

Arbeitsberichte des Instituts für Wirtschaftsinformatik

Herausgeber: Prof. Dr. J. Becker, Prof. Dr. H. L. Grob, Prof. Dr. S. Klein,
Prof. Dr. H. Kuchen, Prof. Dr. U. Müller-Funk, Prof. Dr. G. Vossen

Arbeitsbericht Nr. 80

**Gestaltung von Führungsinformationssystemen
mittels Informationsportalen:
Ansätze zur Integration von Data-Warehouse-
und Content-Management-Systemen**

Jörg Becker, Ralf Knackstedt, Thomas Serries

ISSN 1438-3985

Westfälische Wilhelms-Universität Münster · Institut für Wirtschaftsinformatik
Leonardo-Campus 3, 48149 Münster, Tel. (0251) 83-38100

November 2001

Inhalt

1	Konzepte für Führungsinformationssysteme	3
2	Kritische Bestandsaufnahme zur Gestaltung von Informationsportalen	8
2.1	Data Warehousing und OLAP	8
2.2	Content Management und Informationsportale	10
2.3	Information Retrieval in Texten	11
2.3.1	Automatische Schlagwortidentifikation	12
2.3.2	Information Retrieval mit Schlagworten	13
2.3.3	Strukturierung von Schlagworten	14
2.4	Verbindung von OLAP- und Content-Management-Systemen	15
2.5	Entwicklungsbedarfe	15
3	Ansätze zur Weiterentwicklung von Informationsportalen	20
3.1	Metadaten-basierte Inhalteintegration	20
3.2	Metadaten-basierte Inhaltnavigation	25
3.3	Inhalteklassifikation anhand von Metadaten und Schlüsselwörtern	28
3.3.1	Kollaborative Inhaltebeurteilung	29
3.3.2	Profil-Definition mit Wenn-Dann-Regeln	29
3.3.3	Kollaborative Profile	30
3.3.4	Profilentwicklung mittels lernender Verfahren	31
3.4	DV-Konzeption zur Umsetzung der Ansätze	34
4	Verwandte Arbeiten	38
5	Zusammenfassung und Ausblick	40
	Literaturverzeichnis	42

1 Konzepte für Führungsinformationssysteme

Das *Führungssystem* eines Unternehmens hat die Aufgabe, auf der Grundlage von Informationen¹ über den Stand und die Bedingungen des Ausführungssystems Ausführungshandlungen möglichst effizient und effektiv zu lenken.² Informationen bilden nicht nur die Grundlage für die Vorbereitung und Ausgestaltung von Transaktionen, sondern stellen einen wichtigen Faktor im Wettbewerb zwischen Unternehmen dar.³ „Unternehmerische Ideen und unternehmerischer Erfolg resultieren aus einem Informationsvorsprung gegenüber anderen Wirtschaftssubjekten. Dieser äußert sich in drei wichtigen Funktionen des Unternehmertums im Marktprozess, nämlich in der Innovations-, der Koordinations- und der Arbitragefunktion“.⁴

Führungsinformationssysteme sollen den Aufgabenträgern des Führungssystems (Manager) technisch unterstützt die für ihre Steuerungs- und Regelungsaufgaben relevanten Informationen liefern.⁵ Um diesem Anspruch umfassend gerecht werden zu können, müssen unternehmensinterne und -externe sowie quantitative und qualitative Informationen zur Verfügung gestellt werden.⁶

Bereits 1958 diskutierten LEAVITT und WHISLER den Einfluss informationstechnologischer Innovationen auf die Arbeit von Managern.⁷ Seitdem wurden in Abhängigkeit von den hard- und softwaretechnischen Rahmenbedingungen regelmäßig neue Konzepte zur Gestaltung von Führungsinformationssystemen entwickelt.⁸ In einer ersten Euphorie wurde unter dem Begriff *Management Information System* (MIS) die Vision einer weitreichenden Rationalisierung der Managementtätigkeit durch Informationssysteme auf der Basis eines Total Systems Approach diskutiert. Nach dem zwangsläufigen Scheitern dieser Ansätze wurden die Konzepte für Führungsinformationssysteme auf realistischere Zielsetzungen ausgerichtet. *Management Reporting Systems* (MRS) berücksichtigten die batch-orientierte Verarbeitung von Massendaten und stellen listenbasierte Standardberichte mit eingeschränkter interaktiver Auswertungsflexibilität zur Verfügung. *Decision Support Systems* (DSS) spezialisierten sich auf die Unterstützung bestimmter semi-strukturierter und hinreichend modellierbarer Entscheidungsprobleme. Mit

1 Vgl. Schütte (1998), S. 1-5.

2 Vgl. Weber (1993), S. 167.

3 Vgl. Weiber, McLachlan (2000), S. 120.

4 Picot, Reichwald, Wiegand, S. 29.

5 Vgl. Becker, Holten (1998), S. 485.

6 Vgl. Mertens (1999), S. 405.

7 Vgl. Leavitt, Whisler (1958), S. 46.

8 Vgl. zur historischen Entwicklung der „Computerunterstützung für das Management“ insb. Oppelt (1995).
Siehe auch Holten, Knackstedt (1997).

dem Aufkommen grafischer Benutzeroberflächen zielten *Executive Information Systems* (EIS) auf die Bereitstellung wichtiger Informationen für das Top-Management in Form leicht bedienbarer, in Analogie zu Flugzeug-Cockpits gestalteter Bildschirmmasken. Das Executive Information System sollte die Arbeit des Managements umfassend unterstützen und weitere für das Tagesgeschäft wichtige Applikationen wie E-Mail, Newsticker, Terminkalender etc. über eine gemeinsame Benutzeroberfläche zugänglich machen.⁹ Die gegenwärtige Diskussion der Gestaltung von Führungsinformationssystemen wird durch auf quantitative Daten spezialisierte Data-Warehouse- und OLAP-Systeme (Online Analytical Processing) dominiert.¹⁰ Mit der Verbreitung der Internet-Technologie wurde die Zielsetzung der Executive Information Systems wieder aufgegriffen, eine integrierte, personalisierte Schnittstelle zu Inhalten und Anwendungen unterschiedlicher Herkunft zur Verfügung zu stellen. Ihre Umsetzung wird heute unter dem Begriff des *Enterprise Information Portal* oder kurz *Informationsportal* diskutiert. Der Begriff wurde von SHILAKES und TYLMAN wie folgt geprägt: "Enterprise Information Portals enable companies to unlock internally and externally stored information and provide users a single gateway to personalized information needed to make informed business decisions."¹¹

Zu Recht wird der Versuch kritisiert, ex-post dogmatische Begriffsnormen für Konzeptionen von Führungsinformationssystemen definieren zu wollen.¹² Nutzenstiftender ist sicherlich die Identifikation von Vorteilen verschiedener Konzepte, die sich zu verbesserten Lösungen der Managementunterstützung kombinieren lassen. In dieser Arbeit werden daher die Stärken und Schwächen bestehender Systeme zum Aufbau von Informationsportalen für das Management untersucht und darauf aufbauend Vorschläge für eine vorteilhafte Integration dieser Lösungen entwickelt. Das Integrationskonzept soll insbesondere den folgenden Anforderungen gerecht werden.

Zusammenführen von internen und externen Informationen

Die Unternehmenssituation lässt sich weder aufgrund ausschließlich externer Informationen noch aufgrund ausschließlich interner Informationen beurteilen. So stellt MERTENS fest, dass Führungsentscheidungen zum Großteil basierend auf externen, qualitativen Informationen vorbereitet werden.¹³ Informationsportale müssen somit einerseits sowohl Informationen aus unternehmensinternen als auch Inhalte aus unternehmensexternen Quellen in einem System

⁹ Vgl. Back-Hock (1991), S. 49.

¹⁰ Vgl. hierzu ausführlich Kapitel 2.

¹¹ Shylakes, Tylman (1998), S. 3.

¹² Vgl. Bullinger, Koll (1992), S. 50.

¹³ Vgl. Mertens (1999), S. 405.

zusammenführen. Unabhängig von der Herkunft können Informationen sowohl qualitativer als auch quantitativer Natur sein. Dabei ist dem Umstand Rechnung zu tragen, dass externe Informationen selten in einer Strukturierung vorliegen, die für eine Verwaltung in einer relationalen Datenbank geeignet ist.

Gemeinsamer Zugang zu allen Informationen

Im Rahmen einer Entscheidungsvorbereitung ist eine angemessene Versorgung mit Informationen unterschiedlichen Typs unverzichtbar. Auf der Suche nach entscheidungsrelevanten Informationen sind daher alle Informationstypen gleichermaßen zu berücksichtigen. Mit der Bereitstellung von qualitativen und quantitativen bzw. unternehmensexternen und unternehmensinternen Informationen in einem System ist jedoch nur die Grundvoraussetzung hierfür geschaffen. Darüber hinaus muss es das Ziel sein, alle Informationen in einer einheitlichen, vom Informationstyp unabhängigen Gliederung zu strukturieren und so eine Integration der quantitativen und qualitativen Datenbestände zu erreichen. Ein Mehrwert der Integration muss dabei sein, dass alle Inhalte unabhängig von ihrem Typ und ihrer Herkunft zueinander in Beziehung gesetzt werden und so ein Netzwerk aus miteinander in Beziehung stehenden Inhalten entsteht. Um die Akzeptanz der Struktur durch die Benutzer sicherzustellen, ist diese an die im Unternehmen verwendeten Begriffe anzupassen. Durch die bedarfsgerechte Anordnung soll eine Reduzierung des Einarbeitungsaufwandes erzielt werden.

Personalisierte Informationsversorgung

Mit zunehmender Verbreitung digitaler Informationsverarbeitung und digitaler Geräte haben sowohl die Verfügbarkeit von Informationen stark zugenommen als auch die Zeitspannen, in denen Wissen veraltet bzw. neues Wissen generiert wird, stark abgenommen. Die begrenzte Informationsverarbeitungskapazität des Menschen lässt eine vollständige Verarbeitung der verfügbaren Informationen nicht zu. Daher müssen Selektionsverfahren diese Menge auf ein angemessenes Maß reduzieren.

Aus Gründen der Komplexitätsbeherrschung schaffen sich Entscheider eine pragmatisch bedingte Vorstellung ihrer Umwelt. Diese nach LUHMANN *Subjektivierung* genannte Strategie äußert sich in einer subjektabhängigen Verarbeitung von Informationen und bedingt subjektabhängige Erfordernisse bei der Erschließung von Informationen.¹⁴ Schon die Tatsache, dass Menschen unterschiedliches Vorwissen besitzen, lässt die Forderung nach einer individualisierten Informationsversorgung sinnvoll erscheinen. Um Unterschiede bei der Wahrnehmung

¹⁴ Vgl. zur Subjektivierung Luhmann (1999), S. 179-183.

von Informationen zu berücksichtigen, scheint es unverzichtbar, die Art der Informationspräsentation vom Benutzer mitbestimmen zu lassen.

Push-/Pull-Kommunikation

Aktive und passive Informationsbereitstellung wird häufig mit Push- bzw. Pull-Mechanismen gleich gesetzt.¹⁵ Die Prinzipien Pull und Push werden in der Literatur unter Vermengung technischer und organisatorischer Charakteristika eher unpräzise abgegrenzt. Unter Rückgriff auf die Sprechakttheorie¹⁶ können die Prinzipien folgendermaßen unterschieden werden: Das *Push-Prinzip* stellt eine durch Aufforderung initiierte Bereitstellung von Inhalten dar. Aufforderungen gehören nach AUSTIN zu den exerzitiven Äußerungen und zeichnen sich dadurch aus, dass sie Entscheidungen für oder gegen etwas darstellen und andere verpflichten oder ermächtigen können¹⁷. Unter dem *Pull-Prinzip* wird dagegen die Bereitstellung von Inhalten in Form eines Frage-Antwort-Dialogs verstanden. Als das charakteristische Merkmal einer Frage wird der Versuch angesehen, dem Gefragten gewisse, diesem voraussichtlich bekannte Informationen zu entlocken.¹⁸ AUSTIN zählt Fragen zu den expositiven Äußerungen, für die der Ursprung in der Wechselrede im Allgemeinen typisch ist.

Um Informationsbereitstellungsservices zur Verfügung stellen zu können, werden unabhängig vom Bereitstellungsprinzip ergänzend Verwaltungsfunktionen benötigt (vgl. Abb. 1). Die Verwaltungskomponenten gehen daher nicht in die Unterscheidung zwischen Pull- und Push-Prinzip ein. Zu ihnen gehören insbesondere die Kommunikation des Serviceangebots (Inhaltemarketing), die Vergabe von Zugriffsrechten (Inhaltevertrieb) und die Zahlungsabwicklung (Kreditoren- bzw. Debitorenbuchhaltung).¹⁹

Für vom Manager initiierte Planungen haben sich Pull-Verfahren, bei denen der Entscheidungsträger gezielt nach Informationen sucht, durchgesetzt. Mit der Navigation anhand von Gliederungsstrukturen oder einer Volltextsuche in allen Informationen bieten sie die Möglichkeit, Informationen zu einer aktuell gegebenen Situation zu recherchieren. Push-Verfahren hingegen basieren auf systemseitigen Hypothesen über den Informationsbedarf des Anwenders und ermöglichen darauf aufbauend eine an seinen Bedürfnissen ausgerichtete Information des Anwenders. In Informationsportalen können Push-Verfahren die Ergebnismenge einer Anfrage bzw. Suche automatisch nach den Bedürfnissen des Anwenders sortieren. Darüber hinaus können sie situationsabhängig ein geeignetes Kommunikationsmedium auswäh-

¹⁵ Vgl. Lüdi (1997).

¹⁶ Vgl. hierzu Austin (1979).

¹⁷ Vgl. Austin (1974), S. 173.

¹⁸ Vgl. Searle (1997), S. 102.

¹⁹ Anders z. B. Kurz (1999), S. 485.

len: Beispielsweise werden schnellere Medien (Telefon, SMS, E-Mail) eingesetzt, wenn das Informationsportal anhand der ihm vorliegenden Informationen eine Ausnahmesituation feststellt, in der die Benachrichtigung der Entscheidungsträger zur Vermeidung größerer Nachteile erforderlich ist.

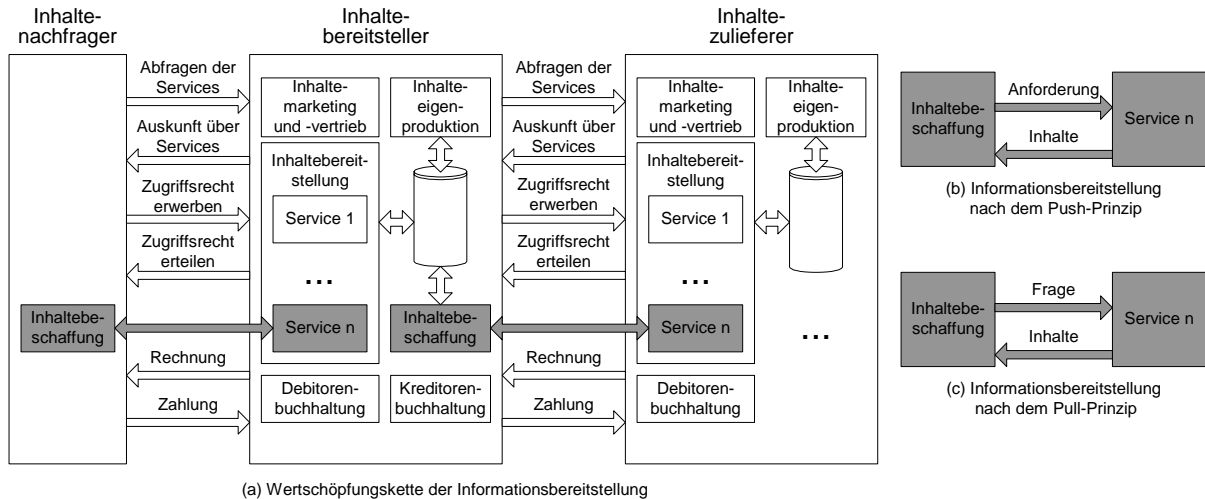


Abb. 1: Pull- und Push-Informationsbereitstellung

Aufbau der Arbeit

Aufgrund der identifizierten Anforderungen an ein umfassendes Informationsportalkonzept erscheint es notwendig, dass Führungsinformationssysteme verschiedene zum Teil bisher isoliert verfolgte Ansätze zur Umsetzung der Informationsbereitstellung kombinieren. Zunächst wird daher im Folgenden eine kritische Bestandsaufnahme existierender Lösungskonzepte vorgenommen (Kapitel 2). Hierauf aufbauend werden Ansätze zur Integration der Konzepte vorgeschlagen, die zur Entwicklung einer neuen Generation von Informationsportalen beitragen (Kapitel 3). Als Basiskonzepte werden Data-Warehouse- und OLAP-Systeme betrachtet, da deren Stärke im Umgang mit quantitativen Daten liegt. Um der Anforderung gerecht zu werden, auch qualitative Daten zur Verfügung zu stellen, wird die Integration mit Content-Management-Systemen untersucht. Als Grundlage für die Personalisierung der Informationsbereitstellung werden Konzepte des Information Retrieval und der Forschung im Bereich der Künstlichen Intelligenz berücksichtigt. Das Konzept der Informationsportale liefert das Leitmotiv, den Managern einer integrierte Oberfläche für deren Informationsversorgung bereitzustellen. Die Arbeit endet mit einem Verweis auf verwandte Arbeiten (Kapitel 4) und einer Zusammenfassung der Ergebnisse (Kapitel 5).

2 Kritische Bestandsaufnahme zur Gestaltung von Informationsportalen

2.1 Data Warehousing und OLAP

Data Warehousing zielt auf die Bereitstellung einer integrierten Sicht auf unterschiedliche Datenquellen in einer auf die Steuerungs- und Regelungsaufgaben der Anwender abgestimmten Form (Dualismus von Datenintegration und Analysefunktion).²⁰ Für die Analyse der Data-Warehouse-Daten werden im Kontext von Führungsinformationssystemen *OLAP-Systeme* eingesetzt. PENDSE und CREETH fassen die wesentlichen Charakteristika von OLAP unter dem Akronym FASMI (Fast Analysis of Shared Multidimensional Information) zusammen.²¹ Die performante, multidimensionale Analyse relevanter Unternehmensdaten im Mehrbenutzerbetrieb hat den Zweck, Analysethemen aus verschiedenen Blickwinkeln zu betrachten bzw. flexibel zu ändern. Die mittels OLAP auswertbaren Daten sind zu diesem Zweck in Form von Fakten zu strukturieren. *Fakten* stellen Verknüpfungen von Bezugsobjekten (z. B. Filialen, Artikel, Kunden, Mitarbeiter, Monate) und Kennzahlen (z. B. Umsatz, Deckungsbeitrag, Kosten) dar. Sowohl über Bezugsobjekte als auch über Kennzahlen werden regelmäßig hierarchische Strukturen definiert.

OLAP-Systeme stellen Projektionen auf den aus der Verknüpfung von Bezugsobjekten und Kennzahlen resultierenden Faktenraum in Form von Tabellen zur Verfügung. OLAP-Berichte stellen Tabellendefinitionen dar, die als Ausgangspunkt für Manipulationen an der Tabellenstruktur im Zuge der Navigation durch den Faktenraum dienen. Werden die Daten durch ein relationales Datenbankmanagementsystem gespeichert, fungiert das OLAP-System als SQL-Generator, der die durch den Benutzer definierten Sichten in Datenbankabfragen umwandelt.²² Die Benutzeroberflächen typischer OLAP-Systeme stellen mit den Funktionen Rotation, Slicing und Ranging/Dicing komfortable Möglichkeiten zur Eingrenzung des relevanten Bezugsobjektraumes zur Verfügung. Roll-up und Drill-down ermöglichen das Navigieren durch die hierarchischen Strukturen der Bezugsobjekte, wodurch Aggregationen bzw. Disaggregationen veranlasst werden.²³ Während die Navigation durch den Bezugsobjektraum umfassend unterstützt wird, sind die rechentechnischen bzw. sachlogischen Kennzahlenbeziehungen für den Anwender in den meisten Systemen nur unzureichend nachvollziehbar. Die für den Tabellenaufbau gewählten Kennzahlen lassen sich in der Regel nicht entsprechend visualisierter Rechen- bzw. Ordnungssysteme per Mausklick zerlegen. Executive Information

²⁰ Vgl. Bauer, Günzel (2001).

²¹ Zitiert nach Jahnke, Groffmann, Kruppa (1996).

²² Vgl. o. V. (2000), S. 39f.

²³ Vgl. Holten (1999), S. 49ff.

Systems wie z. B. das Produkt Forest und Trees können hier Anregungen für eine Erweiterung der Funktionalität liefern.²⁴

Der Entwurf des Navigationsraumes eines OLAP-Systems ist eine Konstruktionsleistung von Modellierern.²⁵ Für die fachkonzeptionelle Spezifikation von Data-Warehouse-Systemen wird es allgemein befürwortet, eine explizite Trennung von Kennzahlen- und Bezugsobjektstrukturen vorzunehmen.²⁶ Dieses Konstruktionsprinzip kann auf die von RIEBEL entwickelten Grund- und Auswertungsrechnungen der relativen Einzelkosten- und Deckungsbeitragsrechnung zurückgeführt werden.²⁷ *Bezugsobjekte* sind „alle selbstständigen Maßnahmen, Vorgänge und Tatbestände, die eigenständiges Dispositionsobjekt oder Untersuchungsobjekt sein können“²⁸. Im Rahmen der Grundrechnungen wurden den Bezugsobjekten zunächst Einzelkosten zugerechnet. Später wurden auch Einzelerlöse berücksichtigt. Mit der Einbeziehung von Mengengrößen wurde das Zurechnungsprinzip zur Einzelwertphilosophie verallgemeinert. Die Wertgrößen stellen – dem üblichen Sprachgebrauch folgend – Kennzahlen dar. *Kennzahlen* erfassen quantitativ darstellbare Sachverhalte in konzentrierter Form.²⁹

Dem Konstruktionsprinzip der Einzelwertphilosophie folgend wird bei der fachkonzeptionellen Spezifikation von Data Warehouses der Raum der relevanten Untersuchungsobjekte in Form von Bezugsobjekten strukturiert. Kennzahlen definieren die interessierenden Aspekte, Merkmale und Eigenschaften der Untersuchungsobjekte. Kennzahlensysteme ordnen Kennzahlen nach rechentechnischen (Rechensystem) oder rein sachlogischen (Ordnungssystem) Zusammenhängen.³⁰ Bezugsobjekte lassen sich – in Abhängigkeit vom Modellierungszweck – zu Dimensionen zusammenfassen. Bezugsobjekte, die genau einer Dimension zugeordnet sind, werden *Dimensionsbezugsobjekte* genannt. Über Dimensionsbezugsobjekte werden regelmäßig hierarchische Ordnungen definiert, deren Hierarchiestufen explizit benannt werden können. Dimensionen lassen sich zudem in Dimensionsgruppen ordnen, wenn eine Art von Dimensionsbezugsobjekten nach verschiedenen Kriterien zu analysieren ist. Die Dimensionen spannen einen mehrdimensionalen Raum auf. Dimensionsbezugsobjekte fungieren als Punkte auf den Achsen. Der Raum selbst wird durch kombinierte Bezugsobjekte gebildet, die sich als Vektoren denken lassen, deren Koordinaten durch Dimensionsbezugsobjekte festgelegt werden. Der Inhalt des Data Warehouses wird durch Fakten spezifiziert, die Verknüpfungen von Bezugsobjekten (in der Regel kombinierte Bezugsobjekte) und Kennzahlen

²⁴ Vgl. Heneböle (1995), S. 135-149.

²⁵ Vgl. zum konstruktionsorientierten Modellbegriff Schütte (1998), S. 59ff.

²⁶ Vgl. z. B. Gabriel, Gluchowski (1997).

²⁷ Vgl. Holten (1999).

²⁸ Vgl. Riebel (1979), S. 869.

²⁹ Vgl. Reichmann (1997), S. 19.

³⁰ Vgl. Groffmann (1992), S. 76.

darstellen. (Eine weitere Verfeinerung der Spezifikation ist durch die Verwendung von Dimensionsausschnitten und deren Kombinationen zu erreichen.)

Anleihe bei Executive Information Systems haben OLAP-Systeme bereits insbesondere bei der Integration des *Exception Reportings* genommen, das eine „periodische Durchführung von Soll-Ist-Vergleichen durch das System und die Meldung signifikanter Abweichungen an den Benutzer“³¹ darstellt. Die Signifikanzbeurteilung wird über manuell festgelegte oder automatisiert ermittelte Toleranzgrenzen operationalisiert.

2.2 Content Management und Informationsportale

Content Management hat die Administration des Lebenszyklus von Inhalten zum Gegenstand. Als Inhalte kommen beliebige digitale Objekte wie z. B. Texte, Bilder, Sounds, Videos und URLs in Frage. Der *Content Life Cycle* gliedert die Informationsverarbeitung in idealtypische Phasen.³² Inhalte werden zunächst durch Autoren erstellt. Anschließend sind die Inhalte solange nachzubearbeiten, bis sie die inhaltliche und gestalterische Kontrolle durch besonders autorisierte Mitarbeiter bestehen und freigegeben werden. Die freigegebenen Inhalte werden publiziert und damit extern verfügbar. Nach erfolgter Publikation sind ggf. Anpassungen z. B. zur Gewährleistung der Aktualität vorzunehmen, die eine nochmalige Kontrolle und Freigabe notwendig machen können. Die Inhalte sind aus Sicherheitsgründen und zum Nachvollzug der Änderungshistorie regelmäßig zu archivieren. Der Content Life Cycle endet mit der Eliminierung der Inhalte.

Content-Management-Systeme dienen der technischen Unterstützung der mit dem Content Life Cycle verbundenen Aufgaben. Sie werden auch unter dem Begriff *Redaktionssysteme* diskutiert. Content-Management-Systeme beruhen auf dem Prinzip der Trennung von Inhalt und Repräsentation. Unter Verwendung unterschiedlicher Layoutschablonen lassen sich so gleiche Inhalte in verschiedenen Ausgabemedien publizieren (z. B. Papiausdruck, CD-ROM, PDA, Handy etc.).³³ Zu den Inhalten können folgende Attribute verwaltet werden: Versionsnummer, Berechtigungen, Veröffentlichungs- und Verfallsdaten sowie Metaattribute, die eine grobe Beschreibung des Inhalts darstellen. Über die Metaattribute werden den Inhal-

³¹ Vetschera (1995), S. 37.

³² Vgl. Büchner, Zschau, Traub, Zahradka (2001), S. 83ff.

³³ Vgl. Büchner, Zschau, Traub, Zahradka (2001), S. 108.

ten Thesaurusklassen³⁴ (Kategorien, Schlagworte) zugeordnet, die bei der Suche nach bestimmten Inhalten verwendet werden können.³⁵

Content-Management-Systeme sind insbesondere für den Betrieb von WWW-basierten Enterprise Information Portals notwendig, was sich aus dem Umfang und der Änderungshäufigkeit ihrer redaktionellen Inhalte begründet.

Die heutigen Realisierungen von Enterprise Information Portals erinnern in ihrem Aufbau an Setzkästen. Die Portalseite ist schematisch in einzelne Sektionen eingeteilt. Die Sektionen beinhalten entweder Anwendungen oder Publikationen. Das Grundmuster wird personalisiert, indem der Benutzer durch die Wahl entsprechender Optionen die Anwendungen wie Wetterbericht, Börsennotierungen, Kalender etc. ein- bzw. ausblenden kann. Die Auswahl der Publikationen kann über die manuelle Angabe von Interessenprofilen bzw. durch die automatisierte Auswertung des Nutzerverhaltens gesteuert werden. Grundlage für die Personalisierung der angezeigten Publikationen ist die Kategorisierung der Inhalte durch Metaattribute, wie sie von Content-Management-Systemen unterstützt wird. Die Aufnahme von Suchmaschinen aus Content-Management-Systemen (z. B. Volltextsuche oder Suche unter Verwendung thesaurusbasierter Ersetzungsregeln) in die Portalseite ermöglicht die eigenständige Navigation des Benutzers durch die verfügbaren Inhalte.

2.3 Information Retrieval in Texten

Eine wichtige Herausforderung bei der Informationsbeschaffung aus externen, elektronischen Quellen wie dem Internet besteht darin, in den verfügbaren Inhalten die gewünschten oder benötigten Informationen zu finden. Dabei können Content-Management-Systeme nur eingeschränkt behilflich sein, da ihre Hauptaufgabe die Unterstützung des Lebenszyklus von Inhalten ist, nicht aber die Katalogisierung von externen Informationen. Schon vor dem Aufkommen von Content-Management-Systemen haben sich daher im Internet Hilfen wie Suchmaschinen und Verzeichnisdienste etabliert, die den Inhalt des Internets klassifizieren und für Suchanfragen zugänglich machen.

Die folgende ausführliche Bestandsaufnahme zu den Methoden des Information Retrieval und der Klassifikation textueller Inhalte bildet die Grundlage für die sich anschließende Suche nach Strukturanalogien zwischen den Bereichen Data Warehousing/OLAP und Content Management.

³⁴ Vgl. Kapitel 3.1.

³⁵ Vgl. Büchner, Zschau, Traub, Zahradka (2001).

2.3.1 Automatische Schlagwortidentifikation

Eine reine Volltextsuche in Dokumenten ist in der Regel sehr ineffizient, da hierbei alle Texte ausgewiesen werden, in denen der gesuchte Begriff verwendet wird. Außerdem werden die unterschiedlichen Flexionsformen eines Begriffes (z. B. „Häuser“ statt „Haus“) als unterschiedlich angesehen und somit als nicht gleichbedeutend erkannt.

Computerlinguistische Verfahren verwenden statistische Methoden und versuchen, Wörter in Texten nicht als Zeichenketten sondern als Flexionsformen eines Wortes aufzufassen. Diese werden dabei auf ihre Grundform (Nominativ Singular oder Infinitiv) oder ihre Stammform (beugungsresistenter Wortteil) zurückgeführt.³⁶

Bei einer Volltextsuche werden alle Dokumente, die eine Menge von Termen enthalten, als Ergebnis einer Anfrage zurückgegeben. Um die geringe Aussagekraft eines booleschen Retrieval zu verbessern, werden Verfahren eingesetzt, durch die Terme eines Dokuments unterschiedlich stark bei der Suche berücksichtigt werden. Begriffe, die den Inhalt eines Textes besonders gut beschreiben, werden *Schlagworte* genannt. Eine manuelle Bestimmung und Gewichtung dieser Schlagworte ist aufgrund des hohen Arbeits-, Zeit- und Kostenaufwands vielfach nicht sinnvoll. Automatische Gewichtungsverfahren verwenden numerische Maße, anhand derer die Bedeutung von Termen in einem Dokument oder einer Dokumentensammlung formal bestimmt werden kann. Möglich ist dies, da die Häufigkeit von Wörtern in Texten nicht gleich verteilt ist. In einer Textsammlung kann die Häufigkeitsverteilung der Wörter nach dem ZIPF'schen Gesetz beschrieben werden:

Für einen repräsentativen Textkorpus C bezeichne $W(C)$ die Menge der Wörter, die in C vorkommen, und $h(w)$ die Häufigkeit, mit der das Wort $w \in W(C)$ in dem Korpus vorkommt. $r(w)$ bezeichne den Rangplatz von w , wenn die Wörter nach fallender Häufigkeit sortiert werden. Dann gilt:

$$r(w) * h(w) \approx c = konst \text{ für alle } w \in W(C)$$

Häufig in Texten vorkommende Wörter, also die Wörter mit kleinen Rangzahlen, sind annähernd gleich über alle Texte verteilt. Eine Häufigkeitsbetrachtung dieser Begriffe sagt daher wenig über den Inhalt eines einzelnen Textes aus. Aus diesem Grund werden sie in sog. *Stoppwort-Listen* zusammengefasst und von der weiteren Analyse ausgeschlossen. Beispiele für mögliche Stoppworte in der deutschen Sprachen sind: „auf“, „der“, „in“, „ein“, etc.

³⁶ Für die Begriffe Grundform und Stammform wird im Folgenden der Begriff „Term“ generalisierend verwendet.

Je seltener Begriffe in Dokumenten vorkommen, desto weniger gilt die Gleichverteilungsannahme.³⁷ Die Häufigkeit, in der Begriffe mit höheren Rangzahlen in einem Dokument vorkommen, variiert und kann somit als ein Indikator dienen, in welchem Ausmaß der einzelne Begriff repräsentativ für den Inhalt des Gesamtdokumentes ist. Aus dieser Erkenntnis heraus wurden unterschiedliche statistische Größen entwickelt, mit denen Schlagworte bestimmt werden können. Beispielhaft wird hier ein Maß von SALTON und MCGILL vorgestellt: der *Diskriminanzwert*. Er bestimmt, wie stark ein Term ein Dokument von den anderen Dokumenten einer Dokumentensammlung unterscheidet. Basierend auf einer Ähnlichkeitsfunktion wird berechnet, wie sich die durchschnittliche Ähnlichkeit der Dokumente zueinander verändern würde, wenn der untersuchte Term nicht in diesen vorkommen würde. Nur Begriffe, durch deren Vorhandensein die Dokumente sich mehr unterscheiden, werden zur Klassifikation verwendet. Die so gebildete Schlagwortmenge wird auch als *Klassifikationswörterbuch T* bezeichnet.

2.3.2 Information Retrieval mit Schlagworten

Basierend auf den vorbereitenden Analyseverfahren wurden Methoden entwickelt, mit denen Texte anhand von Schlagworten gesucht werden können. Gegenüber einer booleschen Volltextsuche bieten diese Verfahren den Vorteil, dass sie sich nach den gewichteten Schlagworten richten. Sie versuchen die Texte zu bestimmen, deren Inhalt sich am deutlichsten auf die gesuchten Schlagworte bezieht. Häufig wird hierbei das Vektorraummodell eingesetzt.

Sei $D = \{d_1, \dots, d_m\}$ die Menge der Dokumente und $A = \{A_1, \dots, A_n\}$ eine Menge von Attributen zu diesen. Die Attributwerte $A_j(d_i) = w_{i,j}$ des Dokuments d_i (beispielsweise der Diskriminanzwert) lassen sich als Gewichte auffassen. Diese werden dann zu einem Vektor $w_i = (w_{i,1}, \dots, w_{i,n}) \in \mathbb{R}^n$, dem sog. *Dokumentenvektor*, zusammengefasst.

Anfragen werden ebenfalls durch einen Vektor dargestellt: $q \in \mathbb{R}^n$. Zur Bestimmung der Ergebnismenge einer Suchanfrage wird eine Ähnlichkeitsfunktion $s: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ definiert, mit der jedem Paar aus zwei Vektoren $x, y \in \mathbb{R}^n$ ein reeller Ähnlichkeitswert $s(x,y)$ zugewiesen wird. Ein einfaches Beispiel hierfür ist das Skalarprodukt.³⁸ Die Dokumente, deren Ähnlichkeit zur Anfrage einen Schwellenwert überschreiten, bilden die Ergebnismenge.

Verbesserungen der Ergebnisqualität lassen sich dadurch erzielen, dass im Vektor der Anfrage nicht nur die Werte „Eins“ (das Wort kommt in der Anfrage vor) und „Null“ (das Wort kommt in der Anfrage nicht vor) verwendet werden. Verfahren, wie sie zur Gewichtung von

³⁷ Vgl. Salton, McGill (1987), S. 65f.

³⁸ Weitere mögliche Ähnlichkeitsmaße finden sich z. B. in Salton, McGill (1987).

Termen in Texten eingesetzt werden, können jedoch bei Anfragen nicht sinnvoll verwendet werden, weil die Anfragen ausschließlich die relevanten Terme enthalten. (Alle Terme wären mit „Eins“ zu gewichten.) Neben der manuellen Gewichtung der gesuchten Terme kann in interaktiven Systemen das Verfahren des *Relevance Feedback* eingesetzt werden. Die Anfrage wird zuerst in einen booleschen Vektor umgewandelt. Der Suchende kann anschließend die Texte der Ergebnismenge nach ihrer Relevanz beurteilen. Die Dokumentenvektoren der dabei als relevant markierten Texte werden zur Anpassung des Vektors der Anfrage verwendet. Iterativ können dann mit dem modifizierten Vektor erneute Anfragen durchgeführt werden.

2.3.3 Strukturierung von Schlagworten

Schlagworte dienen der Erfassung der Inhalte von Texten und können wie beschrieben verwendet werden, um Texte zu suchen. Thesauren fassen diese Schlagworte zu Bedeutungsklassen (Thesaurusklassen) zusammen und setzen die Klassen zueinander in Beziehung. Zu den in Thesauren abbildbaren Beziehungen zählen:

- Synonyme: Unterschiedliche Begriffe besitzen die gleiche Bedeutung.
- Antonyme: Unterschiedliche Begriffe stellen Gegensätze dar („heiß“ und „kalt“).
- Verwandte Terme: Unterschiedliche Begriffe haben eine ähnliche aber nicht gleiche Bedeutung.
- Generalisierung/Spezialisierung: Ein Begriff stellt eine Spezialisierung eines anderen Begriffs dar. (Ein „Mikrocomputer“ ist ein „Computer“ mit bestimmten Eigenschaften.)
- Polysemie: Zeichenketten, die mehrere Bedeutungen haben können, werden in einem Thesaurus dadurch abgebildet, dass der entsprechende Term in unterschiedlichen Thesaurusklassen gleichzeitig enthalten sein kann. Welche Bedeutung eines Terms in einem Text oder einer Anfrage gemeint ist, kann mit einem Thesaurus jedoch nicht ermittelt werden.

Neben der manuellen Erstellung eines Thesaurus existieren automatische Verfahren. Basierend auf den Dokumentenvektoren werden Verfahren der statistischen Datenanalyse eingesetzt, welche die Schlagworte des Vokabulars zu Synonymen gruppieren und so das Grundgerüst des Thesaurus erstellen. Eine darüber hinausgehende Klassifizierung (Generalisierung/Spezialisierung, Antonyme) hat manuell zu erfolgen.

Der Einsatz eines Thesaurus kann bei der Suche nach relevanten Inhalten dazu beitragen, die Genauigkeit zu verbessern, indem sowohl Synonyme also auch Spezialisierungen der Terme einer Anfrage mit berücksichtigt werden.

2.4 Verbindung von OLAP- und Content-Management-Systemen

Auf den Markt der Informationsportal-Server drängen sowohl Hersteller von Content-Management- als auch von OLAP-Systemen. Die Einsetzbarkeit von Thin-Client-Architekturen, die Realisierbarkeit von Mobile-Access-Konzepten und die Erleichterung der Administration von Arbeitsplatzumgebungen haben ursprünglich motiviert, den Zugriff auf OLAP-Server über Internet-Browser zu ermöglichen. Die Möglichkeit, einen OLAP-Bericht über eine URL-Adresse aufzurufen, stellt heute eine Standardfunktionalität vieler OLAP-Systeme dar.³⁹ Auch in der Web-Umgebung sind die OLAP-Funktionen zur Manipulation des Tabellenaufbaus verfügbar.

Der integrierte Zugriff auf quantitative und qualitative Inhalte kann hergestellt werden, indem die URL-Adressen von OLAP-Berichten als Inhalte vom Content-Management-System verwaltet werden. Die URL-Adresse bzw. der OLAP-Bericht wird über Metaattribute kategorisiert.⁴⁰ Suchfunktionen des Content-Management-Systems bzw. des Informationsportals liefern dann sowohl OLAP-Berichte als auch Texte, Bilder, Sounds, Videos etc. zurück. Das Prinzip dieser Verbindung wird in Abb. 2(a) anhand eines Datenmodells in ERM-Notation⁴¹ und in Abb. 2(b) mittels einer Softwaresystemarchitektur veranschaulicht.

2.5 Entwicklungsbedarfe

Nach dem State-of-the-Art wird die Inhalteintegration ausschließlich über eine Zuordnung von Themenwörtern (Termen) zu OLAP-Berichten realisiert. Diese Vorgehensweise birgt die Gefahr von Inkonsistenz in sich, da OLAP-Berichte im Zuge der Tätigkeiten des Data Warehousing verändert werden können. Die gegebenenfalls notwendige Anpassung der Kategorisierung ist nicht sichergestellt. Die *redundante Zuordnung von Themenwörtern* zu OLAP-Berichten macht zudem eine Vervielfachung des Synonymen-/Homonymen-Problems wahrscheinlich, das betrieblichen Begriffssystemen inhärent anlastet.

³⁹ Vgl. Kurz (1999), S. 377-452.

⁴⁰ Vgl. Rieger, Kleber, van Maur (2000).

⁴¹ Vgl. Holten, Ehlers (1998), S. 126.

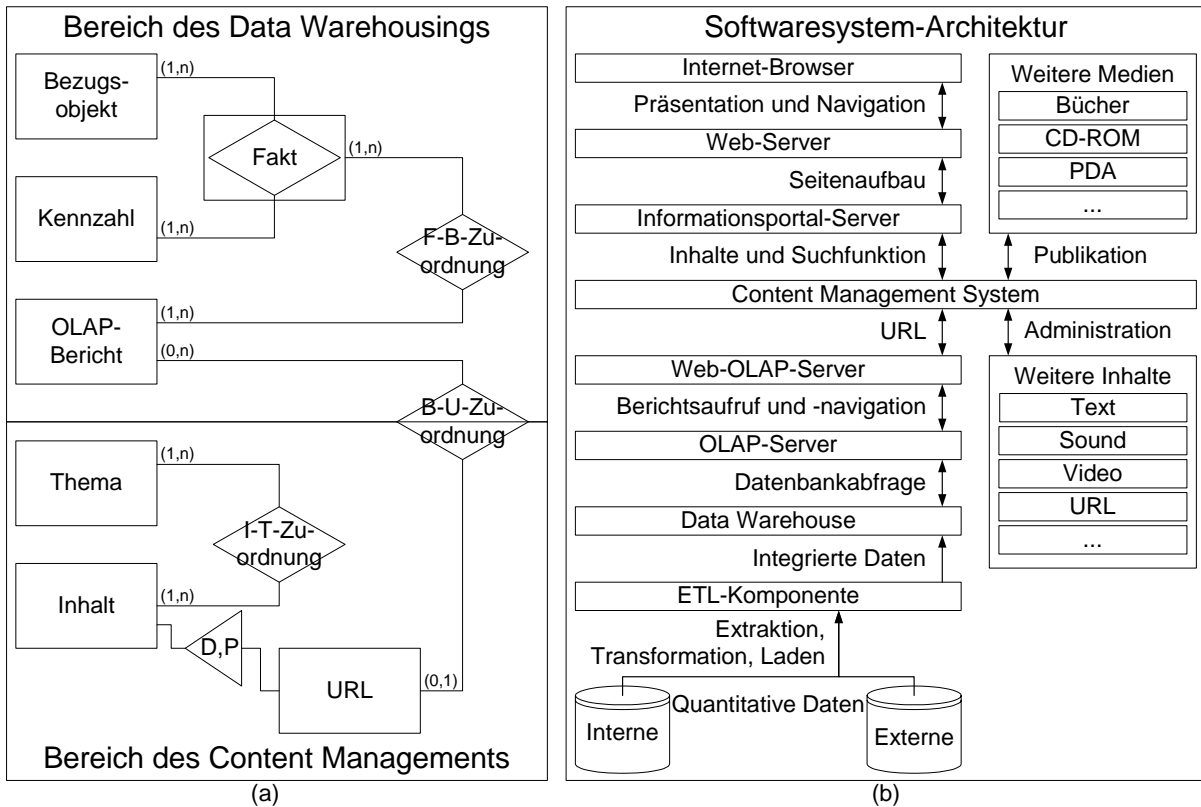


Abb. 2: Aktuelles Prinzip zur Verbindung von Data Warehousing und Content Management

Darüber hinaus macht der Anspruch eines integrierten Informationsportals die einheitliche Inhaltebereitstellung für den Anwender erforderlich. Insbesondere soll dabei keine Unterscheidung zwischen quantitativen oder qualitativen Inhalten gemacht werden. Um diesem Anspruch gerecht werden zu können, reichen die von den vorgestellten Teilsystemen isoliert bereitgestellten Möglichkeiten nicht aus.

Unterstützung bei der Definition von OLAP-Berichten

Während der Anwender in einem Data Warehouse bei der Navigation in einem Bericht durch unterschiedliche Navigationsschritte den Betrachtungsbereich verändern kann sowie bei Einschränkungen und Ausweitungen der analysierten Datenbasis gut unterstützt wird, ist die Suche nach einem Bericht vergleichsweise wenig unterstützt. Berichte können in der Regel zwar gruppiert und in Hierarchien verwaltet werden; aber auf der Suche nach einem Bericht mit bestimmten Eigenschaften – z. B. betrachtete Kennzahlen und Bezugsobjekte – muss diese Kategorisierung nicht immer hilfreich sein. Eine nicht nach den Vorstellungen des Anwenders erstellte Gliederung behindert unter Umständen die Suche. In der Regel ist dann der Aufwand, einen entsprechenden Bericht neu zu definieren, geringer, als unter allen vorhandenen Berichten einen passenden zu suchen. Dies kann eine unnötige Mehrfachdefinition von Berichten zur Folge haben.

Eine Verbesserung der Suchgenauigkeit lässt sich durch neue Navigationssysteme erzielen. Neben der manuellen Strukturierung nach unternehmensspezifischen Kriterien sollte die Navigation für Berichte automatisch nach den Eigenschaften der Berichte strukturiert werden. So können sowohl die Bezugsobjekthierarchien als auch das Kennzahlensystem als Navigationsrahmen dienen. Hierdurch kann sichergestellt werden, dass ein bereits mit den gewünschten Eigenschaften definierter Bericht gefunden wird.

Sollte kein entsprechender Bericht definiert sein, können die im Rahmen der Suche angegebenen Eigenschaften die Definition eines neuen Berichts erleichtern. Sie können als Vorgabe bzw. vorläufige Eigenschaften des neuen Berichts aufgenommen werden, aus denen ein neuer, vorläufiger Bericht erstellt wird. Für eine vollständige Definition des Berichts ist dieser ggf. anschließend noch zu verfeinern bzw. zu erweitern.

Benutzer-individuelle Informationsbereitstellung

Content-Management-Systeme arbeiten nach dem Prinzip der Trennung von Inhalten und deren Präsentation. Zum einen kann durch die Gestaltung von Vorlagen die Darstellung einzelner Inhalte beeinflusst werden. Zum anderen besteht in der Regel die Option, die verfügbaren Inhalte nach individuellen Kriterien zu selektieren. Beide Aspekte tragen dazu bei, eine Benutzer-individuelle Darstellung der Inhalte zu ermöglichen.

Demgegenüber arbeiten Data-Warehouse-Systeme in der Regel mit Berechtigungskonzepten, um die Informationen, die sich ein Anwender anzeigen lassen kann, einzuschränken. Da die Berechtigungen jedoch nicht von den Anwendern selbst bestimmt werden, kann hier nicht von der geforderten Individualisierung die Rede sein. Konzepte, mit denen Anwender selbst bestimmen können, welche Inhalte für sie interessant sind (z. B. welche Berichte, Bezugsobjekte, Kennzahlen, etc.), sind nur in beschränktem Umfang vorgesehen.

Als eine Möglichkeit der Individualisierung lassen sich in Data-Warehouse-Systemen Berichte definieren, die beschreiben, unter welchen Umständen eine oder mehrere Fakten eine Ausnahmesituation darstellen, die einer genaueren Untersuchung bedarf. Durch periodische Auswertung wird sichergestellt, dass die Benutzer über entsprechende Situationen informiert werden.

Derartige Push-Mechanismen lassen sich für qualitative Inhalte nur eingeschränkt umsetzen. So ist das Berechnen einer eindimensionalen Kennzahl, anhand derer das Überschreiten eines Schwellenwertes überprüft werden kann, nur schwer realisierbar. Vielfach wird stattdessen der Weg beschritten, dass die Anwender definieren, welche der vom System bereitgestellten Kategorien für sie eine besonders hohe Bedeutung haben. Entsprechend dieser Bewertung werden die Inhalte selektiert oder neu hinzugefügte Inhalte an den Anwender geschickt.

Klassifikationsschemata eignen sich jedoch nur bedingt für eine individualisierte Informationsversorgung. Schwierigkeiten können während der Gestaltung des Systems bei der Bestimmung der richtigen Anzahl an Kategorien sowie während des Betriebs bei der konsistenten Zuordnung von Inhalten zu diesen auftreten. Für die bedarfsgerechte Informationsversorgung auch in Reaktionssituationen besteht der Bedarf, Push-Mechanismen aus den Bereichen OLAP und Data Warehousing (Exception Reporting) in geeigneter Weise auf qualitative, externe Daten zu übertragen.

Integrierte Betrachtung von quantitativen und qualitativen Informationen

Das Ziel der Zusammenführung quantitativer und qualitativer Informationen besteht aus zwei Teilen. Zum einen muss eine Datenstruktur gestaltet werden, die alle Informationen in gleicher Weise verwalten kann und somit das Finden erleichtert. Die zu schaffende Navigationsstruktur muss die Metadaten beider Informationstypen so zusammenführen, dass keine Unterscheidung bzgl. der Herkunft vorgenommen werden muss. Dabei sind Analogien in den Strukturen der Metadaten zu finden und auszunutzen. Hierauf aufbauend ist als zweites Teilziel anzustreben, dass die Ausprägungen der Metadaten ebenfalls in einer Beziehung zu einander stehen. Die verwalteten Inhalte können so semantisch verknüpft werden. Eine einheitliche Navigation und Suche wird hierdurch ermöglicht. Eine Unterscheidung der Informationsart ist dabei nicht mehr erkennbar.

Verändern sich die fachlichen Hierarchien über Bezugsobjekte und Kennzahlen oder verändert sich der Thesaurus, so muss sichergestellt werden, dass diese Änderungen sowohl im Data-Warehouse- als auch im Content-Management-System berücksichtigt werden. Beispielsweise bewirkt das Anlegen eines neuen Kunden im OLTP-System, dass aus der Sicht des OLAP-Systems ein neues Bezugsobjekt eingeführt wird. Aus Sicht des Content-Management-Systems kann die Anlage eines neuen Kunden die Einführung einer neuen Ausprägung eines Metaattributs zur Folge haben. Während sich die Ergänzung einzelner Kunden als Bezugsobjekte aus dem operativen Geschäft ergibt, begründet sich die Veränderung von Bezugsobjekthierarchien regelmäßig aus neuen Analyseerfordernissen. Um die Vergleichbarkeit der Analysen über die Zeit zu gewährleisten, wird in der Regel angestrebt, die Bezugsobjekt- und Kennzahlenhierarchien unverändert zu lassen, indem zum Beispiel weitere (parallele) Hierarchien angelegt werden. Die von der Integration zu berücksichtigenden Veränderungen unterscheiden sich daher hinsichtlich ihrer Volatilität (zeitliche Länge der Veränderungsintervalle), wodurch die Aufgabe der Konsistenzerhaltung erschwert wird.

In einem Data Warehousing und Content Management integrierenden System werden Bezugsobjekte und Kennzahlen zur Klassifikation von Inhalten herangezogen. Sie werden dabei als Thesaurusklassen verwendet. Über die Verfahren des Information Retrieval können The-

saurusklassen Inhalten automatisiert zugeordnet werden. Die Veränderung von Bezugsobjekten und Kennzahlen sowie deren Beziehungen hat zur Folge, dass die automatisierten Verfahren zu abweichenden Ergebnissen gelangen. Im Rahmen der Integration ist daher festzulegen, ob Modifikationen in den Bezugsobjekt- und Kennzahlenmengen mit einer erneuten Klassifikation von Inhalten verbunden werden soll.

3 Ansätze zur Weiterentwicklung von Informationsportalen

Zielsetzung von Informationsportalen ist es, sowohl quantitative Daten, die heute hauptsächlich durch OLAP-Systeme verwaltet werden, als auch qualitative Inhalte, deren Verwaltung in den Aufgabenbereich von Content-Management-Systemen fällt, über eine integrierte Oberfläche verfügbar zu machen. In Analogie zum für Data-Warehouse-Systeme gültigen Dualismus muss eine entsprechende informationstechnische Infrastruktur einerseits für eine *Inhalteintegration* sorgen und andererseits geeignete Funktionen zur *Bereitstellung* der Inhalte zur Verfügung stellen.

Grundlage für die in dieser Arbeit entwickelte integrierte Inhaltebereitstellung ist eine integrierende Metadaten-Struktur (Kapitel 3.1). Hierauf aufbauend wird ein Vorschlag für zu unterstützende Navigationstechniken gemacht (Kapitel 3.2). Anschließend werden die Möglichkeiten zur Individualisierung der Informationsbereitstellung sowie zur Umsetzung von Push-Mechanismen in einem integrierten Informationsportal erörtert (Kapitel 3.3). Das Kapitel wird durch einige Vorschläge zur DV-konzeptionellen Umsetzung der Konzepte ergänzt (Kapitel 3.4).

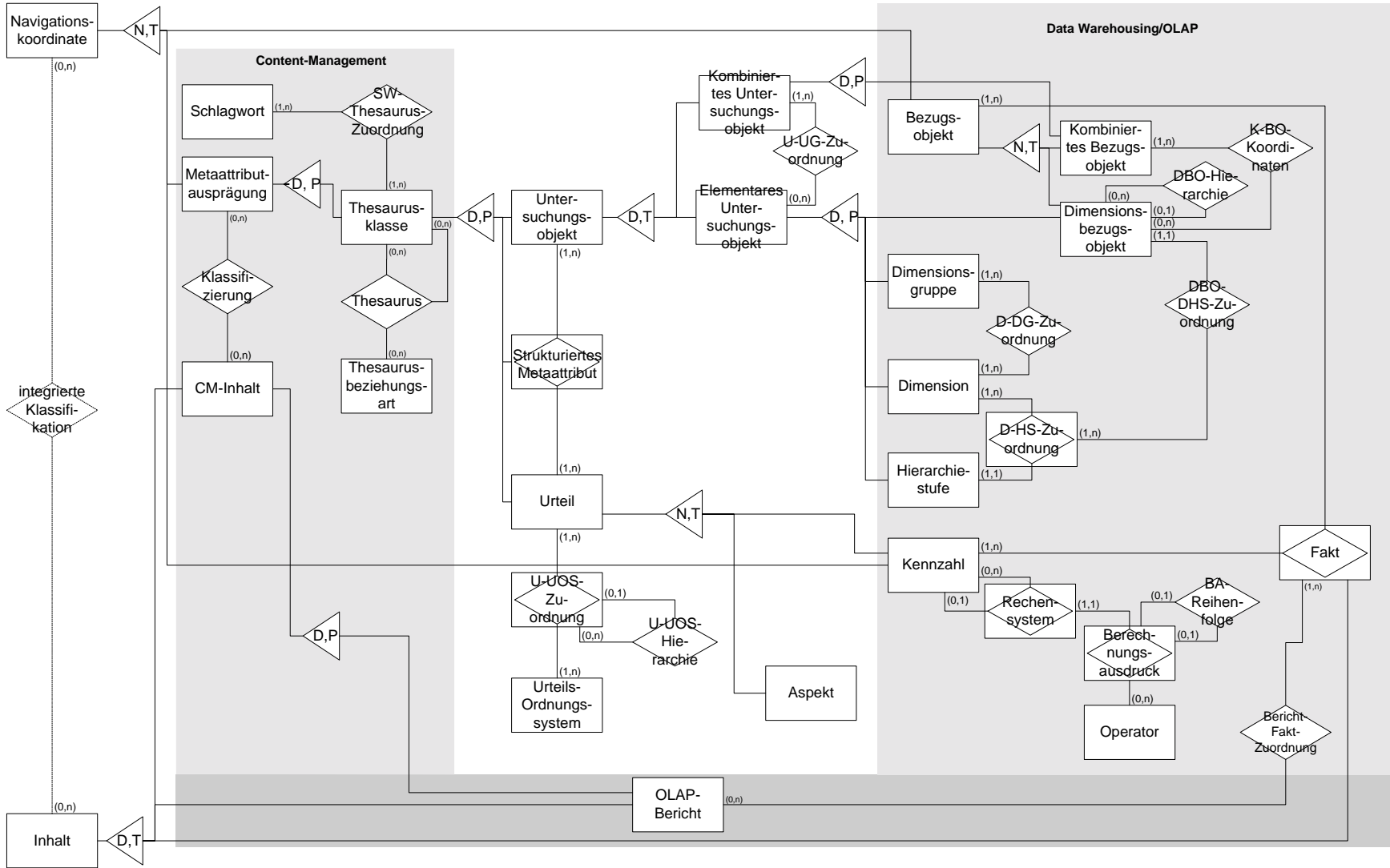
3.1 Metadaten-basierte Inhalteintegration

Für die Integration ist das Konstruktionsprinzip der Spezifikation quantitativer Daten-Navigationsräume auf die Konstruktion von Inhalträumen zu erweitern. Die Struktur des Inhaltraums kann analog zu Metamodellen von Führungsinformationssystemen⁴² konstruiert und mit einem Metamodell beschrieben werden. Ausgehend von den Grundbegriffen aus den Metamodellen des Data Warehousing und des Content Management werden im Folgenden die Begriffe zur Spezifikation eines Enterprise Information Portal konstruiert (vgl. Abb. 3).

Das Informationsportal verwaltet *Inhalte*. Dem Integrationsanspruch des Konzepts des Enterprise Information Portal wird entsprochen, indem *Fakten*, *OLAP-Berichte* und sonstige *CM-Inhalte* zu Inhalten generalisiert werden.

⁴² Vgl. Becker, Holten (1998), S. 486ff.

Abb. 3: Integrierte Metadatenstruktur



Fakten stellen Kombinationen aus *Bezugsobjekten* und *Kennzahlen* dar (vgl. nochmals Kapitel 2.1), für die sich jeweils detaillierte Strukturierungen ergeben: *Dimensionen* fassen Bezugsobjekte zusammen, die untereinander einen starken Bezug aufweisen und in der Regel nach bestimmten Kriterien in *Hierarchien* angeordnet werden (z. B. „Filialen nach Geografie“, „Filialen nach Alter“). Dimensionen, die sich auf gleiche Bezugsobjekte beziehen, bilden *Dimensionsgruppen* (z. B. „Filialen“, „Kunden“, „Artikel“). Über der hierarchischen Gliederung der Bezugsobjekte entstehen *Hierarchiestufen* (z. B. „Warengruppen“, „Warenobergruppen“). Die Bezugsobjekte sind entweder als *Dimensionsbezugsobjekt* („Kunde XY“, „Filialen in NRW“) direkt einer Kombination aus Dimension und Hierarchiestufe zugewiesen oder setzen sich als *kombinierte Bezugsobjekte* (z. B. „Filiale Münster im 4. Quartal 2001“) aus mehreren Dimensionsbezugsobjekten zusammen.

Wie sich *Kennzahlen* durch rechentechnische Operationen zu neuen Kennzahlen kombinieren lassen, wird über *Rechensysteme* abgebildet. Die Reihenfolge, in der dabei die *Operatoren* angewendet werden, wird durch den *Berechnungsausdruck* und die *BA-Reihenfolge* ausgedrückt. So können die Kennzahlen „Umsatz“ und „Eigenkapital“ jeweils der „Eigenkapitalrentabilität“ zugeordnet werden. Über den Operator wird festgelegt, dass die Rentabilität durch Division ermittelt wird. Über die Angabe der Reihenfolge erfolgt die Zuordnung der Rollen „Divisor“ bzw. „Dividend“.

In Content-Management-Systemen können zu CM-Inhalten Metaattribute gepflegt werden. *Metaattributausprägungen* klassifizieren CM-Inhalte. Zu diesen zählen z. B. Versionsnummer, Erstellungsdatum, Autorennamen oder Angaben zum Inhalt. Für Angaben zum Inhalt werden *Thesaurusklassen* verwendet. Die Generalisierung/Spezialisierung zwischen der Metaattributausprägung und der Thesaurusklasse ist somit partiell.

Den Thesaurusklassen werden *Schlagworte* über die Beziehung *SW-Thesaurus-Zuordnung* zugewiesen. Die Kardinalitäten (1,n):(1,n) stellen sicher, dass Homonyme und Synonyme abgebildet werden können. Die Beziehung *Thesaurus* ermöglicht die Abbildung von unterschiedlichen Relationen, in denen Thesaurusklassen zueinander stehen können. Die Art der Relation wird durch die *Thesaurusbeziehungsart* spezifiziert.

Die Voraussetzung für eine einheitliche Behandlung von Metadaten wird im Folgenden durch einen Analogieschluss geschaffen, über den die Metadaten des Data-Warehouse-Systems mit den Metadaten des Content-Management-Systems in Beziehung gesetzt werden. Für die Inhalteintegration wird das Konstruktionsprinzip der Spezifikation quantitativer Daten-Navigationsräume auf die Konstruktion von Inhalträumen erweitert. Es ergeben sich folgende Besonderheiten:

Fakten beziehen sich ausschließlich auf als Bezugsobjekte spezifizierte Untersuchungsobjekte. Für die Beschreibung von Inhalten sollten zusätzliche Bezugsobjektmenge(n) zugelassen werden. Beispielsweise können sich Texte allgemein auf alle Warengruppen und damit auf eine spezielle Hierarchiestufe beziehen. Oder Inhalte können nur eine Bedeutung für die Betrachtung von Kunden allgemein und damit für eine Dimensionsgruppe haben. Für die Spezifikation der Untersuchungsobjekte qualitativer Inhalte sind daher neben Bezugsobjekten auch Hierarchiestufen, Dimensionen und Dimensionsgruppen zu kombinieren. Dimensionsbezugsobjekt, Dimensionsgruppe, Dimension und Hierarchiestufe werden daher zum Begriff *elementares Untersuchungsobjekt* generalisiert. (Eine Verfeinerung kann durch die zusätzliche Berücksichtigung von Dimensionsausschnitten und Dimensionsausschnittskombinationen erreicht werden.) Da die CM-Inhalte nicht auf die Daten aus OLTP-Quellen angewiesen sind, können CM-Inhalte zusätzliche Untersuchungsobjekte betrachten. Beispiele für Untersuchungsobjekte aus CM-Inhalten, die sich nicht in OLAP-Berichten finden, sind z. B. „Konkurrenz“, „Außenwirtschaft“ oder „Steuerpolitik“. Somit ist die Spezialisierung des Untersuchungsobjekts partiell.

Kombinierte Untersuchungsobjekte werden eingeführt, um der Tatsache Rechnung zu tragen, dass ein CM-Inhalt ein Thema behandeln kann, das sich aus elementaren Untersuchungsobjekten sowohl des Data-Warehouse- als auch des Content-Management-Systems zusammensetzt. Ein Beispiel hierfür ist die „nationale Steuerpolitik“. „National“ kann dabei als Bezugsobjekt aus der Dimensionsgruppe „Geographie“, Dimension „Topologie“, Hierarchiestufe „Staaten“ betrachtet werden. „Steuerpolitik“ ist ein Untersuchungsobjekt, das in der Regel nicht im OLAP-System vorkommt. Ein kombiniertes Untersuchungsobjekt setzt sich aus einem oder mehreren Untersuchungsobjekten zusammen. Würde auf die Einführung des kombinierten Untersuchungsobjekts verzichtet, müssten die Untersuchungsobjekte dem Inhalt einzeln zugewiesen werden. Es wäre nicht mehr zu erkennen, welche der möglichen Kombinationen der Untersuchungsobjekte zulässig wären. Situationen, in denen ein CM-Inhalt mehrere kombinierte Untersuchungsobjekte behandelt, ließen sich nicht korrekt abbilden. *Untersuchungsobjekte* fassen die elementaren und die kombinierten Untersuchungsobjekte zusammen.

Die Einführung des Begriffs Untersuchungsobjekt ist nur dann sinnvoll, wenn neben den Untersuchungsobjekten auch spezifiziert wird, welche *Urteile* über diese gefällt werden. Aus Sicht des Data-Warehouse-Systems ist die Kennzahl als ein Urteil über ein Bezugsobjekt anzusehen. Fakten können ausschließlich mit Kennzahlen verbunden sein. Texte sind nicht an die Beschränkung der OLAP-Systeme gebunden, wonach die Sachverhalte quantitativ repräsentiert werden müssen. Urteile wie das Image (eines Produktes) oder die Motivation (von Mitarbeitern) erfüllen die Repräsentationserfordernisse der OLAP-Systeme im Allgemeinen nicht. Um sie dennoch im Portal aufnehmen zu können, wird der Begriff *Aspekt* eingeführt.

Die nicht disjunkte Spezialisierung des Urteils in Kennzahl und Aspekt trägt dem Umstand Rechnung, dass bestimmte Eigenschaften von Untersuchungsobjekten sowohl in Form quantitativer als auch qualitativer Inhalte festgehalten werden. Der Sachverhalt Umsatz kann zum Beispiel quantitativ exakt in Form von Fakten angegeben werden. Zudem können Textpassagen zu erwartende Entwicklungen des Umsatzes beschreiben. Der Text beinhaltet zwar ein Urteil über den Umsatz, dieser muss aber nicht zwangsläufig in einer Zahl fixiert sein.

Urteile (und somit auch jede Kennzahl) können einem oder mehreren *Urteils-Ordnungssystemen* zugeordnet werden. Die Urteils-Ordnungssysteme stellen eine Verallgemeinerung der aus der Theorie über Kennzahlensysteme bekannten Ordnungssysteme dar. Ordnungssysteme ordnen Kennzahlen nach sachlogischen Kriterien. Rechensysteme können dagegen ausschließlich über Kennzahlen gebildet werden, da sie auf rechentechnischen Verknüpfungen beruhen. Die hierarchischen sachlogischen Beziehungen zwischen den Urteilen innerhalb eines Ordnungssystems werden über die *U-UOS-Hierarchie* abgebildet. Über das Ordnungssystem lassen sich dem Urteil „Bonität“ zum Beispiel die Kennzahlen „Liquidität des ersten Grades“ und „Liquidität zweiten Grades“ und der Aspekt „Zahlungsmoral“ zuordnen.

Die Unterscheidung zwischen Untersuchungsobjekt und Urteil ist ein kreativer Akt des Modellierers. „Wirtschaftswachstum“ kann sowohl als Untersuchungsobjekt als auch als Urteil „Wirtschaftswachstum“ aufgefasst werden. Als dritte Möglichkeit kann der Begriff in das Untersuchungsobjekt „Wirtschaft“ und das Urteil „Wachstum“ geteilt werden.⁴³ Die Entscheidung zwischen den Alternativen ist nach Zweckmäßigkeit zu treffen. Durch die Unterscheidungsanweisung erhält der Modellierer eine methodische Anleitung für die Konstruktion der Navigationsräume.

In Analogie zur Verknüpfung von kombinierten Bezugsobjekten mit Kennzahlen zu Fakten im Data-Warehouse-Bereich wird die Kombination aus Untersuchungsobjekt und Urteil unter der Bezeichnung „*strukturiertes Metaattribut*“ eingeführt. Strukturierte Metaattribute erlauben es beispielsweise, für CM-Inhalte zu spezifizieren, dass sie „Filialumsätze“ oder „Monatsgewinne“ thematisieren.

Der Bezug von CM-Inhalten zu den zugehörigen Untersuchungsobjekten und Urteilen erfolgt über die *Thesaurusklassen*. Eine Thesaurusklasse spezialisiert sich in kombinierte Untersuchungsobjekte, Urteile und strukturierten Metainhalt.

⁴³ Siehe auch die sich anschließenden Ausführungen zu „strukturiertes Metaattribut“.

Da kombinierte Untersuchungsobjekte aus mehreren Worten bestehen können, muss zugelassen werden, dass Schlagworte ebenfalls aus mehreren Worten zusammengesetzt sein können, was über die Kardinalitäten der *SW-Thesaurus-Zuordnung* berücksichtigt wird.

Zusammenfassend stellt das vorgestellte Konzept zur Inhalteintegration *Navigationskoordinaten* zur Verfügung, die den Inhalteraum des Informationsportals strukturieren. Die Navigationskoordinaten ergeben sich aus einer nicht disjunkten, totalen Generalisierung über Bezugsobjekte, Kennzahlen und Metaattributausprägungen, womit die Such- bzw. Analyse-kriterien von Content-Management- und Data-Warehouse-Systemen gleichermaßen berücksichtigt sind. In Analogie zu der klassifizierenden Zuordnung von Metaattributausprägungen zu CM-Inhalten dienen die Navigationskoordinaten der Klassifikation der Inhalte. Diese Klassifikation wird hier *integrierte Klassifikation* genannt, da sie die Inhalte der verschiedenen Systeme berücksichtigt. Die Zuordnungen der integrierten Klassifikation lassen sich aus den bisher eingeführten Beziehungen ableiten. (Die Ableitbarkeit wird im Modell über gestrichelte Linien repräsentiert.)

- Jedem Inhalt des Enterprise Information Portal können beliebig viele Navigationskoordinaten zugewiesen werden.
- Auf der Ebene der Navigationskoordinaten ist die Herkunft der Koordinaten transparent.
- Die Instanzen der Navigationskoordinaten stehen unabhängig von ihrer Herkunft zueinander in Beziehung. Die Beziehungen bilden eine Struktur, die den Anwender bei der Suche nach Inhalten unterstützt.

Abschließend ist speziell auf *OLAP-Berichte* einzugehen, da sie Bezüge zu beiden Clustern des Integrationsmodells aufweisen. Einerseits kann ein OLAP-Bericht z. B. in Form einer URL durch Content-Management-Systeme verwaltet werden und ist somit als Spezialisierung eines CM-Inhaltes aufzufassen (vgl. Kapitel 2.4), der über Metaattributausprägungen klassifiziert wird. Andererseits stellt ein OLAP-Bericht Mengen von Fakten dar, die Bezugsobjekte mit Kennzahlen verknüpfen. Die Konsistenz der Klassifizierung sollte sichergestellt werden, indem man bei OLAP-Berichten keine manuelle Klassifikation über Thesaurusklassen zulässt und stattdessen die automatisierte Ableitung der relevanten Untersuchungsobjekte und Urteile unterstützt.

3.2 Metadaten-basierte Inhaltsnavigation

Die im vorherigen Abschnitt vorgestellte Metadatenintegration schafft die Voraussetzung für ein flexibles Zusammenspiel von Push- und Pull-Mechanismen zur Informationsbereitstel-

lung. Im Folgenden wird ein Vorschlag zur Gestaltung der Informationsbereitstellungsfunktionen in Informationsportalen unterbreitet (vgl. Abb. 4).

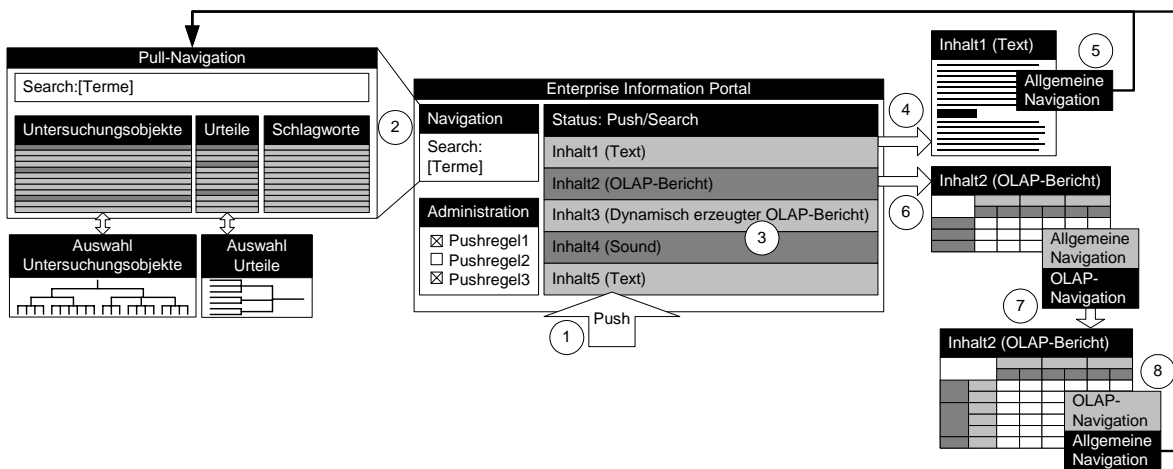


Abb. 4: Umsetzung des Pull- und Push-Prinzips im Enterprise Information Portal

1. *Einstieg in das Portal.* Der Aufbau der Portalseite sollte eine große Liste aufweisen, in welche die in der jeweiligen Benutzerphase relevanten Inhalte eingestellt werden. Beim initialen Einstieg in das Portal weist die Liste alle Inhalte auf, die über Push-Mechanismen zur Verfügung gestellt werden. Die Liste enthält Verweise auf quantitative und qualitative Inhalte. Die Bezeichnungen der Verweise können den Informationstyp ausweisen, der gegebenenfalls auch durch Icons visualisiert wird.⁴⁴ Um den Listenbereich herum sind zusätzliche Applikationen angeordnet. Zu den zusätzlichen Funktionen gehört insbesondere die Möglichkeit zur manuellen Administration von Benutzerprofilen, die der Steuerung des Push-Mechanismus dienen. Die Benutzerprofile können zudem zur relevanzorientierten Sortierung der Inhalte herangezogen werden (vgl. zur Ermittlung von Benutzerprofilen ausführlich Kapitel 3.3). Ein Durchgriff auf das Data-Warehouse-System sollte die Pflege der Exception-Reporting-Regeln des OLAP-Systems erlauben.
2. *Aufruf der Suche.* Als weitere ergänzende Funktion erhält der Anwender die Möglichkeit, Anfragen zu formulieren (Pull-Mechanismus). Eine Anfrage wird durch einen Termvektor repräsentiert, den der Anwender durch Auswahl von Untersuchungsobjekten, Urteilen und Schlagworten (per Mausklick) zusammenstellt. Die Definition der Anfrage wird erleichtert, indem Tree-View-Komponenten eine Navigation durch die über Untersuchungsobjekte und Urteile definierten Hierarchien unterstützen. Die Ergebnisse der Suche können in die Inhalteliste der Hauptseite eingestellt werden.

⁴⁴ Vgl. Rieger, Kleber, von Maur (2000), S. 375.

Eine Statuszeile dokumentiert, ob die Liste Ergebnisse der Suche oder der Push-Mechanismen ausweist.

3. *Dynamisches Erzeugen von OLAP-Berichten.* Für den Fall, dass zu einem Termvektor kein vordefinierter OLAP-Bericht gefunden wird, sollte die dynamische Definition einer geeigneten Einstiegstabelle die Möglichkeit zu einer Analyse quantitativen Materials eröffnen. Der mit der Ergebnisliste verknüpfte Termvektor enthält Informationen über die für den Anwender aktuell relevanten Untersuchungsobjekte (Bezugsobjekte, Dimensionsgruppen, Dimensionen und Hierarchiestufen) und Urteile (Kennzahlen). Dem Fehlen geeigneter Einschränkungen ist über die Vereinbarung von Default-Regeln zu begegnen: Sollte der Term beispielsweise keine Kennzahlen enthalten, kann für die Zusammenstellung der Ausgangstabelle standardmäßig z. B. der Umsatz gewählt werden. Ist dem Term nur die Relevanz einer Dimensionsgruppe zu entnehmen, kann für den Einstieg in das OLAP-System zunächst die Hierarchie mit den meisten Hierarchiestufen gewählt werden. Auf die mit den Default-Regeln verbundenen Ungenauigkeiten hinsichtlich der Informationsbedarfsdeckung wird der Benutzer über eine entsprechende Kennzeichnung des Listeneintrags hingewiesen. Zur Generierung des entsprechenden Berichts schreibt der Informationsportal-Server in die Metadatenbank des Data-Warehouse-Systems, dessen Schnittstelle in der Regel öffentlich ist. Für die dynamisch erzeugten OLAP-Berichte kann vorgesehen werden, dass sie wieder gelöscht werden, z. B. nachdem sich der Benutzer vom System abgemeldet hat. Alternativ bietet es sich an, die dynamisch erzeugten Berichte innerhalb einer festgelegten Latenzzeit als reguläre Berichte im System zu verwenden. Wird der Bericht innerhalb der Latenzzeit hinreichend häufig verwendet, kann er endgültig in den Bestand der regulären, vordefinierten Berichte übernommen werden. Wird der Bericht dagegen nicht häufig genug verwendet, wird er gelöscht.
4. *Auswahl qualitativer Inhalte.* Wählt der Anwender qualitative Informationen, werden diese mit der für den entsprechenden Informationstyp vorgesehenen Anwendung angezeigt bzw. wiedergegeben. Zusätzlich sollte er sich die zu dem Inhalt verfügbaren Metainformationen anzeigen lassen können.
5. *Kontextsensitive Suche.* Basierend auf dem Termvektor qualitativer Inhalte kann eine erneute Suche angestoßen werden. Entweder aus der anzeigenden bzw. wiedergebenden Anwendung heraus oder als Kontextmenü im Informationsportal wird die unter (2.) beschriebene Suchmaske aufgerufen. Dabei wird der Termvektor des Dokuments als Vorgabewert für die neue Suche verwendet. Der Benutzer kann die Vorgabe wiederum verändern und eine neue Suche starten.

6. *Aufruf von OLAP-Berichten.* Bei der Auswahl von quantitativen Informationen wird der entsprechende OLAP-Bericht angezeigt.
7. *OLAP-Navigation.* Die OLAP-Operationen (Slicing, Dicing, etc.) ermöglichen dem Anwender die Navigation durch den Faktenraum. Bei jedem Navigationsschritt wird der Termvektor aktualisiert. Untersuchungsobjekte und/oder Kennzahlen, die nicht mehr ausgewertet werden, werden entfernt, während neu zur Auswertung hinzugekommene Untersuchungsobjekte und/oder Kennzahlen aufgenommen werden.
8. *Wechsel aus dem OLAP-System in die kontextsensitive Suche.* In jeder Phase der OLAP-Navigation kann der Anwender in die Suchfunktion springen. Der Termvektor der aktuellen Tabellenansicht wird wie unter (5.) beschrieben zur Vorbelegung der Anfrage verwendet.

3.3 Inhalteklassifikation anhand von Metadaten und Schlüsselwörtern

Voraussetzung für eine integrierte Klassifikation von Inhalten ist eine einheitliche Bewertungsgrundlage. Als solche können die Inhalte der oben beschriebenen Metadatenstruktur verwendet werden, indem die Navigationskoordinaten für eine Inhaltebeurteilung herangezogen werden. Die Definition des Dokumentenvektors aus Kapitel 2.3 kann auf die Beschreibung von Inhalten verallgemeinert werden:

Sei $I = \{i_1, \dots, i_m\}$ die Menge der *Inhalte* (hier z. B. OLAP-Berichte) und $A = \{A_1, \dots, A_n\}$ eine Menge von Attributen zu diesen. Die Attributwerte $A_j(i_i) = w_{i,j}$ des *Inhalts* i_i lassen sich als Gewichte auffassen. Die Gewichte werden dann zu einem Vektor $w_i = (w_{i,1}, \dots, w_{i,n}) \in \mathbb{R}^n$, dem sog. *Inhaltevektor* zusammengefasst.

Generell lassen sich die Informationsbedarfe für Aufgabengruppen zwar festlegen. Jedoch sind auch die persönlichen Vorlieben der Führungskräfte zu berücksichtigen.⁴⁵ Unabhängig von der Art der Informationsbedarfsbestimmung stehen unterschiedliche Verfahren zur Inhalteklassifikation zur Verfügung. Neben Verfahren, die auf einer vollständigen, manuellen Klassifikation aller Inhalte basieren, verwenden andere Metainformationen zur automatischen Klassifikation.

⁴⁵ Vgl. Becker, Holten (1998) und Mertens (1999), S. 415.

3.3.1 Kollaborative Inhaltebeurteilung

Die kollaborative Inhaltebeurteilung ist ein gruppenorientiertes Inthaltelassifikationsverfahren. Dabei werden den Anwendern generell alle Inhalte zur Verfügung gestellt. Die Anwender können die von ihnen bearbeiteten oder gelesenen Inhalte beurteilen. Von einem bestimmten Quorum der Gruppenmitglieder als relevant beurteilte Inhalte gelten für die gesamte Gruppe als wichtig.⁴⁶ Alternativ kann der Inhalt an den Vorgesetzten der Gruppe weitergeleitet werden.⁴⁷

Da alle Inhalte, bevor sie gezielt an andere weitergeleitet werden, von mehreren Gruppenmitgliedern als relevant eingestuft wurden, ist eine sehr hohe Qualität der Beurteilung für die Zwecke der Gruppe zu erwarten. Bei von den Gruppenbewertungen abweichenden individuellen Informationsbedürfnissen werden die Anwender jedoch nicht unterstützt. Eine gezielte individuelle Informationssuche bleibt aber möglich, da weiterhin alle Inhalte verfügbar sind. Inwieweit dabei bisher nicht als relevant eingestufte Inhalte für den Einzelnen interessant sein können, ist bei diesem Verfahren nicht zu ermitteln. Nachteilig ist zudem, dass alle Inhalte zumindest von einer Teilmenge einer Gruppe manuell klassifiziert werden müssen, bevor eine Relevanzbeurteilung für die Gruppe vorliegt. Hierdurch werden Ressourcen gebunden und es geht vom Eintreffen eines Inhalts bis zu dessen aktiver Verteilung Zeit verloren. Durch die reine Relevanzanalyse werden vorhandene Beziehungen zwischen Inhalten nicht berücksichtigt und mögliche Strukturierungen unterlassen. Es findet keine integrierende Gliederung der Inhalte statt.

3.3.2 Profil-Definition mit Wenn-Dann-Regeln

In Expertensystemen werden häufig regelbasierte Verfahren zur Darstellung des Wissens über eine Domäne verwendet.⁴⁸ Überträgt man dies auf den Anwendungsfall des Informationsportals, können Wenn-Dann-Regeln basierend auf den Metadaten formuliert werden, durch die eine Inthaltelassifikation vorgenommen wird. Hierbei stehen die Wenn-Klauseln für das Auftreten von Metattribut-Ausprägungen, für welche die in der Dann-Klausel angegebenen Aktionen (hier Klassifikation von Inhalten) ausgeführt werden.

Neben den einzelnen Untersuchungsobjekten (einschließlich der Stammdaten aus den Transaktionssystemen) sind auch die Urteile und die erkannten Schlagworte Bestandteil der Metadatenstruktur. Die Anzahl der zur Bewertung nutzbaren Attributausprägungen ist entspre-

⁴⁶ Vgl. Boger, Kuflik, Shapira, Shoyal (2000), S. 380.

⁴⁷ Vgl. Meier, Fülleborn (1999), S. 455.

⁴⁸ Vgl. Kurbel (1992).

chend groß. Um eine vollständige Vorgabe der Klassifikationsregeln für alle Attributausprägungen zu vermeiden, kann die aggregierende Eigenschaft der Metadatenstruktur verwendet werden: Die durch die Bezugsobjekthierarchien, das Kennzahlensystem und den Thesaurus definierten Generalisierungen/Spezialisierungen können genutzt werden, um mit einer angemessenen Anzahl an Regeln den Informationsbedarf abzubilden. Unter Verwendung der Metaattributausprägungen höherer Ebenen wird der Informationsbedarf grob abgebildet. Inhalte, die einer spezifischeren Differenzierung benötigen, können mit Spezialisierungen der jeweiligen Ausprägungen erfasst werden. Vorteile dieses Ansatzes sind insbesondere darin zu sehen, dass für den ersten Einsatz des Systems kein umfassendes Profil erstellt werden muss. Basierend auf dem grob erfassten Informationsbedarf kann dieser sukzessiv durch Hinzufügen spezialisierter Regeln verfeinert bzw. angepasst werden. Allerdings sind die von der direkten Wissensakquisition für Expertensysteme bekannten Schwierigkeiten⁴⁹ auch bei der Profilerstellung im Informationsportal zu erwarten.

Zusätzlich wirkt sich bei regelbasierten Ansätzen nachteilig aus, dass die Regeln statisch sind und sich nicht automatisch an veränderte Informationsbedarfe anpassen. Nach umfangreichen Anpassungen der Regeln über die Zeit sind ihre Zusammenhänge nur noch schwer zu überblicken und die Regelmenge ist somit kaum noch wartbar.

Durch die Verwendung der Metaattribute zur Inthelteklassifikation wird die Integration der unterschiedlichen Inhaltstypen unterstützt und die Beziehungen zwischen diesen bleiben erhalten. Etabliert haben sich regelbasierte Verfahren bei der Umsetzung von Push-Mechanismen in OLAP- bzw. Data-Warehouse-Systemen. Bei der Verallgemeinerung der Schwellwertregeln des Exception Reportings besteht die Schwierigkeit einer analogen Regeldefinition für Informationen, die als qualitative Aussagen vorliegen bzw. die als quantitative Aussagen in Texten eingebettet sind und daher zunächst extrahiert werden müssen.

3.3.3 Kollaborative Profile

Kollaborative Profile vereinen regelbasierte mit kollaborativer Informationsklassifikation. Die individuellen (regelbasierten) Profile der Gruppenmitglieder werden auf Bereiche gleicher Beurteilungen untersucht. Bereiche, die von mehreren Personen gleich eingeschätzt werden, werden in ein Gruppenprofil aufgenommen, das allen Gruppenmitgliedern zugeordnet ist. Die Klassifikation von Inhalten für ein Subjekt ist somit sowohl von der Gruppen- als auch von der individuellen Beurteilung abhängig.

⁴⁹ Vgl. Kurbel (1992), S. 70.

Durch die Kombination des individuellen Profils mit den Profilen der übrigen Gruppenmitglieder wird die Schwierigkeit der direkten Wissensakquisition (Wissensextraktion) verringert, da nicht mehr jeder ein umfassendes individuelles Profil erstellen muss. Vielmehr ist davon auszugehen, dass vom Einzelnen nicht betrachtete aber aus Sicht der Gruppe relevante Bereiche durch das Gruppenprofil abgedeckt werden. Sowohl der Forderung nach einer Metadaten-basierten Inhalteintegration und -klassifikation als auch nach einer möglichen Individualisierung der Informationsbeschaffung wird Rechnung getragen. Das Problem des sich ändernden Informationsbedarfs wird hierdurch jedoch nicht oder nur unzureichend gelöst.

3.3.4 Profilentwicklung mittels lernender Verfahren

Die bisher vorgestellten Verfahren beruhen auf der direkten Wissensexplikation oder der manuellen Klassifikation von Informationen und setzen somit die aktive Mitarbeit der Benutzer voraus. Lernende Verfahren hingegen beobachten den Benutzer bei der Arbeit mit dem System mit dem Ziel, das Verhalten zu erlernen und so die Klassifikation von Informationen vorhersagen und den Anwender gezielt zu den für ihn relevanten Informationen führen zu können. Beispielhaft für die Menge der lernenden Verfahren seien hier das Relevance Feedback (s. Kapitel 2.3.2) und künstliche neuronale Netze genannt. Auf die Verwendung künstlicher neuronaler Netze zur Profilentwicklung wird im Folgenden ausführlich eingegangen.

Ein *künstliches neuronales Netz* (KNN) besteht aus Neuronen und gerichteten, gewichteten Verbindungen zwischen diesen.⁵⁰ Neuronen können über die *Verbindungen* eingehende Signale (mathematisch) verarbeiten und daraus eine Ausgabe erzeugen. Über die Verbindungen werden diese an benachbarte Neuronen weitergegeben. Die Gewichte der Verbindungen bestimmen, mit welchem Einfluss die Information bei den nachgelagerten Neuronen ankommt. Durch Veränderung der Verbindungsgewichte bzw. der Verarbeitungsfunktion in den Neuronen sind KNN in der Lage, komplizierte Funktionen zu erlernen. Für den Einsatz in einem Informationsportal zeichnen sie sich unter anderem dadurch aus, dass sie robust gegenüber verrauschten Daten sind: Auch bei Eingabewerten, die nur annähernd mit der Realität übereinstimmen (geschätzte Daten), ist das Netz in der Lage, angemessene Ausgaben zu erzeugen.⁵¹ So lassen sich die Dokumentenvektoren als geschätzte Werte interpretieren, da statistische Verfahren keine korrekte Erfassung der Bedeutung der Schlagworte für den Inhalt von Texten garantieren können. Weitere geschätzte Daten sind die Urteile des Benutzers über die Inhalte.

⁵⁰ Vgl. z. B. Rojas (1996), Zell (1994).

⁵¹ Vgl. Zell (1994), S. 26ff.

Die Inhaltebeurteilungen sind zur Vermeidung von zusätzlichem Aufwand weitestgehend automatisiert zu erfassen. Relativ einfach lassen sich hierfür Erkenntnisse aus dem Navigationsverhalten des Benutzers gewinnen. Bei den Inhalten, die sich der Anwender anzeigen lässt, ist davon auszugehen, dass diese als relevant angesehen werden. Ebenso sind die in der Navigation und der Suche verwendeten Terme als Indikator für Relevanz zu interpretieren. Neben der automatischen Erfassung der Klassifikationen ist dem Anwender aber immer die Möglichkeit einzuräumen, Inhalte explizit zu klassifizieren. Auf diese Weise besteht die Gelegenheit, Inhalte als nicht-relevant zu beurteilen; insbesondere, wenn das KNN fälschlicherweise eine hohe Relevanz vorhergesagt hat und somit eine Fehleinschätzung vorliegt.

Die Trainingsmenge muss für eine gute Qualität der Vorhersage bestimmte Kriterien erfüllen. Zu diesen zählen:

- Die Mengen müssen groß genug sein, um ein sog. *Overfitting* zu vermeiden.
- Die Merkmale (Navigationskoordinaten) der Inhalte sollten sich stark unterscheiden bzw. die gesamte Spannweite des Spektrums abdecken.
- Die Inhalte sollten das gesamte Spektrum der Relevanz abdecken.

Das Lernverfahren ist vergleichbar mit dem Lösen eines Gleichungssystems: Gesucht werden die Parameterwerte des Netzes (Verbindungsgewichte, Berechnungsfunktionen), die das gesuchte Ergebnis am besten beschreiben. Analog zu Gleichungssystemen können KNN nicht eindeutig bestimmt werden, wenn die Menge der verfügbaren Trainingsdaten nicht ausreicht, um die Parameter des KNN mit eindeutigen Werten zu belegen. In solchen Fällen kommt es zu einem *Overfitting*. Dabei verringert das Netz seinen Vorhersagefehler (basierend auf den Trainingsdaten) zu Lasten seiner Verallgemeinerungsfähigkeit. Mit zunehmender Einsatzdauer des KNN im Informationsportal nimmt jedoch die Menge der klassifizierten Inhalte zu und die erste Anforderung wird somit automatisch erfüllt. Die Gewährleistung der beiden anderen geforderten Eigenschaften ist im normalen Betrieb jedoch nicht trivial zu erfüllen: Anwender arbeiten in der Regel in einem bestimmten Aufgabenbereich. Die hierbei benötigten (verarbeiteten) Inhalte sind durchweg als relevant zu betrachten und weisen dabei eine große inhaltliche Ähnlichkeit auf. Sowohl der fehlenden Streuung der Navigationskoordinaten als auch der durchweg hohen Relevanz der Inhalte kann das System durch gezielte Aufforderung zur Inhaltebeurteilung entgegenwirken. Hierbei sind Inhalte mit nicht betrachteten Navigationskoordinaten sowie mit als gering eingeschätzter Relevanz zu präsentieren.

Eine Unterstützung von Personengruppen bei der Informationsbewertung kann durch die Bereitstellung eines zusätzlichen KNN für die Gruppe geboten werden. Dem Netz werden die Relevanzbeurteilungen aller Gruppenmitglieder für den Lernprozess vorgelegt und so das der Gruppe gemeinsame Klassifikationsmuster erlernt. Informationen, die durch das Netz der

Gruppe als wichtig eingestuft wurden, können dann unabhängig von der individuellen Vorhersage an alle Mitglieder der Gruppe weitergegeben werden. Des Weiteren kann das Gruppennetz neuen Mitgliedern der Gruppe als initiale Kopie zugewiesen werden. Es stellt sicher, dass zu Beginn alle wirklich relevanten Informationen erkannt werden. Die individuellen Präferenzen des neuen Gruppenmitglieds erlernt das Netz im Einsatz.

Durch die implizite Erfassung der Klassifikationsvorschriften/-funktionen wird auf eine explizite, formale Beschreibung des Informationsbedarfs durch den Benutzer verzichtet. Schwierigkeiten bei der Formulierung des Bedarfs aufgrund der eingeschränkten Mächtigkeit der Repräsentationsform oder der (empfundenen) Schwierigkeit bei der direkten Explikation durch den Anwender werden somit vermieden. Die kontinuierliche Erfassung der Präferenzen bei der Navigation und die kontinuierliche Beurteilung von Inhalten ermöglichen eine Anpassung an sich ändernde Umweltbedingungen und somit an Informationsbedarfe. Eine mittel- bzw. langfristige Erfassung der Interessen kann gewährleistet werden.

Ein den KNN immanentes Problem besteht darin, dass diese nur schwer darauf zu untersuchen sind, was sie gelernt haben.⁵² Sie stellen eine Art „black box“ dar, deren Verarbeitungsschritte von außen nicht interpretiert werden können. Wenn auch die einzelnen Gewichte des Netzes bekannt sind und somit die Berechnung nachvollzogen werden kann, so ist doch die semantische Bedeutung der Gewichte nicht direkt bestimmbar.⁵³

Die fehlende Möglichkeit, künstliche neuronale Netze auf die Bedeutung des Erlernenen zu untersuchen, erschwert das Erkennen von falsch Erlernenem oder von Gründen für Fehlverhalten. Lediglich die mangelnde Qualität der Vorhersage kann Indizien für falsch trainierte Netze liefern. Um diesem Problem weitestgehend technisch vorzubeugen, können Verfahren zur Größen- und Strukturadaption eingesetzt werden.⁵⁴

Könnten Bewertungsregeln extrahiert werden, wäre die Voraussetzung zur gezielten Beeinflussung der Vorhersage des KNN gegeben, indem die extrahierten Regeln modifiziert werden. Die korrigierten Regeln bilden die Grundlage zur Neudefinition des Netzes. Die Möglichkeit, Regeln direkt in das KNN einzupflegen, bietet zudem die Voraussetzung, Regelmengen als repräsentative Ausgangslösungen bereitzustellen, um so den Trainingsaufwand für ein neues Netz zu reduzieren.

⁵² Vgl. z. B. Alexander, Mozer (1999), S. 479.

⁵³ Zu ersten Ansätzen zur Regelextraktion aus KNN in anderen Anwendungsbereichen vgl. z. B. Alexander, Mozer (1999) und Omlin, Giles (1996).

⁵⁴ Vgl. Zell (1994), S. 319-334; Rojas (1996), S. 392-397.

Der numerische Trainingsaufwand eines künstlichen neuronalen Netzes wächst stark mit der Anzahl seiner Neuronen und Verbindungen.⁵⁵ Einem Neuron, das Informationen von einem zu bewertenden Inhalt entgegennimmt, muss immer die Kennzahl der gleichen Navigationskoordinate vorgelegt werden. Schon die Anzahl der Bezugsobjekte lässt somit die Netzgröße sehr stark anwachsen. Grundlage für eine Reduktion der Eingangsneuronen kann dabei – wie schon in Kapitel 3.3.2 beschrieben – die aggregierende Eigenschaft der Inhalte der Metadatenstruktur sein.

3.4 DV-Konzeption zur Umsetzung der Ansätze

Die gleichzeitige Etablierung von Data-Warehouse- und Content-Management-Systemen kann darauf zurückgeführt werden, dass die Verwaltung der unterschiedlichen Inhaltstypen jeweils eigene Methoden erfordert. Für die nahe Zukunft ist zu erwarten, dass auch weiterhin die jeweiligen Anwendungssysteme parallel zueinander eingesetzt werden. Ein integriertes Enterprise Information Portal kann durch eine Software-System-Architektur, welche die Metadaten-Strukturen beider Systeme zusammenführt, umgesetzt werden. Über die Einführung einer zentralen Koordinationsinstanz kann die Integration der Metadatenbasen sichergestellt werden (vgl. Abb. 5).⁵⁶ Sie gewährleistet, dass Veränderungen in einem System konsistent in das andere System übertragen werden. Varianten dieser Architektur lassen sich durch die Realisierung dezentraler oder teilzentralisierter Metadaten-Synchronisation erzielen.⁵⁷

Solange die Differenzierung der Systemklassen bestehen bleibt, wird die Integration der Anwendungsbereiche durch folgende Eigenschaften unterstützt:

- Data-Warehouse- und OLAP-Systeme müssen die Inhalte ihrer Metadaten-Strukturen (Bezugsobjekt-Hierarchien und Kennzahlensysteme) dem Content-Management-System oder einem Informationsportal-Server zur Verfügung stellen.
- Das Content-Management-System muss in der Lage sein, sowohl Metadaten-Strukturen anderer Systeme als Basis für eigene Strukturen zu verwenden als auch diese anschließend zu erweitern und die Erweiterungen anderen Systemen (z. B. einem Informationsportal-Server) zur Verfügung zu stellen.

⁵⁵ Vgl. Rojas (1996), S. 212ff.

⁵⁶ Vgl. Rieger, Kleber, von Maur (2000).

⁵⁷ Vgl. Do, Rahm (2000), S. 8ff.

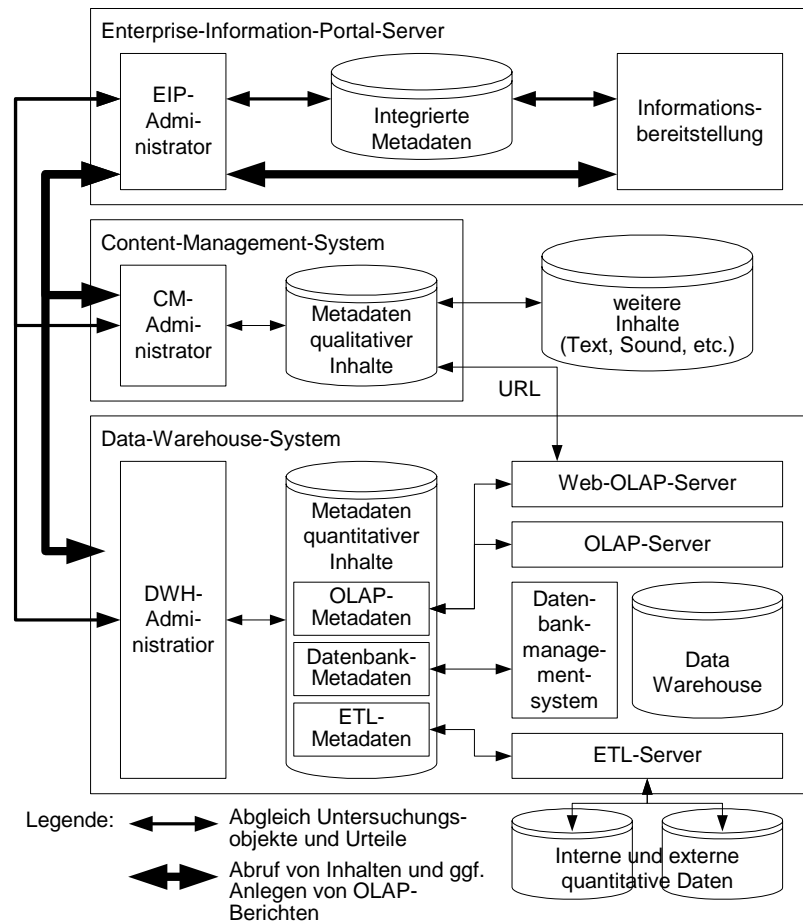


Abb. 5: DV-konzeptionelle Informationssystemarchitektur zur Metadaten-Integration

Das Informationsportal nimmt Inhalte der Metadaten-Strukturen des Data-Warehouse- und des Content-Management-Systems auf und führt sie in einer integrierten Metadaten-Struktur zusammen. Aus dem Content-Management-System werden die Metaattributausprägungen in das Informationsportal übernommen. Die Hierarchien über die Bezugsobjekte und Kennzahlen des Data-Warehouse-Systems bilden das Gerüst für einen Thesaurus gemäß den fachkonzeptionellen Vorgaben aus dem Data Warehouse bzw. gleichen einen vorhandenen Thesaurus entsprechend an. Das Informationsportal überträgt den neuen Thesaurus in die Metadaten des Content-Management-Systems. Auf dem angepassten Thesaurus basierend kann eine erneute Klassifikation der CM-Inhalte durchgeführt werden. Wird durch das Informationsportal das Erstellen eines neuen OLAP-Berichts veranlasst (vgl. Kapitel 3.2), so werden die zum Bericht gehörenden Informationen über Bezugsobjekte und Kennzahlen in die Metadaten-Struktur des Data Warehouse geschrieben und dort der neue Bericht angelegt.

Die Inhaltebereitstellung des Informationsportals greift auf die integrierten Metadaten zurück. Für den Abruf der Inhalte selbst erfolgt ein Rückgriff auf die spezialisierten Systeme. Aus den Spezialsystemen heraus kann jederzeit wieder in die Navigationsstrukturen des Informationsportals zurückgesprungen werden. Hierbei werden die in der Anwendung vollzogenen

Navigationsschritte berücksichtigt: Die Metadaten des jeweils angezeigten Inhalts bestimmen die Menge der im Informationsportal angezeigten Inhalte.

Die integrierte Verwaltung von Metadaten sowie von quantitativen und qualitativen Inhalten in einem gemeinsamen Datenbanksystem wird in Abb. 6 auf DV-konzeptueller Ebene verdeutlicht. Das relationale Datenbankschema ist zur besseren Anschaulichkeit mit exemplarischen Datensätzen gefüllt. Die eigentlichen Inhalte werden in der Datenbanktabelle „Inhalte“ gespeichert. Je nach Inhaltstyp bietet sich eine *Partitionierung* dieser Tabelle an. Die übrigen Tabellen dienen der Strukturierung der Untersuchungsobjekte und Urteile.

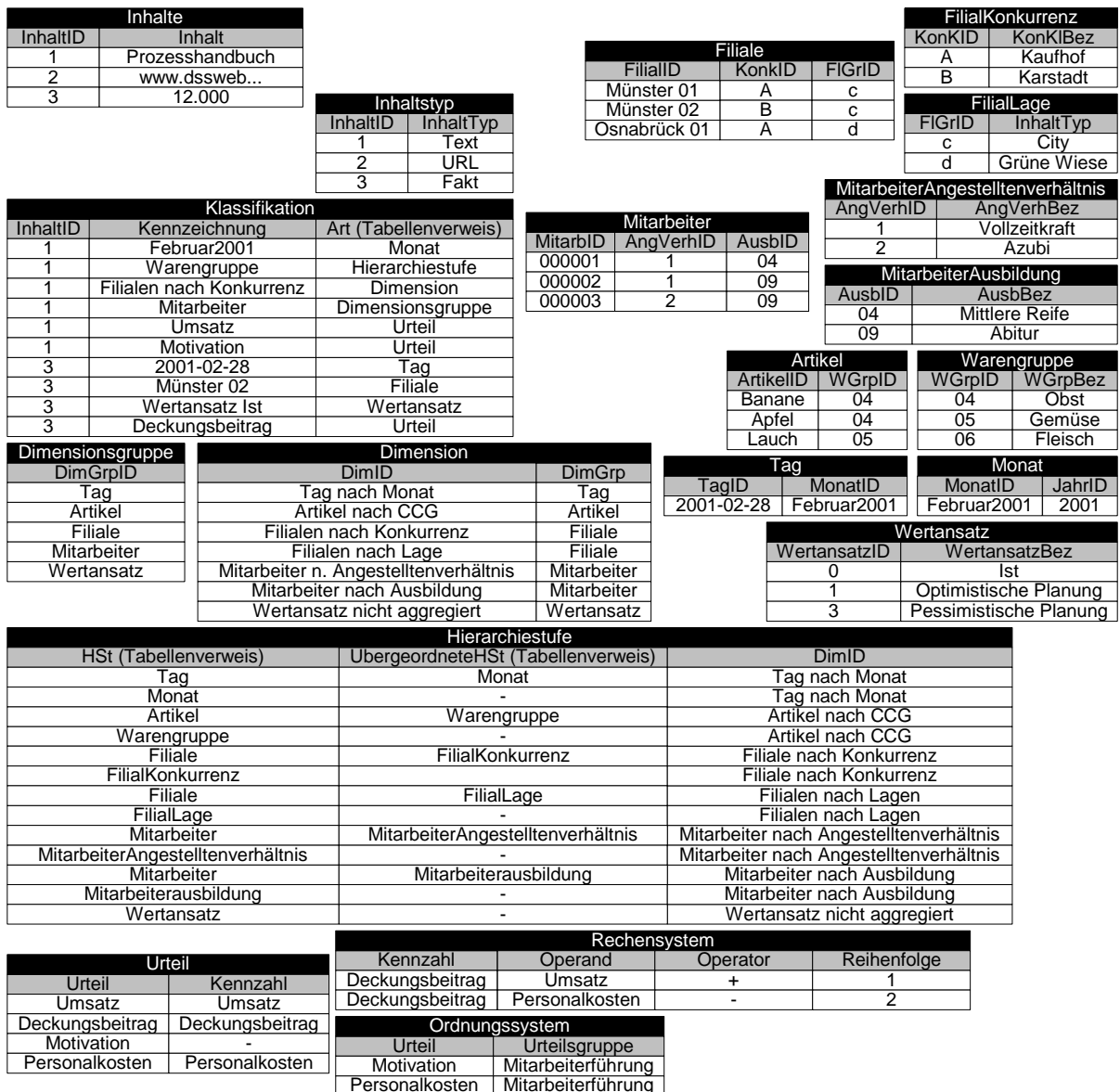


Abb. 6: DV-Konzept zur Metadaten-Integration

Für die fernere Zukunft kann erwartet werden, dass die Trennung von Data-Warehouse- und Content-Management-Systemen überwunden wird und sich eine neue Klasse von Systemen

herausbildet. Die Abb. 7 zeigt eine mögliche Architektur eines solchen Informationssystems. OLAP-Server und CM-Server werden dabei bewusst explizit genannt, um zu verdeutlichen, dass trotz des integrierten Systems weiterhin die unterschiedlichen Methoden für unterschiedliche Inhaltstypen durch eigene Subsysteme bereitgestellt werden können. Der ETL-Server importiert dabei sowohl quantitative als auch qualitative Inhalte aus internen als auch externen Quellen.

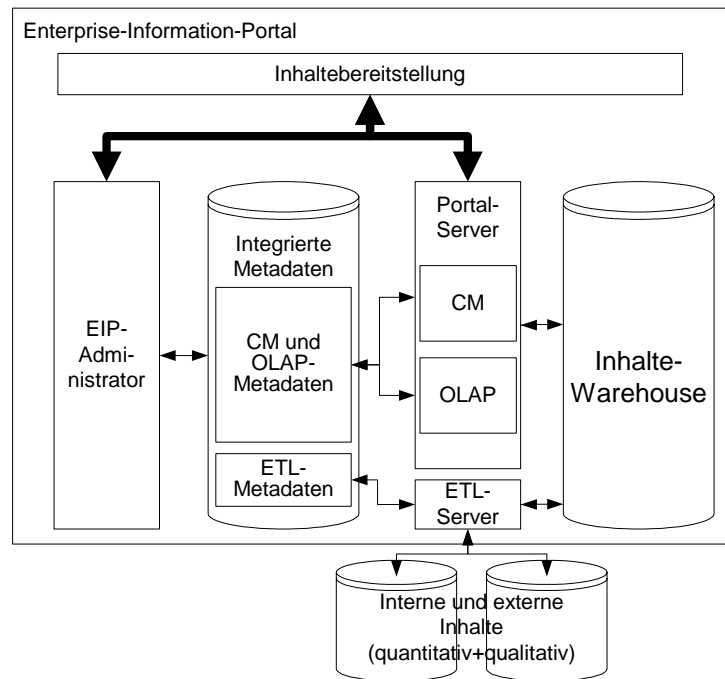


Abb. 7: DV-konzeptionelle Architektur eines vollständig integrierten Informationsportals

4 Verwandte Arbeiten

In der Forschung existieren zurzeit mehrere Strömungen, die sich mit der Integration von quantitativen und qualitativen Informationen sowie mit Benutzermodellierungsproblemen in Verbindung mit Information Retrieval und Information Filtering auseinander setzen.

Ein Vorschlag zur Metadaten-basierten Integration qualitativer und quantitativer Informationsressourcen liegt von RIEGER ET AL. vor.⁵⁸ Die von ihnen entwickelte Systemarchitektur ist mit der in dieser Arbeit vorgeschlagenen vergleichbar. Die weiterhin vorhandenen separaten Datenbanken für die Domänen-spezifische Verarbeitung der Metadaten von qualitativen und quantitativen Inhalten werden über eine Metadatenbank zusammengeführt, in der die Relationen („links“) zwischen diesen verwaltet werden. Unterschiede zum vorliegenden Beitrag bestehen insbesondere darin, dass die Inhalteintegration nicht über ein analogiebasiertes gemeinsames Konstruktionsprinzip für die Metaattributausprägungen vorgenommen wird. Die Autoren beschränken sich zudem auf die Einführung einer Instanz zur Metadatenintegration. Eine Inhaltnavigation, welche die Grenzen zwischen den zugrundeliegenden Inhalteverwaltungssystemen weitreichend überwindet, wird nicht vorgeschlagen.

MEIER und FÜLLEBORN entwickeln zur Integration externer Führungsinformationen aus dem Internet in das Modul SEM von SAP einen Redaktionsleitstand.⁵⁹ Dieser wird sowohl von Informationsrechercheuren als auch von den Informationssuchenden verwendet. Erstgenannte haben dabei die Aufgabe, zum einen zu den verfügbaren Informationsquellen beschreibende Profile anzulegen und zum anderen die Informationsbedarfe für die unterschiedlichen Rollen innerhalb des Unternehmens zu definieren. Der Forderung nach einer Individualisierung des Informationsbedarfs wird dadurch Rechnung getragen, dass die Anwender (Informationssuchenden) ihr eigenes Profil anpassen können. Grundlage für die Definition des Profils ist eine integrierte Metadatenstruktur. Diese wird federführend aus den Metadaten des Data-Warehouse-Systems der SAP (Business Information Warehouse) gebildet. Die Integration von qualitativen Informationen wird über ein an das DWH angepasstes Indexierungsvokabular sichergestellt. Ein Profil kann neben diesen Metadaten auch weitergehende Eingrenzungen in Form von Metaattribut-Ausprägungen wie die Selektion von Produktgruppen, Kunden oder Zeiträumen beinhalten. Die Aufnahme von neuen Informationen in das System erfolgt halbautomatisch durch Werkzeuge des Information Retrieval. Hierbei kommen höherwertige Tools zum Einsatz, die auch Ansätze semantischer Textanalysen durchführen (z. B. überführen sie eine relative Zeitangabe wie „diesen Monat“ in absolute Zeitangaben).

⁵⁸ Vgl. Rieger, Kleber, von Maur (2000).

⁵⁹ Vgl. Meier, Füllerborn (1999).

BOGER ET AL. haben ein künstliches neuronales Netz entwickelt, mit dem sie E-Mail-Nachrichten nach der Wichtigkeit für den Empfänger beurteilen.⁶⁰ Sie verwenden dazu ein Back-propagation-Netz mit drei Schichten. Die Eingabeschicht erhält den Dokumentenvektor (mit reellen Zahlen) einer Nachricht als Eingabe. Die versteckte Schicht besteht aus maximal 6 Neuronen und liefert ihre Ausgabewerte an genau ein Neuron in der Ausgabeschicht weiter. Mit diesem Netz wurde eine Vorhersagegenauigkeit von 76 % bis 99 % erzielt. Im Vergleich zu zuvor gemachten Untersuchungen, die maximal 79 % bei Bewertung mit manuell erstellten Regeln erreichten, ist dieser Ansatz als sehr überzeugend einzustufen.⁶¹

Am Institut für Wirtschaftsinformatik der Westfälischen Wilhelms-Universität Münster wird mit dem Projekt „PI-Agent“ untersucht, inwieweit sich die Technologie der künstlichen neuronalen Netze bei der Bewertung von Nachrichten aus dem Internet mit einem Agenten-Ansatz kombinieren lässt.⁶² Da es jedoch keine dominante Referenztopologie für künstliche neuronale Netze gibt, sind noch zusätzliche Arbeiten zur Auswahl der geeigneten Netzarchitektur (z. B. Anzahl der verborgenen Schichten) und der optimalen Parametrisierung der Netze durchzuführen.

⁶⁰ Vgl. Boger, Kuflik, Shapira, Shoval (2000).

⁶¹ Vgl. Shapira, Shoval, Hanani (1999).

⁶² Vgl. ausführlich Kuropka, Serries (2001).

5 Zusammenfassung und Ausblick

Die Kombination bestehender Instrumente zur Informationsbereitstellung in einer gemeinsamen Benutzeroberfläche allein wird dem Anspruch eines integrierten Informationsportals nicht gerecht. Vielmehr muss sichergestellt werden, dass sowohl quantitative Inhalte aus Data-Warehouse- oder OLAP-Systemen als auch qualitative Inhalte aus Content-Management-Systemen miteinander in Beziehung gesetzt werden. Der Beitrag zeigt, wie über eine Metadaten-Struktur, die Ausprägungen aus Data-Warehouse- und Content-Management-Systemen zusammenführt, eine weitreichende Integration realisiert werden kann. Die Metadaten-Struktur dient als ein Gerüst zur Navigation in den Inhalten beider Teilsysteme. Für die Spezifikation dieser Navigationsstruktur schlägt der Beitrag ein Konstruktionsprinzip vor, welches die für die Konzeption von Data Warehouses bewährte Trennung zwischen Untersuchungsobjekten und Urteilen verallgemeinert. Dem Konstruktionsprinzip folgend wurde eine Metadatenbasis entworfen, welche die technische Grundlage für eine integrierte Bereitstellung von quantitativen und qualitativen Inhalten darstellt. Neben Pull-Mechanismen – beispielsweise in Form von einer Suche über Navigationskoordinaten – können Push-Mechanismen ebenfalls auf den Metadaten basierend umgesetzt werden. Als wichtige Ausprägungsform des Push-Prinzips wird die über Benutzerprofile gesteuerte Bereitstellung von Inhalten in Informationsportalen ausführlich behandelt.

Das vorgestellte Zusammenspiel von Push- und Pull-Mechanismen bei der Informationsbereitstellung bietet das Potenzial, die Effektivität und Effizienz der Entscheidungsunterstützung deutlich zu fördern. Es wird erwartet, dass die realisierbaren komparativen Wettbewerbsvorteile einer solchen Lösung für Softwarehersteller Anreiz sein werden, ihre auf bestimmte Informationstypen spezialisierten Data-Warehouse- bzw. Content-Management-Systeme funktional in die eine oder andere Richtung zu erweitern bzw. bei der Realisierung einer integrierten Lösung zu kooperieren. Es kann sich eine neue Systemklasse herausbilden, welche die Funktionalitäten der beiden heute unterschiedenen Systemklassen abdeckt. Solange die Differenzierung der Systemklassen bestehen bleibt, wird die Integration der Anwendungsbereiche durch offene Systeme gefördert. Aktuell am Markt verfügbare Systeme erfüllen die Anforderungen an die notwendige Offenheit der Systeme nur unzureichend. Die Entwicklung von standardisierten Austauschformaten bleibt daher Ziel aktueller Forschung.

In weiteren Arbeiten soll die Realisierbarkeit des vorgestellten Konzepts durch den Bau von Prototypen, die Einzelprobleme fokussieren, untersucht werden. Insbesondere die Überlegungen zur automatisierten Generierung von OLAP-Berichten anhand von Termen sind auszuarbeiten. Hierbei können Vorarbeiten des MetaMIS-Projektes genutzt werden, das ebenfalls die

automatische Erstellung von OLAP-Berichten zum Gegenstand hat.⁶³ Im Unterschied zum hier behandelten Konzept basiert die Berichtsgenerierung im Rahmen des MetaMIS-Projektes allerdings nicht auf Termvektoren sondern auf fachkonzeptionellen Data-Warehouse-Modellen.

Darüber hinaus ist zu klären, wie auf Veränderungen der Thesaurusklassen, wie sie z. B. aus Modifikationen von Warengruppenhierarchien oder der Akquisition neuer Kunden resultieren, zu reagieren ist. Da die Klassifikationsverfahren aufwändig bzw. rechenintensiv sind, ist eine Neuklassifikation aller Inhalte möglichst zu vermeiden. Es erscheint ein lohnendes Ziel zu sein, Heuristiken zu entwickeln, welche die Menge der neu zu klassifizierenden Inhalte auf ein möglichst geringes Maß beschränken.

Neben der Verbesserung der Verfahren zur automatisierten Klassifikation erscheint es sinnvoll, den bisherigen Fokus auf Einzelpersonen oder Gruppen auf die Gestaltung der Wertschöpfungskette der Informationsbereitstellung auszuweiten. Es ist zu untersuchen, inwieweit die einzelnen Benutzerprofile Aufschluss geben können, welche Informationszulieferer (z. B. in Form der Auswahl geeigneter Nachrichtendienste) von einem Unternehmen überhaupt berücksichtigt werden sollten. Neben der Unterstützung einer Lieferantenbewertung wären zudem Aussagen darüber von Interesse, welche Auswirkungen auf das Marktgeschehen mit der Weitergabe von (aggregierten, anonymisierten) Benutzerprofilen an Informationslieferanten verbunden sind.

Neben der Entwicklung technischer Komponenten bietet auch die methodische Unterstützung der Konstruktion der Untersuchungsobjekt- und Urteilsstrukturen Ansätze zu weiterer Forschung. Der grundsätzlich kreative Akt sollte durch Ausgangslösungen unterstützt werden, die durch automatisierte, gegebenenfalls vorhandene Modelle auswertende Methoden generiert werden.

⁶³ Vgl. z. B. Holten, Knackstedt, Becker (2001), S. 60-62.

Literaturverzeichnis

- Austin, J. R.: Theorie der Sprechakte. 2. Aufl., Reclam, Ditzingen 1979.
- Alexander, J. A.; Mozer, M. C.: Template based procedures for neural network interpretation. In: Neural Networks, 12 (1999): S. 479-498.
- Back-Hock, A.: Executive Information Systeme (EIS). Kostenrechnungspraxis (1991) 1, S. 48-50.
- Bauer, A.; Günzel, H.: Begriffliche Einordnung. In: A. Bauer, H. Günzel (Hrsg.): Data-Warehouse-Systeme. Architektur, Entwicklung, Anwendung. Dpunkt, Heidelberg 2001, S. 5-10.
- Becker, J.; Holten, R.: Fachkonzeptuelle Spezifikation von Führungsinformationssystemen. Wirtschaftsinformatik 40 (1998) 6, S. 483-492.
- Boger, Z.; Kuflik, T.; Shapira, B.; Shoval, P.: Information Filtering and Automatic Keyword Identification by Artificial Neural Networks. In: H. R. Hansen, M. Bichler, H. Mahrer: Proceedings of the 8th European Conference on Information Systems (ECIS 2000), Volume 1, Wien 2000, S. 379-385.
- Büchner, H.; Zschau, O.; Traub, D.; Zahradka, R.: Web Content Management. Websites professionell betreiben. Galileo, Bonn 2001.
- Bullinger, H.-J.; Koll, P.: Chefinformationssysteme (CIS). In: H. Krallmann (Hrsg.): Rechnergestützte Werkzeuge für das Management: Grundlagen, Methoden, Anwendungen. Berlin 1992, S. 49-72.
- Do, H. H.; Rahm, E.: On Metadata Interoperability in Data Warehouses. Report Universität Leipzig (2000) 1, <http://dol.uni-leipzig.de/pub/2000-13> [05.02.2001].
- Gabriel, R.; Gluchowski, P.: Semantische Modellierungstechniken für multidimensionale Datenstrukturen. Theorie und Praxis der Wirtschaftsinformatik, 34 (1997) 195, S. 18-37.
- Groffmann, H.-D.: Kooperatives Führungsinformationssystem. Grundlagen, Konzept, Prototyp. Gabler, Wiesbaden 1992.
- Henneböle, J.: Executive Information Systems für Unternehmensführung und Controlling. Strategie – Konzeption – Realisierung. Gabler, Wiesbaden 1995.
- Holten, R.: Entwicklung von Führungsinformationssystemen. Ein methodenorientierter Ansatz. Gabler, Wiesbaden 1999.
- Holten, R.; Ehlers, L.: Repository zur Integration externer Informationen in Führungsinformationssysteme. In: W. Uhr, S.-E. Breuer (Hrsg.): Integration externer Informationen in Management Support Systems. Wirtschaftsinformatik Fachtagung, Technische Universität Dresden 08.-09.10.1998. Dresden 1998, S. 123-136.
- Holten, R.; Knackstedt, R.: Führungsinformationssysteme – Historische Entwicklung und Konzeption. Arbeitsbericht Nr. 55 des Instituts für Wirtschaftsinformatik. Hrsg.: J. Becker, H.L. Grob, U. Müller-Funk, G. Vossen. Münster 1997.
- Holten, R.; Knackstedt, R.; Becker, J.: Betriebswirtschaftliche Herausforderungen durch Data-Warehouse-Technologien. In: R. Schütte, T. Rotthowe, R. Holten (Hrsg.): Data Warehouse Managementhandbuch. Konzepte, Software, Erfahrungen. Berlin u. a. 2001, S. 41-64.

- Jahnke, B.; Groffmann, H.-D.; Kruppa, S.: On-Line Analytical Processing (OLAP).
Wirtschaftsinformatik, 38 (1996) 3, S. 321-324.
- Kurbel, K.: Entwicklung und Einsatz von Expertensystemen. 2. Aufl., Springer, Heidelberg
1992.
- Kuropka, D.; Serries, T.: Personal Information Agent. In: Knorr, K. (Hrsg.): Agents in E-
Business (AgEB'01), Zürich 2001.
- Kurz, A.: Data Warehousing. Enabling Technology. MITP, Bonn 1999.
- Leavitt, H. J.; Whisler, T. L.: Management in the 1980's: New Information Flows Cut New
Organization Channels. Harvard Business Review, 36 (1958) 6, S. 41-48.
- Lüdi, A.: Personalize or Perish. Electronic Markets, 7 (1997) 3, S. 22-25.
- Luhmann, N.: Zweckbegriff und Systemrationalität. 6. Aufl., Suhrkamp, Frankfurt am Main,
1999.
- Meier, M.; Fülleborn, A.: Integration externer Führungsinformationen aus dem Internet im
SAP Strategic Enterprise Management (SEM). Wirtschaftsinformatik, 41 (1999) 5, S.
449-457.
- Mertens, P.: Integration interner, externer, qualitativer und quantitativer Daten auf dem Weg
zum Aktiven MIS. Wirtschaftsinformatik, 41 (1999) 5, S. 405-415.
- Omlin, C. W.; Giles, C. L.: Extraction of Rules from Discrete-time Recurrent Neural
Networks, In Neural Networks, 9 (1996) 1: S. 41-52.
- Oppelt, R.U.G.: Computerunterstützung für das Management: Neue Möglichkeiten der
computerbasierten Informationsunterstützung oberster Führungskräfte auf dem Weg
von MIS zu EIS? München, Wien 1995.
- o. V. (MicroStrategy): Microstrategy 7. Administrator Guide. Juni 2000.
- Picot, A.; Reichwald, R.; Wigand, R. T.: Die grenzenlose Unternehmung: Information,
Organisaiton und Management. Lehrbuch zur Unternehmensführung im
Informationszeitalter. 2. Aufl., Gabler, Wiesbaden 1996.
- Reichmann, T.: Controlling mit Kennzahlen und Managementberichten. Grundlagen einer
systemgestützten Controlling-Konzeption. 5. Aufl., Vahlen, München 1997.
- Riebel, P.: Gestaltungsprobleme einer zweckneutralen Grundrechnung. ZfbF, 31 (1979),
S. 863-893.
- Rieger, B.; Kleber, A.; von Maur, E.: Metadata-Bases Integration of Qualitative and
Quantitative Information Resources Approaching Knowledge Management. In: H. R.
Hansen, M. Bichler, H. Mahrer: Proceedings of the 8th European Conference on
Information Systems (ECIS 2000), Volume 1, Wien 2000, S. 372-378.
- Rojas, R.: Theorie der neuronalen Netze. Eine systematische Einführung. 4. Aufl., Springer,
Berlin u.a. 1996.
- Salton, G.; McGill, M. J.: Information Retrieval. Grundlagen für Informationswissenschaftler.
McGraw-Hill, Hamburg 1987.
- Shilakes, C. C.; Tylman, J.: Enterprise Information Portals. New York 1998.
- Shapira, B.; Shoval, P.; Hanani, U.: Experimentation with an Information Filtering System
that Combines Cognitive and Sociological Filtering Integrated with User
Stereotypes, In Journal: Decision Support Systems, 27 (1999), S. 5-24.

- Searle, J. R.: Sprechakte. Ein sprachphilosophischer Essay. 7. Aufl., Suhrkamp, Frankfurt am Main 1997.
- Schütte, R.: Grundsätze ordnungsmäßiger Referenzmodellierung. Konstruktion konfigurations- und anpassungsorientierter Modelle. Gabler, Wiesbaden 1998.
- Vetschera, R.: Informationssysteme der Unternehmensführung. Berlin, Heidelberg 1995.
- Weber, J.: Einführung in das Controlling. 4. Aufl., Schäffer-Poeschel, Stuttgart 1993.
- Weiber, R.; McLachlan, C.: Wettbewerbsvorteile im Electronic Business. In: R. Weiber (Hrsg.): Handbuch Electronic Business. Informationstechnologien – Electronic Commerce – Geschäftsprozesse. Gabler, Wiesbaden 2000, S. 117-148.
- Zell, A.: Simulation Neuronaler Netze. Addison-Wesley, Bonn u. a. 1994.

Arbeitsberichte des Instituts für Wirtschaftsinformatik

- Nr. 1 Bolte, Ch., Kurbel, K., Moazzami, M., Pietsch, W.: Erfahrungen bei der Entwicklung eines Informationssystems auf RDBMS- und 4GL-Basis; Februar 1991.
- Nr. 2 Kurbel, K.: Das technologische Umfeld der Informationsverarbeitung - Ein subjektiver 'State of the Art'-Report über Hardware, Software und Paradigmen; März 1991.
- Nr. 3 Kurbel, K.: CA-Techniken und CIM; Mai 1991.
- Nr. 4 Nietsch, M., Nietsch, T., Rautenstrauch, C., Rinschede, M., Siedentopf, J.: Anforderungen mittelständischer Industriebetriebe an einen elektronischen Leitstand - Ergebnisse einer Untersuchung bei zwölf Unternehmen; Juli 1991.
- Nr. 5 Becker, J., Prischmann, M.: Konnektionistische Modelle - Grundlagen und Konzepte; September 1991.
- Nr. 6 Grob, H. L.: Ein produktivitätsorientierter Ansatz zur Evaluierung von Beratungserfolgen; September 1991.
- Nr. 7 Becker, J.: CIM und Logistik; Oktober 1991.
- Nr. 8 Burgholz, M., Kurbel, K., Nietsch, Th., Rautenstrauch, C.: Erfahrungen bei der Entwicklung und Portierung eines elektronischen Leitstands; Januar 1992.
- Nr. 9 Becker, J., Prischmann, M.: Anwendung konnektionistischer Systeme; Februar 1992.
- Nr. 10 Becker, J.: Computer Integrated Manufacturing aus Sicht der Betriebswirtschaftslehre und der Wirtschaftsinformatik; April 1992.
- Nr. 11 Kurbel, K., Dornhoff, P.: A System for Case-Based Effort Estimation for Software-Development Projects; Juli 1992.
- Nr. 12 Dornhoff, P.: Aufwandsplanung zur Unterstützung des Managements von Softwareentwicklungsprojekten; August 1992.
- Nr. 13 Eicker, S., Schnieder, T.: Reengineering; August 1992.
- Nr. 14 Erkelenz, F.: KVD2 - Ein integriertes wissensbasiertes Modul zur Bemessung von Krankenhausverweildauern - Problemstellung, Konzeption und Realisierung; Dezember 1992.
- Nr. 15 Horster, B., Schneider, B., Siedentopf, J.: Kriterien zur Auswahl konnektionistischer Verfahren für betriebliche Probleme; März 1993.
- Nr. 16 Jung, R.: Wirtschaftlichkeitsfaktoren beim integrationsorientierten Reengineering: Verteilungsarchitektur und Integrationsschritte aus ökonomischer Sicht; Juli 1993.
- Nr. 17 Miller, C., Weiland, R.: Der Übergang von proprietären zu offenen Systemen aus Sicht der Transaktionskostentheorie; Juli 1993.
- Nr. 18 Becker, J., Rosemann, M.: Design for Logistics - Ein Beispiel für die logistikgerechte Gestaltung des Computer Integrated Manufacturing; Juli 1993.
- Nr. 19 Becker, J., Rosemann, M.: Informationswirtschaftliche Integrationsschwerpunkte innerhalb der logistischen Subsysteme - Ein Beitrag zu einem produktionsübergreifenden Verständnis von CIM; Juli 1993.
- Nr. 20 Becker, J.: Neue Verfahren der entwurfs- und konstruktionsbegleitenden Kalkulation und ihre Grenzen in der praktischen Anwendung; Juli 1993.

- Nr. 21 Becker, K.; Prischmann, M.: VESKONN - Prototypische Umsetzung eines modularen Konzepts zur Konstruktionsunterstützung mit konnektionistischen Methoden; November 1993
- Nr. 22 Schneider, B.: Neuronale Netze für betriebliche Anwendungen: Anwendungspotentiale und existierende Systeme; November 1993.
- Nr. 23 Nietsch, T.; Rautenstrauch, C.; Rehfeldt, M.; Rosemann, M.; Turowski, K.: Ansätze für die Verbesserung von PPS-Systemen durch Fuzzy-Logik; Dezember 1993.
- Nr. 24 Nietsch, M.; Rinschede, M.; Rautenstrauch, C.: Werkzeuggestützte Individualisierung des objektorientierten Leitstands ooL; Dezember 1993.
- Nr. 25 Meckenstock, A.; Unland, R.; Zimmer, D.: Flexible Unterstützung kooperativer Entwurfsumgebungen durch einen Transaktions-Baukasten; Dezember 1993.
- Nr. 26 Grob, H. L.: Computer Assisted Learning (CAL) durch Berechnungsexperimente; Januar 1994.
- Nr. 27 Kirn, St.; Unland, R. (Hrsg.): Tagungsband zum Workshop "Unterstützung Organisatorischer Prozesse durch CSCW". In Kooperation mit GI-Fachausschuß 5.5 "Betriebliche Kommunikations- und Informationssysteme" und Arbeitskreis 5.5.1 "Computer Supported Cooperative Work", Westfälische Wilhelms-Universität Münster, 4.-5. November 1993; März 1994.
- Nr. 28 Kirn, St.; Unland, R.: Zur Verbundintelligenz integrierter Mensch-Computer-Teams: Ein organisationstheoretischer Ansatz; März 1994.
- Nr. 29 Kirn, St.; Unland, R.: Workflow Management mit kooperativen Softwaresystemen: State of the Art und Problemabriß; März 1994.
- Nr. 30 Unland, R.: Optimistic Concurrency Control Revisited; März 1994.
- Nr. 31 Unland, R.: Semantics-Based Locking: From Isolation to Cooperation; März 1994.
- Nr. 32 Meckenstock, A.; Unland, R.; Zimmer, D.: Controlling Cooperation and Recovery in Nested Transactions; März 1994.
- Nr. 33 Kurbel, K.; Schnieder, T.: Integration Issues of Information Engineering Based I-CASE Tools; September 1994.
- Nr. 34 Unland, R.: TOPAZ: A Tool Kit for the Construction of Application Specific Transaction; November 1994.
- Nr. 35 Unland, R.: Organizational Intelligence and Negotiation Based DAI Systems - Theoretical Foundations and Experimental Results; November 1994.
- Nr. 36 Unland, R.; Kirn, St.; Wanka, U.; O'Hare, G.M.P.; Abbas, S.: AEGIS: AGENT ORIENTED ORGANISATIONS; Februar 1995.
- Nr. 37 Jung, R.; Rimpler, A.; Schnieder, T.; Teubner, A.: Eine empirische Untersuchung von Kosteneinflußfaktoren bei integrationsorientierten Reengineering-Projekten; März 1995.
- Nr. 38 Kirn, St.: Organisatorische Flexibilität durch Workflow-Management-Systeme?; Juli 1995.
- Nr. 39 Kirn, St.: Cooperative Knowledge Processing: The Key Technology for Future Organizations; Juli 1995.
- Nr. 40 Kirn, St.: Organisational Intelligence and Distributed AI; Juli 1995.

- Nr. 41 Fischer, K.; Kirn, St.; Weinhard, Ch. (Hrsg.): Organisationsaspekte in Multiagentensystemen; September 1995.
- Nr. 42 Grob, H. L.; Lange, W.: Zum Wandel des Berufsbildes bei Wirtschaftsinformatikern, Eine empirische Analyse auf der Basis von Stellenanzeigen; Oktober 1995.
- Nr. 43 Abu-Alwan, I.; Schlagheck, B.; Unland, R.: Evaluierung des objektorientierten Datenbankmanagementsystems ObjectStore, Dezember 1995.
- Nr. 44 Winter, R., Using Formalized Invariant Properties of an Extended Conceptual Model to Generate Reusable Consistency Control for Information Systems; Dezember 1995.
- Nr. 45 Winter, R., Design and Implementation of Derivation Rules in Information Systems; Februar 1996.
- Nr. 46 Becker, J.: Eine Architektur für Handelsinformationssysteme; März 1996.
- Nr. 47 Becker, J.; Rosemann, M. (Hrsg.): Workflowmanagement - State-of-the-Art aus Sicht von Theorie und Praxis, Proceedings zum Workshop vom 10. April 1996; April 1996.
- Nr. 48 Rosemann, M.; zur Mühlen, M.: Der Lösungsbeitrag von Metadatenmodellen beim Vergleich von Workflowmanagementsystemen; Juni 1996.
- Nr. 49 Rosemann, M.; Denecke, Th.; Püttmann, M.: Konzeption und prototypische Realisierung eines Informationssystems für das Prozeßmonitoring und -controlling; September 1996.
- Nr. 50 v. Uthmann, C.; Turowski, K.; unter Mitarbeit von Rehfeldt, M.; Skall, M.: Workflow-basierte Geschäftsprozeßregelung als Konzept für das Management von Produktentwicklungsprozessen; November 1996.
- Nr. 51 Eicker, S.; Jung, R.; Nietsch, M.; Winter, R.: Entwicklung eines Data Warehouse für das Produktionscontrolling: Konzepte und Erfahrungen; November 1996.
- Nr. 52 Becker, J.; Rosemann, M., Schütte, R. (Hrsg.): Entwicklungsstand und Entwicklungsperspektiven der Referenzmodellierung, Proceedings zur Veranstaltung vom 10. März 1997; März 1997.
- Nr. 53 Loos, P.: Capture More Data Semantic Through The Expanded Entity-Relationship Model (PERM); Februar 1997.
- Nr. 54 Becker, J., Rosemann, M. (Hrsg.): Organisatorische und technische Aspekte beim Einsatz von Workflowmanagementsystemen. Proceedings zur Veranstaltung vom 10. April 1997; April 1997.
- Nr. 55 Holten, R., Knackstedt, R.: Führungsinformationssysteme - Historische Entwicklung und Konzeption; April 1997.
- Nr. 56 Holten, R.: Die drei Dimensionen des Inhaltsaspektes von Führungsinformationssystemen; April 1997.
- Nr. 57 Holten, R., Striemer, R., Weske, M.: Ansätze zur Entwicklung von Workflow-basierten Anwendungssystemen - Eine vergleichende Darstellung; April 1997.
- Nr. 58 Kuchen, H.: Arbeitstagung Programmiersprachen, Tagungsband, Juli 1997.
- Nr. 59 Vering, O.: Berücksichtigung von Unschärfe in betrieblichen Informationssystemen - Einsatzfelder und Nutzenpotentiale am Beispiel der PPS, September 1997.

- Nr. 60 Schwegmann, A., Schlagheck, B.: Integration der Prozeßorientierung in das objektorientierte Paradigma: Klassenzuordnungsansatz vs. Prozeßklassenansatz, Dezember 1997.
- Nr. 61 Speck, M.: In Vorbereitung.
- Nr. 62 Wiese, J.: Ein Entscheidungsmodell für die Auswahl von Standardanwendungssoftware am Beispiel von Warenwirtschaftssystemen, März 1998.
- Nr. 63 Kuchen, H.: Workshop on Functional and Logic Programming, Proceedings, Juni 1998.
- Nr. 64 v. Uthmann, C.; Becker, J.; Brödner, P.; Maucher, I.; Rosemann, M.: PPS meets Workflow. Proceedings zum Workshop vom 9. Juni 1998.
- Nr. 65 Scheer, A.-W.; Rosemann, M.; Schütte, R. (Hrsg.): Integrationsmanagement, Januar 1999.
- Nr. 66 zur Mühlen, M.: Internet - Technologie und Historie, Juni 1999.
- Nr. 67 Holten R.: A Framework for Information Warehouse Development Processes, Mai 1999.
- Nr. 68 Holten R.; Knackstedt, R.: Fachkonzeption von Führungsinformationssystemen - Instanziierung eines FIS-Metamodells am Beispiel eines Einzelhandelsunternehmens, Mai 1999.
- Nr. 69 Holten, R.: Semantische Spezifikation Dispositiver Informationssysteme, Juli 1999.
- Nr. 70 Becker, J.; zur Mühlen, M.; Rosemann, M. (Eds.): Workflow Management 1999. Proceedings of the 1999 Workflow Management Conference: Workflow-based Applications; November 1999.
- Nr. 71 Klein, S.; Schneider, B.; Vossen, G.; Weske, M.; Projektgruppe PESS: Eine XML-basierte Systemarchitektur zur Realisierung flexibler Web-Applikationen, Juli 2000.
- Nr. 72 Klein, S.; Schneider, B. (Hrsg): Negotiations and Interactions in Electronic Markets, Proceedings of the Sixth Research Symposium on Emerging Electronic Markets, Muenster, Germany, September 19 - 21, 1999, August 2000.
- Nr. 73 Becker, J.; Bergerfurth, J.; Hansmann, H.; Neumann, S.; Serries, T.: Methoden zur Einführung Workflow-gestützter Architekturen von PPS-Systemen; November 2000.
- Nr. 74 Terveer, I.: (In Vorbereitung).
- Nr. 75 Becker, J. (Ed.): Research Reports, Proceedings of the University Alliance Executive Directors Workshop – ECIS 2001; Juni 2001.
- Nr. 76 Klein, St.; u.a. (Eds.): MOVE: Eine flexible Architektur zur Unterstützung des Außendienstes mit mobile devices. (In Vorbereitung.)
- Nr. 77 Becker, J.; Knackstedt, R.; Holten, R.; Hansmann, H.; Neumann, St.: Konstruktion von Methodiken: Vorschläge für eine begriffliche Grundlegung und domänenspezifische Anwendungsbeispiele; Juli 2001.
- Nr. 78 Holten, Roland: Konstruktion domänenspezifischer Modellierungstechniken für die Modellierung von Fachkonzepten; August 2001.
- Nr. 79 Vossen, G.; Hüsemann, B.; Lechtenbörger, J.: XLX – Eine Lernplattform für den universitären Übungsbetrieb; August 2001.

- Nr. 80 Becker, J.; Knackstedt, R.; Series, Th.: Gestaltung von Führungsinformationssystemen mittels Informationsportalen; Ansätze zur Integration von Data-Warehouse- und Content-Management-Systemen, November 2001.