

Themen und Trends der Wirtschaftsinformatik – Eine Analyse unter Einsatz von Big-Data-Technologie

Jens Kaufmann

Universität Duisburg-Essen, Department of Technology and Operations Management, 47057
Duisburg, jens.kaufmann@uni-due.de

Lars Bathen

Universität Duisburg-Essen, Department of Technology and Operations Management, 47057
Duisburg, lars.bathen@cundus.de

Research in Progress Paper

Abstract

Die Wirtschaftsinformatik unterliegt in Forschung und Praxis starken sowie stark wechselnden Einflüssen durch technischen Fortschritt, betriebswirtschaftliche Interessen und konsequenten Druck der Internationalisierung von Methoden, Lösungen und Publikationen. Eine Unterscheidung der aufgetragenen Themen in langfristige Trends und kurzfristige Moderscheinungen wurde durch einzelne Untersuchungen mit geringem Automatisierungsgrad in der Vergangenheit bereits vorgenommen. Aktuelle Verfahren im Kontext von Big Data stellen Werkzeuge für eine stärker automatisierte und umfangreichere Analyse relevanter Publikationen bereit. Der vorliegende Beitrag zeigt, wie diese Verfahren eingesetzt werden können, um Themen zu identifizieren und zu klassifizieren. Es wird dazu eine erste Einordnung von Publikationen mit Fokus auf die Wirtschaftsinformatik anhand einer prototypischen Umsetzung vorgenommen.

1 Einleitung und Motivation

Eine wissenschaftliche Disziplin muss für einen Erkenntnisgewinn, oder allgemeiner einen messbaren Fortschritt, einen Weg anstreben, der ein langfristiges Ziel verfolgt und nicht um einen status quo oszilliert [9]. Die Wirtschaftsinformatik, insbesondere im deutschsprachigen Raum, versucht daher beständig ein Selbstbewusstsein im eigentlichen Sinne zu erreichen. Es wurden dazu in den vergangenen Jahren immer wieder Studien und Beobachtungen zur aktuellen Themensetzung und Positionsbestimmung der Disziplin durchgeführt [5, 10, 11]. Dabei wurde zum einen auf die Abgrenzung zu verwandten Gebieten der Geistes-, Sozial-, Wirtschafts- und Ingenieurwissenschaften Bezug genommen, zum anderen insbesondere das Verhältnis zum angelsächsischen *Information Systems* thematisiert – sowohl aus methodischer wie auch aus inhaltlicher Sicht [16, 19].

Zur Identifikation von Themenschwerpunkten der Wirtschaftsinformatik existiert eine wiederholt überarbeitete und zitierte Studie [9, 10]. Mertens führt dort eine Stichwort-basierte Untersuchung einer populären, praxis-orientierten Fachzeitschrift (*Computerwoche*) durch und identifiziert unterschiedliche Themen aus dem IT-Umfeld, die maßgebliche Trends (nachhaltige Entwicklungen) oder Moden (kurzfristige, ggf. wiederholende Interessensschwerpunkte) darstellen. Ein anderer Beitrag zeigt durch eine umfangreiche Literaturanalyse, dass sich Wirtschaftsinformatik und Information Systems bei den Themenschwerpunkten erkennbar unterscheiden [16]. So weist die Wirtschaftsinformatik einen in der Regel geringeren Abstraktionsgrad und eine größere Themenvielfalt auf. Die Autoren betonen, dass das Spektrum der untersuchten Zeitschriften breiter ist als bei Mertens, fokussieren im Gegenzug aber auf wissenschaftlich anerkanntere Publikationen.

In beiden Fällen wurde eine händische Analyse durchgeführt, die Titel, Schlagwörter und Abstract der Artikel beinhaltet, diese auf Schlüsselwörter untersucht und nur mit hohem Aufwand durchgeführt werden konnte [9], was eine schnelle Aktualisierung der Analysen behindert. Zeitschriftenartikel in natürlicher Sprache, z.B. als HTML- oder PDF-Dokumente, stellen im Sinne der maschinellen Auswertbarkeit semi- oder unstrukturierte Daten dar [4]. Die analytische Verarbeitung großer Datenmengen dieser Art, wird derzeit im Fachgebiet Business Intelligence unter dem Begriff *Big Data* diskutiert. Die bisher gängigste Big-Data-Definition liefert Russom mit Verweis auf Gartner [8] mit dem 3V-Modell: „Size matters, but there are other important attributes of big data, namely data variety and data velocity“ [14]. Big-Data-Technologien umfassen somit neben Techniken der Verarbeitung großer Datenmengen (Volume) auch den Umgang mit semi-strukturierten Daten (Variety) und deren stetige Veränderung (Velocity). Eine Prüfung der Einsetzbarkeit dieser Methoden auf die oben beschriebene Problemstellung liegt nahe.

Im Gegensatz zu ihrer angelsächsischen Schwesterdisziplin ist die Wirtschaftsinformatik stark konstruktiv oder gestaltungsorientiert ausgerichtet. Die bei diesem Ansatz geschaffenen Artefakte, z.B. Software, sollen zwar der Überprüfung von vorher aufgestellten Thesen dienen [6], das Prototyping, also ein schneller Entwurf mit anschließender Implementierung, kann aber auch neue Erkenntnisse aus dem Erstellungsprozess selbst heraus generieren [19]. Für den vorliegenden Beitrag wurde daher ein prototypischer Ansatz zur Prüfung des Einsatzes von Big-Data-Technologien gewählt. In Abschnitt 2 werden zunächst Forschungsfragen und –vorgehen detailliert. Anschließend wird ein neuentwickelter Prototyp vorgestellt, der mit Hadoop das derzeit populärste Framework zur Aufbereitung von Big Data [15] verwendet (Abschnitt 3). Die Ergebnisse der Analyse werden in Abschnitt 4 thematisiert. Abschnitt 5 gibt einen Ausblick auf die zukünftige Forschung im Hinblick auf den vorliegenden Research-in-progress-Beitrag und schließt mit einem Fazit.

2 Forschungsfragen und Forschungsvorgehen

Für den Beitrag ergeben sich auf Basis der bisherigen Überlegungen zwei Forschungsfelder. Das fachlich-inhaltlich orientierte Forschungsfeld setzt sich mit den Themen der Wirtschaftsinformatik, soweit sie als Ergebnis einer Analyse identifiziert werden können, auseinander. Es werden dazu drei Forschungsfragen formuliert:

- a1. Was sind aktuelle Themen in der Wirtschaftsinformatik und welche Trends oder Moden sind dabei erkennbar?
- a2. In welchem Maße unterscheiden sich diese Themen, Trends und Moden bei der differenzierten Betrachtung wissenschaftlicher und praxis-orientierter Publikationen?
- a3. Inwieweit besteht eine Abhängigkeit (z.B. temporaler Art) der Themenbehandlung über die genannten Publikationsarten?

Das technisch orientierte Forschungsfeld nimmt Bezug auf die Eignung von Big-Data-Technologie (konkret realisiert durch Hadoop) zur Beantwortung der fachlich-inhaltlichen Fragestellungen und deren Verallgemeinerbarkeit auf andere Themenfelder. Es ergeben sich zwei Kernfragestellungen:

- b1. Ist das Framework Hadoop ein geeignetes Werkzeug zur Bearbeitung der fachlichen Fragestellungen und wenn ja, mit welchen Elementen des Frameworks?
- b2. Liefert der Einsatz im Vergleich zu klassischen BI-Methoden einen Mehrwert für die Analyse, insbesondere bei Ressourceneinsatz, zeitlichem Umfang und Erkennung neuer Muster?

Zur Durchführung der Untersuchung mit Hilfe eines Prototyps wird die Design Science Research Methodology herangezogen [13] (siehe Bild 1). Problemidentifikation, Motivation und Zielbeschreibung wurden bereits gegeben. Design und Umsetzung der prototypischen Lösung werden in Abschnitt 3 behandelt und demonstriert. Eine Evaluation der Ergebnisse findet mit der Beantwortung der Fragen b1 und b2 statt und wird letztlich im vorliegenden Beitrag vorgestellt.

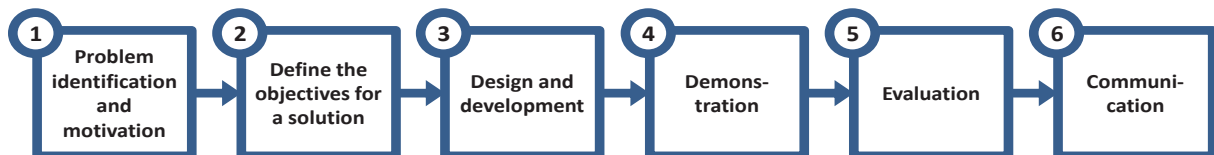


Bild 1: Design Science Research Methodology (nach Peffers [13])

Die initiale Auswahl des Datenbestandes besteht im vorliegenden Fall aus der Auswahl der zu verwendenden Publikationen nach Spektrumsbreite, Relevanz und Verfügbarkeit. Um sowohl praxis-orientierte wie auch wissenschaftliche Publikationen zu berücksichtigen, sind zunächst die Computerwoche (CW), die wöchentlich erscheint, sowie die Reihe HMD – Praxis der Wirtschaftsinformatik (HMD), die zwei-monatig mit Schwerpunktheften erscheint, berücksichtigt worden. Die Relevanz der Medien ist dabei durch Auflage und/oder wissenschaftliches Ranking belegt [20]. Die Publikationen sind elektronisch bis in die Jahre 1974 (CW), und 2000 (HMD) verfügbar und ermöglichen so eine Erkennung auch länger anhaltender Trends oder wiederkehrender Muster. Zudem besteht eine Vergleichbarkeit der Ergebnisse zu den genannten bisherigen Arbeiten.

3 Technische Umsetzung des Prototyps

Die zu Grunde liegenden Daten erfüllen nicht die vollständige Definition des 3V-Modells. Geprüft wird daher explizit der Einsatz von Big-Data-Technologie, die eben auch ‚Variety‘ gezielt behandeln kann, für die beschriebenen semi-strukturierten Daten. Im Vergleich zu klassischen Text-Mining-Werkzeugen, wie z.B. der KNIME Text Processing Engine, die grundsätzlich für die Fragestellung geeignet wären, bietet die Verwendung des Big-Data-Frameworks Hadoop für die Implementierung komfortablere und umfangreichere Möglichkeiten. Stärken von Hadoop sind insbesondere Schnittstellen, Verarbeitung der Rohdokumente und Skalierbarkeit. Dazu kommen die kostenlose Verfügbarkeit, die Möglichkeit einer einfachen Integration in alle relevanten analytischen Informationssysteme (Datenbanken, BI-Tools, Excel), die hohe Skalierbarkeit und Plattformunabhängigkeit der Software, die Möglichkeit, eine Vielzahl von Varianten gespeicherter Daten auszuwerten – Beispiele sind Pig, HQL, R und MapReduce-Jobs in einer Vielzahl von Programmiersprachen – sowie die Verwendbarkeit von Mahout als umfangreicher und etablierter Data-Mining-Suite zur Auswertung.

Für die tatsächliche Implementierung wurde auf die Distribution von Microsoft HDInsight [12] zurückgegriffen. Diese erleichtert Konfiguration sowie Management der Softwareumgebung erheblich und bietet eine zuverlässige Anbindung an die Microsoft-Produktpalette, insbesondere an den etablierten SQL Server, was eine leichte Überprüfbarkeit der Ergebnisse gewährleistet.

Zur Aufbereitung der Rohdaten wurde hier das MapReduce-Framework von Hadoop eingesetzt. Dieses erlaubt es, die Daten aus den Rohdatendateien massiv-parallel zu verarbeiten [3] und im Hadoop File System (HDFS) zu speichern [18]. Dort werden die Daten in Form von standardisierten Textdateien abgelegt. So harmonisiert lassen sich mit MapReduce-Algorithmen Analysen auf dem Datenbestand durchführen. Die Analyseergebnisse werden in Hive, der relationalen Engine von Hadoop, abgelegt und sind von dort über eine standardisierte Schnittstelle mit gängigen Analysewerkzeugen auslesbar (vgl. Bild 2).

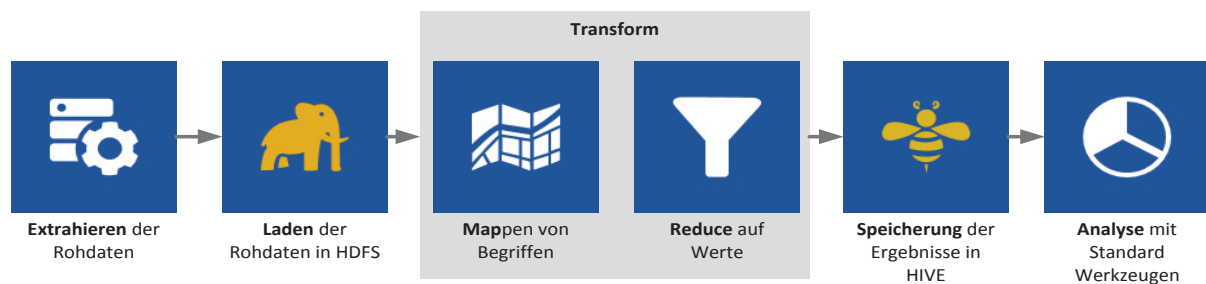


Bild 2: Schematischer Datenverarbeitungsprozess

Die Anbindung unterschiedlicher Datentypen wurde bei der Einbindung der Datenquellen berücksichtigt. Die Artikel der Computerwoche lagen als HTML-Dokumente vor, die Daten der Reihe HMD als PDF-Dokumente. In beiden Fällen wurden die Volltexte von den Metadaten (Autor, Datum, Schlagworte, etc.) separiert und in die Datenbank geladen. Der eigentliche Analyseprozess basierte auf einer Wörterbuch-gestützten Schlagworterkennung und -zählung, die die Daten aufbereitete und bereinigte, und setzte damit Verfahren ein, die im Text Mining üblich sind [17]. Technisch wurden mehrere MapReduce-Jobs verwendet, deren Grundprinzip eine Parallelisierung von Aufgaben über die vorliegenden Daten und die anschließende Zusammenführung der Ergebnisse ist [3].

4 Analyse der bisherigen Ergebnisse

In der manuellen Textanalyse kann ein Katalog von gefundenen Begriffen, die relevant für die Wirtschaftsinformatik sind, explorativ erstellt werden. Eine rein computergestützte Analyse bedingt dagegen die (vorherige) Festlegung von Suchbegriffen. Als Basis-Wörterbuch für die Suche wurden sämtliche Schlagwörter der Enzyklopädie der Wirtschaftsinformatik [7] als redaktionell geführtes und regelmäßig aktualisiertes Lexikon verwendet. Hier sei auf die kritische Betrachtung dieses Vorgehens in Abschnitt 5 verwiesen. Als Betrachtungszeitraum wurde der Zeitraum von 2000 bis 2012 gewählt, den beide betrachteten Publikationen vollständig abdecken. Es wurde eine Schlagwortsuche nach allen Begriffen im Volltext der Artikel durchgeführt und so die Entwicklung der Begriffe über die Jahre aufgezeigt. Für eine Einschätzung der Begriffe im Sinne ihrer Entwicklung wurde auf die Einteilung von Steiniger et al. [16] zurückgegriffen, die Trends, negative Trends, Moden, wiederkehrende Moden, oszillierende Themen und sonstige Verläufe als idealtypische Verläufe anhand des Vergleichs mit einer linearen Regression über die Anzahl gefundener Worte identifizieren. Tabelle 1 zeigt die jeweiligen Top 20 Schlagwörter der betrachteten Publikationen und ihre Klassifikation. Schattiert gekennzeichnet sind die Begriffe, die in beiden Fällen unter der Top 20 sind.

#	CW			HMD		
	Begriff	rel. Häuf.	Klassifikation	Begriff	rel. Häuf.	Klassifikation
1	System	11,58%	Sonstige	System	12,23%	Sonstige
2	Software	11,17%	Neg. Trend	Anwendung	7,22%	Osz. Thema
3	Internet	5,37%	Mode	Software	6,66%	Osz. Thema
4	Anwendung	5,13%	Sonstige	Wissen	4,19%	Osz. Thema
5	Microsoft	4,92%	Sonstige	Internet	3,91%	Wiederk. Mode
6	SAP	4,85%	Mode	Modell	3,64%	Osz. Thema
7	NC	2,60%	Neg. Trend	SAP	2,13%	Osz. Thema
8	Applikation	2,52%	Neg. Trend	Applikation	1,95%	Osz. Thema
9	Cloud Computing	2,17%	Sonstige	Outsourcing	1,88%	Sonstige
10	Plattform	2,16%	Wiederk. Mode	Plattform	1,75%	Sonstige
11	Oracle	2,04%	Sonstige	CRM	1,73%	Osz. Thema
12	ERP-System	2,00%	Osz. Thema	Cloud Computing	1,73%	Sonstige
13	E-Mail	1,93%	Wiederk. Mode	ERP-System	1,69%	Sonstige
14	Betriebssystem	1,91%	Sonstige	SOA	1,62%	Sonstige
15	Java	1,86%	Sonstige	Netzwerk	1,62%	Osz. Thema
16	Modell	1,76%	Wiederk. Mode	XML	1,54%	Sonstige
17	CRM	1,73%	Sonstige	Portal	1,49%	Sonstige
18	Portal	1,41%	Neg. Trend	Transaktion	1,45%	Sonstige
19	Outsourcing	1,35%	Mode	Governance	1,34%	Sonstige
20	Business Intelligence	1,24%	Osz. Thema	Wiki	1,24%	Sonstige

Tabelle 1: Top 20 Schlagwörter Computerwoche und HMD (2000 bis 2012)

Deutlich erkennbar ist die hohe Übereinstimmung (13 von 20) der Schlagwörter sowie eine ähnliche Verlaufskurve der relativen Häufigkeiten, wenngleich die Reihenfolge der vier Top-Begriffe System, Software, Internet und Anwendung differiert. Die Untersuchung der nicht übereinstimmenden Schlagwörter hingegen kennzeichnet deutlich die Art der Publikation. So weist die CW eine deutlich stärkere Nennung von Anbieter- und Produktnamennamen (Microsoft, Oracle, Java) auf, während in den HMD-Publikationen weitere Technologie- und Methodenbezeichnungen geführt werden (SOA, Netzwerk, XML, Governance), was eine neutralere, wissenschaftsorientiertere Darstellung stützt.

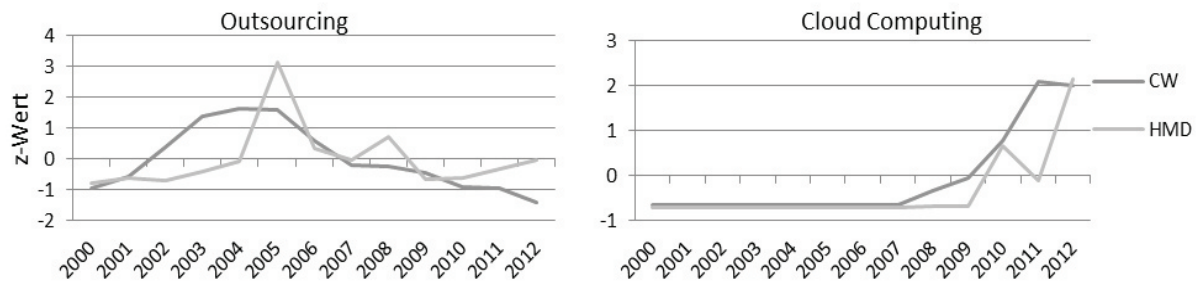


Bild 3: Ausgewählte Themen im zeitlichen Verlauf im Vergleich (standardisierte Werte [16])

Der Vergleich der Zeitverläufe zweier Begriffe über die Publikationen (Bild 3) wurde mit Hilfe einer Standardisierung auf z-Werte erzeugt [1]. Er zeigt eine typische Eigenschaft, die auch bei anderen Begriffen zu finden ist und sich in der Klassifikation der Begriffe widerspiegelt. Obwohl Outsourcing und Cloud Computing in ihrer Tendenz (einmal steigend und wieder fallend, einmal steigend) gleichartig erscheinen, wirkt der Verlauf in der CW im Vergleich zu HMD geglättet. Eine Erklärung dafür kann die größere Anzahl an Ausgaben pro Jahr sein (52 gegen 6), so dass einzelne Schwerpunkte weniger stark gewichtet werden. Eine ausführliche Analyse über alle Themen und mit mehreren Publikationen ist hier in jedem Fall vonnöten.

5 Fortgang der Forschung und Fazit

Die prototypische Implementierung von Big-Data-Technologie hat in kurzer Zeit gezeigt, dass eine technische Analyse der Daten möglich ist und auswertbare Ergebnisse generiert. Des Weiteren bieten die Methodik und die technische Plattform die Möglichkeit, weitere Quellen unter Beibehalt der Analysewerkzeuge anzubinden. In Bezug auf die technisch-orientierten Forschungsfragen kann daher konstatiert werden, dass Hadoop als Big-Data-Technologie geeignet erscheint, um die durchgeführten Analysen zu unterstützen und dass durch den Einsatz der Technik der zeitliche Aufwand und die personelle Belastung gering gehalten werden können.

Der bisherige Umfang von lediglich zwei Publikationen kann nur einen ersten Einblick in die Thematik der Wirtschaftsinformatik bieten (Forschungsfrage a1), insbesondere da bisher noch keine umfangreiche Datenbasis streng wissenschaftlich-orientierter Literatur verwendet wurde. Bei der Bewertung der Ergebnisse ist zu berücksichtigen, dass hier das Aufkommen von Begriffen untersucht wird, nicht die Relevanz der Thematik im Alltag. So zeigt sich, dass sich die Wirtschaftsinformatik stark und recht konstant mit allgemeinen Themen wie ‚System‘ und ‚Anwendung‘ befasst. Die Kennzeichnung von ‚Internet‘ (Position 3, bzw. 5) als (wiederkehrende) Mode macht aber auch deutlich, dass das Thema zwar relevant ist, als eigenständiger Forschungsgegenstand aber offenbar nur zu bestimmten Zeiten von Interesse war.

Generell weisen die bisher identifizierten Themen eine Übereinstimmung mit den bisherigen Untersuchungen auf. Sie unterscheiden sich, wie bei Betrachtung einer aktuelleren Zeitspanne zu erwarten, in Rang, Reihenfolge und Klassifikation der Themen (z.B. wird Outsourcing bei [16] auf Platz 4 als „Sonstiges“ geführt, ERP-System auf Platz 5 als „Wiederkehrende Mode“). Die dedizierte Analyse der unterschiedlichen Publikationsarten ist auf Basis weiterer Quellen und entsprechender Auswertungen durchzuführen, dies stellt sich vor dem bisherigen Erkenntnisstand als vielversprechend dar, da die bisherigen Ergebnisse auf Unterschiede schließen lassen (Frage a2). Speziell der zeitliche Verlauf der Themen (Bild 3) lässt eine zeitliche Verschiebung zwischen praxisorientierten und Publikationen mit wissenschaftlichem Anspruch vermuten. Es ist weiter zu analysieren, ob dies eine generelle Eigenschaft ist und, wenn ja, wie diese begründbar wäre (Frage a3). Denkbar ist z.B., dass die in der Wissenschaft üblichen Review-Prozesse die Publikation von Themen, die gleichzeitig in Wissenschaft und Praxis Interesse wecken, gegenüber Praxiszeitschriften verlangsamen.

Die bisherigen Ergebnisse zeigen, vor allem in fachlicher Hinsicht, nur einen kleinen Ausschnitt der möglichen Analysen. Kritisch zu hinterfragen ist die Verwendung eines Schlagwortkatalogs. Dieser basiert i.d.R. auf bereits etablierten Themen. Wird dazu ein redaktionell geführter Katalog verwendet, besteht die Gefahr, kurzfristige Moden zu ‚verpassen‘ oder vollständig zu ignorieren. Ebenfalls ist die Semantik der Begriffe zu prüfen. So wird z.B. das ‚System‘ nicht nur im WI-Kontext verwendet. Für die vorliegende Arbeit wird eine kontextuelle Fehlinterpretation durch die Auswahl fachspezifischer Datenbasen im Wesentlichen verhindert. Sollen umfangreichere Werke berücksichtigt werden, sind ggf. aufwändigere Verfahren vonnöten. Dies ist auch der Fall, wenn Synonyme, Flexionen und Internationalisierung (z.B. Anglizismen) mit einbezogen werden sollen. Dem erhöhten Mehraufwand dieser Konfiguration im Vergleich zur manuellen Abarbeitung der Quellen und der automatischen Berücksichtigung der Parameter durch den fachkundigen Bearbeiter steht die vielversprechende Möglichkeit gegenüber, eine wesentliche umfangreichere Menge an Quellen zu berücksichtigen. Die Ergebnisse dieser und zukünftiger Arbeiten können dann nicht nur der Wirtschaftsinformatik im Speziellen, sondern einer automatischen Trend- und Modenanalyse im Allgemeinen dienen, wie sie in der Konsumgüterindustrie durch die Analyse von (digitalen, sozialen) Medien gefordert wird [2].

6 Literatur

- [1] Bortz, J; Schuster, C (2010): Statistik für Human- und Sozialwissenschaftler. 7. Auflage. Springer, Berlin.
- [2] Chen, H; Chiang, Roger H. L.; Storey, VC (2012): Business intelligence and analytics. *MIS Quarterly* 36(4): 1165–1188.
- [3] Dean, J; Ghemawat, J (2004): "MapReduce: Simplified Data Processing on Large Clusters". http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en/archive/mapreduce-osdi04.pdf. Abgerufen am 17.08.2013.
- [4] Felden, C (2006): Text Mining als Anwendungsbereich von Business Intelligence. In: Chamoni, P; Gluchowski, P (Hrsg), *Analytische Informationssysteme*. 3. Auflage. Springer, Berlin [u.a.].
- [5] Heilmann, H; Heinrich, LJ (2006): Erkenntnisobjekte der Wirtschaftsinformatik. In: Fröschle, H; Strahringer, S (Hrsg), *HMD - Praxis der Wirtschaftsinformatik*. dpunkt.verlag.
- [6] Hevner, AR; March, ST; Park, J; Ram, S (2004): Design Science in Information Systems Research. *MIS Quarterly* 28(1): 75–105.

- [7] Kurbel, K; Becker, J; Gronau, N; Sinz, E; Suhl, L (2013): "Enzyklopädie der Wirtschaftsinformatik". <http://enzyklopaedie-der-wirtschaftsinformatik.de>. Abgerufen am 12.09.2013.
- [8] Laney, D (2001): "3D Data Management: Controlling Data Volume, Velocity and Variety.". <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>. Abgerufen am 12.12.2013.
- [9] Mertens, P (1995): Wirtschaftsinformatik: Von den Moden zum Trend. In: König, W (Hrsg), *Wirtschaftsinformatik '95*. Physica-Verl, Heidelberg.
- [10] Mertens, P (2006): Moden und Nachhaltigkeit in der Wirtschaftsinformatik. Arbeitsbericht des Lehrstuhls für BWL, insbes. Wirtschaftsinformatik I, Universität Erlangen-Nürnberg.
- [11] Mertens, P; Barbian, D (2012): Materialien zum Forschungs- und Diskussionsthema „Grand Challenges“. Arbeitsbericht, Universität Erlangen-Nürnberg.
- [12] Microsoft (2013): "Microsoft Big Data". <http://www.microsoft.com/en-us/sqlserver/solutions-technologies/business-intelligence/big-data.aspx>. Abgerufen am 12.09.2013.
- [13] Peffers, K; Tuunanen, T; Rothenberger, MA; Chatterjee, S (2007): A Design Science Research Methodology for Information Systems Research. *JMIS* 24(3): 45–77.
- [14] Russom, P (2011): Big Data Analytics. tdwi.
- [15] Russom, P (2013): Integrating Hadoop Into Business Intelligence And Data Warehousing. tdwi.
- [16] Steininger, K; Riedl, R; Roithmayr, F; Mertens, P (2009): Moden und Trends in Wirtschaftsinformatik und Information Systems. *Wirtschaftsinformatik* 51(6): 478–495.
- [17] Weiss, SM (2005): Text mining. Springer, New York.
- [18] White, T (2012): Hadoop. 3. Auflage. O'Reilly, Farnham, Sebastopol, Calif.
- [19] Wilde, T; Hess, T (2007): Forschungsmethoden der Wirtschaftsinformatik. *Wirtschaftsinformatik* 49(4): 280–287.
- [20] WKWI (2008): WI-Orientierungslisten. *Wirtschaftsinformatik* 50(2): 155–163.