Ba21]	X	Produktion	X X	X	X	X X	X X	X
[BAE19]	X	Gesundheitswesen	X	X X X	X	X	X X X	X
[BI19]	X	Automobilessektor	X	X	X	X X	X X X	X
[BM21]	X	-	X X	X X X	X	X	X X X	X
[Bu21]	X	Produktion	X	X	X	X	X X X	X
[Da20]	X	-	X	X X	X	X	X X X	X X
[EGC18]	X	Bildungswesen	X	X	X	X	X X	X X
[ERH19]	X	-	-	X	X X	X X	X X	X X
[FNv21]	X	-	X	X	X	X	X X	X X
[Gi18]	X	-	X	X	X	X X X	X X	X X X
[Gu20]	X	-	X	X	X X X	X X	X X	X X
[Ha21]	X	-	X X	X X	X X X	X X	X X X	X X
[HK21]	X	Sport	X	X X	X	X X	X X X	X X
[Hu21]	X	-	X X	X	X X X	X X X	X X X	X X
[JSK21]	X	-	X X	X X X	X X X	X X	X X	X X
[Ks20]	X	-	X	X	X X	X X	X X	X X
[KG19]	X	-	X	X	X X	X X	X X	X X
[KZE20]	X	Energiesektor	X X	X X	X X	X X	X X	X X
[LH21]	X	-	X X	X X	X X	X X	X X X	X X X
[LHK21]	X	-	X X	X X	X X	X X	X X X	X X X
[Li20]	X	Produktion	X	X	X	X X	X X	X X
[LZZ20]	X	-	X	X	X	X X	X X	X X
[MG20]	X	-	X	X	X X	X X	X X	X X
[MUU21]	X	Heritage	X	X	X	X X	X X	X X
[NB19]	X	-	X	X	X	X X	X X X	X X X
[NBK21]	X	-	X	X	X	X X	X X X	X X X
[Sb18]	X	Bildungswesen	X	X X X	X X	X X	X X X	X X X
[Sj21]	X	Produktion	X	X	X	X X	X X	X X
[Ta20]	X	Supply Chain	X	X	X	X X	X X	X X
[TMT20]	X	-	X	X	X X	X X	X X X	X X X
[Ti21]	X	Personalwesen	X X	X	X X X	X X X	X X X	X X X
[Ya19]	X	Automobilessektor	X X	X	X	X X X	X X X	X X X
[YY20]	X	-	X X	X	X	X X X	X X X	X X X
[Zh21]	X	Produktion	X	X	X	X X	X X	X X
[ZLT21]	X	Supply Chain	X	X	X X	X X	X X X	X X X
[ZYC20]	X	Robotik	X	X	X X	X X	X X	X X
[ZZB18]	X	-	X	X	X X	X X X	X X X	X X X

Tab. 1: Konzeptmatrix (eigene Darstellung)

4.2 Forschungsschwerpunkte im IM

Zur ganzheitlichen Beantwortung der Forschungsfrage ist es wichtig, die Publikationen hinsichtlich ihrer Fokussierung innerhalb des IM zu untersuchen. Betrachtet man die Innovationstypologie, so sind fast alle Publikationen (89,5 %) thematisch der technischen Innovation zuordenbar. Einige Publikationen wie bspw. [KZE20, ZYC20] verweisen dabei explizit auf ihre Fokussierung in diesem Bereich, die meisten Publikationen sind dieser Typologie aber durch die thematisierte Prozessoptimierung zuzuordnen. So verweisen zahlreiche Publikationen auf die Potenziale von ML zur Automatisierung der Trendforschung. Hierbei ist es wichtig anzumerken, dass ML bislang nicht für die eigentliche Generierung innovativer neuer Ideen, sondern vielmehr für die Identifikation spannender Ideen aus bisherigen Konzepten eingesetzt wird.

Ein Anteil von 13 Publikationen lässt sich der organisationalen Innovation zuordnen. Hierbei werden übergreifende Themen beschrieben, welche das gesamte Unternehmen prägen. Als Beispiel können [ERH19] gelten, welche ein Bewertungsschema zur Messung des Innovationsfortschritts sowie zur Ableitung weiterer Maßnahmen beschreiben. Ferner betrachten [ZYC20] die Auswahl sowie Suche geeigneter Talente für die Umsetzung von technologischen Innovationsprojekten. 11 (84,6 %) der 13 Publikationen, die dieser Typologie zuzuordnen sind, sind gleichzeitig der technischen Innovation angehörig. Den geschäftsbezogenen Innovationen lassen sich lediglich zwei Publikationen zuordnen. Beide erwähnen keine ML-Verfahren explizit, sondern weisen allgemein auf die Potenziale von KI im Kontext des Business Model Innovation hin. Hierbei verweisen bspw. [Bu21] auf die Analyse von Nutzungsdaten zur Antizipation von Nutzerwünschen und der Ableitung von neuen Funktionen oder Produkten. Gleichzeitig kann potenziell ebenfalls eine Integration von ML-Verfahren im eigentlichen Geschäftsmodell erfolgen, um Konzepte wie dynamische Preisgestaltung oder Massenindividualisierung zu ermöglichen.

Neben der Innovationstypologie lässt sich die Zuordnung zu einem oder multiplen Prozessschritten im IM-Prozess nach [Th92] zur Klassifizierung der Publikationen betrachten. Hierbei muss jedoch berücksichtigt werden, dass 28,9 % der Publikationen keiner Phase zugeordnet werden können. Grund hierfür ist die übergeordnete oder allgemeine Betrachtung des Themenfelds. Die erste Phase der Ideengenerierung wird mit einem besonders hohen Anteil von 47,4 % angewendet. Hierbei wird in den Publikationen vor allem die automatisierte Erkennung von Trends thematisiert. Die folgende Phase der Ideenakzeptierung wird in 28,9 % der Stichprobe thematisiert. Ein Kernthema, das hierbei durch mehrere Publikationen diskutiert wird, ist die Prädiktion des Erfolgs einer Innovation. Sechs Publikationen lassen sich mehr als einer Phase zuordnen. Weiterhin werden hierbei immer mindestens die ersten zwei Phasen zusammen genannt, in drei Fällen sogar alle drei Phasen. Diese betrachten die Anwendung von ML ganzheitlich im Kontext des IM. Exemplarisch können hierfür [JSK21] genannt werden, welche das Einsatzpotenzial von Chatbots entlang des IM-Prozesses ermitteln. Die Ideenrealisierung ist mit einem Anteil von 21,1 % am wenigsten in der Stichprobe vertreten. In diesem Anteil wird mit 62,5 % deutlich häufiger als in der Gesamtstichprobe auf eine spezifische Domäne verwiesen.

4.3 Anwendung von ML-Methoden im Kontext des IM

Zur ganzheitlichen Betrachtung der Forschungsfrage ist es wichtig, neben dem IM ebenfalls die Verwendung von ML in diesem Kontext detailliert zu betrachten. Eine wesentliche Differenzierung kann hierbei durch die Untersuchung des Lernverfahrens erfolgen. Hierbei ist festzustellen, dass im Wesentlichen nur auf die zwei populären Ansätze des überwachten (36,8 %) bzw. unüberwachten Lernens (39,5 %) verwiesen wird. Diese sind nahezu gleich häufig vertreten, wobei keine signifikante Häufung eines Ansatzes für spezifische IM Teilbereiche vorhanden ist. Sechsmal wird gleichzeitig auf beide Verfahren verwiesen. Dies liegt zum einen an den unterschiedlichen Aufgaben, die mithilfe der Verfahren umgesetzt werden und zum anderen an Publikationen wie [Ha21], die einen allgemeinen Überblick über Implementierung liefern. Alle weiteren untersuchten ML-Verfahren werden zusammengerechnet, lediglich in fünf Publikationen erwähnt und sind folglich für eine weitere Betrachtung nicht relevant.

Eine weitere Informationsquelle ist die verwendete Datenart, da hiermit weitere Informationen zu den durchgeföhrten Analysen gewonnen werden können. In diesem Zusammenhang ist die hohe Anzahl (60,5 %) an Publikationen auffällig, die auf natürlich sprachliche Texte als Analyseobjekt zurückgreift. Bei genauerer Betrachtung lassen sich diese nochmals in Patente, wissenschaftliche Publikationen sowie weitere Dokumente wie bspw. unternehmensinterne Dokumentationen untergliedern. In der Phase der Ideengenerierung wird diese Datenform insbesondere bei der Trenderkennung zu 77,8 % genutzt. Numerische Daten als weitere Datenart wurden im Vergleich insgesamt neunmal und Bild bzw. Videodaten dreimal verwendet.

Mit Ausnahme der präskriptiven Analyse, welche lediglich zwei Mal genannt wird, werden die weiteren Analysetypen in einem ähnlichen Umfang verwendet. Innerhalb der untersuchten Publikationen zeigt sich, dass die deskriptive sowie prädiktive Analyse in 52,6 % und die diagnostische Analyse in 42,1 % der Publikationen angewandt wurde. Acht Publikationen nutzen alle drei zuvor genannten Analysetypen. Weiterhin ist bei der Untersuchung der Analyseaufgabe ein besonders hoher Anteil (26 Nennungen) der Clusteranalyse zu erkennen. Gefolgt wird diese von der Regression (12 Nennungen) sowie der Klassifikation (elf Nennungen). Mit acht Nennungen wird die visuelle Analyse sowie die Konzeptbeschreibung gleich oft erwähnt. Die weiteren Aufgaben sowie der Anteil an nicht zuordnabaren Publikationen beträgt fünf.

5 Diskussion und Ausblick

Durch die Analyse der Stichprobe konnten aktuelle Trends im Einsatz von ML für das IM identifiziert werden. Im Allgemeinen ist eine steigende Anzahl an Publikationen zu verzeichnen, wobei diese zumeist die technische Innovation fokussieren. Dies kann darin begründet sein, dass ML häufig in Kombination mit der Optimierung bzw. Automatisierung von Prozessen als Teilgebiet der technischen Innovation in Kontext gebracht wird. Hierbei werden im aktuellen Status-Quo insbesondere die initialen Prozesse des IM betrachtet und dabei die erste Phase der Ideengenerierung unterstützt. Diese Erkenntnis manifestiert sich, da die Publikation von [Ha21] ebenfalls diese Phase in ihrer Analyse fokussieren. Das Handlungsobjekt in diesen Prozessen stellen zumeist Texte in natürlicher Sprache, überwiegend in Form von Patenten, aber auch wissenschaftlichen Publikationen oder sonstigen, internen Unternehmensdaten dar. Schwerpunkte der Analysen sind dabei zumeist die Selektion, Aggregation oder Identifikation interessanter Konzepte innerhalb der Vielzahl an bereits existierenden Dokumenten. Auch im Zusammenhang mit numerischen Daten beschränkt sich der Einsatz von ML auf die Identifikation spannender Zusammenhänge, die wiederum als Grundlage für die Konzeptionierung neuer Innovationen dienen können. Damit lässt sich in beiden Fällen folgern, dass im aktuellen Status Quo ML den Menschen nur durch Impulse unterstützt, selbstständig allerdings keine Erfindungen erzeugen kann. Stützen lässt sich diese Aussage außerdem durch die Eigenschaften der meistbetrachteten Dokumente, da Patente und wissenschaftliche Publikationen natürlichweise neuartige Innovationen und Fortschritte verkörpern.

Eine weitere Dimension, die diese Erkenntnis ebenfalls unterstützt, ist die Analyseaufgabe, bei der eine besonders starke Ausprägung auf Clustering-Algorithmen, zur Identifikation neuartiger Trends, festzustellen ist. Hinsichtlich der verwendeten ML-Verfahren ist eine Fokussierung auf die beiden traditionellen Lernverfahren des unüberwachten bzw. überwachten Lernens zu erkennen. In der untersuchten Stichprobe sind die Verfahren des bestärkenden, selbstüberwachten oder Transferlernens bislang noch schwach ausgeprägt. Dies könnte durch die relative Neuheit dieser Ansätze sowie der limitierten Auswahl etablierter Modelle im Vergleich zum überwachten und unüberwachten Lernen erklärt werden. Die Publikation von [HiKa21]technical advances in machine learning techniques for natural language understanding, such as semantic word space models and semantic network analytics, have made it practical to capture descriptions of early-stage, need-solution pairs mentioned anywhere in the open, textual content of the Internet. Producers - and anyone - can now thus look for user innovations posted on the web that may involve either known or newly defined needs coupled to new solutions that are gaining traction. This is important because, as is now understood, users, rather than producers, tend to pioneer functionally new products and services for which both the need and the solution may be novel. In this paper, we demonstrate via a case study both the practicality and the value of searching for early-stage need-solution pairs via machine learning methods and assessing the likely general interest in each user-generated innovation by also identifying the trends in posting and query frequencies related to it. The new need-solution pair search method we describe and test here can, we claim, serve as a very valuable complement to traditional market research techniques and practices.“,“author“:[{„dropping-particle“:“,“fa-

mily“:“Hippel“,“given“:“Eric“,“non-dropping-particle“:“von“,“parse-names“:false,“suffix“:““}, {„dropping-particle“:““,“family“:“Kaulartz“,“given“:“Sandro“,“non-dropping-particle“:““,“parse-names“:false,“suffix“:““}],“container-title“:“Research Policy“,“id“:“ITEM-1“,“issue“:“8“,“issued“:{„date-parts“:[[],“2021“]}],“page“:“104056“,“publisher“:“Elsevier B.V.“,“title“:“Next-generation consumer innovation search: Identifying early-stage need-solution pairs on the web“,“type“:“article-journal“,“volume“:“50“},“uris“:[„<http://www.mendeley.com/documents/?uuid=f21b5abd-acb5-4363-ac9f-29fc1c96841f>“}],“mendeley“:{„formattedCitation“:“[HiKa21]“,“plainText-FormattedCitation“:“[HiKa21]“,“previouslyFormattedCitation“:“[HiKa21]“},“properties“:{„noteIndex“:0},“schema“:“<https://github.com/citation-style-language/schema/raw/master/csl-citation.json>“} verweist in ihrer Future Work auf eine potenzielle Anwendung von BERT, einem neuronalen Netz zur Verarbeitung natürlicher Sprache, das auf dem Ansatz des selbstüberwachten Lernens sowie dem Transferlernen beruht. Folglich ist in Zukunft ein steigender Trend dieser neueren Ansätze zu erwarten.

Limitationen der dargestellten Erkenntnisse ergeben sich zum einen aus der beschränkten Stichprobengröße sowie zum anderen aus der uneinheitlichen sowie nicht trennscharfen Definition des Begriffs ML, welcher in den untersuchten Publikationen zumeist ebenfalls als Synonym für KI verwendet wird. Folglich kann eine vollständige Identifikation aller relevanter Publikationen nicht gewährleistet werden. Außerdem bietet die Analyse der Publikationen zur Befüllung der Konzeptmatrix Interpretationsspielraum, wobei die Zuordnung zu den Konzepten nach bestem Wissen und Gewissen erfolgte.

Ein weiterer Forschungsbedarf besteht für die präskriptive Analyse sowie die geschäftsbezogenen Innovationen, da diese bislang beide in lediglich zwei Publikationen thematisiert werden. Außerdem könnten die genauen ML-Verfahren untersucht werden und eine Validierung dieser Arbeit erfolgen.

Literaturverzeichnis

- AbPH21 Abdel-Karim, Benjamin M.; Pfeuffer, Nicolas; Hinz, Oliver: Machine learning in information systems - a bibliographic review and open research issues. In: *Electronic Markets* vol. 31, Electronic Markets (2021), Nr. 3, pp. 643–670
- AT18 Aristodemou, L.; Tietze, F.: The state-of-the-art on Intellectual Property Analytics (IPA): A literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (IP) data. World Patent Information October/55, S. 37–51, 2018.
- Ba20 Balali, F. et al.: Data Intensive Industrial Asset Management. Data Intensive Industrial Asset Management, S. 105–113, 2020.
- Ba21 Babu, M. M. et al.: Exploring big data-driven innovation in the manufacturing sector: evidence from UK firms. Annals of Operations Research 0123456789, 2021.
- BAE19 Baskici, C.; Atan, S.; Ercil, Y.: Forecasting of innovation in the light of semantic networks. Procedia Computer Science 158, S. 443–449, 2019.
- Bl19 Bloomfield, R. et al.: Disruptive Innovations and Disruptive Assurance: Assuring Machine Learning and Autonomy. Computer 9/52, S. 82–89, 2019.
- BM21 Brandtner, P.; Mates, M.: Artificial Intelligence in Strategic Foresight – Current Practices and Future Application Potentials: The 2021 12th International Conference on E-Business, Management and Economics (ICEME 2021). Association for Computing Machinery, New York, NY, USA, S. 75–81, 2021.
- Bo21 Boston Consulting Group: Overcoming the Innovation Readiness Gap, 2021.
- Bu21 Burström, T. et al.: AI-enabled business-model innovation and transformation in industrial ecosystems: A framework, model and outline for further research. Journal of Business Research June 2020/127, S. 85–95, 2021.
- Da20 Daxenberger, J. et al.: ArgumenText: Argument Classification and Clustering in a Generalized Search Scenario. Datenbank-Spektrum 2/20, S. 115–121, 2020.
- EGC18 Escandón-Quintanilla, M. L.; Gardoni, M.; Cohendet, P.: Improving concept development with data exploration in the context of an innovation and technological design course. International Journal on Interactive Design and Manufacturing 1/12, S. 161–172, 2018.
- ERH19 Eljasik-Swoboda, T.; Rathgeber, C.; Hasenauer, R.: Artificial intelligence for innovation readiness assessment. 2019 IEEE International Symposium on Innovation and Entrepreneurship, TEMS-ISIE 2019, S. 14–19, 2019.
- Ev10 Eveleens, C.: Innovation management; a literature review of innovation process models and their implications. Science 2010/800, S. 900–916, 2010.
- Fe06 Fettke, P.: State-of-the-art des state-of-the-art: Eine untersuchung der Forschungsmethode „review“ innerhalb der Wirtschaftsinformatik. Wirtschaftsinformatik 48/4, S. 257–266, 2006.
- FF11 Franken, R.; Franken, S.: Innovationsmanagement. Integriertes Wissens- und Innovationsmanagement, S. 192–298, 2011.

- FNv21 Falco, B. D.; Neubert, M.; van der Krogt, A.: Usage of Machine Learning To Predict Market Attractiveness in the Context of Internationalization. Proceedings of the 8Th International Conference Innovation Management, Entrepreneurship and Sustainability (Imes 2020) Imes, S. 140–152, 2021.
- Gi18 Giaccardi, E.: Things making things: Designing the internet of reinvented things. IEEE Pervasive Computing 3/17, S. 70–72, 2018.
- Gr14 Grewal, D. S.: A Critical Conceptual Analysis of Definitions of Artificial Intelligence as Applicable to Computer Engineering. IOSR Journal of Computer Engineering 2/16, S. 9–13, 2014.
- Gu20 Guarino, G. et al.: SummaTRIZ Summarization Networks for Mining Patent Contradiction. 2019 IEEE International Symposium on Innovation and Entrepreneurship, TEMS-ISIE 2019, S. 979–986, 2020.
- GWR09 Gaubinger, K.; Werani, T.; Rabl, M.: Praxisorientiertes Innovations- und Produktmanagement. Praxisorientiertes Innovations- und Produktmanagement, S. 4–16, 2009.
- Ha16 Hauschildt, J. et al.: Innovationsmanagement. Vahlen, 2016.
- Ha18 Hanschke, I.: Digitalisierung und Industrie 4.0-einfach und effektiv: Systematisch und lean die Digitale Transformation meistern. Carl Hanser Verlag GmbH Co KG, München, 2018.
- Ha21 Haefner, N. et al.: Artificial intelligence and innovation management: A review, framework, and research agenda. Technological Forecasting and Social Change October 2020/162, S. 120392, 2021.
- HK21 Hippel, E. von; Kaulartz, S.: Next-generation consumer innovation search: Identifying early-stage need-solution pairs on the web. Research Policy 8/50, S. 104056, 2021.
- Hu21 Hutchinson, P.: Reinventing Innovation Management: The Impact of Self-Innovating Artificial Intelligence. IEEE Transactions on Engineering Management 2/68, S. 628–639, 2021.
- JM15 Jordan, M. I.; Mitchell, T. M.: Machine learning: Trends, perspectives, and prospects. Science 6245/349, S. 255–260, 2015.
- JSK21 Johannsen, F.; Schaller, D.; Klus, M. F.: Value propositions of chatbots to support innovation management processes. Information Systems and e-Business Management 1/19, 2021.
- Kh20 Khiat, A.: Learning a lightweight representation: First step towards automatic detection of multidimensional relationships between ideas. 2019 IEEE International Symposium on Innovation and Entrepreneurship, TEMS-ISIE 2019, S. 230–233, 2020.
- KZE20 Khalyasmaa, A. I.; Zinovieva, E. L.; Eroshenko, S. A.: Intelligent Data Analytics System for Innovation Efficiency Assessment Based on Risk Analysis. 2020 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT), S. 479–482, 2020.
- LH21 Luo, X.; Huang, R.: Application of Adaptive Co-evolutionary Algorithms to Technology Innovation Management. Wireless Personal Communications 0123456789, 2021.
- LHK21 Lee, C; Hong, S.; Kim, J.: Anticipating multi-technology convergence: a machine learning approach using patent information. Scientometrics 3/126, S. 1867–1896, 2021.

- Li20 Liu, J. et al.: Influence of artificial intelligence on technological innovation: Evidence from the panel data of china's manufacturing sectors. *Technological Forecasting and Social Change* May 2019/158, 2020.
- LZZ20 Li, H.; Zhang, Q.; Zheng, Z.: Research on enterprise radical innovation based on machine learning in big data background. *Journal of Supercomputing* 5/76, S. 3283–3297, 2020.
- MG20 Mühlroth, C.; Grottke, M.: Artificial Intelligence in Innovation: How to Spot Emerging Trends and Technologies. *IEEE Transactions on Engineering Management* 2/69, S. 493–510, 2020.
- MRT18 Mohri, M.; Rostamizadeh, A.; Talwalkar, A.: Foundations of machine learning. MIT press, 2018.
- Mu12 Murphy, K. P.: Machine learning: a probabilistic perspective. *Chance Encounters: Probability in Education*, 2012.
- Mü20 Mühlroth, C.: Artificial Intelligence as Innovation Accelerator: Proceedings of the 2020 on Computers and People Research Conference. Association for Computing Machinery, New York, NY, USA, S. 6–7, 2020.
- MUU21 Münster, S.; Utescher, R.; Ulutas Aydogan, S.: Digital topics on cultural heritage investigated: how can data-driven and data-guided methods support to identify current topics and trends in digital heritage? *Built Heritage* 1/5, 2021.
- NB19 Nazemi, K.; Burkhardt, D.: A Visual Analytics Approach for Analyzing Technological Trends in Technology and Innovation Management. In (Bebis George et al. (eds) Hrsg.): *Advances in Visual Computing. ISVC 2019. Lecture Notes in Computer Science*, vol 11845. Springer, Cham. Springer International Publishing, S. 283–294, 2019.
- NBK21 Nazemi, K.; Burkhardt, D.; Kock, A.: Visual analytics for technology and innovation management. *Multimedia Tools and Applications*, 2021.
- Po90 Porter, M.: Competitive Advantage of Nations. *Competitive Intelligence Review* 1/1, S. 14, 1990.
- PS96 Pleschak, F.; Sabisch, H.: Innovationsmanagement. Stuttgart Schäffer-Poeschel, 1996.
- Ro18 Rossetto, D. E. et al.: Structure and evolution of innovation research in the last 60 years: review and future trends in the field of business through the citations and co-citations analysis. *Scientometrics* 3/115, S. 1329–1363, 2018.
- Ro88 Roberts, E. B.: What We've Learned: Managing Invention and Innovation. *Research-Technology Management* 1/31, S. 11–29, 1988.
- Sh18 Shams, R.: Developing Machine Learning Products Better and Faster at Startups. *IEEE Engineering Management Review* 3/46, S. 36–39, 2018.
- SJ07 Stern, T.; Jaberg, H.: Erfolgreiches Innovationsmanagement: Erfolgsfaktoren - Grundmuster - Fallbeispiele, 2007.
- Sj21 Sjödin, D. et al.: How AI capabilities enable business model innovation: Scaling AI through co-evolutionary processes and feedback loops. *Journal of Business Research* April/134, S. 574–587, 2021.

- Ta20 Tarasova, V. N. et al.: Model of intelligent system for quality evaluation of services on design stage (transport and logistic business as an example). 2019 IEEE International Symposium on Innovation and Entrepreneurship, TEMS-ISIE 2019, S. 294–298, 2020.
- Th92 Thom, N.: Innovationsmanagement. Schweizerische Volksbank, 1992.
- TMT20 Tan, J.; Miao, D.; Tan, Q.: Empirical Study on Influencing Factors of Knowledge Product Remixing in OIC. IEEE Access 8, S. 34215–34224, 2020.
- Tr21 Trocin, C. et al.: How Artificial Intelligence affords digital innovation: A cross-case analysis of Scandinavian companies. Technological Forecasting and Social Change July/173, S. 121081, 2021.
- WW02 Webster, J.; Watson, R. T.: Analyzing the Past To Prepare for the Future Writing a Literature Review. MIS Quarterly 26/2, S. 13–23, 2002.
- Ya19 Yamamura, C. L. et al.: The front-end of product development as systems thinking and predictive learning. Procedia Manufacturing 2019/39, S. 1346–1353, 2019.
- YX20 Yu, S.; Xiao, Q.: Research on the Construction of Strategic Decision-making Platform of Science and Technology Think Tank Based on Multi-agent Collaboration. 2019 IEEE International Symposium on Innovation and Entrepreneurship, TEMS-ISIE 2019, S. 319–324, 2020.
- Zh21 Zhou, Y. et al.: A deep learning framework to early identify emerging technologies in large-scale outlier patents: an empirical study of CNC machine tool. Wireless Personal Communications 2/126, S. 969–994, 2021.
- ZLT21 Zhang, Q.; Liu, S.; Tu, Q.: Knowledge organization of node enterprises' technological innovation under supply chain environment. Complex & Intelligent Systems 0123456789, 2021.
- ZW95 Zahn, E.; Weidler, A.: Integriertes Innovationsmanagement. Handbuch Technologiemanagement, Stuttgart 1995, S. 351, 1995.
- ZYC20 Zhao, N.; Yang, G.; Cao, Y.: Mining Technological Innovation Talents Based on Patent Index using t-SNE Algorithms*: Take the Field of Intelligent Robot as an Example. 2019 IEEE International Symposium on Innovation and Entrepreneurship, TEMS-ISIE 2019, S. 595–601, 2020.
- ZZB18 Zhao, Y.; Zhou, C.; Bellonio, J. K.: New value metrics using unsupervised machine learning, lexical link analysis and game theory for discovering innovation from big data and crowd-sourcing. IC3K 2018 - Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management 2, S. 327–334, 2018.

OpenPredict - An Open Research Dataset and Evaluation Protocol for Fine-grained Predictive Testing

David Brodmann¹, Erik Rodner²

Abstract: Systematic testing of every single component and interface is undoubtedly an important measure to handle the complex nature of current software systems. However, this comes with often neglected computational costs. The aim of this paper is therefore to cut time and resource needs by predictive testing, i.e., predicting test failures with machine learning using a surprisingly simple statistical feature representation. Furthermore, we present the first open research benchmark for predictive testing to enable and foster future research in this area.

Keywords: machine learning; software testing; research dataset; predictive testing

1 Introduction

Software testing is crucial for the success of a professional software project. However, with a growing code complexity, regression testing for changes gets more time and resource consuming resulting in hours of software tests for evaluating code changes for every new commit. A time delay, which is simply impractical to handle during development and a waste of expensive resources. A classical solution to the problem is to track test dependencies and code coverage of each test over time. However, this is time-consuming and sometimes impossible to achieve especially within heterogeneous repositories with several language barriers [MSPC19].

Because predictive testing is heavily project dependent, our contribution to the field of predictive testing is not a new machine learning approach, but rather a curated and open dataset to allow future comparison and evaluation of predictive testing methods.

¹ DResearch Fahrzeugelektronik, Prüflabor / KI, Spenerstr. 38, 10557 Berlin, Germany,
david.brodmann@luckycloud.de

² Hochschule für Technik und Wirtschaft Berlin, Fachbereich 2 / KI-Werkstatt, Wilhelminenhofstr. 75A, 12459 Berlin, Germany, erik.rodner@htw-berlin.de

In summary our main contributions are as follows:

- Repository containing all the databases and code: https://gitlab.com/nexxtnit/predictive_testing
- Ready-to-use database of test results and metadata on a commit basis for four open-source projects (Chap. 2).
- A new baseline method for predictive testing based on simple statistics (Chap. 3).
- Evaluation of a baseline approach (statistical feature extraction) for predictive testing to set the bar for future work (Chap. 5.2)
- In-depth analysis of feature relevance on the new dataset to provide insights concerning major relevant statistics (Chap. 5.3).

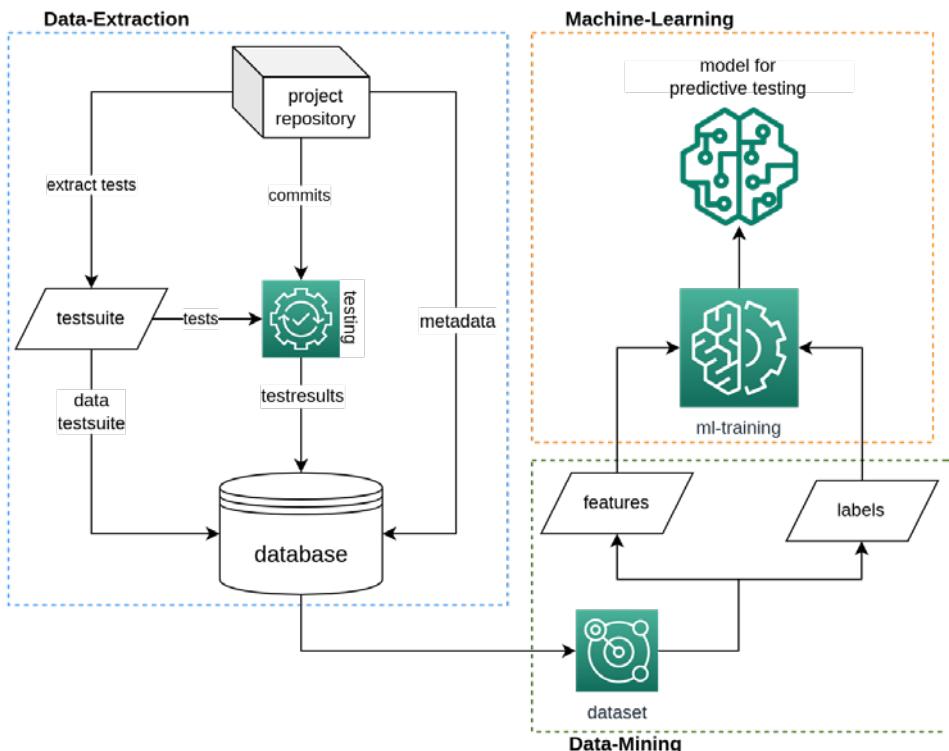


Figure 1 Illustration of predictive testing using supervised machine learning: past code changes and corresponding test results are used for training a machine learning model able to predict test failure probabilities for future code changes.

2 OpenPredict - A New Dataset for Predictive Testing

The problem of current papers and implementations of predictive testing is that the data is often very project specific and not granular enough to make general assumptions. Because predictive testing is often tuned to specific projects, publishing only the machine learning model has no benefit for further research. Furthermore, releasing predictive testing training data of closed-source implementations could reveal insights about the software product. The data used from continuous integration pipelines is often taken as is, there is no possibility in tuning the scope of tests or choose the granularity of test execution.

In contrast, the goal of this work is to create an open research dataset for experiments in the field of predictive testing that can be systematically compared also in future work. Our dataset tries to illustrate the minimal requirements of data that needs to be available to implement predictive testing for your own project. The dataset itself was generated from open-source projects and could be theoretically extended further with the data extraction pipeline we build.

Our data is comprised of four curated open-source repositories of different nature. To ease evaluation and allow stable statistical evaluation, we used the following criteria to select repositories:

- Written in Python to allow for applying language-specific features in the future
- Minimum of 400 commits to get enough data points over time
- Minimum of 10 tests to allow for multi-task prediction
- Testing happens with pytest or tox to ease the automatic data extraction

These requirements, yielded in four repositories outlined in Table 1. In the following, we explain the process of data extraction and dataset curation in detail.

Project	Description	Commits	Testing	Link
Flit	” Simplified packaging of Python modules” [Pypa21]	1017	Tox with Py-test	Github
Mock	” The Python mock library” [Test21]	1277	Pytest	Github
Flake8	”flake8 is a python tool that glues together pycodestyle, pyflakes, mccabe, and thirdparty plugins to check the style and quality of some python code.” [Pycq21]	2074	Tox with Py-test	Github
Nox	” Flexible test automation for Python” [Thea22]	439	Pytest	Github

Table 1 Repositories used in our dataset for predictive testing.

Data Extraction for Test Change Representation: Our method for data extraction is illustrated in Figure 2. The approach is identical for all four projects being tested in this work. All information is gained from the project repository in an automatic fashion. The test suite is extracted from the existing tests and test variations of the repository. We do not generate synthetic code defects as in [Lund19], to obtain more test results. The commit hash provides us with a unique identifier for the test result as well as the related code changes.

In contrast to [PaPr21], our predictive testing paradigm is to forecast the test result of every commit. *Every commit in the repository gets tested by every test in the test suite. As a result of that approach, we gain precise data on code changes and test results.*

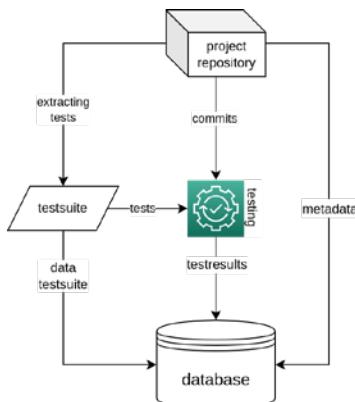


Figure 2 Data extraction method, every commit is tested with all test variations. The data gathered from the test run and the repository is saved in an SQL database keeping the correspondence between them.

Database: All meta information and main statistical features are stored in an SQL database, which is outlined in Figure 3 and identical for all projects. The scripts used for automatically transforming a git repository to the database are available in our associated code repository.

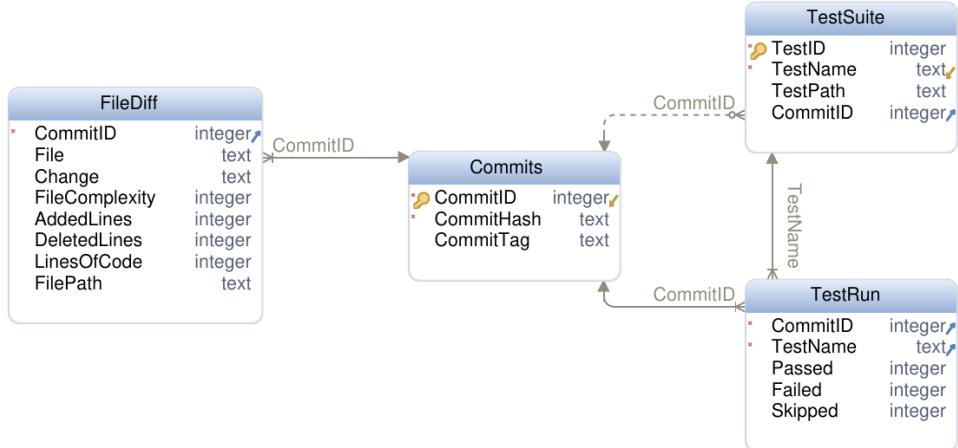


Figure 3 Database architecture, the commit-id, based on the commit hash, is used for synchronization throughout the database

Test Suite: The goal of our predictive testing approach is to predict the test result of all tests in the test suite. However, tests can change over time, and these variations require special care. We need to add each test variation of a test as an additional target to prevent the model from learning wrong statistical dependencies of older code changes and test results not related to the modified tests. We therefore include all changes of a test and create a separate test target for each change (see Table 2). All tasks for a fixed will be not statistically independent, since test changes will consist of tiny changes over time. This fact can be exploited by a multi-task approach for learning.

With including every change in a test case as an own test variation, we increased the amount of test cases from 12 tests to 329 - phrasing each of these variations as independent prediction tasks.

Tests Newest Commit	Test Variations in Test Suite
test_parametrize.py	test_parametrize_354.py
	test_parametrize_359.py
	test_parametrize_367.py
test_version.py	test_version_348.py
	test_version_367.py
test_command.py	test_command_4.py
	test_command_4.py
	test_command_9.py
	test_command_24.py
...	...

Table 2 Left: tests from the newest commit of Flit; Right: test variations

For the flit project the dataset statistics can be checked in Table 3.

Version (Commits)	All (Passed/Failed)	Training Set (P/F)	Testing Set (P/F)
Flit.V0.0 (634)	208'586 (33'391/175'195)	183'253 (28'472/154'781)	25'333 (4919/20'414)
Flit.V1.0.WV (538)	177'002 (28'764/148'238)	157'920 (28'764/143'611)	19'082 (3812/15'270)
Flit.V1.0 (538)	10'760 (7591/3169)	9600 (6830/2770)	1160 (761/399)

Table 3 Dataset statistics of the Flit dataset variations. The newest 10% of the commits are for the testing set, the other 90% go in the training set, see Chapter 5.1 for details.

3 Predictive Testing with Statistical Features

In the following, we outline our baseline machine learning approach for predictive testing, which will be evaluated in Sect. 5.

A dataset consists of three parts, the one-hot encoded test results, the one-hot encoded file changes, and metadata per commit. As label data the pass/fail column is used. A snippet of the complete dataset can be examined in Table 4.

CommitID	test_build_422.py	test_build_465.py	...	pyproject.toml
1028	1	0	...	1
1028	0	1	...	1
...

FileComplexity	AddedLines	DeletedLines	PassFail
27	5	5	0
27	5	5	0
...

Table 4 Complete dataset from the Flit database for machine learning. All file changes and tests are one-hot encoded. The PassFail column is the label we try to predict, and the CommitID will be removed for training and testing.

As a classification model we use a random forest classifier to predict the outcome of a test-case as pass or fail. The random forest classifier consists of several decision trees trained with a random fraction of the training data. A decision tree is made up of inner nodes (split nodes) and leaves. At each split node a single feature is evaluated since we make use of simple decision stumps. After traversing a tree, the final estimate is given by an empirical probability stored at each leaf during learning. A random forest decision is determined by the average of all trees estimates and a majority vote is used for the final classification decision.

Due to the simple decision stumps used, a random forest classifier can easily cope with features of different magnitudes. Furthermore, we also exploit multi-task learning of all different target tasks (every test variation is a single target) since the tasks are highly related. This is done in our case by a one-hot-encoding of the task itself and a single binary variable as a target. Some branches in a single decision tree can therefore focus on an arbitrary set of tasks by using their one-hot encodings as features in early split nodes. This is beneficial compared to learning each predictive testing task independently from each other with only a few training examples available (each test variation might only relate to a few commits).

For our experiments, we use a random forest with 1000 decision trees and a Gini split criterion without a restriction of a maximum number of examples in a leaf.

4 Related Work

Reducing the number of executed tests in regression test suites is an established method for reducing costs and time [YoHa12]. Test selection is a very efficient method for reducing the executed tests [LHSL16]. Combining test selection with machine learning is the usually referred to as predictive Testing. Despite its potential, relevance for sustainable development, and partial adoption at large tech companies [RiMS21], predictive testing is a rather unexplored area of research. In [MSPC19], Machalica et al. also used historical data to implement a data-driven test selection strategy. This allowed the authors to reduce the total number of test executions by 2/3 while still finding 95% of all failed tests. Unfortunately, no open dataset was provided, rendering reproduction of the results impossible. [Lund19] implemented a predictive test selection tool by creating synthetic test results. They altered the code with small synthetic modifications to obtain labeled data with a decent number of test failures. Classifying tests results was done using a random forest classifier, which allows for reducing test executions by 50% at again 95% true positive detection rate for all failures. A severe disadvantage of this approach is that its performance highly depends on the realistic nature of the synthetic code modifications. In complex software systems, these are difficult to design and might be in addition subject to change over time. In contrast, the approach of [SKPS20] uses simple text similarity metrics to rank tests related to a code change. In comparison to the impressive dataset gathered in [YBKB22] such as Test Case Prioritization (TCP from 25 open-source projects, we concentrate on the basic needs to implement predictive testing. In their work, Java projects are exclusively used along with their continuous integration history. As a result, many additional data features like code coverage are available. Furthermore, using only continuous integration data has the disadvantage that the granularity of code changes cannot be chosen by oneself. The code changes do differ from build to build and can range from one to several commits. In our work, we concentrated on changes and test results per single commit - providing predictive testing at a fine-grained level. However, the three “high level features” having the most impact on prediction in [YBKB22] such as Test Case Prioritization (TCPare like the ones used in our work. The work of Sharif et al. [ShML21] focuses on deep learning techniques for regression test prediction. They show the benefits of their approach especially for large datasets.

A related problem to ours is test suite failure prediction [PaPr21], which tries to predict the failure of the whole test suite rather than for each test case individually. A further overview of predictive methods for software engineering is given in [YXLB22].

5 Experiments

In the following, we evaluate our baseline approach on OpenPredict showing the power of simple statistical features for predictive testing.

5.1 Experimental Setup

For experimenting with machine learning methods for test case prediction, the dataset needs to be separated first in training and testing data. Instead of selecting random cases from the dataset, we decided to split the data by its commit-id, i.e., trying to predict the behaviour of newer commit from the history of older ones. This reflects the use case for predictive testing, where the model is trained continuously and applied to the most recent commit. While analyzing the databases, it became obvious that some commits, did not have any test results. This is because early commits in the repository have often not been able to execute tests at all, therefore they contain no test results.

The dataset is split into testing and training data by the CommitID. From the sum of commits containing test data, the newest 10% are used for testing, the other 90% for training. Please note that the commit-id is of course not used as a feature.

Since our prediction tasks are binary, we used an ROC analysis as the main performance metric for our experiments. Instead of calculating ROC results for each task, we evaluated the performance of the random forest classifier, which provides us with an aggregated performance metric over all test cases.

Dataset	Specifications	Features
Flit.V0.0	no metadata all commits, starting at init all test cases, no filtering of variations	test changed files in commit test result
Flit.V1.0.WV	metadata <i>FileComplexity</i> and <i>AddedLines</i> starting at commit tag <i>V1.0</i> all test cases, no filtering of variations	test changed files in commit test result <i>FileComplexity</i> <i>AddedLines</i>
Flit.V1.0	Metadata <i>FileComplexity</i> and <i>AddedLines</i> added Starting at commit tag V1.0 Combining all test variations to one test case	test changed files in commit test result <i>FileComplexity</i> <i>AddedLines</i>

Table 5 Dataset variations of the Flit database: Flit.V0.0 has all raw data and no metadata, Flit.V1.0.WV has metadata added, Flit.V1.0 contains structured test results and metadata.

5.2 Analysis on the flit dataset

First, we evaluate some dataset aspect on the Flit dataset with our baseline approach. In particular, we want to evaluate the impact of skipping the very first commits and dealing with test case variations as individual tasks (see Figure 5 for an illustration). Our experiments are performed with three different versions of the Flit dataset in Table 5. The results are given in Figure 4.

The AUC improves about 6% when adding more relevant data to the dataset and about 8% when structuring the test results of the testcases according to their commit-id.

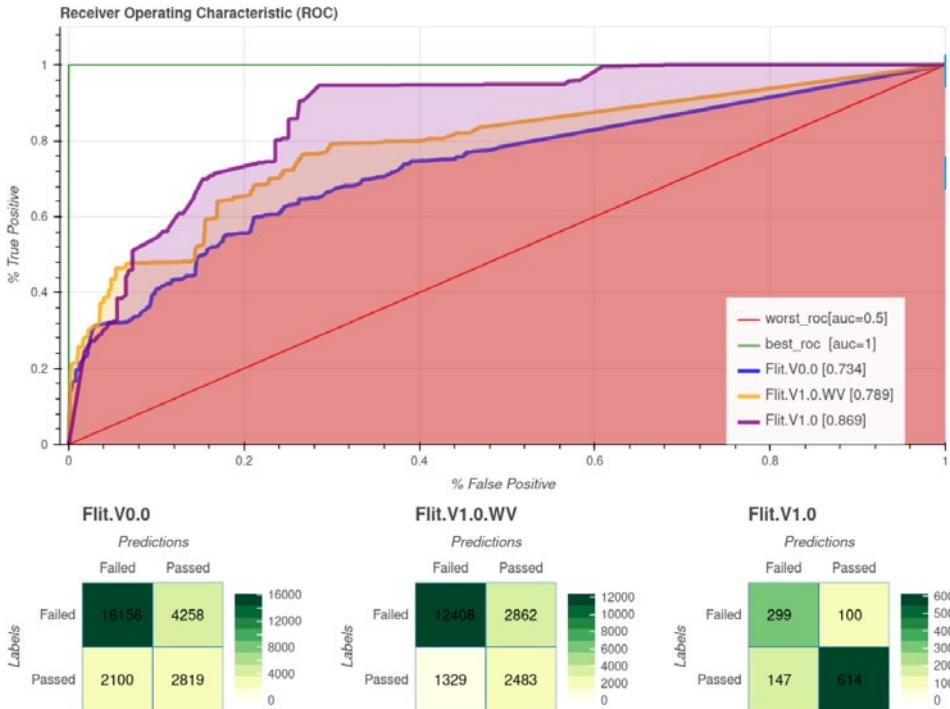


Figure 4 Different classification for the three dataset variations of Flit from Tab. 5

For improving classification an attempt is made to restructure the data of all test case variations. Because all test case variations test the same code functionality, logically it makes sense to only use the result of one test variation at a time. At every commit the result of the matching test variation must be used.

As a result of the previous observations, a test variation is used as long as the commit-id of the current commit to test, is one greater than the commit-id of the test variation (Figure 5).

The diagram illustrates the restructuring of test variations. On the left, a 'TestSuite' table lists multiple test variations across different commit IDs (991, 990, 989). Each variation includes the commit ID, test name, and statistics for Passed, Failed, and Skipped tests. Colored boxes highlight specific rows: a green box for commit 991, a blue box for commit 990, and a red box for commit 989. Arrows point from these highlighted rows to a second table on the right. This second table, titled 'PassFail', contains only the commit ID, test name, and a single column for 'PassFail' (either 1 or 0). The commit IDs in this table correspond to the ones in the original TestSuite table, with the same color coding (green for 991, blue for 990, red for 989).

CommitID	TestName	Passed	Failed	Skipped
991	test_wheel_990.py	18	0	0
	991 > 990 989.py	17	0	0
	991 test_wheel_988.py	16	1	0
	990 test_wheel_990.py	16	2	0
	990 > 989 988.py	16	1	0
	989 test_wheel_990.py	16	2	0
	989 test_wheel_988.py	17	0	0
989	989 > 988 ...	16	1	0

CommitID	test_wheel.py	PassFail
991	1	1
...
990	1	1
...
989	1	0

Figure 5 Test variations being restructured for a dataset. The Results of every test variation is used if the commit currently testing is greater than the commit the test variation is created.

5.3 In-depth analysis of feature relevance

The metadata features we identified in Sect. 3 are contributing strongly to the classification. The importance is visible in Table 6. The feature importance is measured as the impurity decrease within each tree. The absolute value of the accumulated impurity decrease is the importance of each feature. Besides the obvious test cases as features, the metadata features are the ones with the biggest impact.

Nr.	Importance	Feature
1	0.137	test_importable_.py
2	0.119	test_metadata_.py
3	0.098	FileComplexity
4	0.055	test_config_.py
5	0.052	AddedLines
6	0.049	DeletedLines

Table 6 Feature importance of Flit.V1.0 dataset. The Metadata features FileComplexity, AddedLines and DeletedLines are within the 6 most important features.

When analyzing the classification results, the question emerged what probably would happen, if the commit-id is kept as a feature in the dataset. The results can be seen in Figure 6.

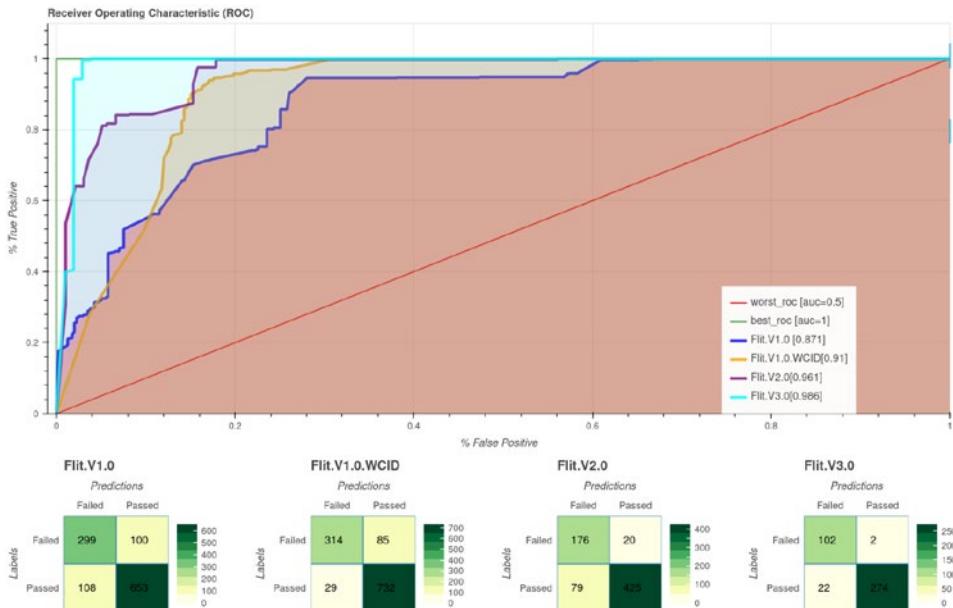


Figure 6 Importance of commits included in dataset

Flit.V1.0: The blue line in the roc-analysis (Figure 6) and the classification matrix on the left border, is the best performing dataset from Chap. 5.1 (Flit.V1.0).

Flit.V1.0.WCID: The orange line and second classification matrix from the left (Figure 6), is the same dataset as Flit.V1.0 but with the feature commit-id added.

The AUC of the random forest classifier is improving by around 5% by adding the commit-id as a feature to the dataset.

The reason for the performance improvement can be found in a deeper analysis of the nodes of the feature commit-id in the balanced random forest classifier, visible in Figure 7.

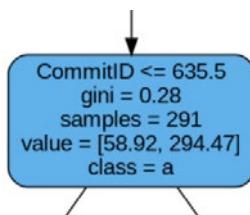


Figure 7 Node for feature commit-id in tree of balanced random forest classifier

The whole dataset contains around 1000 commits. Starting with commit-id 1 for the oldest commit, the newest commit has the id 1000. The node in Figure 7 is separating the commits for the commit-id value $\leq 635,5$. Because our test set contains only the newest 10% of the commits, the commit-id in the test set will always be greater than 635,5. This observation leads to the idea that fewer commits in the dataset could actually benefit the classifier.

Flit.V2.0: The purple line in the ROC analysis and the classification matrix second from the right border, is the same dataset as the first one. The commit-id is not included as a feature, but commits were only used starting with commit-tag V2.0.

Flit.V3.0: The cyan line in the ROC analysis and the classification matrix first from the right, is the same dataset as the first one. The commit-id is not included as a feature, but commits were only used starting with commit-tag V3.0.

In conclusion, it becomes clear that a logical connection of the code, represented with version tags, is important to train a machine learning model for predictive testing. This leads to the theory that not the whole history of a repository is necessary for predictive testing, but rather the last one or two versions of the project code.

Version (Commits)	All (Passed/Failed)	Training Set (P/F)	Testing Set (P/F)
Flit.V1.0 (538)	10'760 (7591/3169)	9600 (6830/2770)	1160 (761/399)
Flit.V1.0.WCID (538)	10'760 (7591/3169)	9600 (6830/2770)	1160 (761/399)
Flit.V2.0 (347)	6940 (4948/1992)	6240 (4444/1796)	700 (504/196)
Flit.V3.0 (193)	3860 (2678/1182)	3460 (2382/1078)	400 (296/104)

Table 7 Statistical features of the datasets used to analyze the commit-id feature

5.4 Full evaluation on OpenPredict

In Table 8 and Figure 8 the size of all datasets and the performance of the classifiers for the different projects can be examined. All four projects we tested in our work provided an acceptable classification. The visible differences in Figure 8 are explainable because of the code complexity, size, and test amount of the different projects. The worst classification project, *Flake8*, also offers space to further improve classification performance by increasing the commit tag.

Training a predictive testing model for different repositories shows that our approach is not project specific and offers the possibility to implement predictive testing in a general manner, as long as the correct data is provided.

Version (Commits)	All (Passed/Failed)	Training Set (P/F)	Testing Set (P/F)
Flit.V2.0 (347)	6940 (4948/1992)	6240 (4444/1796)	700 (504/196)
Mock.V0.8 (464)	4640 (2636/2004)	4170 (2254/1916)	470 (382/88)
Nox.V0.0 (436)	6976 (2964/4012)	6272 (2486/3786)	704 (478/226)
Flake8.V3.7 (613)	27'585 (10'178/17'407)	24'795 (8564/16'231)	2790 (1614/1176)

Table 8 Statistical features of the datasets of all different projects

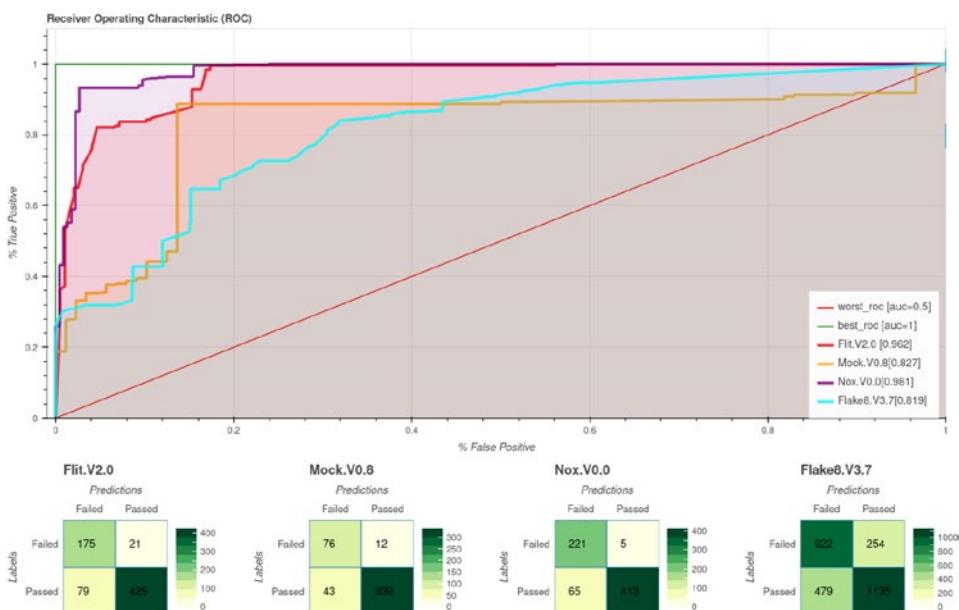


Figure 8 Overview of the best possible classification with optimizations discovered in this paper.

Conclusions and Future Work

We presented the first open research dataset for predictive testing on a fine-granular basis. With this minimalistic approach we identified the core features necessary to predict test results. With a detailed analysis of feature importance, it was even possible to conclude that not the whole code history is necessary for gathering training data. Furthermore, we developed a baseline approach based on statistical features from the repository and a multi-task random forest as predictor. Applied to the dataset, the approach showed surprisingly high prediction performance, even though no detailed code analysis was used to enrich the feature representation of the related code changes.

There is a multitude of research ideas to boost the performance of predictive testing, including using recent language models trained on source code [CTJY21] or even directly on code changes. Although we use a simple multi-task technique (using one-hot-encoded task descriptions) for prediction, this representation lacks the information that variations of the same underlying unit test are related differently based on their commit history. In addition, the baseline approach and related ones should be further evaluated in a continuous learning setting, where the model is learned from last K commits and applied to the most recent one.

Bibliography

- [CTJY21] Chen, Mark ; Tworek, Jerry ; Jun, Heewoo ; Yuan, Qiming ; Pinto, Henrique Ponde de Oliveira ; Kaplan, Jared ; Edwards, Harri ; Burda, Yuri ; u. a.: Evaluating large language models trained on code. In: *arXiv preprint arXiv:2107.03374* (2021)
- [LHSL16] Legusen, Owolabi ; Hariri, Farah ; Shi, August ; Lu, Yafeng ; Zhang, Lingming ; Marinov, Darko: An extensive study of static regression test selection in modern software evolution: 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering, FSE 2016. In: Su, Z. ; Zimmermann, T. ; Cleland-Huang, J. (Hrsg.) *FSE 2016 - Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering, Proceedings of the ACM SIGSOFT Symposium on the Foundations of Software Engineering.*, Association for Computing Machinery (2016), S. 583–594
- [Lund19] Lundsten, Erik: *EALRTS : A predictive regression test selection tool*, KTH, School of Electrical Engineering and Computer Science (EECS), Master's Thesis, 2019. — Backup Publisher: KTH, School of Electrical Engineering and Computer Science (EECS)
- [MSPC19] Machalica, Mateusz ; Samykin, Alex ; Porth, Meredith ; Chandra, Satish: Predictive Test Selection. In: *Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Practice, ICSE-SEIP '19* : IEEE Press, 2019, S. 91–100
- [PaPr21] Pan, Cong ; Pradel, Michael: Continuous test suite failure prediction. In: *Proceedings of the 30th ACM SIGSOFT International Symposium on Software Testing and Analysis*, 2021, S. 553–565
- [Pycq21] Pycqa: flake8. *GitHub*.
- [Pypa21] pypa: flit. *GitHub*.
- [RiMS21] Ricca, Filippo ; Marchetto, Alessandro ; Stocco, Andrea: AI-based Test Automation: A Grey Literature Analysis. In: *2021 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW)* : IEEE, 2021, S. 263–270
- [ShML21] Sharif, Aizaz ; Marijan, Dusica ; Liaen, Marius: DeepOrder: Deep Learning for Test Case Prioritization in Continuous Integration Testing. In: *2021 IEEE International Conference on Software Maintenance and Evolution (ICSME)* : IEEE, 2021, S. 525–534
- [SKPS20] Sutar, Shantanu ; Kumar, Rajesh ; Pai, Sriram ; Shwetha, BR: Regression test cases selection using natural language processing. In: *2020 International Conference on Intelligent Engineering and Management (ICIEM)* : IEEE, 2020, S. 301–305
- [Test21] testing-cabal: mock. *GitHub*.
- [Thea22] theacodes: nox. *GitHub*.
- [YBKB22] Yaraghi, Ahmadreza Saboor ; Bagherzadeh, Mojtaba ; Kahani, Nafiseh ; Briand, Lionel: Scalable and Accurate Test Case Prioritization in Continuous Integration Contexts. In: *IEEE Transactions on Software Engineering* (2022), S. 1–24
- [YoHa12] Yoo, S. ; Harman, M.: Regression testing minimization, selection and prioritization: a survey. In: *Software Testing, Verification and Reliability* Bd. 22 (2012), Nr. 2, S. 67–120
- [YXLB22] Yang, Yanming ; Xia, Xin ; Lo, David ; Bi, Tingting ; Grundy, John ; Yang, Xiaohu: Predictive Models in Software Engineering: Challenges and Opportunities. In: *ACM Transactions on Software Engineering and Methodology* Bd. 31 (2022), Nr. 3, S. 56:1–56:72

Prototypen

BPMN-CREATOR – Ein innovatives Tool zur vollautomatischen Digitalisierung zuvor manuell erstellter Geschäftsprozessmodelle

Selcan Ipek-Ugay¹, Tabea Herrmann² und Eric Siegeris³

Abstract: Modellierung ist ein wichtiger Bestandteil der Anforderungserhebung und passiert häufig in interdisziplinären Teams an der Schnittstelle zwischen Fachlichkeit und IT. Dabei werden die Modelle typischerweise zunächst gemeinsam und analog (z.B. an Whiteboards) erfasst und erst dann in ein digitales Tool überführt, damit sie weiterverwendet werden können. Die manuelle Umwandlung ist mit erheblichem Zeit- und Arbeitsaufwand verbunden und führt darüber hinaus zu unerwünschten Reibungen im Modellierungsworkflow. In diesem Beitrag wird eine neue Anwendung *BPMN-Creator* vorgestellt, die zuvor manuell erstellte Geschäftsprozessmodelle vollautomatisch erkennen und in digital weiterverwendbare BPMN-Modelle transformieren kann. Mit nur drei Klicks kann die Anwendung digitale BPMN-Diagramme aus gemeinsam skizzierten Prozessmodellen erzeugen: Fotografieren der Whiteboard-Skizze, Zuschneiden des Bildes und Generieren des Modells. Danach kann das Modell in einem beliebigen BPMN-Tool weiterverwendet werden. Die entwickelte Technik wurde mit einem Datensatz von 70 manuell erstellten Prozessmodellen evaluiert. Es hat sich gezeigt, dass unsere Technik die Lücke von der fertigen Skizze zum digital weiterverwendbaren Modell schließt und somit eine wertvolle Grundlage zur Optimierung des Modellierungsprozesses liefert.

Keywords: Collaborative Modeling, Business Process Modeling, Handcrafted Models, Digital Image Processing

1 Einleitung

In den letzten Jahren hat die Dokumentation von Geschäftsprozessen für die Analyse von prozessorientierten Informationssystemen sowie die Erfassung von Anforderungen in Softwareentwicklungsprojekten erheblich an Aufmerksamkeit gewonnen [We19, BR14]. Diese Tendenz hat dazu geführt, dass diverse digitale Modellierungstools entwickelt wurden, um Modellierer:innen bei der Erstellung und Pflege von Geschäftsprozessmodellen zu unterstützen. Obwohl diese Werkzeuge im Allgemeinen als unverzichtbar angesehen werden können, ist es wichtig zu beachten, dass die Entwicklung der meisten Prozessmodelle nicht

¹ Vostura GmbH, Residenzstrasse 25, 13409 Berlin selcan.ipek-ugay@vostura.com,
<https://orcid.org/0000-0002-4132-1199>

² Vostura GmbH, Residenzstrasse 25, 13409 Berlin tabea.herrmann@vostura.com

³ Vostura GmbH, Residenzstrasse 25, 13409 Berlin eric.siegeris@vostura.com

mit Software beginnt. In kollaborativen Umgebungen werden Prozessmodelle typischerweise auf Whiteboards oder Papier gezeichnet, was die Beteiligung der Prozessbeteiligten fördert [Lü11, Du05, FV05, CVDK07]. Am Ende eines solchen Vorgehens liegen die Ergebnisse als handgefertigte Modelle vor und müssen mit einem Modellierungstool in ein digitales Gegenstück umgewandelt werden. Dieser Transformationsschritt wird derzeit manuell realisiert und ist mit einem erheblichen Zeit- und Arbeitsaufwand verbunden. Darüber hinaus führt es zu Unterbrechungen sowie Reibungen im gesamten Modellierungsprozess. Um dieses Transformationsproblem zu umgehen und die kollaborative Modellierung [RKV08] zu unterstützen wurden bereits einige Methoden erforscht ME22, Wi12, BRW11]. Existierende Werkzeuge, die den interaktiven Modellierungsprozess unterstützen sind einerseits mit erheblichen Anschaffungskosten verbunden und andererseits verlangen sie von den Benutzer:innen sich an vordefinierte Konstrukte und Funktionen zu halten. Dadurch werden die, mit der kollaborativen Modellentwicklung verbundenen Vorteile limitiert, sodass diese Tools nur bedingt eine geeignete Alternative zum Whiteboard darstellen. Ferner konnte festgestellt werden, dass Gesprächsphasen von Beteiligten bei der Modellierung von BPMN-Prozessen am Whiteboard größer sind als bei der Nutzung von Notebook und Beamer oder eines Multi-Touch-Geräts [KKRR15]. Mit der Entwicklung des BPMN-Creators wurde die wissenschaftliche Fragestellung geprüft, inwieweit sich unter Verwendung eines begrenzten Satzes von BPMN-Elementen analoge Skizzen automatisiert in digitale BPMN-Modelle transformieren lassen. In Anbetracht der Bedeutung und der Vorteile von gemeinsam handgefertigten Prozessmodellen sowie des Aufwands, der mit ihrer manuellen Umwandlung verbunden ist, schlagen wir in diesem Beitrag diesen Ansatz vor, der diesen Schritt automatisiert, indem kollaborativ gefertigte Prozessmodelle in digitale BPMN-Modelle umgewandelt werden. Unser Ansatz, BPMN-Creator, nutzt innovative Bildverarbeitungs- und Objekterkennungsalgorithmen, um wesentliche BPMN-Elemente zu erkennen und erweitert den Stand der Technik [SALS21] im Bereich der Modelltransformation, indem ein am Whiteboard gefertigtes BPMN-Modell abfotografiert und daraus automatisch eine digital weiterverarbeitbare BPMN-XML Datei erzeugt wird (vgl. Abb.1). Unsere Evaluationen zeigen die hohe Genauigkeit, die unser Ansatz erreicht. Diese Präzision wird einerseits ohne komplexe sowie ressourcenaufwändige Deep-Learning basierende Ansätze erreicht (vgl. [SALS21]) und liefert sehr zuverlässige Ergebnisse. Andererseits können die Skizzen erstmals in Echtzeit und benutzerunabhängig digitalisiert werden, indem unser Tool unmittelbar am Ende der kollaborativen Modellierung eingesetzt wird. Das Ergebnis kann direkt digital weiterverarbeitet werden.

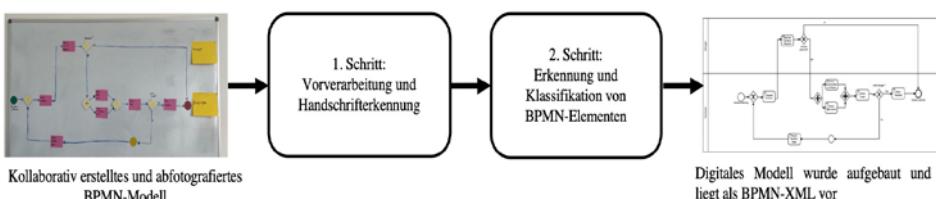


Abb. 1: BPMN-Creator Übersicht

2 Materialien und Methoden

Um die Bedeutung unserer Arbeit zu verdeutlichen, werden im Folgenden zuerst die Herausforderungen eingeführt, die **(a)** mit der Optimierung des gemeinsamen Modellierungsprozesses in Teams und **(b)** mit der einfachen und automatischen Digitalisierung des Prozessmodells verbunden sind. Dann wird unser innovativer Ansatz zur Lösung dieser Probleme eingeführt.

2.1 Problembeschreibung

(a) Modellierungsworkflow innerhalb von interdisziplinären Teams: Untersuchungen haben bestätigt, dass Geschäftsprozessmodellierungen mit vereinfachter Notation an Whiteboards zu durchdachteren Modellen führen und den Modellierenden mehr Spaß bereiten [Lü11]. Zudem werden Maßnahmen zur Involvierung aller Beteiligten sowie Beschleunigung der gesamten Modellierungsphase begrüßt. In der Regel ist es dann jedoch sehr mühsam am Whiteboard gebaute Prozesse zu digitalisieren. Zusätzlich zum unnötigen Zeit- und Ressourcenaufwand können in diesem manuellen Digitalisierungsschritt neue Prozessfehler entstehen. Es muss also eine Technologie entwickelt werden, welche sowohl die kreative gemeinsame Modellierung fördert als auch die Dokumentation automatisiert.

(b) Digitalisierung manuell ersteller BPMN-Modelle: Abbildung 2 zeigt ein reales Prozessmodell aus unserem Testdatensatz, das sich mit einem Prozess zur Bearbeitung von Urlaubsplanung befasst. Hierbei wurde Testpersonen folgendes Setup zum Modellieren bereitgestellt: Whiteboard, Whiteboard-Marker, Post-its und optionale magnetische Elemente (Kreise und Gateways) zur Beschleunigung des Modellierungsprozesses.

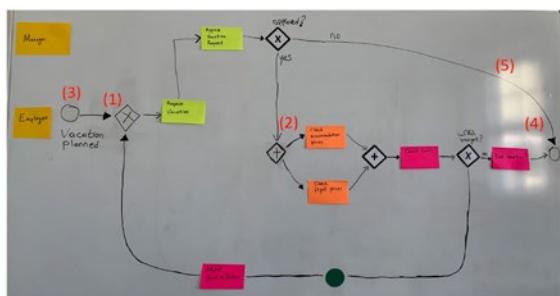


Abb. 2: Beispielhaftes BPMN-Modell aus dem Testdatensatz

Obwohl das dargestellte Modell syntaktisch korrekt ist, veranschaulicht die Abbildung Herausforderungen, die von einem automatischen Erkennungsansatz bewältigt werden müssen. Problem (1) zeigt, dass der Sequenzfluss nicht richtig mit seiner Ausgangs- und

Zielform verbunden ist. Problem (2) verdeutlicht, dass Sequenzflussspitzen unterschiedlich dargestellt werden können. Ein weiteres Problem bezieht sich auf Modellelemente, die unvollständig gezeichnet sein können, wie z. B. das Start event (3). Zudem fällt auch die Verwendung von gebogenen und geraden Linien (4) auf. Eine große Herausforderung stellt auch die Belichtung des Whiteboards dar, bei der unerwünschte Spiegelungen (5) aufgetreten sind. Unabhängig vom gemeinsam manuell erstellten Prozessmodell, spielen folgende Faktoren eine wichtige Rolle: Aufnahmequalität, Perspektive, Belichtungen (natürliche, schlechte und/oder teilweise Belichtung) sowie Whiteboardzustand (Edding-spuren, Elemente, die nicht zum BPMN gehören, Ränder usw.). Obwohl sich die obigen Ausführungen nur auf ein einziges Beispiel beziehen, ist das Spektrum handgefertigter BPMN-Modelle aufgrund der Kombination verschiedener Zeichenstile, Stifte und Digitalisierungsmethoden noch komplexer.

2.2 Problembewältigung

Im Folgenden wird unser Vorgehen zur Bewältigung der oben beschriebenen Probleme erläutert. Zur Überprüfung der Machbarkeit unseres Ansatzes haben wir uns zunächst auf die wesentlichen BPMN-Elemente fokussiert: Lanes, Aktivitäten, Start events, End events, Intermediate events, XOR-Gateways, AND-Gateways, Sequenzflüsse.

(a) Förderung des Modellierungsworkflows innerhalb von interdisziplinären Teams: Soll ein Geschäftsprozess im kollaborativen Rahmen skizziert werden, bietet sich unter anderem die Nutzung eines Whiteboards an. Ein Whiteboard ist in fast jedem Büro vorhanden, sodass Modellierungsideen schnell und einfach umgesetzt sowie diskutiert werden können. Darüber hinaus empfehlen wir den Einsatz von Post-its zur Abbildung von Aktivitäten und Lanes. Somit haben alle Beteiligten die Möglichkeit Ihre Ideen festzuhalten und dem Team mitzuteilen. Zudem werden durch Post-its (im Gegensatz zu Texten, die direkt am Whiteboard geschrieben werden) Verschmierungen reduziert. Um den gesamten Modellierungsprozesses zu beschleunigen sowie den Spaßfaktor zu erhöhen, können für weitere essentielle BPMN-Elemente, die intern nicht beschriftet werden müssen, wie z.B. Events und Gateways magnetische Formen eingesetzt werden. Zur Darstellung der Sequenzflüsse sowie weiterer Zeichnungen können Marker eingesetzt werden. Die haptische Wahrnehmung der BPMN-Elemente macht die digitalen Inhalte spürbar und fördert die Kreativität während der Modellierungsphase, weil die Post-its und Magnete ganz einfach und beliebig am Whiteboard platziert werden können. Unser vorgeschlagenes *Setup bestehend aus Whiteboard, Marker und Post-its* ist in sehr vielen Büroräumlichkeiten vorhanden. Optionale haptische Magnelemente (Kreis- und Gateway-Formen) unterstützen den reibungslosen Modellierungsprozess.

(b) Automatische Digitalisierung manuell erster BMPN-Modelle: Um eine robuste, einfache und vollautomatische Transformation der abfotografierten BPMN-Modelle in digitale BPMN-XML zu ermöglichen, wurde folgende mehrstufige und aufeinander aufbauende Verarbeitungspipeline entwickelt.

1. Schritt: (Vor-)verarbeitung und Handschrifterkennung: Im obigen Abschnitt wurden relevante Aspekte erwähnt, die einen unmittelbaren Einfluss auf die Objekterkennung haben, nämlich Aufnahmegerätqualität (Auflösung, unterschiedliche Aufnahmegeräte), Belichtung (natürliche, schlechte und/oder teilweise Belichtung), die daraus folgende Spiegelung und der Zustand des Whiteboard (ungleichmäßig schmutziger Hintergrund). Um diese Einflussfaktoren zu eliminieren und das abfotografierte Input-Bild I_{in} zu normieren wird zuerst die Größe des Bildes angepasst und dann eine automatische und aufnahmespezifische Helligkeitsoptimierung mittels Gammakorrektur -eine Potenzfunktion mit einem Exponenten Gamma (γ) als einzigem Parameter- durchgeführt:

$$I_{ou} t = I_{in} \gamma, \text{ wobei } \gamma = \log(\text{mean}(I_{in})) / \log(255/2).$$

Im Anschluss werden die im Bild des Prozessmodells I_{out} enthaltenen Zeichen und Buchstaben ausgelesen. Es ist eine große Herausforderung, eine funktionierende Handschrifterkennung zu entwickeln, da bereits für das Training einer solchen Erkennung immense Datenmenge notwendig sind. Eine Eigenentwicklung für ein solches Vorhaben wurde daher zunächst durch die optische Zeichenerkennung der Google Cloud Vision API ersetzt. Diese liefert zuverlässige Ergebnisse und erkennt Wörter häufig fehlerfrei. Danach wurde das Grauwert-trasformierte I_{out} mit morphologischen Operatoren sowie Medianfilterung gefiltert, um Rauschanteile und unerwünschte Strukturen zu eliminieren. Abbildung 3 zeigt die Resultate der bisherigen Schritte bei dem oben eingeführten Modell.

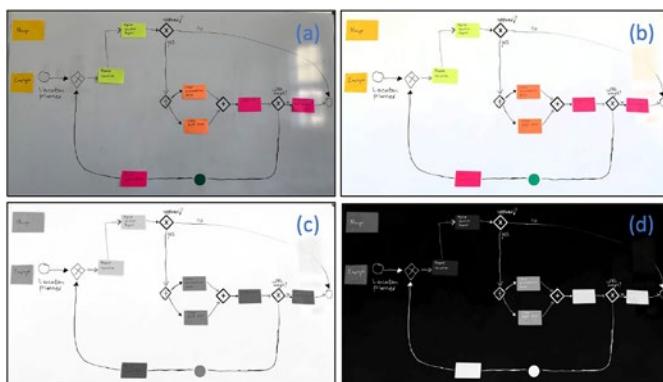


Abb. 3: Originalaufnahme der am Whiteboard erstellten BPMN-Skizze (a), Ergebnis nach der Helligkeitskorrektur mittels Potenzfunktion (b), Grauwerttransformation (c) und anschließende Filterung des Bildes (d).

2. Schritt: Erkennung und Klassifikation von BPMN-Elementen:

Im Folgenden unser Vorschlag zur Erkennung und anschließender Klassifikation der wesentlichen BPMN-Elemente.

Zur Extraktion von **elliptischen Formen** haben wir das globale Verfahren *Hough-Transformation* für Kreise verwendet [IK87]. Die detektierten Kreise werden aus dem Inputbild $I_{\text{preprocessed}}$ entfernt und das Ergebnisbild (=img_deletedCircles) so gefiltert, dass alle Stör-elemente eliminiert werden (vgl. Abb. 4 a). In dem vorliegenden Bild img_deletedCircles_filtered (=tmp_img) sind nun ausschließlich rechteckige Formen und Pfeile vorhanden. Zur Extraktion der **rechteckigen Formen** aus tmp_img wird eine mehrstufige Konturdetektion basierend auf [Su85] durchgeführt. Im Anschluss werden die einzelnen Konturen approximiert. Wenn das untersuchte Medium vier Eckpunkte hat, wird auch das Verhältnis der Längen und der Breiten des Mediums untersucht. Wenn die Seiten nahezu gleich lang sind und innerhalb des Mediums ein X oder + Symbol detektiert wurde, dann handelt es sich um ein Gateway, ansonsten ist es ein Rechteck (vgl. Abb. 4 b). Die detektierten viereckigen Formen (Rechtecke und Quadrate) werden aus dem Bild img_deletedCircles entfernt. Zur Extraktion **linearer Formen** wird das Bild mehrfach hintereinander mit den morphologischen Operatoren *Dilation* und *Erosion* vorverarbeitet (vgl. Abb. 4 c), sodass im Anschluss mit der Hough-Transformation für Linien und eines Glättungsfilters alle Sequenzflüsse ermittelt werden können (vgl. Abb. 4 d).

Nun müssen die detektierten Formen bezüglich relevanter BPMN-Elemente korrekt klassifiziert werden. Dazu schlagen wir folgendes Vorgehen vor: **Klassifikation detekтирter elliptischer Formen:** $\text{Shape}_{\text{elliptic}} = \{\text{Start event}, \text{End event}, \text{Intermediate event}\}$ mit dem Shape-Kandidaten $\epsilon \in \text{Shape}_{\text{elliptic}}$. Für alle s_e wird überprüft, ob die detektierten Pfeile eingehend oder ausgehend sind. Sind die detektierten Pfeile ausgehend, dann handelt es sich beim s_e um einen Start event, sind die Pfeile eingehend, dann ist s_e ein End event. Andernfalls ist es ein Intermediate event.

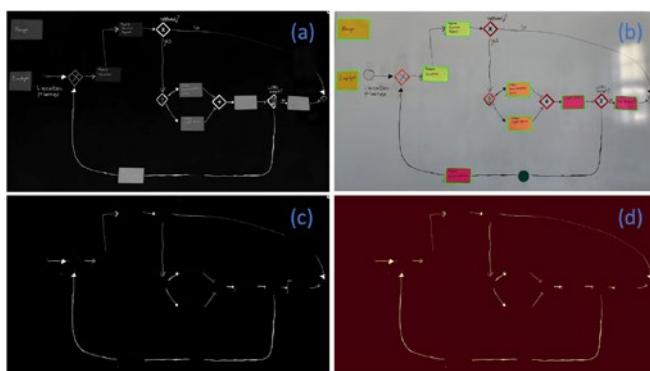


Abb. 4: (a) Kreisförmige Elemente, (b) Quadrate (rot) und Rechtecke (grün) wurden detektiert und aus dem Bild gelöscht, sodass nur lineare Formen erhalten bleiben (c). Hough-Transformation und GlättungsfILTER haben aus diesen linearen Formen alle Pfeilelemente extrahiert (d).

Für alle s_r wird überprüft, ob die rechteckigen Formen (z.B. Post-its) in dem Bereich zwischen Start- und End events positioniert wurden. Falls ja, dann sind es Aktivitäten (tasks). Falls diese Rechtecke (Post-its) rechts von End events bzw. links von Start events platziert sind, handelt es sich um Lanes. **Klassifikation detektierter quadratischer Formen:** $Shape_{square} = \{XOR-Gateway, AND-Gateway\}$ mit dem Shape-Kandidaten $s_s \in Shape_{square}$. Für alle s_s wird mittels eines Template Matching Algorithmus [Br09, OP22] überprüft, ob in dem Quadrat ein X oder ein + Symbol vorliegt. Symbolfreie Quadrate werden als XOR-Gateways klassifiziert. **Klassifikation detektierter Pfeile:** $Shape_{arrows} = \{Eingehende Pfeile, ausgehende Pfeile\}$ mit dem Shape-Kandidaten $s_a \in Shape_{arrows}$. Für alle s_a werden zuerst die möglichen Start- und Endpunkte festgehalten. Daraufhin wird die Pfeilspitze ermittelt, indem das Verhältnis der Pixelintensität dieser beiden Punkte berechnet wird. Der Punkt mit der höchsten Pixelintensität ist die Spitze von s_a . Aus dem Zielelement heraus betrachtet ist dieser Pfeil somit ein eingehender Pfeil. Andernfalls wäre es ein ausgehender Pfeil.

Output Generation: Im letzten Schritt wird aus den klassifizierten Elementen ein Prozessmodell im BPMN 2.0 XML-Format erstellt. Zur Erstellung des XMLs wurde die externe Camunda BPMN Model API eingebunden.

3 Ergebnisse

Merkmalsextraktion: Das Ziel des vorgestellten Ansatzes ist es aus dem vorliegenden Testdatensatz (70 abfotografierte BPMN-Skizzen) folgende Merkmale (geometrische Formen) zu detektieren: Kreise, Rechtecke, Quadrate und Geraden, die zwei Elemente verbinden (=Pfeile). Folgende Tabelle 1 stellt die Ergebnisse dar. Zusätzlich zu Post-its und Marker wurden den Modellierungsteams zwei haptische Elemente (Magnetkreise und Magnetquadrate) zur Verfügung gestellt, die optional eingesetzt werden konnten, um die Events (Kreise) und Gateways (Quadrate) nicht zeichnen zu müssen. Eine genauere Untersuchung der BPMN-Skizzen hat gezeigt, dass diese Elemente bevorzugt eingesetzt wurden (nur 45 von 231 Kreisen, 15 von 420 Quadrate wurden händisch gezeichnet).

Klassifikation: Hierbei fällt auf, dass die korrekte Klassifikation mit Post-its dargestellter rechteckiger Formen (Lanes vs. Activities) am besten funktioniert und sehr robust ist (99 %). Die Klassifikation der kreisförmigen Elemente liefert bei Intermediate events die besten Ergebnisse (96 %). Die Klassifikation der Sequenzflüsse (eingehend und ausgehend) basiert auf den Pixelintensitätvergleich der Start- und Endpunkte bei jeweils einzelnen Geraden. Dieses Kriterium funktioniert an sich gut, liefert jedoch im Gegensatz zur Klassifikation der anderen Elemente schlechtere Ergebnisse (74 % bei eingehenden Pfeilen, 70 % bei ausgehenden Pfeilen).

In Testbildern vorhandene geometrische Formen (Gesamtzahl n)	Merkmalsextraktion			Klassifikation vorliegender Formen
	Anzahl detektierter Formen	BPMN-Element-Klasse		
Kreis (n = 231)	224 (96.67 %)	Start event	n = 68	63 (92.65 %)
		End event	n = 129	104 (80.62 %)
		Intermediate event	n = 27	26 (96.30 %)
Rechteck (n = 868)	864 (99.54 %)	Activity (Task)	n = 621	620 (99.84 %)
		Lane	n = 243	241 (99.17 %)
Quadrat (n = 420)	413 (98.13 %)	Parallel Gateway	n = 165	138 (83.64 %)
		Exclusive Gateway	n = 248	218 (87.90 %)
Pfeil (n = 931)	807 (86.68 %)	Sequence Flow (eingehend)	n = 409	318 (77.75 %)
		Sequence Flow (ausgehend)	n = 398	312 (78.39 %)

Tabelle 1: Ergebnisse der Merkmalsextraktion und der anschließenden Klassifikation

4 Diskussion

Tabelle 2 stellt eine Übersicht der Gesamtergebnisse des neuen Ansatzes dar. Der Überblick kann entnommen werden, dass die entwickelten Algorithmen zur vollautomatischen Digitalisierung von händisch gefertigter BPMN-Modelle gut geeignet sind. Am besten funktionieren die Verarbeitungspipelines bei *Tasks*, *Lanes*, *Intermediate events* und *Start events* mit über 90 % Erfolgsrate. Auch *Gateways* (parallel und exclusiv) und *End events* werden relativ sicher ermittelt und digitalisiert (Erfolgsrate im Bereich von 79 % - 87 %). Bei Sequenzflüssen ist die absolute korrekte Extraktion und anschließende Klassifikation insgesamt vergleichsweise nicht sehr robust (68 %).

Manuell gezeichnete BPMN-Elemente (Gesamtzahl n)	Detektierte und korrekt klassifizierte BPMN-Elemente in der digitalen Datei
Start event (n = 70)	63 (90.10 %)
End event (n = 133)	104 (78.95 %)
Intermediate event (n = 28)	26 (92.85 %)
Activity (Task) (n = 623)	620 (99.50 %)
Lane (n = 245)	241 (98.36 %)
Parallel Gateway (n = 168)	138 (82.14 %)
Exclusive Gateway (n = 252)	218 (86.51 %)
Sequence Flow (n=931)	585 (67.84 %)

Tabelle 2: Gesamtergebnisse (Extraktion und anschließende Klassifikation) der zuvor händisch gefertigten BPMN-Elemente

Während die Erkennung im ersten Schritt der Verarbeitungsspipeline viel bessere Ergebnisse liefert (87 %), erfolgt die darauf folgende Klassifikation (eingehende vs. ausgehende Pfeile) nicht so gut. Bei der Grundidee zur Unterscheidung zwischen den beiden Pfeiltypen (Pixelintensitätsverhältnis zwischen den jeweiligen Start- und Endpunkten, um die Pfeilspitze zu detektieren) besteht Optimierungsbedarf. Die entwickelten Verarbeitungspipelines lassen sich einerseits reibungslos auf andere Nutzungsszenarien übertragen, wie z.B. Abfotografieren von existierenden Modellen (aus Anforderungsdokumenten oder Präsentationen), bei denen man die elektronische Quelle nicht hat bzw. die Quelldatei nicht importiert werden kann (z.B. Word-Dokumente) oder das aktuelle Werkzeug den Dateityp nicht unterstützt. Andererseits lässt sich der Ansatz zur Digitalisierung der geometrischen Element problemlos auf andere Diagrammtypen (Klassendiagramme, Use-Case-Diagramme etc.) übertragen, weil solche Diagramm in der Regel mit denselben geometrischen Formen dargestellt werden.

Die ersten Analysen haben gezeigt, dass der vorgeschlagene Ansatz mit optionalen haptischen Elementen zur Förderung des kreativen Modellierungsprozesses gerne aufgegriffen wird. Die Modellierungsteams haben fast ausschließlich magnetische Elemente verwendet, anstatt z.B. die Gateways selbst auf das Whiteboard zu zeichnen (nur 15 von 420 Gateways wurden von Hand gezeichnet). Der Grund dafür war zum einen, dass es schneller und sauber ist (ohne Verschmieren) und zum anderen, dass es angenehmer ist, die haptischen Elemente in der Hand zu halten. Kreise hingegen wurden im Vergleich zu den Gateways häufiger mit der Hand gezeichnet (45 von 231), weil sie einfacher zu zeichnen sind. Es wurde jedoch keine systematische Nutzererfahrung durchgeführt, um den Spaßfaktor sowie die Modellierungsdauer genauer zu messen.

5 Fazit

In dieser Arbeit haben wir unseren BPMN-CREATOR Tool zur vollautomatischen und robusten Digitalisierung von zuvor manuell erstellten BPMN-Modellen vorgestellt. Benutzer müssen ein Foto des kollaborativ erstellten Whiteboard-Diagramms machen, es ggf. zuschneiden und „Create BPMN“ anklicken. Daraufhin wird die Skizze vollautomatisch und in Echtzeit zu einem BPMN-XML Dokument transformiert. Diese Datei kann heruntergeladen und in anderen BPMN-Tools, wie z.B. Signavio, SAP Solution Manager weiterverwendet werden. Zur Untersuchung unseres Ansatzes haben wir 70 BPMN-Modelle erstellt und diese Skizzen durch Postprocessing (Rauschsimulation, Disrektisierung etc.) vervielfacht. Zur Erkennung der Handschrift haben wir zunächst eine handelsübliche Lösung verwendet. Für die Zukunft planen wir (1) eine Eigenentwicklung der Handschrifterkennung, (2) Testen unterschiedlicher Hintergründe (z.B. karierte Blätter), (3) Verbesserung der aktuellen Pfeilspitzenerkennung, (4) Erweiterung detektierbarer BPMN-Elemente, (5) User Experience Studien, um den Spaß- und Kreativitätsfaktor bei optionalen haptischen Elementen genauer zu erfassen und (6) einen Syntax-Checker, um die skizzierten Geschäftsprozessmodelle zu bewerten (bereits in der Implementierungsphase).

Literaturverzeichnis

- [We19] Weske, M.: Business Process Management. Concepts, Languages, Architectures, ISBN:978-3-662-59431-5, Springer Berlin, 2019.
- [BR14] v. Brocke, J; Rosemann, M.: Handbook on Business Process Management 1&2, ISBN: 978-3-642-54915-1. Springer Berlin, Heidelberg, 2014.
- [Lü11] Lübbecke, A.: Tangible Business Process Modeling, 2011.
- [Du05] Dumas, M. et.al.: Fundamentals of Business Process Management, ISBN: 978-3-662-56508-7, Springer Berlin, 2018.
- [FV05] Fruhling, A.; d. Vreede, G.-J.: “Collaborative Usability Testing to Facilitate Stakeholder Involvement,” Value Based Software Engineering. Springer, Berlin, pp. 201-223, 2005.
- [CVDK07] Cherubini, M.; Venolia, G.; DeLine, R.; Ko, A.J.: “Let’s Go to the Whiteboard: How and Why Software Developers Use Drawings,” SIG CHI, pp. 557-566, 2007.
- [RKV08] Renger, M.; Kolfschoten, G.; d. Vreede, G.-J.: “Challenges in collaborative modeling: A literature review,” in Advances in Enterprise Engineering I, Lecture Notes in Business Information Processing, Springer, 2008, pp. 61-77.
- [ME22] Allgeier IT Solutions GmbH: metasonic Process Touch: Business Software zum Anfassen, URL: <https://www.metasonic.de/en/products/metasonic-process-touch/overview/>, 2022.
- [Wi12] Wittern, H.: “Devices, Empirical Study Evaluating Business Process Modeling on Multi-touch,” in IEEE International Conference on Software Science, Technology and Engineering, 2012.
- [BRW11] Brown, R.; Recker, J.; West, S.: “Using virtual worlds for collaborative business process modeling,” Business Process Management Journal, pp. 546-564, 2011(3).
- [KKRR15] Kammerer, K.; Kolb, J.; Ronis, S.; Reichert, M.: “Collaborative process modeling with tablets and touch tables - A controlled experiment,” in IEEE 9th International Conference on Research Challenges in Information Science, 2015.
- [SALS21] Schäfer, B.; van der Aa, H.; Leopold, H.; Stuckenschmidt, H.: „Sketch2BPMN: Automatic Recognition of Hand-Drawn BPMN Models,“ in International Conference on Advanced Information Systems Engineering, CAiSE, 2021.
- [IK87] Illingworth, J.; Kittler, J.: “The Adaptive Hough Transform,” IEEE Transactions on Pattern Analysis and Machine Intelligence , Issue: 5, pp. 690-698, 1987.
- [Su85] Suzuki, S.: „Topological structural analysis of digitized binary images by border following,“ Computer Vision, Graphics, Image Processing, 30(1), p. 32–46, 1985.
- [Br09] Brunelli, R.: Template Matching Techniques in Computer Vision: Theory and Practice, ISBN: 978-0-470-51706-2, Wiley, 2009.
- [OP22] Template Matching in OPEN CV (Python). URL: https://docs.opencv.org/2.4/doc/tutorials/imgproc/histograms/template_matching/template_matching.html, 2022.

Data Synthesis for Fairness Audits of Learning Analytics Algorithms

Linda Fernsel¹ and Katharina Simbeck²

Abstract: The purpose of methods of fairness auditing is to uncover to what extent Learning Analytics algorithms are fair. Fairness auditing methods often rely on pre-existing test data. In the context of Learning Analytics auditing, learning data is needed for testing. However, learning data might not be available (in large quantities) due to privacy concerns. Our poster shares our findings on how relational data for fairness audits of Learning Analytics systems can be synthesized from little pre-existing data, using the most promising available data synthesizers.

Keywords: Learning analytics, synthetic data, fairness audit

1 Introduction

Learning Analytics is the field concerned with making sense of learning data, enabling stakeholders in learning processes (students, teachers, administrators) to better understand a learning process or environment and possibly adapt it to individual learners (cf. [La22] (p. 8ff)). Learning management systems like Moodle can integrate Learning Analytics, connecting “delivery of educational materials” to “student activity” [La22](p. 10). For instance, Moodle tries to predict which students might fail to complete a course [Mo18]. By conducting fairness audits one can find out whether Learning Analytics algorithms, like Moodle’s dropout prediction model, discriminate against groups of learners (see for example [RS19]).

Data-based fairness audits require test data. The challenge is that test data is not always available, especially not in the context of Learning Analytics [Be16][Do19]: Learning data contains sensitive user data generated over long periods. Therefore the need to synthesize realistic learning data arises.

In section 2 we give a brief overview of related work and highlight how we will contribute to the existing research body. In section 3 we outline our approach. We end with a summary of the content of this poster submission in section 4.

¹ University of Applied Sciences (HTW), 10313 Berlin, fernsel@htw-berlin.de,
<https://orcid.org/0000-0002-0239-8951>

² University of Applied Sciences (HTW), 10313 Berlin, simbeck@htw-berlin.de,
<https://orcid.org/0000-0001-6792-461X>

2 Related Work

The challenge of synthesizing data can be approached via statistical sampling methods (statistical methods) or methods from the field of machine learning (neural methods) [Fa20]. Examples of statistical methods are “copulas, Bayesian Networks, Gibbs sampling and Fourier decomposition” [Fa20]. Variational auto-encoders and generative adversarial networks are examples of neural methods [Fa20]. There are various tools available that implement statistical and/or neural data synthesis methods for realistic tabular data, “AIT-EST” [SAH22] or the “Synthetic Data Vault” for instance [Pa16].

Learning data have been synthesized on several occasions. [Wu15] devises a method, Latent-Explicit Markov Model, for synthesizing realistic MOOCdb learning activity data that can be used in Learning Analytics model training. “DATASIM” is a software for generating “xAPI” (a learning record standard) data, parameterized with information on potential users and user groups, with the goal of testing systems using xAPI [Bl20]. [WLS13] use synthetic learning data for machine learning model performance testing. [Vo21] evaluate one neural- and one statistical method’s usefulness for the generation of privacy-preserving learning data and find that the neural method creates more useful results.

Even though advances have been made in the field of data synthesis, the generation of realistic tabular remains challenging, e.g. because columns have different data types or unknown distributions [Fa20], [Xu19]. None of the named methods and tools can synthesize Moodle learning data in an out-of-the-box way.

2.1 Our Contribution

The topic of our poster is how to synthesize realistic, tabular, relational data that can be used for the fairness audit of a machine learning model in the field of Learning Analytics.

Our work focuses on applying available method(s) of data synthesis for generating realistic, tabular learning data for fairness audits of Moodle’s classification model of students at risk of dropping out. We will evaluate whether different learning styles are represented in the synthetic data and whether the synthetic data can be used for auditing the dropout prediction model.

As synthetic data is necessary for a variety of goals, including but not limited to machine learning model training [Fa20][Bo21][Ur21], machine learning model evaluation [WLS13][Be16][SAH22], and software testing [LMG14][Be16][EMH20], our findings apply in general to the use of synthetic data in machine learning, quality assurance or in learning analytics.

3 Approach

To synthesize learning data specific to the Moodle Learning Management System, we follow the general approach for data synthesis suggested in [EMH20], combined with suggestions from [Fa20] and [SAH22]. [EMH20] break down a general data synthesis process into three basic steps. First, if available, real data or data from “existing models or simulations” is prepared [EMH20]. Next, synthetic data is generated. In the third and last step, the synthesized data’s utility is assessed [EMH20].

The data preparation step can include removing data errors, making data consistent, harmonization, linking data [EMH20], encoding and normalization of pre-existing or simulated data values, and transforming the data into the necessary shape [Fa20].

The data generation step concretely consists of three parts: (1) formulating data requirements in the form of column and association constraints [SAH22], (2) creating data, e.g. by training a model with real data and then generating data with the trained model, and (3) post-processing created data [Fa20].

We select the most promising available solutions for synthesizing tabular data and try them in the use case of generating data for Moodle’s Learning Analytics algorithm for dropout prediction. Utility assessment (cf. [EMH20]) allows evaluating how useful each tried data synthesis method is.

Challenges are (1) that only little pre-existing data is available for seeding and utility assessment, (2) that data needs to realistically represent different learning styles, and (3) that created data needs to be integrated into the Moodle Learning Management System to be used in the fairness audit of Moodle’s dropout prediction model.

4 Summary

We share experiences and findings from following a three-step data synthesis approach - seed data preparation, data generation, and utility assessment – for creating realistic, tabular learning data.

5 Acknowledgments

This research is funded by the Federal Ministry of Education and Research of Germany (project nr: 16DHB4002). Thanks to André Selmanagic and three anonymous reviewers for their feedback.

Literature

- [Be16] Berg, A. M. et. al.: The Role of a Reference Synthetic Data Generator within the Field of Learning Analytics. *Journal of Learning Analytics*, Vol. 3 (1), P. 107-128, 2016.
- [Bl20] Blake-Plock, S: DATASIM – Data and Training Analytics Simulated Input Modeler. Technical Report, Yet Analytics, Baltimore, Maryland, USA, 2020.
- [Bo21] Bourou, S. et. al.; A Review of Tabular Data Synthesis Using GANs on an IDS Dataset. *Information*, Vol 12 (9), 2021.
- [Do19] Dorodchi, M. et. al.: Using Synthetic Data Generators to Promote Open Science in Higher Education Learning Analytics. In (IEEE): International Conference on Big Data (Big Data), Los Angeles, 2019. P. 4672-4675.
- [EMH20] El Emam, K.; Mosquera, L.; Hoptroff, R.: Practical Synthetic Data Generation, O'Reilly Media, Inc., Sebastopol, USA, 2020.
- [Fa20] Fan, J. et. al: Relational Data Synthesis Using Generative Adversarial Networks: A Design Space Exploration. *Proceedings of the VLDB Endowment*, 2020. Vol. 13 (12), P. 1962-1975.
- [La22] Lang, C. et. al. (Eds): *Handbook of Learning Analytics* (2nd ed.). Society of Learning Analytics Research (SoLAR), Vancouver, 2022.
- [LMG14] Lu, W.; Miklau, G.; Gupta, V.: Generating private synthetic databases for untrusted system evaluation. In (IEEE): 30h International Conference on Data Engineering, Chicago, 2014. P. 652-663.
- [Mo18] Monllaó Olivé, D. et. al.: A supervised learning framework for learning management systems. In: *Proceedings of the First International Conference on Data Science, E-learning and Information Systems (DATA)*, Madrid, 2018.
- [Pa16] Patki, N.; Wedge, R.; Veeramachaneni, K.: The Synthetic Data Vault. In (IEEE): International Conference on Data Science and Advanced Analytics (DSAA), Montreal 2016. P. 399-410.
- [RS19] Riazy, S.; Simbeck, K.: Predictive Algorithms in Learning Analytics and their Fairness. In (Gesellschaft für Informatik e.V. (GI)): DeLF1 – Die 17. Fachtagung Bildungstechnologien, Berlin, 2019. P. 223-228.
- [SAH22] Saha, D.; Aggarwal, A.; Hans, S.: Data Synthesis for Testing Black-Box Machine Learning Models. In (ACM): 5th Joint International Conference on Data Science & Management of Data (9th ACM IKDD CODS and 27th COMAD), Bangalore, India, 2022. P. 110-114.
- [Ur21] Urbina, F. et. al.: MegaSyn: Integrating Generative Molecule Design, Automated Analog Designer and Synthetic Viability Prediction. *ACS Omega*, Vol 7 (22), P. 18699-18713, 2022.
- [WLS13] Waters, A. E.; Lan, A. S.; Studer, C.: Sparse probit factor analysis for learning analytics. In (IEEE): International Conference on Acoustics, Speech and Signal Processing, Vancouver, 2013. P. 8776-8780.
- [Wu15] Wu, M. Q.: The synthetic student: a machine learning model to simulate MOOC data. Master thesis, MIT, Department of Electrical Engineering and Computer Science, Cambridge, 2015.

- [Vo21] Vostres, Y.: Generierung synthetischer Daten – Betrachtung hinsichtlich Datenschutz, Innovationspotential und technischer Umsetzung. Bachelor thesis, University of Applied Sciences (HTW), Berlin, 2021.
- [Xu19] Xu, L. et. al: Modeling Tabular data using Conditional GAN. In (Curran Associates, Inc.): Advances in Neural Information Processing Systems 32 (NeurIPS), Vancouver, 2019. P. 7335-7345.

Instrumente zur aktiven und sicheren Verbraucherteilhabe an öffentlichen Online-Diensten

Esther Ruiz Ben¹, Anne Jellinghaus², Britta Hesse³, Margit Scholl⁴, Sebastian Möller⁵

Abstract: Das Projekt „Instrumente zur aktiven und sicheren Verbraucherteilhabe an Online Public Services“ (IVTOPS) analysiert die Verbraucherakzeptanz von öffentlichen Online-Diensten am Beispiel von Nutzerkonten in Berlin und Brandenburg, um geeignete Instrumente zur Überwindung von Akzeptanzbarrieren bzw. Informations- und digitalen Kompetenzasymmetrien bezüglich Datensicherheit und Datenschutz bei der Nutzung von öffentlichen Online-Diensten zu schaffen und zu testen.

Keywords: Öffentliche Online-Dienste, Datensicherheit und -schutz, Usable Privacy and Security, Digital Divide.

1 Projektziele

Die Digitalisierung von Public Services ist für Verbraucher*innen eine große Herausforderung. In einer sich immer stärker diversifizierenden Angebots- und Anbietervielfalt von digitalen öffentlichen Diensten benötigen Verbraucher*innen Orientierung, um Nutzungsentscheidungen treffen zu können. Speziell Datensicherheit und -schutz spielen dabei eine ebenso wesentliche Rolle wie Informationen und digitale Kompetenzen, die zur inklusiven Teilhabe in der Nutzung von digitalen Public Services beitragen. Wie die Ergebnisse des jährlich durchgeführten eGovernment MONITOR⁶ zeigen, erweist sich unter anderem die Angst vor Datendiebstahl als wesentliche Barriere für die Nutzung von E-Government Online-Angeboten bzw. digitalen Public Services. Auch wenn diese Angst der Verbraucher*innen als Akzeptanzbarriere für die Nutzung von digitalen Public Services bekannt ist, wurde die Ursache dieser Akzeptanzbarriere bzw. den Informations- und Kompetenzstand der Verbraucher*innen bezüglich Datenschutz und -sicherheit kaum analysiert.

1 TH Wildau, Wirtschaft, Informatik Recht, Technische Hochschule Wildau
Hochschulring 1, 15745 Wildau, ruiz_ben@th-wildau.de

2 TH Wildau, Wirtschaft, Informatik Recht, Technische Hochschule Wildau
Hochschulring 1, 15745 Wildau, anne.jellinghaus@th-wildau.de

3 TU Berlin, Institut für Softwaretechnik und Theoretische Informatik, Quality and Usability Lab, TU-Hochhaus Ernst-Reuter-Platz 7, 10587 Berlin, britta.hesse@tu-berlin.de

4 TH Wildau, Wirtschaft, Informatik Recht, Technische Hochschule Wildau
Hochschulring 1, 15745 Wildau, margit.scholl@th-wildau.de

5 TU Berlin, Institut für Softwaretechnik und Theoretische Informatik, Quality and Usability Lab, TU-Hochhaus Ernst-Reuter-Platz 7, 10587 Berlin, sebastian.moeller@tu-berlin.de

6 <https://initiatived21.de/publikationen/egovernment-monitor-2019/> (Zugriff: 20.04.2022)

Instrumente zur Lösung von diesen Akzeptanzproblemen bei öffentlichen Online-Diensten wurden bisher nicht umgesetzt. Heuwing and Maletz [HM19] erklären, dass partizipative Entwicklungsmethoden in der Gestaltung von öffentlichen Online-Diensten nicht etabliert sind. Diese Autor*innen haben partizipative Ansätze für die Umsetzungsempfehlungen von öffentlichen Online-Diensten in Ländern und Kommunen verwendet. Usable Security Aspekte werden dabei jedoch nicht genutzt.

Das laufende Projekt IVTOPS⁷ analysiert die Verbraucherakzeptanz von öffentlichen Online-Diensten am Beispiel von Nutzerkonten in Berlin und Brandenburg, um geeignete Instrumente zur Überwindung von Akzeptanzbarrieren bzw. Informationsdefiziten und digitalen Kompetenzasymmetrien bezüglich Datensicherheit und -schutz bei der Nutzung von öffentlichen Online-Diensten zu schaffen und zu testen.

2 Methodologie

Es werden qualitative und quantitative Methoden angewandt, die sukzessiv die Entwicklung der verschiedenen Methodeninstrumente unterstützen. In der explorativen Phase wurde mit Hilfe einer Dokumentenanalyse der Stand der Entwicklung und Nutzung von Servicekonten bzw. eIDAS in Berlin und Brandenburg sowie von Messinstrumenten für die Akzeptanzanalyse eruiert. Basierend auf diesen Ergebnissen wird die Akzeptanz von Servicekonten aus der Perspektive des Digital Divide in der Analysephase mit qualitativen (Experteninterviews) und quantitativen (Umfrage bei Nutzer*innen von Servicekonten bzw. eIDAS in Berlin und Brandenburg) Methoden genau analysiert. Aus den Erkenntnissen dieser Analysen werden Usable Security-Instrumente für eIDAS bzw. Servicekonten partizipativ (Fokusgruppen und Design Thinking) entwickelt und evaluiert. Die Abb. 1 zeigt eine Zusammenfassung der angewandten Methodologie:



Abb. 1: IVTOPS-Methodologie

⁷ Das diesem Poster zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz unter dem Förderkennzeichen 28V1403A20 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autor*innen.

3 Usable Privacy und Security Instrumente

Zusammen mit Verbraucher*innen werden „Usable Privacy und Security Instrumente“ (Privacy Enhancing Technologies – PETs [AS17], [Be20], [SD18], [SB15]) zur Überwindung von Informationsdefiziten und digitalen Kompetenzasymmetrien bezüglich Datensicherheit und Datenschutz bei der Nutzung von öffentlichen Online-Diensten entwickelt. Dazu werden Design-Thinking-Workshops mit Verbraucher*innen organisiert, um Usable Privacy und Security Instrumente als Prototypen partizipativ zu entwickeln. Die gemeinsam partizipativ entwickelten Prototypen werden in Fokusgruppen danach diskutiert bzw. getestet.

Wir gehen davon aus, dass ein partizipatives Design (z. B. Fokusgruppen und Design-Thinking-Methoden) von Usable Security Instrumenten notwendig ist, um mögliche Intransparenzen von „Security Nudges“ zu vermeiden [Fu20].

4 Vorläufige Ergebnisse

Die vorläufigen Ergebnisse der Akzeptanzuntersuchung zeigen u. a., dass öffentliche Online-Dienste und speziell Nutzerkonten wenig bekannt sind und kaum genutzt werden. Auch wenn die Mehrheit der Befragten den Zugang zu diesen Diensten als unproblematisch einschätzt, fühlen sich viele unsicher hinsichtlich der Selbstbestimmung über ihre persönlichen Daten sowie bezüglich der Verarbeitung ihrer Daten in den Nutzerkonten. Hinsichtlich der Benutzerfreundlichkeit der Nutzerkonten wird die Registrierung von der Mehrheit der Befragten als umständlich gewertet. Bei dem Design der „Usable Security Prototypen“ werden wir u. a. diese Aspekte als Grundlage verwenden.

Literaturverzeichnisses

- [AS17] Acquisti, M.; Sleeper, Y.; Wang, S.; Wilson, I.; Adjerid, R.; Balebako, L.; Brandimarte, L. F.; Cranor, S.; Komanduri, P. G.; Leon, N. Sadeh; Schaub, F.: Nudges for privacy and security. In: ACM Computer Surv.: 50: 3: 1–41, 2017
- [Be20] Bertino, E.: Privacy enhancing techniques for security. In (S. E. Woo, L. Tay, & R. W. Proctor Hrsg.): Big data in psychological research (411–424). American Psychological Association, 2020
- [Fu20] Fuhrberg, R.: Verhaltensökonomie in der Verwaltungskommunikation – Der Staat als Entscheidungsarchitekt. In: K. Kocks et al. Öffentliche Verwaltung – Verwaltung in der Öffentlichkeit. Berlin: Springer, 2020
- [HM19] Heuwing, B.; Maletz, C.: Bürger-Services nutzerzentriert gestalten. Mit Digitalisierungslaboren Verwaltung und Bürger zusammenbringen. In: S. Hess; H. Fischer (Hrsg.): Mensch und Computer 2019 – Usability Professionals, 08.-11. September 2019, Hamburg.
- [SD18] Sunyaev, A.; Dehling, T.; Schmidt-Kraepelin, M.: Verbraucherorientierter Datenschutz: Identifizierung von Verbraucherarchetypen zur effektiven Kommunikation von Datenschutzpraktiken. In (Bala, C.; Schuldzinski, W. Hrsg.), Jenseits des Otto Normalverbrauchers: Verbraucherpolitik in Zeiten des „unmanageable consumer“ (S. 163-179). Düsseldorf: Kompetenzzentrum Verbraucherforschung NRW. https://doi.org/10.15501/978-3-86336-920-0_8 (20.06.2020), 2018
- [SB15] Schaub, F.; Balebako, R.; Durity, A. L.; Cranor, L. F.: A design space for effective privacy notices. In: Eleventh symposium on usable privacy and security (SOUPS 2015), 1–17. Ottawa: USENIX Association.<https://www.usenix.org/conference/soups2015/proceedings/presentation/schaub> (Zugriff: 20.04.2022), 2015

Serious Games als Lernmethode zur Steigerung der Informationssicherheit

Frauke Prott¹, Ulrike Küchler², Regina Schuktomow³ und Margit Scholl⁴

Abstract: Sensibilisierung für Informationssicherheit und die Förderung entsprechender Kenntnisse sind essentiell für eine gelungene Digitalisierung. Im vorliegenden Projekt werden dafür u. a. digitale Serious Games entwickelt und erprobt. Serious Games bieten in der Nachbildung realistischer Alltagssituationen einen geschützten Raum, in dem Fehler gemacht, Konsequenzen von getroffenen Entscheidungen ohne Folgen für das wahre Leben erlebt und verschiedene Wege ausprobiert werden können. So kann sicherheitsbewusstes Verhalten im Falle eines Angriffs auf die Informationssicherheit eingeübt werden. Die Ergebnisse eines Usertests zeigen, dass narratives Lernen mit Serious Games zum Thema Informationssicherheit positiv angenommen wird und zum Lernerfolg beiträgt.

Keywords: Serious Game, digitale Lernszenarien, Informationssicherheit, Sensibilisierung, KMU

1 Einleitung

Erfolgreiche Digitalisierung benötigt ein hohes Niveau an Informationssicherheit [BMI21]. Da Informationssicherheits-Vorfälle häufig durch Unwissenheit und/oder aktives Handeln ermöglicht werden (z. B. Klicken eines Links, Öffnen eines Anhangs), ist die Sensibilisierung von Nutzenden für ein sicheres Verhalten im digitalen Umfeld unerlässlich [BMI21]. Insbesondere kleine und mittlere Unternehmen (KMU) sind aufgrund mangelnden Bewusstseins für die Risiken des Einsatzes von Informationstechnologie, knapper Ressourcen und fehlenden Wissens besonders gefährdet [BMI21] [BSI21].

Daher verfolgt das Projekt „Awareness Labor KMU (ALARM) Informationssicherheit“, gefördert vom Bundesministerium für Wirtschaft und Klimaschutz, das Ziel, ein innovatives Gesamtszenario zur Unterstützung von KMU zur Erhöhung ihrer Informationssicherheit und zur Etablierung einer nachhaltigen Sicherheitskultur zu entwickeln und zu erproben. Das Gesamtszenario umfasst analoge und digitale Lernszenarien (Serious Games) sowie „Vor-Ort-Angriffe“ (z. B. Phishing-Simulation) zur Sensibilisierung für Informa-

1 Technische Hochschule Wildau (TH Wildau), Fachbereich Wirtschaft, Informatik, Recht (WIR), Hochschulring 1, 15745 Wildau, frauake.prott@th-wildau.de

2 Gamebook Studio HQ GmbH, Wilhelm-Kabus-Straße 77, 10829 Berlin, ulrike.kuechler@gamebook.studio

3 TH Wildau, Fachbereich WIR, Hochschulring 1, 15745 Wildau, regina.schuktomow@th-wildau.de

4 TH Wildau, Fachbereich WIR, Hochschulring 1, 15745 Wildau, margit.scholl@th-wildau.de

tionssicherheit sowie Befragungen, Wissenstests, Awareness-Messungen und Reifegrad-aussagen. Informationssicherheit soll (be-)greifbar und emotional erlebbar werden. Der vorliegende Beitrag widmet sich den digitalen Serious Games in diesem Projekt.

2 Literaturüberblick

2.1 Game-based Learning

Game-based Learning (GBL) wird als unterhaltsame und motivierende Form des Lernens beschrieben [LA09]. Lernförderliche Eigenschaften von Spielen sind klare Zielvorgaben und direktes Feedback [FZC13]. Die Teilnehmenden arbeiten auf ein Ziel hin, wählen und führen Aktionen aus und erleben unmittelbar die daraus resultierenden Konsequenzen. GBL ermöglicht, in einem geschützten Raum Fehler zu machen sowie zu experimentieren [Tr14]. Die gemachten Erfahrungen sind gut von der simulierten in die reale Welt zu übertragen [Tr14]. Game-based Learning ermöglicht statt einer passiven Informationsaufnahme (Vortrag, Video, Lesen) eine aktive Auseinandersetzung mit den Lerninhalten.

Aufgrund dieser Eigenschaften belegen zahlreiche Studien positive Wirkungen des Einsatzes von Spielen und Spielelementen: GBL-Umfelder sind hoch involvierend und unterstützen daher effektiv den Lernprozess [Bu15]. Spiele als Lernmethode verbessern kurzfristige und langfristige Lernergebnisse [Wo13]. Spielebasierte Lernszenarien erhöhen die Motivation und fördern Verhaltensänderungen [BK11] [Hs08]. Wichtig ist die spezifische Ausrichtung an der Alltagswirklichkeit der Zielgruppen, da die Verbindung zu realen Situationen und Herausforderungen den Lernerfolg verbessert [Lo07].

2.2 Serious Games

Serious Games nutzen (Computer-)Spiele, um für ein Thema zu sensibilisieren sowie Wissen und Fähigkeiten zu entwickeln, indem sie den Lernenden ermöglichen, in Situationen einzutauchen, die sonst schwer oder selten erlebt werden können [Yp14]. Im Gegensatz zu Unterhaltungsspielen zielen Serious Games neben Unterhaltung auf die Vermittlung von Lerninhalten [MM16]. Serious Games fördern den Lernerfolg durch die Ansprache mehrerer Sinne (z. B. visuell, auditiv) und somit verschiedener Lerntypen sowie durch die aktive Einbindung der Spielenden, die dadurch relevante Bezüge zu realen Situationen herstellen können [Yp14].

Narrative – sinnstiftende Erzählungen – fördern das Eintauchen in ein Serious Game, was entscheidend für den Lernerfolg ist. Narrative laden die Spielenden ein, an der Geschichte teilzuhaben sowie ihren Verlauf mitzubestimmen, und fördern die intrinsische Motivation

zu lernen. Dabei sollten die Geschichten unter anderem die Fantasie anregen und empathische Charaktere beinhalten [NL20].

2.3 Sensibilisierung für Informationssicherheit

Klassische Lernmethoden, bei denen Fachpersonen in z. B. Vorträgen Lerninhalte an Lernende vermitteln, scheinen eher ineffektiv und passiv im Vergleich zu moderneren Methoden, die auf „learning by doing“ oder experimentellem Lernen basieren [Yp14]. Gleichwohl erfolgen die Sensibilisierung für Informationssicherheit und die Schulung entsprechender Kenntnisse in der Praxis oftmals nach diesem Prinzip durch die Bereitstellung von Web-Based-Trainings (WBT) oder durch Vorträge [Al16].

Gerade ein oftmals abstraktes und komplexes Thema wie Informationssicherheit erfordert eine aktive Auseinandersetzung mit Lerninhalten, wie es Game-based Learning ermöglicht. Im Ernstfall eines Angriffsversuchs auf die Informationssicherheit sollten Nutzende fähig sein, richtig zu handeln. Um das richtige Verhalten für den Ernstfall, bei Bedarf auch wiederholend, einzuüben, empfehlen sich Serious Games. Sie bieten in der Nachbildung realistischer Alltagssituationen einen geschützten Raum, in dem Fehler gemacht, Konsequenzen von getroffenen Entscheidungen ohne Folgen für das wahre Leben erlebt und verschiedene Wege ausprobiert werden können [Tr14]. Im Gegensatz zu kognitiven Informationsvermittlungen ermöglichen Serious Games ein stärkeres Involvement der Teilnehmenden, mehr Lebendigkeit sowie Interaktion und Lernen gemäß der von Pestalozzi geprägte Formel mit „Kopf, Herz und Hand“ [BroJ].

Daher haben in den vergangenen Jahren vermehrt Forschungsprojekte [Ha20] [HAB16] und kommerzielle Anbieter [Ka22] [FaoJ] Serious Games zur Sensibilisierung für Informationssicherheit und Vermittlung entsprechender Kenntnisse entwickelt, um Schulungen mit reiner Wissensvermittlung zu ergänzen. Sowohl analogen als auch digitalen Serious Games werden positive Wirkungen auf die Beteiligung der Teilnehmenden und den Lernerfolg zugeschrieben [Ha20] [Ya19] [GS18].

3 Entwicklung von digitalen Serious Games zur Steigerung der Informationssicherheit

3.1 Storykonzept

Die digitalen Serious Games im vorliegenden Projekt sind immersive Geschichten, die Alltagssituationen aus dem Berufsleben in KMU darstellen. Die Spielenden erleben die Geschichten in der Ich-Perspektive. Dies ermöglicht eine intensive Auseinandersetzung und Identifikation mit den enthaltenen Lerninhalten – und damit auch eine besondere Effizienz und Nachhaltigkeit in der Wissensvermittlung. Für die Entwicklung der Serious Games wurde das Visual Novel Format gewählt – eine Art interaktives Buch, in dem die Spielenden Entscheidungen treffen und damit den weiteren Verlauf der Geschichte bestimmen [Ch19]. Die Lernszenarien sind von Bildern, geschriebenen Dialogen und Situationsbeschreibungen geprägt. Dreidimensionalität wird durch Ambient-Hintergrundmusik und Soundeffekte (z. B. Räuspern, Telefonklingeln, Türzuschlagen) erzeugt.

Das Storykonzept der insgesamt sieben Serious Games zeichnet sich durch Vielfalt, Individualität und Kontinuität aus. Jedes Lernszenario behandelt schwerpunktmäßig ein anderes informationssicherheitsrelevantes Thema: von Passwörtern, über Datenschutz in der Cloud und CEO Fraud⁵, bis zu Informationsklassifizierung. Dabei nehmen die Spielenden wechselnde Rollen ein – einmal agieren sie als Sicherheitsexpertin, ein anderes Mal als Hackende oder Ermittelnde oder sogar als Künstliche Intelligenz. Dies erlaubt nicht nur ein abwechslungsreiches Spiel, sondern ermöglicht, die Themen aus verschiedenen Blickwinkeln kennenzulernen und zu verstehen. Die sieben Serious Games können unabhängig voneinander und in beliebiger Reihenfolge gespielt werden. Jede Geschichte und jeder Lerninhalt sind in sich geschlossen. So können die einzelnen Themen in einer größeren Breite und Tiefe behandelt werden. Gleichwohl sind die einzelnen Geschichten durch eine übergreifende Gesamtstory, die in einem fiktiven Unternehmen spielt, miteinander verknüpft und die Spielenden begegnen in jedem Serious Game denselben Personen (Chef, Disponentin, Auszubildender, Werkstattleiter) und lernen sie immer besser kennen.

⁵ Betrugsmethode, bei der Kriminelle sich als Führungskraft der Organisation ausgeben und Mitarbeitende auffordern, Geld zu überweisen [LS20]

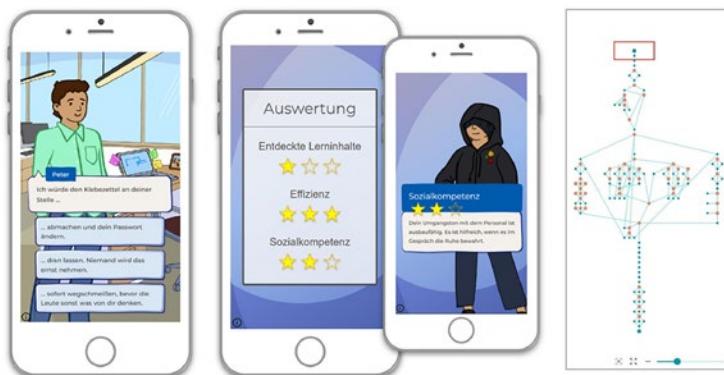


Abb. 1: Beispiele einer Entscheidungsauswahl, eines Feedbacks und eines Entscheidungsbaums

Jedes Lernszenario bietet ein personalisiertes Lernerlebnis. Zu Beginn wählen die Spielenden ihren Charakter und ihren Namen. Mit jeder Entscheidung, die echte Konsequenzen im Spielverlauf mit sich bringt, begeben sich die Spielenden auf ihre ganz persönliche Lernreise, die von ihrem Wissen und ihren Präferenzen bestimmt wird. Jedes Serious Game enthält zwei bis drei Lernpfade, die die Spielenden durch ihre Entscheidungen einschlagen. Die gewählten Wege bestimmen, wie viele Lerninhalte den Spielenden begegnen und welche Schwierigkeiten sie zu bewältigen haben. In jedem Serious Game können die Spielenden unterschiedliche Fähigkeiten (z. B. Sozialkompetenz) stärken und Kenntnisse (z. B. Sicherheitsverständnis) vertiefen, die für Informationssicherheit wichtig sind.

Die Spielenden werden stets von einem Charakter der Gesamtstory begrüßt und darauf aufmerksam gemacht, worum es im vorliegenden Spiel geht, was konkret bewertet wird und worauf sich die Spielenden konzentrieren sollten. Am Ende gibt derselbe Charakter Feedback zu den erzielten Punkten. Dieses beinhaltet zum einen Erläuterungen, wie die Punkteanzahl zustande kommt, und zum anderen Vorschläge und Aufforderungen an die Spielenden. Auch bereits im Laufe des Spiels werden Feedbacknachrichten eingeblendet, die auf vorteilhafte oder nachteilige Entscheidungen und Verhaltensweisen aufmerksam machen. Zudem bietet ein Lexikomodul die Möglichkeit, wichtige Begriffe der Informationssicherheit vor und nach dem Spiel nachzulesen. Die Spielenden werden nicht alleine gelassen, sondern auch in diesem digitalen Format beim Lernen begleitet.

3.2 Technische Anforderungen

Die Serious Games wurden mit der Gamebook Technologie von Gamebook Studio erstellt. Die Gamebook Technologie ist ein Werkzeug zur vereinfachten und schnellen Produktion von interaktiven seriellen Inhalten. Sie verbindet klassische Erzähltechniken aus Literatur und Film mit modernen digitalen Produktionsmethoden und ermöglicht, aus einem analogen Drehbuch eine interaktive Sequenz zu erstellen. Die Lernszenarien werden als statische Inhalte in einer abgeschlossenen HTML5 Webapplikation bereitgestellt und können per iFrame in die Projektwebseite eingebunden werden und sind dort über den Browser abspielbar. Aufgrund der Allgegenwärtigkeit des Smartphones sind die Zielplattform mobile Endgeräte. Somit können Nutzende auf die Serious Games direkt über die Webseite des Projekts zugreifen.

3.3 Usertest

Mitarbeitende der Pilot- und kooperierenden Unternehmen des Projekts evaluieren die Serious Games aus Sicht der Praxis. Die Rückmeldungen aus den Usertests sind die Grundlage für die finale Erstellung der Serious Games. Die Usertests werden von kurzen, standardisierten, schriftlichen Vorher- und Nachher-Befragungen begleitet, die deskriptiv ausgewertet werden. Die Teilnehmenden werden gebeten, auf einer fünf-Punkte-Likert-Skala zu verschiedenen Aspekten (z. B. Inhalt, realistische Darstellung, Spielzeit, Schwierigkeitsgrad) ihre Zustimmung mitzuteilen. Zur Messung des Lernerfolgs dienen Selbsteinschätzungen zur Erfahrung mit dem Thema Informationssicherheit sowie zum Wissenserwerb. Im Folgenden sind die Ergebnisse des Usertests zum Serious Game „Der Hackerangriff“ dargestellt. Hier nehmen die Spielenden die Rolle eines Hackenden ein und versuchen, sich auf dem Firmenserver des fiktiven Unternehmens einzuloggen. Beurteilt werden die Spielenden nach Effizienz – gelingt ihnen der Hack – und Variabilität – wie viele Wege sind die Spielenden zur Erreichung ihres Ziels gegangen.

An dem Usertest nahmen mehrheitlich Männer (65%, N=26)⁶ im Alter von 25 bis 50 Jahren teil (69%, N=26). Die Testpersonen arbeiten hauptsächlich im Vertrieb/Außendienst (23%), in der IT (19%) und im Personalwesen (19%) (N=26). Im Durchschnitt benötigten sie zehn Minuten für das Serious Game (N=22). Diese Spielzeit wurde im Mittel als „genau richtig“ beurteilt (N=31). Auch der Schwierigkeitsgrad tendiert mit einem Mittel von 2,81 zu „genau richtig“ (Skala 1=zu einfach, 3=genau richtig, 5=zu schwer, N=31).

6 Die unterschiedliche Stichprobengröße N ist dadurch bedingt, dass nicht alle Befragten an der Vorher- und Nachher-Befragung teilnahmen bzw. nicht zu allen Fragen Angaben machten.



Abb. 2: Beurteilung des Serious Games „Der Hackerangriff“ im Usertest

Abbildung zwei zeigt eine gute Bewertung des Serious Games „Der Hackerangriff“, da die Testpersonen nahezu allen Aussagen im Durchschnitt mit größer drei zustimmen (Skala 1=überhaupt nicht bis 5=sehr). Wir interpretieren Mittelwerte über dem Skalenmittelpunkt (=3) als zufriedenstellend in der Beurteilung der Nutzenden aufgrund der Heterogenität der Testpersonen im Hinblick auf Branchen, Tätigkeiten und Präferenzen in Bezug auf (digitale) Spiele sowie der Tatsache, dass das linke Ende einer Antwortskala häufiger als das rechte Ende gewählt wird [MB15]. Die geringe Zustimmung zur Aussage „Das Lernszenario möchte ich erneut spielen“ wird relativiert durch die höhere Zustimmung zur Aussage „Das Lernszenario macht Lust, weitere Lernszenarien dieser Art zu spielen“. Gleichwohl wurden im Feedback im betreffenden Serious Game Formulierungsänderungen vorgenommen, um Spielende stärker zu motivieren, das Lernszenario erneut zu spielen. Zudem attestieren die Testpersonen dem Serious Game eine lernförderliche Wirkung. Teilweise wurde neues Wissen erworben und/oder bestehendes vertieft (jeweils M=3,17, N=30). Die eigene Erfahrung mit dem Thema Informationssicherheit wurde nach dem Serious Game leicht höher wahrgenommen ($M_{\text{vorher}} = 3,12$, $M_{\text{nachher}} = 3,32$, N=25).

4 Mehrwert für Wirtschaft und Gesellschaft

Gleichwohl die Serious Games für Mitarbeitende in KMU entwickelt wurden, ist die Sensibilisierung für die behandelten Informationssicherheits-Themen für das Berufsleben insgesamt als auch für das Privatleben (z. B. Passwörter, Telefonbetrug) relevant. Somit besteht der Mehrwert darin, dass die Serious Games für alle interessierten Personen sukzessive auf der Projektwebseite kostenfrei zur Verfügung stehen. Organisationen können damit ihre Mitarbeitenden mit nur geringem Einsatz eigener Ressourcen sensibilisieren und schulen. Ferner können die Serious Games in die Ausbildung von jungen Menschen integriert werden, sodass Auszubildende als auch Studierende bereits mit einem Bewusstsein für Informationssicherheit in die Arbeitswelt starten.

5 Ausblick

Es wird angestrebt, Lernenden auf Basis ihrer erzielten Ergebnisse in den Serious Games einerseits Empfehlungen für das Spielen weiterer Serious Games zu geben, die, im Falle einer geringen Punktzahl, dieselben oder, im Falle einer hohen Punktzahl, andere Fähigkeiten stärken. Andererseits sollen die Teilnehmenden an einen Wissenstest verwiesen werden, in dem sie ihre Kenntnisse zu den im Serious Game behandelten Thema ausbauen können. Aufgrund begrenzter Ressourcen im vorliegen Projekt konnten nicht alle gewünschten Features in den Serious Games umgesetzt werden. In zukünftigen Projekten sollen weitere Ansätze zur Erhöhung der Interaktivität entwickelt und erprobt werden.

Literaturverzeichnis

- [Al16] Allianz für Cyber-Sicherheit: Awareness-Umfrage 2015. Ergebnisse, Stand 5.4.2016. Bonn.
- [BfV22] Bundesamt für Verfassungsschutz: Sicherheitshinweis für die Wirtschaft, 02/2022, 23. März 2022, Betreff Krieg in der Ukraine. 2022.
- [BK11] Bösche, W.; Kattner, F.: Fear of (Serious) Digital Games and Game-based Learning? Causes, Consequences and a Possible Countermeasure. International Journal of Game-Based Learning, 1/3, S. 1–15, 2011.
- [BMI21] Bundesministerium des Innern, für Bau und Heimat (BMI): Cybersicherheitsstrategie für Deutschland 2021. Berlin, 2021.
- [BroJ] Brühlmeier, A.: Heinrich Pestalozzi: Grundgedanken: Erziehung / Bildung. <https://www.heinrich-pestalozzi.de/grundgedanken/erziehung-bildung>, Stand: 8.7.2022
- [BSI21] Bundesamt für Sicherheit in der Informationstechnik (BSI): Die Lage der IT-Sicherheit in Deutschland 2021. Bonn, 2021.
- [Bu15] Buffum, P. S.; Boyer, K. E.; Wiebe, E. N.; Mott, B. W.; Lester, J. C.: Mind the Gap: Improving Gender Equity in Game-Based Learning Environments with Learning Companions. AIED: International Conferences on Artificial Intelligence in Education, 2015.
- [Ch19] Choi, C.: Bigger on the Inside: A History of Visual Novels. <https://medium.com/@cecilchoi/bigger-on-the-inside-a-history-of-visual-novels-981e42f43608>, 22. Februar 2019, Stand: 25.4.2022.
- [FaoJ] Fabula Games: Security Games. <https://fabula-games.de>, Stand: 28.6.2022.
- [FZC13] Fang, X.; Zhang, J.; Chan, S.S.: Development of an Instrument for Studying Flow in Computer Game Play. International Journal of Human-Computer Interaction, 29/7, S. 456–470, 2013.
- [GS18] Ghazvini, A.; Shukur, Z.: A Serious Game for Healthcare Industry: Information Security Awareness Training Program for Hospital Universiti Kebangsaan Malaysia. International Journal of Advanced Computer Science and Application, 9/9, S. 236–245, 2018.
- [Ha20] Hart, S. et al.: Riskio: A Serious Game for Cyber Security Awareness and Education. Computer & Security, 95/August 2020, Artikel 101827, 2020.
- [HAB16] Hendrix, M.; Al-Sherbaz, A.; Bloom, V.: Game Based Cyber Security Training: Are Serious Games suitable for cyber security training? International Journal of Serious Games, 3/1, S. 53–61, 2016.
- [Hs08] Hsu, S. H.; Wu, P. H.; Huang, T. C.; Jeng, Y. L.; Huang, Y. M.: From Traditional to Digital: Factors to Integrate Traditional Game-based Learning into Digital Game-based Learning Environment. Proceedings 2nd IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning, DIGITEL, S. 83–89, 2008.
- [Ka22] Kaspersky: [Dis]connected. A mobile cybersecurity quest.
[https://media.kaspersky.com/en/business-security/enterprise/Kaspersky_\[dis\]connected_datasheet_0121EN_Gl.pdf](https://media.kaspersky.com/en/business-security/enterprise/Kaspersky_[dis]connected_datasheet_0121EN_Gl.pdf), 2022, Stand: 28.6.2022.

- [LA09] Linek, S. B.; Albert, D.: Game-based Learning: Gender-Specific Aspects of Parasocial Interaction and Identification. Conference: International Technology, Education and Development Conference (INTED), 2009.
- [Lo07] Lombardi, M.: Authentic Learning for the 21st Century: An Overview. 2007.
- [LS20] Luber, S.; Schmitz, P.: Definition CEO Fraud / CEO-Betrug – Was ist CEO-Fraud?. Security Insider, 14.12.2020, <https://www.security-insider.de/was-ist-ceo-fraud-a-991462/>, Stand: 12.4.2022.
- [MB15] Menold, N.; Bogner, K.: Gestaltung von Ratingsskalen in Fragebögen. Mannheim, GESIS – Leibniz-Institut für Sozialwissenschaften (SDM Survey Guidelines), 2015.
- [MM16] Mildner, P.; Mueller, F.: Design of Serious Games. In (Dörner, R.; Göbel, S.; Effelsberg, W.; Wiemeyer, J. Hrsg.): Serious Games: Foundations, Concepts and Practice. Springer International Publishing, Cham, S. 57–82, 2016.
- [NL20] Naul, E.; Liu, M.: Why Story Matters: A Review of Narrative in Serious Games. Journal of Educational Computing Research, 58/3, S. 687–707, 2020.
- [Tr14] Trybus, J.: Game-Based Learning: What it is, Why it Works, and Where it's Going. New Media Institute, 2014.
- [Wo13] Wouters, P.; van Nimwegen, C.; van Oostendorp, H.; van der Spek, E. D.: A Meta-analysis of the Cognitive and Motivational Effects of Serious Games. Journal of Educational Psychology, 105/2, S. 249–265, 2013.
- [Ya19] Yasin, A. et al.: Improving software security awareness using a serious game. IET Software, 13/2, S. 159–169, 2019.
- [Yp14] Ypsilanti, A.; Vivas, A.B.; Räisänen, T.; Viitala, M.; Ijäs, T.; Ropes, D.: Are Serious Video Games Something More than a Game? A Review on the Effectiveness of Serious Games to Facilitate Intergenerational Learning. Education and Information Technologies, 19, S. 515–529, 2014.

Userorientierter Prototyp einer Open-Source App für die Paketzustellung

Daniel Quiter¹, Birte Malzahn², Maximilian Engelhardt³, Stephan Seeck⁴

Abstract: Das stark ansteigende Paketvolumen in den letzten Jahren führt zu einer Verschlechterung der Services in der Paketzustellung und zu höherer Unzufriedenheit bei Empfänger*innen. Die Kund*innen fordern eine flexiblere Lieferung von Paketen, die sich an die individuellen Alltagsplanungen anpasst. Die Digitalisierung bietet hierfür zahlreiche Ansätze (z. B. Ankunftszeit voraussagen, Lieferzeitfenster anpassen), die jedoch in den Apps der großen Paketdienstleister nur unzureichend umgesetzt werden. In diesem Beitrag wird ein App-Prototyp vorgestellt, der eine kundenorientierte Paketzustellung auf der „letzten Meile“ ermöglicht. Der Prototyp richtet sich an Endkonsument*innen und legt den Fokus auf Funktionalitäten und Usability. Im Beitrag werden auf Basis von Personas und User Experience-Research User Flows und Auszüge des Prototyps vorgestellt. Zudem werden die Ergebnisse des durchgeföhrten Usability Tests dargestellt. Durch die Open-Source Publikation des Prototyps wird Anbietern von Logistikdienstleistungen die Umsetzung kundenorientierter Logistikkonzepte erleichtert, da bisher am Markt keine geeigneten Lösungen existieren.

Keywords: E-Commerce, Paketzustellung, Letzte Meile, Softwareentwicklung, Prototyp, Open-Source

1 Einleitung und Motivation

Infolge des stetigen Anstiegs des Paketvolumens und den damit verbundenen Herausforderungen werden die bestehenden Belieferungskonzepte der Paketzustellung im urbanen Raum zunehmend in Frage gestellt. Verschiedene Studien belegen, dass nur jedes zweite Paket den/die Empfänger*in wie gewünscht an der Haustür erreicht [SeGö18]. So sind zwischen 2017 und 2019 die Anzahl der Beschwerden über Paketdienstleister bei der Verbraucherzentrale um 800 Prozent v.a. bezüglich der Haustürzustellung gestiegen [PeLi20] [Ve19]. Jedoch bleibt die Akzeptanz für alternative Belieferungsorte wie Paketautomaten und -shops oder neuartige Formen wie die Kofferraumzustellung hinter den Erwartungen zurück [Sp22][PKP17]. Die Kundschaft fordert flexiblere Paketzustellungen, sodass Lieferungen – angepasst an individuelle Alltagsplanungen – persönlich entgegengenommen werden können [He21]. Die Apps der großen Paketdienstleister und die dahinterliegenden

1 HTW Berlin, Treskowallee 8, 10318 Berlin, daniel.quiter@gmail.com

2 HTW Berlin, Treskowallee 8, 10318 Berlin, birte.malzahn@htw-berlin.de,
<https://orcid.org/0000-0001-5361-1794>

3 HTW Berlin, Ostendstraße 1, 12459 Berlin, maximilian.engelhardt@htw-berlin.de

4 HTW Berlin, Ostendstraße 1, 12459 Berlin, stephan.seeck@htw-berlin.de

logistischen Systeme verfügen jedoch nur unzureichend über die Möglichkeit, die Zustellung auf der letzten Meile durch den/die Empfänger*in zu steuern [Qu21]. Neben den Apps der etablierten Paketdienstleister, bei denen lediglich ein Ablageort oder ein Wunschliefertag angegeben werden kann, gibt es Versuche durch andere Unternehmen, den Empfänger*innen flexible Zustelloptionen zu bieten. So bietet das Unternehmen fairsenden eine Lösung, bei der Empfänger*innen für integrierte Online-Shops ein Zeitfenster innerhalb der kommenden drei Tage im Check-Out-Prozess oder in der bereitgestellten App buchen und bis zu 30 Minuten vor der Lieferung umbuchen können. Daneben gibt es anbieterunabhängige Paketshops, u.a. das Unternehmen Paketconcierge, zu denen sich Kund*innen Pakete aller Paketdienstleister liefern lassen können. Der Anbieter dropfriends ermöglicht, Pakete anbieterunabhängig durch Privatpersonen annehmen zu lassen. Eine Auslieferung und Zustellung findet bei den beiden letztgenannten Konzepten jedoch nicht statt. Eine kundenorientierte Alternative zur heutigen Paketzustellung wurde im Forschungsprojekt „Kundenorientierte Paketzustellung durch den Kiez-Boten“ erfolgreich erprobt: Endkund*innen konnten Pakete aller Versender und Paket-Dienste an ein Mikro-Hub (kleines Lager in der Stadt) liefern lassen, um sich diese dann im gewünschten Zeitfenster umweltfreundlich per Lastenrad nach Hause bringen zu lassen. Die Abstimmung der Wunschzeitfenster zwischen Zusteller*in und Empfänger*in wurde durch eine von einem weiteren Forschungsprojekt bereitgestellte App sichergestellt. Während die Kundenzufriedenheit bezüglich der Paketzustellung durch den Service deutlich gesteigert werden konnte, wurden eine unzureichende Usability sowie fehlende oder mangelhafte Funktionen an der eingesetzten App kritisiert, beispielsweise die fehlende Möglichkeit zur Umbuchung von Lieferzeitfenstern [SeEn21]. Die App bildet die wichtigste Schnittstelle zu den Kund*innen und ist daher elementar für einen erfolgreichen Rollout dieser kundenorientierten Alternative der Paketzustellung. Da, wie oben gezeigt, am Markt keine geeigneten Apps zur Verfügung stehen, welche Zeitfensterzustellung und gleichzeitig Bündelung von Paketen ermöglichen, wurde im Rahmen eines Folgeprojektes mit der Entwicklung einer Open-Source-Lösung zur Steuerung der Paketzustellung auf der letzten Meile begonnen. Im Folgenden werden theoretische Grundlagen, die Entwicklung und der Usability-Test des entsprechenden Prototyps beschrieben.

2 Grundlagen

2.1 Distributionslogistik und letzte Meile in der Zustellung von Paketen

„Die Distributionslogistik umfasst alle Aktivitäten, die in einem Zusammenhang mit der Belieferung des Kunden [...] stehen.“ [Pfl18]. Paketsendungen werden hierbei über mehrstufige Transportnetze verteilt. Im Vorlauf werden Sendungen durch die Paketdienste einem Verteilzentrum zugeführt. Im Hauptlauf werden sie gebündelt per Fernverkehr zu empfängernahen Verteilzentren transportiert. Im Nachlauf bzw. auf der „letzten Meile“, dem Tätigkeitsumfeld des vorliegenden Forschungsprojektes, werden Sendungen an geschäftliche und private Empfänger*innen verteilt [Th17]. Aufgrund der hohen Kosten für die „letzte Meile“ von bis zu 77 Prozent der gesamten Transportkosten [Br20] wird dieser Abschnitt in Forschung und Praxis besonders fokussiert betrachtet.

2.2 App-Entwicklung und Usability-Testing

Mobile Applikationen (Kurzform **App**) sind Softwareanwendungen, die als gekapselte Programme auf mobilen Endgeräten laufen [AiSc14]. **Prototypen** simulieren eine Anwendung, mit der u. a. die technische Umsetzbarkeit von Anforderungen sowie die Usability überprüft werden können [JaMe19]. Mit Hilfe einer **User Story Map** kann auf horizontaler Ebene der Flow von Anwender*innen durch das System dargestellt werden [WiMa17]. Gezeigt werden dabei die einzelnen Schritte, die Benutzer*innen im System durchlaufen müssen, um mit dem System eine bestimmte Aufgabe zu erfüllen [WiMa17].

Personas stellen wichtige unterschiedliche Typen von Anwender*innen und deren Anforderungen dar [Co14]. Sie können bei der Entwicklung von Apps dazu dienen, die unterschiedliche Interaktion und Navigation von Anwendertypen zu untersuchen [Ni19]. In **Fokusgruppen** arbeitet man mit potenziellen Nutzer*innen [JaMe19], um Erwartungen an eine neue Anwendung herauszuarbeiten. Repräsentative Nutzer*innen werden dazu in einem Raum versammelt [Co14].

Usability Testing wird dafür eingesetzt, die Eigenschaften einer Interaktion von Anwender*innen mit einem Produkt zu erfassen [Co14]. Testpersonen können zuvor definierten Personas zugeordnet werden [To18]. In den Versuchen wird untersucht, wie gut die Anwender*innen vorgegebene Standardaufgaben mit dem Produkt lösen können und welche Probleme sich dabei ergeben [Co14]. Bei der **Think-Aloud-Methode** werden sämtliche Gedanken durch Tester*innen laut ausgesprochen, um mögliche Probleme identifizieren zu können [Ba11]. Auf Basis von Usability Tests kann ein Prototyp iterativ verbessert werden, bis das finale Produkt entwickelt werden kann [Wo21].

3 Ergebnisse

3.1 Entwicklung des Prototyps

Auf Basis aktiver Kund*innen aus dem Forschungsprojekt wurden drei Personas entwickelt. Die erste Persona ist 36 Jahre, bestellt zweimal im Monat Pakete und legt Wert auf eine unkomplizierte und flexible Lieferung. Die zweite Persona ist 48 Jahre und bestellt mehrfach pro Woche. Sie ist selten zu Hause und probiert gerne neue Technik zur Entlastung des Alltags aus. Die dritte Persona ist 62 Jahre, bestellt einmal in der Woche und setzt sich aktiv für eine verbesserte Verkehrssituation im Kiez ein. Vor neuen Apps schreckt sie zurück, sollten sie nicht intuitiv funktionieren.

Um ein geeignetes UX-Design zu entwickeln, wurden zwei Fokusgruppen-Interviews und eine quantitative Online-Befragung durchgeführt. Dadurch wurde ersichtlich, dass essenzielle Funktionen wie eine einfache Registrierung sowie Buchung der Zeitfenster durch die vorhandene App bezüglich des User Interface gut gelöst, funktional jedoch nur unzureichend umgesetzt wurden. Des Weiteren konnten Verbesserungsvorschläge für die bisher verwendete App und Interesse an folgenden neuen Features identifiziert werden [SeEn21]:

- Tracking des Auslieferprozesses (z. B. auf Live-Karte)
- Umbuchung oder Stornierung von gebuchten Zeitfenstern
- Ankündigung des voraussichtlichen Lieferzeitpunkts innerhalb des Zeitfensters
- Übersichtliche Darstellung von zugestellten Lieferungen
- Buchungsmöglichkeit von Retouren



Abb. 1: User Flow (Quelle: Eigene Darstellung)

Aufbauend auf den Erkenntnissen wurde iterativ ein User-Flow entwickelt, der die Prozessabläufe in der App wiedergibt. In Abbildung 1 wird ein möglicher User Flow von der erstmaligen Registrierung bis zum Empfang einer Lieferung dargestellt. Abgebildet sind die Hauptfunktionen und deren Abfolgen innerhalb der App, mit denen die Anwender*innen sich registrieren, ein Onboarding durchlaufen, eine Sendung hinzufügen, ein Zeitfenster buchen und bezahlen sowie die Lieferung verfolgen und abändern können.

Auf Basis der User Flows wurden Mock-Ups erstellt, welche die Anordnung der Inhalte und Funktionen auf den benötigten Screens und das Grundgerüst des User Interfaces darstellen. Exemplarisch sind in Abbildung 2 die Screenshots des Prototyps der Registrierung, der Paketübersicht sowie der Lieferoptionen dargestellt. Der Prototyp wurde in mehreren Iterationen, in die sowohl Feedback von potenziellen Nutzer*innen sowie vom Projektteam einfloss, von einem UX-Designteam ausgearbeitet. In Vorbereitung für den Usability-Test wurden die gestalteten Screens entsprechend der User Flows zusammengefügt, sodass ein klickbarer Prototyp entstand. Im Folgenden werden die Funktionen des Prototyps näher erläutert.

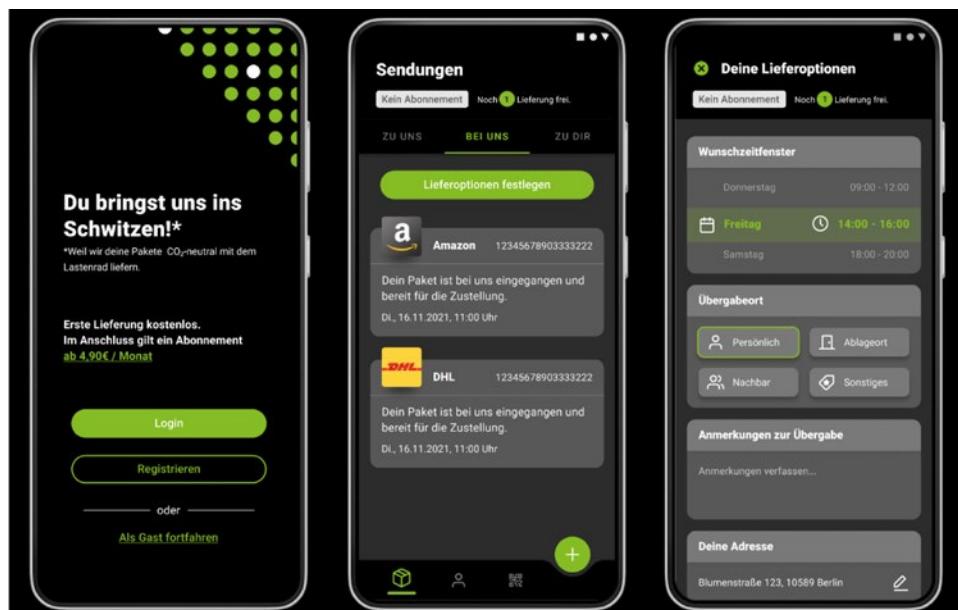


Abb. 2: Screenshots des Prototyps (Quelle: Eigene Darstellung)

3.2 Bausteine der Applikation

Registrierung, Login und Onboarding

Für Nutzer*innen ohne Registrierung besteht die Möglichkeit, einen Einblick in den Funktionsumfang der App zu erhalten: Sie können u.a. für bereits hinterlegte Dummy-Pakete Liefer-Buchungen vornehmen und sich über angebotene Abomodelle informieren. Bei der Registrierung sind von den Anwender*innen Name, Adresse, E-Mail-Adresse sowie ein Passwort anzugeben. Außerdem besteht die Möglichkeit, eine Telefonnummer und Lieferhinweise zu hinterlegen. Nach Abschluss der Registrierung können die Login-Daten für den Zugang zur App genutzt werden. Neuen Nutzer*innen wird ein Onboarding-Prozess zur Erklärung und Nutzung der Dienstleistung angezeigt. Für registrierte Nutzer*innen befindet sich eine Nutzungsanleitung in der Hauptansicht, solange noch kein Paket für die Zustellung vorhanden ist.

Sendungsübersicht

Die Sendungsübersicht stellt die Hauptansicht der App dar und besteht aus drei Unteransichten, die den jeweiligen Paketstatus abbilden. Das Fenster ZU UNS zeigt alle Pakete an, die sich in der Zustellung zum jeweiligen Mikro-Hub befinden. Mittels Sendungsnummer können über einen Button Pakete hinzugefügt werden. Die Übersicht BEI UNS bündelt alle Pakete, die sich für den/die jeweilige Nutzer*in im Hub befinden und bereit für die Zustellung sind. Das Fenster ZU DIR zeigt die aktuell gebuchte Zustellung sowie die sich darin befindlichen Pakete an. Bis zu zwei Stunden vor Zustellung kann hier die getätigte Buchung angepasst werden.

Zeitfensterbuchung

Nutzer*innen können in der App Zustelltag und Zeitfenster festlegen sowie weitere Angaben zur Zustellung machen. Eine kurzfristige Änderung der Lieferadresse ist möglich, solange sich diese im Zustellgebiet befindet. Gesonderte Wünsche oder Hinweise an die Zusteller*innen können über ein Freifeld kommunizieren werden, die den Zusteller*innen bei der Übergabe angezeigt werden. Die gebuchten Pakete werden übersichtlich nach Online-Shop gruppiert aufgelistet.

Abomodelle und Zahlungsarten

Nutzer*innen können in der App zwischen drei Abomodellen wählen, die sich hinsichtlich der möglichen Lagerdauer im Mikro-Hub, inkludierter Lieferungen, App-Features wie wöchentlich fixierte Liefertermine, zusätzlich angebotenen Lieferzeitfenstern und -tagen sowie dem Preis unterscheiden. Die Bezahlung ist über Kreditkarte, SEPA-Lastschriftmandat sowie Google- und Apple-Pay möglich.

3.3 Usability-Test mit Endanwender*innen

Der klickbare Prototyp wurde einem moderierten Usability-Test mit der Think-Aloud-Methode unterzogen. Die Rekrutierung der Stichprobe erfolgte unter Berücksichtigung der zuvor entwickelten Personas. Die Proband*innen setzten sich aus elf Kund*innen zusammen, die die bisherige App bereits genutzt hatten, sowie aus neun Personen, die bisher nicht mit der Thematik vertraut sind. Die Testszenarien bestanden aus den vier wichtigsten Prozessen der App: Registrierung/Login, Ermittlung der Lieferadresse des Mikro-Hubs, Buchung eines Wunschzeitfensters sowie Bezahlung.

3.4 Ergebnisse des Usability-Tests

Die Usability Tests legten zwei zentrale Probleme bzgl. „Registrierung/Login“ und „Ermittlung der Lieferadresse des Mikro-Depots“ sowie kleinere Anpassungsbedarfe offen. Die im ursprünglichen Design der Registrierung vorhandene Karte mit den Hub-Standorten wurde von 14 der 20 Proband*innen bemängelt, da keine freie Auswahl der Hubs vorgesehen war. Der Prozess wurde dahingehend angepasst, dass die Registrierung lediglich auf Eingabefelder beschränkt und die Zuordnung von Zustelladresse und Hub im Backend vorgenommen wird. Zudem hatten 65% der Proband*innen Probleme, die zu Beginn in der Top-Bar angezeigte Lieferadresse und den Nutzen hinter dem für die Übergabe benötigten QR-Code zu verstehen. Daher wurde eine Kundenkarte als gesonderter Menüpunkt erstellt, die die persönliche Lieferadresse und den QR-Code sowie jeweils eine Beschreibung enthält. Abbildung 3 zeigt die Screens der Registrierung und der Paketübersicht/Kundenkarte vor und nach dem Usability-Test.

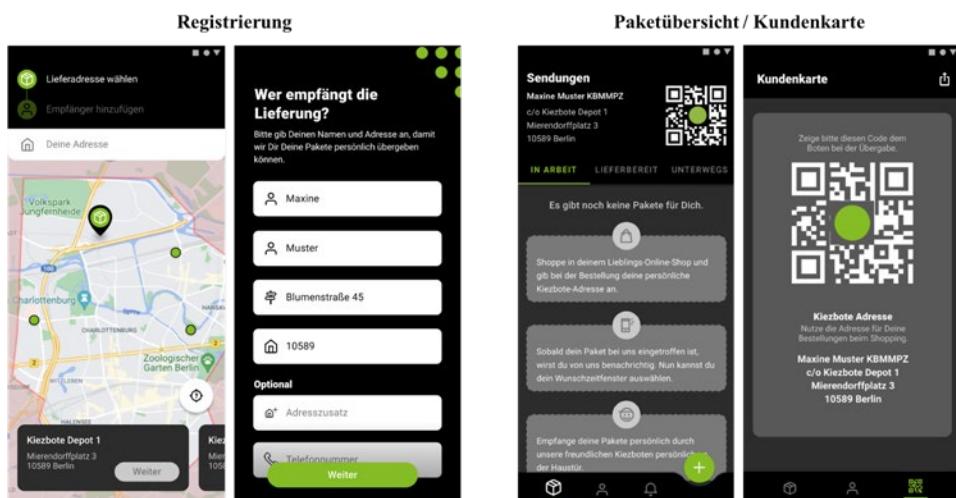


Abbildung 3: Registrierung und Paketübersicht/Kundenkarte (initial und nach Anpassung, v.l.n.r.)
(Quelle: Eigene Darstellung)

Weitere kleinere Änderungen wurden vorgenommen, u.a. einheitliche Wortwahl, verbesserte Sichtbarkeit von Preisen für Abonnements, verbesserte Beschreibung der angebotenen Abonnements und die Möglichkeit, den Account zu löschen.

4 Zusammenfassung und Ausblick

In diesem Beitrag wurde ein App-Prototyp vorgestellt, der eine kundenorientierte Paketzustellung auf der „letzten Meile“ ermöglicht. Die Funktionalitäten der App wurden unter Berücksichtigung von Usability-Aspekten entwickelt. Am Markt existieren bisher keine vergleichbaren App-Lösungen. Die Open-Source Publikation ermöglicht Anbietern von entsprechenden Logistikdienstleistungen den Zugang zu dem entwickelten Prototyp. Desse n Weiterentwicklung und Nutzung als betriebsfähige App kann die wirtschaftliche Umsetzung des Logistikkonzepts unterstützen, da ein höheres Service-Level, bessere Kundenbindung und effizientere Prozesse in der Kommunikation mit den Kund*innen erreicht werden können. Sofern nachhaltige Transportmittel wie z. B. Lastenfahrräder oder E-Fahrzeuge eingesetzt werden, sind auch signifikant positive Umweltauswirkungen zu erwarten. Die Umsetzung des Prototyps in eine funktionsfähige App sowie deren Einsatz im betrieblichen Umfeld ist geplant.

Literaturverzeichnis

- [AiSc14] Aichele, Christian; Schönberger, Marius: App4U: Mehrwerte durch Apps im B2B und B2C. Springer Fachmedien Wiesbaden, 2014.
- [Ba11] Barnum, Carol M.: Usability Testing Essentials: Ready, Set... Test!. Burlington, MA, Morgan Kaufmann Publishers, 2011.
- [Br20] Brabänder, Christian: Die Letzte Meile - Definition, Prozess, Kostenrechnung und Gestaltungsfelder. Springer Gabler, Wiesbaden, 2020.
- [Co14] Cooper, Alan et al.: About Face: The Essentials of Interaction Design. John Wiley and Sons, Indianapolis, 2014.
- [He21] Hermes Germany GmbH: Für mehr als jede*n zweite*n Deutsche*n tragen Paketdienstleister positiv zur Lebensqualität bei, <https://bit.ly/3NwuiY4>, Stand: 28.02.2022.
- [JaMe19] Jacobsen, Jens; Meyer, Lorena: Praxisbuch Usability & UX. Rheinwerk Verlag, 2019.
- [Ni19] Nielsen, Lene: Personas - User Focused Design. Springer London, 2019.
- [PeLi20] Pecanic, Ana; Lippegaus, Olaf: Klagen über nicht zugestellte oder beschädigte Pakete nehmen zu, <https://bit.ly/3IORwp1>, Stand: 28.02.2022.
- [PfI18] Pfohl, Hans-Christian: Logistiksysteme. Springer Vieweg, 2018.
- [PKP17] Prümm, Dietmar; Kauschke, Peter; Peiseler, Hanna: Aufbruch auf der letzten Meile - Neue Wege für die städtische Logistik, <https://pwc.to/3uEQ3MI>, Stand: 28.02.2022.
- [Qu21] Quiter, Daniel; Seeck, Stephan; Engelhardt, Maximilian; Malzahn, Birte: Ermittlung von Kundenanforderungen an eine app-basierte Paketzustellung im urbanen Raum mittels Conjoint-Analyse. In: Anwendungen und Konzepte der Wirtschaftsinformatik, Nr. 13, 2021.
- [SeEn21] Seeck, Stephan; Engelhardt, Maximilian: New Opportunities for Smart Urban Logistics - Results of a Field Study. In: Proceedings of the Hamburg International Conference of Logistics (HICL), 2021.
- [SeGö18] Seeck, Stephan; Göhr, Marco: Paketzustellung und Kundenservice – ein Widerspruch in sich? In: Digitales trifft Reales: 35. Deutscher Logistik-Kongress, DVV Media Group, 2018.
- [Sp22] Spectos GmbH: E-Commerce Monitor 2022, <https://bit.ly/3K1aCt8>, Stand: 29.03.2022.
- [To18] Tomlin, W. Craig: UX Optimization: Combining Behavioral UX and Usability Testing Data to Optimize Websites. Apress, 2018.
- [Th17] Thaller, Carina et al.: KEP-Verkehr in urbanen Räumen. In (Proff, H.; Fojcik, T. M., Hrsg.): Innovative Produkte und Dienstleistungen in der Mobilität. Springer Gabler, Wiesbaden, S. 443–458, 2017.
- [Ve19] Verbraucherzentrale Nordrhein-Westfalen e.V.: Statistik zu Paket-Beschwerden: Juli bis Einschließlich Dezember 2018, <https://bit.ly/3uF2p7s>, Stand: 29.03.2022.
- [WiMa17] Wirdemann, Ralf; Mainusch, Johannes: Scrum mit User Stories. 3., erweiterte Auflage, Hanser, München, 2017.
- [Wo21] Wood, Brian: Figma Essential Training: The Basics, <https://bit.ly/3DkmRP6>, Stand: 21.02.2022.

Das Process-Simulation.Center

Prototyp eines Integrierten Managementsystems für Lehre, Forschung und Praxis

Carlo Simon¹, Lara Zakfeld¹ und Stefan Haag¹

Abstract: Um die Aufbau- und Ablauforganisation einer Unternehmung, etwa im Rahmen eines Managementsystems, ganzheitlich integriert und eventuell sogar simulierbar abzubilden, müssen Modelle übergreifenden Regeln folgen und zueinander kompatibel sein. Das hier vorgestellte *Process-Simulation.Center* ist ein prototypisches Integriertes Managementsystem zur unternehmensweiten Dokumentation und Simulation von Prozessen und Verantwortlichkeiten mittels Petri-Netzen, Prozesslandkarten und Organigrammen. Hierzu speichert es die Modelle in einer zentralen Datenbank, der *P-S.C-Cloud*, und erlaubt allen Mitgliedern der Organisation den Zugriff auf die Modelle gemäß eines umfassenden Rechtekonzepts. Um regulatorische Anforderungen an die Modelle organisatorisch sicherzustellen, durchlaufen sie einen Freigabeprozess. Hierdurch unterscheidet sich das *Process-Simulation.Center* von anderen Hochschulprojekten zur Entwicklung von Software für Modellierung und Simulation. Dieser Beitrag stellt den Prototyp vor und wie er in Lehre, Forschung und Praxis eingesetzt wird.

Keywords: Prototyp, Petri-Netz, Prozesslandkarte, Organigramm, Integriertes Managementsystem

1 Einleitung

Zentraler Bestandteil des Wirtschaftsinformatik-Studiums ist es, Unternehmensprozesse, Daten und Organisationsstrukturen multiperspektivisch zu modellieren. Mit Blick auf die Prozesse haben insbesondere die Methoden BPMN, EPK und zum Teil Flussdiagramme eine hohe Relevanz [SHZ21]. Eine besondere Rolle spielen zudem Petri-Netzen aufgrund ihrer formalen Semantik, die Analysen und komplexe Simulationen ermöglicht.

Eine anschauliche Lehre zum Thema Petri-Netze setzt die Verwendung von Software zur Modellierung, Simulation und Analyse voraus. Eine in [Si18] veröffentlichte Auswertung von 91 Petri-Netz-Werkzeugen zeigt aber, dass diese an Hochschulen entwickelten Tools weder zeitgemäß weiterentwickelt noch gewartet werden. In [SHZ21] wurden darüber hinaus auch Defizite in (zum Teil kommerzieller) Prozessmanagement-Software bzgl. ressourcenabhängiger Simulation aufgezeigt.

¹ Hochschule Worms, Fachbereich Informatik, Erenburgerstraße 19, 67549 Worms
{simon,zakfeld,haag}@hs-worms.de

Diese Situation führte zur Entwicklung des *Process-Simulation.Center (P-S.C)*, das unter <https://process-simulation.center/> aufgerufen werden kann und von Anwendern aus dem akademischen Umfeld kostenfrei genutzt werden darf. Anderen Usern steht ein kostenfreier Testzugang für drei Monate zur Verfügung. In Anlehnung an [Be21] handelt es sich beim *P-S.C* derzeit um einen Prototyp, der wesentliche Elemente und Funktionen eines Hochintegrierten Managementsystems, insbesondere prozessübergreifende und ressourcenabhängige Simulationen, überprüft und Nutzer-Interaktionen zu testen erlaubt.

Die erstellten Modelle werden in der *P-S.C-Cloud* gespeichert, in der User in Mandanten gemeinsam Modelle bearbeiten können. Konkret wurden die folgenden Funktionalitäten realisiert, die im Beitrag näher vorgestellt werden:

1. Es lassen sich Mandanten für unterschiedliche Zielgruppen definieren, wodurch Produktvarianten realisiert werden.
2. Innerhalb eines Mandanten gibt es verschiedene User-Berechtigungen.
3. Durch die unterschiedlichen Berechtigungskonzepte kann ein Freigabeworkflow implementiert werden, wodurch Modelle einen Lebenszyklus durchlaufen. Eine Historie ermöglicht das Nachverfolgen von Änderungen an den Modellen.
4. Zudem lassen sich User in Gruppen zusammenzufassen. Über Modellrechte lässt sich dann zusätzlich der Zugriff von Usern auf die Modelle differenziert steuern.

Diese Konzepte erlauben die Verwendung des *P-S.C* als Integriertes Managementsystem. Hierdurch können neue Lehr- und Lerninhalte erlebbar gemacht werden. Die konkrete technische und didaktische Umsetzung wird in den folgenden Abschnitten thematisiert, wodurch Kolleginnen und Kollegen eingeladen werden sollen, das *P-S.C* ebenfalls in der Lehre und in eigenen Projekten zu verwenden.

Der restliche Beitrag gliedert sich wie folgt: Der zweite Abschnitt erklärt den bisherigen Entwicklungstand des *P-S.C* bis zur Einführung der *P-S.C-Cloud*. Danach wird das neue Mandanten-Konzept vorgestellt und die Möglichkeiten, User innerhalb eines Mandanten zu organisieren. Anschließend wird dargestellt, wie der neue Freigabeprozess über ein Berechtigungskonzept realisiert wird. Die Arbeit endet mit einem Fazit.

2 Bisheriger Entwicklungsstand des P-S.C

Das *P-S.C* ist ein Werkzeug zur Modellierung und Simulation der Aufbau- und Ablauforganisation von Unternehmen und wird seit mehreren Jahren an der Hochschule Worms in Lehre, Forschung und Transferprojekten mit der Industrie prototypisch eingesetzt. Der Entwicklungsprozess kombiniert Empfehlungen der Design Science Research Guidelines [He04] und des evolutionären Prototyping [Fl84].

Das Tool lässt sich in Kurzform durch folgende Merkmale beschreiben:

1. Das *P-S.C* ist ein webbasiertes Werkzeug zur Modellierung und Simulation von einfachen und höheren Petri-Netzen mit Integration von Swimlanes, zur Erstellung von Organigrammen sowie der Kombination der Modelle mit Hilfe von Prozesslandkarten (vgl. [SHZ22]).
2. Über eine CSV-Schnittstelle können Realdaten für höhere Petri-Netze importiert und Simulationsergebnisse exportiert werden (vgl. [SH20a]).
3. Teil des *P-S.C* ist ein Tutorial mit mehr als 80 erklärten Beispielmodellen. Neben technischen Grundlagen des *P-S.C* vermittelt es anhand von drei Fallstudien, wie man unternehmensweite Prozessmodelle erstellt, wie Engpässe gemäß der Theory of Constraints identifiziert und wie Push- und Pull-Strategien in Produktionsprozessen simuliert werden können.
4. Wird das *P-S.C* lokal auf einem Raspberry Pi installiert, kann der Simulator über die GPIO-Schnittstelle zur Anlagensteuerung verwendet werden (vgl. [SH20b]).

Zwischenzeitlich haben mehr als 300 Studierende mit einem eigenen Account im *P-S.C* gelernt und gearbeitet. Ihre Erfahrungen sowie solche, die durch Transferprojekte mit Unternehmen gesammelt werden, sind wertvolle Quellen für die Anforderungen an den Prototyp. Um die Sichtbarkeit dieser Forschungsarbeiten zu erhöhen, wurde an der Hochschule Worms die *Gruppe für Angewandte Prozesssimulation (GAPS)* gegründet.

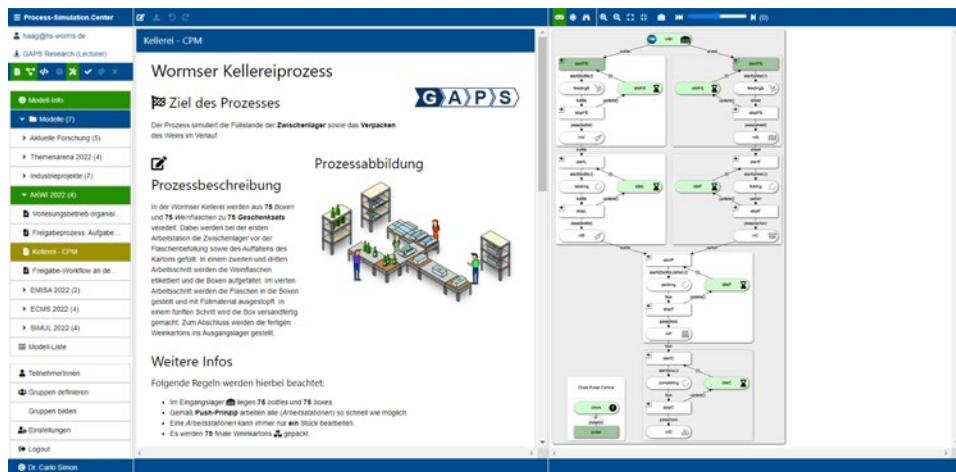


Abb. 1: P-S.C-Screenshot zum Clock Pulse Model eines Kellereiprozesses [SHZ22]

Abbildung 1 zeigt einen Screenshot des P-S.C mit dem (einklappbaren) Menü an der linken Seite. Je Modell gibt es drei Ansichten: eine Erklärung zum Modell, das Modell als Graph sowie (hier nicht gezeigt) eine Spezifikation mittels einer eigens entwickelten Spezifikationssprache. Dieses Bedienkonzept, das sich von üblichen Zeichenoberflächen unterscheidet, erlaubt das Entwickeln komplexer Szenarien, sowie durch die Verwendung automatischer Layoutalgorithmen eine hohe visuelle Konsistenz.

Mit der nun neu entwickelten *P-S.C-Cloud* können neben der eigentlichen Modellierung und Simulation auch die komplexen organisatorischen und fachlichen Maßnahmen gezeigt werden, die bei der Einführung eines Integrierten Managementsystems zu beachten sind. In Analogie zum Freigabeprozess für Modelle in Unternehmen erhalten die Studierenden ihr Feedback zu Übungsaufgaben über das *P-S.C*. Dabei erbitten die Studierenden für Lösungen, bei denen sie glauben, sie richtig modelliert zu haben, um Freigabe durch die Tutorin oder den Tutor. Hierzu ist es notwendig, für verschiedene Vorlesungen Mandanten einzurichten, die Studierenden in Gruppen einzuteilen und den beteiligten Akteuren ein Durchlaufen des Freigabeprozesses zu ermöglichen.

3 Mandanten, Rollen, Gruppen

Die *P-S.C-Cloud* ist so eingerichtet, dass sie gleichzeitig unterschiedliche Mandanten bedienen kann. In einem mehrmandantenfähigen System können die User der verschiedenen Mandanten ohne wechselseitigen Einblick in fremde Daten, Benutzer oder Einstellungen so arbeiten, als würde ihnen die Umgebung exklusiv zur Verfügung stehen (in Anlehnung an [HMN19]).

Das *P-S.C* unterstützt verschiedene Mandantentypen, wodurch sich perspektivisch die Produktkategorien des *P-S.C* ergeben. Ein besonderer (interner) Mandant ist das *Tutorial*.

Mandant *Public* erlaubt es Usern, die Möglichkeiten des *P-S.C* bzgl. Modellierung und Simulation zu erleben. Testuser erhalten in diesem Mandanten einen drei Monate gültigen *Guest-Account* mit eigenen, von anderen Usern getrennten Modellen.

Weitere Mandanten sind die Typen *Academic*, *Consulting* und *Company*. In diesen Mandanten können mehrere User zusammenarbeiten. Über Nutzerrollen lassen sich hierbei Rechte und Pflichten vergeben, was beispielhaft für *Academic* gezeigt wird.

In einem *Academic*-Mandanten lassen sich die Rollen *Admin*, *Lecturer* und *Student* unterscheiden: *Admin*-User können andere User aufnehmen und deren Rolle festlegen. *Lecturer* definieren Lerngruppen und weisen diesen User zu. Außerdem legen sie Modelle an und können so etwa in deren Beschreibung eine Modellierungsaufgabe formulieren. Auch dürfen *Lecturer* die Rechte an den Modellen und damit deren Sichtbarkeit innerhalb des Mandanten festlegen. Schließlich spielen sie eine entscheidende Rolle beim Freigabeprozess. *Student*-User arbeiten in Gruppen zusammen, können Lösungen einreichen, indem sie um Freigabe ihrer Modelle bitten, oder als Prozesseigner innerhalb ihrer Gruppe besondere Verantwortung übernehmen.

Die Abläufe rund um die Einschreibung und die Organisation der Lerngruppen zeigt Abbildung 2. Die Stellen repräsentieren hierbei die anfallenden Informationen: *Guest im Mandant Public* die User, die sich über die Webseite zunächst als *Guest*-User registriert haben; *Student & Lecturer im Mandanten* die *Guest*-User, die in den Mandanten aufgenommen wurden; *Gruppe im Mandanten* die von *Lecturer*-Usern gebildeten Gruppen; *Lerngruppen im Mandanten* die Gruppen zusammen mit den Studierenden.

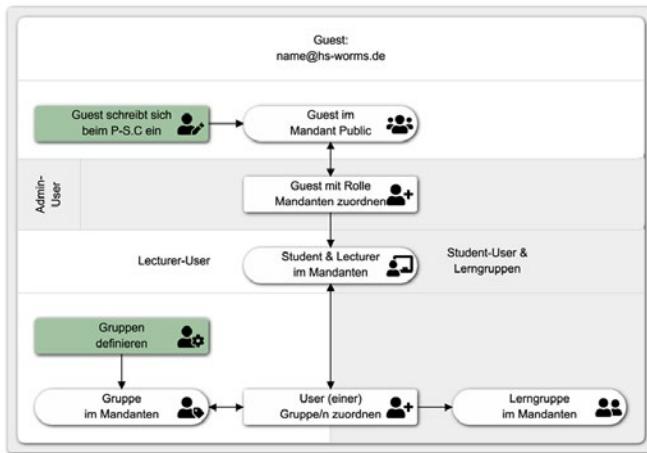


Abb. 2: Petri-Netz-Modell: In Mandanten einschreiben und Lerngruppen organisieren

4 Modelle, Rechte und Freigabeprozess

Modelle sind einem Mandanten zugeordnet, in dem die Rechte an den Modellen nach den sich aus den User-Rollen ergebenden Interessen differenziert sind. Dabei wird zwischen folgenden Rechten unterschieden: Das Recht, (1) ... ein Modell ohne Spezifikation zu sehen. (2) ... ein Modell zu kommentieren. (3) ... ein Modell mit Spezifikation zu sehen. (4) ... ein Modell zu simulieren. (5) ... die Arbeitsversion eines Modells zu verändern. (6) ... um Freigabe eines Modells zu bitten. Im aktuellen Prototyp ist die Möglichkeit, Modelle zu kommentieren, bislang noch nicht realisiert.

User können in folgenden Beziehungen zu Modellen stehen: Prozesseigner sind verantwortlich für einen spezifischen Prozess; User können aber auch einer Gruppe angehören, der ein Modell zugeordnet ist; schließlich kann es auch sein, dass User einfach aufgrund ihrer Zugehörigkeit zum Mandanten ein Interesse am Modell haben. Für alle drei Arten der Zugehörigkeit können die genannten Rechte vergeben werden, wobei für einen User in Bezug auf ein Modell das allgemeinste gültige Recht wirkt. Schließlich können User als Executive benannt sein, wodurch sie die Möglichkeit erhalten, ein zur Freigabe anstehendes Modell innerhalb des Mandanten zu veröffentlichen.

Im Mandant *Academic* dürfen *Lecturer* neue Modelle erstellen, sie Usern als Prozesseigner oder auch Lerngruppen zuordnen und zudem die Rechte je Interessengruppe vergeben. Die Modelle werden hierbei zunächst als *Arbeitsversion* erstellt. Haben die Studierenden die Modellierungsaufgabe gelöst, stellen sie diese *Zur Freigabe* ein. Diese kann widerrufen oder ihr kann stattgegeben werden. Im letztgenannten Fall ist die Version dann im Zustand *Freigegeben*. Von freigegebenen Modellen können aber auch neue Arbeitsversionen

abgeleitet werden. Durch neuerliche Freigaben werden frühere Versionen in den Zustand *Archiviert* überführt.

5 Fazit

Das *P-S.C* und die neu entwickelte *P-S.C-Cloud* führen Studierende ganzheitlich an das Thema Prozessmanagement heran, indem sie Anforderungen und Arbeitsweisen eines Integrierten Managementsystems verdeutlichen. Dies führt dazu, dass immer mehr Studierende auch für Abschlussarbeiten auf das *P-S.C* zugreifen, was ebenfalls zu positiven Lerneffekten bzgl. neuartiger Modellierungstechniken führt. So verbindet das *P-S.C* Lehre, Forschung und Praxis zu einer Einheit.

Literaturverzeichnis

- [Be21] Bendel, O.: Prototyp. Gabler Wirtschaftslexikon. Zuletzt aufgerufen am 10.7.2022, <https://wirtschaftslexikon.gabler.de/definition/prototyp-122469/version-384553>.
- [Fl84] Floyd, C.: A Systematic Look at Prototyping. In (Budde, R.; Kuhlenkamp, K.; Mathiasen, L.; Züllighoven, H., Hrsg.): Approaches to Prototyping, Namur (BEL) 1984. Springer, Berlin, Heidelberg, S. 1-18, 1984.
- [He04] Hevner, A.R.; March, S.T.; Park, J.; Ram, S.: Design Science in Information Systems Research. MIS Quarterly 28/1, S.75-105, 2004.
- [HMN19] Hansen, H.R.; Mendling, J.; Neumann, G.: Wirtschaftsinformatik. De Gruyter Oldenbourg, Berlin, 2019.
- [Si18] Simon, C.: Web-Based Simulation of Production Schedules with High-Level Petri Nets. In (Nolle, L.; Burger, A.; Tholen, C.; Werner, J., Hrsg.): ECMS 2018: 32th International ECMS Conference on Modelling and Simulation, Wilhelmshaven, S. 275-281, 2018.
- [SH20a] Simon, C.; Haag, S.: Simulatable Reference Models to Transform Enterprises for the Digital Age. In (Steglich, M.; Müller, C.; Neumann, G.): ECMS 2020: 34th International ECMS Conference on Modelling and Simulation, Wildau, S. 294-300, 2020.
- [SH20b] Simon, C.; Haag, S.: Digitale Zwillinge modellieren und verstehen: Eine Fallstudie zum problembasierten und forschenden Lernen. In (Michael, J.; Bork, D.; Rehse, J.-R.; Striewe, M.; Ullrich, M., Hrsg.): Joint Proceedings of Modellierung 2020, Wien (AUT) 2020. CEUR-WS, Aachen, S. 101-112, 2020.
- [SHZ21] Simon, C.; Haag, S.; Zakfeld, L.: Anforderungen an hochintegrierte Management-systeme. In (Nees, F.; Stengel, I.; Meister, V.G.; Barton, T.; Herrmann, F.; Müller, C.; Wolf, M., Hrsg.): Angewandte Forschung in der Wirtschaftsinformatik 2021: Tagungsband zur 34. AKWI-Jahrestagung, Karlsruhe 2021. Mana-Buch, S. 56-65, 2021.
- [SHZ22] Simon, C.; Haag, S.; Zakfeld, L.: Stratification of Timed Petri Nets at the Example of a Production Process. In: (Hameed, I.; Hasan, A.; Alaliyat, S. A.-A., Hrsg.): ECMS 2022: 36th International ECMS Conference on Modelling and Simulation, Ålesund (NOR), S. 128-134, 2022.

Studentische Beiträge

Evaluierung des Dense Passage Retrievals zur Dokumentensuche in Behörden im Vergleich zu BM25

Anket Mehra¹, Malte Prieß², Tom Vincent Peters³

Abstract: Die Suche nach Dokumenten ist eine der Aufgaben, die Maschinen am häufigsten ausführen. Täglich genutzte Suchmaschinen suchen nach Dokumenten, um das Informationsbedürfnis eines Anfragenden zu erfüllen. Aktuelle Forschungsergebnisse zeigen, dass der Bewertungsansatz des Dense Passage Retrieval (DPR) – welcher die zu suchenden Dokumente mittels tiefer neuronaler Netze in Dense-Vektoren umwandelt – in englischer ebenso wie in deutscher Sprache bessere Ergebnisse liefert als aktuell implementierte Algorithmen wie der BM25-Algorithmus. Die Dokumentensuche in deutschen Behörden stellt eine besondere Aufgabe dar, da in diesen komplexe behördenspezifische Fachausdrücke verwendet werden. Im Rahmen dieser Arbeit wurde für die Dokumentensuche zu Corona-Regelungen in der öffentlichen Verwaltung ein DPR mit einem öffentlichen deutschsprachigen Datensatz trainiert und mit dem BM25-Algorithmus von Elasticsearch anhand der Relevanz als Gütemaß verglichen. Die Ergebnisse zeigen, dass das DPR mehr Dokumente mit gesuchten Informationen findet, Synonyme in Suchanfragen berücksichtigen kann und damit das Potenzial zur Verbesserung der Dokumentensuche in Behördenanwendungen hat.

Keywords: Dense Passage Retrieval, BM25, Natürliche Sprachverarbeitung, Künstliche Intelligenz, Neuronale Netze, Information Retrieval, Dokumentensuche

1 Einleitung

Die Dokumentensuche ist eine wichtige Aufgabe in der natürlichen Sprachverarbeitung. Um Inhalte maschinell gezielt und effizient zu finden, werden schon seit den 1970er Jahren an Ansätzen für die Dokumentensuche geforscht [Sp72].

Viele kommerziell eingesetzte Suchmaschinen basieren auf der Konvertierung von Suchanfragen sowie der Dokumente in Sparse-Vektoren, die miteinander verglichen werden [RZ09]. Aktuelle Forschungsergebnisse zeigen, dass die Ergebnisse einer Dokumentensuche eine höhere Genauigkeit und Sensitivität besitzen, wenn die Nutzereingabe und die zu suchenden Dokumente statt in Sparse- in Dense-Vektoren umgewandelt und mit einem

¹ Duale Hochschule Schleswig-Holstein, Studium der Wirtschaftsinformatik, Hans-Detlev-Prien-Straße 10, 24106 Kiel, anket.mehra@dataport.de

² Duale Hochschule Schleswig-Holstein, Professur im Fachbereich Wirtschaftsinformatik, Hans-Detlev-Prien-Straße 10, 24106 Kiel, malte.priess@dhsh.de

³ Dataport AöR, Machine Learning Engineer, Altenholzer Straße 10-14, 24161 Altenholz, tomvincent.peters@dataport.de

Proximitätsmaß nach der Ähnlichkeit gemessen werden [Ka20; MRP21]. Die Dense-Vektoren werden mit Hilfe von tiefen neuronalen Netzen erstellt, die auf der BERT-Architektur basieren [De18]. Diese sind hier die Dense Passage Retriever (DPR). Die Ergebnisse von [Ka20] sowie [MRP21] zeigen, dass das DPR in englischer sowie deutscher Sprache bessere Ergebnisse als aktuell implementierte Algorithmen wie der BM25-Algorithmus erzeugt.

Die Dokumentensuche in der öffentlichen Verwaltung stellt eine besondere Aufgabe dar. Der Suchalgorithmus muss Wörter präzise generalisieren können. Dies ist in der deutschen Sprache schwieriger als in der englischen Sprache. Die Gründe dafür sind das Gendern sowie der Kasus der deutschen Sprache, die eine wichtigere Rolle als im Englischen besitzen [Ja00]. Zudem besitzen Behördendokumente oftmals aus mehreren Nomina bestehende Fachausdrücke, die in Queries von Nutzenden umschrieben werden und daher die Suche erschweren können. Ein bekanntes Beispiel ist das mittlerweile aufgehobene Rindfleischetikettierungsüberwachungsaufgabenübertragungsgesetz (RfLEttÜAÜG) von 1999 [In00]. Dementgegen ist ein aktuelles Beispiel das aktuell im Bundestag zur Debatte stehende Bundeswehrbeschaffungsbeschleunigungsgesetz [Ta22]. Mit einem besseren Suchalgorithmus finden Benutzende relevante Dokumente in kürzerer Zeit, was zu einer Zeiter sparnis führt. Suchalgorithmen, welche Synonyme erkennen, können zudem semantisch zusammenhängende Dokumente finden, welche im Laufe der Zeit beispielsweise durch Anglicizismen andere Wörter benutzen.

In dieser Arbeit wird ein Dense Passage Retriever mit Open Source Quellen entwickelt und trainiert. Die Ergebnisse des DPR sowie des BM25-Algorithmus werden hinsichtlich ihrer Relevanz in den ersten 10 und 20 Dokumenten verglichen. Die Relevanz ist hier das Verhältnis zwischen Dokumenten mit der Antwort auf eine Suchanfrage und allen Dokumenten, die für eine Suchanfrage gefunden werden.

Im Folgenden wird im Kapitel 2 die grundsätzliche Problematik in der Dokumentensuche skizziert und der BM25-Algorithmus sowie das DPR erläutert. In Kapitel 3 und 4 wird eine beispielhafte Implementierung eines Dense Passage Retriever mit der Python-Bibliothek Farm-Haystack beschrieben sowie die Relevanz der Ergebnisse beider Algorithmen an hand einer Dokumentensuche zu Corona-Regelungen in einer Behörde verglichen. Kapitel 5 schließt mit einem Fazit und Ausblick ab.

2 Suche mit BM25 und DPR

2.1 Problemstellung

Ausgangspunkt der Dokumentensuche ist die Beantwortung einer Query q mit einer Menge an Dokumenten D . Die Dokumente sind nach einem Bewertungswert eingestuft, der den Platz eines Dokumentes in den Suchergebnissen definiert. Ein hoher Wert lässt ein Dokument früher erscheinen.

Die Inhalte der gefundenen Dokumente besitzen bestenfalls die Informationen zur Beantwortung einer Query. Gesucht wird in allen Dokumenten eines Datensatzes. Alle Begriffe aller Dokumente bilden den Textkorpus C . Die Anzahl der Begriffe von C (im Folgenden n), stellt die Länge der Sparse-Vektoren dar (vgl. (7)). Mathematisch formuliert, kann die Dokumentensuche als eine Funktion S betrachtet werden, welche als Parameter eine Query q sowie den Textkorpus C übermittelt bekommt. Als Ausgabe wird eine auch mitgegebene Anzahl n an Dokumenten D zurückgegeben.

$$S(q, C, n) = D \quad (1)$$

wobei

$$D := \{d_1; d_2; \dots; d_n\} \quad (2)$$

Die Berechnung des Bewertungswertes unterscheidet sich zwischen den angewandten Algorithmen für die Dokumentensuche.

2.2 TF-IDF und BM25

Ein Ansatz zum Suchen von Dokumenten ist das Term-Frequency-Inverse-Document Frequency Maß, kurz TF-IDF [Sp72]. Das Maß wird verwendet, um eine Menge von Dokumenten D anhand einer Query q zu bewerten, die aus qn Termen besteht [LZ11]:

$$TFIDF(d_i, q) = \sum_{j=1}^{qn} idf(t_j) * tf(t_j, d_i) \quad (3)$$

wobei

$$idf(t_j) = \log \frac{D}{n_j} \quad (4)$$

Hierbei wird q in einzelne Terme t_j aufgeteilt [Ro04]. Die Definition eines Terms ist dabei nicht festgelegt. Es können einzelne Wörter oder Gruppen von Wörtern sein [Ro04]. Für die Berechnung des TF-IDF Maßes wird durch die einzelnen Dokumente d_i iteriert und jeweils für jeden Term t_j der Query q die Häufigkeit tf im Dokument d_i mit der inversen Dokumentenhäufigkeit idf multipliziert [Ro04]. Die inverse Dokumentenhäufigkeit ist der Logarithmus zwischen der Gesamtanzahl an Dokumenten D sowie der Anzahl an Dokumenten n_j , in welchen der Term t_j mindestens einmal enthalten ist [Ro04].

Die somit berechneten TF-IDF Werte für alle Dokumente d_i werden im Folgenden in absteigender Reihenfolge sortiert. Das Dokument mit dem höchsten Wert beinhaltet mit hoher Wahrscheinlichkeit die relevanten Wörter der Query q und bietet eine mögliche Antwort auf die gestellte Query. Relevante Wörter eines Dokumentes sind in hoher Anzahl in diesem enthalten, dabei aber selten in anderen Dokumenten.

Als Erweiterung des TF-IDF-Maßes gilt die BM25-Berechnung (s. (5)). Im Kern wird hier analog zu (3) die Term-Häufigkeit tf mit der inversen Dokumentenhäufigkeit idf multipliziert. Vom Algorithmus existieren eine Reihe von Abwandlungen [RZ09]. Grundlage der folgenden Experimente ist die Implementierung in Elasticsearch [Co18]:

$$\begin{aligned} BM25_{IDF}(d_i, t_j) &= \ln \left(1 + \frac{\text{docCount} - tf(t_j, d_i) + 0.5}{tf(t_j, d_i) + 0.5} \right) \\ BM25_{TF}(d_i, t_j, k_1, b) &= \frac{tf(t_j, d_i) * (k_1 + 1)}{tf(t_j, d_i) + k_1 * 1 - b + \frac{b * \text{amountTerms}(d_i)}{\text{avgTermAmount}}} \\ BM25(d_i, t_j, k_1, b) &= BM25_{IDF}(d_i, t_j) \cdot BM25_{TF}(d_i, t_j, k_1, b) \end{aligned} \quad (5)$$

Hierbei werden analog zum TF-IDF-Maß in (4) Bewertungswerte für Dokumente d_i berechnet.

Dabei ist der BM25-Wert umso höher, je häufiger der jeweilige Term t_j im Dokument d_i und gleichzeitig seltener im gesamten Datensatz vorkommt. Ein grundsätzlicher Vorteil vom BM25-Algorithmus ist die leichte Anpassbarkeit [LZ11]. Über die Hyperparameter b und k_1 kann der Einfluss der Dokumentenlänge bzw. der Häufigkeit eines Wortes auf den Bewertungswert reguliert werden [Li15]. Nachteile sind schlechte Ergebnisse bei langen Dokumenten [Li15]. Zudem werden Synonyme nicht erkannt, weil keine semantische Verbindung hergestellt werden kann.

Beim BM25 Algorithmus handelt es sich um einen Algorithmus basierend auf Sparse-Vektoren⁴. Für die Suche nach den relevantesten Dokumenten wird für jedes einzelne Dokument d_i ein Vektor $\mathbf{v}_{d,i}$ der Länge n mit den jeweiligen BM25-Werten (s. (5)) erzeugt, wobei n die Anzahl einzigartiger Terme im Datensatz ist⁵:

$$\mathbf{v}_{d,i} = [BM25(d_i, t_1), BM25(d_i, t_2), \dots, BM25(d_i, t_n)] \quad (7)$$

Weiterhin wird ein Queryvektor \mathbf{v}_q mit der Länge n erstellt, dessen Elemente binär kodiert sind⁶:

$$\mathbf{v}_q = [v_q^{(1)}, v_q^{(2)}, \dots, v_q^{(n)}], \text{ mit } v_q^{(i)} = \{0,1\} \quad (8)$$

Bei Termen, die in der Suchanfrage nicht auftauchen oder die nicht im Dokument vorhanden sind, wird das jeweilige Element im Vektor \mathbf{v}_q bzw. $\mathbf{v}_{d,i}$ gleich null sein.

Um für ein Dokument letztlich die Relevanz bezüglich der Query q zu berechnen (vgl. das TF-IDF-Maß in (3)), wird das Skalarprodukt aus Query- und Dokumentenvektor gebildet:

$$BM25(d_i, q) = \mathbf{v}_q \circ \mathbf{v}_{d,i} \quad (9)$$

2.3 Dense Passage Retrieval

Das Dense Passage Retrieval (DPR) wurde von [Ka20] vorgestellt. Dieser Ansatz zur Bewertung von Dokumenten basiert auf neuronalen Netzen, die Vektoren \mathbf{v}_q , $\mathbf{v}_{d,i}$ für die Query q sowie die Dokumente $d_i \in D$ erstellen:

$$\begin{aligned} \mathbf{v}_q &= [v_q^1, v_q^2, \dots, v_q^m] \\ \mathbf{v}_{d,i} &= [v_{d,i}^1, v_{d,i}^2, \dots, v_{d,i}^m] \\ DPR(q, d_i) &= \mathbf{v}_q \circ \mathbf{v}_{d,i} \end{aligned} \quad (10)$$

Die Länge m der erstellten Dense-Vektoren \mathbf{v}_q und $\mathbf{v}_{d,i}$ wird als Hyperparameter mitgegeben. Diese Vektoren stellen die Dense-Vektor Repräsentationen der Query q sowie der Dokumente mit dar. Die Ähnlichkeit zwischen Query- und jeweiligem Dokumentenvektor wird mit Hilfe eines Proximitätsmaßes – ursprünglich dem Skalarprodukt – berechnet [Ka20].

⁴ Ein Vektor, der sparse ist, besitzt viele Werte, die gleich null sind.

⁵ Der Übersichtlichkeit halber wurden in (7) die Hyperparameter des BM25 Algorithmus weggelassen.

⁶ Eine 1 bzw. 0 im Vektor bedeutet, dass der jeweilige Term t_i in q vorhanden bzw. nicht vorhanden ist.

Das Dokument mit dem höchsten Ähnlichkeitswert stellt dabei das relevanteste Dokument bezüglich der Suchanfrage dar (vgl. (9)). Die neuronalen Netze sind Transformer [Va17] nach der BERT-Architektur [De18]. Das Ziel ist die Bildung eines Vektorraumes, in der die Vektoren v_q und $v_{d,i}$ für ähnliche Query-/Dokumentenpaare niedrige Distanzen bzw. hohe Similaritäten aufweisen. Mit Hilfe der Attentionberechnung in den Transformern sollen semantische Zusammenhänge in den Dense-Vektoren gespeichert werden [Va17, De18].

Dies stellt ein sogenanntes metric learning Problem dar [Ku13]. Hierfür werden während des Trainings der neuronalen Netze Beispielqueries mit einer richtigen sowie mehreren falschen Dokumenten trainiert. Das Ziel ist es hierbei, eine Loss-Funktion zu minimieren, so dass der Abstand zwischen den Queryvektoren und den jeweiligen Dokumentenvektoren möglichst niedrig ist [Ka20].

3 Dense Passage Retrieval basierendes Suchmodell

Für die Entwicklung des DPRs in der öffentlichen Verwaltung wurde die Farm-Haystack Bibliothek von Deepset.ai verwendet⁷. Für den Enkodierer der Query sowie der Dokumente wurde das deepset/gbert-base vortrainierte neuronale Netz von [MRP21] benutzt⁸. Die Implementierung des Dense Passage Retriever besitzt folgende Kernelemente:

```
query_embedder_model = "deepset/gbert-base"
passage_embedder_model = "deepset/gbert-base"
retriever = DensePassageRetriever(
    document_store = InMemoryDocumentStore(),
    query_embedding_model = query_embedder_model,
    passage_embedding_model = passage_embedder_model,
    max_seq_len_query = 128, max_seq_len_passage = 711 )
```

Zum Trainieren der neuronalen Netze wurden die Hyperparameter von [MRP21] und [Ka20] übernommen. Eine Ausnahme stellt die maximale Passagenlänge dar. Da die Dokumententexte des in dieser Arbeit verwendeten Datensatzes längere Beschreibungen von Behördendienstleistungen sind, wurde der Hyperparameter auf maximal 711 Terme erhöht und bei der Query auf 128 Terme. Allerdings wurde die Batch-Size auf vier verringert, um ein stabileres Training zu ermöglichen. Weiterhin wurde 20 Epochen trainiert. Als Trainingsdatensatz wurde der frei verfügbare GermanDPR Datensatz von Deepset.ai verwendet⁹. Trainiert wurde mit der train() Methode von farm-haystack¹⁰.

⁷ <https://github.com/deepset-ai/haystack>

⁸ <https://huggingface.co/deepset/gbert-base>

⁹ <https://www.deepset.ai/germanquad>

¹⁰ <https://haystack.deepset.ai/reference/retriever>

4 Experimente

Um die zurückgegebenen Dokumente von Elasticsearch sowie des DPRs zu untersuchen, wurden in Elasticsearch zwei Indizes¹¹ mit den gleichen Dokumenten erstellt, der erste Index für den BM25-Algorithmus und der Zweite für die Dense–Vektoren der Dokumente. Als Proximitätsmaß für die Vektoren dient die Euklidische Distanz.

4.1 Validierungsdatensatz

Für den Vergleich von BM25 und DPR wurden 50 Sätze ausgesucht, die als Queries für den Vergleich dienen. Dabei wurden die Sätze nicht per Zufall ausgewählt, sondern anhand von Feedback, das von Bürger:innen über eine dedizierte Eingabemaske in der produktiven Chat-Oberfläche gegeben wurde. Es wurden Sätze gewählt, die häufig eingegeben werden oder deren zurückgegebene Dokumente das Nutzerbedürfnis nicht erfüllt haben, weil diese die gesuchten Informationen nicht beinhalten. Weiterhin wurde darauf geachtet, dass statt bekannten Wörtern des Textkorpus Synonyme benutzt wurden, um auch die Qualität der Algorithmen mit nicht in den Dokumenten vorhandenem Vokabular zu überprüfen.

4.2 Auswahl des Gütemaßes

In vorausgegangenen Veröffentlichungen zum DPR wurde die Dokumentensuche als Klassifizierungsaufgabe betrachtet [Ka20, MRP21]. [Ka20] haben die Genauigkeit und [MRP21] die Sensitivität der Ergebnisse untersucht. Grundlage der folgenden Ergebnisse ist die Relevanz der Dokumente, um den prozentualen Anteil an wahrhaft richtigen Dokumenten in der Rückgabe zu einer Query zu ermitteln [Gé18]:

$$\text{Relevanz} = \frac{TP}{FP+TP} \quad (11)$$

Dafür wurden die zurückgegebenen ersten $k = \{10, 20\}$ Dokumente dahingehend überprüft, ob die Query beantwortet wurde. Diese Dokumente waren die wahrhaft richtigen (TP) Dokumente, während die restlichen Dokumente falsch richtig (FP) sind.

¹¹ <https://www.elastic.co/de/blog/what-is-an-elasticsearch-index>

4.3 Ergebnisse

Das DPR hat in Suchanfragen durchschnittlich mehr relevante Dokumente in den ersten k zurückgegebenen Dokumenten im Behördendatensatz gefunden als der BM25-Algorithmus (vgl. Tab. 1).

	BM25	Dense Retrieval	Passage
k	Relevanz in %		
10	12.9	20.5	
20	9.6	14.6	

Tab. 1: Relevanz der Algorithmen bei k zurückgegebenen Dokumenten

In Suchanfragen, in denen Synonyme verwendet werden, sind die Abstände zwischen den Relevanzwerten teils doppelt so hoch. Ein Beispiel sind Corona-Tests und PCR-Tests, wo das DPR Dokumente über PCR-Tests auch mit Corona-Tests verbinden konnte. Dies lässt vermuten, dass das DPR durch das Training auch Synonyme in den Vektoren erlernen und semantisch ähnliche Begriffe identifizieren konnte. Weiterhin werden die Ergebnisse von [Ka20] sowie [MRP21] bestätigt, dass das DPR in der Dokumentensuche den BM25-Algorithmus übertrifft.

5 Fazit und Ausblick

In dieser Arbeit wurde nach dem Beispiel von [Ka20] ein auf DPR basierendes Suchmodell für die Dokumentensuche implementiert. Dabei wurden die dahinterliegenden neuronalen Netze mit minimalen Abweichungen von den originalen Parametereinstellungen trainiert. Das Training wurde mit dem GermanDPR Trainingsatz von Deepset.ai durchgeführt. Die damit trainierten neuronalen Netze wurden für die Erstellung von Vektorrepräsentationen von Dokumenten der öffentlichen Verwaltung benutzt. Weiterhin wurden die Suchergebnisse dieses Modells mit dem aktuell eingesetzten BM25-Algorithmus von Elasticsearch verglichen.

Die Ergebnisse zeigen, dass der DPR eine höhere Relevanz in den zurückgegebenen Dokumenten erzielen und damit im selben Datensatz mehr wahrhaft relevante Dokumente finden kann als der BM25-Algorithmus. Die Ergebnisse von [Ka20] und [MRP21] konnten damit auch hinsichtlich der Relevanz als Gütemaß bestätigt werden.

Es konnte gezeigt werden, dass das DPR für Behörden das Potenzial hat, die Genauigkeit der internen Dokumentensuche sowie die Nutzerfreundlichkeit von Anwendungen der öffentlichen Verwaltung durch Beachtung von Synonymen zu verbessern. Für Behörden

bietet sich mit dem DPR ein Algorithmus an, der in Software-Anwendungen für die Dokumentensuche implementiert werden kann, um die Relevanz der Ergebnisse zu erhöhen. Wenn Behörden die neuronalen Netze zudem mit alten Dokumenten trainieren (die Wörter enthalten, die im heutigen Sprachgebrauch nicht mehr angewendet werden) können zudem Archive effektiver digitalisiert werden – die Suche kann hier potenzielle neue Begriffe mit alten Begriffen derselben Bedeutung gleichsetzen. Endanwendende können somit einen Zeitgewinn in der Dokumentensuche erzielen. Für eine umfassendere Empfehlung muss die Eignung des DPRs für den Produktiveinsatz überprüft werden. Herausforderungen hierbei sind stetig neu eingefügte Dokumente, die indiziert werden müssen. Zudem besteht die Möglichkeit, dass die neu hinzugefügten Dokumente nicht abgedeckte Themengebiete mit neuen Begriffen besitzen. Dies könnte die Ergebnisse dieser Arbeit sowie von [Ka20] und [MRP21] widerlegen. Weiterhin sollte die Relevanz der zurückgegebenen Dokumente von einem DPR mit Abwandlungen des BM25-Algorithmus wie BM25f verglichen werden. Für den Produktiveinsatz sollte zudem der Ressourcenverbrauch und die Handhabung der Kodierer bei stetig neuen Dokumenten in der Dokumentendatenbasis untersucht werden.

Literaturverzeichnis

- [Co18] Conelly, S.: Practical BM25 - Part 2: The BM25 Algorithm and its Variables. <https://www.elastic.co/de/blog/practical-bm25-part-2-the-bm25-algorithm-and-its-variables>, Stand: 11.04.2022.
- [De18] Devlin, J. et al.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Computing Research Repository (CoRR), abs/1810.04805, 2018.
- [Gé18] Géron, A.: Praxiseinstieg Machine Learning mit Scikit-Learn und TensorFlow: Konzepte, Tools und Techniken für intelligente Systeme. O'Reilly, 2018.
- [In00] Innenministerium des Landes Mecklenburg-Vorpommern, Gesetz zur Übertragung der Aufgaben für die Überwachung der Rinderkennzeichnung und Rindfleischetikettierung (Rinderkennzeichnungs- und Rindfleischetikettierungsüberwachungsaufgabenübertragungsgesetz - RkReÜAÜG M-V), 2000
- [Ja00] Jamtawee, T.: Reading a Foreign Language: Similarities and Differences. Thammasat Review, 5/1, S. 134–146, 2000.
- [Ka20] Karpukhin, V. et al.: Dense Passage Retrieval for Open-Domain Question Answering, 2020.
- [Ku13] Kulis, B.: Metric Learning: A Survey. Foundations and Trends® in Machine Learning 5/4, S. 287–364, 2013.
- [Li15] Lipani, A. et al.: Verboseness Fission for BM25 Document Length Normalization: Proceedings of the 2015 International Conference on The Theory of Information Retrieval. ACM, S. 385–388, 2015.
- [LZ11] Lv, Y.; Zhai, C.: When documents are very long, BM25 fails!: Proceedings of the 34th international ACM SIGIR conference on Research and development in Information - SIGIR '11. ACM Press, S. 1103, 2011.
- [MRP21] Möller, T.; Risch, J.; Pietsch, M.: GermanQuAD and GermanDPR: Improving Non-English Question Answering and Passage Retrieval, 2021.
- [Ro04] Robertson, S.: Understanding inverse document frequency: on theoretical arguments for IDF. 0022-0418 60/5, S. 503–520, 2004.
- [RZ09] Robertson, S.; Zaragoza, H.: The Probabilistic Relevance Framework: BM25 and Beyond. 1554-0669 3/4, S. 333–389, 2009.
- [Sp72] Sparck Jones, K.: A STATISTICAL INTERPRETATION OF TERM SPECIFICITY AND ITS APPLICATION IN RETRIEVAL. 0022-0418 1/28, S. 11–21, 1972.
- [Ta22] Tagesschau, Wie die Beschaffung schneller werden soll, <https://www.tagesschau.de/inland/bundeswehr-beschaffung-103.html>, Stand: 13.07.2022
- [Va17] Vaswani, A. et al.: Attention Is All You Need, 2017.

Strategische Technologiefrüherkennung für die Intralogistik 4.0 in der Nutzfahrzeugindustrie

Entwicklung einer Vorgehensweise zur Gestaltung von Technologie-Zielbildern

Christina Westermaier¹, Jörg Puchan²

Abstract: Die heutigen Bestrebungen zur Digitalisierung und Autonomisierung im Kontext von Industrie 4.0 forcieren auch in der Logistik grundlegende Veränderungen. Neue Ansätze und technologische Weiterentwicklungen in diesem Zusammenhang versprechen dabei ein großes Potenzial zur Bewältigung der steigenden Komplexität und zur Lösung aktueller und künftiger Herausforderungen. Unternehmen müssen dazu jedoch über den aktuellen Hype hinausblicken und die Chancen und Gefahren der Industrie 4.0 erkennen, um Zielbilder zu definieren, die den Weg in die Welt von morgen vorzeichnen. Ziel der diesem Beitrag zugrunde liegenden Masterarbeit war es deshalb, eine systematische Vorgehensweise zu entwickeln, um Unternehmen bei der Identifikation und Kommunikation von relevanten Informationen zu unterstützen. Dabei werden konventionelle Ansätze des strategischen Technologiemanagements um eine anwendungs- und prozessorientierte Sichtweise auf die Intralogistik erweitert. Die Ergebnisse entstanden in Zusammenarbeit mit einem Unternehmen der Automobilindustrie sowie einem Forschungsprojekt an einer Universität und wurden sowohl wissenschaftlich als auch anwendungsbezogen validiert und verprobt.

Keywords: Intralogistik, Industrie 4.0, Logistik 4.0, Strategisches Technologiemanagement, Technologiefrüherkennung

1 Einleitung

Bestehende Methoden der Wertschöpfung in der produzierenden Industrie stoßen unter anderem angesichts steigender Anforderungen an die Kosteneffizienz, Flexibilität, Anpassungsfähigkeit, Stabilität und Nachhaltigkeit sowie eines allgemein verschärften Wettbewerbs an ihre Grenzen [Ar08] [BHV14] [Ny08]. Aus diesem Grund werden seit einigen Jahren neue Einsatzmöglichkeiten von Technologien im Rahmen einer vierten industriellen Revolution diskutiert [Ro16]. Vor allem im Bereich der Intralogistik bzw. „internen Logistik“ verspricht die damit verbundene organisatorische und technologische Neuausrichtung ein hohes wirtschaftliches Potenzial. So soll die Etablierung von cyber-physi-

1 Hochschule München, Fakultät für Wirtschaftsingenieurwesen, Lothstraße 64, 80335 München, christina.westermaier@web.de

2 Hochschule München, Fakultät für Wirtschaftsingenieurwesen, Lothstraße 64, 80335 München, puchan@hm.edu

schen Systemen, die Verbindung zwischen realer und virtueller Welt ermöglichen und die notwendige Basis für autonome Funktionalitäten schaffen [ES18] [Fo21]. Aufgrund des rasanten technologischen Fortschritts in der jüngeren Vergangenheit und der vielen Trends und Schlagwörter, die im wissenschaftlichen Diskurs thematisiert werden, fehlt Unternehmen jedoch eine konkrete Vorstellung, wie eine Industrie 4.0 realisiert werden soll [Ca17]. Ziel der diesem Beitrag zugrunde liegenden Masterarbeit war es deshalb, eine systematische Vorgehensweise zu entwickeln, um Unternehmen frühzeitig bei der Technologie-Identifikation und -folgenabschätzung in den Prozessen der Intralogistik zu unterstützen. Dazu werden in den folgenden Abschnitten zunächst das methodische Vorgehen (Kapitel 2) und der theoretische Bezugsrahmen (Kapitel 3) dargelegt. Anschließend wird die entwickelte Vorgehensweise mit ihren vier Schritten beschrieben (Kapitel 4) und im Weiteren diskutiert (Kapitel 5).

2 Methodisches Vorgehen

Um die Konzeption einer Vorgehensweise methodisch und wissenschaftlich zu fundieren, wurde der iterative Problemlösungsansatz „Design Science“ zur Anwendung gebracht. Entlang von sechs sequenziellen Phasen und Fragestellungen nach dem Prozessmodell der Design Science Research Methodology sollte so ein Artefakt entworfen werden [Pe07]:

Problem identifizieren und motivieren

Um eine Vorgehensweise zu entwickeln, welche die Herausforderungen bei der strategischen Planung von Technologien in der Intralogistik adressiert, war zunächst eine ausführliche Literaturrecherche zum aktuellen Stand der Forschung durchzuführen (Vgl. Kapitel 3). Ein besonderes Interesse lag dabei auf dem aktuellen Methodenbestand. Hierzu wurden die Begriffe Vorgehensweise, Technologiefreherkennung, Logistik 4.0 und entsprechende Synonyme in die Suche einbezogen und mit Hilfe von AND/UND Operatoren verknüpft. Ergebnisse konnten vor allem durch eine weiterführende Vorwärts- und Rückwärtsverkettung erzielt werden, weshalb hier auf eine exakte Darstellung eines Schlagwortverzeichnisses und Trefferzahlen verzichtet wird.

Definition der Ziele einer Lösung

Anschließend wurden neun Kriterien definiert, um die Anforderungen an eine ideale Lösung zu beschreiben. Diese beziehen sich zum einen auf den Anwendungsbereich, in welchem die Vorgehensweise eingesetzt werden soll (Technologischer Wandel, langfristige Planung, produzierendes Unternehmen mit mehreren Standorten, Intralogistik) und zum anderen auf die Inhalte und Charakteristik der Herangehensweise (Rechercheprozess, ressourcenschonend, updatefähig, Bewertungsansatz, Dokumentation / Kommunikation).

Entwerfen und Entwickeln / Demonstration

Auf dieser Basis erfolgte die Bewertung und der Vergleich der bestehenden Ansätze, die im Rahmen der Literaturrecherche als relevant eingestuft wurden. Die Bewertung basiert auf einer dreistufigen Skala, welche den Grad der Erfüllung der Kriterien zum Ausdruck bringt. Die Ausprägungen unterscheiden dabei zwischen „Kriterium nicht erfüllt“, „Kriterium teilweise erfüllt“ und „Kriterium voll erfüllt“. Unter Berücksichtigung der resultierenden Ergebnisse entstand das grundsätzliche Konzept der Methode, welches anschließend in iterativen Zyklen mit Fachkräften aus den Bereichen Logistik, Technologieinnovation und Human Resources überarbeitet wurde. Die Demonstration der Vorgehensweise erfolgte mit ihrer Anwendung in der Intralogistik eines Unternehmens der Nutzfahrzeugindustrie.

Validierung

Für eine abschließende Bewertung der Vorgehensweise, war zu beurteilen, inwiefern die heutigen Schwierigkeiten im Umgang mit neuen Technologien durch den Einsatz der entwickelten Vorgehensweise behoben werden können. Dazu erfolgte zum einen die Präsentation und offene Diskussion der Ergebnisse beim Praxispartner. Zum anderen wurden fünf semistrukturierte Experteninterviews durchgeführt, die entlang von neun Fragen die konzipierten Schritte sowie den Gesamteindruck der Methode beleuchten. Zudem erfolgt eine Rückführung zu den vorab definierten Kriterien einer Lösung.

Publikation

Die Publikation der Ergebnisse, erfolgte durch die Masterarbeit und den Ergebnisbericht beim Praxispartner. Darüber hinaus entstanden ein wissenschaftlicher Beitrag für die International Conference of Supply Chain Management [Ko22] sowie dieser Beitrag.

3 Stand der Technik und Forschung

Um die zugrunde liegende Problemstellung folgerichtig abzuleiten, bedarf es zunächst einer kurzen Beschreibung der theoretischen Grundlagen.³

Prozesse in der Intralogistik

Den Untersuchungsgegenstand im Rahmen dieser Arbeit bildet die Intralogistik, die sich mit der Organisation, Steuerung, Durchführung und Optimierung der innerbetrieblichen Materialflüsse, der Informationsströme und dem Warenumschlag in Industrie, Handel und öffentlichen Einrichtungen beschäftigt [Ar06]. Um die damit verbundenen Geschäftsprozesse im Weiteren zu systematisieren und einheitlich darzustellen, konnten die Vorteile von

³ Stark verkürzte Darstellung der theoretischen Grundlagen aufgrund des limitierten Beitragumfangs. Zugang zur Vollversion der Forschungsarbeit auf Anfrage bei den Autoren.

sogenannten Referenzmodellen herausgestellt werden, die auf Erfahrungswissen zurückgreifen [Ny08]. Im Bereich der operativen Prozesse der Intralogistik lassen sich so nach der Bundesvereinigung Logistik sechs Prozessbausteine und Aufgabengebiete untergliedern: Wareneingang, innerbetrieblicher Transport, Lagerung, Kommissionierung, Verpackung, Warenausgang [Bu19].

Industrie 4.0 und Logistik 4.0

Unter dem Begriff Industrie 4.0 wird im Kern die technische Integration von cyber-physischen Systemen (CPS) in die Produktion und Logistik sowie die Anwendung des Internets der Dinge und Dienste in industriellen Prozessen verstanden [KWH13]. Der Begriff Logistik 4.0 umfasst dabei allein den Einfluss der entsprechenden technologischen Veränderungen auf den speziellen Bereich der Logistik [SS19]. Entwicklungen sind hier z. B. Technologien zur Identifikation, Augmented und Virtual Reality, Big Data Analytics, autonome Roboter und künstliche Intelligenz [St17].

Strategisches Technologiemanagement

Die Technologiefrüherkennung stellt eine Teilaufgabe des Technologiemanagements dar und verfolgt das Ziel, Chancen und Risiken im Unternehmensumfeld frühzeitig zu erkennen, um Entwicklungspotenziale abzuleiten [SK11] [Go20]. Im Rahmen dieser Arbeit konnten dazu insgesamt sieben existierende Ansätze identifiziert werden [We11] [SL10] [Re06] [Cu17] [WPU17] [Te15] [Be19].

Forschungsbedarf

Obwohl Einigkeit darüber besteht, dass Industrie 4.0 eine disruptive Auswirkung auf die produzierende Industrie nehmen wird, fehlt in Unternehmen ein einheitliches Verständnis darüber, wie das Konzept konkret umzusetzen ist [Ca17]. Die existierenden Methoden zeigen bisher einen starken Fokus auf die Dimension Technologie und vernachlässigen eine anwendungsorientierte Sicht auf das Zusammenspiel von Prozess und Technologie in der logistischen Kette [BSP18]. Um das „Was“ und „Warum“ der Veränderungen jedoch aktiv zu adressieren und eine Basis für die Zustimmung und Bereitschaft der Mitarbeitenden zu schaffen, gewinnen klar definierte und einheitliche Zielbilder immer mehr an Bedeutung [Pe19]. In der bisherigen Forschung finden sich nach Recherchen dieser Arbeit aber noch keine Ansätze, die derartige Dokumente bzw. Informationen hervorbringen. Aus diesem Grund soll eine systematische Methode entwickelt werden, die Unternehmen frühzeitig bei der Identifikation neuer technologischer Möglichkeiten für die Intralogistik unterstützt. Der Fokus liegt somit auf der entsprechenden Informationsrecherche und -aufbereitung für die Kommunikation.

4 Entwicklung einer Vorgehensweise

Für die Konzeption der Vorgehensweise wurden insgesamt sechs der bestehenden Ansätze berücksichtigt. Die Basis in Form der prozessualen Strukturierung in vier Schritten stammt dabei aus den Konzepten [SL10] [We11] mit den zutreffendsten Gesamtergebnissen in Hinblick auf die vorab definierten Anforderungen. Erweiterungen wurden darüber hinaus durch Ansätze mit einer herausragenden Bewertung in einzelnen Kriterien (z.B. Intralogistik) vorgenommen [Cu17] [WP17] [Te15] [Be19]. So entstand ein technologie- und anwendungsorientiertes Konzept für die Intralogistik, wobei dennoch Defizite in einer ausführlichen und auf die Logistik abgestimmten Kommunikation der Ergebnisse festzustellen waren. Um diese Lücke zu schließen, wurden eigenständig Konzepte für die Entwicklung von Technologie-Steckbriefen und Technologie-Zielbildern integriert. Die untenstehende Abbildung 1 zeigt dazu einen Überblick zum Gesamtergebnis in Form der vier entwickelten Schritte, den dazugehörigen Arbeitspaketen und beteiligten Personen.

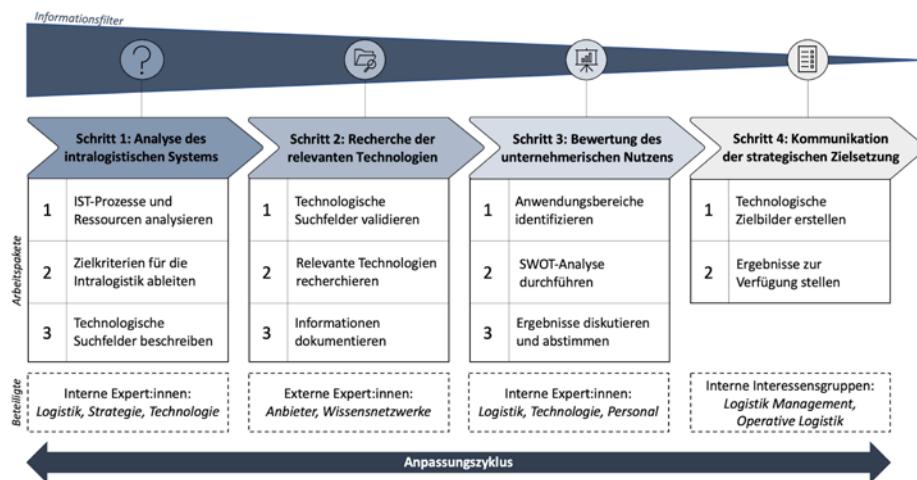


Abb. 1: Vorgehensweise zur strategischen Technologiefrüherkennung

Erster Schritt: In einem ersten Schritt ist es erforderlich, den Grundstein für eine zielgerichtete und effektive Herangehensweise zu legen, indem der Informationsbedarf für das weitere Vorgehen so genau wie möglich bestimmt wird. Dazu ist die Ausgangssituation der Logistik mit Hilfe von standardisierten Prozessbausteinen für den Material- und Informationsfluss zu erfassen (1) sowie die strategische Zielsetzung für die Intralogistik abzuleiten (2). Auf dieser Basis kann sodann eine Eingrenzung der relevanten Informationen für die Intralogistik der Zukunft vorgenommen werden (3). Um die erforderlichen Informationen zu sammeln, sollten im Unternehmen bereits vorhandene Dokumente wie Prozessdokumentationen oder Strategiedokumente berücksichtigt und Workshops mit relevanten Expertinnen und Experten der verschiedenen Werke durchgeführt werden. So resultieren als Gesamtergebnis technologische Suchfelder (z.B. autonome Entladung), welche die strate-

gischen Zielsetzungen und neuen technologischen Anwendungsmöglichkeiten in den einzelnen Prozessen (z.B. Wareneingang, Kommissionierung) zusammenfassen und als Basis für die anschließende Recherche fungieren.

Zweiter Schritt: Das Ziel des zweiten Schritts ist es, eine informatorische Grundlage bereitzustellen, die dem in Schritt 1 definierten Informationsbedarf gerecht wird und eine Bewertung der Informationen im Unternehmenskontext ermöglicht. Dazu steht die strukturierte Einbindung von externen Quellen (z.B. Technologieradar der Bundesvereinigung Logistik, Gartner Hype Cycle) im Fokus, um eine fundierte Wissensbasis zum aktuellen Stand der Technik zu generieren. Die Arbeitspakete umfassen deshalb die Überprüfung der gewählten Eingangsfilter in Form der technologischen Suchfelder (1), um darauf basierend eine gezielte aber hinreichend umfassende Recherche zu den relevanten Technologien (z. B. Drohnen, fahrerlose Transportsysteme) vorzunehmen (2). Im Anschluss können die identifizierten Technologien zu Technologiefeldern (z.B. autonomes Fahren) zusammengefasst werden, um Informationen zu Technologien mit einem ähnlichen Wirkprinzip in Form von Technologie-Steckbriefen im Teil 1 zu dokumentieren (3). Diese repräsentieren auf übersichtliche Art und Weise die charakterisierenden Informationen des technologischen Wirkprinzips sowie Umsetzungsmöglichkeiten und Besonderheiten je Technologiefeld.

Dritter Schritt: In einem dritten Schritt sind die erhobenen Informationen nun weiter aufzubereiten und zu verdichten, um die bisherigen Informationen auf das Unternehmen anzuwenden. Hierzu ist der erweiterte Wissensstand aus Schritt 2 zu nutzen, um einen Überblick über die Anwendungsmöglichkeiten je Technologiefeld und die damit verbundenen Veränderungen in den individuellen Logistikprozessen des Unternehmens zu schaffen (1). Auf dieser Basis kann der unternehmerische Nutzen durch einen qualitativen Vergleich der heutigen und zukünftigen Logistikabläufe in Form einer Strengths – Weaknesses – Opportunitites – Threats – Analyse (SWOT) bewertet werden (2). Diese Ergebnisse sind anschließend beispielsweise mit Hilfe von internen Experteninterviews oder Workshops in interdisziplinären Teams zu diskutieren und abzustimmen, um eine Rückführung der Informationen in den Unternehmenskontext vorzunehmen (3). Die Dokumentation aller Ergebnisse erweitert die bisherigen Technologie-Steckbriefe um die Unternehmenssicht (Teil 2).

Vierter Schritt: Abschließend wird in einem letzten Schritt das Ziel verfolgt, die Erkenntnisse zu den Veränderungen in der Intralogistik des Unternehmens zu kommunizieren. Für eine sukzessive Einbindung der Mitarbeitenden sind die bisherigen Ergebnisse deshalb in Form von Technologie-Zielbildern aufzubereiten (1). Diese Dokumente verweisen zum einen auf weiterführende Informationen in Form der Technologie-Steckbriefe und ermöglichen zum anderen die einheitliche Kommunikation umfassender Informationsunterlagen über verschiedene Kanäle (z. B. Arbeitskreise, Abteilungsleitungen) im Unternehmen (2). So soll mithilfe eines Technologie-Zielbilds sowohl die zu erwartende Veränderung entlang der strategischen Ziele im Unternehmen und zum anderen die konkrete Auswirkung auf die logistischen Abläufe im Material- und Informationsfluss abgebildet werden. Dadurch wird ein anwendungsbezogenes Verständnis für die technologischen Veränderungen in der Intralogistik gefördert.

5 Diskussion und Fazit

In einem letzten Kapitel sollen die Ergebnisse dieser Forschungsarbeit kritisch diskutiert und ein Ausblick auf den zukünftigen Forschungsbedarf gegeben werden. Zur Validierung der Ergebnisse werden dabei die vier Schritte der entwickelten Vorgehensweise herangezogen, die im Rahmen von Interviews und Gruppendiskussionen betrachtet wurden. Die im Rahmen der Validierung festgestellten kritischen bzw. erfolgsentscheidenden Faktoren sollten in der Methode entsprechend prominent platziert und unbedingt beachtet respektive durchgeführt werden.

Schritt 1: Der Einsatz von neuen Technologien hat das Potenzial, ganze Prozessketten zu verändern. Um erste Ansatzpunkte zu identifizieren, ist es jedoch unerlässlich zu bestimmen, welche Material- und Informationsflussaktivitäten im Unternehmen mit welcher Intention verändert werden sollen.

Schritt 2: Die Recherche ist aufgrund der Schnelllebigkeit der Entwicklungen auf dem Markt ein sehr wichtiger Schritt. Eine zentrale Stelle für diese Aufgabe ist deshalb erfolgsentscheidend, um den hohen Aufwand bei der Informationsbeschaffung aufgrund der starken Technologievielfalt zu minimieren und das Wissen unternehmensweit zu nutzen.

Schritt 3: Bei der Entscheidung für neue Technologien müssen eine Vielzahl von internen und externen Einflussfaktoren berücksichtigt werden. Eine SWOT-Analyse ist daher ein geeignetes und ergänzendes Instrument, um einen prägnanten ersten Eindruck zu erzeugen.

Schritt 4: Ohne gute Kommunikation und die Einbeziehung aller betroffenen Mitarbeiter sind Veränderungsprojekte zum Scheitern verurteilt. Da sich durch den Einsatz neuer Technologien auch Tätigkeiten und Prozesse und damit Qualifikationen verändern, ist es notwendig, die Veränderungen aktiv zu kommunizieren und zu veranschaulichen, z. B. mit Videomaterial, um Ängsten und Bedenken zu begegnen.

Zukünftiger Forschungsbedarf

In Hinblick auf den zukünftigen Forschungsbedarf sollte die entwickelte Vorgehensweise auf die Intralogistik weiterer Unternehmen im Bereich der Automobil- und Nutzfahrzeugindustrie angewendet werden. Nur so kann geprüft werden, ob die erarbeiteten Hilfestellungen ausreichend unterstützen, um ein besseres Verständnis für die zukünftigen Entwicklungen in der Intralogistik eines Unternehmens zu erzeugen. Darüber hinaus konnten im Rahmen dieser Arbeit starke Veränderung durch neue Technologien für den Arbeitsalltag des Menschen festgestellt werden. In diesem Zusammenhang sollte kritisch geprüft werden, ob die Inhalte der Methode ausreichen, um Qualifizierungsbedarfe der Zukunft abzuleiten und ein adäquates Change-Management umzusetzen.

Literaturverzeichnis

- [Ar06] Arnold, D.: Intralogistik. Potentiale, Perspektiven, Prognosen, Springer Verlag, VDI-Buch, 2006.
- [Ar08] Arnold, D. et. al.: Handbuch Logistik, 3. Auflage, Springer-Verlag, Berlin, Heidelberg, 2008.
- [Be19] Behrendt, F. et. al.: Der Intelligente Logistikraum: Neue Gestaltungsformen im Kontext der digitalen Transformation. Industrie 4.0 Management, 2019(4), S. 35-38, 2019.
- [BHV14] Bauernhansl, T., ten Hompel M., Vogel-Heuser, B.: Industrie 4.0 in Produktion, Automatisierung und Logistik, Springer Vieweg, Wiesbaden, 2014.
- [BSP18] Blöchl, S., Schneider, M., Pichlmaier, S.: Industrie 4.0 in der logistischen Kette. Zeitschrift für wirtschaftlichen Fabrikbetrieb 113(12), S. 863-868, 2018.
- [Bu19] Bundesvereinigung Logistik: Life-Cycle Costing und Performance-quantifizierung von Cyber-Physicalen Systemen in der Intralogistik. https://www.bvl.de/files/1951/2125/2131/Abschlussbericht_Industrie4.0Profitabel.pdf, 2019.
- [Ca17] Capgemini Consulting: Industry 4.0 – Eine Einschätzung von Capgemini Consulting, https://www.capgemini.com/consulting/wp-content/uploads/sites/30/2017/07/capgemini-consulting-industrie-4.0_0_0.pdf, Stand: 27. September.2021.
- [Cu17] Cuhls, K.: Unternehmensstrategische Auswertung von Foresight-Ergebnissen. In Möhrle, M., Isenmann, R.: Technologie-Roadmapping, Springer Berlin Heidelberg, 2017.
- [ES18] Endres, F., Sejdic, G.: Cyber-Physische Systeme in der Intralogistik. Zeitschrift für wirtschaftlichen Fabrikbetrieb 113(5), S. 346-349, 2018.
- [Fo21] Fottner, J. et. al.: Autonomous Systems in Intralogistics. State of the Art and Future Research Challenges. Logistics Research 14:2, S. 1-41, 2021.
- [Go20] Goehermann, J.: Technologiemanagement, Springer-Gabler, Wiesbaden, 2020.
- [Ko22] Kohl, M., Westermaier, C., Puchan, J., Fottner, J.: Developing a Company-Specific Technology Vision for Intralogistics Processes. In Association for Computing Machinery: 2022 5th International Conference on Computers in Management and Business. Singapore, S. 144-150, 2022.
- [KWH13] Kagermann, H., Wahlster, J., Helbig, J.: Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0, Plattform Industrie 4.0, 2013.
- [Ny08] Nyhuis,P.: Beiträge zu einer Theorie der Logistik, Springer-Verlag, Berlin, 2008.
- [Pe07] Peffers, K. et. al.: A Design Science Research Methodology for Information Systems Research. Journal of Management Informations Systems 24(3), S. 45-77, 2007.
- [Pe19] Petry, T.: Digital Leadership. Erfolgreiches Führen in Zeiten der Digital Economy, 2. Auflage, Haufe Group, Freiburg, 2019.
- [Re06] Reger, G.: Technologie-Früherkennung: Organisation und Prozess. In Gassmann, O., Kobe, C.: Management von Innovation und Risiko, 2. Auflage, Springer-Verlag, Berlin, Heidelberg, 2006.

- [Ro16] Roth, A.: Einführung und Umsetzung von Industrie 4.0. Grundlagen, Vorgehensmodell und Use Cases aus der Praxis, Springer Gabler, Berlin, Heidelberg, 2016.
- [SK11] Schuh, G., Klappert, S.: Technologiemanagement. Handbuch Produktion und Management 2, 2. Auflage, Springer-Verlag, Berlin, Heidelberg, 2011.
- [SL10] Schimpf, S., Lang-Koetz, C.: Technologiemonitoring, <http://publica.fraunhofer.de/dokumente/N-146352.html>, 2010.
- [SS19] Speringer, M., Schnelzer, J.: Differentiation of Industry 4.0 Models, <http://www.ra-un.org/publications>, 2019.
- [St17] Strandhagen, J. et. al.: The fit of Industry 4.0 applications in manufacturing logistics: a multiple case study. Advances in Manufacturing 5(4), S. 344-358, 2017.
- [Te15] Technische Universität Darmstadt: Generisches Vorgehensmodell zur Einführung von Industrie 4.0 in mittelständischen Unternehmen der Serienfertigung, <https://www.darmstadt.ihk.de/blueprint/servlet/resource/blob/3344884/7fe3af982b9147b216d9da67db0a1b34/pilotprojekt-cypiflex-data.pdf>, 2015.
- [We11] Wellensiek, M. et. al.: Technologiefrüherkennung. In Schuh, G., Klappert, S.: Handbuch Produktion und Management 2, 2. Auflage, Springer-Verlag, Berlin, Heidelberg, 2011.
- [WPU17] Weinert, N., Plank, M., Ullrich, A.: Metamorphose zur intelligenten und vernetzten Fabrik. Ergebnisse des Verbundforschungsprojekts MetamoFAB. Springer Vieweg, Berlin, Heidelberg, 2017.

Untersuchung des Gender Bias im deutschsprachigen Wikipedia

Sarah Krause¹

Abstract: Stereotypen und Vorurteile bezüglich Frauen und Männer herrschen trotz gesetzlicher Gleichstellung in den Köpfen vieler Menschen. Dieser sogenannte Gender Bias hat Einfluss auf viele Bereiche des gesellschaftlichen Lebens, auch auf vermeintlich neutrale Medien wie die Online-Enzyklopädie Wikipedia. Der Gender Bias in Wikipedia wurde in verschiedenen Abhandlungen anhand von Biographien untersucht und insbesondere hinsichtlich der Sichtbarkeit und Sprache durch verschiedene Parteien festgestellt. Diese Arbeit baut auf vorangegangenen Untersuchungen auf und betrachtet die Ausprägung des Gender Bias im deutschsprachigen Wikipedia hinsichtlich drei verschiedener Aspekte: Textanzahl, Textlänge und Sprache. Insgesamt bestätigten die Untersuchungen einen Gender Bias im deutschsprachigen Wikipedia. Nur 29,16% der betrachteten Biographien handeln von Frauen, auch die Textlänge der Artikel über Frauen ist durchschnittlich um 12,3% kürzer. Hinsichtlich der Sprache konnten weniger Unterscheidungen in der Beschreibungen der Geschlechter festgestellt werden als in vorherigen Abhandlungen. Dies lässt auf einen positiven Trend hin zur Minimierung des Gender Bias hoffen.

Keywords: Gender Bias, Wikipedia

1 Einleitung

Heute sind Frauen und Männer in Deutschland gesetzlich gleichgestellt. Diese Gleichstellung ist jedoch noch nicht in allen Bereichen umgesetzt. Beispielsweise verdienen Männer immer noch mehr als Kolleginnen in vergleichbaren Positionen [Fü21]. Die früheren Rollenbilder von Mann und Frau sind noch nicht aus allen Bereichen der Gesellschaft und allen Köpfen ausgeradiert. Geschlechterstereotypen und Vorurteile führen zu einer unterschiedlichen Betrachtung der Geschlechter, dem sogenannten Gender Bias. Gender Bias hat, oft unbemerkt, Einfluss auf verschiedene Alltagssituationen. Auch in einigen vermeintlich neutralen Quellen kann Gender Bias nachgewiesen werden, ein Beispiel ist die Online-Enzyklopädie Wikipedia.

Der Gender Bias in Wikipedia wurde bereits durch verschiedene Abhandlungen betrachtet, diese Arbeit soll darauf aufbauen. Ziel ist es zu untersuchen, inwiefern der Gender Bias auch im deutschsprachigen Wikipedia vorhanden ist.

¹ Hochschule für Technik und Wirtschaft (HTW) Berlin, Treskowallee 8, 10318 Berlin,
Sarah.Krause@student.htw-berlin.de

Zur Beantwortung der Forschungsfrage wurden im Rahmen dieser Arbeit Biographien in Form von Wikipediaartikeln hinsichtlich verschiedener Aspekte untersucht.

In den folgenden Kapiteln wird zunächst eine thematische Einführung bezüglich des Gender Bias auf Wikipedia, dessen Ausprägungen und Gründe gegeben, bevor auf die praktischen Untersuchungen eingegangen wird. In Kapitel 3 folgt die Erläuterung der Methode. Es wird beschrieben, welche Daten wie ausgewertet werden sollen und wie die Daten gesammelt wurden. Die Ergebnisse der Untersuchungen werden in Kapitel 4 beschrieben. Abschließend wird ein Fazit gezogen und ein Ausblick auf Anknüpfungspunkte der Arbeit gegeben.

2 Hintergrund

Als Online-Enzyklopädie sammelt Wikipedia Artikel zu verschiedenen Themenbereichen und veröffentlicht sie kostenlos, mit einer hohen Reichweite, im Internet [CB12, Gr15]. Es enthält Beiträge über verschiedene Epochen, Disziplinen und Personen und möchte, nach eigener Aussage, das Wissen der Welt darstellen. Die Besonderheit von Wikipedia, im Vergleich zu herkömmlichen Enzyklopädien, ist, dass die Erstellung der Artikel Community basiert ist [Wa16].

Benutzer*innen können sich freiwillig als Autor*innen beteiligen und müssen keine Experten in dem Gebiet ihres Beitrags sein [Gr15]. Ihre Beiträge müssen lediglich durch Quellen gestützt sein.

Trotz dieser Art der Informationssammlung zeigen Studien, dass Wikipediaartikel eine vergleichbare Qualität mit herkömmlichen Enzyklopädien aufweisen. Gegen einen Grundsatz von Enzyklopädien verstößt Wikipedia jedoch: Das Verfassen aus einem neutralen Standpunkt (englisch: neutral point of view, NPOV) heraus. Auf der Plattform kommt es verstärkt zur Überrepräsentation von bestimmten Meinungen bzw. Sichtweisen (Englisch: point of view, POV), dem sogenannten Bias. Bias tritt dabei in verschiedenen Ausprägungen auf, in kultureller, politischer und geschlechterspezifischer Sicht [Hu17]. Diese Arbeit fokussiert sich auf die unterschiedliche Betrachtungsweise der Geschlechter, dem Gender Bias, in Wikipedia. Gender Bias beschreibt die asymmetrische Assoziation von Ereignissen mit Frauen und Männern, die zu Geschlechterstereotypen führen können [SP11].

Der Gender Bias in Wikipedia wurde bereits in verschiedenen wissenschaftlichen Abhandlungen betrachtet [CB12, Gr15, Hu17, RR11, SP11, Wa15, Wa16]. Dabei wurden verschiedene Ausprägungen des Bias in Wikipedia untersucht und festgestellt. Als Grundlage der Untersuchungen dienen Wikipediaartikel in Form von Biographien. Da in diesen einzelne Personen thematisiert werden, eignen sie sich, um Unterschiede in der Behandlung und Betrachtung der Geschlechter festzustellen [Gr15]. Unterschiede zwischen den Biographien der Geschlechter wurden insbesondere in der Sprache beziehungsweise der Inhalte der Texte und der Sichtbarkeit der Artikel festgestellt [CB12, Gr15, Hu17, RR11, SP11, Wa15, Wa16].

Die Sprache, die Autor*innen innerhalb eines Textes verwendet, kann, trotz des in Wikipedia üblichen neutralen Schreibstils, ein Indikator für die Meinung der Autor*innen über das Thema des Artikels sein. Dabei kann zwischen explizitem und implizitem Bias unterschieden werden. Explizit beschreibt, dass Autor*innen ihre Meinung direkt in den Text einfließen lassen und ist dadurch oft erkennbar. Bei einem impliziten Bias hingegen werden bestimmte Fakten oder Aspekte eines Themas weggelassen, die eigentlich zum Sachverhalt gehören. Damit wird, durch die Leser*innen des Textes unerkannt, der Fokus nur auf die beschriebenen Inhalte gelegt [Hu17]. Gerade in einer Enzyklopädie wie Wikipedia ist ein impliziter Bias besonders gefährlich, da Leser*innen auf die Korrektheit und Vollständigkeit der Inhalte vertrauen und auf deren Basis Meinungen bilden und sich Wissen aneignen [SP11].

In mehreren Untersuchungen zur Ausprägung des Gender Bias in der Sprache bzw. in den Themen der Wikipediaartikel-Biographien wurden Unterschiede zwischen Männern und Frauen festgestellt [Gr15, SP11, Wa15, Wa16]. In Artikeln über Männern werden häufig Themen wie Sport und ihr Beruf in den Fokus gestellt und verstärkt kognitive Begriffe genutzt [Gr15, Wa15]. In Biographien über Frauen wird hingegen verstärkt auf ihren Familienstand (insbesondere verheiratet oder geschieden) und ihre Kinder eingegangen [Wa15]. Bei Artikeln über Frauen werden häufig persönliche Ereignisse mit beruflichen vermischt, was bei Biographien über Männer nicht der Fall ist. Diese Vermischung stärkt Geschlechterstereotypen in Bezug auf Karriere, Ehe und Elternschaft und gefährdet die Gleichstellung der Geschlechter in der Gesellschaft [SP11]. In Biographien über Frauen gibt es außerdem häufiger geschlechtsbezogene Wörter und Inhalte als in denen über Männer [Gr15, Wa15, Wa16]. Frauenartikel enthalten eher Wörter wie Frau, weiblich oder Dame (Original in Englisch: woman, female, lady) als Wörter wie Mann, männlich oder Herr (Original in Englisch: man, masculine, gentleman) in Artikeln über Männern vorkommen [Wa15]. Wagner et al. (2015) bringen dies mit der Theorie in Zusammenhang, dass das männliche Geschlecht in vielen Situationen als Null-Gender (Standardgeschlecht) angesehen wird. Männliche Wörter scheinen in diesen Situation als nicht notwendig, da der Kontext das männliche Geschlecht vorzugeben scheint [Wa15].

Diese unterschiedliche Beschreibung der Geschlechter ist wahrscheinlich nicht bei Betrachtung einer einzelnen Biographie erkennbar, doch es zeigt in der Gesamtheit einen Spiegel der Meinung der Autor*innen.

Artikel über Männer und Frauen in Wikipedia unterscheiden sich neben der Sprache, die in ihren Biographien verwendet wird, auch in ihrer Sichtbarkeit [CB12, Gr15, RR11, Wa15, Wa16]. Um Sichtbarkeit zu erreichen ist einerseits entscheidend, dass Artikel zum jeweiligen Geschlecht existieren und vorhandene Artikel eine gute Netzwerkstruktur innerhalb Wikipedias haben.

Graells-Garrido et al. stellten in ihren Untersuchungen fest, dass nur 15,5% der Biographien in ihren Daten von Frauen handeln [Gr15]. Auch die Betrachtung von Regale und Rhue bestätigen, dass in Wikipedia und anderen Online Datenbanken weniger Einträge über Frauen als über Männer existieren. [RR11]. Doch woran liegt das? Es ist wahrschein-

licher, dass eine Biographie über eine bedeutende Frau fehlt, als über einen bedeutenden Mann [CB12]. Wagner et al. (2016) interpretieren dies als Ergebnis des Effekts der gläsernen Decke. Der Begriff der gläsernen Decke wird vorrangig im beruflichen Kontext genutzt. Sie beschreibt eine unsichtbare Decke, die es Frauen unmöglich, beziehungsweise schwerer macht, in höheren Positionen eingestellt zu werden. Wagner et al. (2016) vermuten, dass Wikipedia diese gläserne Decke in Bezug auf die Erstellung von Biographien besitzt. So werden Biographien nur für Frauen erstellt, die höhere Fähigkeiten und Fertigkeiten haben als Männer über die eine Biographie verfasst wird [Wa16].

Ein Merkmal von Wikipedia ist die Verlinkung zwischen den Artikeln. Werden in Artikeln Personen oder Themen genannt, die ebenfalls einen Wikipediaartikel besitzen, werden sie verlinkt, sodass der Leser*innen sie direkt aufrufen kann. Umso öfter ein Artikel in anderen verlinkt ist, desto wahrscheinlicher ist es, dass ein Wikipedianutzer auf ihn aufmerksam wird. Eine häufige Verlinkung ist deshalb ein Merkmal für die Relevanz eines Artikels. Untersuchungen zeigen, dass Biographien von Männern eine größere Netzwerkstruktur haben und deswegen häufiger zu finden sind [Gr15]. Frauenartikel werden hingegen weniger häufig verlinkt [Wa16]. Wagner et al. (2015) fanden außerdem heraus, dass Frauenartikel häufiger zu Biographien über Männer verlinken als umgedreht. Dies hat einen negativen Effekt auf die Sichtbarkeit von Frauen in Wikipedia [Wa15].

Es gibt verschiedene Erklärungen, warum es in Wikipedia einen Gender Bias hinsichtlich der beschriebenen Aspekte gibt. Ein Ansatz ist die unterschiedliche Sicht auf die Geschlechter in der Vergangenheit [Wa15]. Frauen sind in der Geschichte oft unterrepräsentiert, da sie einer bestimmten Rolle (oft der Hausfrau und Mutter) zugeschrieben wurden und oft ausschließlich hinsichtlich dieser betrachtet wurden. Biographien zu Personen aus dieser Zeit sind deshalb oft auf diese Thematik beschränkt [Wa15].

Der aber wohl entscheidendere Ansatz ist die Übertragung des Bias der Autor*innen auf ihre Wikipediabeiträge [CB12, Gr15, Wa15, Wa16]. In der Autorenschaft von Wikipedia herrscht ein Gender Gap (Geschlechterunterschied), nur 16,1% der Autor*innen sind weiblich [HS13]. Es gibt aber nicht nur weniger Frauen, die zu Wikipediaartikeln beitragen, Autorinnen schreiben durchschnittlich weniger und hören häufig früher auf, Beiträge auf Wikipedia zu verfassen und zu bearbeiten. Da Autorinnen sich auf andere Themen fokussieren als männliche Autoren, ist der Gender Gap dahingehend ein Problem, dass diese Themen und Sichtweisen vernachlässigt werden [Gr15]. Warum der Anteil der Autorinnen in Wikipedia so gering ist, begründen Collier und Bear anhand von zwei Faktoren: Konflikten und Wissen. Autoren in Wikipedia erstellen nicht nur neue Artikel, die Hauptarbeit besteht in der Erweiterung oder Bearbeitung bereits bestehender Beiträge. Das Löschen und Aktualisieren von Texten anderer Autor*innen führt zu Konflikten auf der Plattform. Frauen neigen wiederum eher dazu als Männer, Konflikte zu vermeiden [CB12]. Der Faktor Wissen bezieht sich nicht auf tatsächliche Unterschiede in Fähigkeiten und Kenntnissen von Männern und Frauen. Männer haben durchschnittlich ein höheres Selbstvertrauen bezüglich ihres Fachwissens, was in einer stärkeren Motivation mündet, ihr Wissen mit der Welt teilen zu wollen [CB12].

3 Methode

Dieses Kapitel beschreibt die Ziele der Arbeit sowie die zu erwartenden Ergebnisse. Es wird außerdem der Erhalt der Daten und die Vorgehensweise bei der Auswertung erläutert.

3.1 Untersuchung

Im Großteil der Abhandlungen, die sich mit dem Gender Bias in Wikipedia auseinandersetzen, wird das englischsprachige Wikipedia als Datenquelle genutzt, da es die größte Anzahl von Biographien enthält. Mit ca. 102.000 Biographien ist das deutschsprachige Wikipedia auf dem dritten Platz. Bezuglich des Anteils der Biographien über Frauen bildet es das Schlusslicht unter den 20 größten Wikipedia Editionen mit 13,2 % [Wa16].

Diese Arbeit untersucht das bisher weniger betrachtete deutschsprachige Wikipedia. Basis der Untersuchungen bilden Biographien. Dabei werden nur Biographien von Personen betrachtet, die in den Jahren 1996 – 2000 geboren sind. Umso später die Geburtsjahre, desto weniger Einfluss sollte die unterschiedliche historische Betrachtung der Geschlechter auf die Inhalte der Biographien haben. Würden noch aktuellere Geburtsjahre gewählt werden, so würde die Stichprobe zu klein werden.

Anhand der ausgewählten Biographien soll überprüft werden, inwiefern in den modernen Biographien des deutschsprachigen Wikipedias ein Gender Bias zu erkennen ist. Dazu werden die Biographien hinsichtlich drei Aspekten untersucht: Textanzahl, Textlänge und Sprache. Die Auswahl der untersuchten Gesichtspunkte erfolgte basierend auf den Methoden bisheriger Untersuchungen.

Die Textanzahl wird genutzt, um einen Teil der Sichtbarkeit zu überprüfen. Es wird gezählt, wie viele Wikipediaartikel Biographien über Frauen und Männer enthalten. Aufgrund des beschriebenen Effekts der gläsernen Decke wird erwartet, dass es weniger Biographien über Frauen als über Männer gibt.

Neben der Textanzahl wird die Textlänge der vorhandenen Beiträge gesammelt und verglichen. Die Textlänge gibt Auskunft darüber, wie viel Beachtung der beschriebenen Person gegeben wird.

Mit Hilfe der Sprache soll untersucht werden, wie Personen beschrieben werden und somit die Meinung der Autor*innen und ein möglicher Bias ermittelt werden. Die Untersuchung der Sprache erfolgt über Worthäufigkeiten, welche ein Indikator für Unterschiede in Texten sind [Gr15]. Für die Auswertung von Worthäufigkeiten wird das Programm Linguistic Inquiry Word Count (LIWC) verwendet. LIWC ist ein Textanalyseprogramm, welches automatisiert Wörter in mehr als 80 psychologische Kategorien einordnet [TP09]. Die deutsche Version DE-LIWCV2015 enthält 18.711 Wörter. Wörter können in mehr als einer Kategorie vorkommen [Me17]. In dieser Arbeit werden insbesondere die Kategorien bezüglich

sozialer, kognitiver und biologischer Prozesse (Original Englisch: social, cognitive and biological processes) betrachtet. Abbildung 1 zeigt Beispielwörter der drei Kategorien sowie ihren Unterkategorien.

Social processes	social	gesellig, kumpel, reden
Family	family	papa, tochter, tante
Friends	friend	bro, kumpel
Female references	female	Frau, mädchen, weiblich
Male references	male	bruder, mann, onkel
Cognitive processes	cogproc	denken, weil, Wissen
Insight	insight	denken, realisieren
Causation	cause	deswegen, grund
Discrepancy	discrep	sollte, wollte
Tentative	tentat	eventuell, vielleicht
Certainty	certain	immer, sicher
Differentiation	differ	aber, sonst
Biological processes	bio	essen, blut, schmerz
Body	body	arm, kopf, muskel
Health	health	erkältet, klinik, medikament
Sexual	sexual	geil, heiß, nackt
Ingestion	ingest	hunger, mahlzeit, pizza

Abb. 1: Betrachtete LIWC Hauptkategorien und deren Unterkategorien [Me17].

Analog zu den Ergebnissen von Graells-Garrido. et al. wird erwartet, dass in Biographien über Frauen häufiger Begriffe der Kategorien soziale und biologische Prozesse auftreten, bei Biographien über Männer hingegen Wörter die Kategorie kognitive Prozesse [Gr15].

3.2 Datenquelle

Als Datenquelle für die Untersuchungen dienen Wikipediaartikel in Form von Biographien von Personen, die in den Jahren 1996 bis 2000 geboren sind. Die Daten aus den Biographien wurden mit Hilfe eines Web Scrapers direkt aus Wikipedia extrahiert. Der Web Scraper wurde in Python programmiert.

Wikipediaartikel sind Kategorien zugeordnet, die am Ende des Artikels aufgeführt werden. Bei Artikeln über Personen enthalten diese Kategorien unter anderem das Geburtsjahr (z.B. Geboren 1998) und das Geschlecht (z.B. Frau). Wird eine der Kategorien ausgewählt, werden alle Wikipediaartikel, die dieser Kategorie zugeordnet sind, angezeigt. Die Liste eines Geburtsjahrs dient als Ausgangspunkt für die Ausführung des Web Scrapers.

Der Web Scraper durchläuft automatisch die Liste der Namen der in dem ausgewählten Jahr geborenen Personen und öffnet die individuellen Wikipediaartikel. Aus dieser Seite

wird der Text extrahiert und geprüft, welches Geschlecht die Person hat. Dafür wird überprüft ob die Kategorie Frau oder Mann zugeordnet ist. Das Geschlecht sowie die Länge des extrahierten Textes werden in einer CSV-Datei gespeichert. Die Länge des Textes wird in Anzahl der Zeichen gemessen. Für die Analyse mit LIWC wird nicht nur die Textlänge sondern der gesamte Text benötigt. Alle Texte einer Geschlechts werden in einer TXT-Datei gespeichert. Dieser Vorgang wird für die fünf ausgewählten Geburtsjahre wiederholt.

Die Datensammlung fand im Januar 2022 statt.

3.3 Vorgehen in der Auswertung

Für die Auswertung der Textanzahl und der Textlänge wurden die CSV-Dateien der einzelnen Jahre in Excel importiert und mit Hilfe von Pivot-Tabellen ausgewertet und visualisiert.

Wie schon beschrieben, wurde die Sprache in Form der Worthäufigkeiten mit Hilfe des DELIWCV2015 ausgewertet. Dazu wurden die TXT-Dateien in das Programm geladen. LIWC gibt die Worthäufigkeiten in Form von Prozent wieder [TP09]. Die Ergebnisse wurden wiederum in Excel exportiert. Dort wurde der Mittelwert zwischen den Ergebnissen der einzelnen Jahre pro Geschlecht gebildet. Anhand dieser Mittelwerte wurden die Ausprägungen der Kategorien verglichen.

Die Ergebnisse der Datenauswertung werden im folgenden Kapitel erläutert.

4 Ergebnisse

Insgesamt wurden 16.311 Biographien für die Geburtsjahre 1996 bis 2000 extrahiert. Die Ergebnisse der Analyse hinsichtlich Textanzahl, Textlänge und Sprache werden in diesem Kapitel beschrieben.

4.1 Textanzahl

Die Anzahl der Biographien pro Jahr nimmt, wie in Abbildung 2 zu erkennen, stetig ab. Waren es für im Jahr 1996 geborene Personen noch 4528 Biographien, waren es nur noch 2295 für Personen die 2000 geboren sind. Dies ist nicht verwunderlich, da die in diesen Jahren geborenen Personen noch relativ jung sind. Mit den Jahren steigt die Anzahl an Personen, deren Biographien in Wikipediaartikeln abgebildet sind. In Zukunft werden die Anzahlen der Biographien für diese Geburtsjahre also weiter steigen.

Der Anteil von Biographien über Frauen hat sich, wie in Abbildung 2 zu erkennen, während der betrachteten Jahre kaum verändert. Insgesamt beträgt der Frauenanteil der Biographien 29,16%. Die Hypothese, dass es weniger Biographien über Frauen gibt, ist somit bestätigt.

Geschlecht	1996	1997	1998	1999	2000	Gesamt
Keine Angabe	0.16%	0.29%	0.18%	0.07%	0.09%	0.17%
Frau	29.87%	28.67%	28.80%	28.05%	30.46%	29.16%
Mann	69.96%	71.05%	71.02%	71.87%	69.46%	70.67%

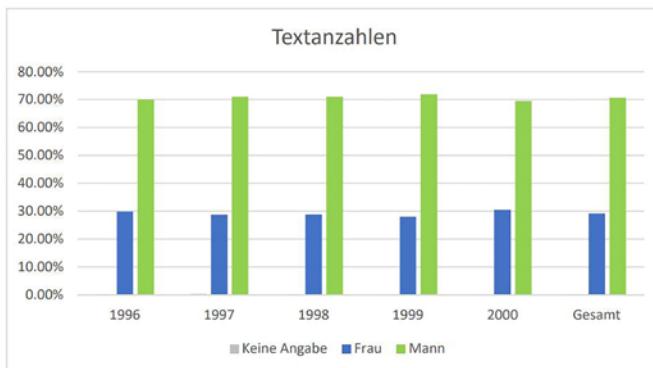


Abb. 2: Übersicht der Textanzahlen (absolut und prozentual) und Visualisierung

Damit liegt der Anteil an Frauen deutlich über dem für das gesamte deutschsprachige Wikipedia gemessenen Anteil von 13,2% [Wa16]. Dies könnte daran liegen, dass ausschließlich jüngere Jahrgänge in die Betrachtung eingeflossen sind. Das frühere Rollenbild, was es vielen Frauen erschwert hat in der Geschichte sichtbar zu werden, ist für diese Geburtsjahre nicht mehr vorhanden.

Trotzdem sind die erhaltenen 29,16% noch weit von 50% entfernt, was dem ungefähren realen Frauenanteil entspricht. Dies kann darauf hindeuten dass die gläserne Decke, die die Erstellung von Biographien über Frauen verhindert, weiterhin besteht.

4.2 Textlänge

Auch die Hypothese hinsichtlich der Textlänge hat sich bestätigt. Die Biographien über Frauen sind in jedem Jahr durchschnittlich kürzer als die über Männer.

Biographien über Frauen sind im Durchschnitt 1.711 Zeichen lang, Biographien über Männer hingegen durchschnittlich 1.921 Zeichen. Das ist ein Unterschied von 12,3%. Die genauen Ergebnisse sind in Abbildung 3 abgebildet.

Geschlecht	1996	1997	1998	1999	2000	Gesamt
Frau	1799	1739	1625	1590	1752	1711
Mann	1968	1996	1889	1843	1847	1921
Differenz	169	257	264	253	94	210
Differenz in %	9.4%	14.8%	16.3%	15.9%	5.4%	12.3%

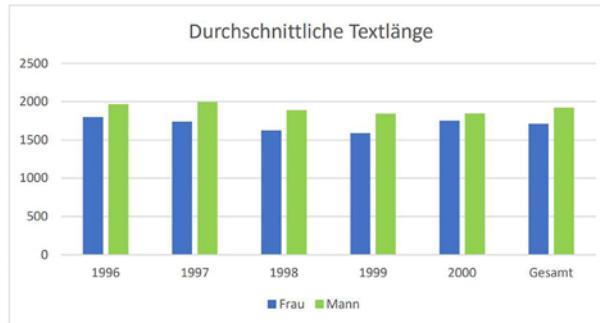


Abb. 3: Übersicht der durchschnittlichen Textlängen und Visualisierung

Einen großen Unterschied gibt es in der maximalen Textlänge. Der längste Artikel über einen Mann ist mit 39.180 Zeichen um 10.524 Zeichen länger als der längster Artikel über eine Frau, was einen Unterschied von 36,7% ergibt. Die kürzesten Artikel umfassen sowohl bei Frauen als auch bei Männern 65 Zeichen. Maximale und minimale Textlänge sind in Abbildung 4 abgebildet.

	Durchschn. Textlänge	Minimum Textlänge	Maximum Textlänge
Frau	1711	65	28656
Mann	1921	65	39180

Abb. 4: Übersicht der Textlängen

4.3 Sprache

Die Ergebnisse der Sprachanalyse mittels Worthäufigkeiten stimmen nicht für alle Kategorien mit den Erwartungen überein. Abbildung 5 zeigt das Ergebnis der Analyse für die drei untersuchten Hauptkategorien soziale, kognitive und biologische Prozesse inklusive ihrer Unterkategorien.

Geschlecht	Jahr	social	family	friend	female	male	cogproc	insight	cause	discrep	tentat	certain	differ	bio	body	health	sexual	ingest
Frau	1996	0.69	0.02	0.04	0.14	0.05	0.19	0.06	0.05	0.02	0.03	0.02	0.04	0.22	0.07	0.05	0.03	0.07
Frau	1997	0.70	0.02	0.04	0.15	0.05	0.17	0.05	0.03	0.01	0.03	0.02	0.04	0.20	0.06	0.06	0.03	0.06
Frau	1998	0.70	0.03	0.05	0.15	0.05	0.16	0.05	0.03	0.01	0.03	0.02	0.03	0.17	0.05	0.04	0.02	0.07
Frau	1999	0.73	0.02	0.06	0.17	0.04	0.18	0.06	0.03	0.01	0.03	0.03	0.03	0.20	0.07	0.04	0.02	0.07
Frau	2000	0.72	0.02	0.05	0.13	0.08	0.15	0.04	0.03	0.02	0.02	0.03	0.03	0.15	0.05	0.03	0.01	0.06
Mann	1996	0.55	0.02	0.03	0.01	0.08	0.15	0.04	0.03	0.01	0.03	0.03	0.02	0.14	0.05	0.04	0.01	0.05
Mann	1997	0.52	0.02	0.02	0.01	0.07	0.16	0.03	0.03	0.01	0.03	0.04	0.03	0.15	0.05	0.04	0.02	0.05
Mann	1998	0.53	0.01	0.03	0.01	0.07	0.14	0.04	0.02	0.01	0.03	0.03	0.02	0.13	0.04	0.04	0.02	0.04
Mann	1999	0.51	0.01	0.03	0.01	0.07	0.12	0.03	0.02	0.02	0.03	0.03	0.02	0.14	0.05	0.03	0.02	0.04
Mann	2000	0.53	0.02	0.03	0.01	0.08	0.12	0.04	0.02	0.01	0.02	0.03	0.02	0.12	0.04	0.03	0.01	0.04
Frau		0.708	0.022	0.048	0.148	0.054	0.17	0.052	0.034	0.014	0.028	0.024	0.034	0.188	0.06	0.044	0.022	0.066
Mann		0.528	0.016	0.028	0.01	0.074	0.138	0.036	0.024	0.012	0.028	0.032	0.022	0.136	0.046	0.036	0.016	0.044
Differenz		0.18	0.006	0.02	0.138	-0.02	0.032	0.016	0.01	0.002	0	-0.008	0.012	0.052	0.014	0.008	0.006	0.022

Abb. 5: Ergebnisse der LIWC Analyse

Wie erwartet werden für die Artikel über Frauen mehr Wörter verwendet, die der Kategorie soziale Prozesse zugeordnet werden können, als in Artikeln über Männern. Der Unterschied beträgt 0,18%. Besonders auffällig ist dabei die Betrachtung der Anteile der weiblichen und männlichen Begriffe. Die Ergebnisse decken sich mit den Ergebnissen von Wagner et al. (2016), dass Frauenartikel eher weibliche Wörter enthalten als Männerartikel männliche Wörter enthalten. Dies kann durch ihre Hypothese des männlichen Geschlechts als Null-Gender begründet werden.

Die Annahme, dass Wörter der Kategorie kognitive Prozesse häufiger in Biographien über Männer verkommen, konnte nicht bestätigt werden. Der Anteil bei Artikeln über Frauen liegt mit 0,17%, 0,032% über dem Anteil bei Artikeln über Männer.

Ebenso konnte nicht bestätigt werden, dass bei Biographien über Frauen deutlich mehr Wörter der Kategorie biologische Prozesse verwendet werden. Zwar ist der Anteil bei Frauen größer als bei Männern, jedoch gibt es keinen signifikanten Unterschied.

5 Grenzen

In dieser Arbeit wird nur zwischen zwei Geschlechtern unterschieden: Frau und Mann. Dies entspricht nicht der Realität. Der Anteil von Artikeln, die keinem Geschlecht zugeordnet werden konnten, die also nicht die Kategorie Frau oder Mann besitzen, beträgt 0,17% (siehe Abbildung 2). Es wurde nicht untersucht ob diese Personen sich keinem der zwei Geschlechter zugehörig fühlen, die Kategorie Frau oder Mann unabsichtlich fehlt, oder ob diese Artikel keine Biographien sind und fälschlicherweise in die Kategorie des Geburtsjahrs eingeordnet wurden. Diese Artikel wurden in der Arbeit nicht betrachtet.

6 Fazit

Der Gender Bias in Wikipedia ist eine vielbetrachtete Thematik, die in verschiedenen Ebenen untersucht und festgestellt wurde. Wikipediaartikel in Form von Biographien über Frauen und Männer unterscheiden sich insbesondere hinsichtlich ihrer Sprache und Sichtbarkeit. Während bei Frauen familienbezogene Themen dominieren, sind bei Artikeln über Männer die Themen Sport und Beruf im Vordergrund. Die Sichtbarkeit unterscheidet sich vor allem in der Anzahl der Artikel pro Geschlecht und der Netzwerk Struktur der Artikel auf der Wikipedia Plattform. Der Gender Bias kann größtenteils auf die Autor*innen der Beiträge zurückgeführt werden. Durch eine starke Mehrheit männlicher Autoren (Gender Gap) wird ihr Bias verstärkt auf die Artikel übertragen. Insbesondere der implizite Bias, das Weglassen von Informationen, die eine Biographie komplett machen würde, ist bei der Betrachtung des Gender Bias wahrscheinlich.

In der Betrachtung deutschsprachiger Biographien der Geburtsjahre 1996 bis 2000 auf Wikipedia konnten Ergebnisse vorheriger Untersuchungen teilweise bestätigt werden. Es gibt auch in diesen Geburtsjahren einen starken Unterschied zwischen der Anzahl weiblicher und männlicher Biographien. Auch hinsichtlich der Textlänge ist bei Artikeln über Frauen ein Defizit zu erkennen. Dies bestätigt die zuvor getroffenen Annahmen, unterstützt die Theorie der gläsernen Decke und zeigt, dass Frauen im Durchschnitt nicht so ausführlich betrachtet werden wie Männer.

Hinsichtlich der Sprache konnten nicht alle Annahme bestätigt werden. Wie vermutet konnte in Texten über Frauen eine Häufung von sozial assoziierten Wörtern festgestellt werden. Insbesondere die Theorie des Mannes als Null-Gender konnte durch die Verhältnisse weiblicher und männlicher Wörter bestätigt werden. Die Annahmen, dass Männer im Anteil von Wörtern kognitiver Prozesse und Frauen im Anteil von Wörtern biologischer Prozesse höher abschneiden, konnte nicht nachgewiesen werden.

Insgesamt bestätigen die Untersuchen, dass ein Gender Bias im deutschsprachigen Wikipedia vorhanden ist. Hinsichtlich der Textanzahlen wurde eine Verbesserung im Vergleich zum gesamten deutschsprachigen Wikipedia festgestellt. Dies lässt eine weitere Angleichung des Verhältnisses, von Artikel über Frauen und Männer, für spätere Generationen vermuten. Die Untersuchung der Sprache zeigt, dass Unterschiede in den Kategorien kleiner werden, was bedeutet, dass sich die Sprache, die genutzt wird um Frauen und Männer zu beschreiben, immer weiter angleicht. Dies ist wichtig, um den Gender Bias zu minimieren und Leser*innen eine neutrale Sicht auf beschriebene Personen zu bieten.

Initiativen von Wikipedia, mehr Autorinnen für ihre Plattform zu gewinnen, würde die Angleichung der Betrachtung der Geschlechter weiter vorantreiben.

6.1 Ausblick

Diese Arbeit bietet Punkte, die zur Anknüpfung weiterer Untersuchungen geeignet sind.

Es könnte ein Vergleich zwischen verschiedenen Generationen durchgeführt werden. Werden verschiedene Generationen separat betrachtet, kann eine Trendanalyse in der Entwicklung der Betrachtung der Geschlechter durchgeführt werden.

Ein anderer Ansatzpunkt könnte die erneute Betrachtung der in dieser Arbeit untersuchten Geburtsjahre, nach einer gewissen Anzahl von Jahren, sein. So könnte festgestellt werden, ob neue Artikel und Bearbeitungen von Artikeln zu einer ausgeglichenen Sicht auf die Geschlechter geführt haben

Literaturverzeichnis

- [CB12] Collier, B.; Bear, J. (2012): Conflict, criticism, or confidence: an empirical examination of the gender gap in wikipedia contributions. Proceedings of ACM 2012 conference on Computer Supported Cooperative Work.
- [Fü21] Fünffinger, A. (2021): Frauen verdienen 18 Prozent weniger, <https://www.tagesschau.de/inland/lohn-frauen-maenner-101.html>, 22.01.2022.
- [Gr15] Graells-Garrido, E. et al. (2015): First Women, Second Sex: Gender Bias in Wikipedia. Proceedings of 26th ACM Conference on Hypertext & Social Media.
- [HS13] Hill, B.; Shaw, A. (2013): The Wikipedia Gender Gap Revisited: Characterizing Survey Response Bias with Propensity Score Estimation.
- [Hu17] Hube, C. (2017): Bias in Wikipedia. Proceedings of the 26th International Conference on World Wide Web Companion.
- [Me17] Meier, T. et al. (2018): “LIWC auf Deutsch”: The Development. Psychometrics, and Introduction of DE-LIWC2015.
- [RR11] Reagle, J.; Rhue, L. (2011): Gender Bias in Wikipedia and Britannica. International Journal of Communication 5, 2011.
- [SP11] Sun, J.; Peng, N. (2011): Men Are Elected, Women Are Married: Events Gender Bias on Wikipedia.
- [TP09] Tausczik, Y.; Pennebaker, J. (2009): The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods.
- [Wa15] Wagner, C. et al. (2015): It’s a Man’s Wikipedia? Assessing Gender Inequality in an Online Encyclopedia.
- [Wa16] Wagner, C. et al. (2016): Women through the glass ceiling: gender asymmetries in Wikipedia.

Entwicklung einer mHealth Anwendung zur psychosozialen Unterstützung von pflegenden Angehörigen

Janina Krassa¹, Pia Christina Fisahn² und David Koschnick³

Abstract: Für Eltern, die ihre pflegebedürftigen Kinder (0-20 Jahre) meist viele Jahre lang zuhause pflegen, gibt es derzeit keine adäquate psychosoziale Unterstützung. Ziel ist die Entwicklung eines Designkonzepts für eine mHealth App zur psychosozialen Unterstützung von pflegenden Angehörigen. Mithilfe eines User-zentrierten Design-Prozesses und semistrukturierten qualitativen Interviews wurde ein Design-Prototyp für eine Anwendung zur psychosozialen Unterstützung dieser Zielgruppe erstellt. Zentrale Bedürfnisse pflegender Eltern sind Bedürfnisse nach Informationen, Austausch sowie Dokumentation des Pflegeverlaufs. Es konnte gezeigt werden, dass ein User-zentrierter Ansatz zur Entwicklung einer mHealth Anwendung zur psychosozialen Unterstützung von pflegenden Eltern sinnvoll ist und dass niederschwellige digitale Interventionen das Potential haben, die Probleme dieser Zielgruppe anzugehen und so zu einer Unterstützung sowie zur Verbesserung der klassischen medizinischen Versorgung beitragen können.

Keywords: mHealth, psychosoziale Unterstützung, pflegende Angehörige, User-zentriertes Design

1 Einleitung

In den letzten Jahren hat die Entwicklung, Bereitstellung und Verwendung von mHealth Apps stetig zugenommen [St18]. Unter mHealth wird der Einsatz mobiler Technologien, wie beispielsweise Smartphones, zur Unterstützung der medizinischen Praxis sowie zur Bereitstellung von Angeboten im Gesundheitsbereich verstanden [WH11].

Akteure in Politik, Wirtschaft und Gesundheitswesen zeigen sich hoffnungsvoll, dass mHealth Angebote eine effiziente Versorgungsverbesserung herbeiführen können. Neben den Chancen für die Gesundheitsprävention liegt der Fokus zunehmend auch auf der Verbesserung der Versorgung chronisch kranker Menschen [Al16].

¹ Freie Universität Berlin, Gesundheitspsychologie, Habelschwerdter Allee 45, 14195 Berlin,
janina.krassa@fu-berlin.de

² Hochschule Furtwangen, Industrial Technologies, Kronenstraße 16, 78532 Tuttlingen,
pia.christina.fisahn@hs-furtwangen.de

³ Freie Universität Berlin, Gesundheitspsychologie, Habelschwerdter Allee 45, 14195 Berlin,
david.koschnick@fu-berlin.de

Jedoch benötigen oftmals nicht nur die Erkrankten selbst, sondern auch deren Angehörige Unterstützung. So kann beispielsweise die Unterstützung von pflegenden Angehörigen zu einer Entlastung derer und somit direkt zu einer Verbesserung der Versorgung der Pflegebedürftigen selbst beitragen [WRL15]. Besonderen Unterstützungsbedarf benötigen hierbei Eltern, die ihre pflegebedürftigen Kinder im Alter von 0 bis 20 Jahren meist über viele Jahre rund um die Uhr zuhause pflegen [BB19]. In Deutschland werden rund 210.000 pflegebedürftige Kinder und Jugendliche zuhause gepflegt. Davon werden ca. 90% allein von den Eltern und nur 10% zusätzlich durch ambulante Pflegedienste versorgt [St21]. Rund 70% der pflegenden Eltern fühlen sich durch die Pflege physisch und psychisch stark belastet [BB19]. Die langjährige Pflege führt bei einem Großteil der pflegenden Eltern zu erheblichen gesundheitlichen Folgen [BP19]. Derzeit gibt es jedoch noch keine adäquate psychosoziale Unterstützung für diese Zielgruppe. Hierbei könnten niedrigschwellige Angebote [KFS11], wie beispielsweise mHealth-Anwendungen, einen sinnvollen Beitrag zur Verringerung dieser Versorgungslücke leisten. In diesem Beitrag wird die Entwicklung eines Design-Konzepts für eine mHealth Anwendung zur psychosozialen Unterstützung von pflegenden Eltern beschrieben.

2 Theoretischer Hintergrund

2.1 mHealth Apps zur psychologischen Unterstützung

mHealth Anwendungen können von den Nutzenden größtenteils eigenverantwortlich verwendet werden, um einzelne Teilbereiche ihrer Gesundheit oder Erkrankung zu überwachen bzw. zu verbessern. Hier können mHealth Anwendungen durch die Bereitstellung von Informationen, Handlungsempfehlungen oder Unterstützungs-funktionen, eine sinnvolle Ergänzung zur klassischen medizinischen Versorgung darstellen [Br+20].

Spezifische Formen von mHealth-Apps sind Anwendungen zur Verbesserung der psychischen Gesundheit und des Wohlbefindens. Diese können ein breites Spektrum der psychischen Gesundheit unterstützen: von der Unterstützung der Genesung bei psychischen Erkrankungen, über die Verbesserung des allgemeinen Wohlbefindens, bis hin zur Förderung von hilfreichen Gewohnheiten z.B. zur Verbesserung der emotionalen Gesundheit [Ba+16; Do+13].

So konnte die Wirksamkeit von mHealth-Apps zur Unterstützung bei der Behandlung von psychischen Störungen im Erwachsenenalter für die häufigsten Störungen, wie beispielsweise Depressionen [Kö+17] und Angststörungen [Ar+14], nachgewiesen werden.

Außerdem konnte der Nutzen von mHealth-Apps zur Behandlung von psychosozialen Problemen im Zusammenhang mit somatischen Erkrankungen, wie beispielsweise Diabetes [No+15], nachgewiesen werden [An16; Be+18].

2.2 User-zentriertes Design zur Entwicklung von Gesundheitsapps

Bei der Entwicklung von mHealth-Anwendungen spielt nicht nur die medizinische und inhaltliche Konzeption und Umsetzung der Gesundheitsapps eine Rolle, sondern insbesondere auch das Design, die Usability sowie die technische Implementierung haben großen Einfluss auf die Dauer der Nutzung und somit die Wirksamkeit sowie den Erfolg dieser Anwendungen [Ph+14].

Dies zeigt sich auch daran, dass die Nutzenden heruntergeladene Apps nach der ersten Verwendung nicht erneut öffnen, wenn diese nicht ihren Bedürfnissen entsprechend gestaltet sind. Es ist daher notwendig, bei der Konzeption und Entwicklung einer mHealth App einen User-zentrierten Design-Prozess zu verwenden, der sich an den Vorstellungen und Bedürfnissen der späteren Nutzenden orientiert [Mc+12]. Hierdurch kann ein besseres Nutzungsverhalten der App erreicht werden. Dies steigert somit auch die Wahrscheinlichkeit, dass die Intervention wirksam ist [Ts+07]. Aus diesem Grund empfiehlt auch die Weltgesundheitsorganisation, die Nutzenden kontinuierlich in den Entwicklungszyklus einer mHealth-App mit einzubeziehen [WH11].

3 Methodik

Bei der Wahl der Forschungsmethoden wurde ein exploratives Vorgehen gewählt. Zur Erfassung der Bedürfnisse der Zielgruppe wurde eine semistrukturierte Literaturrecherche durchgeführt. Des Weiteren wurden semistrukturierte qualitative Interviews durchgeführt, um die bisherigen Erkenntnisse aus der Literatur zu validieren und zu ergänzen. Auf Grundlage der Ergebnisse der Literaturrecherche, wurde ein semistrukturierter Interviewleitfaden mit folgenden Themenkomplexen erstellt: Demografische Daten, Diagnose und Krankheit des Kindes, pflegerischer Alltag und Bedürfnisse der Pflegenden sowie existierende Probleme. Hierdurch kann sichergestellt werden, dass interessierende Fragen durch die Teilnehmenden beantwortet werden, Rückfragen bezüglich der Antworten der Teilnehmenden möglich sind [SB15] sowie ein Einblick in die subjektiven Sichtweisen und Erfahrungen der angesprochenen Personen gewonnen werden können [Fl11].

Im Mittelpunkt des Forschungsvorhabens stehen pflegende Eltern, die als Hauptpflegepersonen ein pflegebedürftiges Kind mit einer schweren chronischen Erkrankung oder Behinderung zuhause versorgen. Im Zeitraum von Mai bis Juni 2022 wurden telefonische Interviews mit vier pflegenden Müttern durchgeführt. Die befragten Mütter waren zwischen 36 und 51 Jahren alt ($M = 44,5$). Deren pflegebedürftige Kinder waren zwischen 4 und 18 Jahren alt ($M = 13,5$). Die Befragten pflegten ihre Kinder zwischen 1 Jahr und 16 Jahren ($M = 6$). Drei der Mütter waren zum Zeitpunkt der Interviews teilzeitbeschäftigt, eine Mutter befand sich in Elternzeit.

Auf Basis der gewonnenen Erkenntnisse wurden drei Personas erstellt, welche die zukünftigen Nutzenden sowie deren Bedürfnisse und Probleme abbilden. Im weiteren Verlauf

wurde auf Grundlage der gesammelten Informationen im Rahmen einer Anforderungsanalyse ein Anforderungskatalog zur Entwicklung des Appkonzepts sowie des Design-Prototyps erstellt. Dieser bietet einen Überblick über die gesammelten Daten und enthält Nutzeranforderungen sowie Systemanforderungen, die bei der Entwicklung einer digitalen Anwendung beachtet werden müssen [GT19]. Abschließend wurde ein erstes Appkonzept sowie darauf aufbauend ein erster Design-Prototyp entwickelt.

4 Ergebnisse

4.1 Probleme und Bedürfnisse pflegender Eltern

Die Literaturrecherche zeigte, dass sich 70% der pflegenden Angehörigen durch die Pflege physisch und psychisch stark belastet fühlen [BB19]. Zentrale pflegebedingte Einflussfaktoren auf die Gesundheit der Pflegenden sind u.a. erhebliche körperliche Anstrengungen, der konstante Hilfebedarf des Kindes - auch während der Nacht, psychische Belastungen, fehlende Auszeiten für die eigene Regeneration, Auseinandersetzungen mit Kostenträgern und Behörden sowie Einschränkungen in der gesellschaftlichen Teilhabe [BP19]. Die langjährige Pflege ihrer schwerkranken Kinder führt somit bei einem Großteil der pflegenden Eltern im Laufe der Zeit zu erheblichen körperlichen und psychischen Folgen, wie z.B. chronischen Schmerzen, Verschleißerscheinungen des Bewegungsapparats, chronischen Schlafstörungen, depressiven Verstimmungen, chronischer Erschöpfung, Zukunftsängsten, abnehmender physischer und psychischer Belastbarkeit sowie sozialer Isolation [BP19]. Dies konnte auch in den Interviews bestätigt werden.

Die Literaturrecherche ergab, dass sich etwa 70% der pflegenden Eltern mehr Informationen wünschen, z.B. über die Erkrankung ihres Kindes und die daraus resultierenden Konsequenzen [KL14], zur Bewältigung der Pflegesituation und zum Erhalt der eigenen Gesundheit [BB19]. Außerdem wünschen sich über 70% der pflegenden Angehörigen Austauschmöglichkeiten mit anderen Betroffenen [BB19], um sich auszutauschen, zu beraten, zu unterstützen oder zu entlasten. Dies deckt sich auch mit den Ergebnissen der durchgeführten Interviews. Ergänzend wurden Wünsche nach krankheitsspezifischer/behinderungsgerechter Freizeitgestaltung sowie Angebote zur Förderung der eigenen Gesundheit genannt. Zudem bestätigten die befragten Eltern den Bedarf und das Interesse an einer digitalen Lösung zur psychosozialen Unterstützung.

4.2 User-zentrierter Design-Prozess

Es wurden drei Personas erstellt. Hierbei wurden zwei Benutzergruppen identifiziert: langer Pflegende (Pflege über drei Jahre) und kürzer Pflegende (Pflege unter drei Jahren). Die Benutzergruppen unterscheiden sich in ihren Wünschen und Bedürfnissen innerhalb der Pflege. So haben kurzzeitig Pflegende meist ein höheres Bedürfnis nach Informationen sowie Austausch mit anderen Betroffenen, während die länger Pflegenden ein größeres Bedürfnis zur Dokumentation des Pflegeverlaufs haben und vermehrt über körperliche Beschwerden klagen, welche durch die langjährige Pflege hervorgerufen werden.

Die darauf aufbauende Anforderungsanalyse ergab einen Anforderungskatalog mit 52 Anforderungen. Diese lassen sich in Nutzungsanforderungen und Allgemeine Systemanforderungen unterteilen.

Formulierte Anforderung	Mögliche Feature
Der Benutzer muss am System eine Möglichkeit des Austauschs erkennen können.	Chat
Der Benutzer muss am System relevante Informationen zur Pflege erkennen können.	Information
Der Benutzer muss am System nach weiteren Betroffenen suchen können.	Suchfunktion
Der Benutzer muss am System seine aktuelle Stimmung dokumentieren können.	Stimmungstracking

Tab. 1: Auszug aus den Nutzungsanforderungen nach festgelegter Syntax (nach [Ge+16])

Vier beispielhafte Nutzungsanforderungen mit daraus abgeleiteten möglichen App-Features sind in Tabelle 1 dargestellt. Die allgemeinen Systemanforderungen beschreiben, welche nichtfunktionalen Anforderungen erfüllt sein sollten. Hierzu zählen beispielsweise die einfache und intuitive Bedienbarkeit, adressatengerechte Sprache sowie ein geringer Zeitaufwand zur Nutzung der Anwendung.

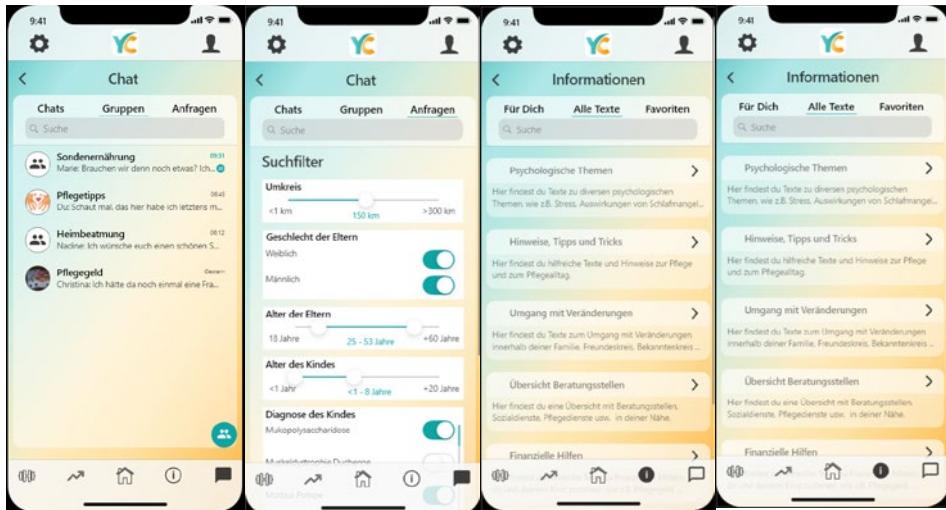


Abb. 1: Auszug aus dem Design-Prototypen (von links nach rechts): Dashboard inkl. Stimmungs-tracking, Informationsfeature, Suchfunktion nach anderen Betroffenen und Austauschfeature

Anhand der beschriebenen Ergebnisse wurde ein erstes Appkonzept sowie darauf aufbauend ein erster Design-Prototyp entwickelt. Dieser konzentriert sich zunächst auf die gestalterische Umsetzung der Anforderungen. Technische Limitationen wurden hierbei nicht berücksichtigt. Das Design ist einfach und übersichtlich gehalten, um der Zielgruppe gerecht zu werden. Beispielhaft wird in Abbildung 1 die Umsetzung des Austauschfeatures in Form eines Chats dargestellt. Dieser bietet die Möglichkeit, mit einzelnen Betroffenen oder in themenspezifischen Gruppen zu kommunizieren.

5 Diskussion

Die Ergebnisse zeigen einen großen Bedarf sowie ein großes Interesse an einer mHealth App zur psychosozialen Unterstützung für pflegende Eltern. Auf Grundlage der erfasssten Probleme und Bedürfnisse dieser Zielgruppe konnte ein erster Design-Prototyp erstellt werden. Die Ergebnisse der Literaturrecherche konnten durch die Interviews bestätigt werden. Das explorative Vorgehen mithilfe der Interviews ergab zudem spezifische Bedürfnisse sowie neue Anforderungen, die bisher noch nicht in der Literatur erwähnt wurden. Aufgrund der geringen Anzahl der Stichprobe empfiehlt sich jedoch die Durchführung weiterer Interviews bis eine inhaltliche Sättigung erreicht ist und keine weiteren neuen Erkenntnisse aus den Interviews hervorgehen. Zudem ist hierbei auf eine größere Heterogenität der Stichprobe zu achten, wie beispielsweise männliche oder alleinerziehende Pflegende sowie Eltern von pflegebedürftigen Kindern über 18 Jahren, um auch hier weitere spezifische Anforderungen zu generieren.

Insgesamt belegen die über die Literatur hinaus gefundenen Erkenntnisse die Notwendigkeit von User-zentrierten Entwicklungsprozessen bei der Entwicklung von mHealth Anwendungen. Das User-zentrierte Vorgehen erwies sich als geeigneter Ansatz, um die App angepasst an die Bedürfnisse der pflegenden Eltern zu entwickeln und so potentielle Hindernisse für die regelmäßige und längerfristige Nutzung der App frühzeitig zu identifizieren und zu berücksichtigen.

Mithilfe des entwickelten Design-Prototyps ist im weiteren Verlauf die Durchführung eines User-Testing geplant, um weiteres Feedback der Zielgruppe einzuholen. Anschließend sollte die Entwicklung eines MVP auf Grundlage der Anforderungsanalyse sowie des Design-Prototyps erfolgen. Mithilfe des MVP kann die Durchführung einer Studie zur Untersuchung der psychologischen Wirksamkeit der Anwendung realisiert werden, um den aus den bisherigen Ergebnissen, zu erwartenden Nutzen der Anwendung nachzuweisen. Bei positiver Evaluation ist ein Weg der Anwendung in das medizinische Versorgungssystem vorstellbar. Hierzu müssen jedoch weitere gesetzliche, ökonomische sowie datenschutzrechtliche Überlegungen angestellt werden. Weiterhin ist anzunehmen, dass die verwendeten Methoden auch zur Entwicklung weiterer Anwendungen zur psychosozialen Unterstützung neuer Zielgruppen, wie beispielsweise Menschen mit chronischen Erkrankungen sowie deren Angehörigen, geeignet sind.

Es konnte gezeigt werden, dass ein User-zentrierter Ansatz zur Entwicklung einer mHealth Anwendung zur psychosozialen Unterstützung von pflegenden Eltern sinnvoll ist und dass niederschwellige digitale Interventionen das Potential haben, die Probleme dieser Zielgruppe anzugehen und so zu einer Unterstützung sowie zur Verbesserung der klassischen medizinischen Versorgung beitragen können.

Literaturverzeichnis

- [Al16] Albrecht, U. V.: Chancen und Risiken von Gesundheits-Apps (CHARISMHA). Medizinische Hochschule Hannover, Hannover, 2016.
- [An16] Andersson, G.: Internet-delivered psychological treatments. Annual Review of Clinical Psychology 12/16, S. 157–179, 2016.
- [Ar+14] Arnberg, F. K. et. al.: Internet- delivered psychological treatments for mood and anxiety disorders: A systematic review of their efficacy, safety, and cost-effectiveness. PloS One 9(5), e98118/14, 2014.
- [Ba+16] Bakker, D. et. al.: Mental Health Smartphone Apps: Review and Evidence-Based Recommendations for Future Developments. JMIR Mental Health, 3(1), e7/16, 2016.
- [BB19] Bohnet-Joschko, S.; Bidenko, K.: Pflegende Angehörige: Hoch belastet und gefühlt allein gelassen. Deutsches Ärzteblatt International - Supplement: Perspektiven der Onkologie, 2019.
- [Be+18] Bendig, E. et. al.: Internet-based interventions in chronic somatic disease. Deutsches Aerzteblatt Online, 2018.
- [BP19] Büker, C.; Pietsch, S.: Abschlussbericht des Forschungsprojekts Gesundheitsbezogene Lebensqualität von Müttern mit einem pflegebedürftigen Kind (GesuLeM). Fachhochschule Bielefeld, Fachbereich Wirtschaft und Gesundheit, Institut für Bildungs- und Versorgungsforschung im Gesundheitsbereich, 2019.
- [Br+20] Brönneke, J. B. et. al.: DiGA VADEMECUM: Was man zu Digitalen Gesundheitsanwendungen wissen muss. MWV, 2020.
- [Do+13] Donker, T. et. al.: Smartphones for Smarter Delivery of Mental Health Programs: A Systematic Review. Journal of Medical Internet Research, 15(11), e247/13, 2013.
- [Fl11] Flick, U.: Qualitative Sozialforschung. Eine Einführung. 4. Auflage. Rowohlt, Reinbek bei Hamburg, 2011.
- [Ge+16] Geis, T. et al.: CPUX-UR Curriculum: Certified Professional for Usability and User Experience – User Requirements Engineering, 2016.
- [GT19] Geis, T.; Tesch, G.: Basiswissen Usability und User Experience: Aus- und Weiterbildung zum UXQB Certified Professional for Usability and User Experience (CPUX) - Foundation Level (CPUX-F) (1. Auflage), dpunkt.verlag, 2019.
- [KL14] Kofahl C.; Lüdecke D.: Die Lebens- und Versorgungssituation von Familien mit chronisch kranken und behinderten Kindern in Deutschland. knw kinder spezial, 48, 13–9/14, 2014.
- [KFS11] Klie T.; Frommelt M.; Schneekloth U.: Evaluation der Pflegeberatung gem. § 7a Abs. 7 Satz 1 SGB XI. GKV-Spitzenverband, Berlin, 2011.
- [Kö17+] Königbauer, J.: Internet- and mobile-based depression interventions for people with diagnosed depression: A systematic review and meta-analysis. Journal of Affective Disorders, 223/17, S. 28–40, 2017.
- [No15+] Nobis, S., et. Al: Efficacy of a web-based intervention with mobile phone support in treating depressive symptoms in adults with type 1 and type 2 diabetes: A randomized controlled trial. Diabetes Care, 38/15, S. 776–783, 2015.

- [St18] Statistisches Bundesamt (Destatis), Anzahl der Downloads von mHealth-Apps weltweit in den Jahren 2013 bis 2018 (in Milliarden). <https://de.statista.com/statistik/daten/studie/695434/umfrage/anzahl-der-weltweiten-downloads-von-mhealth-apps/>, Stand: Aufruf am 30.06.2022.
- [St21] Statistisches Bundesamt (Destatis), Pflegebedürftige: Deutschland, Stichtag, Geschlecht, Altersgruppen, Art der Versorgung von Pflegebedürftigen. <https://www-genesis.destatis.de/genesis/online?sequenz=tabelleErgebnis&selectionname=22421-0001>, Stand: 30.06.2022.
- [Ph14+] Phillips, R., et. Al: Randomized controlled trial of computerized cognitive behavioural therapy for depressive symptoms: effectiveness and costs of a workplace intervention. *Psychological Medicine*, 44(04)/14, S. 741-752, 2014.
- [Mc12+] McCurdie, T., et. Al: mHealth Consumer Apps: The Case for User-Centered Design. *Bio-medical Instrumentation & Technology*, 46(s2)/12, S. 49-56, 2012.
- [Ts07+] Tsai, C., et. Al: Usability and Feasibility of PmEB: A Mobile Phone Application for Monitoring Real Time Caloric Balance. *Mobile Networks and Applications*, 12/07, S. 173-184, 2007.
- [SB15] Sarodnik, F.; Brau, H.: Methoden der Usability Evaluation: Wissenschaftliche Grundlagen und praktische Anwendung (3., unveränderte Auflage). Hogrefe, 2015.
- [WH11] WHO Global Observatory for eHealth, mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organization. <https://apps.who.int/iris/handle/10665/44607>, Stand: 30.06.2022.
- [WRL15] Wetzstein, M.; Rommel, A.; Lange, C.: Pflegende Angehörige–Deutschlands größter Pflegedienst. <https://edoc.rki.de/bitstream/handle/176904/3137/3.pdf>, Stand: 30.06.2022.

Autor*innen

Prof. Dr. Thomas Barton

Hochschule Worms
Erenburgerstr. 19
67549 Worms
barton@hs-worms.de

Prof. Dr. Helmut Beckmann

Hochschule Heilbronn
Max-Planck-Str. 39
74081 Heilbronn
Helmut.beckmann@hs-heilbronn.de

Marcel Biewendt

Fachhochschule Aachen
Eupener Str. 70
52066 Aachen
biewendt@fh-aachen.de

David Brodmann

DResearch Fahrzeugelektronik
Spenerstraße 38
10557 Berlin
David.brodmann@luckycloud.de

Prof. Dr. Johannes Busse

Hochschule Landshut
Am Luzernhof 1
84036 Landshut
jbusse@jbusse.de

Maria Margarita Cabrera Frias

bbw Hochschule
Leibnizstraße 11-13
10625 Berlin
margarita.cabrera023@gmail.com

Mazlum Copurkuyu

Hochschule Worms
Erenburgerstr. 19
67549 Worms
Mazlum.Copurkuyu@service-haus.de

Simone Dogu

Hochschule Karlsruhe
Moltkestraße 30
76133 Karlsruhe
simone.dogu@h-ka.de

Prof. Dr. Katrin Dziergwa

Hochschule für Technik und Wirtschaft
Treskowallee 8
10318 Berlin
Katrin.Dziergwa@htw-berlin.de

Prof. Dr. Mathias Eggert

FH Aachen
Eupener Str. 70
52066 Aachen
eggert@fh-aachen.de

Prof. Dr. Sandy Eggert

Hochschule für Wirtschaft und Recht
Badensche Straße 52
10825 Berlin
sandy.eggert@hwr-berlin.de

Maximilian Engelhardt

Hochschule für Technik und Wirtschaft
Ostendstraße 1
12459 Berlin
Maximilian.Engelhardt@htw-berlin.de

Linda Fernsel

Hochschule für Technik und Wirtschaft
Treskowallee 8
10318 Berlin
Linda.Fernsel@htw-berlin.de

Pia Christina Fisahn

Hochschule Furtwangen
Kronenstraße 16
78532 Tuttlingen
pia.christina.fisahn@hs-furtwangen.de

Prof. Dr. Selcan Ipek-Ugay

Vostura GmbH
Residenzstraße 25
13409 Berlin
selcan.ipek-ugay@vostura.com

Kendime Ismailji

PwC
Friedrichstraße 14
70174 Stuttgart
kendime.ismailji@pwc.com

Stefan Haag

Hochschule Worms
Erenburgerstraße 19
67549 Worms
haag@hs-worms.de

Laura Haase

Hochschule für Wirtschaft und Recht
Badensche Straße 52
10825 Berlin
laura.haase@hwr-berlin.de

Tabea Herrmann

Vostura GmbH
Residenzstraße 25
13409 Berlin
tabea.hermann@vostura.com

Britta Hesse

Technische Universität Berlin
Ernst-Reuter-Platz 7
10587
britta.hesse@tu-berlin.de

Sebastian Homer

Hochschule für Technik und Wirtschaft
Ostendstraße 25
12459 Berlin
Sebastian.Homer@htw-berlin.de

Anne Jellinghaus

Technische Hochschule Wildau
Hochschulring 1
15745 Wildau
anne.jellinghaus@th-wildau.de

Prof. Dr. Andreas Johannsen

Technische Hochschule Brandenburg
Magdeburger Straße 50
14770 Brandenburg
johannse@th-brandenburg.de

Silvia Knittl

PwC
Bernhard-Wicki-Straße 8
80636 München
silvia.knittl@pwc.com

Prof. Dr. Peter Konhäusner

GISMA Hochschule
Konrad-Zuse-Ring 11
14469 Potsdam
peter.konhaeusner@gisma.com

Janina Krassa

Freie Universität Berlin
Habelschwerdter Allee 45
14195 Berlin
janina.krassa@fu-berlin.de

Sarah Krause

Hochschule für Technik und Wirtschaft
Treskowallee 8,
10318 Berlin
SarahKrause@Student.HTW-Berlin.de

David Koschnick

Freie Universität Berlin
Habelschwerdter Allee 45
14195 Berlin
david.koschnick@fu-berlin.de

Ulrike Küchler

Gamebook Studio HQ GmbH
Wilhelm-Kabus-Straße 77
10829 Berlin
ulrike.kuechler@gamebook.studio

Prof. Dr. Claudia Lemke

Hochschule für Wirtschaft und Recht
Alt-Friedrichsfelde 60
10315 Berlin
claudia.lemke@hwr-berlin.de

Linda Mai

Heinrich-Heine-Universität
Universitätsstraße 1
40225 Düsseldorf
linda.mai@hhu.de

Prof. Dr. Birte Malzahn

Hochschule für Technik und Wirtschaft
Treskowallee 8
10318 Berlin
Birte.Malzahn@htw-berlin.de

Robert Maurer

Technische Hochschule Brandenburg
Magdeburger Straße 50
14770 Brandenburg
robert.maurer@th-brandenburg.de

Anket Mehra

Duale Hochschule Schleswig-Holstein
Hans-Detlev-Prien-Straße 10
24106 Kiel
anket.mehra@stud.dhsh.de

Prof. Dr. Dagmar Monett Diaz

Hochschule für Wirtschaft und Recht
Alt-Friedrichsfelde 60
10315 Berlin
dagmar.monett-diaz@hwr-berlin.de

Prof. Dr. Christof Mosler

Hochschule für Technik Stuttgart
Schellingstraße 24
70174 Stuttgart
christof.mosler@hft-stuttgart.de

Prof. Dr.-Ing. Sebastian Möller

Technische Universität Berlin
Ernst-Reuter-Platz 7
10587
sebastian.moeller@tu-berlin.de

Prof. Dr.-Ing. Frank Neumann

Hochschule für Technik und Wirtschaft
Wilhelminenhofstraße 75A
12459 Berlin
Frank.Neumann@htw-berlin.de

Prof. Dr. Malte Prieß

Duale Hochschule Schleswig-Holstein
Hans-Detlev-Prien-Straße 10
24106 Kiel
malte.priess@dhsh.de

Tom Vincent Peters

Dataport AöR
Altenholzer Straße 10-14
24161 Altenholz
tomvincent.peters@dataport.de

Frauke Prott

Technische Hochschule Wildau
Hochschulring 1
15745 Wildau
frauke.prott@th-wildau.de

Prof. Dr. Jörg Puchan

Hochschule München
Lothstraße 64
80335 München
puchan@hm.edu

Daniel Quiter

Hochschule für Technik und Wirtschaft
Treskowallee 8
10318 Berlin
daniel.quiter@gmail.com

Lucas Rahn

Hochschule für Wirtschaft und Technik
Wilhelminenhofstraße 75A
12459 Berlin
Lukas.Rahn@htw-berlin.de

Prof. Dr. Stefanie Regier

Hochschule Karlsruhe
Moltkestraße 30
76133 Karlsruhe
stefanie.regier@h-ka.de

Prof. Dr. Erik Rodner

Hochschule für Wirtschaft und Technik
Wilhelminenhofstraße 75A
12459 Berlin
Erik.Rodner@htw-berlin.de

Juan Carlos Rodríguez-del Pino

University of Las Palmas de Gran Canaria
C. Juan de Quesada 30
35001 Las Palmas de Gran Canaria
jc.rodriguezdelpino@ulpgc.es

PD Dr. Esther Ruiz Ben

Technische Hochschule Wildau
Hochschulring 1
15745 Wildau
ruiz_ben@th-wildau.de

Tobias Schmieg

Hochschule Heilbronn
Max-Planck-Str. 39
74081 Heilbronn
schmieg.stud@hs-heilbronn.de

Lynn Schmodde

Heinrich-Heine-Universität
Universitätsstraße 1
40225 Düsseldorf
lynn.schmodde@hhul.de

Prof. Dr. Marianne Schmolke

Hochschule Wismar
Phillipp-Müller-Straße 14
23966 Wismar
Marianne.Schnmolke@hs-wismar.de

Prof. Dr. Stephan Seeck

Hochschule für Technik und Wirtschaft
Ostendstraße 1
2459 Berlin
Stephan.Seeck@Htw-berlin.de

Eric Siegeris

Vostura GmbH
Residenzstraße 25
13409 Berlin
eric.siegeris@vostura.com

Prof. Dr. Katharina Simbeck

Hochschule für Technik und Wirtschaft
Treskowallee 8
10318 Berlin
Katharina.Simbeck@htw-berlin.de

Prof. Dr. Margit Scholl

Technische Hochschule Wildau
Hochschulring 1
15745 Wildau
margit.scholl@th-wildau.de

Regina Schuktomow

Technische Hochschule Wildau
Hochschulring 1
15745 Wildau
regina.schuktomow@th-wildau.de

Prof. Dr. Carlo Simon
Hochschule Worms
Erenburgerstraße 19
67549 Worms
simon@hs-worms.de

Giuseppe Sortino
Hochschule für Wirtschaft und Recht
Badensche Straße 52
10825 Berlin
giuseppe.sortino@outlook.de

Prof. Dr. Ingo Stengel
Hochschule Karlsruhe
Moltkestraße 30
76133 Karlsruhe
ingo.stengel@hs-ka.de

Pablo Stockhausen
Universität Potsdam
August-Bebel-Straße 89
14482 Potsdam
stockhausen017@gmail.com

Daniel Szafarski
Hochschule Heilbronn
Max-Planck-Str. 39
74081 Heilbronn
daniel.szafarski@hs-heilbronn.de

Prof. Dr. Irene Weber
Hochschule für angewandte
Wissenschaften Kempten
Bahnhofstraße 61
87435 Kempten
irene.weber@hs-kempten.de

Prof. Dr. Marius C. Wehner
Heinrich-Heine-Universität
Universitätsstraße 1
40225 Düsseldorf
marius.wehner@hhu.de

Prof. Dr. Thomas Wengerek
Hochschule Stralsund
Zur Schwedenschanze 15
18435 Stralsund
thomas.wengerek@hochschule-stralsund.de

Christina Westermaier
Hochschule München
Lothstraße 64
80335 München
christina.westermaier@web.de

Prof. Dr.-Ing. Martin R. Wolf
Fachhochschule Aachen
Eupener Str. 70
52066 Aachen
m.wolf@fh-aachen.de

Lara Zakfeld
Hochschule Worms
Erenburgerstraße 19
67549 Worms
lara-zakfeld@hs-worms.de

Philipp M. Zäh
Fachhochschule Aachen
Eupener Str. 70
52066 Aachen
philipp.zaehl@alumni.fh-aachen.de

Angewandte Forschung in der Wirtschaftsinformatik 2022

Tagungsband zur 35. Jahrestagung
des Arbeitskreises Wirtschaftsinformatik an Hochschulen für Angewandte
Wissenschaften im deutschsprachigen Raum (AKWI)
vom 11.09. bis 13.09.2022, ausgerichtet von
der Hochschule für Technik und Wirtschaft Berlin (HTW Berlin) und
der Hochschule für Wirtschaft und Recht Berlin (HWR Berlin)

Autor*innen

Thomas Barton, Helmut Beckmann, Marcel Biewendt, David Brodmann,
Johannes Busse, Maria Margarita Cabrera Frias, Mazlum Copurkuyu, Simone Dogu,
Katrin Dzergwa, Mathias Eggert, Sandy Eggert, Maximilian Engelhardt,
Linda Fernsel, Pia Christina Fisahn, Selcan Ipek-Ugay, Kendime Ismailji,
Stefan Haag, Laura Haase, Tabea Herrmann, Britta Hesse, Sebastian Homer,
Anne Jellinghaus, Andreas Johannsen, Silvia Knittl, Peter Konhäusner,
David Koschnick, Janina Krassa, Sarah Krause, Ulrike Küchler, Claudia Lemke,
Birte Malzahn, Linda Mai, Robert Maurer, Anket Mehra, Dagmar Monett,
Christof Mosler, Sebastian Möller, Frank Neumann, Tom Vincent Peters,
Malte Prieß, Frauke Prott, Jörg Puchan, Daniel Quiter, Lucas Rahn, Stefanie Regier,
Erik Rodner, Juan Carlos Rodríguez-del-Pino, Esther Ruiz Ben, Tobias Schmieg,
Lynn Schmodde, Marianne Schmolke, Margit Scholl, Regina Schuktomow,
Stephan Seeck, Eric Siegeris, Katharina Simbeck, Carlo Simon, Giuseppe Sortino,
Ingo Stengel, Pablo Stockhausen, Daniel Szafarski, Irene Weber, Marius C. Wehner,
Thomas Wengerek, Christina Westermayer, Martin Wolf, Lara Zakfeld, Philipp Zähl



ISBN 978-3-95545-409-8

