



literature, but the potentials of text analytical methods and application systems (text mining, text analytics) are usually not covered. As a result, the digital competencies required for system-supported analysis and processing of literature data are not sufficiently developed. In order to address this competence gap, an adequate course has been developed at Osnabrück University of Applied Sciences and implemented as part of a project-oriented study format. This article documents the professional and technical design of this course and shows potential for future development.

**Keywords:** e-learning, Text Mining, Text Analytics, Literature Review, Literature Analysis, Scientific Work, Digital Competences, Data Literacy

## 1 Problemstellung

---

Die Durchführung einer systematischen Literaturanalyse ist eine zentrale Kompetenz wissenschaftlichen Arbeitens und bildet daher einen festen Ausbildungsbestandteil von Bachelor- und Masterstudiengängen. Die Literaturanalyse gehört damit zum grundlegenden wissenschaftlichen „Handwerkszeug“ und wird an Hochschulen unter anderem im Rahmen von Lehrveranstaltungen zu Techniken des wissenschaftlichen Arbeitens oder auch durch Kurse an Hochschulbibliotheken vermittelt [11]. In diesen Veranstaltungen werden Studierende zwar mit den etablierten Hilfsmitteln zur Suche nach Literatur und deren Verwaltung vertraut gemacht, allerdings werden die Potenziale textanalytischer Methoden und entsprechender analytischer Anwendungssysteme meist nicht thematisiert. Solche Anwendungssysteme sind in der Lage, bislang unbekannte Muster und Zusammenhänge in Textdaten zu identifizieren, wobei das Spektrum textanalytischer Methoden von einfachen lexikometrischen Verfahren (z. B. Frequenz- und Konkordanzanalysen für unterschiedliche Wortformen) bis hin zu komplexen, multivariaten Ansätzen zur Segmentierung und Klassifikation von Textdokumenten mithilfe maschineller Lernverfahren reicht [10].

Der zurückhaltende Einsatz textanalytischer Methoden für die Literaturrecherche ist angesichts der allgemeinen Entwicklung textbasierter Datenbestände kritisch zu sehen. So ist generell festzustellen, dass digitalisierte, textbasierte Daten mittlerweile einen hohen Anteil an sämtlichen weltweit verfügbaren Datenbeständen ausmachen [17]. Zudem verdoppelt sich der Bestand an wissenschaftlichen Publikationen innerhalb von ca. 24 Jahren bei einer jährlichen Wachstumsrate von etwa 3 % [6], sodass auch in Zukunft mit weiterhin stark wachsenden Literaturbeständen zu rechnen ist.

Angesichts dieser quantitativen Entwicklung wird in der Nutzung textanalytischer Methoden und unterstützender, analytischer Anwendungssysteme ein hohes Potenzial gesehen, um die Produktivität und Qualität wissenschaftlicher Arbeitsprozesse zu steigern. Diese sind vielfach dadurch gekennzeichnet, dass Aufgaben der Literatursuche und -verwaltung zwar bereits in hohem Maße digitalisiert sind – etwa durch die Nutzung etablierter Literaturdatenbanken und Literaturverwaltungsprogramme – indes die inhaltliche Analyse und Synthese von Literaturquellen weitestgehend manuell erfolgen [25]. Eine Digitalisierung dieser Kernaktivitäten kann einerseits dazu beitragen, dass die bestehenden Arbeitsabläufe in literaturintensiven Forschungsprozessen unterstützt werden, sodass personelle Ressourcen entlastet werden. Andererseits kann die Digitalisierung auch zur

Etablierung innovativer Arbeitsabläufe beitragen, indem beispielsweise großvolumige Literaturrecherchen zur Beantwortung neuer, datengetriebener Forschungsfragen ermöglicht werden.

Um die Datenkompetenzen (Data Literacy [15]) der Studierenden für die Literaturrecherche zu stärken, ist an der Hochschule Osnabrück eine neue Lehrveranstaltung entwickelt und im Rahmen eines projektorientierten Veranstaltungsformats umgesetzt worden. Diese Veranstaltung richtet sich schwerpunktmäßig an Studierende wirtschaftswissenschaftlicher Studiengänge, die bereits über IT-Grundkompetenzen zum Umgang mit strukturierten Datenbeständen verfügen, und diese um Fähigkeiten zur Handhabung unstrukturierter Textdatenbestände ergänzen möchten. Im Rahmen dieses Beitrags wird die fachliche und technische Ausgestaltung dieser Veranstaltung vorgestellt. Zu diesem Zweck wird in einem ersten Schritt auf die Ziele und Aufgaben der Literaturrecherche im Forschungsprozess eingegangen, aus denen inhaltliche Anhaltspunkte für den Ausbildungsbedarf hervorgehen. Anschließend werden das Veranstaltungskonzept und exemplarische, analytische Aufgabenstellungen vorgestellt. Abschließend werden die Erfahrungen aus der Umsetzung (Lessons Learned) expliziert.

## 2 Die Literaturanalyse im Forschungsprozess

### 2.1 Ziele und Aufgaben der Literaturanalyse

Literaturrecherchen dienen in Forschungsprozessen dazu, den aktuellen Stand (State of the Art, State of Research) oder die historische Entwicklung der Forschung in einem thematisch abgegrenzten Forschungsfeld widerzuspiegeln [11]. Diese Aufgabenstellung kann in unterschiedlichen Phasen des Forschungsprozesses anfallen. Einen Bezugsrahmen hierfür liefert Abb. 1, die einen generischen Forschungsprozess darstellt.

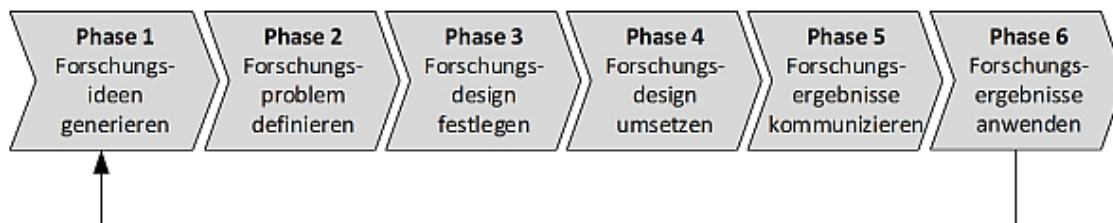


Abbildung 1: Phasenmodell für einen generischen Forschungsprozess in Anlehnung an [20]

Die einzelnen Phasen dieses idealtypischen Forschungsprozesses sind kurz zu erörtern:

1. Der Forschungsprozess beginnt mit der Generierung von Forschungsideen als Impulsgeber für neue Forschungsprojekte. Dies kann sowohl durch kreativ-intuitive Ansätze (z. B. Brainstorming) als auch logisch-systematische Verfahren (z. B. Inhaltsanalysen) unterstützt werden.
2. In der zweiten Phase ist das Forschungsproblem zu konkretisieren, indem der Wissensstand mit Rückgriff auf die dokumentierten wissenschaftlichen Erkenntnisse erhoben wird.

3. Aufbauend auf der Formulierung des Forschungsproblems erfolgt in Phase 3 die Fixierung eines Forschungsdesigns, in dem – in Abhängigkeit vom Forschungsziel – die anzuwendenden Forschungsmethoden festzulegen sind.
4. Die vierte Phase hat die Umsetzung des ex ante definierten Forschungsdesigns zum Inhalt. In dieser Phase wird das Set der im Forschungsdesign fixierten Forschungsmethoden ausgeführt.
5. Anschließend findet die Kommunikation der generierten Erkenntnisse an Wissenschaftler, Studierende und weitere Adressaten (z. B. Drittmittelgeber) statt, die in der Regel durch Publikation der Forschungsergebnisse realisiert wird.
6. Abschließend sieht der Forschungsprozess auch die Nutzung der generierten Forschungsergebnisse vor. Die daraus gewonnenen Erkenntnisse können wiederum Impulse zur Generierung weiterführender Forschungsideen liefern, sodass der Forschungszyklus erneut zu durchlaufen ist.

Aus der skizzenhaften Darstellung wird deutlich, dass die Literaturanalyse im Forschungsprozess an mehreren Stellen von Bedeutung ist. So kann eine Literaturanalyse erste Orientierung bei der Generierung von Forschungsideen liefern (Phase 1), und auch zur systematischen Identifikation von Erkenntnis- und Gestaltungsdefiziten bei der Formulierung von Forschungsproblemen dienen (Phase 2). Darüber hinaus wird die Literaturanalyse auch als eine eigenständige Forschungsmethode aufgefasst, die im Rahmen des Forschungsdesigns (Phase 4) umgesetzt wird und zur Erkenntnisgewinnung beiträgt [28].

## 2.2 Prozessmodell zur Literaturanalyse

Zur operativen Durchführung von Literaturanalysen sind unterschiedliche Aktivitäten zu verknüpfen, die auch Aufgaben der Suche und Beschaffung von relevanter Literatur zum Gegenstand haben. Zu diesem Zweck liegen mittlerweile zahlreiche Vorschläge und Handlungsempfehlungen vor, die teils durch die jeweilige Fachdisziplin geprägt sind (z. B. [26][24][7]). Aufgrund der Orientierung an Studierenden der Wirtschaftswissenschaften werden hier die Handlungsempfehlungen nach [11] zugrunde gelegt, die anhand eines Prozessmodells in Abbildung 2 dargestellt werden.

Der dargestellte Literaturanalyseprozess beginnt mit einem bereits definierten Forschungsfeld, das anschließend mithilfe eines Sets thematischer Schlagwörter bzw. Stichwörter zu beschreiben ist. Oberbegriffe sind primäre Suchbegriffe, die das Forschungsfeld generell beschreiben, während Unterbegriffe als sekundäre Suchbegriffe das Forschungsfeld einschränken. Anschließend sind Literaturdatenbanken auszuwählen, in denen die Recherche nach relevanter Literatur erfolgen soll. Für diese sind Datenbankabfragen zu formulieren, die an den Anforderungen der jeweiligen Literaturdatenbank auszurichten sind [24].

Die Ausführung dieser Abfragen liefert Fundstellen, die im Kontext des subjektiven Informationsbedarfs des Forschenden bewertet werden. Dabei besteht die grundsätzliche Zielsetzung darin, möglichst sämtliche relevanten Literaturquellen zu identifizieren (High Recall), aber auch den Anteil unbrauchbarer Fundstellen zu reduzieren (High Precision) [5]. Die Sichtung der resultierenden Fundstellen kann dazu führen, dass die gewählten Begriffe

zur Konzeptualisierung des Forschungsfelds zu überarbeiten oder aber das Set an Literaturdatenbanken bzw. deren Abfragen zu modifizieren sind, etwa durch die thematische, zeitliche oder räumliche Einschränkung bzw. Erweiterung des Suchraums.

Nach Ausführung der Abfragen liegen Fundstellen vor, die zur Weiterverarbeitung – z. B. mithilfe gängiger Literaturverwaltungssoftware – gesichert werden. In der Regel werden dabei die zentralen bibliografischen Angaben sowie die Kurzzusammenfassung (Abstract) exportiert, zudem stehen bisweilen auch Volltexte in Literaturdatenbanken zur Verfügung. Hieran schließen sich Aktivitäten zur Bereinigung der Literaturliste (z. B. Dubletteneliminierung, Priorisierung) und zur Beschaffung fehlender Volltexte an. Abschließend ist der Forschungsstand zu erarbeiten, wobei in hohem Maße strukturierende und bewertende Aktivitäten erforderlich sind. Eine Kernaufgabe bei der Erarbeitung des Forschungsstands besteht nach [26] in der Identifikation und Interpretation zentraler, forschungsleitender Konzepte. Zur Schaffung eines strukturierten Gesamtüberblicks über die Konzepte des Forschungsfelds wird daher die Konstruktion einer Konzeptmatrix vorgeschlagen, die in ihrer einfachen Form die Abdeckung unterschiedlicher Konzepte (z. B. Theorien, Methoden, Modelle) durch die untersuchten Literaturquellen eines Forschungsfelds widerspiegelt. Zur Konstruktion der Konzeptmatrix ist ein systematisches Lesen der Literaturquellen erforderlich, sodass sämtliche relevanten Konzepte identifiziert werden können. Zu diesem Zweck werden Lesetechniken empfohlen, wie z. B. die SOAR-Methode [19]. Nach der Erarbeitung des Forschungsstands können die Ergebnisse – etwa in Form einer Konzeptmatrix – weiterverarbeitet werden (z. B. für Abschlussarbeiten, Publikationen), alternativ können die Ergebnisse auch dazu führen, dass eine Neuausrichtung des Forschungsfelds stattfindet.

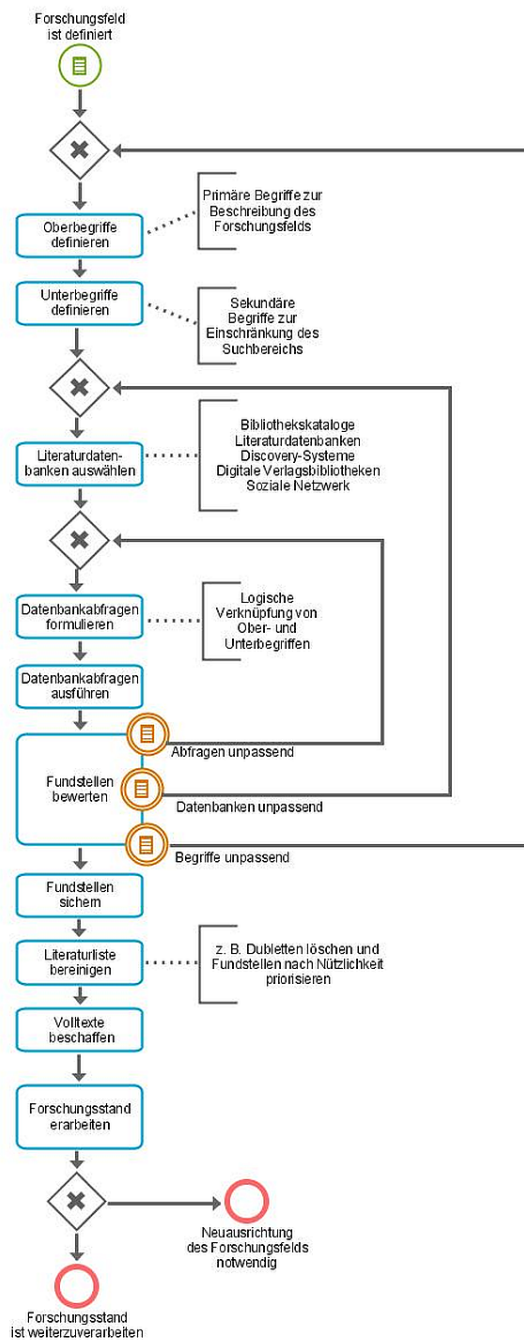


Abbildung 2: Literaturanalyseprozess in Anlehnung an [4]

## 2.3 Digitalisierungspotenziale der Literaturanalyse

Hinsichtlich der Werkzeugabdeckung des Literaturanalyseprozesses ist festzustellen, dass insbesondere die Suchaktivitäten mittlerweile durch eine Vielzahl unterschiedlicher Recherchelösungen unterstützt werden. So stehen neben traditionellen

Bibliothekskatalogen (Online Public Access Catalogue, OPAC) auch spezielle Onlinedienste unterschiedlichster Anbieter zur Verfügung. Zu unterscheiden sind beispielsweise:

- auf Literatursuchen spezialisierte Online-Literaturdatenbanken (z. B. Google Scholar),
- Meta-Suchmaschinen (z. B. Karlsruher Virtueller Katalog),
- suchmaschinenbasierte Discovery-Systeme für Bibliotheken (z. B. EBSCO Discovery Service, Primo),
- digitale Bibliotheken von Verlagen (z. B. SpringerLink) oder Fachgesellschaften (z. B. ACM Digital Library), sowie
- soziale Netzwerke für Scientific Communities (z. B. ResearchGate).

Zur Verwaltung der mit diesen Lösungen identifizierten Fundstellen stehen Literaturmanagementsysteme (z. B. EndNote, Zotero) mit Funktionalitäten zum Sichern, Ordnen und Zitieren in gängigen Textverarbeitungsprogrammen zur Verfügung. Hingegen ist im Hinblick auf die Praxis der Literaturanalyse festzustellen, dass die inhaltliche Verarbeitung der Literaturquellen zur Ermittlung des Forschungsstands bislang überwiegend ungestützt stattfindet [25]. Zur Unterstützung dieser kognitiv anspruchsvollen Aktivität werden in der Literatur unterschiedliche Ansätze diskutiert. Einerseits wird die Verwendung von Softwarepaketen für die qualitative Datenanalyse (QDA-Software, wie z. B. Nvivo, MAXQDA) vorgeschlagen, die eine schemabasierte Codierung von Texten gestatten [2]. Auf diese Weise können unterschiedliche Konzepte bzw. Konzeptkategorien in Texten annotiert und untersucht werden, wobei allerdings ein hoher manueller Aufwand erforderlich ist. Andererseits werden textanalytische Verfahren aus dem Umfeld des Text Mining [18] diskutiert, mit denen die Inhalte und insbesondere auch Zitationen von Literaturquellen zur Handhabung explorativer Aufgabenstellungen verarbeitet werden (z. B. [10][13]). In der Anwendung solcher textanalytischer Verfahren ist hohes Potenzial zu sehen, da hiermit generell die Arbeitsproduktivität von Fachkräften in wissensintensiven Wertschöpfungsprozessen gesteigert werden kann [16].

Für die Umsetzung dieser Verfahren stehen unterschiedliche Kategorien von Anwendungssystemen zur Verfügung. Einerseits sind vertikale Lösungen verfügbar, die dediziert für den Bereich der Literaturdatenanalyse entwickelt worden sind. Hierzu zählen beispielsweise die Standardsoftwareprodukte VOS Viewer, Pajek, Citnetexplorer und Bibexcel, die ihren Schwerpunkt in der Visualisierung von Literaturdaten besitzen und teils auf spezifische Literaturdatenbanken als Datenquelle ausgerichtet sind [21]. Andererseits besteht die Möglichkeit, horizontale Softwareprodukte für die Textanalyse einzusetzen, die domänenunabhängig genutzt werden können [1]. Insbesondere für die Ausbildung von Studierenden der Wirtschaftswissenschaften wird diese Option als attraktiv gesehen, da die Analyse von Textdaten in den unterschiedlichen betriebswirtschaftlichen Teildisziplinen zunehmend an Bedeutung gewinnt, etwa im Umfeld der Marktforschung [12], im Controlling [23] und im Dienstleistungsmanagement [8]. Im Folgenden wird ein Veranstaltungskonzept beschrieben, das zur Vermittlung entsprechender Datenkompetenzen zur Literaturdatenanalyse mithilfe von horizontaler Standardsoftware entwickelt worden ist.

### 3 Konzept der Lehrveranstaltung

---

An der Fakultät Wirtschafts- und Sozialwissenschaften der Hochschule Osnabrück sind mehr als 5.000 Studierende in 37 Bachelor- und Masterstudiengängen eingeschrieben, die von Betriebswirtschaft und Management über Gesundheit und Soziales sowie Öffentliches Management bis hin zu Wirtschaftsrecht und Wirtschaftsinformatik reichen. Um Studierende der wirtschaftswissenschaftlichen Studiengänge zur selbständigen Analyse von Literaturdaten zu befähigen, ist eine projektorientierte Lehrveranstaltung konzipiert und in Form einer Blockveranstaltung umgesetzt worden. Blockveranstaltungen ergänzen regelmäßig stattfindende Lehrveranstaltungen und haben eine Dauer von einer Woche. In diesen Blockwochen können insbesondere auch IT-Themen behandelt werden, für die längere zusammenhängende Bearbeitungszeiten erforderlich sind.

Die Veranstaltung fokussiert die von der Bibliothek der Hochschule Osnabrück angebotene Recherchelösung scinos (Scientific Information Osnabrück) [3], die auf dem EBSCO Discovery Service (EDS) basiert. Diese Lösung bietet wissenschaftliche Inhalte aus sämtlichen Fachdisziplinen unter einer einheitlichen Suchoberfläche an, wobei kommerzielle Ressourcen sowie internationale Open-Access-Inhalte erfasst werden. Mit scinos können Fundstellen mit ihren bibliografischen Daten gesichert und in verschiedenen Formaten (z. B. RIS, BibTeX) exportiert werden, sodass eine empirische Basis für die Anwendung textanalytischer Verfahren entsteht.

Zur Auswertung dieser Datenbasis ist eine Reihe von Text-Mining-Systemen untersucht worden, wobei neben der Methodenunterstützung und der Bedienbarkeit auch die kostenfreie Verfügbarkeit für Lehrzwecke im Mittelpunkt standen. Insgesamt hat diese Untersuchung dazu geführt, dass das Analysesystem IBM Watson Explorer (WEX) gewählt wurde [27]. Diese Lösung steht einerseits über das akademische Programm des Herstellers (IBM Academic Initiative) Hochschulen für Lehrzwecke zur freien Verfügung, und unterstützt andererseits ein umfangreiches Set an Analysemethoden sowie Dokumentsprachen. Darüber hinaus bietet sie eine komfortable, browserbasierte Benutzeroberfläche und gestattet auch die Analyse großvolumiger Dokumentkollektionen.

Zur Anbindung von IBM WEX an die Recherchelösung scinos ist eine datenbasierte Kopplung realisiert worden. So können Studierende ihre Datenbankabfragen zunächst in scinos realisieren und die Ergebnisse anschließend als RIS-Dateien (Research Information System Format) sichern, wobei die Anzahl der exportierbaren Fundstellen je Export auf eine Obergrenze von 25.000 Sätze beschränkt ist. Diese RIS-Dateien werden anschließend mithilfe eines Konverters in ein XML-Format umgewandelt, das für jede Fundstelle den Fundstellentyp (z. B. Journal, Book), den Haupttitel/Buchtitel, die Autor(en), das Publikationsjahr, die Schlagwörter und Kurzfassung (Abstract) beinhaltet.

Die inhaltliche Strukturierung der Lehrveranstaltung ist anhand der funktionellen Architektur von Text-Mining-Systemen nach [14] und [27] vorgenommen worden (siehe Abbildung 3). So wird beispielsweise das Arbeiten mit fokussierten Crawlern und die Konfiguration von Dokumentprozessoren thematisiert, die der Vorbereitung der Literaturdaten dienen. Den thematischen Schwerpunkt bildet hingegen die Ausführung unterschiedlicher Analysemethoden zur Exploration von Textdatenbeständen. Nach der Einführung in die Funktionalitäten von IBM WEX ist den Studierenden die Möglichkeit gegeben worden, eigene Problemstellungen mithilfe des Systems in Form projektorientierter Gruppenarbeit



zu bearbeiten sowie zum Abschluss der Blockwochenveranstaltung zu präsentieren und zu diskutieren. Aufbauend auf dem skizzierten Veranstaltungskonzept werden im folgenden zentrale Aufgaben und Methoden zur Analyse von Literaturdaten mit IBM WEX demonstriert.

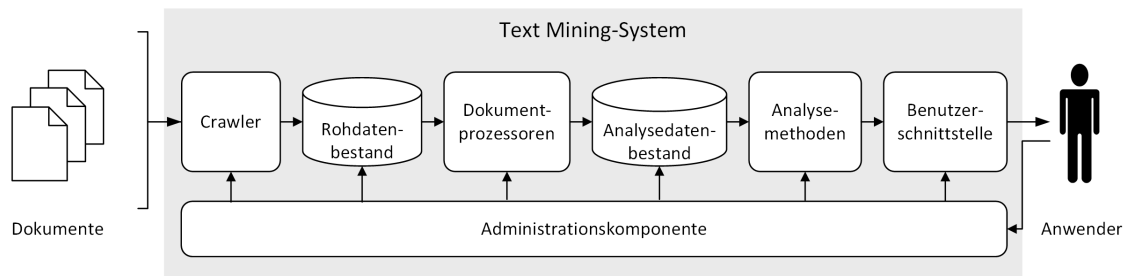


Abbildung 3: Elemente eines Text-Mining-Systems (in Anlehnung an [14] und [27])

## 4 Analysemethoden zur Erschließung von Literaturdaten

Zur Demonstration der Analysemethoden des Text Mining wird hier eine einfache, exemplarische Literaturrecherche für das Forschungsfeld Robotic Process Automation (RPA) (z. B. [9][22]) verwendet. Die folgende Abfrage sucht in scinos nach Publikationen, die im Langtext einen expliziten RPA-Bezug besitzen:

```
TX "robotic process automation" OR
TX "robotergesteuerte prozessautomatisierung"
```

Hiermit wurden in scinos 11.191 Fundstellen mit Kurzzusammenfassung (Abstract) identifiziert und anschließend exportiert (Stand März 2019). Nach dem Import und der Vorbereitung dieser Daten stehen in IBM WEX die im folgenden beschriebenen Techniken zur explorativen Analyse zur Verfügung.

### Suchanfragen

Die Literaturdaten können mit WEX durch eine Suchmaschine abgefragt werden. Abbildung 4 zeigt die trunkierte Suche nach dem Schlagwort *business process* mit der entsprechenden Ergebnismenge (n=760) und der markierten Verwendung des Schlagworts im Kontext des Abstracts der jeweiligen Literaturquelle (Keyword in Context, KWIC-Analyse). Dabei können auch komplexere Abfragen mit logischen Operatoren (NOT, AND, OR) formuliert werden, wobei die Suche auf einzelne Felder oder auch Dokumentsprachen (z. B. Englisch, Deutsch) beschränkt werden kann. Mithilfe dieser Analysemethode können Literaturdatenbestände in Bezug auf Begriffe für a priori definierte Konzepte überprüft werden, sodass eine theoriegeleitete Untersuchung der Literaturdaten möglich wird.

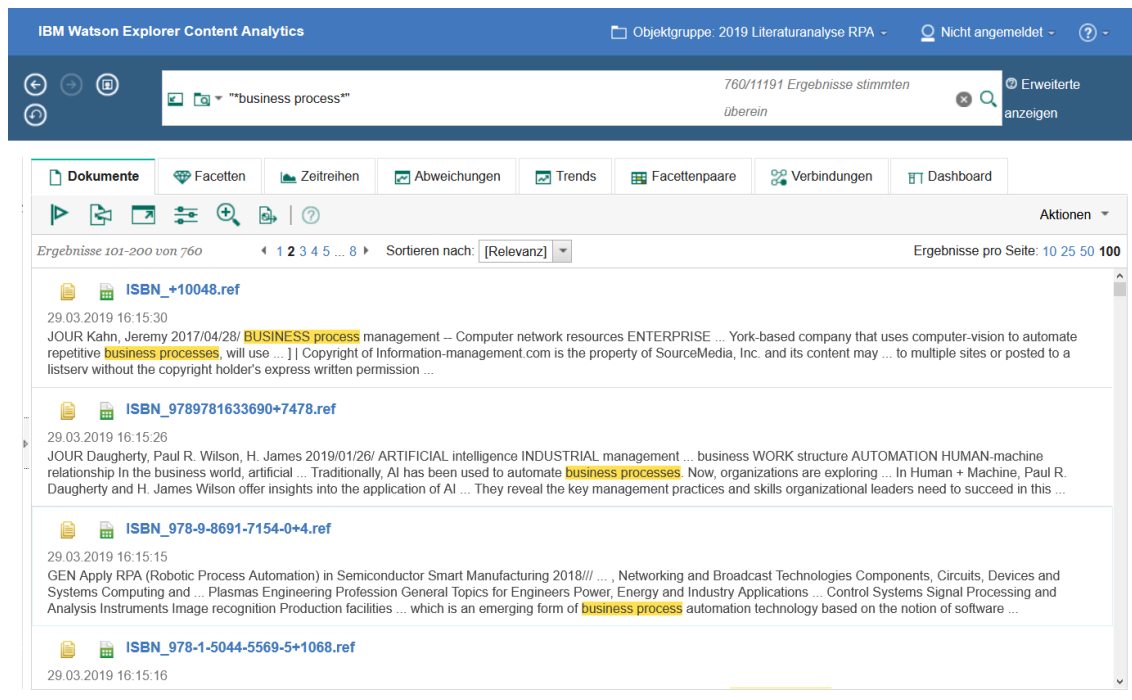


Abbildung 4: Suchanfrage und Trefferliste

### Analyse von Wortarten und Subgruppen

Außerdem besteht die Möglichkeit, unterschiedliche Wortarten (z. B. Substantive, Verben) und Satzteile (z. B. Substantivfolgen, Verb-Substantiv- und Adjektiv-Substantiv-Kombinationen) zu identifizieren. Diese Funktionalität ist relevant, da Fachterminologien in der Regel durch Substantive bzw. komplexere Substantivkonstruktionen gebildet werden [16]. Abbildung 5 zeigt einen Ausschnitt der in den Fundstellen enthaltenen Substantivfolgen, in denen der Begriff *learning* auftritt, mit Angabe der absoluten Häufigkeiten (Frequenzanalyse). Auf diese Weise können frequente Konzepte (z. B. *machine learning*, *survey machine learning*) in der Literaturdatenbasis identifiziert werden, sodass ein exploratives Instrument zur induktiven Ableitung potenziell interessanter Konzepte zur Verfügung steht.

Anzeigen: Werte Operator: OR Filter: learning

Werte	Häufigkeit
machine learning	330
technology application machine learning	68
report machine learning	23
intelligence machine learning	19
sap Leonardo machine learning	15
machine learning algorithm	12
machine learning technology	11
statistic machine learning	11
learning Robotic process automation	10
automation mechanization machine learning	10
trend machine learning	10
machine learning solution	10
survey machine learning	9
intelligence machine learning data	9

Abbildung 5: Gefilterte Frequenzanalyse von Substantivfolgen

Die Recherche in den Literaturdaten kann auch auf Subgruppen beschränkt werden. In diesem Fall besteht die Möglichkeit, Zusammenhänge zwischen Konzepten in den gebildeten Subgruppen zu identifizieren. Abbildung 6 zeigt Substantivfolgen für sämtliche Literaturquellen (n=301), in denen das Konzept *machine learning* auftritt. Diese Liste ist nach der Kennzahl Korrelation absteigend sortiert, die mit Werten größer 1 eine besondere Häufung (Dichte) der jeweiligen Substantivfolge in der Subgruppe anzeigt. So wird aus dem Beispiel deutlich, dass das Konzept *machine learning* mit Substantivfolgen wie *speech Technology*, *enterprise mobility* und *analytics Technology* verknüpft ist.

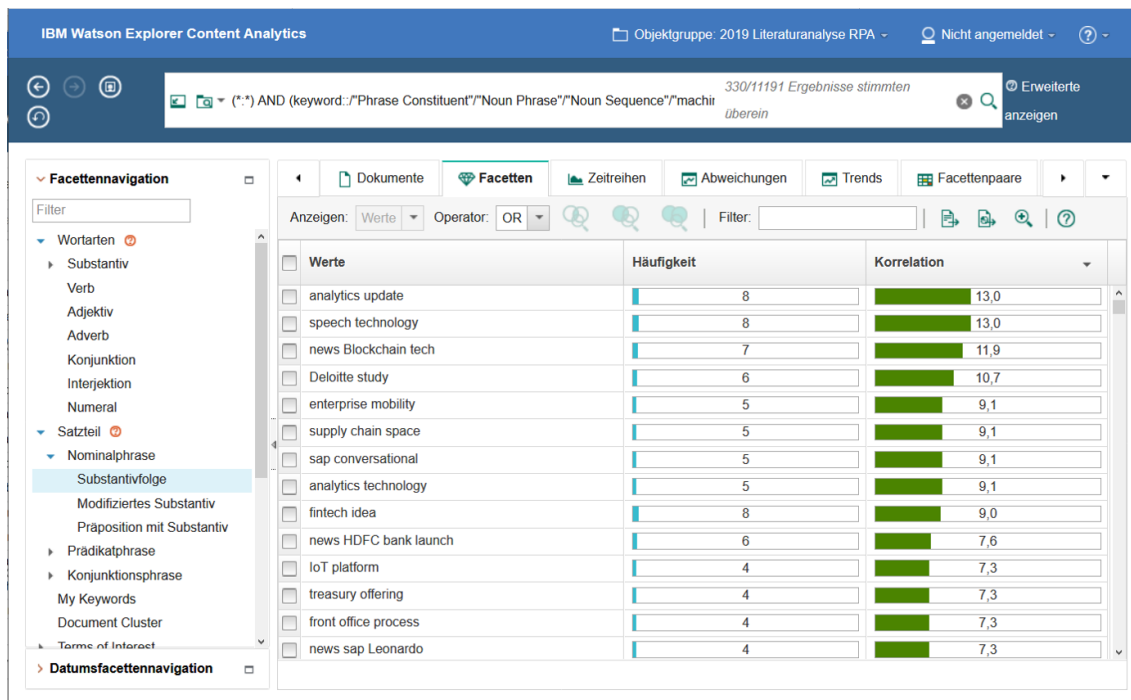


Abbildung 6: Subgruppenanalyse für Literaturquellen mit der Substantivfolge machine learning

## Zeitreihenanalyse

Mithilfe einer Zeitreihenanalyse kann die Fragestellung beantwortet werden, wie sich themenspezifische Publikationsaufkommen im Zeitablauf entwickelt haben, wobei als Datumsattribut das scinos-Publikationsjahr verwendet wird. Dabei kann nach beliebigen Begriffen gefiltert werden, sodass die zeitliche Diffusion einzelner Konzepte transparent wird. Abbildung 7 verdeutlicht als Balkendiagramm die starke Zunahme an RPA-Publikationen im Zeitablauf. Diese Zeitreihenanalyse kann auch verfeinert werden, sodass die zeitliche Entwicklung einzelner Konzepte in der Literaturlbasis transparent wird.

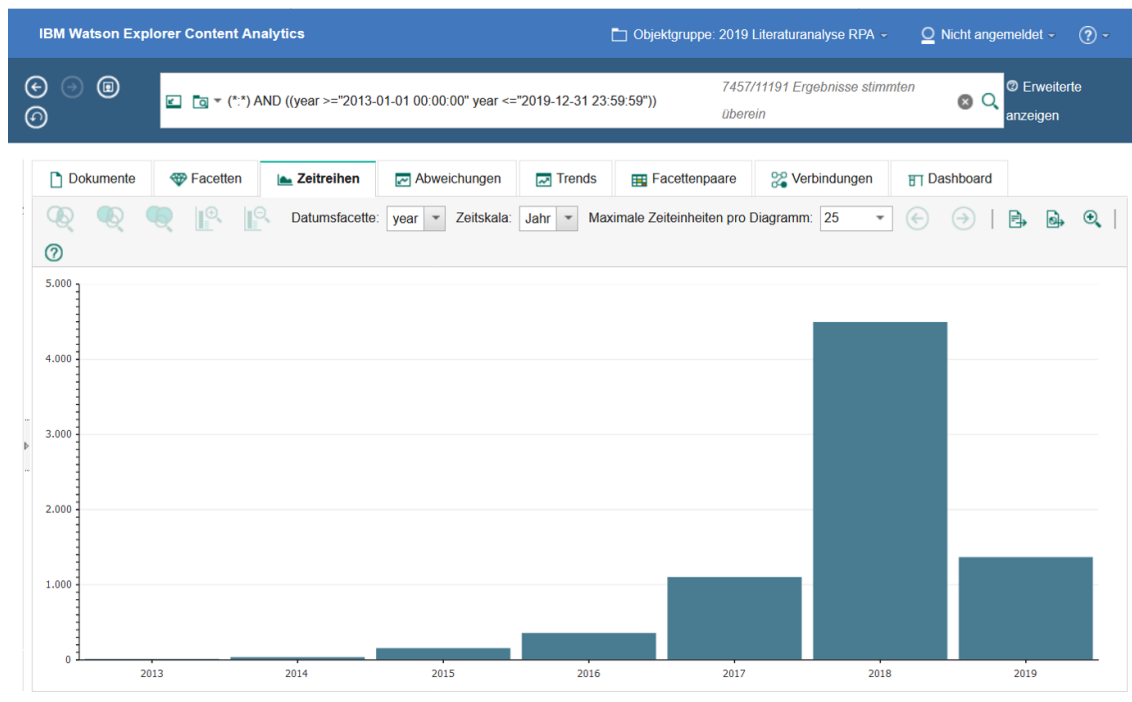


Abbildung 7: Zeitliche Entwicklung der Literaturlbasis (Anzahl Publikationen nach Publikationsjahr)

### Korrelationsanalyse

Durch Korrelationsanalysen können Zusammenhänge zwischen unterschiedlichen Konzepten der Fachdomäne ermittelt werden. Abbildung 8 zeigt ein Netz hochkorrelierter Substantivfolgen und modifizierter Substantive, aus denen fachliche Zusammenhänge zwischen den Konzepten deutlich werden. Die Analyse ist auf Grundlage sämtlicher Dokumente der Literaturlbasis erstellt worden, in denen der Fachbegriff *digital Transformation* auftritt. Aus den Zusammenhängen wird beispielsweise deutlich, dass agilem Projektmanagement (agile...project) in dieser Subgruppe zentrale Bedeutung zukommt. Auch werden Hinweise dazu geliefert, welche Branchen bzw. Akteure in diesem Umfeld relevant sind (z. B. *airline Company, consulting...firm, management...Consulting*) und welche ökonomischen Zielsetzungen verfolgt werden (*operational...Efficiency*). Diese thematischen Zusammenhänge liefern eine potenzielle Informationsgrundlage zur Prüfung und Entwicklung von Forschungsfragen.

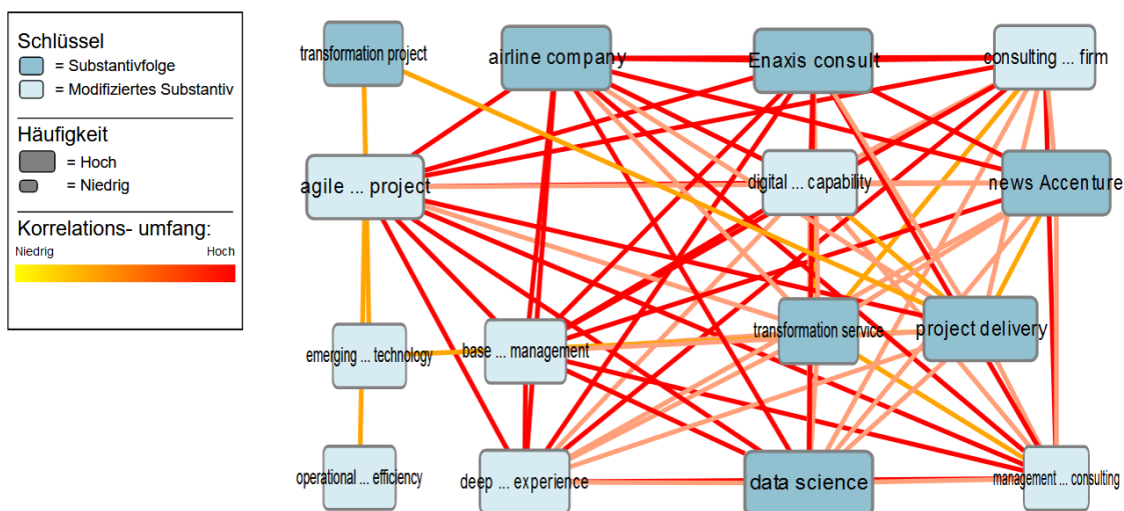


Abbildung 8: Korrelationsanalyse für Substantivfolgen und modifizierte Substantive

## 5 Lessons Learned

Das vorgestellte Veranstaltungskonzept ist an der Hochschule Osnabrück im Sommersemester 2018 im Rahmen einer Blockveranstaltung mit 22 Studierenden aus unterschiedlichen Bachelorstudiengängen durchgeführt und durch das Qualitätsmanagement der Hochschule evaluiert worden. Die Evaluationsergebnisse haben dabei gezeigt, dass die Veranstaltung von den Studierenden in Bezug auf den Schwierigkeitsgrad und den Umfang als überwiegend passend wahrgenommen wurde, während in Bezug auf die Interessanztheit der Lehrinhalte deutlich bessere Ergebnisse im Vergleich zum Durchschnitt über sämtliche Blockveranstaltungen des Semesters erzielt werden konnten. Allerdings ist auch festzustellen, dass die Bewertung in Bezug auf die inhaltliche Relevanz für das individuelle Studienziel hinter dem Durchschnitt zurückbleibt. Ein maßgeblicher Einflussfaktor hierfür wird darin gesehen, dass die teilnehmenden Studierenden eine geringe Fachsemesterzahl aufweisen ( $\mu=3,09$  Fachsemester), sodass sich die konkrete Herausforderung zur Erstellung einer systematischen Literaturanalyse – etwa zur Vorbereitung der Bachelorarbeit – noch nicht gestellt hat. Hieraus lässt sich die Empfehlung ableiten, die Veranstaltung künftig für Studierende höherer Fachsemester anzubieten, insbesondere auch für Masterstudierende mit einem tendenziell intensiveren Forschungsbezug.

Aus Dozentsicht hat sich der Einsatz des ausgewählten Analysesystems IBM WEX zur Umsetzung der im vorangehenden Abschnitt vorgestellten Aufgabenstellungen zur Literaturdatenanalyse bewährt. Aufgrund der leicht bedienbaren, dialogorientierten Benutzeroberfläche konnten die Studierenden nach einer kurzen Einführung eigenständig Analysen durchführen und in Form von projektorientierter Gruppenarbeit vertiefen. Die generierten Ergebnisse – etwa Häufigkeitstabellen für bestimmte Fachbegriffe – konnten aufgrund der bereits vorhandenen Kompetenzen im Umgang mit gängigen Endbenutzerwerkzeugen wie MS Excel weiterverarbeitet und visualisiert werden. Dabei hat sich allerdings auch herausgestellt, dass häufig der Wunsch zur Analyse weiterer Datenquellen (z. B. Nachrichtenströme aus Twitter, Social News aus Reddit) entsteht. Im

Rahmen künftiger Veranstaltungen kann dies durch die Konfiguration bzw. Entwicklung fokussierter Crawler berücksichtigt werden, die allerdings ein erhebliches Maß an informationstechnischen Kompetenzen erfordert.

Diese Ergebnisse werfen insgesamt die Fragestellung auf, in welchem Umfang Datenkompetenzen für die Auswertung unstrukturierter Textdaten im Rahmen wirtschaftswissenschaftlicher Studiengänge zu vermitteln sind. Traditionell konzentriert sich die informationstechnische Ausbildung von Wirtschaftswissenschaftlern auf die Verarbeitung gut strukturierter Daten zur Ermittlung betriebswirtschaftlicher Kennzahlen. Aufgrund der zunehmenden Digitalisierung der Wirtschaft ist indes damit zu rechnen, dass auch die Kompetenzen zur Analyse von Textdaten in den betriebswirtschaftlichen Teildisziplinen von steigender Relevanz für die Beschäftigungsbefähigung von Absolventen sein werden. Infolgedessen sollte kritisch reflektiert werden, inwieweit eine Ergänzung bestehender Curricula sinnvoll ist.

## Literaturverzeichnis

---

1. Alpar, P.; Alt, R.; Bensberg, F.; Grob, H.L.; Weimann, P.; Winter, R.: Anwendungsorientierte Wirtschaftsinformatik. Springer Fachmedien Wiesbaden, Wiesbaden, 2016.
2. Bandara, W.; Miskon, S.; Fiert, E.: A systematic, tool-supported method for conducting literature reviews in information systems. In: Proceedings of the 19th European Conference on Information Systems. Helsinki, 2011.
3. Bartlakowski, K.: Make the library really look more like google. In: Bibliotheksdienst, Vol. 49, 2015, pp. 643–648. doi: 10.1515/bd-2015-0073 (last check 2019-04-18)
4. Bensberg, F.; Auth, G.; Czarnecki, C.: Einsatz von Text Analytics zur Unterstützung literaturintensiver Forschungsprozesse. Arbeitskreis Wirtschaftsinformatik AKWI, 2018, J 1–6. <https://ojs-hslu.ch/ojs302/index.php/AKWI/article/view/128/114> (last check 2019-04-18)
5. Blümle, A.; Lagrèze, W.A.; Motschall, E.: Systematische Literaturrecherche in PubMed. In: HNO, Vol. 66, 2018, Iss. 8, pp 631–648. doi: 10.1007/s00106-018-0538-x (last check 2019-04-18)
6. Bornmann, L.; Mutz, R.: Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. In: J Assoc Inf Sci Technol, 66, 2015, pp. 2215–2222. doi: 10.1002/asi.23329 (last check 2019-04-18)
7. vom Brocke, J.; 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. In: Communications of the Association for Information Systems, Vol. 37, 2015. doi: 10.17705/1CAIS.03709 (last check 2019-04-18)
8. Cardoso, J.; Hoxha, J.; Fromm, H.: Service Analytics. In: Cardoso, J.; Fromm, H.; Nickel, S.; Satzger, G.; Studer, R.; Weinhardt, C. (eds): Fundamentals of Service Systems. Springer International Publishing, Cham, 2015. pp. 179–215.

9. Czarnecki, C.; Auth, G.: Prozessdigitalisierung durch Robotic Process Automation. In: Barton, T.; Müller, C.; Seel, C. (eds): Digitalisierung in Unternehmen: Von den theoretischen Ansätzen zur praktischen Umsetzung. Springer Fachmedien Wiesbaden, Wiesbaden, 2018, pp 113–131
10. Dann, D.; Hauser, M.; Hanke, J.: Reconstructing the Giant: Automating the Categorization of Scientific Articles with Deep Learning Techniques. In: Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik. St. Gallen, 2017, pp. 1538–1549. <https://pdfs.semanticscholar.org/4a9a/d285d789c39e32760c78e5a11c48e523eab0.pdf> (last check 2019-04-18)
11. Döring, N.; Bortz, J.: Forschungsmethoden und Evaluation. Springer Berlin Heidelberg, Berlin, Heidelberg, 2016.
12. Ebener, S.; Ebener, M.: Der Einsatz von Text Mining zur Bestimmung des Diffusionsprozesses von Produkten. In: Gansser O, Krol B (eds) Moderne Methoden der Marktforschung. Springer Fachmedien Wiesbaden, Wiesbaden, 2017, pp. 247–271.
13. van Eck, N.J.; Waltman, L.: Text mining and visualization using VOSviewer. 2011. <https://arxiv.org/abs/1109.2058> (last check 2019-04-18)
14. Feldman, R.; Sanger, J.: The text mining handbook: advanced approaches in analyzing unstructured data. Cambridge University Press, Cambridge, New York, 2007.
15. Grillenberger, A.; Romeike, R.: Developing a theoretically founded data literacy competency model. In: Proceedings of the 13th Workshop in Primary and Secondary Computing Education on - WiPSCE '18. ACM Press, Potsdam, Germany, 2018, pp. 1–10. [https://www.computingeducation.de/pub/2018\\_Grillenberger-Romeike\\_WiPSCE2018.pdf](https://www.computingeducation.de/pub/2018_Grillenberger-Romeike_WiPSCE2018.pdf) (last check 2019-04-18)
16. Heyer, G.; Quasthoff, U.; Wittig, T.: Text Mining: Wissensrohstoff Text: Konzepte, Algorithmen, Ergebnisse. W3L-Verlag, 2006.
17. Hilbert, M.: What Is the Content of the World's Technologically Mediated Information and Communication Capacity: How Much Text, Image, Audio, and Video? In: The Information Society - An International Journal Inf Soc, Vol. 30, 2014, Iss. 2, pp. 127–143. doi: 10.1080/01972243.2013.873748 (last check 2019-04-18)
18. Hippner, H.; Rentzmann, R.: Text Mining. Informatik-Spektrum, Vol 29, 2006, Iss. 4, pp. 287–290. doi: 10.1007/s00287-006-0091-y
19. Jairam, D.; Kiewra, K.A.; Rogers-Kasson, S.; Patterson-Hazley, M.; Marxhausen, K.: SOAR versus SQ3R: a test of two study systems. In: Instructional Science, Vol. 42, 2014, Iss. 3, pp. 409–420. doi: 10.1007/s11251-013-9295-0 (last check 2019-04-18)
20. Kalb, H.; Bukvova, H.: The Digital Researcher: Exploring the Use of Social Software in the Research Process, 2009. [https://pdfs.semanticscholar.org/1e70/9f9549253902b17bd7574f5246438918aa52.pdf?\\_ga=2.18702905.1246137489.1555575760-928083673.1555575760](https://pdfs.semanticscholar.org/1e70/9f9549253902b17bd7574f5246438918aa52.pdf?_ga=2.18702905.1246137489.1555575760-928083673.1555575760) (last check 2019-04-18)
21. Mas-Tur, A.; Modak, N.M.; Merigó, J.M.; Roig-Tierno, N.; Geraci, M.; Capecchi, V.: Half a century of Quality & Quantity: a bibliometric review. In: Qual Quant, Vol. 53, 2019, , Iss. 2, pp. 981–1020. doi: 10.1007/s11135-018-0799-1 (last check 2019-04-18)



22. Schmitz, M.; Dietze, C.; Czarnecki, C.: Enabling Digital Transformation Through Robotic Process Automation at Deutsche Telekom. In: Urbach, N.; Röglinger, M. (eds): *Digitalization Cases: How Organizations Rethink Their Business for the Digital Age*. Springer International Publishing, Cham, 2019, pp. 15–33.
23. Seufert, A.; Oehler, K.: Controlling und Big Data: Anforderungen an die Methodenkompetenz. In: Schäffer, U.; Weber, J. (eds): *Controlling & Management Review. Sonderheft 1-2016*. Springer Fachmedien Wiesbaden, Wiesbaden, 2016, pp 74–81.
24. Shakeel, Y.; Krüger, J., Nostitz-Wallwitz, I. v.; Lausberger, C.; Durand, G.C.; Saake, G.; Leich, T.: (Automated) Literature Analysis - Threats and Experiences. In: *2018 IEEE/ACM 13th International Workshop on Software Engineering for Science (SE4Science)*. 2018, pp 20–27
25. Sturm, B.; Sunyaev, A.: Design Principles for Systematic Search Systems: A Holistic Synthesis of a Rigorous Multi-cycle Design Science Research Journey. In: *Business & Information Systems Engineering*, Vol. 61, Iss. 1, 2019, pp. 91–111. doi: 10.1007/s12599-018-0569-6
26. Webster, J.; Watson, R.T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, Vol. 26, 2002, Iss. 2, pp. xiii-xxiii.
27. Wei-Dong (Jackie) Zhu; Foyle, Bob; Gagné, Daniel; Gupta, Vijay; Magdalen, Josemina; Mundi, Amarjeet S.; Nasukawa, Tetsuya; Paulis, Mark; Singer, Jane; Triska, Martin: *IBM Watson Content Analytics: Discovering Actionable Insight from Your Content*. IBM, 2014.
28. Wilde, T.; Hess, T.: Forschungsmethoden der Wirtschaftsinformatik. In: *Wirtschaftsinformatik*, Vol. 49, 2007, Iss. 4, pp. 280–287. doi: 10.1007/s11576-007-0064-z (last check 2019-04-18)