Potenziale wissenbasierter Analytischer Informationssysteme am Beispiel der Krankenhausmarktanalyse

Matthias Mertens
OFFIS - Institut für Informatik
Escherweg 2
26121 Oldenburg, Germany
mertens@offis.de

Abstract:
Durch einen höheren Bedarf an Gesundheitsleistungen bei gleichzeitig sinkender Zahl an Beitragszahlern hat sich die Marktsituation des deutschen Gesundheitssystems und insbesondere des Krankenhausmarktes verändert. Ein stärkerer Wettbewerbs- und Kostendruck zwingt Krankenhaus-Controller neue Potentiale zu erschließen und Leistungsangebote konkurrenzfähig auszubauen. Zur Sicherung der Wettbewerbsfähigkeit und zur Entscheidungsfindung müssen Informationen über das eigene Krankenhaus, über Wettbewerber und über allgemeine Statistiken zu aktuellen Entwicklungen vorliegen. Die Nutzung von analytischen Informationssystemen bietet sich aufgrund von multidimensionalen explorativen Analysen auf integrierten qualitätsgesicherten Datenbeständen zur Klärung fachspezifischer Fragestellungen an. Problematisch ist jedoch ihre Komplexität, wodurch nur geschulte Analysten mit entsprechendem Analyse- und Domänenwissen adäquate Analysen durchführen können. Schwerpunkt des hier vorgestellten Ansatzes ist die Konzeption einer semantischen Metadatenebene in einem analytischen Informationssystem, um explizit modelliertes Analyse- und Domänenwissen zu importieren, zu verwalten und für weiterführende Funktionalitäten der Analyse- und Entscheidungsunterstützung im Kontext der Krankenhausmarktanalyse zu nutzen.

Keywords
Analytische Informationssysteme, Data Warehouse, semantisches Metadatenmodell, Analyse- und Domänenwissen, Krankenhausmarktanalyse

1 Einleitung

Vor dem Hintergrund des steigenden Gesundheitsbewusstseins und des demografischen Wandels ist zu beobachten, dass der Bedarf an Gesundheitsleistungen zunimmt. Steigende Kosten müssen von immer weniger Beitragszahlern geschultert werden, wodurch sich finanzielle Auswirkungen im Gesundheitssektor ergeben. Besonders betroffen ist der Krankenhausmarkt [TRK+10]. Krankenhäuser (KH) sehen sich finanziellen Herausforderungen gegenüber, die auch aus der Einführung der Diagnosis Related Groups (DRG), d. h. eines auf diagnosebezogenen Fallgruppen basierenden Vergütungssystems, im Jahr 2003

erschienen im Tagungsband der INFORMATIK 2011
Lecture Notes in Informatics, Band P192
ISBN 978-3-88579-286-4
weitere Artikel online:
http://informatik2011.de/519.html

resultieren. Der verschärfte marktwirtschaftliche Kosten- und Wettbewerbsdruck bewirkt, dass Krankenhäuser sich nun verstärkt erfolgreich als wirtschaftliche Unternehmen am Markt positionieren und von ihren Konkurrenten abgrenzen müssen. Zur Sicherung der Wettbewerbsfähigkeit müssen die Potenziale zielgerichtet erschlossen und die Leistungsangebote konkurrenzfähig ausgerichtet werden.

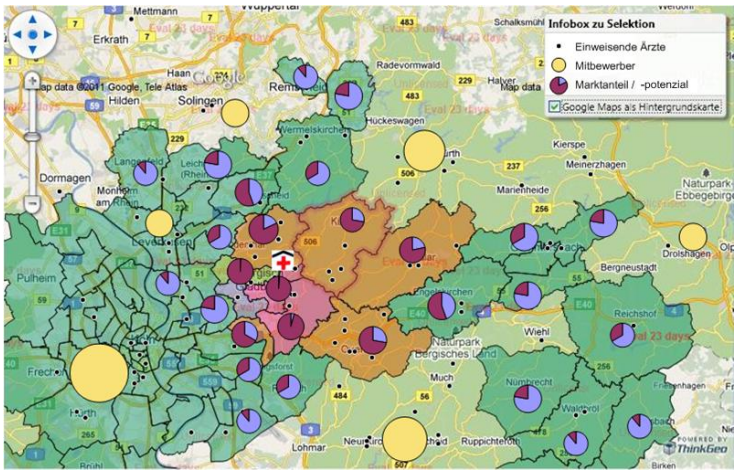

Abbildung 1: Marktpotenzial, Einweiser, Mitbewerber im Einzugsgebiet eines Krankenhauses

Zu diesem Zweck benötigen zuständige strategische Krankenhaus-Controller, sogenannte Business User, Informationen, auf deren Basis Entscheidungen zur Anpassung der eigenen Marktstrategie vorgenommen werden können. Von Interesse ist z. B. die aktuelle Markt- und Wettbewerbssituation - z. B. Einzugsgebiet, Marktanteil, Marktpotenzial, Wettbewerber (s. Abb. 1). Entsprechend müssen Krankenhäuser ihre internen §21 KHEntgG.-Daten mit verschiedenen zusätzlichen Daten, unter anderem von statistischen Ämtern (Mortalitäts- und Diagnosestatistiken), soziodemografische Daten (Bevölkerungsstatistiken), Einweiserdaten sowie Wettbewerbsdaten z. B. aus strukturierten Qualitätsberichten (SQB) über gemeinsame Dimensionen und Attribute integrieren und darauf basierend relevante Analysen durchführen. Mit Hilfe geografischer / statistischer Verfahren kann der Business User sich einen Überblick über die aktuellen und langfristigen Marktbedingungen verschaffen. Es können Informationen über Änderungen im Patientenspektrum, Einweiserverhalten, Versorgungslücken oder Regionen mit stillen Reserven gewonnen werden [SGDT11]. Diese Informationen können als Basis für weiterführende strategische Entscheidungen, wie z. B. dem Ausbau des eigenen Leistungsspektrums oder für explizites Einweisermarketing dienen.

Die Technologien der Business Intelligence (BI) können zur Informationsversorgung von Business User beitragen. Schlüsseleigenschaften der BI sind die Unterstützung des Managements und die Verbesserung der Entscheidungsfindungsprozesse. Um alle benötigten Informationen zur richtigen Zeit zu erhalten, haben sich im Kontext der BI Analytische Informationssysteme (AIS) entwickelt, die eine Bereitstellung, Visualisierung und Analyse von großen Datenmengen ermöglichen und im Rahmen der Krankenhausmarktanalyse

(KMA) genutzt werden könnten. AIS bestehen aus einem Data Warehouse [BG08] und darauf aufbauenden Analysekomponenten. Während ein Data Warehouse es ermöglicht, verschiedene Datenquellen qualitätsgesichert zu integrieren und multidimensional aufzubereiten, erlauben es Analysekomponenten Kennzahlen hinsichtlich verschiedener Dimensionen (z. B. Alter, Region, Geschlecht, ...) in einer zugehörigen Ergebnisvisualisierung (z. B. Tabellen, Diagrammen, Karten, ...) aufzuspannen und diese mit Hilfe von Online Analytical Processing (OLAP) Operationen sowie weiterführenden geografischen und statistischen Verfahren (z. B. Clusteranalysen, ...) multidimensional zu untersuchen.

Neben den genannten nützlichen Eigenschaften von AIS sind jedoch verschiedene Mängel zu beobachten. Diese werden im folgenden Abschnitt 2 beschrieben, bevor im Abschnitt 3 Anforderungen an einen Lösungsansatz diskutiert werden. Den Schwerpunkt des Abschnitts 4 bildet eine Architektur, die eine semantische Metadatenebene in einem AIS vorsieht. Anhand dieser Architektur werden die verschiedenen Konzepte zur Überwindung der genannten Mängel am Beispiel der KMA erläutert. In Abschnitt 5 werden im Anschluss verwandte Arbeiten in diesem Forschungsfeld aufgezeigt, bevor abschließend in Abschnitt 6 eine Zusammenfassung und ein Ausblick des Forschungsvorhabens gegeben werden.

2 Probleme von Analytischen Informationssystemen

Mit Hilfe von AIS können Analysen auf integrierten Datenbeständen durchgeführt werden, um Antworten auf Fragestellungen in verschiedenen Domänen zu erhalten. In dem am OFFIS entwickelten AIS - Multidimensional Statistical Data Analysis Engine (MUSTANG) [MTA09] - konnte beobachtet werden, dass in AIS verschiedene Mängel bestehen, die insbesondere Business User bei der Durchführung von Analysen vor signifikante Herausforderungen stellen.

Komplexität: Ein großes Problem stellt die hohe Nutzungs-Komplexität der AIS dar, die sich aus deren Flexibilität und Mächtigkeit ergibt. Um adäquate ad-hoc Analysen durchführen zu können, wird ein entsprechendes Analyse- und Domänenwissen sowie ein tiefer gehendes technisches und konzeptionelles Verständnis des multidimensionalen Datenmodells (MDM) vorausgesetzt. Über dieses verfügen Business User im Gegensatz zu Analysten jedoch nur selten. Hierbei enthält das Domänenwissen Informationen darüber, welche Fragestellungen einer zu untersuchenden Analyse zu Grunde liegen, mit welchen Kennzahlen Analysen zu spezifischen Fragestellungen möglich sind, welche explizite Semantik diese Kennzahlen haben und in welchen Beziehungen diese zueinander stehen. In Abgrenzung dazu umfasst Analysewissen Informationen zu den Analyseinstrumenten, d. h. den Analyseoperationen, Analyseverfahren und den möglichen Ergebnisvisualisierungen. Werden beide Wissensarten kombiniert, so lassen sich Aussagen zu den jeweiligen Analyseergebnissen erstellen. Schließlich wird eine hohe Interaktion zwischen den Business Usern und dem AIS benötigt, wodurch diese ohne entsprechende Schulung schnell überfordert sind [Mer11].

Erweiterte Metadaten: Die geringe Berücksichtigung von Metadaten, die zusätzliche In-

formationen über die quantitativen DWH-Daten sowie die DWH Struktur (Kennzahlen und Klassifikationselemente) bereitstellen, wird als weiterer Mangel gesehen [BRZ+10]. Zu diesen zählen Annahmen, Definitionen, Business Regeln, Terminologien und Hintergrundinformationen [ONe07]. Daher müssen Business User sich die Semantiken von Daten und Strukturen durch Zuhilfenahme externer Quellen selbst erschließen [BRZ+10]. Die Nutzung von externen strukturierten Daten, d. h. Daten die nicht Bestandteil des AIS sind, wird im Analysekontext des AIS bisher kaum berücksichtigt. Beispiele sind Daten des Projekts „Linked Open Data“ [BHIB08].

Verwaltung und Nutzung von „Know How“: Verbunden mit dem vorherigen Mangel ist auch das Problem zu sehen, dass AIS derzeit nicht in der Lage sind, modelliertes Analyse- und Domänenwissen von Analysten zu importieren, zu verwalten und für eine weiterführende Analyseunterstützung der Business User zu nutzen [Mer11]. Insbesondere sind die Analysestrategien, sogenannte „Best Practices“ für Analysen von Interesse.

Wünschenswert wäre es, wenn Business User mit geringerem Analyse- und Domänenwissen in einem komplexen AIS befähigt werden, sich selbst mit Informationen auf effiziente und intuitive Art versorgen zu können, ohne auf Analysten angewiesen zu sein. Dazu soll das AIS den Business User in seinen explorativen Analysen unterstützen, indem es modelliertes semantisches, maschinenlesbares und -verständliches Wissen importiert, verwaltet und in verschiedenen analyseunterstützenden Funktionen ausnutzt. Diese bilden einen Business User Self-Information Service [SKGS08, BRZ+10], der die wahrgenommene Komplexität des AIS senkt.

3 Anforderungen

Aus den in Abschnitt 2 beschriebenen Mängeln von AIS können verschiedene Anforderungen an einen Lösungsansatz abgeleitet werden. Diese sollen es allgemein und im speziellen in der Domäne erlauben, das Ziel der Informations-Selbstversorgung von Business Usern mit Hilfe von AIS zu erreichen.

Domänenunabhängiges AIS: Im Ansatz soll ein AIS zum Einsatz kommen, dessen Datenstrukturen, Analyseoperationen und -verfahren sowie Visualisierungen nicht auf eine spezifische Domäne eingeschränkt sind. Erweiterungen des AIS zur Erreichung des definierten Ziels sollen ebenfalls domänenunabhängig sein. Für eine Evaluation der Konzepte soll eine konkrete Domäne, z. B. KMA, instanziiert werden.

Komplexitätsreduktion: Der Ansatz soll eine Komplexität eines AIS reduzieren, ohne jedoch die Flexibilität und die Mächtigkeit des AIS einzuschränken. Der Ansatz soll es ermöglichen, dass das AIS für Business User mit geringer Expertise intuitiv nutzbar ist. Diese sollen befähigt werden, adäquate explorative ad-hoc Analysen auf dem DWH durchzuführen. Des Weiteren soll der Ansatz die Anzahl der benötigten Nutzer-Interaktionen, für das Erreichen äquivalenter Analyseergebnisse wie zuvor, reduzieren. Hierdurch soll die Effektivität des AIS gesteigert werden.

Analyseunterstützende Funktionen: Für die Informations-Selbstversorgung von Busi-

ness Usern werden unterstützende Funktionen auf Basis von explizit modelliertem, semantischen, maschinenlesbaren und -verständlichen Wissen in Form von Metadaten angestrebt. Daher soll das Konzept eine semantische Suche umsetzen, die es dem Business User ermöglicht, spezifische Metadaten zur Analyse, wie z. B. Fragestellungen, Analysevisualisierungen und -ergebnisse anhand ihrer Semantik zu suchen. Daneben soll der Nutzer entlang semantischer Beziehungen in den Metadaten navigieren können und es sollen ihm weiterführende sinnvolle Analyseschritte unterbreitet werden.

Verschiedenen Arten von Metadaten: Das Konzept soll es ermöglichen, mit verschiedenen Metadatenarten umzugehen und diese in einer intelligenten Art in den unterstützenden Funktionen zu verknüpfen. In der Literatur gibt es Ansätze [Inf04, SCM+05, BRZ+10], die weiterführende Informationen über die DWH Struktur und Dimensionselemente bereithalten. Dies können z. B. Hintergrundinformationen, synonyme Bezeichnungen, semantische Beziehungen zwischen den Dimensionselementen oder Business Rules sein. Ferner sollen Metadaten zum Analyseprozess berücksichtigt werden, die den Analyseprozess von einer Fragestellung über verschiedene Analyseschritte hin zu einem Analyseergebnis beschreiben. Diese Metadaten können synonym auch als Strategien oder „Best Practices“ bezeichnet werden und beschreiben das Analyse- und Domänenwissen eines Analysten. Schließlich soll das AIS Metadaten zu quantitativen Daten des DWH verarbeiten. Diese können Analyseergebnisse, wie z. B. Trends, Korrelationen, Auffälligkeiten oder geografische Ballungen beschreiben.

Flexibles Metadatenmodell: Zur Erfassung und Verwaltung sowie zur Nutzung verschiedener Metadatenarten in unterstützenden Funktionen wird ein flexibles Metadatenmodell benötigt. Dieses soll es auf einer generischen Ebene ermöglichen allgemein die genannten Metadatenarten zu beschreiben. Zusätzlich sollen darauf basierend domänenspezifische Instanzen der generischen Konzepte modelliert, ins AIS importiert, verwaltet und für die genannten Funktionen genutzt werden. Durch die instanziierten Konzepte wird konkretes Analyse- und Domänenwissen repräsentiert.

Performance: Nur wenn Anfragen an das Metadatenmodell, darauf basierende unterstützende Funktionen für den Business User sowie grundlegende OLAP Operationen auf dem multidimensionalen Modell des AIS im Zusammenspiel hinreichend gute Antwortzeiten liefern, ist eine interaktive Arbeit mit dem System möglich. Insbesondere vor dem Hintergrund eines explorativen Umgangs mit dem AIS ist dies von besonderer Bedeutung.

Unterstützung des Lernprozesses: Durch Nutzung des im AIS gespeicherten Wissens durch analyseunterstützende Funktionen, soll potenziell fehlendes Analyse- und Domänenwissen von Business Users ausgeglichen werden. Lernprozesse von Business Usern sollen unterstützt werden. Initialer Einarbeitungs- und Trainingsaufwand ließe sich durch das Erlernen des im System gespeicherten Analyse- und Domänenwissens reduzieren.

Domänenspezifische Instanziierung: Um domänenspezifisches Analyse- und Domänenwissen durch ein AIS nutzen zu können, muss dieses zuvor hergeleitet, auf Basis eines generischen Modells modelliert und in das AIS eingebracht werden. Soll beispielsweise die Domäne Krankenhausmarktanalyse durch das Konzept umgesetzt werden, so muss konkretes Analyse- und Domänenwissen von Analysten in Expertengesprächen sowie aus der Literatur erhoben werden.

Motiviert aus der in Abschnitt 2 gegebenen Problemstellung und den hier aufgeführten Anforderungen an einen Lösungsansatz werden im folgenden Abschnitt 4 eine Architektur, darin enthaltende Konzepte für ein Metadatenmodell sowie analyseunterstützende Funktionen vorgestellt.

4 Architektur und Konzepte

In diesem Abschnitt wird eine Architekturskizze vorgestellt, die den Aufbau eines AIS mit semantischer Metadatenebene zeigt. Ein Überblick über die enthaltenen Layer und ihre Services wird in Abschnitt 4.1 gegeben. Im anschließenden Abschnitt 4.2 werden die zugrunde liegenden Konzepte des semantischen Metadatenmodells näher betrachtet sowie darauf aufbauende unterstützende Funktionen im AIS erläutert.

4.1 Architekturskizze eines wissensbasierten AIS

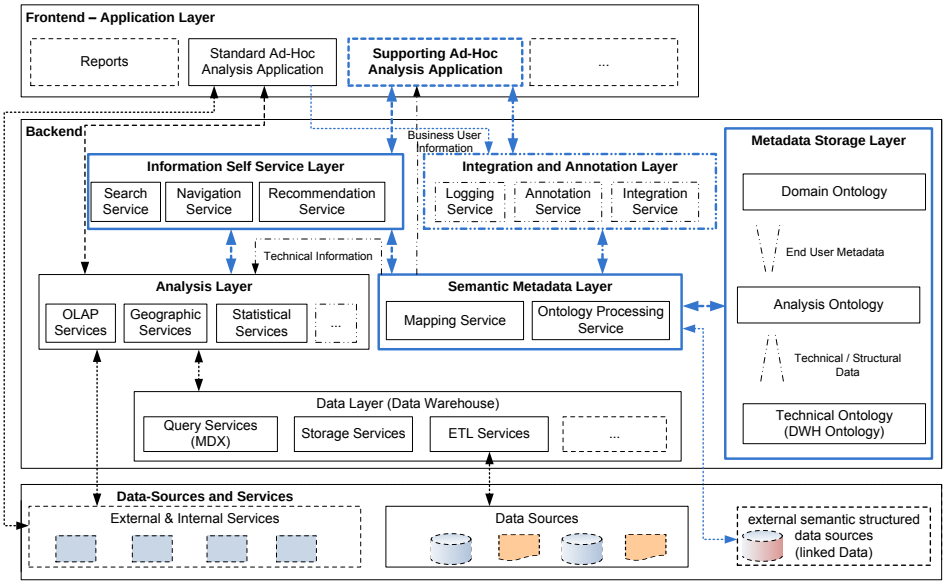

Abbildung 2: Architekturskizze eines um eine semantische Metadatenebene erweiterten AIS

In Abbildung 2 wird eine Architekturskizze gezeigt, die auf einer groben Ebene die verschiedenen „bestehenden“ Layer eines AIS, hier schwarz umrandet, illustriert (vgl. AIS MUSTANG in [TRM10]). Von Interesse sind die in blau dargestellten Layer, die als neue Komponenten in einem AIS konzipiert und entwickelt werden. Ziel ist es, eine Metadatenebene im AIS zu etablieren, die für die Importierung, Verwaltung und Nutzung von

Analyse- und Domänenwissen verwendet werden kann. Die hier dargestellte Architektur ist so aufgebaut, dass der Abstraktionsgrad von unten nach oben zunimmt.

Data-Sources und Services:
-Data-Sources: Auf unterster Ebene sind die externen Datenquellen vertreten, die für eine spätere Analyse qualitätsgesichert ins DWH geladen und miteinander integriert werden. Wie in Abschnitt 1 beschrieben, kommen im KMA Kontext als Datenquellen §21 KHEntgG.-Daten der jeweiligen Krankenhäuser oder der InEK GmbH und Daten der statistischen Ämter zu Mortalitäts-, Diagnose-, Bevölkerungsstatistiken in Frage. Des Weiteren sind die SQB der Krankenhäuser, die Indikatoren der Bundesgeschäftsstelle Qualitätssicherung gGmbH (BQS) sowie Daten zu einweisenden Ärzten zu berücksichtigen.

-Services: Neben den Datenquellen befinden sich auch externe Dienste auf der untersten Ebene. Hierbei kann es sich um Webservices, wie z. B. Google Maps oder um lokal vorliegende Dienste wie Geo-Server, Statistiksoftware, etc., die vom AIS für räumliche statistische Analysen genutzt werden, handeln.

-Linked Data: Strukturierte und semantische verknüpfte Daten, wie sie in der Linked Data Cloud [BHIB08] zu finden sind, bilden eine weitere Datenquelle, welche jedoch im Kontext von AIS kaum Beachtung findet. Eine Berücksichtigung wird im Rahmen dieses Forschungsvorhabens für eine intelligente Auswahl von Klassifikationselementen im MDM angestrebt.

Backend:
-Data Layer: Der Data Layer bildet das zentrale „Datenlager“ und wird im Kontext eines AIS durch ein Data Warehouse gebildet. Dieses wird nach [BG08] wie folgt definiert: „Ein Data Warehouse ist eine physische Datenbank, die eine integrierte Sicht auf beliebige Daten zu Analysezwecken ermöglicht.“ Das DWH verfügt über ein multidimensionales Datenmodell bestehend aus Würfeln, Dimensionen, Klassifikationshierachien, Klassifikationsstufen, Klassifikationselementen und Kennzahlen. Verschiedene Services ermöglichen im Rahmen des Data Warehousings, sogenannte Extraktions-, Transformations- und Ladeprozesse (ETL), um Datenquellen qualitätsgesichert ins DWH zu integrieren. Ein Storage Service verwaltet das MDM mit seinen Daten und ein Query Service, repräsentiert durch einen OLAP-Server, erlaubt die multidimensionale Anfrage der Kennzahlen in den „Würfel“-Zellen mittels standardisierten Sprachen wie z. B. Multidimensional Expressions (MDX).

-Analysis Layer: Der Analysis Layer ermöglicht es, mit dem OLAP-Service Anfragen an das MDM zu formulieren. Komplexere Kennzahlen können unter Berücksichtigung statistischer Verfahren und externer Statistik Software durch den Statistical Service berechnet werden. Sollen geografische Analysen ermöglicht werden, so müssen geografische Koordinaten z. B. von Krankenhäusern aus einem Geo-Server abgefragt und zugehörige Polygone auf einer Karte im Frontend, z. B. Google Maps, dargestellt werden. Diese Aufgaben können von einem Geographic Service übernommen werden.

-Metadata Storage Layer: Der Metadata Storage Layer bildet den Kern der semantischen Metadatenebene. In diesem Layer wird zum einen das generische Metadatenmodell, bestehend aus Klassen, Relationen zwischen diesen und zugehörigen Attributen, verwaltet. Zum anderen werden aber auch die eigentlichen Metadaten, welche als Instanzen der Klassen

ISBN 978-3-88579-286-4

und Beziehungen beschrieben werden, in diesem Layer abgespeichert. Als grundlegende Datenstrukturen werden Ontologien eingesetzt. Diese eignen sich aufgrund ihrer auf Logik basierenden Konzepte zur maschinenverständlichen Beschreibung von Analyse- und Domänenwissen. Die Ontologien sind in diesem Modell in verschiedene Ebenen eingeteilt, um die unterschiedlichen Themenbereiche und Abstraktionsstufen logisch abzugrenzen. Hierbei stehen die Ontologien in vielfältigen semantischen Beziehungen zueinander. Die Technical Ontology, die Analysis Ontology und Domain Ontology werden im Folgenden näher betrachtet:

-Technical Ontology: Die Technical Ontology (DWH-Ontologie) dient der semantisch reichhaltigen Beschreibung von Data Warehouse Metadaten. Die Ontologie modelliert auf einer Klassen-Ebene das für das Konzept benötigte Wissen über die dem AIS zugrunde liegende DWH-Struktur, d. h. sie modelliert die Entitäten des MDM, wie Würfel, Kennzahlen, Dimensionen, etc. sowie deren Beziehungen zueinander (vgl. [MTA09]). Auf einer Instanzebene können die Entitäten realer DWH-Systeme instanziiert werden, so dass domänenspezifische Instanzen der modellierten Klassen zur weiteren Verarbeitung bereitstehen. Im Architekturkonzept nimmt diese Ontologie die Aufgabe einer Vermittlungsschicht ein. Sie erlaubt es, dass in der Analysis und Domain Ontology die Real-Welt-Entitäten des DWH, abstrahiert durch die in der DWH-Ontologie instanziierten Entitäten, zugegriffen werden können. In der Literatur finden sich verschiedene Ansätze zur Bildung und Verwendung von DWH-Ontologien (vgl. [Har08, KGB10]).

-Analysis Ontology: Die Analysis Ontology modelliert abstrakt Entitäten und deren Beziehungen, die Analyseprozesse von multidimensionalen Datenstrukturen beschreiben. Diese decken den gesamten Analyseprozess von Fragestellungen, über Analyseketten hin zu Analyseergebnissen ab, beschreiben aber auch die enthaltenen Operatoren, Visualisierungen, Verfahren, Business Rules sowie die verschiedenen Akteure. Vielfältige semantische Beziehungen zu den Klassen der Technical Ontology sind definiert, um Analyseschritte und -ergebnisse im DWH zu referenzieren. Die Ontologie wird in einer abstrakten domänenunabhängigen Form vorgehalten, d. h. sie verfügt selbst über keine Instanzen. Domänenspezifische Erweiterungen werden in Domain Ontologies umgesetzt. Die einzelnen Klassen und ihre Beziehungen sind in [MTA09] im Detail visualisiert und erläutert.

-Domain Ontology: Sie hält das Analyse- und Domänenwissen für eine eindeutige Domäne vor. Hierbei erbt diese Ontologie die Klassen und deren Beziehungen von der Analyse-Ontologie und erweitert / verfeinert diese um domänenspezifische Konzepte. In ihr werden die konkreten Instanzen gebildet, die das Analyse- und Domänenwissen bilden. Anreicherungen durch das System um neue Informationen werden in dieser Ontologie vorgehalten.

-Semantic Metadata Layer: Der Semantic Metadata Layer ermöglicht dem AIS mit seinem Ontology Processing Service den technischen Zugriff auf die Ontologien. Er ist für die Verwaltung der semantischen Metadaten zuständig. Konkret bietet dieser Service eine API, um Metadatenstrukturen zu ändern und Metadaten abzufragen, einzufügen und zu ändern. Daher ist dies eine zentrale Komponente, die intern vom Mapping Service als auch extern von aufbauenden Layern verwendet wird. Der Mapping Service hat die Aufgabe eine Verbindung zwischen semantischen Metadaten (DWH-Ontology), welche reale Entitäten beschreiben, und den realen Entitäten des MDM herzustellen. Er fungiert als Mediator. Konkret werden durch diesen Service die semantischen Metadaten in techni-

sche Metadaten des zugrunde liegenden DWHs umgewandelt, so dass die Elemente des MDM gezielt angefragt werden können.

-Information Self Service Layer: Dieser Layer verwendet zur Verfügung stehende semantische Metadaten, um diese für verschiedene unterstützende Funktionen (s. Abschnitt 4.2) zu nutzen. Hierbei greift er lesend auf Metadaten zu. Die Funktionen werden durch Services realisiert, die von Frontend Applikationen angefragt werden können. Hierzu stellen sie eine API zur Verfügung. Selbst nutzen diese die API des Semantic Metadata Layer, um semantische und technische Metadaten zu erhalten. Semantische Metainformationen werden als sogenannte „Business User Informationen“ über die jeweiligen Dienste des Information Self Service Layer an den Endnutzer weitergegeben, während technische Metadaten für das Ansprechen von Funktionalitäten im Analysis Layer verwendet werden.

-Integration- und Annotation Layer: Dieser Layer beinhaltet verschiedene Services die primär einen schreibenden Zugriff auf den Metadata Storage Service ausüben und somit für die Weiterentwicklung bzw. die Anreicherung der semantischen Metadaten um neue Inhalte zuständig sind. Der Logging Service dient der Erfassung von Analyseschritten, die während des Analyseprozesses durchlaufen wurden. Die Protokollierung kann direkt auf den OLAP-Operatoren inkl. Angabe von relevanten / nicht relevanten Schritten, Business Rules, Visualisierungen erfolgen und somit die Definition von neuen Analysepfaden durch entsprechend geschulte Analysten ermöglichen. Anzumerken ist, dass der initiale Aufbau der semantischen Metadatenmodelle (Domain Ontology) außerhalb des AIS durchgeführt werden soll und in den Metadata Storage Layer über den Integration Service importiert wird. Der Annotation Service stellt Funktionen bereit, um quantitative Daten mit semantischen Annotationen zu versehen, so dass ein Analyseergebnis - z. B. Erkennen eines „Trends“ - direkt mit quantitativen Daten des DWHs verknüpft werden kann.

Frontend: Reports ermöglichen die automatisierte periodische Erzeugung von Berichten für Business User. Unter der Standard Ad Hoc Analysis Application wird eine Anwendung mit „standard“ Auswertungskomponenten verstanden, wie sie z. B. im MUSTANG AIS derzeit angewendet werden und die eine explorative Analyse ermöglichen. Die Supporting Analysis Application ist eine Standard Ad Hoc Analysis Application, die verschiedene Funktionen zur Analyseunterstützung für Business User bereitstellt. Hierzu nutzt diese im System modelliertes Analyse- und Domänenwissen aus, indem die Services des Information Self Service Layer verwendet werden.

Nachdem in diesem Abschnitt eine Architekturskizze mit verschiedenen Layern und zugehörigen Services vorgestellt wurde, wird im folgenden Abschnitt 4.2 genauer auf die zu speichernden Metadaten, die zu Grunde liegenden Konzepte und darauf aufbauende analyseunterstützende Funktionalitäten im Kontext der KMA eingegangen. Für eine detaillierte Beschreibung sei auch auf [Mer11] verwiesen.

4.2 Konzepte und Funktionen in einem semantischen Metadatenmodell

Im Metadatenmodell zu speicherndes Wissen soll Aussagen zu konkreten Fragestellungen, über durchzuführende Analyseschritte bis hin zu erkennbaren Analyseergebnissen

enthalten. Soll z. B. die Fragestellung „Welchen Marktanteil hat mein Krankenhaus (KH) im Einzugsgebiet?“ untersucht werden, so ist als Wissen modelliert, dass das Einzugsgebiet, bestehend aus Kernmarkt, erweiterter Kernmarkt und Peripheriemarkt [SGDT11], für das KH über dessen Fallzahlen ermittelt werden muss und dass der Marktanteil eine berechnete Kennzahl im DWH ist. Diese berechnet sich aus dem Verhältnis der erwarteten Fallzahl und den behandelten Fällen des KH. Des Weiteren ist z. B. als Wissen modelliert, dass sich aus der ersten Fragestellung weitere relevante Fragestellungen ableiten können, z. B. „Wie verhält sich die Fragestellung für Konkurrenten oder aber für spezifische Fachabteilungen (FA) des eigenen KH?“. Es kann zusätzlich modelliert sein, dass sich spezifische Visualisierungen zur Interpretation der quantitativen Daten im Kontext dieser Fragestellungen eignen und dass spezifische Elemente des MDM im Verlauf der Analyse berücksichtigt werden müssen. Für letztere Fragestellung müssen insbesondere die behandelten Diagnosen (ICD-Codes) und die vorhandene Ausstattung / vorhandenen Verfahren (OPS-Codes) sowie der Versorgungsschwerpunkt der FA analysiert werden. Modelliertes Analyse- und Domänenwissen, das beschreibt wie im Falle von erkannten Analyseergebnissen, wie z. B. erkannte Auffälligkeiten, Änderungen im Patientenspektrum / Einweiserverhalten [vSH09, SGDT11] weiter verfahren werden muss, lässt sich aus dem impliziten Wissen von Analysten ableiten.

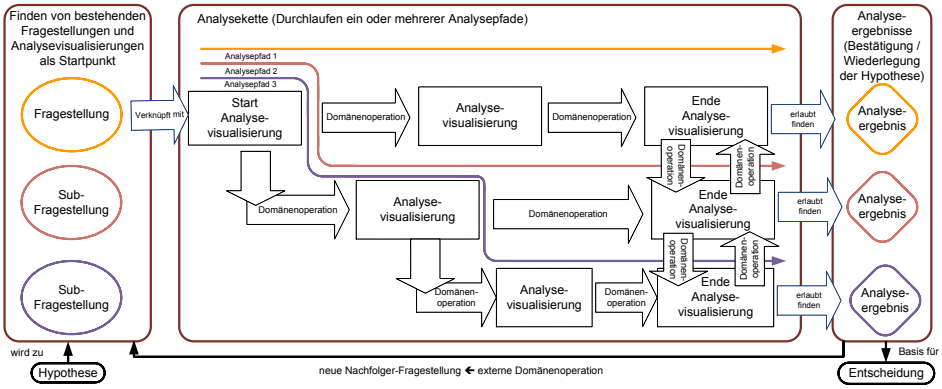

Abbildung 3: Zusammenhang von Konzepten im Analyseprozess

Soll dieses Analyse- und Domänenwissen zu einem Analyseprozess in einem Metadatenmodell explizit modelliert werden, so bedarf es verschiedener Entitäten, die in semantisch definierten Beziehungen zueinander stehen (s. Abb. 3). Hierbei gibt es Fragestellungen, die einer Hypothese Ausdruck geben und die es zu klären gilt, um auf deren Ergebnis eine Entscheidung zu fällen. Um zu einem Analyseergebnis zu gelangen, kann ein spezifischer Analysepfad durchlaufen werden. Dieser wiederum besteht aus einer Reihe von Analysevisualiserungen und Domänenoperationen zwischen diesen. Eine Analysevisualiserung ist als grafische Repräsentation der quantitativen Daten eines Analyseschrittes zu sehen und kann z. B. in Form einer Pivottabelle, eines Diagramms oder einer Karte dargestellt werden. Eine Domänenoperation ermöglicht das Navigieren zwischen zwei Analysevisualisierungen und kommt zum Einsatz, wenn spezifische Ausprägungen des MDM in einer Analysevisualisierung eintreten. Domänenoperationen können aus einer

Menge auszuführender OLAP-Operatoren, Business Rules und Visualisierungswechseln bestehen. Ihr Zweck ist es, die notwendigen Schritte der Kennzahlen-, Dimensionen-, Dimensionselemente- und Visualisierungsauswahl bzw. des -wechsels für den Endanwender durchzuführen. Business Rules repräsentieren konkretes abgeleitetes Analyse- und Domänenwissen. Sie beziehen sich auf eine Fragestellung, eine Analysevisualisierung oder ein Analyseergebnis, oder aber auf eine Kombination aus diesen und lassen sich in unterschiedlichen Domänenoperationen anwenden. Hieraus können sich mehrere Möglichkeiten für den Business User zum Durchlaufen der Analysepfade ergeben. Die Menge aller möglichen Analysepfade wird als Analysekette bezeichnet. Am Ende eines Analysepfades kann ein Analyseergebnis erreicht werden, welches eine Interpretation der quantitativen Daten im MDM des DWH beschreibt. Dies können Korrelationen und Zusammenhänge zwischen den Daten, aber auch Trends, Auffälligkeiten, Einbrüche, Zunahmen, geografische Ballungen, etc. sein.

Basierend auf den Konzepten des Metadatenmodells (s. auch [Mer11]) lassen sich verschiedene Funktionen, mit dem Ziel der domänenspezifischen Analyseunterstützung eines Business Users, im AIS umsetzten:

Semantische Suche: Ziel der semantischen Suche ist es, Instanzen des Metadatenmodells anhand ihrer Semantik zu suchen, zu finden und anschließend zu visualisieren. Die Semantik der Instanzen wird hierbei durch die Klassenzugehörigkeit, durch die Attribute und Attributausprägungen und durch die benannten Relationen zu anderen Klassen sowie deren Ausprägungen zu Instanzen definiert. So lassen sich bestehende Fragestellungen und damit verbundene Analyseketten finden, die sich auf die aktuellen zu Grunde liegenden quantitativen Daten des DWH adaptieren lassen. Neben dem Finden von Fragestellungen spielt aber auch das Finden von Analyseergebnissen und damit verbundenen quantitativen Daten des DWH eine Rolle. Annotationen auf den quantitativen Daten können mit ihrer Semantik gefunden und in der zugehörigen Analysevisualisierung wieder dargestellt werden. Ebenfalls lassen sich die durchlaufende Analysekette samt Fragestellung und analysierendem Business User ermitteln und als Ausgangsbasis für anknüpfende Analysen verwenden.

Semantische Navigation: Durch die semantischen Beziehungen zwischen den verschiedenen Entitäten ergeben sich semantische Netze. Diese erlauben es, von einer Entität zu einer anderen zu navigieren. Im Kontext eines AIS kann diese semantische Navigation zum einen als eine reine Navigation innerhalb der Metadaten erfolgen und zum anderen können zusätzliche assoziierte Operationen des AIS ausgeführt werden. Möchte man eine Antwort darauf erhalten, welche Vorgänger-Fragestellungen vor einer aktuell gewählten Fragestellung zwingend analysiert werden müssen, so kann man der transitiven Beziehung „notwendige Fragestellung“ im Metadatenmodell folgen. Möchte der Business User basierend auf einer Fragestellung eine konkrete Analyse durchführen, so kann er von der Fragestellung zu der verbundenen Start-Analysevisualisierung wechseln und dann entlang eines spezifischen Analysepfades zu einer Ende-Analysevisualisierung navigieren (s. Abb. 3). Diese Art der semantischen Navigation ist aufgrund der Darstellung von quantitativen Daten in Analysevisualisierungen und der Anwendung von OLAP-Operatoren bei Visualisierungswechseln immer mit realen Aktionen und Daten des AIS verbunden.

Vorschlagsgenerierung: Basierend auf dem im AIS gespeicherten Analyse- und Domän-

enwissen lassen sich im Kontext einer Analysevisualisierung Vorschläge zu weiteren sinnvollen folgenden Analysevisualisierungen geben, die der Klärung einer Fragestellung dienen. Zusätzlich können auch verwandte Fragestellungen aufgezeigt und weiterführende Hintergrundinformationen im Analyseprozess gegeben werden. Ersteres kann erreicht werden, indem Domänenoperationen und die enthaltenen Business Rules auf ihre Anwendbarkeit in der jeweiligen Analysevisualisierung überprüft werden. Ziel der Vorschlagsgenerierung soll es sein, dass relevante Analysevisualisierungen erreicht werden, die eine Interpretation der quantitativen Daten hinsichtlich der Fragestellung, zur Erstellung eines Analyseergebnisses zulassen. Zusätzlich sollen weiterführende Analysemöglichkeiten aufgezeigt werden. Eine kurze Erläuterung von Vorschlägen hilft dem Business User diese nachzuvollziehen und sich Analyse- und Domänenwissen anzueignen.

5 Verwandte Arbeiten

Im Kontext der Business Intelligence gibt es vielfältige Forschungsfragen, die von der Erschließung des Analysewissens [Baa06], über die Anpassung der BI-Tools an Business Anforderungen [SCM+05, BRZ+10] und die integrierte Anfrage von BI-Tools [CZL06, SKGS08] bis hin zur Annotation von Datenschemata mit weiterführenden Informationen zur Analyseunterstützung reichen [Inf04, SCM+05, BRZ+10]. In vielen Fällen soll der Analyst und teilweise auch der Business User [SKGS08, BRZ+10] stärker in den Fokus rücken, indem die BI-Software besser an individuelle Informationsbedürfnisse angepasst werden kann oder aber eine Analyseunterstützung zur individuellen Befriedigung des Informationsbedarfs angeboten wird. Viele der verwandten Arbeiten versuchen die jeweiligen adressierten Probleme durch die Nutzung semantischen Metadaten und damit verbundenen Semantic Web Technologien zu lösen.

Im Bereich der Dokumentation von Analyseprozessen und -ergebnissen ist die Arbeit [Baa06] zur Distribution von Business-Intelligence-Wissen zu nennen. Diese zielt auf die kontrollierte organisationsweite Verbreitung und Weiterverwendung von Berichten und Analyseansätzen durch das Einstellen der BI-Inhalte in Wissensmanagementsysteme ab, jedoch ohne auf eine konkrete technische Umsetzung einzugehen.

Im Kontext von Analyseprozessen ist auch die Arbeit von [Bis01] relevant, da diese den Begriff der Analysekette einführt. Fokus der Arbeit ist die Mensch-Maschine Interaktion, wobei das Finden, Aufbereiten und Darstellen von Daten als technisch / operative Tätigkeit gesehen wird. Zu den kognitiven Tätigkeiten zählen die Bildung eines anwenderorientierten Analysemodells sowie dessen Prüfung und Verfeinerung.

Eine andere Art der Analyseunterstützung für eine höhere Nutzerfreundlichkeit wird im Kontext der BI in der Trennung von Business und IT Belangen gesehen. Anders als in den zuvor genannten Arbeiten spielen hier semantische Metadaten eine entscheidende Rolle. Über die Metadatenmodelle können die Entitäten der zugrunde liegenden BI-Systeme mit ihren zugehörigen Relationen modelliert werden. Ebenfalls erlauben diese flexiblen, erweiterbaren Metadatenstrukturen weiterführende Informationen zu den Entitäten, wie beispielsweise Business Rules oder Expertenwissen festzuhalten. In [Inf04] wird diese

Möglichkeiten weiter vertieft und als Anwendung z. B. die Nutzung der Metadaten im ETL-Prozess für ein DWH angeführt.

In der Arbeit von [SCM+05] wird auf dieser Semantic Web Basis eine Architektur für analytische Tools vorgeschlagen, um Effizienzsteigerungen im Entscheidungsfindungsprozess zu erzielen. Durch die Nutzung einer Domänen-Ontologie, welche die Entitäten der zu untersuchenden Domäne, deren Relationen sowie weitere Informationen bereitstellt und durch die Nutzung einer Business-Intelligence-Ontologie, welche Informationen zu den Datenstrukturen des multidimensionalen Modells vorhält, soll eine Unterstützung im Analyseprozess erfolgen. Zum einen soll die Auswahl von Dimensionselementen durch proaktive Vorschläge erleichtert werden und zum anderen sollen durch ein semiautomatisches Umschreiben der Anfragen Analyseergebnisse aufgewertet werden.

Die Unterstützung eines technisch unversierten Business Users wird insbesondere in [SKGS08] und [BRZ+10] fokussiert. In [BRZ+10] liegt der Schwerpunkt auf einer kollaborativen ad hoc Entscheidungsunterstützung, in der Daten integriert aus den verschiedenen BI-Tools dargestellt und über semantische Metadaten mit weiterführenden Informationen versehen werden. Insbesondere sollen über Web 2.0 Technologien Information Mash-Ups gebildet, aber auch eine Kollabration zwischen verschieden Business Usern erzielt werden.

Auch wenn, wie in vielen anderen Arbeiten, die semantischen Metadaten eine entscheidende Rolle spielen, grenzt der eigene Ansatz sich von diesen durch die Modellierung von Konzepten und deren Instantiierung in Form von Analyse- und Domänenwissen ab. Im Fokus steht dabei das Wissen zu Analyseprozessen, das von verschiedenen Komponenten eines AIS zur Unterstützung des Business Users genutzt werden kann. Eine technologisch ähnliche Umsetzung mit Hilfe von verschiedenen Ontologien wie sie in [SCM+05] und [SKGS08] genannt werden, wird angestrebt.

6 Zusammenfassung und Fazit

Ziel dieser Arbeit war es, die Potenziale eines wissensbasierten analytischen Informationssystems für eine Informations-Selbstversorgung von Business Usern aufzuzeigen. Bisherige Analytische Informationssysteme setzen in der Regel aufgrund ihrer hohen Interaktionsmöglichkeiten und Komplexität entsprechendes Analyse- und Domänenwissen sowie ein tiefergehendes technisches und konzeptionelles Verständnis beim Nutzer voraus, um adäquate Analysen durchführen zu können. Dieses ist meist nur bei entsprechend geschulten Analysten vorhanden. Vor dem Hintergrund des sich verändernden Krankenhausmarktes, in dem sich Krankenhäuser verstärkt erfolgreich als wirtschaftliche Unternehmen am Markt positionieren und von ihren Konkurrenten abgrenzen müssen, werden auch die Business User der Krankenhäuser (Krankenhaus-Controller) in die Lage versetzt, adäquate Analysen durchführen zu müssen. Sie müssen Informationen aus ihren Daten gewinnen, um Potentiale zielgerichtet erschließen und die Leistungsangebote konkurrenzfähig ausrichten zu können.

Dieser Bedarf für eine Informations-Selbstversorgung von Business Usern im Kontext der

Krankenhausmarktanalyse wurde in Abschnitt 1 identifiziert. Das Analytische Informationssysteme sich im Kontext der Business Intelligence als Systeme zur Informationsgewinnung im Entscheidungsfindungsprozess etabliert haben, allerdings über verschiedene Mängel verfügen, die eine Informations-Selbstversorgung für Business User erschweren, wurde in Abschnitt 2 erläutert. Eine zu hohe Komplexität und keine Möglichkeiten Metadaten zu Business Regeln, Strategien oder Hintergrundinformationen bereitzustellen waren einige von diesen Mängeln. Im Abschnitt 3 wurden anschließend Anforderungen an einen Lösungsansatz definiert, um den genannten Mängeln zu begegnen und das Ziel einer Informations-Selbstversorgung für Business User in einem AIS zu ermöglichen. Der zugehörige Ansatz wurde anschließend in Abschnitt 4 präsentiert, wobei hier zunächst der Fokus auf einer Architekturskizze lag, die eine Erweiterung eines bestehenden AIS um eine semantische Metaebene, bestehend aus verschiedenen Layern und Services, beschrieb. Den Kern bildeten die verschiedenen, miteinander verknüpften Ontologien, welche das semantische Metadatenmodell aufbauen und die Speicherung von Analyse- und Domänenwissen ermöglichen. In diesem Zusammenhang wurden verschiedene Konzepte zur Modellierung von Analyseprozessen präsentiert sowie darauf aufbauende Funktionalitäten der Analyseunterstützung im Kontext eines AIS erläutert. Hier wurden eine Semantische Suche, eine Semantische Navigation und eine Vorschlagsgenerierung diskutiert. Der Abschnitt 5 gab einen Überblick über die verwandten Arbeiten in diesem Forschungsgebiet und erläuterte, wie sich der eigene Ansatz gegen diese abgrenzt.

Als weiterer Schritt in diesem Forschungsvorhaben werden die vorgestellten Konzepte in einem Metadatenmodell mittels Wissensrepräsentationssprachen modelliert und für die Domäne der Krankenhausmarktanalyse instanziiert. Dieses Metadatenmodell wird in eine zu implementierende semantische Metadatenebene des AIS MUSTANG eingebettet, um Analyse- und Domänenwissen zu verwalten und für die genannten Analyseunterstüzungsfunktionalitäten zu verwenden. Eine Umsetzung und Evaluierung wird im Rahmen einer laufenden Dissertation und einer studentischen Projektgruppe erfolgen.

Literatur

[Baa06] Henning Baars. Distribution von Business-Intelligence-Wissen - Diskussion eines Ansatzes zur Nutzung von Wissensmanagement-Systemen für die Verbreitung von Analyseergebnissen und Analsysetemplates. In Peter Chamoni und Peter Gluchowski, Hrsg., *Analytische Informationssysteme: Business Intelligence-Technologien und -Anwendungen (German Edition)*, Seiten 409–424. Springer, 3., vollst. überarb. aufl.. Auflage, 2006.

[BG08] Andreas Bauer und Holger Günzel, Hrsg. *Data Warehouse Systeme - Architektur, Entwicklung, Anwendung*. Dpunkt.Verlag GmbH, Heidelberg, Germany, 12 2008.

[BHIB08] Christian Bizer, Tom Heath, Kingsley Idehen und Tim Berners-Lee. Linked data on the web (LDOW2008). In *Proceeding of the 17th international conference on World Wide Web*, Seiten 1265–1266, Beijing, China, 2008. ACM.

[Bis01] Nicolas Bissantz. DeltaMiner. *Wirtschaftsinformatik*, Volume 43(1):77–80, 2001.

[BRZ+10] Henrike Berthold, Philipp Rösch, Stefan Zöller, Felix Wortmann, Alessio Carenini, Stuart Campbell, Pascal Bisson und Frank Strohmaier. An architecture for ad-hoc and collaborative business intelligence. In *Proceedings of the 2010 EDBT/ICDT Workshops*, EDBT '10, Seiten 13:1–13:6, New York, NY, USA, 2010. ACM.

[CZL06] Longbing Cao, Chengqi Zhang und Jiming Liu. Ontology-based integration of business intelligence. *Web Intelligence and Agent Systems*, Volume 4(3):313 – 325, 2006.

[Har08] Stefan Hartmann. *Überwindung semantischer Heterogenität bei multiplen Data-Warehouse-Systemen*. University of Bamberg Press, Bamberg, Germany, 2008.

[Inf04] Network Inference. Ontology and Data Warehousing. Technology white paper, NETWORK INFERENCE, INC., 2004.

[KGB10] Christian Kurze, Peter Gluchowski und Martin Bohringer. Towards an Ontology of Multidimensional Data Structures for Analytical Purposes. In *HICSS '10: Proceedings of the 2010 43rd Hawaii International Conference on System Sciences*, Washington, DC, USA, 2010. IEEE Computer Society.

[Mer11] Matthias Mertens. Wissensbasiertes Business Intelligence für die Informations-Selbstversorgung von Entscheidungsträgern. In *Proceedings of the 23. GI-Workshop on Foundations of Databases (Grundlagen von Datenbanken), Obergurgl, Tirol, Österreich, May 31 - June 03, 2011*. Universität Innsbruck, 2011.

[MTA09] Matthias Mertens, Yvette Teiken und H.-Jürgen Appelrath. Semantische Anreicherung von strukturierten Daten und Prozessen in analytischen Informationssystemen am Beispiel von MUSTANG. In *Forschungskolloquium Business Intelligence 2009 der GI Fachgruppe 5.8 - Management Support Systems , Dortmund, Deutschland, October 08 - 09, 2009*. Universität Dortmund, 2009.

[ONe07] B. ONeil. Semantics and Business Meta Data In The Data Administration. Besucht am 28.06.2011; URL: http://www.tdan.com/view-articles/4934, 2007.

[SCM+05] Denilson Sell, Liliana Cabral, Enrico Motta, John Domingue, Farshad Hakimpour und Roberto Pacheco. A Semantic Web Based Architecture for Analytical Tools. In *CEC '05: Proceedings of the Seventh IEEE International Conference on E-Commerce Technology*, Seiten 347–354, Washington, DC, USA, 2005. IEEE Computer Society.

[SGDT11] Rosemarie Stibbe, Johannes Güsgen, Alexandra Dierkes und Melanie Tilgen. Geocodierung als Instrument zur Marktanalyse. In *KU Gesundheitsmanagement, 1/2011, S. 29 - 31*. Mediengruppe Oberfranken Buch- und Fachverlage GmbH u. Co. KG, 2011.

[SKGS08] Michael Spahn, Joachim Kleb, Stephan Grimm und Stefan Scheidl. Supporting business intelligence by providing ontology-based end-user information self-service. In *OBI '08: Proceedings of the first international workshop on Ontology-supported business intelligence*, Seiten 1–12, New York, NY, USA, 2008. ACM.

[TRK+10] Wilfried Thoben, Martin Rohde, Sascha Koch, H.-Jürgen Appelrath und Ralph Stuber. Konzepte und Technologien für die strategische Planung im Krankenhausmarkt. *Krankenhaus IT Journal*, Volume 5, 2010.

[TRM10] Yvette Teiken, Martin Rohde und Matthias Mertens. MUSTANG: Realisierung eines Analytischen Informationssystems im Kontext der Gesundheitsberichtserstattung. In *Informatik 2010: Service Science - Neue Perspektiven für die Informatik*, Seiten 53–68. CEUR Workshop Proceedings, 2010.

[vSH09] Nikolai von Schroeders und Christoph Heller. *Geocoding - Geografische Analyse für Krankenhäuser*. Baumann Fachzeitschriften Verlag, Kulmbach, Germany, 2009.

erschienen im Tagungsband der INFORMATIK 2011
Lecture Notes in Informatics, Band P192
ISBN 978-3-88579-286-4
weitere Artikel online:
http://informatik2011.de/519.html