

Henning Baars (Hrsg.)

**Impulse für die Forschung oder
Impulse durch die Forschung: Entwicklungsperspektiven
der BI-Forschung in Deutschland**

Dritter Workshop „Business Intelligence“

der GI-Fachgruppe Business Intelligence
in Zusammenarbeit mit dem Lehrstuhl für ABWL und
Wirtschaftsinformatik I der Universität Stuttgart

am 13. und 14. Oktober 2011
in Stuttgart

Tagungsband

Vorwort

Das Bild der Business Intelligence (BI) wird zunehmend unschärfer. Während sich im praktischen Betrieb zentralisierte, mehrschichtige Architekturen durchgesetzt zu haben scheinen, werden diese in prominenten Praxis- und Forschungsbeiträgen z.T. grundlegend hinterfragt. In den propagierten neuen Datenhaltungs- und Datenanalysekonzepten verschwimmt die Grenze zwischen operativer und dispositiver Welt. Vermehrt drängen neue Inhalte und Datenformate in die BI, während gleichzeitig das Portfolio an Analyse- und Visualisierungstechniken sowie an Ausgabekanälen und -endgeräten vielfältiger wird.

Ohne Frage entsteht hier Bedarf für wissenschaftlich abgesicherte Konzepte auf Basis einer *anwendungsorientierten* Perspektive, wie sie der deutschsprachigen Wirtschaftsinformatik inhärent ist. Erforderlich sind hierfür grundlegende Erkenntnisse über Informationsbedarfe und Informationsnutzung ebenso wie spezifische Ansätze, Methoden und Architekturen für eine effektive Gestaltung von BI-Lösungen sowie für BI-spezifische Organisationskonzepte. Es ist zu prüfen, inwieweit die BI-Forschung auf etablierte Modelle und Theorien zurückgreifen kann, an welchen Stellen diese adaptiert werden müssen und in welchen Feldern eigene Modelle und Theorien entwickelt werden müssen.

Ziel des Workshops dritten Workshops „Business Intelligence“ der Fachgruppe Business Intelligence der Gesellschaft für Informatik war es, innovative Forschungsansätze und Forschungsergebnisse aus den Themenfeldern Business Intelligence (BI) bzw. Management Support Systeme (MSS) zu präsentieren, zu diskutieren und in Bezug zu setzen. Vorgestellt werden sollen neben originären Forschungsergebnissen bewusst auch Zwischenergebnisse aus Forschungsprojekten (*Research in Progress*), Forschungsideen sowie neue methodische Herangehensweisen.

Dieser Tagungsband enthält die angenommenen und auf dem Workshop vorgestellten fünf Beiträge sowie die Ergebnisse der Eröffnungsdiskussion zum übergeordneten Rahmenthema.

Im Mittelpunkt der Eröffnungsdiskussion standen dabei Fragen zur Positionierung der Business-Intelligence-Forschung in der Wirtschaftsinformatik: Welche Kompetenzen zeichnen die BI-Forschung in der WI besonders aus, welche Forschungsrichtungen erscheinen aus Forschungs- und aus Praxissicht besonders vielversprechend und welche Grenzen und Herausforderungen sollten adressiert werden.

Den zunehmenden Überlappungen zwischen den operativen und den dispositiven Systemlandschaften widmet sich der Beitrag von Tom Hänel und Carsten Felden am Beispiel der Manufacturing Execution Systems (MES). Er zeigt die Potentiale einer Zusammenführung von MES mit Systemen des Operational BI für die integrierte operative Prozesssteuerung.

Robert Krawatzek, Frieder Jacobi, André Müller und Marcus Hofmann diskutieren in Ihrem Artikel, wie der zunehmenden Dynamik in BI-Systemen durch eine automatisierte Erstellung von Systemdokumentationen begegnet werden kann. Sie entwickeln hierzu ein Framework, das auf dem Ansatz der modellgetriebenen Softwareentwicklung beruht.

Jens Lachenmaier analysiert die Potentiale einer Zusammenführung von technischen Daten und Produktmerkmalen mit betriebswirtschaftlichen Key Performance Indicators (KPIs) und schlägt hierfür den Aufbau eines produktorientierten Data Warehouses vor.

Kerstin Schäfer und Oliver Günther widmen sich in ihrem Full-Research-Paper der Methodenwahl und Zusammenstellung im Bereich der zielgerichteten Optimierung großer Webangebote. Sie kombinieren dabei Ansätze des Relational Marketing mit solchen des Web-Mining.

Mit der Analyse von Meinungsbildern in Micro-Blogs setzen sich Andreas Schieber, Kai Heinrich und Andreas Hilbert auseinander. Sie schlagen eine Sequenz verschiedener Verfahren vor – und zwar für die Themenextraktion, die Filtrierung von Meinungsführern, die Extraktion von Themen und zur Abbildung von Meinungstrends im Zeitverlauf.

Inhalte

Henning Baars: Impulse für die Forschung oder Impulse durch die Forschung: Entwicklungsperspektiven der BI-Forschung in Deutschland	1
Tom Hänel, Carsten Felden: Manufacturing Execution Systems und Operational Business Intelligence	7
Robert Krawatzek, Frieder Jacobi, André Müller, Marcus Hofmann: Konzeption eines Frameworks zur automatisierten Erstellung nutzerspezifischer IT-Systemdokumentationen	15
Jens F. Lachenmaier: Development of a Concept to integrate technical Product Data and Business Data: Motivation, conceptual Framework, and consecutive Research	27
Kerstin Schäfer, Oliver Günther: Evaluating the Effectiveness of Relational Marketing Strategies in Corporate Website Performance	35
Andreas Schieber, Kai Heinrich, Andreas Hilbert: Analyse von Konsumentenmeinungen in Microblogs: Topic-based Opinion Mining	50

Impulse für die Forschung oder Impulse durch die Forschung: Entwicklungsperspektiven der BI-Forschung in Deutschland

**Ergebnisse der Eröffnungsdiskussion auf dem 3. Workshop Business
Intelligence der GI-Fachgruppe WI-BI**

Henning Baars

Universität Stuttgart

Lehrstuhl für ABWL und Wirtschaftsinformatik 1

Abstract

„Business Intelligence“ (BI) ist in der deutschen Wirtschaftsinformatik (WI) als Forschungsfeld fest etabliert. Die BI-Forschung ist speziell in Deutschland durch eine hohe Anwendungsorientierung, eine starke Praxisrelevanz sowie gestaltungsorientierte Forschungsansätze gekennzeichnet. Im Rahmen des „3. Workshops Business Intelligence“ wurde diskutiert, welche Kompetenzen die BI-Forschung in der WI besonders auszeichnen, welche Forschungsrichtungen sowohl aus Forschungs- wie auch aus Praxis­sicht besonders vielversprechend erscheinen sowie welche Grenzen und Herausforderungen sie adressieren sollte. Als vorläufiges Fazit kann konstatiert werden, dass es v.a. der integrative Aspekt der BI ist, der im Zentrum der WI-Forschung steht und dass dieser auch Bezugspunkte zu anderen Forschungsgebieten aufzeigt. Gerade angesichts der hohen Reife der BI-Ansätze in der Praxis ist Prototyping ein vielsprechendes Instrument zur überzeugenden Kommunikation innovativer Forschungsideen – allerdings eines, das wohlüberlegt und sehr gezielt eingesetzt werden muss. Auf Seiten der Forschung bringt die inhärente Breite des BI-Gegenstandes erhebliche Herausforderungen bei der Abgrenzung und Publikation scharf definierter Forschungsfragen mit sich.

1 Business Intelligence als Forschungsfeld

Im Rahmen der Eröffnungsdiskussion des dritten „Workshops Business Intelligence“ am 13.10.2011 wurden die Relevanz, die Abgrenzung und die Ausrichtung der Business-

Intelligence-Forschung in der deutschen Wirtschaftsinformatik diskutiert. Teilnehmer waren 14 Vertreter von sechs Universitäten und einem BI-Beratungshaus.

Hierbei wurde zunächst noch einmal konstatiert, dass sich das Thema „Business Intelligence“ (BI) als Forschungsfeld der Wirtschaftsinformatik fest etabliert hat. Als Begriff hat BI mittlerweile das ältere Management Support Systems (MSS) abgelöst, wobei im BI-Begriff eine Reihe von neuen Akzentuierungen mit schwingen. Diese Überlegungen haben u.a. auch zur Umbenennung der Fachgruppe im Frühjahr 2011 geführt – deren ausführliche Begründung in einem Positionspapier festgehalten ist, das auf der Webseite des Fachbereiches abgerufen werden kann. Besonders hervorzuheben sind dabei die Ausrichtung auf eine breite, nicht nur auf das Management bezogene Entscheidungsunterstützung, sowie der Aspekt der Integration. Letzterer umfasst eine technische, konzeptionelle und organisatorische Perspektive (Baars et al., 2011).

Die integrative Ausrichtung unterstützt gemäß den Diskussionsteilnehmern auch bei der Abgrenzung zur Operations-Research- und Data-Mining-Community, in denen die Gestaltung und Evaluation einzelner analytischer Methoden im Mittelpunkt stehen und weniger deren Einbettung in den infrastrukturellen und organisatorischen Kontext. Es wurde jedoch von einzelnen Teilnehmern argumentiert, dass die Auswahl und Kombination bestimmter Methoden bzw. Algorithmen für einen bestimmten Anwendungskontext sehr wohl auch Gegenstände der BI-Forschung bleiben sollten.

Darüber hinaus wurde hervorgehoben, dass ein vergleichbares Begriffsverständnis zwar auch in der Praxis gängig ist, punktuell aber auch eine sehr weit gefasste Nutzung des Begriffs vorgefunden werden kann (BI als generischer „Umbrella Term“).

Unabhängig davon ist das Thema BI (im Sinne der Fachgruppe) in den Unternehmen von hoher Relevanz. Aus Sicht der Diskussionsteilnehmer haben ein dort vorherrschendes Bewusstsein um noch zu adressierende Herausforderungen und bislang unzureichend ausgeschöpfte Potentiale, zu einer besonders fruchtbaren Zusammenarbeit von Praxis und Wissenschaft beigetragen. Das in den Unternehmen aufgebaute Know-How bietet dabei einen wichtigen Resonanzboden für *relevante* anwendungsorientierte Forschung. Aufgesetzt werden kann insbesondere auf etablierte Architektur- und Organisationskonzepte, zu denen in mehreren Jahrzehnten ein stabiler Wissensfundus erarbeitet wurde. Hierzu hat v.a. auch die Wirtschaftsinformatik maßgeblich beigetragen.

2 Stärken und Herausforderungen der WI-basierten BI-Forschung

Die universitäre BI-Kompetenz wird auf Seiten der Praxis wahrgenommen und gewürdigt, was sich beispielsweise auch in einer engen Zusammenarbeit mit Praxisorganisatio-

nen wie dem TDWI Germany niederschlägt. Die Universitäten fungieren als Ideenlieferanten, agieren als Multiplikatoren von best-practices, verdichten und strukturieren relevantes Wissen und erarbeiten und kommunizieren innovative Lösungskonzepte.

Im Kontext der BI ist es v.a. der inhärent ganzheitlich-systemische Ansatz, der als besondere Stärke der Wirtschaftsinformatik wahrgenommen wird. Neben der Abgrenzung zur Informatik und zum Operations Research, deren Fokus eher auf der Gestaltung und der Evaluation einzelner Algorithmen und Verfahren liegt, ist auch eine klare Abgrenzung zu den Disziplinen des Controlling und der Organisation erforderlich. Hier bestehen breite Überlappungsbereiche: Die Konzeption von BI-Lösungen ist insbes. oft verschränkt mit der Erarbeitung von Kennzahlenkonzepten und Ansätzen für das Corporate Performance Management. Grundsätzlich sind es aus Sicht der Mehrheit der anwesenden Teilnehmer nicht die *Inhalte*, die von der BI-Forschung gestaltet werden, sondern die *Systeme*, mit denen diese umgesetzt bzw. transportiert werden. Darin müssen zwar sehr wohl fachliche und domänenspezifische Zusammenhänge *abgebildet* werden, es ist aber nicht Aufgabe der BI-Forschung, spezifische fachliche Konzepte zu *erarbeiten*.

Daneben ist zu beachten, dass die Abgrenzung zwischen dispositiven und operativen Systemen zunehmend verwischt, was zu Herausforderungen bei der Positionierung von Forschungsarbeiten führen kann. Es dürfte eigentlich nicht überraschen, dass so genannte „Operational-BI-Systeme“ oftmals gar nicht die wirtschaftlich sinnvollste Lösung für die operative Entscheidungsunterstützung bieten. Hier sind häufig bereits etablierte „klassische“ operative Systeme (OLTP) vorzufinden, die diese Aufgabe effizient und effektiv erfüllen. BI kommt primär dann zum Tragen, wenn eine (Daten-)Integration angestrebt wird und diese mit den aufgabenspezifischen, operativen Systemen nicht adäquat umgesetzt werden kann. Gerade in Disziplinen wie der Produktion und der Logistik, mit teilweise jahrzehntelang gereiften IT-Lösungen, ist sorgsam abzuwägen, an welchen Stellen tatsächlich die Kompetenzen der BI-Forschung relevant werden.

Daneben birgt der eigentlich erfreuliche „Demand-Pull“ nach Forschungsleistungen auch diverse Risiken. Diese entstehen v.a. bei einer Verwechslung von nur scheinbar „neuen“ Anwendungen mit echtem Erkenntniszuwachs oder bei einer unmittelbaren Adaption fertiger Lösungskonzepte auf neue Anwendungsdomänen. Grundsätzlich besteht das Risiko eines Verweilens in wissenschaftlich uninteressanten „Brot-und-Butter-Themen“, da hier auf Praxisseite in vielen Unternehmen noch Nachholbedarf besteht. Zu Fragen etwa hinsichtlich der Relevanz einer Schichtenarchitektur, zu den Potentialen eines Enterprise DWHs, zur multidimensionalen Modellierung u. ä. liegt bereits ein gefestigter Erkenntnisblock vor. Selbst wenn hier in einigen Unternehmen noch *Handlungsbedarf* besteht, ist ein *Forschungsbedarf* besonders gründlich zu prüfen.

Relativiert wird diese Aussage allerdings durch schleichende Änderungen im BI-Umfeld wie etwa der veränderten Bereitschaft zur Nutzung von Outsourcing-Dienstleistungen und Netzdiensten (z.B. Cloud-BI) oder die gestiegenen IT-Kompetenzen im Top-Management. Auch scheinbar gefestigte Erkenntnisse sind deshalb immer wieder kritisch zu hinterfragen.

Schließlich bringt es der technische Fortschritt mit sich, dass einzelne Forschungsthemen in Schichten unterhalb der WI „abrutschen“ können (z.B. bei Ablösung performanceoptimierender multidimensionaler Snow-Flake-Modellierung durch hardwarebasierte In-Memory-Lösungen). Die Aktualität der in Forschung (und Lehre) behandelten Themen ist deshalb immer wieder neu zu hinterfragen.

3 Forschungsartefakte und deren Praxisrelevanz

Zunächst einmal wurde konstatiert, dass auf akademischem Niveau qualifizierte BI-Fachkräfte ein relevanter Output der deutschen Wirtschaftsinformatik sind. Die Funktion der *Lehre* ist somit nicht zu vernachlässigen.

Daneben werden nach Ansicht der Teilnehmer sehr wohl auch die Forschungsartefakte der Wirtschaftsinformatik ernst genommen. Die Rolle der Forschung kommt dabei gerade im Rahmen der Konzeptualisierung, der Begriffsschärfung und der Strukturierung relevanter Zusammenhänge zum Tragen. Auch die Dokumentation von best practices sowie (durchaus auch deskriptiv orientierter) Empirie wird geschätzt. Anders sieht dies mit theorieorientierten empirischen Studien aus, wie sie in der angelsächsischen IS-Forschungspraxis gängig sind. So wichtig eine theoriegeleitete Forschung auch sein mag, so sollte bei der Wahl des Abstraktionsgrades der Bezug zu den als relevant empfundenen Herausforderungen der Praxis nicht verloren gehen.

Kontrovers diskutiert wurde die Praxisrelevanz von Organisationskonzepten, da hier in den Unternehmen oftmals bereits feste Vorstellungen vorhanden sind, die nicht zuletzt von schwer beeinflussbaren betrieblichen Realitäten geformt werden. Auf gute Resonanz stoßen hingegen nach Ansicht einzelner Diskussionsteilnehmer Modelle und Modellnotationen. Ebenfalls positiv wahrgenommen wird überdies Forschung zu aufkommenden Standards und Regelwerken.

Eine besondere Bedeutung hat die Exploration innovativer Anwendungsszenarien und Technologiekombination, die auch in den Unternehmen als wesentlicher Impulsgeber wahrgenommen wird. Gerade hier werden Universitäten als neutrale Instanzen gesehen, die anders als Berater oder Systemhäuser kein kommerzielles Interesse an der Verbreitung bestimmter Konzepte oder Technologien haben.

4 Konstruktionsorientierte Forschung und die Rolle von Prototypen

Der zuletzt angesprochene Punkt führt auch zur besonderen Rolle konstruktionsorientierter Forschung und zur Bedeutung von Prototypen. Prototypen sind wesentliche Vehikel zur Exploration der Machbarkeit und – stärker noch – zur Kommunikation von Lösungskonzepten. Die Festlegung, wie nahe der Prototyp einem Produktivsystem kommen muss, stellt dabei eine Gratwanderung dar. Dies ist vom Zweck abhängig. Während in einigen Fällen einfache Oberflächen-Mockups („Power-Point-Prototypen“) ausreichen können, um Kernideen zu transportieren, ist es in anderen Fällen erforderlich, funktionsfähige und in Umsysteme integrierte Software zu erstellen (z.B. um potentielle aber kritische Anwender überzeugen zu können).

Die Gefahr, dass eine solche Entwicklung dabei allerdings zum Selbstzweck wird, wobei schnell beachtliche Forschungsmittel vergeudet werden, ist dabei allerdings groß. Zudem sollte sichergestellt werden, dass Forschung nicht als kostengünstige Entwicklungsleistung missverstanden wird – ein Ansatz, von dem im Übrigen bereits aus Kompetenzerwägungen abgesehen werden sollte.

5 Herausforderungen im Forschungsdesign

Wie eigentlich immer in der Wirtschaftsinformatik-Forschung stellt die Herausarbeitung des theoretischen Beitrags der Forschung eine besondere Herausforderung in der BI dar. Hierbei kommt der in der Wirtschaftsinformatik ausführlich diskutierte Konflikt zwischen rigoroser und relevanter Forschung besonders zum Tragen, da die integrative Ausrichtung der BI eine Fokussierung auf eng abgegrenzte Forschungsfragen oft erheblich erschwert.

Die im „Memorandum für eine gestaltungsorientierte Wirtschaftsinformatik“ (Österle et al., 2010) festgehaltene Ausrichtung auf konstruktionsorientierte Forschung ist hierbei ein wesentlicher Anker bei der Entwicklung, Konkretisierung und Strukturierung von Forschungsarbeiten. Dieses befreit jedoch nicht von der Abstraktion konstruktionsorientierter Ergebnisse. Die Frage, welcher Erkenntniszuwachs aus einer BI-Lösung für einen bestimmten Fall oder eine bestimmte Branche resultiert und wo die Generalisierungsmöglichkeiten liegen, ist nicht immer einfach zu beantworten. Aus dem Teilnehmerkreis kam hier der Vorschlag, die Entwicklung von Referenzlösungen und -architekturen zu verfolgen, die als solche sehr wohl als wissenschaftlicher Beitrag gewertet werden könnten. Eng verknüpft mit dem konstruktionsorientierten Ansatz sind die in der internationalen Forschung immer noch weit verbreiteten Vorbehalte gegenüber qualitativer Forschung –

zu der jedoch in den genannten relevanten Gegenstandsbereichen oftmals keine befriedigende Alternative existiert.

Schließlich wurde diskutiert, Forschungsthemen anhand von Querschnittsthemen aus der Wirtschaftsinformatik aufzugliedern und die BI-Forschung hierbei von Standardthemen zu bereinigen, die bereits an anderer Stelle auf übergreifendem Niveau beantwortet wurden. Im Zuge dieses Abgleichs können darüber hinaus Fragestellungen herausgelöst werden, die über die BI hinweg von Relevanz sein könnten. Grundsätzlich wurde der BI-Forschung das Potential zugesprochen, auch die übergreifenden Disziplinen Wirtschaftsinformatik und Information Systems zu befruchten.

6 Fazit

Der Gegenstandsbereich „Business Intelligence“ zeichnet sich primär durch seine integrative Ausrichtung aus. Der Aspekt der Integration stellt auch den Kern der BI-Forschung der Wirtschaftsinformatik dar und ist ein wesentliches Kriterium zur Abgrenzung zu benachbarten Domänen (z.B. Operations Research oder Controlling) und Anwendungsfeldern (insbes. operative Systeme). Die BI-Forschung wird zum einen als wesentlicher Impulsgeber für die Praxis verstanden und hat das Potential wesentliche Forschungserkenntnisse zu erzielen. Besondere Bedeutung kommt konstruktionsorientierter Forschung zu.

Literaturverzeichnis

- Baars, H., Rieger, B., Finger, R., Gluchowski, P., Hilbert, A. & Kemper, H.G. February 2011: *Umbenennung der GI-Fachgruppe „Management Support Systems“ (MSS) in „Business Intelligence“ – Positionspapier des Leitungsgremiums der GI-Fachgruppe Management Support Systems*. URL: [http://www.fg-mss.gi-ev.de/fileadmin/gliederungen/fg-mss/Dateien/GI_FG_BI_Positionspapier Umbenennung_2011.pdf](http://www.fg-mss.gi-ev.de/fileadmin/gliederungen/fg-mss/Dateien/GI_FG_BI_Positionspapier_Umbenennung_2011.pdf)
- Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis & A., Sinz, E.J. 2010: *Memorandum zur gestaltungsorientierten Wirtschaftsinformatik*. In: Österle, H., Winter, R. & Brenner, W. (Hrsg.): *Gestaltungsorientierte Wirtschaftsinformatik: Ein Plädoyer für Rigor und Relevanz*. StGallen 2010.

Manufacturing Execution Systems und Operational Business Intelligence

Zur Notwendigkeit einer integrierten Betrachtung

Tom Hänel, Carsten Felden

*Professur für ABWL insb. Informationswirtschaft/Wirtschaftsinformatik
TU Bergakademie Freiberg*

Abstract

Manufacturing Execution Systems (MES) und Operational Business Intelligence (OpBI) analysieren und koordinieren betriebliche Abläufe in unterschiedlichen Segmenten der Wertschöpfungskette eines Unternehmens. Dies birgt die Gefahr, dass eine Verinselung der Infrastruktur innerhalb der operativen Entscheidungsunterstützung erfolgt. Daher ist das Ziel der Forschung, inwieweit eine gemeinsame Betrachtung von OpBI und MES ein Potenzial im Kontext einer unternehmensweiten Steuerung und Analyse von operativen Prozessen birgt, sodass Unternehmen sich flexibel an sich wandelnde Rahmenbedingungen anpassen können. In diesem Beitrag werden die Forschungsidee und ein methodisches Raster zur Zusammenführung von MES und OpBI in einer operativen Integrationsplattform skizziert. Erste Arbeitsergebnisse einer literaturbasierten Begriffsklassifikation motivieren eine intensivere Auseinandersetzung des Forschungsansatzes, um Erkenntnisse für eine prozessorientierte, BI-gestützte Entscheidungsunterstützung zu gewinnen.

1 Einleitung

Die Gestaltung von Geschäftsprozessen ist ein entscheidender Wettbewerbsfaktor. Dabei werden nicht nur die Prozesse innerhalb eines Unternehmens durch die zunehmende Vernetzung beeinflusst. Es ändern sich ebenfalls die Beziehungen zu Lieferanten und Kunden. Dies zwingt Unternehmen dazu, Geschäftsprozesse flexibel zu halten und immer wieder anzupassen. Konzepte, die wirksame Unterstützung anbieten, sind vorhanden

(z. B. SOA/service-oriented Business Intelligence (Erl, 2005), Real-Time Analytics (Melchert et al., 2004), Active Data Warehousing (Imhoff, 2001)). In den vergangenen zehn Jahren sind dazu ergänzend Manufacturing Execution Systems (MES) und Operational Business Intelligence (OpBI) parallel aus unterschiedlichen Blickrichtungen in die Diskussion getreten. Beide Ansätze versprechen Unterstützung in der Prozessflexibilität, argumentieren jedoch aus ihrer jeweiligen Perspektive heraus – der ingenieurwissenschaftlichen und der entscheidungsunterstützungsorientierten Sichtweise. Im Unklaren verbleibt jedoch, inwiefern OpBI und MES überlappend sind oder zumindest solche Ähnlichkeiten aufweisen, als dass eine kombinierte Betrachtung einen höheren Beitrag für die Flexibilität von Geschäftsprozessen und der Unterstützung unternehmerischer Tätigkeiten leistet. Ein Bedarf an entsprechenden Lösungen ist erkennbar, weil z. B. gemäß einer Studie der Aberdeen Group 85 Prozent der befragten Unternehmen mit ihren Anwendungen keine adäquate Flexibilität erreichen können (Rodriguez, 2007). Das Ziel des Dissertationsprojektes ist eine Zusammenführung von MES und OpBI unter Berücksichtigung der Gemeinsamkeiten aber auch der Unterschiede, um zu einer interdisziplinären, prozessorientierten und flexiblen Entscheidungsunterstützung beizutragen.

OpBI stellt mithilfe von BI-Techniken wie Data Warehousing oder OLAP Echtzeitinformationen zur Verfügung, um Geschäftsprozesse zu koordinieren (Marjanovic, 2007; Cunningham, 2005). Hauptanwendungsgebiete sind Vertrieb und Marketing (Eckerson, 2007). Innerhalb von Produktionsumgebungen erfolgt die IT-Unterstützung der Prozesse vorrangig durch ingenieurwissenschaftlich-orientierte Konzepte, insbesondere dem MES, welche die Abläufe von Produktionsanlagen in Echtzeit organisieren (MESA, 1997a). OpBI und MES sind integrative Ansätze, die eine flexible Echtzeitsteuerung von Geschäftsprozessen fördern. Dabei kommen sie jedoch in verschiedenen Segmenten der Wertschöpfungskette zum Einsatz. In diesem Kontext leistet die Forschung durch die Zusammenführung von MES und OpBI in einer operativen Integrationsplattform einen Beitrag hinsichtlich einer unternehmensweiten Steuerung und Analyse der Geschäftsprozesse.

Kapitel 2 stellt MES und OpBI vor, ordnet die Konzepte in die Organisationsstruktur eines Unternehmens ein und diskutiert Folgen einer gemeinsamen Betrachtung für die Integration der Informationsverarbeitung. Ein methodisches Raster wird in Kapitel 3 beschrieben. Kapitel 4 präsentiert erste Arbeitsergebnisse. Abschließend erfolgen Zusammenfassung und Ausblick.

2 Status Quo

2.1 Operational Business Intelligence und Manufacturing Execution Systems

OpBI und MES analysieren Geschäftsprozesse, um Schwachstellen, Fehlfunktionen oder Störfälle zu erkennen. Unternehmen können damit ihre Abläufe kontinuierlich steuern, überwachen und verbessern. In die Struktur eines Unternehmens lassen sich die vorgestellten Konzepte wie folgt einordnen:

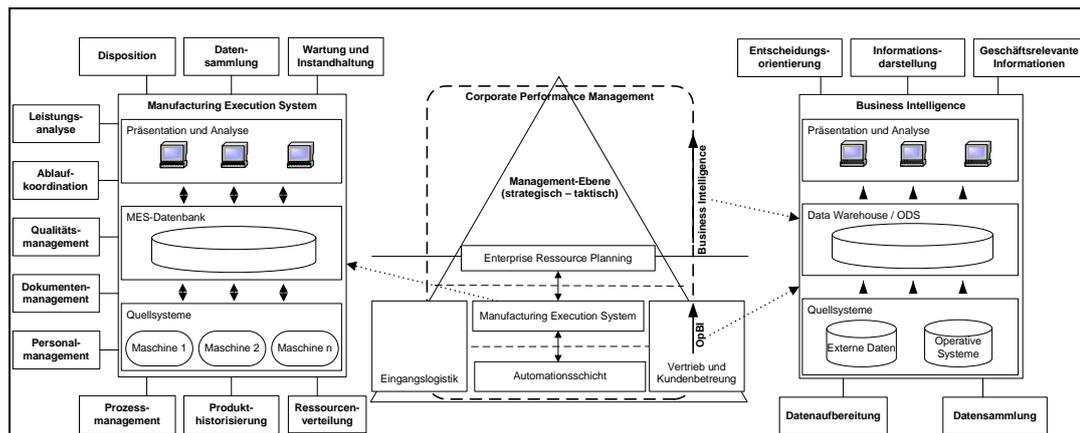


Abbildung 1: Organisationsstrukturelle Einordnung der Konzepte

BI und OpBI bedienen sich Funktionen, die auf Aspekte der Datensammlung, Datenaufbereitung sowie der Informationsdarstellung hinsichtlich Entscheidungsorientierung und geschäftlicher Relevanz fokussieren (Schrödl, 2006). Daten werden aus operativen Quellsystemen in ein zentrales Data Warehouse übertragen (Gangadharan und Swamy, 2004) und z. B. mit OLAP oder Data Mining analysiert (Gluchowski et al., 2008). Die Ergebnisse fließen im Weiteren in die dispositive Entscheidungsunterstützung des strategischen und taktischen Managements ein. Konzepte wie Corporate Performance Management betrachten die dispositive und operative Ebene im Zusammenspiel (Golfarelli et al., 2004). Neben der integrierten Sammlung und Auswertung von operativen Daten im dispositiven Bereich ist ein Steuerungseffekt der operativen Prozesse in zumindest teilautomatisierter Form verbunden. In diesem Kontext findet sich das Konzept der OpBI wieder, welches vorrangig auf die Steuerung von Prozessabläufen auf Basis von Prozesskennziffern zielt. Dies bringt neben Automation und Leistungsmessung von Geschäftsprozessen auch Trends zur Echtzeitanalyse mit sich (Melchert et al., 2004). Hinsichtlich des Anwendungskontextes weiß OpBI einen betriebswirtschaftlichen Fokus auf und findet wenig Beachtung im Fertigungsbereich (Eckerson, 2007).

MES übernehmen die Prozesssteuerung und damit auch die Entscheidungsunterstützung im Fertigungsbereich (Younus et al., 2010). Diese Systeme agieren zwischen dem unternehmensweiten ERP-System sowie der fertigungsspezifischen Automatisierungsebene (ISA 2000) und stellen den Soll-Vorgaben der ERP-Systeme permanent die Ist-Werte der Produktion gegenüber. Funktional müssen MES dazu abhängig von der Komplexität des Produktionsprozesses vielseitig ausgelegt sein. Die MESA (MESA, 1997) schlägt dafür elf Kernfunktionen vor (siehe linker Teil der Abbildung 1).

OpBI und MES ermöglichen beide eine integrierte Bereitstellung von Daten sowie deren zielgerichtetes Berichten und Auswerten in unterschiedlichen Unternehmensbereichen. Ein MES könnte zusätzlich die Aufgaben von OpBI ausführen. Dazu müsste eine höhere Schnittstellenanzahl berücksichtigt werden. Derartige Komplexitätssteigerungen können die Systemleistungsfähigkeit einschränken (Saenz de Ugarte et al., 2009). Zusätzlich stellen die begrenzten Analysefähigkeiten eines MES (Alpar und Louis, 2007) den Nutzen einer solchen Strategie infrage. OpBI dagegen fehlt es an der funktionalen Vielfältigkeit von MES, die speziell auf Produktionsumgebungen ausgerichtet sind (Meyer et al., 2009).

Im Kontext ihrer Anwendungsspezifik können sich OpBI und MES nicht gegenseitig ersetzen, aber im Hinblick auf unternehmensweite Entscheidungsunterstützung nutzenbringend ergänzen. Dazu sind die Konzepte gemeinsam zu betrachten, da sonst Insellösungen das Resultat sind. Eine Integrationsplattform zur Kombination von Informationen aus Lieferkette, Produktion und Kundenbetreuung stellt eine Zusammenführung dieser Verinselungen dar, wobei die spezifischen Funktionalitäten innerhalb der Anwendungsgebiete erhalten bleiben. Prozessleistungskennziffern erfahren eine Anreicherung mit Informationen aus der gesamten Wertschöpfungskette. Dies ermöglicht eine übergreifende Synchronisation der Geschäftsprozesse mit flexibilitätssteigernder Wirkung. Durch die Erhöhung der Transparenz in den operativen Prozessen sind Verbesserungen im Bereich der Leistungsfähigkeit der Prozesse und im Kontext der Prozessgestaltung denkbar.

2.2 Folgen für die Integration der Informationsverarbeitung

In diesem Abschnitt werden die Folgen für die Ausprägungen der Integrierten Informationsverarbeitung (vgl. Mertens 2009) auf Grund einer gemeinsamen Betrachtung von MES und OpBI diskutiert. Die betroffenen Ausprägungen sind grau hinterlegt. Die Kennzeichnung resultiert aus der Zielstellung des Dissertationsprojektes, zu einer interdisziplinären, prozessorientierten und flexiblen Entscheidungsunterstützung beizutragen.

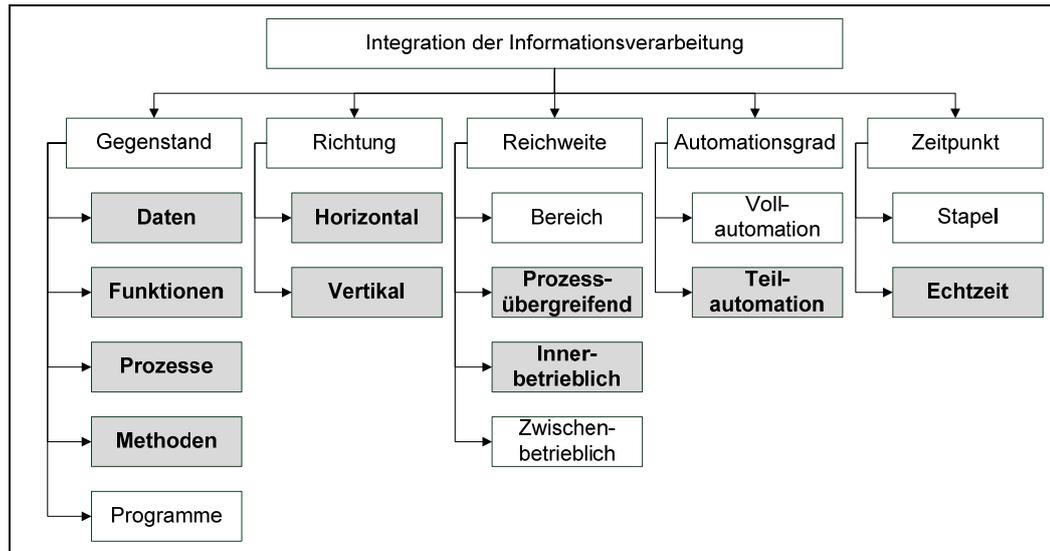


Abbildung 2: Ausprägungen der Integrierten Informationsverarbeitung
(Entnommen und modifiziert nach Mertens 2009)

Gegenstand der Integration sind Daten, Funktionen, und Methoden, um der Forderung nach einer flexiblen Entscheidungsunterstützung nachzukommen. Ein Entscheidungsunterstützungssystem verlangt das Vorhandensein einer Daten-, Methoden- und Modellbank. Der Begriff der Flexibilität erfordert die Anpassungsfähigkeit hinsichtlich sich ändernder Funktions- und Leistungsanforderungen (Heinrich et al. 2004). Der interdisziplinäre Charakter kennzeichnet sich durch horizontale und vertikale Integrationsrichtung sowie durch die prozessübergreifende und innerbetriebliche Integrationsreichweite. Prozessorientierung bedeutet die Ausrichtung des Unternehmens an Geschäftsprozessen (Heinrich et al. 2004) mit dem Ziel, deren Leistungsfähigkeit und die Flexibilität zu Gunsten einer transparenten Prozessbeherrschung zu steigern (Deszteler, 2000). Daher sind auch Prozesse Gegenstand der Integration. Um die Flexibilität auch im Hinblick auf eine erhöhte Anpassungsgeschwindigkeit zu verbessern, erscheint eine teilautomatisierte Integration in Echtzeit sinnvoll.

3 Methodik

Das methodische Raster lässt sich in drei Bereiche differenzieren. Im ersten Bereich wird der Stand der Forschung ermittelt. Dazu erarbeitet ein State-of-the-Art (siehe dazu grundlegend Fettke, 2006) eine literaturbasierte Klassifikation der Begriffe MES und OpBI, um existierende Forschungsergebnisse hinsichtlich einer gemeinsamen Betrachtung der Konzepte zu ermitteln. Im zweiten Bereich erfolgt mithilfe von Fallstudien (siehe dazu grundlegend Yin, 2009) eine Anforderungsbestimmung für eine operative Integrationsplattform

von OpBI und MES zu Gunsten einer Steuerung und Analyse von operativen Geschäftsprozessen. Dabei werden zu berücksichtigenden Prozesse abstrakt definiert sowie erforderliche Datenobjekte und Informationsflüsse analysiert. Auf Basis der Anforderungen erfolgt eine Ableitung von Hypothesen bezüglich eines Modells zur theoretischen Fundierung der Integrationsplattform, die auch einem Hypothesentest zuzuführen sind. Im dritten Bereich wird auf Basis der qualitativen Erkenntnisse eine Implementierung und Validierung der konzipierten Integrationsplattform angestrebt.

4 Erste Ergebnisse

Zum gegenwärtigen Zeitpunkt ist die Literaturrecherche abgeschlossen und publiziert (Hänel und Felden, 2011). Für den zweiten Schritt ist zur Planung der Fallstudien mit der Erarbeitung eines Forschungsprotokolls begonnen worden. Dieses geht auf Problemstellung und Zielsetzung, Definition und Auswahl der Fälle, die anzuwendenden Datenerhebungsmethoden sowie auf die Durchführung einer exemplarischen Studie ein (Borchardt und Göthlich, 2006).

Die Literaturrecherche konnte vier Publikationen identifizieren, die eine gemeinsame Betrachtung von BI und MES thematisieren. Die Beiträge berücksichtigen die multidimensionale Analyse von Produktionsdaten und die Rolle des MES im Rahmen unternehmensweiter Entscheidungsfindung. Dies sind jedoch nur einzelne Facetten der Integration. Der Zusammenhang zwischen MES und OpBI wird lediglich in einer Publikation gestreift. Die unterschiedlichen Wurzeln der Konzepte bieten hierfür ein Erklärungspotenzial: demnach ist die Wahrnehmung von MES im entscheidungsorientierten Umfeld erst in letzter Zeit stärker ausgeprägt. Andererseits zeigt OpBI eine geringe Menge an Suchergebnissen, was auf eine Neuartigkeit des Begriffs schließen lässt. Eine weiterführende Auseinandersetzung mit dem Thema bietet die Möglichkeit, zur Schärfung des Begriffsverständnisses beizutragen.

5 Fazit und Ausblick

MES und OpBI sind nicht identisch – sie liefern aber beide Informationen im Zuge der Prozessanalyse in unterschiedlichen Arbeitsgebieten. Die Konzepte lassen damit eine gemeinsame Betrachtung für eine flexible operative Entscheidungsunterstützung zu, wobei der Integrationsaspekt bisher kaum wissenschaftlich betrachtet ist.

Die Ergebnisse der Literaturrecherche motivieren eine weitere Betrachtung des Themas entsprechend der vorgeschlagenen Methodik. Dabei besteht die Möglichkeit, neue Erkenntnisse bezüglich der Analyse und Steuerung von Geschäftsprozessen im Kontext des

aktuellen Tagesgeschehens zu sammeln. Die Integration von OpBI und MES zu einer operativen Informationsplattform unterstützt die Fähigkeit, schnell und flexibel auf Geschäftsereignisse zu reagieren und damit die unternehmensindividuelle Wettbewerbsfähigkeit zu steigern.

Literaturverzeichnis

- Alpar, P. und Louis, J.P. (2007). Eine empirische Untersuchung der Softwareunterstützung bei der Fertigung und Qualitätssteuerung - Implikationen für Manufacturing Execution Systeme. *Philipps-Universität Marburg*. Marburg.
- Borchardt, A. und Göthlich, S. (2009). Erkenntnisgewinn durch Fallstudien. In Albers, S. et al. (Hrsg.), *Methodik der empirischen Forschung*(S. 33 – 48). Wiesbaden: Gabler.
- Cunningham, D. (2005). Aligning Business Intelligence with Business Processes. *What Works (TDWI)*. 20, S. 50-51.
- Derszteler, G. (2000). Prozeßmanagement auf Basis von Workflow-Systemen: Ein integrierter Ansatz zur Modellierung, Steuerung und Überwachung von Geschäftsprozessen. Lohmar: EUL.
- Eckerson, W.W. (2007). Best Practices in Operational BI: Converging Analytical and Operational Processes. *TDWI Best Practices Report*. Renton: WA.
- Erl, T. (2005). *Service-Oriented Architecture: Concepts, Technology, and Design*. Upper Saddle River: Prentice Hall.
- Felden, C., Chamoni P. und Linden M. (2010 May 3-5). From Process Execution towards a Business Process Intelligence. In Abramowicz, W. and Tolksdorf, R. (Hrsg.), *Business Information Systems 13th International Conference*(S. 195-206). Berlin.
- Fettke, P. (2006). State-of-the-Art des State-of-the-art – Eine Untersuchung der Forschungsmethode ‘Review’ innerhalb der Wirtschaftsinformatik. *Wirtschaftsinformatik*. 4, S. 257-266.
- Gangadharan, G. R. und Swamy, N. S. (2004). Business intelligencesystems: design and implementation strategies. *Proceedings of 26th International Conference on Information Technology Interfaces*. Cavtat.Croatia.
- Gluchowski, P., Gabriel, R. und Dittmar, C. (2008). *Management Support Systeme und Business Intelligence: Computergestützte Informationssysteme für Führungskräfte und Entscheidungsträger*. Heidelberg: Springer.

- Golfarelli, M., Rizzi, S. und Cella, I. (2004). Beyond data warehousing: what's next in business intelligence? *Proceedings of 7th ACM international workshop on Data warehousing and OLAP*. New York: ACM Press. S. 1-6.
- Hänel, T. und Felden, C. (2011). Limits or Integration? – Manufacturing Execution Systems and Operational Business Intelligence. *AMCIS 2011 Proceedings*.
- Heinrich, L.J., Roithmayer, F. und Heinzl, A. (2004). Wirtschaftsinformatik-Lexikon. 7. überarbeitete Auflage. München: Oldenbourg.
- Imhoff, C. (2001). Active Data Warehousing the Ultimate Fulfillment of the Operational Data Store. *Intelligent Solutions, Inc.* Boulder.
- ISA ANSI/ISA-95.00.01-2000 (2000). Enterprise Control System Integration, Part 1: Models and Terminology. *ISA technical paper*.
- Marjanovic, O. (2007). The Next Stage of Operational Business Intelligence: Creating New Challenges for Business Process Management. *Proceedings of the 40th Annual Hawaii International Conference on System Sciences*. Washington DC: IEEE Computer Society.
- Melchert, F., Winter, R. und Klesse, M. (August 2004). Aligning Process Automation and Business Intelligence to Support Corporate Performance Management. *Proceedings of the Tenth Americas Conference on Information Systems*. New York.
- Mertens, P. (2009). Integrierte Informationsverarbeitung 1: Operative Systeme in der Industrie. 17. überarbeitete Ausgabe. Wiesbaden: Springer.
- MESA. (1997). MES Functionalities and MRP to MES Data Flow Possibilities. *MESA International – White Paper Number 2*. Pittsburgh.
- Meyer, H., Fuchs, F. und Thiel K. (2009). Manufacturing Execution Systems (MES): Optimal Design, Planning, and Deployment. Columbus (OH): McGraw Hill.
- Rodriguez, R. A. (2007). Aligning IT to Business Processes: How BPM is Complementing ERP and Custom Applications. *Aberdeen Group*.
- Saenz de Ugarte, B., Artiba, A., und Pellerin, R. (2009). Manufacturing execution system – a literature review. *Production Planning and Control*. 20, 6, S. 525-539.
- Schrödl, H. (2006). Business Intelligence. München, Wien: Hanser.
- Yin, R. K. (2009). Case Study Research: Design and Methods. Los Angeles: SAGE.
- Younus, M., Peiyong, C., Hu, L. und Yuqing, F. (2010). MES Development and Significant Applications in Manufacturing – A Review. *2nd International Conference on Education Technology and Computer (ICETC)*.

Konzeption eines Frameworks zur automatisierten Erstellung nutzerspezifischer IT-Systemdokumentationen

Robert Krawatzek, Frieder Jacobi, André Müller, Marcus Hofmann

Technische Universität Chemnitz

Fakultät Wirtschaftswissenschaften, Professur Wirtschaftsinformatik II

Reichenhainer Str. 70, D-09126 Chemnitz

Abstract

Die Komplexität von BI-Systemen als heterogene IT-Systeme steigt in Folge einer immer dynamischeren Unternehmensumwelt zunehmend. Softwaredokumentationen helfen, trotz steigender Systemkomplexität einen detaillierten Überblick über die Arbeitsweise von BI-Systemen zu erhalten. Der Erstellungsprozess von qualitativ hochwertigen Dokumentationen ist jedoch aufwändig und wird – unter Vernachlässigung nachträglicher Änderungen und Anpassungen – meist nur in der Planungs- oder Entwicklungsphase von BI-Systemen vorgenommen. Dieser Beitrag beschreibt den Entwurf eines Frameworks, welches durch die Anwendung des Ansatzes der modellgetriebenen Softwareentwicklung eine vollständig automatisierte Erstellung von nutzerspezifischen Softwaredokumentationen erlaubt, von denen sowohl BI-Systementwickler als auch Fachanwender profitieren.

1 Problemidentifizierung und Motivation

Der Begriff Business Intelligence (BI) beschreibt einen analytischen Prozess, der die Sammlung, Aufbereitung und Präsentationen von unternehmens- und marktbezogenen Daten unter Verwendung mathematischer Modelle und analytischer Methoden unterstützt, dessen Ergebnisse in ökonomische Entscheidungsprozesse einfließen. Softwarelösungen, die im BI-Umfeld zum Einsatz kommen, werden als Business Intelligence Systeme (BI-Systeme) (Vercellis, 2009) bezeichnet.

Durch die Verwendung von BI-Systemen bei taktischen und strategischen Unternehmensentscheidungen entsteht ein hoher Anspruch auf Korrektheit der präsentierten Informationen. Fehler und daraus resultierende Fehlentscheidungen können weitreichende Folgen haben.

Die Anforderungen an ein BI-System sind durch ein komplexes und dynamisches Unternehmensumfeld geprägt und unterliegen daher im Verlauf der Zeit Veränderungen. Das führt zu einem permanenten Anpassungsprozess des Systems, der von den Datenquellen über das zentrale Data Warehouse und Data Marts bis hin zu den Berichten reicht. Dieser Anpassungsprozess kann durch geeignete Systemdokumentationen positiv beeinflusst werden.

Die Dokumentation von Softwaresystemen hat die Aufgabe, deren Arbeitsweise – sowohl aus einer technischen als auch aus einer Endanwendersicht – zu beschreiben (K. C. Laudon & J. P. Laudon, 2010). Wie bei anderen Software-Produkten verringert sich ohne eine ausreichende und qualitativ gute Dokumentation auch bei einem BI-System dessen Nutzen, denn es lässt sich unter Umständen nur schlecht oder gar nicht beurteilen, ob die Software den Anforderungen genügt (Wallmüller, 2001, S. 150).

Dokumentationen, die in frühen Phasen eines Softwareprojekts, wie dem Design, der Konzeption oder der Entwicklungsphase erstellt und nach Änderungen nicht aktualisiert wurden, veralten, was letztendlich zu einem sinkenden Nutzen der Dokumentationen führt (Forward & Lethbridge, 2002; Hofmann, Gluchowski, Jacobi, & Kurze, 2011; Wallmüller, 2001). Weitere Ursachen für eine unzureichende Dokumentation bilden qualitative Mängel, wie schlecht lesbare oder schwer zu verstehende Inhalte, was in gegenwärtigen Softwareprojekten häufig vorkommt (Wallmüller, 2001, S. 149). Eine vollständig veraltete, nicht sinnvoll nutzbare Dokumentation entspricht dem Fehlen einer Dokumentation. Re-Dokumentation versucht dieses Problem zu lösen, indem es die nachträgliche Erzeugung einer Dokumentation von existierenden Systemen, für die keine Dokumentation verfügbar ist, vorschlägt (Chikofsky & Cross, J. H., 1990, S. 15).

BI-Systeme sind heterogen aufgebaut und bestehen aus mehreren, meist unterschiedlichen Systemen, was die Nachvollziehbarkeit und Transparenz ohne eine ausreichende Dokumentation im Vergleich zu monolithischen Softwaresystemen zusätzlich erschwert. Gerade in einem derart komplexen System ist es wichtig, eine gute Dokumentation bereitzustellen, um das Vertrauen in das System zu stärken. Steht bei monolithischen Systemen die Dokumentation der Systeme *selbst* im Vordergrund, so existieren bei heterogenen Systemen weitere Anforderungen an den Umfang einer Dokumentation: neben der jeweiligen Systembeschreibung der *einzelnen beteiligten* Systeme wird auch eine Beschreibung der *statischen Gesamtarchitektur* sowie der zwischen den beteiligten Systemen existierenden *Datenflüsse* benötigt.

Der Umstand, dass eine Dokumentation im Kontext von BI-Systemen häufig nur einmalig beim Entwurf oder der Entwicklung stattfindet, birgt die Gefahr einer mangelnden Dokumentation. Spätere Dokumentationsanpassungen verursachen oft einen hohen Auf-

wand (Hofmann et al., 2011). Ein Praxisprojekt bei einem großen deutschen Tourismusanbieter zeigte, dass es Probleme mit der Dokumentation des BI-Systems geben kann, wenn „beim Aufbau von BI-Systemen häufig sehr spontan vorgegangen und eine konzeptionelle Systemplanung zugunsten eines iterativen Prototyping eher vernachlässigt“ wird (Gluchowski & Kurze, 2010, S. 672). Es wurde ebenfalls gezeigt, dass sich mit Hilfe von automatisierten Dokumentationsprozessen der Aufwand für die Erstellung von Dokumentationen von Datenstrukturen (Spezifika der Datenhaltungskomponenten) reduzieren lässt (Gluchowski & Kurze, 2010, S. 676).

Die in diesem Paper dargestellte Forschungsarbeit zielt auf die Erhöhung des Nutzens eines BI-Systems durch die Bereitstellung aktueller und qualitativ hochwertiger Dokumentationen. Dadurch können sowohl die verschiedenen Adressaten – Fachanwender, Entwickler und Manager (Pomberger & Blaschek, 1996) – als auch das durch das BI-System unterstützte Unternehmen wie folgt profitieren.

Fachanwender erhalten Unterstützung bei der *sicheren und funktionsgerechten Handhabung* des Systems. Das wird durch eine hohe Transparenz bezüglich der Datenqualität, d.h. der Herkunft, Verarbeitung und Aktualität der durch das System bereitgestellten Informationen ermöglicht.

Entwickler erhalten Informationen über den aktuellen technischen Aufbau, um das System kennenzulernen und beurteilen zu können. Das ermöglicht Entwicklern ein schnelleres und leichteres Umsetzen von *Änderungen und Erweiterungen*. Dem Entwickler hilft eine sowohl detaillierte als auch vollständige Dokumentation über alle Teilbereiche des Systems, wobei verschiedene Abstraktionsstufen (beispielsweise – in steigender Abstraktionsstufe – Attribut, Tabelle, Datenbank, OLTP-System) zur Bewahrung der Übersichtlichkeit beitragen.

Manager erhalten Informationen aus organisatorischer, kalkulatorischer und führungs-politischer Sicht. Die von Ihnen genutzte Projektdokumentation beinhaltet im Wesentlichen Projekt- und Organisationspläne, Berichte, Ressourceneinsatz sowie Zieldefinition und -erreichung. Das ermöglicht die Auswertung des Projektes im Hinblick auf die *bessere Durchführung zukünftiger Projekte*.

Unternehmen profitieren von der automatisierten Erstellung von Dokumentationen durch *Kosteneinsparungen*:

- durch den Wegfall von hohem manuellen Aufwand und der damit verbundenen Zeit bei der Erstellung von Dokumentationen,

- durch eine verbesserte Qualität und daraus resultierender zusätzlich gewonnen Zeit, die andernfalls benötigt wird, um mittels einer mangelhaften Dokumentation das System zu verstehen (Chikofsky & Cross, J. H., 1990, S. 17) – insbesondere neue Mitarbeiter bekommen die Möglichkeit, sich schneller in die Systeme einzuarbeiten um mit diesen produktiv zu arbeiten,
- da Expertenwissen explizit vorliegt und beim Ausscheiden von Wissensträgern keine mit hohem Aufwand zu füllende Wissenslücke verbleibt.

Nach der einleitenden Diskussion der Problemstellung und der Motivation für konfigurierbare, automatisch zu erstellende Dokumentationen werden im Folgenden erst die Zielvorgaben einer möglichen Lösung aufgestellt. Dem anschließend vorgestellten Entwurf eines Frameworks folgen schlussendlich eine Zusammenfassung der wichtigsten Punkte und eine Ausführung des weiteren Forschungsvorgehens.

Bei dem vorliegenden Beitrag handelt es sich um Research-in-Progress, wobei sich zur Sicherstellung der wissenschaftlichen Rigorosität an den Design Science Forschungsprozess von Peffers et al. (2007) in Kombination mit den Handlungsanweisungen für die einzelnen Forschungsaktivitäten von Hevner et al. (2004) gehalten wird.

2 Definition der Zielvorgaben

(Wallmüller, 2001, S. 151) empfiehlt, sich bei der Erstellung von qualitativ guten Dokumentationen an den vorgeschlagenen Qualitätsmerkmalen für Dokumente zu orientieren. Demnach sollte eine hochwertige Dokumentation folgenden acht Merkmalen genügen (Dt. Ges. für Qualität e.V. u.d. Nachrichtentechn. Ges. im VDE (NTG), 1995; zitiert nach Wallmüller, 2001, S. 151): Änderbarkeit, Aktualität, Eindeutigkeit, Identifizierbarkeit, Normkonformität, Verständlichkeit, Vollständigkeit und Widerspruchsfreiheit.

Mit dem Ziel, den Dokumentationsbedarf von verschiedenen Anspruchsgruppen zu erfüllen, ist darauf zu achten, die Dokumentationen entsprechend den Interessen der verschiedenen Gruppen auszurichten (Wallmüller, 2001, S. 153). Um eine hochqualitative, zielgruppengerechte Dokumentation bereitstellen zu können, erfolgt zuerst eine Unterteilung der Dokumentqualitätsmerkmale in zielgruppenabhängige und zielgruppenunabhängige Merkmale.

Unter dem Begriff *Dokumentattribute* fassen Forward & Lethbridge (2002, S. 26) Informationen über Dokumente zusammen, welche über den reinen Dokumentinhalt hinausgehen. Beispiele für solche Attribute sind das Dokumentformat, der Schreibstil, die Dokumentenstruktur und die Länge einzelner Artikel. Um den größtmöglichen Nutzen zu erhalten, sind diese zielgruppengerecht anzupassen. Die Qualitätsmerkmale *Eindeutigkeit*,

Verständlichkeit und *Vollständigkeit* werden als Dokumentattribute, welche zielgruppenabhängig sind, aufgefasst und entsprechend unterschiedlich gestaltet.

Änderbarkeit, *Aktualität*, *Identifizierbarkeit*, *Normkonformität* und *Widerspruchsfreiheit* sind Qualitätsmerkmale die nicht an individuelle Bedürfnisse der Zielgruppen gebunden sind und sollten stets im gleichen Maße erfüllt sein. Analog zu dem Begriff der Dokumentattribute, werden diese zielgruppenunabhängigen Dokumentanforderungen im Folgenden unter *Erstellungsprozessattribute* zusammengefasst.

Unter Berücksichtigung der Wirtschaftlichkeit bei der Erstellung und Pflege von qualitativ hochwertigen Dokumentationen ergibt sich folgendes Forschungsziel:

Entwicklung einer Methode zum kostengünstigen Erstellen hochqualitativer und nutzerspezifischer BI-Systemdokumentationen.

Zur Erreichung dieses Ziels werden im weiteren Verlauf die folgenden Forschungsfragen untersucht:

1. Wie lassen sich die Anforderungen der Erstellungsprozessattribute (Qualität der Dokumentation) bei der Dokumentationserstellung kostengünstig erfüllen?
2. Wie lassen sich die Anforderungen der Dokumentattribute (nutzerspezifische Dokumentation) bei der Dokumentationserstellung kostengünstig erfüllen?

3 Entwurf und Entwicklung

Wie bereits dargestellt ist die manuelle Erstellung von Dokumentationen zeitintensiv und fehleranfällig. Durch einen höheren Automatisierungsgrad lässt sich der manuelle Aufwand jedoch reduzieren (Forward & Lethbridge, 2002, S. 28). Die zu dokumentierenden Systeme und Teilsysteme beinhalten bereits einen großen Teil der für die Erstellung einer Dokumentation nötigen Informationen in Form von Quelltext¹ (Forward & Lethbridge, 2002, S. 30).

Modellgetriebene² Ansätze für die Entwicklung von Data Warehouses (Kurze, 2011; Mazón & Trujillo, 2008) beinhalten – über die im Quelltext enthaltenen Informationen

¹ In BI-Systemen entspricht der Quelltext, der üblicherweise in der niedrigsten Abstraktionsstufe angeordnet ist, beispielsweise Datenbankschemata.

² Innerhalb der modellgetriebenen Softwareentwicklung ist die modellgetriebene Architektur (Model-Driven Architecture, MDA) ein von der Object Management Group (OMG) standardisierter Ansatz, der auf der strikten Trennung zwischen Spezifikation und Implementierung eines Systems basiert (Object Management Group, 2003). Dieses Prinzip wird durch die Einführung verschiedener Abstraktionsebenen – berechnungsunabhängige Modelle (Computation Independent Model, CIM), plattformunabhängige Modelle (Platform Independent Model,

hinaus – weitere, für die Erstellung von nutzerspezifischen Dokumentationen erforderliche Informationen in Form von Metadaten über die Systeme, abgelegt in den Modellen. Durch die automatisierte Erzeugung einer Dokumentation direkt aus den vorliegenden Modellen wird ein Großteil der angestrebten Erstellungsprozessattribute erfüllt:

- *Änderbarkeit* und *Aktualität*: der Wegfall von manuellen Aufwänden erlaubt es, Systemänderungen jederzeit durch eine Neugenerierung der Dokumentation zu veröffentlichen;
- *Identifizierbarkeit*: durch die zur Verfügung stehenden Metadaten ist es möglich, Zusammenhänge zwischen einzelnen Objekten, die im Quelltext u.U. nicht abbildbar sind, in die Dokumentation einfließen zu lassen;
- *Widerspruchsfreiheit*: durch die automatische Erzeugung von Dokumentationen aus den Modellen wird in der Dokumentation genau der Grad an Widerspruchsfreiheit abgebildet, der in den Modellen vorliegt.

Im Folgenden wird dargestellt, wie sich diese Metadaten erweitern und verknüpfen lassen, um eine konfigurierbare Erstellung von Dokumentationen zu ermöglichen. Um die Dokumentation nutzerspezifisch zu generieren, sind die bereits vorhandenen Modelle um eine Möglichkeit der Konfiguration bei der Verknüpfung von Informationen zu erweitern. Diese Konfiguration sollte in das modellgetriebene Vorgehen eingebettet werden, also in Form von Modellen vorliegen. Dieser Ansatz erlaubt die Konfiguration der Nutzerzugehörigkeit zu einem Fachbereich und damit den bereitzustellenden Inhalt (beeinflusst *Vollständigkeit* und *Eindeutigkeit*), des Layouts der Dokumentation (beeinflusst *Verständlichkeit*) und des Ausgabeformats (beeinflusst *Eindeutigkeit*).

Die bisher vollzogenen Überlegungen führen zur Konzipierung des folgenden Frameworks.

3.1 Konzept des Frameworks

Dieser Abschnitt stellt das Konzept eines Frameworks zur automatisierten Dokumentation von BI-Systemen vor, welches Abbildung 1 schematisch darstellt. Nachfolgend werden die Komponenten des Frameworks in Richtung des Datenflusses erläutert.

BI-Systemlandschaft

Die BI-Systemlandschaft kann aus einer Vielzahl verschiedener Systeme bestehen, die je nach Architektur operative Systeme, ETL-Prozesse, Datenhaltungs- und Berichtssysteme enthält. Zusätzlich beinhaltet die BI-Systemlandschaft auch das Architekturmodell, d.h. die Einbettung der Systeme in die Gesamlandschaft, und Beschreibungen über Datenflüsse und Prozesse zwischen den Teilsystemen. Diese Aspekte werden im Folgenden unter dem Begriff „Dokumentationssubjekte“ zusammengefasst, die wiederum selbst aus einer Vielzahl kleinerer Entitäten bestehen, die im Folgenden „Bausteine“ genannt werden.

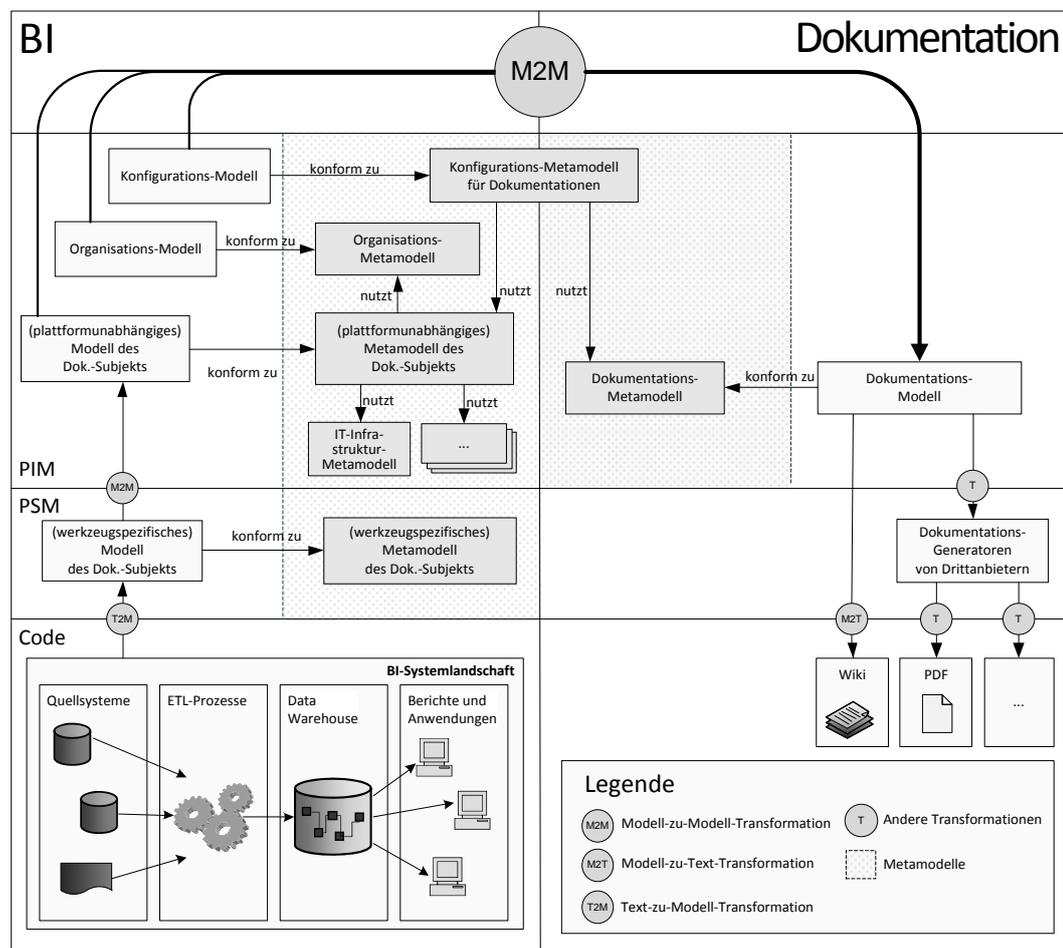


Abbildung 1: Framework zur automatisierten Erstellung nutzerspezifischer BI-Systemdokumentationen

Werkzeugspezifische Metamodelle

BI-Werkzeuge bieten üblicherweise eine Möglichkeit zum Speichern und Laden der modellierten Komponenten, die als Dokumentationsobjekte dienen können. In den meisten Fällen entspricht das Ausgabeformat jedoch keinem Standard wie beispielsweise dem Common Warehouse Model (CWM), sondern basiert auf einem proprietären Format. Dem modellgetriebenen Ansatz folgend ist deshalb für jedes singuläre Format ein Metamodell zu erstellen, welches die jeweiligen Eigenheiten des Werkzeugs abbildet und somit für die Verwendung im Dokumentations-Framework kapselt. Für die Transformation in ein dem Metamodell entsprechendes Modell muss das proprietäre Format gelesen und semantisch korrekt interpretiert werden können. Abhängig von der Art des Formats kann das zu zusätzlichen Entwicklungsaufwänden führen. Im Ergebnis entsteht eine Menge von Metamodellen, von denen jedes spezifisch für ein konkretes Werkzeug und ein damit modellierbares Dokumentationsobjekt ist.

Werkzeugunabhängige Metamodelle

Wie im vorhergehenden Abschnitt dargestellt, kann für jedes Dokumentationsobjekt eine Vielzahl von werkzeugabhängigen Metamodellen existieren. Diese Metamodelle sind naturgemäß nicht kompatibel, da jedes Werkzeug – trotz gleicher Dokumentationsobjekte – verschiedene Attribute und Modellierungsmöglichkeiten bietet. Eine unmittelbare Transformation dieser Artefakte zu einer Dokumentation führte zu einer großen Anzahl von ähnlichen Transformationen, was wiederum einen großen Wartungsaufwand und eine hohe Fehleranfälligkeit bei Anpassungen mit sich brächte. Dem an MDA angelehnten Ansatz weiter folgend werden daher werkzeugunabhängige Metamodelle verwendet, die den Inhalt von Dokumentationsobjekten auf einer konzeptionellen Ebene zusammenführen.³ Werkzeugspezifische Modelle werden dann mit Hilfe von Modell-zu-Modell-Transformationen zunächst in eine werkzeugunabhängige Darstellung überführt. Die werkzeugunabhängigen Modelle bilden somit den Ausgangspunkt für weitere Transformationen.

Transformation aus werkzeugunabhängigen Modellen zur Dokumentation

Das werkzeugunabhängige Modell eines Dokumentationsobjekts enthält dessen vollständige Beschreibung auf konzeptioneller Ebene. Um eine für Nutzer wertvolle Dokumentation zu erstellen, sind jedoch weitere Verarbeitungsschritte notwendig. In Abhän-

³ Darüber hinaus lassen sich mit Hilfe dieser Metamodelle Beziehungen der Dokumentationsobjekte untereinander ausdrücken, was – insbesondere bei der Verwendung von Werkzeugen verschiedener Anbieter – auf werkzeugspezifischer Ebene nur selten möglich ist.

gigkeit des Fachbereichs und der Rolle des Adressaten der Dokumentation im Unternehmen müssen die Informationen unterschiedlich gefiltert und strukturiert werden.

Die dafür notwendigen Daten können nicht aus der Beschreibung des Dokumentations- subjekts gewonnen werden, sondern müssen auf anderem Wege bereitgestellt werden. Dafür sind weitere, dokumentationssubjektunabhängige Metamodelle zu konzipieren, deren Instanzen die nötigen Informationen enthalten:

- Ein Organisationsmodell bildet die Organisationsstruktur eines Unternehmens ab, einschließlich von Abteilungen, Nutzern, Nutzerrollen und deren Beziehungen. In Kombination mit einer Beschreibung der IT-Landschaft lassen sich Zuordnungen zwischen Bausteinen der Dokumentationssubjekte und der Einordnung des Adressaten im Unternehmenskontext bilden, was es ermöglicht, die Inhalte adressatengerecht zu filtern.
- Um die durch die Dokumentattribute spezifizierten Anforderungen der Nutzerrollen an die Dokumentation gerecht zu werden, wird im vorgestellten Konzept ein Konfigurations-Modell verwendet, das es ermöglicht, spezifische Bausteine (beispielsweise Tabellen in einer Beschreibung von Datenbankschemata) bestimmten Dokumentationskonzepten (wie beispielsweise Kapitel, Abschnitte oder Diagramme) in einem dem Informationsbedarf angepassten Detaillierungsgrad zuzuordnen.

Die strukturierte Kombination des werkzeugunabhängigen Modells mit den eben beschriebenen Modellen ermöglicht nun die Erzeugung von adressatengerechten und damit hochwertigen Inhalten.

Dokumentations-Metamodell und Ausgabeformate

Dem Ziel des hier vorgestellten Frameworks nach einer hochgradig flexiblen und gut anpassbaren Lösung folgend ergibt sich die Notwendigkeit zur Unterstützung verschiedenartiger Ausgabeformate. Analog zu dem oben beschriebenen Problem des Mehraufwands bei der Umsetzung durch direkte Transformationen aus den beschriebenen Modellen zu verschiedenen Ausgabeformaten wird auch hier eine plattformunabhängige (also ausgabeformatunabhängige) Abstraktionsebene etabliert. In diesem Dokumentationsmodell liegen die entsprechend aufbereiteten Inhalte in einer dokumentenähnlichen Struktur vor. Dieses Modell lässt sich mit Hilfe von Transformationen einerseits direkt zu verschiedenen Ausgabeformaten (wie beispielsweise einem Mediawiki) oder mittelbar unter Ausnutzung existierender Dokumentationsgeneratoren wie z.B. DITA (DITA, 2011) oder

DocBook (DocBook, 2011) zu den von diesen Drittanbietern unterstützten Formaten transformieren.

4 Zusammenfassung und Ausblick

Der vorliegende Beitrag präsentierte die konzeptionelle Entwicklung eines Frameworks zur automatisierten Erstellung von qualitativ hochwertigen, nutzerspezifischen Software-dokumentationen. Eine praktische Umsetzung dieses Ansatzes verspricht eine signifikante Reduktion des Erstellungs- und Pflegeaufwands von Software-dokumentationen. Weiter zeichnen sich die so erstellten Dokumentationen durch einen hohen Grad an Aktualität aus, da die Anwendung des Paradigmas der modellgetriebenen Softwareentwicklung eine einfache Integration des Dokumentationsprozesses in den Wartungsprozess von BI-Systemen gestattet. Die vollständig automatische Erzeugung von nutzerspezifischen Dokumentationen führt ferner zur Vermeidung manueller Eingriffe sowie Medienbrüche und schließt so potentielle Fehlerquellen aus, was zu qualitativ höherwertigeren Software-dokumentationen führt.

Dem von Peffers et al. vorgeschlagenen Design Science Forschungsprozess folgend, ist als nächstes die Umsetzbarkeit des entwickelten Frameworks zu demonstrieren. Hierfür ist, auf Grundlage des vorgestellten Frameworks, die Entwicklung eines Prototypens für automatisierte, konfigurierbare ETL-Prozess-Dokumentation geplant. Dieser Prototyp stellt weiterhin die Grundlage für die ausstehende Evaluation des Dokumentationsframeworks dar. Der Nachweis der Reduktion des Erstellungs- und Pflegeaufwandes von Dokumentationen geschieht im Rahmen einer Pilotierung (Österle et al., 2010, S. 668). Die in diesem Zusammenhang anfallenden Dokumentationen bilden zusammen mit dem Standard „Systems and software engineering – Requirements for designers and developers of user documentation“ (ISO-IEC 26514; International Organisation for Standardization, 2008) den Ausgangspunkt zum Nachweis des noch ausstehenden Dokumentqualitätsmerkmals der Normkonformität.

5 Anmerkungen

Das „Computer-Aided Warehouse Engineering“ (CAWE) Projekt, in dessen Rahmen der vorliegende Beitrag entstanden ist, wird mit Mitteln des ESF und des Freistaates Sachsen gefördert.

Literaturverzeichnis

- Chikofsky, E. J., & Cross, J. H., I. (1990). Reverse engineering and design recovery: a taxonomy. *IEEE Software*, 7(1), 13-17.
- DITA (2011). <http://www.oasis-open.org/committees/dita>.
- DocBook (2011). <http://www.oasis-open.org/docbook>.
- Forward, A., & Lethbridge, T. C. (2002). The Relevance of Software Documentation, Tools and Technologies: a Survey. *Proceedings of the 2002 ACM symposium on Document engineering - DocEng'02* (pp. 26-33). New York, USA: ACM Press.
- Dt. Ges. für Qualität e.V. u.d. Nachrichtentechn. Ges. im VDE (NTG). (1995). *Software-Qualitätssicherung : Aufgaben, Möglichkeiten, Lösungen*. (2nd ed.). VDE-Verlag.
- Gluchowski, P., & Kurze, C. (2010). Modellierung und Dokumentation von BI-Systemen. *CONTROLLING - Zeitschrift für erfolgsorientierte Unternehmensplanung*, 22, 676-682.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, 28(1), 75-105.
- Hofmann, M., Gluchowski, P., Jacobi, F., & Kurze, C. (2011). Computer-Aided Warehouse Engineering: Dokumentation und Modellierung komplexer Data-Warehouse-Systeme. *Presentation slides of the 11th European TDWI Conference with BARC@TDWI-Track*.
- International Organisation for Standardization. (2008). *Systems and software engineering -- Requirements for designers and developers of user documentation (ISO-IEC 26514)*.
- Kurze, C. (2011). *Computer-Aided Warehouse Engineering: Anwendung modellgetriebener Entwicklungspfadparadigmen auf Data-Warehouse-Systeme*. Verlag Dr. Kovač.
- Laudon, K. C., & Laudon, J. P. (2010). *Management information systems: managing the digital firm* (11th ed.). Upper Saddle River, New Jersey: Pearson.
- Mazón, J.-N., & Trujillo, J. (2008). An MDA approach for the development of data warehouses. *Decision Support Systems*, 45(1), 41-58.
- Object Management Group. (2003). MDA Guide V1.0.1. Retrieved from <http://www.omg.org/cgi-bin/doc?omg/03-06-01>.
- Peppers, K., Tuunanen, T., Rothenberger, M. a, & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45-77.

- Pomberger, G., & Blaschek, G. (1996). *Software Engineering: Prototyping und objektorientierte Software-Entwicklung* (2nd ed.). Carl Hanser Verlag.
- Vercellis, C. (2009). *Business intelligence: data mining and optimization for decision making*. Hoboken, N.J.: Wiley.
- Wallmüller, E. (2001). Die Rolle der Dokumentation in Software-Projekten. *Software-Qualitätsmanagement in der Praxis: Software-Qualität durch Führung und Verbesserung von Software-Prozessen* (2nd ed., pp. 149-156). Hanser Fachbuch.
- Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., et al. (2010). Memorandum zur gestaltungsorientierten Wirtschaftsinformatik. *Angewandte Informatik*, 62(6), 664-672.

Development of a Concept to Integrate Technical Product Data and Business Data

Motivation, Conceptual Framework, and Consecutive Research

Jens F. Lachenmaier

*Lehrstuhl für Wirtschaftsinformatik I
Universität Stuttgart*

Abstract

As Industrial Enterprises have to deal with many challenges, it is no longer affordable to treat the engineering and the business perspective separately. By combining technical product data from the engineering perspective with established key-performance-indicators (KPIs) from the business perspective in a product-oriented Data Warehouse (pDWH), it will be possible to relate the KPIs to properties of a product. The consecutive research establishes a link between the two domains. In this research in progress paper existing literature from these two domains, some basic assumptions, a conceptual framework, as well as the course of action will be discussed.

1 Introduction

Companies struggle with many different issues such as the shortening of product lifecycles and the growing complexity of their business in general. Especially industrial enterprises face fierce competition and have to improve their internal structures to stay competitive. An important part of the internal processes, which influences the company's performance, is decision-making. Therefore, it has to be ensured that decision-makers are provided with reliable, correct, timely, and complete information. This is an important part of information management in a company and essential to create a competitive advantage (Krcmar, 2011).

2 Proceeding Research

2.1 Business Perspective

The business perspective consists on the one hand of operational systems that focus on transactions, like Enterprise-Resource-Planning-, Material-Requirements-Planning-, and Supply-Chain-Management-systems. On the other hand, there is an agreement on the business side, that the data from the different systems should be homogenized in order to analyze the business performance and to support decision-making processes. This could be achieved with business intelligence (BI) systems and especially Data Warehouses that are considered the basis for BI. (Kemper et al., 2010)

2.2 Engineering Perspective

In most industrial companies, there is a separate IT-landscape for the engineering domain. It is often referred to as CAx-systems, as it contains among others computer-aided-design, computer-aided-manufacturing and computer-aided-quality-systems. These systems are integrated in order to automatize certain steps during manufacturing (Asiabanpour et al., 2009). During product design, typically 3D-models of the product are created, which consist of many features. If the design is feature-based, it is possible to exchange the features with other technical CAx systems (Deng, et al., 2002; Dong & Vijayan, 1997).

2.3 Research Gap

Already the Y-Model, which was developed during the 1990s by Scheer, shows that there is no established link between the two perspectives (cf. *Figure 1*). It shows on the left side the business planning functions that are nowadays covered by Enterprise-Ressource-Planning-systems; on the right side, there are the technical functions with the corresponding CAx-systems.

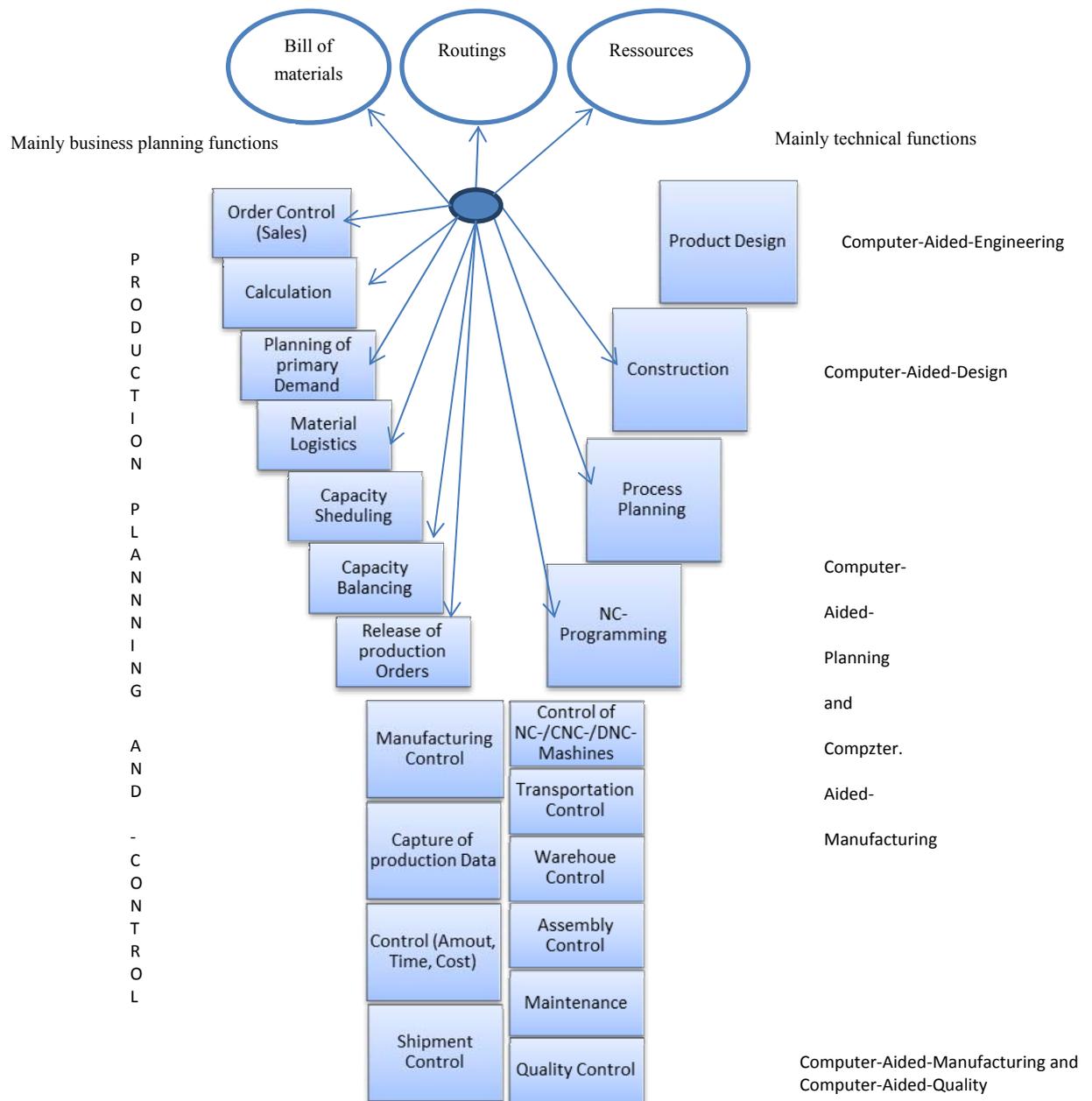


Figure 1: Y-Model – adapted from (Scheer, 1995)

So far, the integration has only been established within the two separate domains. The only shared information in practice is the bill of materials (e.g. Koliha, 2006).

Some authors in literature have identified this gap (N.N. 2011; Zeilhofer-Ficker, 2010; Lasi & Kemper, 2011). There may be different solutions to integrate the two different IT-system landscapes (e.g. concepts that can be subsummed as Computer-Integrated-Manufacturing or workflow management), but they do not focus on the support of deci-

sion-making. This paper proposes to solve the problem using business intelligence. The approach focuses on a product-oriented Data Warehouse (pDWH), which combines the technical construction data of a product with different kinds of transactional business data. Examples might be manufacturing and recycling costs or processing time to identify which properties of products – like the material or geometric features - cause which outcomes on the business side. A pDWH is proposed by Kemper, Baars & Lasi (2011) as it offers the possibility to combine business and technical data in a separate system, aggregate and analyze it using standard Online-Analytical-Processing-Techniques as well as its focus on supporting decision makers.

3 Conceptual Framework for Consecutive Research

One of the first steps in research is structuring the topic in a conceptual framework (Miles, 2009). It contains all relevant factors that should be considered and their assumed relationships.

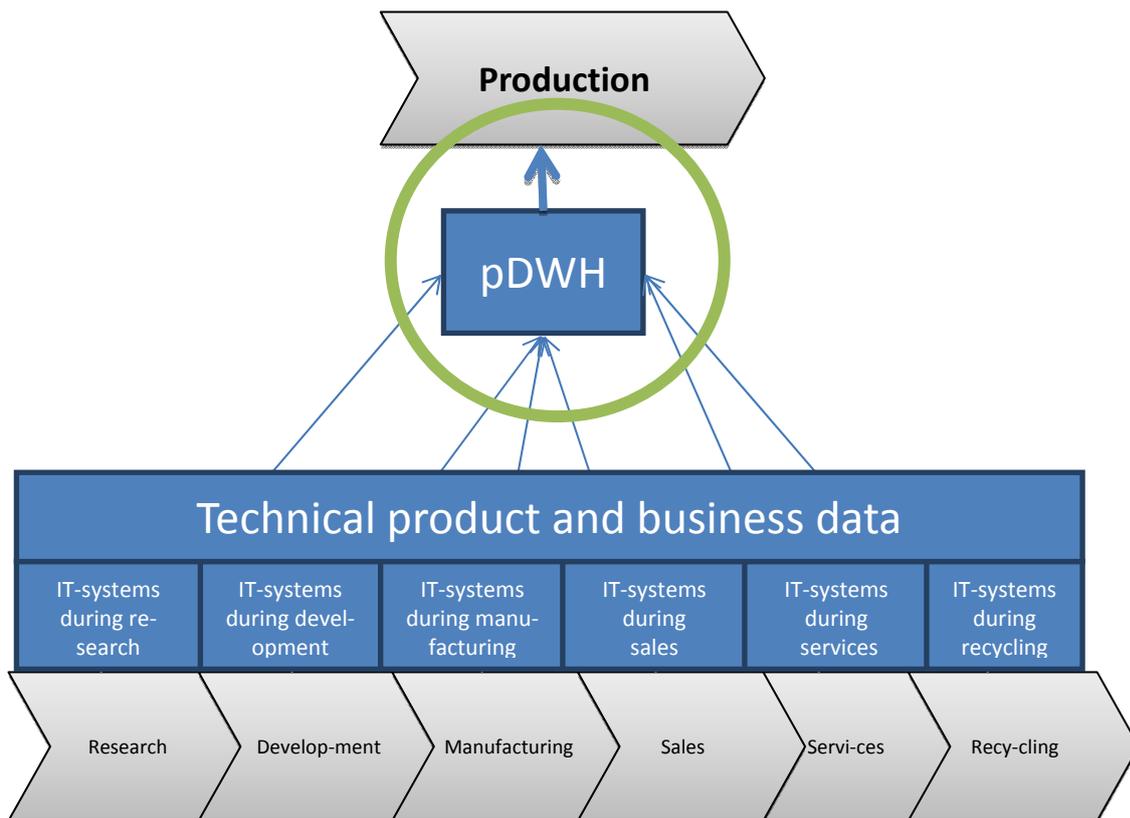


Figure 2: Conceptual Framework

The conceptual framework (cf. *Figure 2*) is based on a typical product lifecycle, starting with research and development, continuing with production sales, services, and finally recycling. During these different phases, the operative IT systems generate and store technical product data and transactional business data. This transactional and feature-based data can be extracted, transformed, and loaded (ETL) into a product-oriented Data Warehouse (pDWH). The focus of the research is the development of such a pDWH. Therefore, ETL-processes have to be defined; the data has to be modeled in star schemas to be suitable for storage and analysis in a Data Warehouse and other requirements have to be collected and fulfilled. (e.g. physical design, non-functional requirements). (Silvers, 2008), (Gluchowski et al., 2008), (Kimball., 2010) With an existing prototype, it will be possible to determine how the pDWH can influence and support decision making during production. Using iterative development cycles, the pDWH will evolve over time (Goeken, 2006; Kemper et al., 2010).

This conceptual framework is meant to structure and define the scope for all related research. Therefore, it has to be ensured that there are not any important parts missing.

4 Consecutive Research

The research follows the design-oriented approach (Österle et al., 2011), which concentrates on developing concepts for innovative information systems that can be implemented by companies. The objective is to design an information system, the result type is to develop a prototype of the pDWH, and the stakeholders are decision-makers who work in production contexts.

After an initial analytics phase, which is necessary to identify and describe the business problem that should be solved, the focus will be on the design of a pDWH. The need for a product-oriented Data Warehouse has been documented by investigating the current state of the art with regard to the integration of technical and business oriented IT-systems. Additionally, interviews will be conducted to ensure that the pDWH will fit the needs of decision makers.

As the pDWH should not only work in a single company, but for a larger variety of companies, it is not possible to ask all prospective users if they have all the information they need and if the Data Warehouse fulfills all their requirements as proposed by Kimball (Kimball et al., 2008). Therefore, this paper proposes a more generic approach: The Data Warehouse will be created based on the information needed to support decision making processes. Gluchowski et al. call this logical deduction of the information demand. (Gluchowski et al., 2008)

Later on, the extracted data from the operative systems has to be matched with the information that is necessary and has to be modeled into star schemas.

5 Discussion and Conclusion

The necessity of a concept for integrating both the business and the technical perspective has been shown, but some aspects of the research are still open for discussion: First, it has to be discussed if all relevant factors are included in the conceptual framework. Second, how to proceed with the design of the DWH.

A product-oriented Data Warehouse is beneficial in all phases of the product lifecycle. It supports decision-making during construction, production, and recycling. In the future, it might even be possible to anticipate the consequences of changes during product design to optimize the product regarding issues that occur later and thereby increasing business performance based on a company's experience with similar products in the past. Not only will this give the management a more complete picture of what is really going on in their company, but it can also help to reduce costs and the time to market.

References

- Asiabanpour, B., Mokhtar, A., Hayasi, M., Kamrani, A., & Nasr, E. A. (2009). An Overview On Five Approaches for Translating CAD Data into Manufacturing Information. *Journal of Advanced Manufacturing Systems*, 8(1), 89-114.
- Deng, Y. M., Britton, G. A., Lam, Y. C., Tor, S. B., & Ma, Y. S. (2002). Feature-based CAD-CAE integration model for injection-moulded product design. *International Journal of Production Research*, 40(15), 3737-3750. doi: 10.1080/00207540210141643
- Dong, J., & Vijayan, S. (1997). Features extraction with the consideration of manufacturing processes. *International Journal of Production Research*, 35(8), 2135-2156. doi: 10.1080/002075497194778
- Gluchowski, P., Gabriel, R. & Dittmar, C. (2008). *Management Support Systeme und Business Intelligence computergestützte Informationssysteme für Fach- und Führungskräfte* (2nd. ed.). Berlin Heidelberg: Springer.
- Goeken, M. (2006). *Entwicklung von Data-Warehouse-Systemen Anforderungsmanagement, Modellierung, Implementierung* (1. ed.). Wiesbaden: Dt. Univ.-Verl.

Kemper, H.-G., Baars, H. & Lasi, H. (2011). *Business Intelligence - Innovative Einsatzfelder in der Industrie*, in: Kemper, H.G., Baars, H., Lasi, H. Felden, C., Krebs, S., Stock, S. (Eds.): *Perspektiven der Business Intelligence - Festschrift anlässlich des 60. Geburtstags von Peter Chamoni*, München, Hanser.

Kemper, H.-G., Baars, H. & Mehana, W. (2010). *Business Intelligence - Grundlagen und praktische Anwendungen eine Einführung in die IT-basierte Managementunterstützung* (3. ed.). Wiesbaden: Vieweg + Teubner.

Kimball. R., Ross, M., Thornthwaite, W., Mundy, J. & Becker, B. (2008). *The data warehouse lifecycle toolkit [practical techniques for building data warehouse and business intelligence systems]* (2. ed.). Indianapolis, Ind.: Wiley.

Kimball, R. (2010). *The Kimball Group reader relentlessly practical tools for data warehousing and business intelligence*. Hoboken, N.J.Chichester: Wiley, John.

Koliha, M. W. (2006). CAD and BOM data get connected. *Design News*, 61(8), 75-80.

Krcmar, H. (2011). *Einführung in das Informationsmanagement*. Berlin Heidelberg Dordrecht [et.al.]: Springer.

Lasi, H., & Kemper, H.-G. (2011). Integrationsansätze zur Verbesserung der Entscheidungsunterstützung im Innovationsmanagement. *HMD - Praxis der Wirtschaftsinformatik* (278), 94-103.

Miles, M. B. (2009). *Qualitative data analysis an expanded sourcebook* (2. ed., [Reprint]). Thousand Oaks, Calif. [u.a.]: Sage.

N.N. (2011), ERP/MES/CAD: Durchgängigkeit der Informationen über alle Ebenen ist Voraussetzung - Industrial IT wird ausgebaut. *Industrieanzeiger* (6).

Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Sinz, E. J. (2011). Memorandum on design-oriented information systems research. *European Journal of Information Systems*, 20(1), 7-10. doi: 10.1057/ejis.20 0,55.

Scheer, A.-W. (1995). *Wirtschaftsinformatik Referenzmodelle für industrielle Geschäftsprozesse*. Berlin Heidelberg [u.a.]: Springer.

Silvers, F. (2008). *Building and maintaining a data warehouse*. Boca Raton, Fla. London: Auerbach: Taylor & Francis.

Zeilhofer-Ficker, I. (2010). Durchgängige Daten für die Produktionsplanung - immer noch Vision statt Wirklichkeit. *Produktion, Materialwirtschaft & Logistik*.

Evaluating the Effectiveness of Relational Marketing Strategies in Corporate Website Performance

Kerstin Schäfer, Oliver Günther

*Institute of Information Systems
Humboldt-Universität zu Berlin*

Abstract

The study aligns marketing intelligence approaches with website performance assessment for evaluating the effectiveness of a company's relational marketing strategies (RMS) in the performance of its corporate website. We develop a web mining based methodology that combines a classification procedure with discriminant analysis for analyzing click-stream data. Furthermore, we integrate managerial perspectives in investigating and interpreting customer website usage for determining RMS effectiveness. As a result, we are able to quantify the moderating effect of the website's structure and content on the effectiveness of Internet based relational marketing (RM) efforts. The applicability of this approach is demonstrated by investigating RMS effectiveness in the performance of a software developer's website. The results of our approach support website optimization and also enable customization for different relational marketing strategies.

1 Introduction

A critical concern of marketing has been how to initiate, customize and strengthen the relationship between a company and its customers, in order to increase customer satisfaction and to achieve a competitive advantage. This is reflected in a conception of relationship marketing (RM) that includes strategies for establishing, developing, and maintaining successful relational exchanges (Morgan & Hunt, 1994). RM has experienced considerable growth in the last few decades, both in business practice and as an object of academic research (Srinivasan & Moorman, 2005). Thus, the evaluation of RM effectiveness provides knowledge for increasing the return of a company's RM investment (Reinartz & Kumar, 2003). One of the most important relational exchange contexts takes place on the Internet, where a company's RM efforts are closely connected to the performance of its corporate website, particularly with respect to website structure, content provision, and

quality of service (Palmatier, Dant, Grewal & Evans, 2006). Thus, the performance of a corporate website moderates the relationship between a company and its customers and thereby affects the strength of RM effectiveness. The question of how to quantify this moderating effect is the research objective of this study.

Marketing and information systems (IS) research addresses this in several ways: RM uses data intensive methodologies based on customer purchase behavior (*marketing intelligence*) to determine RMS effectiveness (Vercellis, 2009). In website performance assessment, RM outcomes such as web-customer satisfaction and loyalty are measured (i.e. Lee, Strong, Kahn & Wang, 2002; McKinney, Yoon & Zahedi, 2002). But IS also investigates website performance by analyzing customer-website interaction using clickstream data in web mining approaches (i.e. Senecal, Kalczynski & Nantel, 2005; Spiliopoulou & Pohle, 2001). Nevertheless, we observe only a small number of methodological approaches that link the evaluation of a company's RM effectiveness to website performance, based empirically on customer website usage.

We attempt to overcome this shortage by presenting a methodological framework that aligns marketing intelligence with web mining approaches. We model, analyze and predict the effectiveness of a company's RM effort in website performance by integrating managerial perspectives in the assessment of customer website usage. The knowledge gained supports website optimization and also enables customization for different relational marketing strategies.

The paper is structured as follows. After a summary of related research, we describe the developed methodology in detail. The practical relevance of our approach will be demonstrated by analyzing historical clickstream data of a corporate web site from the software development sector. Finally, we discuss our results and their limitations.

2 Related Research

Research investigating relational marketing strategy from a marketing intelligence perspective deals mainly with methodology improvement for relational marketing analysis that uses data mining techniques, the information infrastructure involved, and aspects of relationship economics such as customer lifetime value, acquisition, retention and cross-selling (i.e. Büchner & Mulvenna, 1998; Chi & Tavella, 2008).

Website performance assessment is a significant topic in marketing and IS research. Depending on the website type, different measures of effectiveness exist. For transaction-orientated websites in e-commerce, website effectiveness is determined in terms of revenue. Here, metrics are used that describe the customer lifecycle and customer web usage

such as reach, acquisition, conversion, click through and look-to-buy (Lee, Hoch, Podlasek, Schonberg & Gomory, 1999; Teltzrow & Berendt, 2003). The effectiveness of information and service provision orientated websites is often measured by investigating customer web site usage in terms of implicit feedback (Stolz, Viermetz, Neuneier & Skubacz, 2005). In this case, web usage mining approaches are an adequate procedure (i.e. Bucklin & Sismeiro, 2003).

With regard to the research object of this study, only one web mining approach is known that estimates website performance in a corporate context using clickstream data (Hochsztain, Millán, Pardo, Pena & Menasalvas, 2003). Although customer value is estimated in this study from a website provider perspective, no clear conceptual linkage to relationship marketing (RM) and relational marketing analysis methodology is available. The study at hand will overcome this lack and directly integrate relational marketing analysis in an methodological web mining approach for determining the effectiveness of a company's RM efforts in its corporate website.

3 Evaluating the Effectiveness of Relational Marketing Strategies in Corporate Website Performance

Corporate websites are an important channel for Internet based marketing communication in a company's multi-channel marketing efforts (Weinberg, Parise & Guinan, 2007). They support three relational marketing strategies (RMS) in particular: (1) the building of brand equity, which is the sum of the intangible assets of the corporate brand that are supported by factors such as name awareness, perceived quality and customer loyalty (Aaker, 1993); (2) the creation and maintenance of relationships at reduced costs (Shet & Parvatiyar, 2000); and (3) the creation of customer satisfaction by delivering superior products and services (Gale, 1994). Success in any of these strategies leads to an increase in repeat purchases, insulation from price increases and improved responsiveness to marketing communication by customers. Thus, a corporate website can maximize the impact of a company's RM efforts (Argyriou, Kitchen & Melewar, 2006). The strength of this contribution depends on how well the website performs with respect to customer information needs. Therefore it is necessary to develop metrics that evaluate the effectiveness in relation to customer website usage of the RM effort in website performance.

In the following, a general research framework is presented and operationalized that integrates RMS effectiveness determination in a web mining approach for modeling, evaluating and predicting RMS effectiveness in corporate website performance.

3.1 Corporate Websites and Relational Marketing Strategies

According to the functions a corporate website possesses, different relational marketing strategies (RMS) are supported, and different effectiveness metrics can thus be distinguished. Using the typology of website functions and metrics used by Booth & Jansen (2009), we specify firstly that a *commerce* orientation focuses on getting customers who visit the site to purchase goods or services directly from the website. The building of brand equity is the most important RMS here, because it addresses customer loyalty in terms of repeated purchases and perceived product quality. Clickstream-based effectiveness metrics for commerce emphasize transactions such as purchase or downloads. *Content and media provision* focuses on drawing in visitors and immersing them within the site. The relationship between a company and its customers can be strengthened as a result. Clickstream-based RMS effectiveness metrics are in this case concerned with visitor engagement, for instance a high number of visit actions indicating browsing, and long session duration. A *support and service* function helps users to find specialized answers for specific problems. This increases customer satisfaction at reduced external support costs. RMS effectiveness depends largely upon the provision of the searched information with regard to the structure of the website; thus, low page depth is a useful indicator in clickstream data.

3.2 The Evaluation Algorithm

The allocation of relational marketing strategies (RMS) to website functions is necessary for developing a web mining approach that integrates managerial perspectives in the investigation of customer-website interaction. Our approach proceeds in three steps: initially business experts perform a characterization and discrimination of RMS relevant effectiveness measures in historical click stream data; then we use a classification procedure for determining the values of the previously identified metrics in the historical clickstream data; and finally we evaluate RMS effectiveness and identify contributing factors by using discriminant analysis.

In the initial exploratory approach, a business expert identifies RMS relevant metrics for inferring customer behavior from the clickstream data. Here, typical RMS relevant visitor activities are identified, such as information gathering by browsing, fact-finding, and all possible transactions (Kellar, Watters & Shepherd, 2006). Browsing, as an expression of the RMS metric called visitor engagement, is usually defined by a high number of visit actions, and a long session duration in the clickstream data at the user session level. Fact-finding, on the other hand, is defined by a low page depth, which implies a low number of visit actions and a certain style of interacting with the structural elements of the website content (Bucklin & Sismeiro, 2003). Integrating a distinction of different page types, i.e.

into action/navigational and target pages (Spiliopoulou & Pohle, 2001), fact-finding is expressed in clickstream data by visiting only few navigational pages in order to find specific content.

Assuming that the corporate website supports different RMS, the identified visitor activities are then weighted by business experts according to their relevance for each available RMS (table 1). We express this formally, similar to Hochsztain et al. (2003):

Definition 1: The Strategy set $S = \{s_1, s_2, \dots, s_n\}$ represents the relational marketing strategies of a company that are supported by the corporate website at a particular moment.

Definition 2: The Customer Website Usage set $U = \{u_1, u_2, \dots, u_m\}$ represents different activities a visitor can engage in when interacting with the corporate website.

Definition 3: We define w_{ij} as the weight to each relational marketing strategy, where i represents the i^{th} visitor activity and j the j^{th} relational marketing strategy. It is assumed that business experts assign weights differently for each visitor activity in every relational marketing strategy. The weight function has the following properties:

$$0 \leq w_{ij} \leq 1 \text{ where } s_i \in S, u_i \in U \text{ with } \sum_{i \in S, j \in U} w_{ij} = 1.$$

For each RMS the sum of all visitor activities is always 1; when an activity possesses no weight, then it is not relevant for this RMS from a managerial point of view. Table 1 displays an example of such a relevance weighting of customer behavior RMS specific.

		RMS for website functions		
		Commerce	Content/media	Support /service
RMS effectiveness metric	Customer website usage	Brand equity (s_1)	Relationship maintenance (s_2)	Customer satisfaction (s_3)
Visitor engagement	Long session duration (u_1)		0.5	
	Browsing (u_2)		0.3	
Page depth	Finding (u_3)	0.1	0.1	0.9
Transaction	Buy, download (u_4)	0.9		0.1

Table 1: Example of relevance weightings for customer behavior in RMS

This is relevant for the following steps, where we deploy a statistical classification method (decision tree) for inferring visitor activities in the clickstream data (Fox, Karnawat, Mydland, Dumais & White, 2005). Further, the expert discrimination of user activities and their defined thresholds are used for identifying and selecting visitor activity representing nodes of the decision tree. Integrating the RMS weighting for calculating an overall RMS effectiveness value then extends the so enriched classification results. Here, we apply an effectiveness heuristic that determines the number of user sessions with RMS relevant customer behavior, using a linear equation L : if the user session contains any RMS relevant activities, then the website successfully moderates a company's RM effort; if the user session does not contain any RMS relevant activities, than the website fails to moderate the RM effort. This is formally expressed by:

$$L_1 = s_1 u_1 + s_1 u_2 + \dots + s_1 u_j$$

$$L_i = \sum_{j=1}^m s_i u_j$$

- If RMS effectiveness in a user session is determined to be $L > 0$
- If RMS failure in a user session is determined to be $L \leq 0$

This dichotomous RMS effectiveness value is necessary for the last step of our analysis, where discriminant analysis depicts the relative contribution of the modeled and classified visitor activities to overall RMS effectiveness.

For an overview of the developed methodology see figure 1.

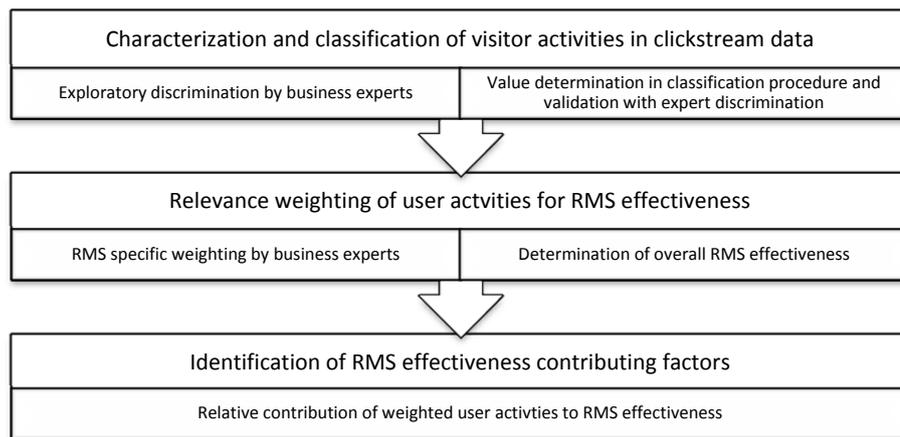


Figure 1: Developed methodology for RMS effectiveness evaluation

4 Implementation and Empirical Results

The developed approach was applied for evaluating the effectiveness of the relational marketing strategies (RMS) in the website of a German software developer. The website is dynamic and contains over 25 billion distinct URLs for content generation. The website offers a commerce function (shop), content and media provision (product catalogue, media center, company, community) and service and support (support). Relevant RMS that are moderated by website performance are: (1) building of brand equity, which refers to the shop area; (2) relationship maintenance in all content providing areas; and (3) creation and increase of customer satisfaction in the service and support area.

4.1 Clickstream Data Collection and Data Preparation

For investigating RMS effectiveness in website performance, we collected server log files of the clickstream over a period of four weeks (6.5 GB). In the following data preparation and transformation process, according to Cooley, Mobasher & Srivastava (1999), and Markov & Larose (2007), data was cleaned and sessionized using MySQL 5.5.9. In total, we identified and prepared for statistical analysis 1,389,413 sessions with a session duration longer than 10 seconds and containing two or more visit actions. Additional data was collected by using a crawler to distinguish navigational and content page view objects. The crawler requested the HTML page title, which offers information about the website area for all included URLs and parsed each URL as navigational or content objects. To this end, we deployed a decision heuristic: page view objects with more than 30 links are navigational; those with less than 30 links are content objects. This classification was successfully evaluated and validated by comparing the mean of session duration values (Spiliopoulou & Pohle, 2001). In total, we identified 143,546 distinct navigational, and 11,075,070 content objects in our data set.

For applying our developed web mining approach, we generated four data sets that allocate user sessions according to their clicks in a website area such as products, shop, community and support (multiple response possible). In further data quality monitoring, outliers were identified and removed using the interquartile range method (Larose, 2005) for the ratios of visit actions and session duration in each data set.

Table 2 displays the data set after outlier removal.

	Means of session ratios per website area			
	Products	Shop	Community	Support
N of sessions	114,264	158,529	20,052	184,626
Session duration per website area (SD)	20.11	156.36	277.14	146.80
Session duration per navigational object (SD_Nav)	14.86	18.82	21.40	33.52
Session duration per content object (SD_Con)	5.25	137.55	255.74	113.28
Visit actions in website area (VA)	2.85	3.85	3.98	3.42
Visit actions navigational objects (VA_Nav)	2.06	0.59	0.56	0.80
Visit actions content objects (VA_Con)	0.79	3.26	3.41	2.62

Table 2: Mean of session ratios per website area

4.2 Data Analysis and Results

Statistical analysis was performed using PASW Statistics 18 and PASW Modeler 13. Data analysis proceeded according to the developed approach.

Initially, the expert executed characterization and discrimination of RMS relevant visitor activities in the sessionized click stream data. It defined the thresholds for the ratios visit actions (VA) and session duration (SB), and the values of the indicator variables that were used for describing visitor activities. In the subsequent classification procedure, a C&RT regression tree was used for predicting the values of the visitor activities browsing and fact-finding, and for classifying visitor activities within the whole data set. The Gini-index was used as a measure of impurity and a test-sample cross validation for tree-selection. Two models for each activity were calculated and compared, using one dependent (DV) and one or more independent variables (IV). For model comparison and evaluation, the model with the lower standard error of the estimate (s) was preferred. For each activity an indicator variable was created (Breiman, Friedman, Stone & Olshen, 1984).

Table 3 displays the regression models, and the results of the model evaluation that prefers model 1, as well as the predicted values of the ratios used for model 1.

		Regression tree				Model evaluation	
		C&RT model 1		C&RT model 2		<i>s</i>	<i>s</i>
		DV	IV	DV	IV	C&RT 1	C&RT 2
Products	Browsing	VA >3.5	SD >12.5	SD	VA	0.02	4.34
	Fact-Finding	VA <3.5	SD_Nav <0.5, SD_Con >12.5	SD	VA-Nav. Con	0.01	4.33
Shop	Fact-Finding	VA <3.5	SD_Nav <0.5, SD_Con >2.5	SD	VA_Nav, VA_Con	0.03	167.44
Comm- unity	Browsing	VA >3.5	SD >39.5	SD	VA	0.10	1836.77
	Fact-Finding	VA <3.5	SD_Nav <0.5, SD_Con >31.5	SD	VA_Nav, VA_Con	0.10	1790.04
Support	Fact-Finding	VA <3.5	SD_Nav <0.5, SD_Co >14.5	SD	VA_Nav, VA_Con	0.02	7196.89

Table 3: Classified user activities

The identified user activities were then relevance weighted as presented in table 4.

	Website area	Products	Shop	Community	Support
	RMS	Relationship maintenance	Building of brand equity	Relationship maintenance	Create customer satisfaction
Weights of RMS relevance	Long session duration	0.5		0.5	
	Browsing	0.2		0.3	
	Fact-Finding	0.2	0.1	0.1	0.9
	Download	0.1	0.1	0.1	0.1
	Purchase		0.8		

Table 4: Relevance weighting of visitor activities

Then, overall RMS effectiveness was calculated, using the weighted activity indicator variables (table 5). Results demonstrate that the company's RM effort was moderated differently in each website area: for relationship maintenance in the product area, only 45 % of all user sessions were determined as effective; whereas in the community area, which provides an exchange forum between customers, a high effectiveness of 77% was achieved. For the building of brand equity in the shop area, the website performed moderately efficiently, with 48% of RMS user sessions effective. The creation of customer satisfaction displays the highest RMS failure, including 75% of all user sessions.

Website area	Products		Shop		Community		Support	
RMS	Relationship maintenance		Building of brand equity		Relationship maintenance		Create customer satisfaction	
Statistics	N	%	N	%	N	%	N	%
RMS failure	62956	55.1	83891	52.9	4473	22.3	138251	74.9
RMS effectiveness	51308	44.9	74638	47.1	15579	77.7	46375	25.1
Total	114264	100	158529	100	20052	100	184626	100

Table 5: Results of RMS effectiveness determination

Finally, discriminant analysis determined the relative relevance of the weighted activities. The assumption of homogeneity of variance was fulfilled; analysis was thus conducted. The quality of discrimination was very high, as shown in high eigenvalues and low values of Wilk’s Lambda (eigenvalues between 5.04 and 8.84; Wilk’s Lambda between 0.10 and 0.17 with $p = 0.00$)

Table 6 presents the average discriminant values. Table 7 shows the relative importance of the weighted visitor activities for RMS effectiveness.

	Group centroids	
	RMS effectiveness	RMS failure
Products	-2.68	3.29
Shop	-2.24	2.52
Community	-4.19	1.20
Support	-1.31	3.91

Table 6: Average discriminant values

	Structure matrix of standardized discriminant coefficients				
	Website area	Products	Shop	Community	Support
	RMS	Relationship maintenance	Building of brand equity	Relationship maintenance	Create customer satisfaction
RMS relevance weighted visitor activities	Long session duration	0.17		0.88	
	Browsing	0.65		0.46	
	Fact-Finding	0.09	0.36	0.12	0.71
	Download	0.11	0.08	0.09	0.15
	Purchase		0.30		

Table 7: Relative importance of user activities for RMS effectiveness

Classification results for the predicted group membership (RMS failure/effectiveness) are good with 95.3 % (products), 95.8 (community), 99% (support), and 100% (shop) for the correct prediction assumed a priori probability.

Results of the discriminant analysis indicate that the contribution of visitor activities to RMS effectiveness varies in the different website areas: for relationship maintenance, visitor engagement is the greatest contributing factor, which is characterized by browsing in the product area and by long session duration in the community area. Effective building of brand equity in the shop area is determined by low page depth, indicating fact-finding as a prominent activity besides purchase. The effective creation of customer satisfaction in the support area is also affected most by low page depth due to fact-finding.

Integrating overall RMS effectiveness into this perspective, our study finally demonstrates that the website moderates the company's overall RM effort differently: It moderates it particularly well for relationship maintenance in the community area, where a high visitor engagement supports the strength of the company-customer relationship. In the product area, the company's RM effort is less effective for relationship maintenance. Further, the website is not successful in moderating the RM effort in terms of customer satisfaction in the support and service area. Fact-finding, the user activity which contributes most to effectiveness, is not supported, indicating that the website does not provide a quick and efficient information provision in this area. Thus, the website needs optimization in navigational aspects, content access and presentation in the support and service area in particular.

5 Discussion

The developed methodology has several implications for practitioners and researchers. For practitioners, a RM effort orientated investigation of website performance using customer web usage data was provided, and knowledge of RM effort orientated website customization and optimization is generated. Further applications of this methodology support online analytical processing by providing user model development, identification and classification. Implications for researchers include the extension of existing approaches for determining website effectivity. We provided a comprehensive methodology for interpreting implicit user data, combining expert discrimination and data mining techniques. Furthermore, this methodology unites marketing and IS research by aligning marketing intelligence with advanced web mining techniques.

The main drawback of the method is the characterization and discrimination of user activities by business experts, which determines the quality and quantity of the activities modeled in the subsequent classification procedure. Further, the selection and weighting of

activities of RMS relevant visitor activities may not be adequate and exhaustive for the RMS in a company's multi-channel marketing strategy.

6 Acknowledgements

We thank Tobias Feldhaus, Martin Huth, and Boris Zielinski for their support during data collection and preparation.

References

- Aaker, D. (1993). Are brand equity investments really worthwhile? In D.A. Aaker and A. Biel (Eds.), *Brand equity and advertising: Advertising's role in building strong brands* (pp. 333-341). Hillsdale, NJ: Erlbaum.
- Argyriou, E., Kitchen, P.J. & Melewar, T.C. (2006). The relationship between corporate websites and brand equity: A conceptual framework and research agenda. *International Journal of Market Research*, 48, 575-599.
- Booth, D. & Jansen, B.J. (2009). A review of methodologies for analyzing websites. In B. J. Jansen, A. Spink and I. Taksa (Eds.), *Handbook of research on web log analysis* (pp. 141-163). Hershey, NY: IGI Global.
- Büchner, A.G. & Mulvenna, M. D. (1998). Discovering Internet marketing intelligence through online analytical web usage mining. *SIGMOD Record* 27, 4, 54-61.
- Breiman, L., Friedman, J., Stone, C.J. & Olshen, R. (1984). *Classification and regression trees*. Wadsworth, Belmont: Chapman & Hall.
- Bucklin, R.E. & Sismeiro, C. (2003). A model of web site browsing behavior estimated on clickstream data. *Journal of Marketing Research*, 40, 3, 249-267.
- Chi, S. & Tavella, D. (2008). Data mining and market intelligence for optimal marketing returns. Oxford, UK: Elsevier.
- Cooley, R., Mobasher, B. & Srivastava, J. (1999). Data preparation for mining world wide web browsing patterns. *Journal of Knowledge and Information Systems*, 1, 5-32.
- Fox, S., Karnawat, K., Mydlands, M., Dumais, S. & White, T. (2005). Evaluating implicit measures to improve web search. *ACM Transactions on Information Systems*, 23, 2, 147-168.

- Gale, B. (1994). *Managing Customer Value: Creating Quality and Service that Customers Can See*. New York, NY: The Free Press.
- Kellar, M, Watters, C. & Shepherd, M. (2006). A goal-based classification of web information tasks. *Proceedings of the American Society for Information Science and Technology, USA, 43, 1*, 1-22.
- Hochsztain, E., Millán, S., Pardo, B., Pena, J.M. & Menasalvas, E. (2003). A framework to integrate business goals in web usage mining. In E. Menasalvas, J. Segovia & P.S. Szczepaniak (Eds.), *Advances in Web Intelligence: First International Atlantic Web Intelligence Conference (AWIC 2003)* (pp. 28-45). Berlin, New York: Springer.
- Larose, D.T. (2005). *Discovering knowledge in data: An introduction to data mining*. New York, NY: Wiley.
- Lee, J., Hoch, R. Podlasek, M., Schonberg, E. & Gomory, S. (1999). Analysis and visualization of metrics for online merchandising, web usage analysis and user profiling. *Proceedings of the International WEBKDD'99 Workshop, USA*, 126-141.
- Lee, Y.W., Strong, D.M., Kahn, B.V. & Wang, R.Y (2002). AIMQ: a methodology for information quality assessment. *Information and Management, 40*, 133-146.
- Markov, Z. & Larose, D.T. (2007). *Data mining the web: Uncovering patterns in web content, structure, and usage*. New York, NY: Wiley.
- McKinney, V., Yoon, K. & Zahedi, F.M. (2002). The measurement of web-customer satisfaction: An expectation and disconfirmation approach. *Information Systems Research, 13, 3*, 296-315.
- Morgan, R. M & Hunt, S. D. (1994). The commitment-trust theory of relationship marketing. *The Journal of Marketing, 58, 3*, 20-38.
- Palmatier, R.W., Dant, R.P., Grwal, D. & Evans, K.R. (2006). Factors influencing the effectiveness of relationship marketing: A meta-analysis. *Journal of Marketing, 70*, 136-153.
- Reinartz, W.J. & Kumar, V. (2003). The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing, 67*, 77-99.
- Senecal, S., Kalczynski, P.W. & Nantel, J. (2005). Consumer's decision-making process and their online shopping behavior: A clickstream analysis. *Journal of Business Research, 58*, 1599-1608.

- Sheth, J. & Parvatiyar, A. (2000). Relationship marketing in consumer markets: Antecedents and consequences. In J. Sheth & A. Parvatiyar (Eds.), *Handbook of Relationship Marketing*. Thousand Oaks, CA: Sage.
- Spiliopoulou, M. & Pohle, C. (2001). Data mining for measuring and improving the success of web sites. *Data Mining and Knowledge Discovery*, 5, 85–114.
- Srinivasan, R. & Moorman, C. (2005). Strategic firm commitments and rewards for customer relationship management in online retailing. *Journal of Marketing*, 69, 193–200.
- Stolz, C., Viermetz, M., Neuneier, R. & Skubacz, M. (2005). Web performance indicator by implicit user feedback – applications and formal approach. *Proceedings of the sixth International Conference of Web Information Systems Engineering (WISE 2005)*, November 20-22, New York, NY, USA, 689-700.
- Teltzrow, M. & Berendt, B. (2003). Web-usage-based success metrics for multi-channel businesses. *Proceedings of the ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (WEBKDD Workshop on Web Mining 2003)*, USA, 17-27.
- Vercellis, C. (2009). *Business intelligence: Data mining and optimization for decision making*. New York, NY: Wiley.
- Weinberg, B. D., Parise, S., & Guinan, P. J. (2007). Multichannel marketing: Mindset and program development. *Business Horizons*, 50, 5, 385–394.

Analyse von Konsumentenmeinungen in Microblogs

Topic-based Opinion Mining

Andreas Schieber, Kai Heinrich, Andreas Hilbert

TU Dresden

Abstract

In diesem Beitrag steht die Analyse von Meinungsbildern in Blog-Communities, insbesondere in Microblogs, im Vordergrund. Motiviert durch die immer größere Bedeutung des Web 2.0 und der steigenden Relevanz des Social Commerce werden Verfahren analysiert, mit denen eine Meinungsanalyse durchgeführt werden kann. Es wird dabei die Übertragbarkeit dieser Verfahren auf die Microblog-Domäne untersucht und eine Empfehlung für geeignete und nicht geeignete Verfahren ausgesprochen. Der Beitrag endet mit einem Ausblick auf die weitere Forschung.

1 Einleitung

„Was mache ich gerade?“ beschreibt am besten die grundlegende Idee von Twitter¹. Über das soziale Netzwerk Twitter tauschen Personen Neuigkeiten oder Meinungen in kurzen Nachrichten aus. Twitter ist ein sogenannter Microblog, das ist eine spezielle Art von Weblog, die einen gewöhnlichen Blog mit Funktionen eines sozialen Netzwerks kombiniert. Twitter war im Jahr 2009 die populärste Microblog-Applikation mit mehr als 1,8 Millionen Nutzern in Deutschland (Petty & Stevens, 2009). Aufgrund der positiven Entwicklung von Microblogs (insbesondere von Twitter) werden diese Dienste zu einer wertvollen Quelle für Unternehmen ((Pak & Paroubek, 2010), (Barnes & Böhringer, 2009)). Denn mittels der Microblogs geben Kunden wertvolles Feedback zu Produkten und Dienstleistungen, indem sie ihre Meinungen untereinander austauschen und dabei auch die Meinungsbildung anderer Kunden beeinflussen (O'Connor, Balasubramanian, Routledge, & Smith, 2010).

¹ <http://www.twitter.com>

Das Ziel muss für Unternehmen daher darin bestehen, die von Kunden geäußerten Meinungen auf solchen Kommunikationsplattformen zu analysieren. Wegen der großen Anzahl an Einträgen auf diesen Plattformen ist es allerdings sehr schwierig, die relevanten Inhalte ohne den Einsatz automatischer Prozeduren zu filtern. Der Beitrag widmet sich dieser Problemstellung, indem er ein Konzept präsentiert, das die Einträge in einem iterativen Verfahren anhand mehrerer Kriterien analysiert. Im Ergebnis sollen daraufhin thematisch relevante Meinungsäußerungen leichter aufzuspüren sowie im Zeitverlauf zu betrachten sein.

2 Forschungsziel und Methodik

Das Ziel dieser Forschungsarbeit ist die Erstellung eines Konzepts, mit dessen Hilfe die Einträge von Microblogs bzgl. mehrerer Kriterien analysiert werden sollen, um relevante Äußerungen von der unübersichtlichen Masse zu trennen. Die Analyse soll dabei einerseits das Referenzpotenzial der Microblog-Nutzer, also von zentralen Personen innerhalb der Kommunikationsplattform, berücksichtigen. Andererseits sollen auch die Inhalte der veröffentlichten Einträge untersucht und zu Themenclustern zusammengefasst werden.

Zur Erreichung des Ziels sollen folgende Forschungsfragen beantwortet werden:

Wie können Meinungsführer in Microblogs identifiziert werden?

Wie können die Inhalte der Einträge in Microblogs zu Themenclustern zusammengefasst werden?

Das Konzept soll unter Nutzung des Design-Science-Forschungsansatzes von (Hevner, March, Park, & Ram, 2004) entwickelt werden, indem ein Artefakt implementiert wird, das die Evaluation der in Frage kommenden Methoden ermöglicht.

Als Datenquelle soll der Microblogging-Dienst Twitter dienen: Twitter bietet die populärste Microblogging-Plattform und verfügt daher über eine große Anzahl an Nutzern. Diese Nutzer generieren täglich eine hohe Zahl an Nachrichten, die meist öffentlich zugänglich publiziert werden (Pak & Paroubek, 2010).

Im weiteren Verlauf des Beitrags wird das Problem zunächst in das Forschungsfeld eingeordnet, und wichtige Begrifflichkeiten werden geklärt. Im Anschluss werden die Charakteristika von Microblogs beschrieben. Abschnitt 4 präsentiert auf Basis dieser Erkenntnisse ein mehrstufiges Prozessmodell für die Analyse von Kundenmeinungen in Microblogs.

3 Einordnung der Problemstellung

Der folgende Abschnitt ordnet die Arbeit in das Opinion Mining ein und zeigt auf, welche Forschungsfelder zur Lösung der Problemstellung beitragen können.

3.1 Opinion Mining

Die Methoden des Opinion Mining erlauben die Extraktion von Meinungen aus Texten. Das Opinion Mining kann daher wertvolle Informationen liefern, wenn Meinungen von Kunden über die Produktpalette eines Unternehmens ausgewertet werden sollen. Die Methoden des Opinion Mining werden wie Data- und Text-Mining-Methoden dem analytisch orientierten Business-Intelligence-Verständnis zugeordnet (Gluchowski, 2001).

In der Wissenschaft finden sich zahlreiche Definitionsversuche zum Begriff Opinion Mining (Xia, Gentile, Munro, & Iria, 2009, Dave, Lawrence, & Pennock, 2003, Kaiser & Bodendorf, 2009, Dey & Haque, 2009). (Liu, 2008) definiert das Opinion Mining als Extraktion von Attributen und Komponenten eines kommentierten Objekts; die Extraktion erfolgt dabei aus Textdokumenten, die positive, negative oder neutrale Meinungen über das Objekt enthalten.

(Liu, 2007) identifiziert im Rahmen des Opinion Mining drei Bereiche mit unterschiedlichen Aufgabenstellungen:

- „sentiment classification“: Die Analyse einer Aussage auf Dokumentenebene und Einstufung der Aussage als positiv oder negativ.
- „feature-based opinion mining and summarization“: Satzbasierete Identifikation von Eigenschaften (features) eines Objekts und Einstufung der Eigenschaften als positiv oder negativ.
- „comparative sentence and relation mining“: Der Vergleich von Objekten und Eigenschaften sowie Identifikation des präferierten Objekts.

Da es sich bei den Nachrichten in Microblogs grundsätzlich um Textdokumente handelt, eignen sich alle genannten Verfahren für die Analyse der Inhalte. Aufgrund der beschränkten Länge der Nachrichten ist es jedoch praktikabel, vor allem die Verfahren der sentiment classification einzusetzen: Über Microblogs werden meistens kurze Statements abgegeben, die selten über einen oder zwei Sätze hinausgehen; dementsprechend dürfte eine Analyse auf Dokumentenebene im Rahmen von Microblogs ähnliche Ergebnisse wie eine Analyse auf Satzebene erbringen.

3.2 Topic Modeling

(Blei & Lafferty, 2009) beschreiben Topic Models als eine leistungsstarke Technik zur unüberwachten Identifizierung von Strukturen in ansonsten unstrukturierten Dokumenten. Die Dokumente werden durch das Verfahren anhand der Verteilung der Wörter gruppiert, welche dazu tendieren, in ähnlichen Dokumenten gemeinsam aufzutreten. Diese Wortgruppen werden anschließend zu Themen zusammengefasst.

In einer vorgelagerten Forschungsarbeit wurde diese Methode im Rahmen der Analyse von Microblogs eingesetzt. (Schieber, Sommer, Hilbert, & Heinrich, 2011) haben mit Hilfe der Verfahren des Topic Modelling Microblog-Einträge über mehrere Zeiträume analysiert. Dabei zeigen die Ergebnisse, dass Themencluster identifiziert werden konnten, welche mit bestimmten, realen Ereignissen in Zusammenhang standen; so verwies ein Themencluster auf den Zusammenschluss von Sony, IMAX und Discovery zu einem Unternehmensnetzwerk (Schieber et al., 2011, S. 6).

Entgegen der Auswertung eines feststehenden Datensatzes von (Schieber et al., 2011) beschreiben (Blei & Lafferty, 2009) die Möglichkeit, neben den Inhalten der Einträge auch die Zeitkomponente zu untersuchen. Die sogenannten „dynamic topic models“ erfassen dabei die Evolution von Themen(clustern) im Verlauf der Zeit und zeigen auf, wie sich die Wortgruppen innerhalb eines Themas während der betrachteten Zeiträume verändern (Blei & Lafferty, S. 14ff.).

3.3 Soziale Netzwerke und Meinungsführer

Microblogs werden zu den sozialen Netzwerken zugeordnet. Ein soziales Netzwerk ist dabei die Komposition einer großen Anzahl von Akteuren, welche durch das Muster der zu Grunde liegenden Interaktion charakterisiert ist. (Iacobucci & Hopkins, 1992, S. 5). Die Nutzer von Microblogs veröffentlichen Nachrichten, die von anderen Nutzern gelesen werden können. Durch entsprechende Funktionen können sich Nutzer zu Netzwerken zusammenschließen, innerhalb derer ihre Nachrichten schneller kursieren. Nutzer, die eine zentrale Rolle in diesen Netzwerken einnehmen, besitzen deswegen eine mitunter große Zuhörerschaft und können Verhalten und Einstellungen von anderen Netzwerkmitgliedern beeinflussen (Kröber-Riel & Weinberg, 2003, S. 518). Im Forschungsfeld der Sozialen Netzwerkanalyse sind Verfahren bekannt, mit denen solche Meinungsführer aufgespürt werden können (vgl. Chin & Chignell, 2007 und Brin & Page, 1998).

3.4 Aufbau und Eigenschaften von Microblogs

Abgeleitet von der Nutzung als private Tagebücher bzw. Journale sind Microblogs eine besondere Form von Weblogs. Im Unterschied zu Weblogs kann hier nur eine beschränk-

te Anzahl von Zeichen zur Kommunikation genutzt werden. (Java, Finn, & Tseng, 2007) beschreiben Microblogging als eine neue Form der Kommunikation, bei welcher ein Nutzer in kurzen Mitteilungen seinen Status bekanntgeben kann.

Die Funktionsweise eines Microblogs soll an dem größten Vertreter Twitter erläutert werden. (Böhringer & Gluchowski, 2009) geben einen Überblick über die Funktionen des Dienstes: Ein Nutzer hat die Möglichkeit neben dem Verfassen von den auf 140 Zeichen begrenzten Statusmeldungen auch auf andere Statusmeldungen zu reagieren. Dazu gibt es zum einen eine Antwortfunktion (REPLY) und zum anderen eine Verbreitungsfunktion (RETWEET). Die Verbreitung der Einträge erfolgt dabei über die Netzwerkstruktur, welches sich durch die Folge-Funktion (FOLLOW) ergibt. Ein User hat die Möglichkeit, die Meldungen eines anderen Nutzer im Auge zu behalten, indem er ihm folgt. So bildet sich das Netzwerk um einen User aus den entsprechenden FOLLOW-Beziehungen.

4 Meinungsanalyse in Microblogs

Im Sinne der in Abschnitt 3.4 erwähnten Netzwerkstruktur ist es sinnvoll, die Meinungen derer zu analysieren, welche eine zentrale Stellung einnehmen und damit ein hohes Referenzpotenzial besitzen (Brüne, 1989). Nach der Anwendung eines entsprechenden Filters werden die Verfahren aus Abschnitt 3.1 zur Meinungsanalyse in Microblogs herangezogen. Im Sinne des Modells der zweistufigen Kommunikation (Brüne, 1989) stehen die resultierenden Meinungen dann stellvertretend für die Masse. Die Begrenzung der Datenmenge ist eine Notwendigkeit, da die algorithmischen Beschränkungen einiger Verfahren die Analyse solch enormer Datenmengen, wie sie bei der Verwendung von Microblogs anfallen, nicht zulassen. Neben der eigentlichen Analyse von Meinungen sollen ebenfalls Themen extrahiert werden, um abschließend eine Trendbetrachtung der unterschiedlichen Meinungen in gewissen Themenbereichen untersuchen zu können.

Entsprechend der Ziele, eine Trendmodellierung mit einer Themenstruktur unter Berücksichtigung der Besonderheiten von Microblogs bzgl. Inhalt und Struktur zu vereinbaren, wird das Modell von (Mei, Ling, Wondra, Su, & Zhai, 2007) auf die abgegrenzten Subnetzwerke der Meinungsführer angewendet.

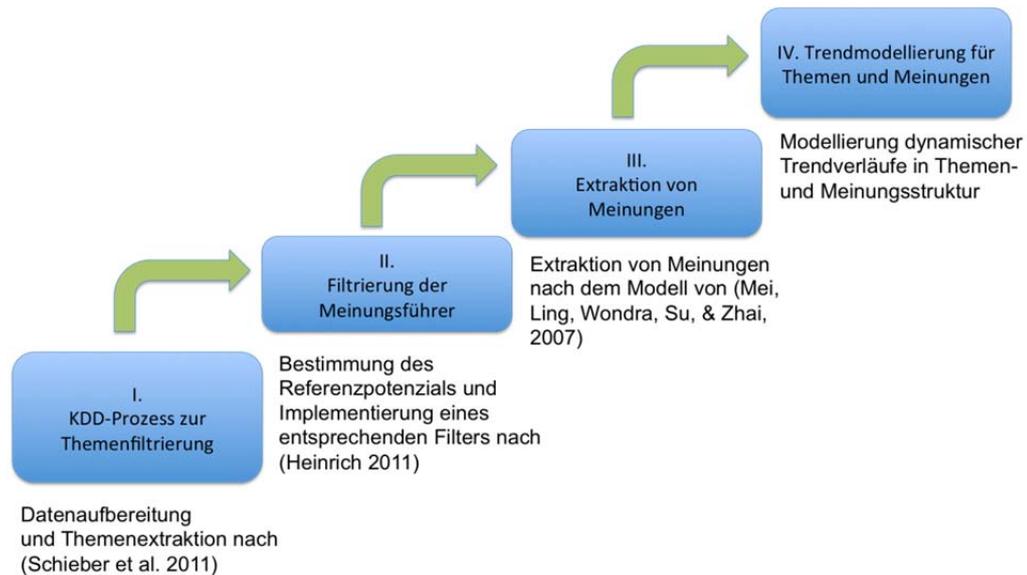


Abbildung 1: Prozessmodell zur Meinungsanalyse in Microblogs

Ziel nach der Datenaufbereitung und Themenfiltrierung (Schieber, Sommer, Heinrich, & Hilbert, 2011) in Schritt I ist die Filtrierung von Meinungsführern (Heinrich, 2011) in Schritt II. Danach können in Schritt III Meinungen extrahiert werden (Mei, Ling, Wondra, Su, & Zhai, 2007). Abschließend wird in Schritt IV die Abbildung von Meinungstrends und vor allem deren Prognose im Zeitverlauf angestrebt. Das komprimierte Prozessmodell ist in Abbildung 1 dargestellt.

5 Ausblick und Fazit

Nachdem die Frage „Worüber reden die Kunden?“ mit der explorativen Extraktion von Themen und die Zuordnung entsprechender beschreibender Wörter beantwortet ist (Schieber, Sommer, Heinrich, & Hilbert, 2011), stellen sich darauf aufbauend die Fragen: „Was denken die Kunden über dieses Thema“ und „Wie verändert sich diese Meinung Laufe der Zeit?“ Diese Arbeit schlägt ein entsprechendes Prozessmodell zur Beantwortung beider Fragen vor, welches auf vier Säulen beruht: Themenextraktion, Filtrierung von Meinungsführern, Extraktion von Themen und schließlich der Abbildung von Meinungstrends im Zeitverlauf.

In zukünftiger Forschungsarbeit sollen auf Basis der Erkenntnisse dieser Arbeit die beiden letzten Säulen ausgestaltet, implementiert und evaluiert werden.

Literaturverzeichnis

- Barnes, S., & Böhringer, M. (2009). Continuance Usage Intention in Microblogging Services: The Case of Twitter. *Proceedings of the 17th European Conference on Information Systems*, S. 1-13.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*.
- Blei, D. & Lafferty, J. *Topic Models*. Retrieved September 25, 2011, from <http://www.cs.princeton.edu/~blei/papers/BleiLafferty2009.pdf>.
- Böhringer, M., & Gluchowski, P. (2009). Microblogging. *Informatik-Spektrum*.
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Seventh International Word-Wide Web Conference*, from <http://ilpubs.stanford.edu:8090/361/1/1998-8.pdf>.
- Brüne, G. (1989). *Meinungsführerschaft im Konsumgütermarketing*.
- Chin, A., & Chignell, M. (2007). Identifying communities in blogs: roles for social network analysis and survey instruments. *International Journal of Web Based Communities*, 3(3), 345–363.
- Dave, K., Lawrence, S., & Pennock, D. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In ACM (Ed.), *Proceedings of the 12th international conference on World Wide Web* (pp. 519–528). New York, New York: ACM.
- Dey, L., & Haque, M. (2009). Opinion mining from noisy text data. *International Journal on Document Analysis and Recognition*, 3(12), 205–226.
- Gluchowski, P. (2001). Business Intelligence - Konzepte, Technologien und Einsatzbereiche. *HMD - Praxis der Wirtschaftsinformatik*, (222), 5–15. Retrieved September 20, 2010.
- Heinrich, K. (2011). Influence Potential Framework: Eine Methode zur Bestimmung des Referenzpotenzials in Microblogs. *Tagungsband zum 14. Interuniversitären Doktorandenseminar Wirtschaftsinformatik*. Universitätsverlag Chemnitz.
- Hevner, A., March, S., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*.
- Iacobucci, D., & Hopkins, N. (1992). Modeling Dyadic Interactions and Networks in Marketing. *Journal of Marketing Research*, 29(1), 5–17.
- Java, A., Finn, T., & Tseng, B. (2007). Why we twitter: understanding microblogging usage and communities. *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007*.

- Kaiser, C., & Bodendorf, F. (2009). Opinion and Relationship Mining in Online Forums. *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, (1), 128–131.
- Kröber-Riel, W., & Weinberg, P. (2003). *Konsumentenverhalten* (3. Auflage). München: Vahlen.
- Liu, B. (2007). *Web Data Mining: Exploring hyperlinks, contents, and usage data* (1. Auflage). Berlin, Heidelberg: Springer-Verlag.
- Liu, B. (2008). *Opinion Mining*. Retrieved July 25, 2011, from <http://www.cs.uic.edu/%7Eliub/FBS/opinion-mining.pdf>.
- Mei, Q., Ling, X., Wondra, M., Su, H., & Zhai, C. (2007). Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs. *Proceedings of the Sixteenth International World Wide Web Conference (WWW2007)*.
- O'Connor, B., Balasubramanyan, R., Routledge, B., & Smith, N. (2010). From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*.
- Pak, A., & Paroubek, P. (2010). 'Twitter as a Corpus for Sentiment Analysis and Opinion Mining'. *Proceedings of the International Conference on Language Resources and Evaluation*, S. 1320-1326.
- Pettey, C., & Stevens, H. (2009). Abgerufen am 12. 2010 von Gartner's Hype Cycle Special Report for 2009: <http://www.gartner.com/it/page.jsp?id=1124212>
- Richter, A., Koch, M., & Krisch, J. (2007). *Social Commerce - Eine Analyse des Wandels im E-Commerce*. Fakultät Informatik, Universität der Bundeswehr München.
- Schieber, A., Sommer, S., Heinrich, K., & Hilbert, A. (2011). Analyzing customer sentiments in microblogs – A topic-model-based approach for Twitter datasets. *Proceedings of the Seventeenth Americas Conference on Information Systems (forthcoming)*.
- Xia, L., Gentile, A., Munro, J., & Iria, J. (2009). Improving Patient Opinion Mining through Multi-step Classification. In V. Matoušek & P. Mautner (Eds.), *Lecture notes in computer science. Text, Speech and Dialogue* (pp. 70–76). Springer Berlin / Heidelberg.
- Zhang, Y., Callan, J., & Minka, T. (2002). Novelty and Redundancy Detection in Adaptive Filtering. *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Tampere, Finland*.