

September 2003

# Ein integrativer Ansatz für unternehmensweite Wissensportale

Torsten Priebe

*Universität Regensburg*, torsten.priebe@wiwi.uni-regensburg.de

Günther Pernul

*Universität Regensburg*

Peter Krause

*EMPRISE Consulting Düsseldorf GmbH*

Follow this and additional works at: <http://aisel.aisnet.org/wi2003>

---

## Recommended Citation

Priebe, Torsten; Pernul, Günther; and Krause, Peter, "Ein integrativer Ansatz für unternehmensweite Wissensportale" (2003).

*Wirtschaftsinformatik Proceedings 2003*. 67.

<http://aisel.aisnet.org/wi2003/67>

This material is brought to you by the Wirtschaftsinformatik at AIS Electronic Library (AISeL). It has been accepted for inclusion in Wirtschaftsinformatik Proceedings 2003 by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

In: Uhr, Wolfgang, Esswein, Werner & Schoop, Eric (Hg.) 2003. *Wirtschaftsinformatik 2003: Medien - Märkte - Mobilität*, 2 Bde. Heidelberg: Physica-Verlag

ISBN: 3-7908-0111-9 (Band 1)

ISBN: 3-7908-0116-X (Band 2)

© Physica-Verlag Heidelberg 2003

# Ein integrativer Ansatz für unternehmensweite Wissensportale

**Torsten Priebe, Günther Pernul**

Universität Regensburg

**Peter Krause**

EMPRISE Consulting Düsseldorf GmbH

*Zusammenfassung: Wissensportale leisten einen wichtigen Beitrag zum Wissensmanagement eines Unternehmens, indem sie es Benutzern ermöglichen, direkt mit Geschäftsprozessen und Wissen zu arbeiten. Dabei bieten Wissensportale durch eine konsolidierte, personalisierte Benutzeroberfläche einen effizienten Zugriff auf die immer unüberschaubarer werdende Flut an Informationen, die die Wissensbasis des Unternehmens bilden. Heutige Portalsysteme beschränken sich jedoch auf die gemeinsame Darstellung unterschiedlicher Wissensressourcen (Data Warehouse, Dokumentenmanagementsystem, Internet, etc.) in Form von sog. Portlets auf einer Portalseite. Eine Interaktion der Portlets untereinander erfolgt nicht. Navigiert ein Benutzer innerhalb eines Portlets, so bleiben die anderen statisch, was bedeutet, dass jede Wissensressource getrennt nach relevanten Informationen durchsucht werden muss. Dieser Beitrag stellt einen Integrationsansatz vor, der im Sinne einer effizienten Wissensnutzung den durch Navigation in einem Portlet offenbarten Informationsbedarf eines Benutzers auch in anderen Portlets zum Auffinden passender Informationen heranzieht. Zur Evaluation des Ansatzes ist ein prototypisches Portalsystem in Entwicklung, welches sich auf die Integration von OLAP und Information Retrieval konzentriert.*

*Schlüsselworte: Wissensmanagement, Portale, Integration, OLAP, Information Retrieval, Semantic Web*

## 1 Einleitung

Unternehmen sind heute und zukünftig Herausforderungen ausgesetzt (z.B. zunehmender internationaler Konkurrenz, Kostendruck und Angleichung der Produkte), denen durch traditionelle Strategien und Organisationsformen nicht mehr in ausreichendem Maße begegnet werden kann. Diese Herausforderungen machen eine konsequente Einbindung des Faktors Wissen und die Etablierung einer Lernkultur (lernende Organisation) notwendig.

Das Konzept der lernenden Organisation kann jedoch im Zielkonflikt zu den verbreiteten erfolgskonzentrierten Managementansätzen stehen, denen eine Optimierung der Geschäftsprozesse entlang der Wertschöpfungskette zugrunde liegt. Aufgrund solcher Restrukturierungsprozesse werden unter Umständen Maßnahmen beschlossen, die einer lernenden Organisation entgegenwirken. Hierzu zählen z.B. das Auseinanderreißen von eingespielten Teams oder Mitarbeiterentlassungen. Dennoch steht die Geschäftsprozessorientierung in engem Zusammenhang mit der Ressource Wissen. Wissensstruktur wird mit den Geschäftsprozessen sichtbar, da sich Wissensflüsse primär um diese anordnen [Lehn00].

Die Informations- und Kommunikationstechnologie kann dabei durch eine integrierte Systemarchitektur dazu beitragen, sowohl die konsequente Nutzung und Entwicklung der Ressource Wissen als auch die Geschäftsprozessorientierung zu adressieren. Unternehmensweite Wissensportale (sog. Enterprise Knowledge Portals) bauen dabei auf dieser Architektur auf und ermöglichen es Benutzern, über eine konsolidierte, personalisierte Benutzeroberfläche direkt mit Geschäftsprozessen und Wissen zu arbeiten. Dabei ermöglichen Wissensportale einen effizienten Zugriff auf die immer unüberschaubarer werdende Flut an Informationen, die die Wissensbasis des Unternehmens bilden.

Abschnitt 2 dieses Beitrages beschäftigt sich mit einem Überblick über den Aufbau solcher Wissensportale. Wissensportale stellen ein Integrationskonzept für unterschiedliche Wissensressourcen im Unternehmen dar. Wichtige Ressourcen sind hierbei üblicherweise Anwendungs- und Legacy-Systeme (Geschäftsprozesse, Transaktionen), strukturierte Data-Warehouse-Daten (Business Intelligence, Berichte, Kennzahlen), sowie unstrukturierte Daten in Form von Dokumenten. Externe Ressourcen können durch Einbindung von Web-Dokumenten und Web Services berücksichtigt werden (vgl. [Popp02]).

Mittlerweile existieren eine Reihe von kommerziellen Produkten als Plattformen zum Aufbau von Wissensportalen. Eine wahre Integration der genannten Ressourcen bieten heutige Systeme jedoch nicht. Es wird zwar eine konsolidierte Benutzerschnittstelle zum Zugriff auf die unterschiedlichen Wissensquellen zur Verfügung gestellt, die einzelnen Komponenten sind dabei jedoch unabhängige, nicht miteinander kommunizierende Einzelteile. Dieses Manko wird in Abschnitt 3 aufgegriffen und ein Lösungsansatz zur Integration von Portalkomponenten durch Kommunikation des Benutzerkontexts erarbeitet. Kernidee ist es, das durch Navigation in einer Portalkomponente offenbarte Informationsbedürfnis eines Benutzers auch in anderen (andere Quellen repräsentierenden) Komponenten zur Suche nach passenden Informationen heranzuziehen.

Wir arbeiten an einem prototypischen Wissensportalsystem zur Evaluation des in Abschnitt 3 dargestellten Ansatzes. Hierbei konzentrieren wir uns auf die Integration von Portalkomponenten zum OLAP-Zugriff auf ein Data Warehouse und zur Suche nach Dokumenten (Information Retrieval). Der Prototyp wird in Abschnitt 4 vorgestellt.

In Abschnitt 5 präsentieren wir verwandte Arbeiten und ordnen unseren Ansatz in diese Forschungslandschaft ein. Abschnitt 6 schließt den Beitrag mit einer Zusammenfassung und einem Ausblick.

## 2 Unternehmensweite Wissensportale

### 2.1 Begriffsbestimmung

Der Begriff des Portals stammt aus dem Lateinischen und bedeutet Tür bzw. Tor. In Anlehnung daran wird der Begriff des (Enterprise) Portals auch für Internet- und Intranet-Einstiegsseiten verwendet. Abbildung 1 zeigt eine Klassifizierung der unterschiedlichen Ausprägungen von Enterprise-Portalen.

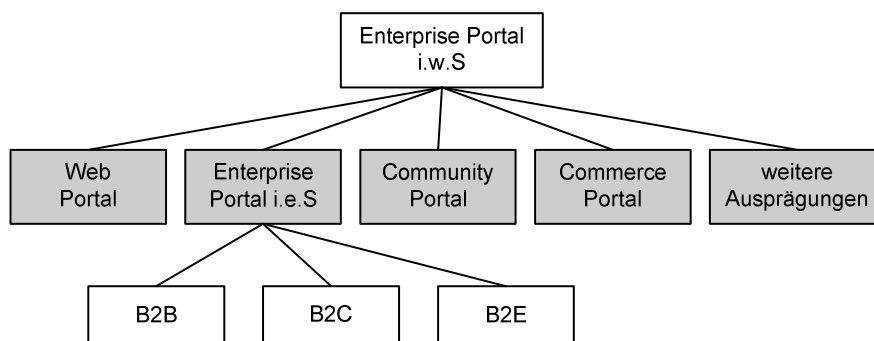


Abbildung 1: Klassifizierung von Enterprise-Portalen

Als Web-Portale werden Dienste bezeichnet, die Internet-Benutzern eine katalogisierte Sicht auf das gesamte Internetangebot bzw. ihren indizierten Bestand an Internetseiten für eine Suche anbieten<sup>1</sup>. Community-Portale sind zentraler Treffpunkt beliebiger virtueller Gemeinschaften zu diversen Themen. Meist werden verschiedene Dienste zur Kommunikation und Gemeinschaftsbildung angeboten. Hierunter fällt z.B. das Portal eines Hardware-Herstellers, der für den Erfahrungsaustausch seiner Kunden entsprechende Diskussionsforen anbietet. Unter Commerce-Portalen werden virtuelle Marktplätze verstanden, auf denen ein oder mehrere Unternehmen ihre Produkte anbieten. Die zahlreichen Web-Shops verschiedener Anbieter fallen in diese Kategorie. Weitere Einteilungen wie z.B. „Health-Care-Portale“ sowie Mischungen der genannten Kategorien sind ebenfalls möglich. Darunter fallen beispielsweise Dienste von MSN oder AOL. Diese Portale werden auch Mega-Portale genannt, die Übergänge sind fließend.

<sup>1</sup> Dies wird häufig auch gleichgesetzt mit Suchmaschinen, wie z.B. Yahoo oder Google.

In diesem Beitrag beschäftigen wir uns mit Enterprise-Portalen im engeren Sinne (auch Unternehmensportale). Der Fokus liegt auf Unternehmensinformationen und Dienstleistungen, Anwendungen und Funktionen, die ein Unternehmen einer bestimmten Zielgruppe anbietet. Wir konzentrieren uns dabei auf die Zielgruppe der Mitarbeiter, also sog. B2E-Portale („Business to Employee“). Die Begriffe Enterprise Portal und Enterprise Information Portal werden dabei synonym verwendet. Hierunter wird eine unternehmensweite und personalisierte Integrationsplattform verstanden, mittels derer ein individueller Zugriff auf diverse Informationen und Applikationen realisiert wird. Das Ziel ist es, dem Benutzer nur noch eine Oberfläche zur Verfügung zu stellen, mit der er seine täglichen Aufgaben erledigen kann, und mit welcher eine transparente Wissensnutzung ermöglicht wird.

Die Abgrenzung zwischen dem Begriff des Enterprise Information Portal und dem des Enterprise Knowledge Portal (unternehmensweites Wissensportal) ist unscharf. Wissensportale setzen den Schwerpunkt eher auf die Zielorientierung im Hinblick auf Wissenserwerb, Wissensentwicklung, Wissensnutzung, und Wissensverteilung. Da wir durch den in diesem Beitrag vorgestellten Integrationsansatz eben diese Wissensprozesse verbessern wollen, verwenden wir hier auch den Begriff des Wissensportals.

## 2.2 Stand der Technik

Wie bereits in der Einleitung dargelegt, existieren eine Reihe von kommerziellen Produkten (z.B. IBM, BEA, Plumtree, etc.) als Portalplattformen. Das Prinzip dieser Plattformen ist es, verschiedene Datenquellen auf einer Portal-Webseite zu vereinen. Technologien wie die J2EE Portlet API<sup>2</sup> werden zur Verbindung der Komponenten verwendet [Wege02]. Die einzelnen Portalkomponenten werden dabei Portlets genannt. Abbildung 2 zeigt eine typische Portalseite mit mehreren Portlets (durch Einrahmung hervorgehoben).

---

<sup>2</sup> Siehe auch das Apache Jetspeed Portal Framework unter <http://jakarta.apache.org/jetspeed/>

The screenshot shows a personalized web portal titled "My Front Page" for Monday, Jan 13. The interface includes several portlets:

- Message Center:** Includes links for "Check Email" and "Check Calendar".
- My Front Page Headlines:** Features a "Check out the new Tech Section at Yahoo! News" link and "Top Stories from Reuters" with headlines like "US Still Willing for North Korea Talks, Energy Aid".
- Weather:** Displays weather for various cities including Hong Kong, London, UK, Los Angeles, CA, New York, NY, Paris, Rome, Sydney, and Tokyo.
- Portfolios:** Shows "My First Portfolio" with stock tickers (DJI, NASDAQ, NIKKEI, DAX, YHOO, MSTR) and their respective values and changes.
- Scoreboard:** Displays sports scores for "TODAY" (NBA: New York vs Chicago) and "YESTERDAY" (NBA: Miami vs LA Lakers).
- Lead Photo:** Features a photo of James Kelly and a headline: "Envoy: N. Korea Could Get U.S. Energy Aid".
- Best Fares:** Lists travel options from New York, NY to Paris, Seattle/Tacoma, WA, and San Francisco, CA.

Abbildung 2: Portalseite mit Portlets

Übliche Portlets in unternehmensweiten Wissensportalen dienen, wie bereits erwähnt, dem Zugriff auf Transaktionen von Anwendungssystemen oder Web Services, OLAP-basiertem Zugriff auf ein Data Warehouse und der Suche nach Intranet- oder Web-Dokumenten. Weitere mögliche Portlets widmen sich dem Zugriff auf Kartenmaterial, wie es von GIS-Systemen bereitgestellt wird, oder anderen entscheidungsunterstützenden (z.B. wissensbasierten) Systemen. Viele Anbieter, beispielsweise von OLAP-Systemen, bieten bereits Portlet-Implementierungen zu verbreiteten Portalplattformen an.

Das Problem dieser existierenden Standardlösungen ist jedoch die fehlende Interaktion der Portlets untereinander. Navigiert ein Benutzer innerhalb eines Portlets, so bleiben die anderen statisch. Da verschiedene Portlets den Zugriff auf verschiedene Wissensressourcen (bzw. -quellen) repräsentieren, bedeutet dies, dass jede dieser Quellen getrennt nach relevanten Informationen durchsucht werden muss.

### 3 Integration durch Kommunikation des Benutzerkontexts

Im Sinne einer effizienten Wissensnutzung wäre es jedoch wünschenswert, wenn der in einem Portlet durch (wie auch immer gearbete) Navigation offenbarte In-

formationsbedarf des Benutzers auch in anderen Portlets zum Auffinden passender Informationen genutzt werden könnte.

Klassisches Beispiel ist ein Finanzportal, bei dem ein Benutzer, der Aktienkurse eines bestimmten Unternehmens betrachtet, automatisch auch Zugriff auf dieses Unternehmen betreffende Nachrichten erhält, die sich beispielsweise bei Presseagenturen wie Reuters finden. Das für die Darstellung der Aktienkurse zuständige Portlet teilt dem News Portlet mit, zu welchem Unternehmen es nach Nachrichten suchen soll. Ein anderes Beispiel ist ein Web-Portal wie Yahoo, das bei Suche nach einer bestimmten Stadt gleichzeitig dort ansässige Restaurants oder das dortige Kinoprogramm anbietet.

Diese genannten Lösungen sind jedoch „fest verdrahtete“ Spezialanwendungen. Dieser Beitrag erarbeitet eine generische Lösung, die in Standard-Portalplattformen integriert werden kann. Ziel ist ein Architekturrahmen, der durch Kommunikation des aktuellen Benutzerkontexts über Portlet-Grenzen hinweg, eine entsprechende Integration ermöglicht.

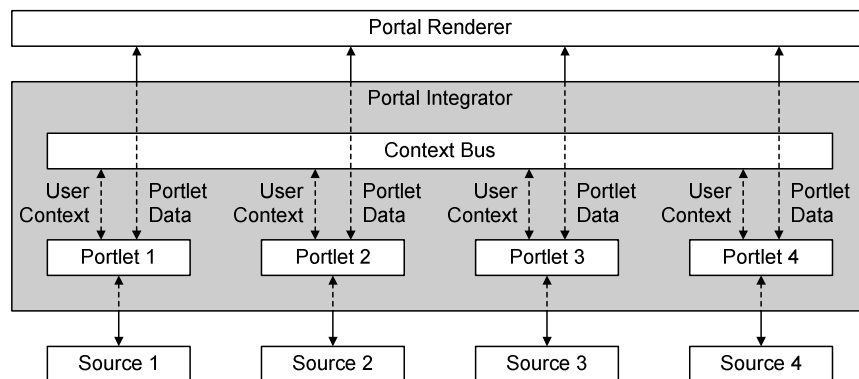


Abbildung 3: Architekturrahmen für kontextbasierte Portlet-Integration

Abbildung 3 zeigt unseren Architekturrahmen zur kontextbasierten Portlet-Integration. Während Portlets in herkömmlichen Portalsystemen nur Ihre Portlet-Daten zur grafischen Darstellung an die Portalplattform weitergeben, sieht unser Ansatz bei Navigation (sprich Veränderung des Kontexts) innerhalb eines Portlets eine Nachricht an einen sog. Context Bus vor. Andere Portlets können diese Nachricht abgreifen und durch Anpassung der dargestellten Informationen reagieren.

Als problematisch stellt sich, besonders im Hinblick auf die angestrebte generische Lösung, die Heterogenität der unterschiedlichen Informationsquellen und damit der Portlets dar. Ein OLAP-Portlet formuliert den Benutzerkontext mit Hilfe von Konstrukten eines OLAP-Datenmodells (siehe nächster Abschnitt). Hiermit wird ein Portlet zum Zugriff auf eine operative Anwendungskomponente wenig anfangen können. Dort wird eher das Datenmodell einer operativen Datenbank



oder ein Anwendungsobjektmodell zur Kontextbeschreibung zum Einsatz kommen.

Unser Ansatz sieht die Verwendung einer Ontologie als „gemeinsame Sprache“ zum Übersetzen der Konstrukte verschiedener Datenmodelle vor. Dies ist schematisch in Abbildung 4 dargestellt.

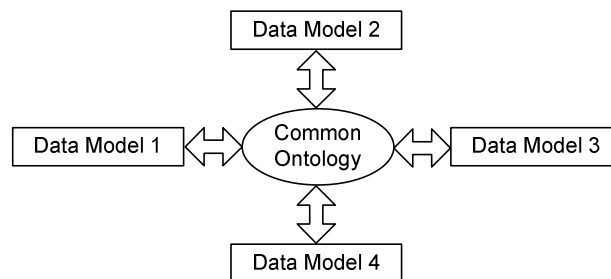


Abbildung 4: Ontologie zur Übersetzung von Konstrukten verschiedener Datenmodelle

## 4 Prototyp: Integration von OLAP und Information Retrieval

Zur Evaluation unseres Ansatzes arbeiten wir an der Entwicklung eines prototypischen Wissensportalsystems, die Architektur ist in Abbildung 5 dargestellt. Zum gegenwärtigen Zeitpunkt konzentrieren wir uns dabei auf ein OLAP-System zum Zugriff auf Data-Warehouse-Daten und auf ein Information-Retrieval-System. Es entsteht ein Portlet zur OLAP-Navigation durch Berichte und Kennzahlen, sowie ein Portlet zur Suche nach in einem Dokumentenmanagementsystem gespeicherten Dokumenten.

Zum einen werden (vordefinierte) Standard-OLAP-Berichte als „Dokumente“ berücksichtigt und vom Information-Retrieval-Portlet als solche in die Suche mit einbezogen. Das OLAP-Portlet kann dabei (nach dem „Best-Fit-Prinzip“) den zu einer Suchanfrage jeweils relevantesten Bericht unaufgefordert darstellen. Dadurch ist eine Kontextintegration von Information Retrieval nach OLAP gewährleistet. Weitaus interessanter ist jedoch die Gegenrichtung: Wenn ein Benutzer durch Drilling oder Slicing/Dicing in einem OLAP-Bericht navigiert, kann die IR-Komponente den Abfragekontext des Benutzers dazu heranziehen, automatisch auch relevante Dokumente anzubieten.

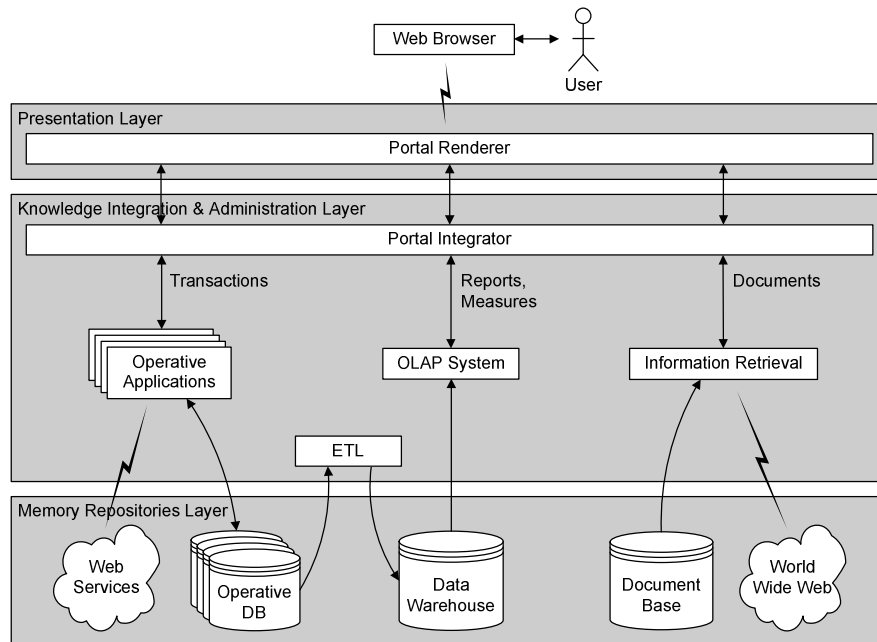


Abbildung 5: Architektur des prototypischen Wissensportalsystems

#### 4.1 Der OLAP-Abfragekontext

OLAP ist eine interaktive Analysetechnik zur Abfrage strukturierter Daten (z.B. Umsatzzahlen), die meist in einem Data Warehouse gespeichert sind [ChDa97, BaGü01]. OLAP basiert auf dem multidimensionalen Datenmodell, d.h. numerische Kennzahlen (z.B. Umsatz) sind nach hierarchischen Dimensionen (Zeit, Geographie, Produkte, etc.) organisiert. Abfrageergebnisse werden als interaktive Tabellen (Berichte) dargestellt. Die Benutzernavigation erfolgt durch Drilling (Veränderung des Detaillierungsgrades, z.B. von Jahr auf Monat) und Slicing/Dicing (Auswahl nur eines Teils der verfügbaren Daten, z.B. nur ein bestimmtes Produkt betreffend).

In unserem Prototyp gehen wir von einem Unternehmen aus, welches diverse Produkte aus dem Consumer-Bereich (Elektronik, Lebensmittel, Geschenkartikel, etc.) über Call Center vertreibt. Eine Beispiel-OLAP-Abfrage ist in Abbildung 6 dargestellt, sie zeigt die Umsätze für Electronics-Produkte im Jahr 1998.

Subcategory	Dollar Sales			
	Q1 1998	Q2 1998	Q3 1998	Q4 1998
Audio	\$ 801,00	\$ 457,00	\$ 85,00	\$ 372,00
Comfort	\$ 10.461,00	\$ 1.794,00	\$ 3.385,00	\$ 6.995,00
Gadgets	\$ 2.508,00	\$ 726,00	\$ 756,00	\$ 959,00

Abbildung 6: Beispielhafte OLAP-Abfrage

Ein Benutzer navigiert durch OLAP-Berichte durch Drilling und Slicing/Dicing. Der Abfragekontext (d.h. die im OLAP-Bericht dargestellten Daten) kann durch drei Komponenten beschrieben werden:

- die im Bericht dargestellten Kennzahlen,
- die auf den Zeilen und Spalten des Berichts dargestellten Dimensionen (und Hierarchieebenen),
- sowie die zur Beschränkung der dargestellten Daten verwendeten Selektionskriterien (d.h. das Slicing/Dicing).

Bei dem für unseren Prototyp verwendeten OLAP-Produkt MicroStrategy<sup>3</sup> spricht man bei den ersten beiden Komponenten auch vom Report Template (da sie eine Art Formatvorlage für den Bericht darstellen) und bei der letzten Komponente von Filter (da sie einen Teil der verfügbaren Daten herausfiltert). Die Beispielabfrage in Abbildung 6 führt zu folgendem Abfragekontext:

- Die abgefragte Kennzahl ist Dollar Sales.
- Die dargestellten Dimensions-Hierarchieebenen sind Subcategory und Quarter.
- Der Benutzer hat nur das Jahr 1998 und nur Produkte der Electronics-Kategorie gewählt.

## 4.2 Information Retrieval und das Semantic Web

Information Retrieval beschäftigt sich mit der Suche nach relevanten Dokumenten in Bezug auf das Informationsbedürfnis eines Benutzers. Standardtechniken sind schlüsselwortbasierte Verfahren (z.B. Boolesches Retrieval) [BaRi99]. Eine Hauptanwendung für Information Retrieval sind Suchmaschinen, wie sie im Web verwendet werden. Schlüsselwort-basierte Suchen haben jedoch ihre Probleme und Grenzen. Jedermann kennt die Qual des Suchens nach relevanten Webdokumenten nur anhand von Schlüsselworten. Daher werden in kontrollierten Umge-

<sup>3</sup> Siehe unter <http://www.microstrategy.com>

bungen (z.B. betrieblichen Intranets oder Dokumentenmanagementsystemen) Dokumente zusätzlich mit Metadaten versehen.

Beispielsweise verwaltet das in unserem Prototyp betrachtete Unternehmen Produktbeschreibungen. Jedes Dokument ist von einem bestimmten Mitarbeiter geschrieben und es behandelt ein bestimmtes Produkt, welches durch eine Item ID identifiziert wird. Auch der Dokumenttyp (die Tatsache, dass es sich um eine Produktbeschreibung handelt) ist Teil der Metadaten.

Mit Tim Berners-Lees Vision des Semantic Web [BeHL01] sind Metadaten auch im Web-Umfeld in das Zentrum der Betrachtung gerückt. Das Resource Description Framework (RDF) [W3C99] definiert eine standardisierte, XML-basierte Form, Metadaten zu repräsentieren. RDF erweitert also das nur auf syntaktischer Ebene definierte XML um formale Semantik. Dazu verwendet es ein einfaches Tripel-basiertes Modell (bestehend aus Ressourcen, Properties und Werten). Graphisch werden RDF-Metadaten als gerichteter Graph dargestellt: Ressourcen als Ovale, Properties als Pfeile und (literale) Werte als Rechteck.

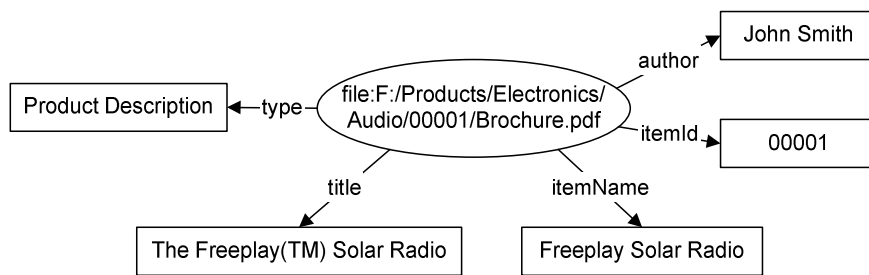


Abbildung 7: Einfache RDF-Metadaten für ein Produktbeschreibungsdokument

In unserem Prototyp verwenden wir RDF-basierte Metadaten für die Verwaltung der Dokumente. Die Metadaten eines beispielhaften Produktbeschreibungsdokumentes sind (vereinfacht) in Abbildung 7 dargestellt. Wie bereits erwähnt, werden RDF-Metadaten mit Hilfe von XML in maschinenlesbarer Form repräsentiert. Die XML-Repräsentation der Metadaten aus Abbildung 7 sehen folgendermaßen aus:

```
<?xml version="1.0">
  <r:RDF xmlns:r=
    "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:s="http://www.mycompany.com/portal/schema#">
  <r:Description r:about="file:F:/Products/Electronics/
    Audio/12345/Brochure.pdf">
    <r:type resource="http://www.mycompany.com/portal/
      schema#ProductDescription"/>
    <s:author>John Smith</s:author>
    <s:title>The Freeplay(TM) Solar Radio</s:title>
```

```
<s:itemID>12345</s:itemID>
<s:itemName>Freeplay Solar Radio</s:itemName>
</r:Description>
</r:RDF>
```

Solche Metadaten können entweder in den Dokumenten selbst (als Meta-Tags) oder in separaten XML-Dokumenten abgelegt werden. Ein anderer – auch von uns verfolgter – Ansatz ist die Verwendung von XML-Datenbanken (vgl. [Thur02]) oder speziellen RDF Repositories, wie Sesame<sup>4</sup>. Es fällt auf, dass anstelle eines Stringliterals „Product Description“ als Dokumententyp ein Verweis auf die RDF-Schema-Klasse mit der URI „[http://www.mycompany.com/portal/schema# ProductDescription](http://www.mycompany.com/portal/schema#ProductDescription)“ aufgeführt ist. Ein RDF Schema beschreibt die unterschiedlichen Metadatenkonzepte (in diesem Fall author, title, itemID und itemName) und ihre Abhängigkeiten. Es stellt somit das „Datenmodell“ der Dokumentenmetadaten dar.

RDF ist selbstverständlich in der Lage auch komplexere Metadaten als im obigen Beispiel darzustellen. So können durch das RDF-Konstrukt „Bag“ mehrwertige Metadatenelemente angegeben werden (z.B. wenn sich ein Dokument auf mehrere Produkte bezieht). Desweiteren können die Werte von RDF Properties einer Ressource neben Literalen auch selbst wiederum Ressourcen sein. Dadurch lassen sich Produkte oder Mitarbeiter als „Objekte“ mit eigenen Metadaten erfassen. Diese können dann von Dokumenten als Thema oder Autor referenziert werden.

Die Suche nach RDF-basierten Metadaten (und damit nach Dokumenten) kann mit Hilfe von Standard XML-Technologien (beispielsweise XQuery) oder speziellen RDF-Abfragesprachen (z.B. RQL oder RDQL [KCPA01, MiSR02]) erfolgen. [ShFM02] schlagen außerdem eine Kombination von Metadaten- und Schlüsselwort basierten Verfahren zum Information Retrieval im Semantic Web vor.

### 4.3 Inferenzregeln

Abbildung 5 zeigt die Architektur unseres prototypischen Portalsystems. Über das Portal-Integrator-Modul wird der Abfragekontext des OLAP-Systems an die Information-Retrieval-Komponente weitergegeben. Auf Seite der Dokumente verwenden wir RDF-basierte Metadaten. Die Metadatenelemente sind in einem RDF Schema definiert. Das Mapping zwischen den Konstrukten des OLAP-Datenmodells und des RDF-Schemas leistet dabei eine (momentan ebenfalls RDF(S) basierte<sup>5</sup>) Ontologie (siehe Abbildung 8).

<sup>4</sup> Siehe unter <http://sesame.aidadministrator.nl>

<sup>5</sup> Aufgrund der größeren Sprachmächtigkeit (z.B. zum Ausdrücken von Synonymbeziehungen) bieten sich jedoch Erweiterungen wie DAML+OIL (siehe unter <http://www.daml.org>) an.

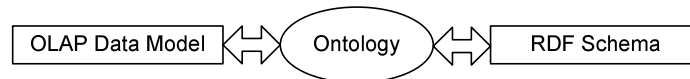


Abbildung 8: Mapping zwischen OLAP-Datenmodell- und RDF-Schema-Konstrukten

Momentan arbeiten wir an der Entwicklung von Heuristiken, wie sich OLAP-seitig gewählte Kennzahlen, Hierarchieebenen und Slicing/Dicing-Filter auf IR-Abfragen und Relevanzmaße abbilden lassen.

Item	Metriken	Dollar Sales		
		Jan 98	Feb 98	Mar 98
Freeplay Solar Radio		\$ 85,00	\$ 425,00	\$ 170,00
Micro FM Radio		\$ 19,00	\$ 19,00	\$ 19,00
Shower Companion			\$ 64,00	

Abbildung 9: Beispielhafte OLAP-Abfrage

Betrachten wir beispielsweise die in Abbildung 9 dargestellte OLAP-Abfrage, in diesem Fall betrachtet der Benutzer Umsatzzahlen für Audio-Produkte im ersten Quartal 1998. Aus dem Abfragekontext lässt sich schließen, dass sich der Benutzer für Produktbeschreibungen aller Audio-Produkte interessieren könnte. Er könnte sich ebenfalls für andere diese Produkte betreffende Dokumente (z.B. interne Nachrichten oder Presseerklärungen) interessieren, besonders solche des ersten Quartals 1998. Dokumente die die Audio-Subkategorie allgemein betreffen könnten ebenfalls relevant sein (sollten aber auf einer Ergebnisliste weiter unten erscheinen, da nicht im analytischen Fokus).

Angenommen, der Benutzer nimmt nun ein Roll-up auf die Ebene Subcategory vor und ändert seine Selektion auf alle Electronics-Produkte. Weiterhin ändert er den Zeitbezug auf ganz 1998 (das entspricht der in Abbildung 6 dargestellten OLAP-Abfrage). Welche Auswirkung hätte das auf die Relevanz von Dokumenten? Zum einen wären nun die Produktbeschreibungen aller Electronics-Produkte relevant. Jedoch wären sie jetzt vermutlich weniger wichtig, als Dokumente über die Subkategorien (da der analytische Fokus nicht mehr auf einzelnen Produkten liegt).

## 5 Verwandte Arbeiten

Die Integration unterschiedlicher Wissensressourcen wird intensiv im Wissensmanagement- und Organizational-Memory-Umfeld untersucht. Autoren wie [Lehn00] schlagen Managementansätze und Systemarchitekturen vor, meist ohne

jedoch näher auf (technische) Details einzugehen. Wissensportalsysteme (z.B. [Popp02]) greifen diese Lücke auf, heutige Systeme leisten jedoch keine kontextbasierte Integration der einzelnen Portalkomponenten, wie sie von uns betrachtet wird.

Im EU-geförderten Projekt GOAL<sup>6</sup> [KoMM00], an dem auch unsere Arbeitsgruppe beteiligt war, wurde die Integration von Data-Warehouse- (bzw. OLAP-) Technologien und geographischen Informationssystemen (GIS) beleuchtet. Die Grundidee hinter einer solchen Integration ist die Abbildung einer geographischen OLAP-Dimension auf Objekte (z.B. Städte, Regionen, Länder) eines GIS, so dass Karten zur OLAP-Navigation verwendet werden können. Im Prinzip lässt sich der in Abschnitt 3 vorgestellte Ansatz als Verallgemeinerung dieser Grundidee verstehen.

Ähnlich wie wir betrachten [RiKM00, MeRi01] die Verbindung unstrukturierter (bzw. semistrukturierter) Dokumente mit strukturierten OLAP-Daten. Der dort verfolgte Weg unterscheidet sich jedoch in einem Punkt gravierend von unserem. Während in [RiKM00] OLAP-Würfel als in einer Art Digital Library gespeicherte „Dokumente“ betrachtet werden, die über Metadaten mit anderen Dokumenten verlinkt sind, gehen wir einen Schritt weiter. Wir betrachten OLAP-Würfel nicht als „Black Box“, sondern nutzen auch die Benutzernavigation innerhalb der Würfel (den Abfragekontext) als Grundlage zum Auffinden relevanter Dokumente.

Solch ein kontextbasiertes Information Retrieval wird auch von [HeMo02] untersucht. Die Autoren schlagen vor, in Client-Anwendungen sog. Plug-ins zu integrieren, die den Arbeitskontext des Benutzers an ein Information-Retrieval-System melden. Dieses kann dann mit Hilfe dieser Kontextinformationen potentiell relevante Dokumente anbieten. Solch eine generische Architektur ist sicher wünschenswert, viele Anwendungsprogramme werden jedoch eine Verwendung von Plug-ins nicht ermöglichen. Daher konzentrieren wir uns auf die (kontrollierbare) Umgebung innerhalb eines Wissensportals. Weiterhin scheint die Interpretation des Benutzer-Arbeitskontexts und dessen Übersetzung in eine IR-Abfrage in einer völlig allgemeingültigen Form nicht unproblematisch. Aus diesem Grund beschränken uns in unserem in Abschnitt 4 vorgestellten Prototyp zunächst auf den (auch von [HeMo02] als vielversprechend angesehenen) Abfragekontext eines OLAP-Systems.

## 6 Zusammenfassung und Ausblick

In diesem Beitrag haben wir einen Integrationsansatz für unternehmensweite Wissensportale vorgestellt. Kernidee ist die Kommunikation des Benutzerkontexts ü-

---

<sup>6</sup> Geographic Information Online Analysis, INCO COPERNICUS Projekt-Nr. 977071

ber Portlet-Grenzen hinweg um passende Informationen aus unterschiedlichen Quellen zusammenzuführen. Gerade bei der immer größer werdenden Flut an Informationen, gespeichert in unterschiedlichsten Systemen mit unterschiedlichsten Schnittstellen, wird ein effizienter integrierter Zugriff immer wichtiger.

Der in Abschnitt 3 dargestellte allgemeine Ansatz konkretisiert sich in dem als „Proof of Concept“ in Entwicklung befindlichen, in Abschnitt 4 vorgestellten, prototypischen Portalsystem, welches die Integration von OLAP und Information Retrieval als Teilproblem herausgreift. Momentan arbeiten wir, wie erwähnt, an der Entwicklung von Regeln zur Ableitung von IR-Abfragen und Relevanzmaßen aus einem OLAP-Abfragekontext. Die ersten Ergebnisse sind vielversprechend.

Allerdings haben sich dabei die üblichen Suchverfahren für RDF-Metadaten (beispielsweise RQL-Anfragen) aufgrund ihrer scharfen Trennung zwischen Treffer und Nicht-Treffer als eher ungeeignet für das Information Retrieval herausgestellt. Dies trifft sowohl auf Benutzeranfragen, insbesondere jedoch auf Kontext-generierte automatische Anfragen zu. Benötigt wird ein (tolleranteres) unscharfes Suchverfahren. Erste Lösungsansätze hierzu liegen bereits vor.

Neben einer Integration weiter Datenquellen, beispielsweise operativer (Legacy-) Systeme, ist eine Verbindung mit dem in Portalsystemen üblichen Konzept der Personalisierung interessant und Bestandteil zukünftiger Arbeiten. Schließlich ist die Rolle (bzw. andere Attribute) des Benutzers ebenfalls Bestandteil des Benutzerkontexts.

Ein weiterer offener Punkt ist Sicherheit. In früheren Forschungsarbeiten (beispielsweise im erwähnten GOAL-Projekt) haben wir uns mit Zugriffskontrollen in Data-Warehouse- und OLAP-Systemen beschäftigt. Die meisten kommerziellen OLAP-Systeme verfügen heutzutage über recht ausgereifte Sicherheitsmechanismen. Auf der Gegenseite gibt es jedoch noch eine Menge ungelöster Probleme bei der Beschränkung des Zugriffs auf unstrukturierte (textuelle) Daten. Ein weiteres Forschungsprojekt unserer Arbeitsgruppe beschäftigt sich mit einer auf Benutzer-Credentials und Dokumentenmetadaten basierenden Zugriffskontrolle. Eine Einbindung dieses Sicherheitsmechanismus in das angesprochene Portalsystem ist geplant. Dabei ist die Integration unterschiedlicher Sicherheitspolitiken der diversen Systemkomponenten (beispielsweise des OLAP-Systems und des Dokumentenmanagementsystems) eine nicht triviale Herausforderung.

## Literatur

- [BaGü01] Bauer, A.; Günzel, H. (Hrsg.): Data-Warehouse-Systeme: Architektur, Entwicklung, Anwendung. Dpunkt-Verlag, 2001.
- [BeHL01] Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. In: Scientific American, Mai 2001, S. 35-43.



- [BaRi99] Baeza-Yates, R.; Ribeiro-Neto, B. (Hrsg.): *Modern Information Retrieval*. Addison Wesley Longman Limited, 1999.
- [ChDa97] Chaudhuri, S.; Dayal, U.: An Overview of Data Warehousing and OLAP Technology. *SIGMOD Record* 26(1), März 1997, S. 65-74.
- [HeMo02] Henrich, A.; Morgenroth, K.: Integration von kontextunterstütztem Information Retrieval in Portalsysteme. Teilkonferenz Management der Mitarbeiter-Expertise in IT-Beratungsunternehmen (MKWI 2002), Nürnberg, 2002.
- [KCPA01] Karvounarakis, G.; Christophides, V.; Plexousakis, D.; Alexaki, S.: Querying RDF Descriptions for Community Web Portals. In: Proc. French National Conference on Databases (BDA'01), Agadir, Maroc, November, 2001, S. 133-144.
- [KoMM00] Kouba, Z.; Matousek, K.; Miksovsky, P.: On Data Warehouse and GIS Integration. In: Proc. 11th International Conference on Database and Expert Systems Applications (DEXA 2000), Greenwich, UK, September 2000, S. 604-613.
- [Lehn00] Lehner, F.: *Organizational Memory*. Carl Hanser Verlag, München, 2000.
- [MeRi01] Mentrup, A.; Rieger, B.: MSS und Wissensmanagement: Dimensionen und Perspektiven der Integration. In: *Professionelles Wissensmanagement – Erfahrungen und Visionen (WM 2001)*, Baden-Baden, März 2001, S. 99-112.
- [MiSR02] Miller, L.; Seaborne, A.; Reggiori, A.: Three Implementations of SquishQL, a Simple RDF Query Language. In: *First International Semantic Web Conference (ISWC2002)*, Sardinia, June, 2002, S. 423-435.
- [Popp02] Popp, K.: Nutzbarmachung von Portaltechnologie: mySAP Enterprise Portals. In Meinhardt, S.; Popp, K. (Hrsg.): *Enterprise-Portale & Enterprise Application Integration*. HMD 225, Dpunkt-Verlag, Juni 2002, S. 21-29.
- [RiKM00] Rieger, B.; Kleber, A.; von Maur, E.: Metadata-based Integration of Qualitative and Quantitative Information Resources Approaching Knowledge Management. In: Proc. 8th European Conference of Information Systems, Wien, Juli 2000, S. 372-378.
- [ShFM02] Shah, U.; Finin, T.; Mayfield, J.: Information Retrieval on the Semantic Web. In: Proc. Eleventh International Conference on Information and Knowledge Management (CIKM 2002), McLean, VA, USA, November 2002, S. 461-468.
- [Thur02] Thuraisingham, B.: *XML Databases and the Semantic Web*. CRC Press, Boca Raton et al., 2002.
- [W3C99] N.N.: Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation, 1999. <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>
- [Wege02] Wege, C.: Portal Server Technology. *IEEE Internet Computing*, Mai/Juni 2002, S. 73-77.