

**"Speicherung schwach strukturierter
Daten für die Verarbeitung in MUS"**

Seminararbeit

an der
Technischen Universität Dresden

im Rahmen des Hauptseminars
„Die Wirtschaftsinformatik auf dem Weg zur Unternehmensspitze“

Fakultät Wirtschaftswissenschaften
Lehrstuhl Wirtschaftsinformatik,
insbesondere Informationssysteme
in Industrie und Handel

eingereicht von
Mathias Paul Eckert
aus Bischofswerda

Betreuer: Prof. Dr. Wolfgang Uhr, Dipl.-Wirtsch.-Inf. Frank Wenzke
Bearbeitungszeitraum: 13.10.03 – 18.12.03

Autorenreferat

Diese Arbeit beschäftigt sich mit Möglichkeiten zur Speicherung von schwach strukturierten Daten für deren anschließende Nutzung in Management-Unterstützungssystemen (MUS). Ausgangspunkt sind dabei die besonderen Merkmale solcher Daten und die sich daraus ergebenden Anforderungen an Systeme zu deren Speicherung. Mit Multimedia-Datenbanken und dem Konzept des Document Warehouse wird danach näher auf zwei konkrete Speichersysteme eingegangen. Dabei werden auch die relevanten, der Speicherung vor- und nachgelagerte Komponenten, dieser Systeme aufgezeigt. Abschließend werden die vorgestellten Konzepte auf die im ersten Teil erarbeiteten Anforderungen geprüft und untersucht inwieweit Multimedia-Datenbanken und Document Warehouses als Bestandteile eines MUS geeignet sind.

Inhaltsverzeichnis

Abkürzungsverzeichnis.....	V
Abbildungsverzeichnis	VI
Tabellenverzeichnis	VII
1. Einführung	1
2. Anforderungen an MUS-Datenbasen	3
2.1 Anforderungen an klassische Datenbanken	3
2.2 Anforderungen an die Speicherung schwach strukturierter Daten	3
3. Multimedia-Datenbanken	4
3.1 Multimedia-Datentypen	5
3.2 Besonderheiten von Multimedia-Datenbanken	6
3.3 Relationale Multimedia-Datenbanken	7
3.4 Objektorientierte Multimedia-Datenbanken	8
3.5 Speicherkonzepte in Multimedia-Datenbanken	9
3.6 Abfrage und Content-Retrieval in Multimedia-Datenbanken	9
3.7 Multimedia-Speichermanagement	11
4. Document-Warehouse	13
4.1 Definition und Abgrenzung	13
4.2 Architektur eines Document-Warehouse-Systems	15
4.2.1 Dokumentenquellen	17
4.2.2 Textverarbeitungsserver	17
4.2.3 Speichermöglichkeiten	18
4.2.4 Metadatenbanken	19
4.2.5 Komponenten zur Integration mit dem Data Warehouse	20
5. Eignung der Konzepte zur Nutzung in MUS	21
5.1 Nutzung von Multimedia-Datenbanken in MUS	21
5.2 Nutzung von Document-Warehouses in MUS	22
5.3 Vergleichendes Fazit	22

Quellenverzeichnis27
Selbständigkeitserklärung.....28

Abkürzungsverzeichnis

ASCII	= American Standard Code for Information Interchange
DMS	= Dokumentenmanagementsystem
GIF	= Graphics Interchange Format
JPEG	= Joint Photographic Experts Group
LOB	= Large Object
MPEG	= Motion Picture Experts Group
MUS	= Management-Unterstützungssysteme
OCR	= Optical Character Recogniton
RAID	= Redundant Array of Inexpensive Disks
RAM	= Random Access Memory
SQL	= Structured Query Language
WWW	= World Wide Web

Abbildungsverzeichnis

Abbildung		Seite
1	Aufbau der vorliegenden Arbeit	2
2	Hierarchisches Speichermanagement	11
3	Document Warehouse als Sammlung von Führungsinformationen	13
4	Document-Warehouse-Systemkomponenten und Prozess	15
5	Analyse als Zusammenspiel zwischen Data Warehouse und Document Warehouse	20

VII

Tabellenverzeichnis

Tabelle		Seite
1	Multimedia-Daten im Vergleich	4
2	Multimedia-Datenbanken und Document Warehouse im Vergleich	23

1. Einführung

Management-Unterstützungssysteme (MUS) als entscheidungsorientierte Sammlung und Aufbereitung von Daten hin zu geschäftsrelevanten Informationen, erlangen eine immer stärkere strategische Bedeutung für Unternehmen (BANGE 03, S. 2). Die Nutzung strukturierter Daten, in Form von Kennzahlen, als Träger von quantitativen Führungsinformationen hat sich dabei schon seit einiger Zeit, nicht zuletzt durch das Data-Warehouse-Konzept, etabliert. Zunehmend wird aber auf die Wichtigkeit von Dokumenten, die qualitative Informationen in Form von unstrukturierten Daten zur Verfügung stellen, als eine zweite wichtige Informationsquelle verwiesen (vgl. BANGE 02, S. 16; UHR 99, S. 403). Unter Dokumenten können dabei neben Textdaten wie etwa Berichte, E-Mails oder Internet-Seiten auch Bilder, Tonaufzeichnungen oder Videos verstanden werden. Diese un- oder schwach strukturierten Daten erschweren durch ihre geringere gemeinsame innere Struktur eine maschinelle Auswertung und erfordern deshalb besondere Konzepte um sie für MUS nutzen zu können (vgl. MERTENS 99, S. 415).

Ziel dieser Arbeit ist es Konzepte zur Speicherung von schwach strukturierten Daten zu identifizieren und vorzustellen, sowie diese auf eine mögliche Nutzung in MUS hin zu überprüfen. Zunächst werden dafür in Kapitel 2 Anforderungen an die Speicherung von schwach strukturierten Daten in MUS definiert. Ziel der Arbeit kann es nicht sein einen kompletten Überblick über alle Möglichkeiten zur Speicherung von schwach strukturierten Daten zu geben. Deshalb werden anschließend mit Multimedia-Datenbanken (Kapitel 3) und dem Document-Warehouse (Kapitel 4) zwei ausgewählte Konzepte näher vorgestellt. Abschließend wird in Kapitel 5 untersucht inwieweit die vorgestellten Konzepte die in Kapitel 1 definierten Anforderungen erfüllen und darauf aufbauend die Eignung der Konzepte für MUS überprüft.

Einen Überblick über den Aufbau der Arbeit gibt Abbildung 1.

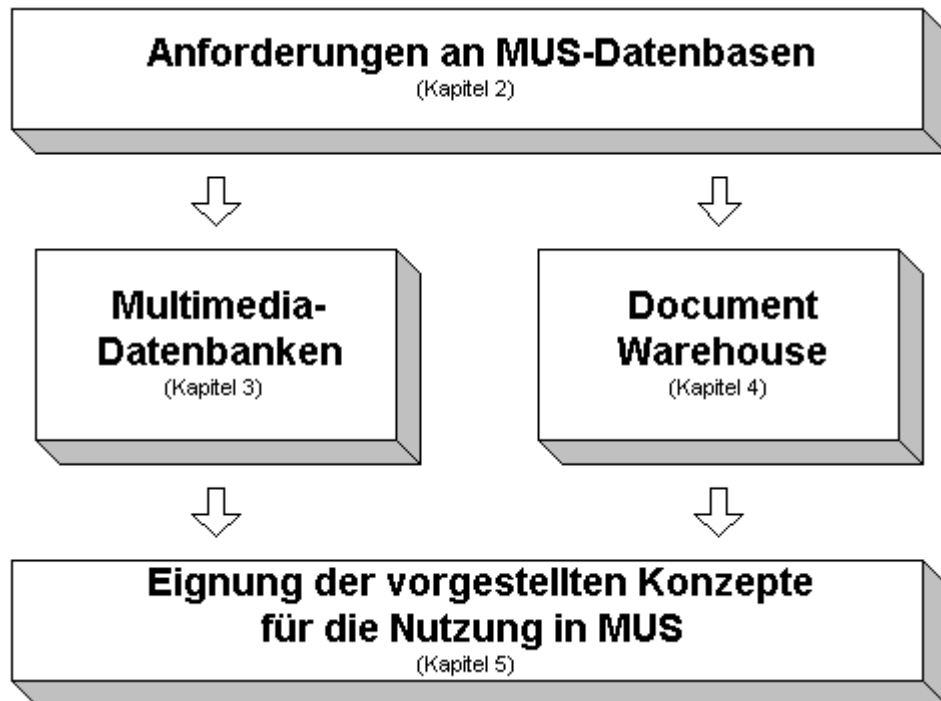


Abbildung 1: Aufbau der vorliegenden Arbeit

Durch die Arbeit wird versucht die folgenden Forschungsfragen zu beantworten:

- Welche besonderen Anforderungen stellen schwach strukturierte Daten an die für deren Speicherung und Verarbeitung relevanten Teile eines MUS ?
- Welche konkreten Möglichkeiten zur Speicherungen von schwach strukturierten Daten existieren ?
- Inwieweit sind die identifizierten Konzepte für die Speicherung und Verarbeitung von schwach strukturierten Daten in MUS geeignet ?

2. Anforderungen an MUS-Datenbasen

Bevor in den nachfolgenden Kapiteln auf konkrete Konzepte zur Speicherung von schwach strukturierten Daten eingegangen wird, werden in diesem Kapitel zunächst Anforderungen an solche Konzepte definiert. In Kapitel 2.1 wird dabei kurz auf allgemeine Anforderungen an klassische Datenbanken eingegangen während im Kapitel 2.2 die Besonderheiten von Systemen zur Speicherung von schwach strukturierten Daten für die Verarbeitung in MUS beleuchtet werden um daraus spezielle Anforderungen abzuleiten.

2.1 Anforderungen an klassische Datenbanken

Schwach strukturierte Daten können in Textbasen oder Dokumentenmanagementsystemen (DMS), in Datenbanken mit Unterstützung von multimedialen Datenformaten oder in Dateisystemen gespeichert werden (vgl. SULLIVAN 01, S. 92). Im Falle einer Speicherung in speziellen Datenbanken werden an diese die gleichen Anforderungen gestellt wie an klassische Datenbanken. Im Folgenden sollen diese Anforderungen kurz genannt werden (vgl. MATTHIESSEN 00, S. 17ff.):

- Fähigkeit zur Verwaltung großer Datenbestände
- Beliebige Möglichkeiten der Verknüpfung der Datenbestände nach inhaltlichen Gesichtspunkten
- Flexible Abfragbarkeit (Recherchemöglichkeiten)
- Redundanzfreiheit
- Datenkonsistenz
- Mehrbenutzerbetrieb
- Trennung von Datenbank-Anwendern und Datenbank-Administratoren

2.2 Anforderungen an die Speicherung schwach strukturierter Daten

Aus den speziellen Eigenschaften schwach strukturierter Daten ergeben sich besondere Anforderungen an Systeme zu deren Speicherung. Zu nennen sind hier insbesondere die

Fähigkeit zur Verarbeitung und Speicherung verschiedensten Datenformaten sowie Möglichkeiten zur Umwandlung von Daten in bestimmte Datenformate. Ebenfalls beachtet werden muß die, im Vergleich zu klassischen Datenbanken, tlw. immense Größe einzelner Datenobjekte. Ein effizientes Retrieval in Beständen von unstrukturierte Daten stellt ebenfalls besondere Anforderungen an die Datenbank. Es werden neben speziellen Werkzeugen zur direkten Recherche in den Datenobjekten auch Mechanismen zur Gewinnung und gesonderten Speicherung von Metadaten über die eigentlichen Datenobjekte benötigt, auf deren Basis dann eine indirekte Recherche erfolgen kann. Wie bei klassischen Datenbanken muß auch eine beliebige Verknüpfbarkeit der Datenbestände nach inhaltlichen Gesichtspunkten möglich sein. Dies umfaßt Datenbestände in den verschiedensten Formaten inklusive die Verknüpfung strukturierter und schwach strukturierter Daten. MERTENS (vgl. 99, S. 405ff.) kritisiert in diesem Zusammenhang die derzeit verbreitete separate Speicherung (z. B. in Data Warehouse, Dokumentenmanagementsystemen, Document Warehouse) von unterschiedlichen Datentypen. Er fordert eine Abkehr von Konzept verschiedener Datenbasen, hin zu einer integrierten MUS-Datenbasis zu Speicherung von qualitativen, quantitativen, externen und internen Daten. Ein solches Information Warehouse kann dann als leistungsfähige Informations-Drehscheibe eines Unternehmens fungieren (MERTENS 03, S. 67).

3. Multimedia-Datenbanken

Die rasante Entwicklung der Computer-Technologie im letzten Jahrzehnt ist verbunden gewesen mit Veränderungen in der Art, in der Informationen präsentiert werden (vgl. DUNCKLEY 03, S. 1) und einer steigenden Bedeutung von „Nicht-Textdaten“. Diese sogenannten Multimedia-Daten haben einige etablierte Bereiche der Informatik, wie z.B. die Datenbanktechnologie, vor neue Herausforderungen gestellt (vgl. SUBRAHMANIAN 98, S. XV). Im folgenden Kapitel sollen Möglichkeiten zur Speicherung von Multimedia-Daten näher untersucht werden. Zunächst werden die verschiedenen Multimedia-Datentypen kurz vorgestellt (Kapitel 3.1) um anschließend

den Begriff Multimedia-Datenbank abzugrenzen und zu definieren (Kapitel 3.2). Die Verträglichkeit mit relationalen und objektorientierten Datenbankkonzepten wird dann in den Kapiteln 3.3 und 3.4 beleuchtet, bevor auf Besonderheiten bei der Speicherung bzw. Abfrage in Multimedia-Datenbanken eingegangen wird (Kapitel 3.5 und 3.6). Abschließend wird in Kapitel 3.7 kurz die Problematik des Multimedia-Speichermanagements vorgestellt.

3.1 Multimedia-Datentypen

Medium	Elemente	Anordnung	typische Größe	zeitabhängig ?	Sinn
Text	druckbare Zeichen	Folge	10 KB (5 Seiten)	Nein	Visuell/akustisch
Graphik	Vektoren, Flächen, Körper (3D), Texturen	Menge oder Folge	10-100 KB	Nein	visuell
Rasterbild	Bildpunkte (Pixel)	Matrix	1 MB (1024x1024)	Nein	visuell
Tonaufnahme (Audio)	Lautstärkepegel	Folge	600 MB (Audio-CD)	Ja	akustisch
Bewegtbild (Video-Clip)	Rasterbild, Graphik	Folge	2 GB (30 Min.)	Ja	visuell

Tabelle 1: Multimedia-Daten im Vergleich (in Anlehnung an Meyer 02, S. 13)

Es existiert eine Anzahl von Datentypen (vgl. Tabelle 1) die als Multimedia-Datentypen charakterisiert werden können und demzufolge für eine Speicherung in Multimedia-Datenbanken in Frage kommen. Im Folgenden werden diese aufgeführt: (vgl. KOSHAFIAN 96, S. 4ff.):

- *Text*: Text besteht aus einer Folge von druckbaren Zeichen (vgl. MEYER 91, S. 31). Es bestehen verschiedene Möglichkeiten der Speicherung. Von einfachen ASCII-basierten Textdateien bis hin zu komplexen Word-Dateien. Trotz der zunehmenden Dateigröße durch immer komplexere Möglichkeiten der Textgestaltung (Schriftarten, Stile, Farben etc.) ist Text der am wenigsten speicherintensivste Multimedia-Datentyp.
- *Tonaufnahme (Audio-Daten)*: Nach vorheriger Digitalisierung (vgl. MEYER 91, S. 59) existieren auch hier eine Reihe von möglichen Formaten zur Speicherung

(WAV, MPEG Layer 3 etc.). Die Speicherintensivität hängt stark von der Qualität der Aufnahme ab, ist aber im Vergleich zur Speicherung von Text, trotz Einsatz von unterschiedlichen Komprimierungsverfahren recht hoch.

- *Graphik/Rasterbild*: Digitalisierte Graphiken werden als Sequenz von Pixeln gespeichert. Auch hier gibt es große Unterschiede die Speicherintensivität betreffend. Diese ist abhängig von Auflösung, Größe, Farben oder der verwendeten Kompression. Es existieren verschiedene gängige Graphikformate wie z.B. JPEG oder GIF.
- *Bewegtbild (Video-Clip)*: Digitalisierte Videos sind die speicherintensivsten Multimediaobjekte. Videoaufzeichnungen werden als Daten einer höheren Stufe bezeichnet, da sie eine komplizierte Aggregation von Bildern und Tonaufnahmen darstellen (vgl. MEYER 91, S. 63). Sie werden unter Einsatz verschiedener Kompressionstechniken als eine Sequenz von sogenannten Frames gespeichert.

Die oben genannten Multimedia-Datentypen treten meist nicht einzeln auf, sondern sind Elemente von komplexen Multimedia-Objekten.

3.2 Besonderheiten von Multimedia-Datenbanken

Konventionelle relationale und objektorientierte Datenbanken unterstützen ein Reihe von einfachen Datentypen wie z.B. Integerwerte, Fließkommazahlen oder Strings. Auch die Speicherung von großen Objekten ist seit einigen Jahren möglich (vgl. KOSHAFIAN 96, S. 21). So unterstützt die Abfragesprache SQL ab ihrer dritten Version von 1999 die Speicherung von sogenannten „large objects“ (LOBs) (vgl. SCHNEIDER 00, S. 527). Trotzdem werden Multimedia-Objekte in solchen Datenbanken meist nicht direkt gespeichert, sondern nur die Referenzen zu externen Speicherorten festgehalten (vgl. KOSHAFIAN 96, S. 22). Außerdem sind Content-Retrieval-Funktionen, speziell für Multimedia-Inhalte, kaum implementiert. Eine Verwaltung von Multimedia-Objekten in herkömmlichen relationalen oder

objektorientierten Datenbanken ist also konzeptionell zwar möglich, kann jedoch nicht den gewünschten Funktionsumfang realisieren.

Mit speziellen Multimedia-Datenbanken soll den oben genannten Defiziten begegnet werden. Diese erweitern herkömmliche Datenbanken um einige Funktionalitäten (vgl. SCHNEIDER 00, S. 527). Nach KHOSAFIAN (vgl. 96, S. 264) besitzt eine ideale Multimedia-Datenbanken die folgenden Eigenschaften:

- Eigenschaften herkömmlicher Datenbanken
- Unterstützung von Multimedia-Datentypen
- Kapazität zur Verarbeitung einer sehr großen Anzahl von Multimediaobjekten
- Unterstützung von hierarchischer Speicherstruktur und Archivierung
- Information-Retrieval-Funktionalität

Auch DUNCKLEY (vgl. 03, S. 7) zählt Funktionalitäten auf, um die traditionelle Datenbanken im Kontext der Multimediaeignung zu erweitern sind. Dazu zählen:

- Techniken zur Reduzierung der immensen Größe von Multimediaobjekten
- Funktionen zur Darstellung bzw. Synchronisation von zeitsensitiven Daten (Audio, Video)
- Metadatenkonzepte zur Erschließung der Semantik von Multimediadaten

Zusammenfassend können Multimedia-Datenbanken also als Datenbanksysteme mit Verwaltungsmöglichkeiten für multimediale Strukturen beschrieben werden (vgl. SCHEIDER 00, S. 527).

3.3 Relationale Multimedia-Datenbanken

Relationale Datenbanken sind die bis heute dominierende Form von Datenbanken (vgl. KOSHAFIAN 96, S. 274). Die Unterstützung von Multimedia-Objekten wurde anfangs hier durch die Existenz von Feldern mit variabler Länge gewährleistet. Oft fand jedoch trotzdem eine Beschränkung auf eine bestimmte Größe statt, wodurch von einer hinreichenden Multimediaunterstützung nicht gesprochen werden konnte. Erst die

Einführung von LOBs (siehe Kapitel 2.2) ermöglichte eine Verarbeitung von multimedialen Inhalten. Trotz dieser Fortschritte können relationale Datenbanken nur strukturierte Datentypen optimal verarbeiten. Einem effizienten Arbeiten mit unstrukturierten Daten sind dagegen schon bei einfachen SQL-Abfragen Grenzen gesetzt (vgl. DUNCKLEY 03, S. 28).

3.4 Objektorientierte Multimedia-Datenbanken

Objektorientierte Datenbankkonzepte werden oft als ideale Grundlage für Multimedia-Datenbanken bezeichnet (vgl. KOSHAFIAN 96, S. 31), da sowohl die Objektorientierung als auch Multimedia an sich einige Gemeinsamkeiten besitzen. So ist es Ziel des objektorientierten Paradigmas Dinge der realen Welt abzubilden. Genauso geben Multimedia-Daten wie Bilder oder Tonaufnahmen die reale Welt wieder. Die drei fundamentalen Konzepte der Objektorientierung, abstrakte Datentypen, Vererbung und Objektidentität sind auch wichtige Eigenschaften von objektorientierten Datenbanken. Im Hinblick auf Multimedia-Datentypen entstehen dadurch einige Vorteile. Es besteht die Möglichkeit der einfachen Erweiterung bzw. dem Hinzufügen von Datentypen (vgl. KOSHAFIAN 96, S. 293ff.). Dadurch können die oft sehr komplexen Multimediaobjekte als Ganzes verarbeitet werden und müssen nicht wie in relationalen Datenbanken geteilt und anschließend in verschiedenen Tabellen gespeichert werden. Ähnliche Objekte werden durch Klassen (vgl. MEYER 91, S. 169f) repräsentiert und Verbindungen zwischen verschiedenen Objektarten durch Klassenhierarchien abgebildet (vgl. SUBRAHMANIAN 98, S. 44f.). Positiv wirkt sich auch die einheitliche Verwendung eines Paradigmas sowohl in der Datenbank als auch in den darauf zugreifenden Anwendungen aus. Trotz der augenscheinlichen Vorteile des objektorientierten Ansatzes gegenüber dem relationalen Konzept wird es in seiner Bedeutung jedoch immer noch als recht gering einzuschätzen (vgl. SUBRAHMANIAN 98, S. 38).

3.5 Speicherkonzepte in Multimedia-Datenbanken

Spezielle Techniken zur Speicherorganisation und Indexerstellung sind die Grundlage für effektive Abfragen in Datenbanken (vgl. KOSHAFIAN 96, S. 415). Die Alternative, nacheinander alle vorhandenen Datensätze zu durchsuchen ist gerade im Hinblick auf die Größe von Multimedia-Objekten nicht realistisch. Es existieren verschieden Formen der Indexerstellung. Dabei werden wichtige Attribute der Objekte als sogenannte Metadaten separat gespeichert. Zu nennen sind hier z.B. B-Bäume (vgl. FALOUTOS 98, S. 13f.). Diese Methode unterstützt nicht nur exakte Treffer sondern auch sogenannte „Spielräume“, also unscharfe Anfragen. Eine andere Methode, das Hashing, (vgl. FALOUTOS 98, S. 11ff.) ist dagegen nur für genaue Treffer geeignet (vgl. KOSHAFIAN 96, S. 416). Für Multimedia-Datenbanken besonders geeignet erscheint das Konzept der multidimensionalen Indexerstellung, da hier eine Indexstruktur mehrere Attribute eines Objektes gleichzeitig indiziert. Beim Indizieren, egal nach welchem Modell, bieten sich grundsätzlich zwei Vorgehensweisen an. Das manuelle setzen von Indexen und das automatische Indizieren (vgl. KOSHAFIAN 96, S. 334). Während beim Ersteren der Nutzer verschiedene Eigenschaften des Objektes eingibt und anschließend indiziert, werden bei letzterem Ansatz durch einfache Berechnungen oder Mustererkennungen die Metadaten automatisch erzeugt und indiziert. Technologien für letzteren Ansatz im Kontext von Multimedia-Daten sind jedoch recht anspruchsvoll und noch nicht ausgereift (vgl. DUNCKLEY 03, S. 206).

Auf Performanceaspekte bei der Suche nach Daten richtet sich dagegen das Clustering. Dabei werden Objekte auf die häufig gleichzeitig oder nacheinander zugegriffen wird physisch in gleichen oder nahen Speicherbereichen gespeichert (vgl. KOSHAFIAN 96, S. 416).

3.6 Abfragen und Content-Retrieval in Multimedia-Datenbanken

Aufbauend auf den verschiedenen oben beschriebenen Speichertechniken existieren verschiedene Möglichkeiten Objekte in Multimedia-Datenbanken abzufragen oder zu

finden. Grundsätzlich unterscheidet man dabei zwei Konzepte (vgl. KOSHAFIAN 96, S. 325ff.)

Beim der Navigation (oder Browsing) springt der Nutzer ausgehend von einem Wurzelobjekt solange in mit dem jeweils aktuellen Objekt referenzierte Objekte weiter bis das gewünschte Ziel erreicht ist. Verschiedene Objekte sind hier also in einer hypertextähnlichen Struktur miteinander verbunden. Nachteile ergeben sich durch lange Suchwege nach Objekten die in sehr tiefen Ebenen liegen.

Eine andere Möglichkeiten zum Finden von bestimmten Objekten sind Suchanfragen. Dieses Konzept basiert auf den Werten bestimmter Attribute oder Eigenschaften des gesuchten Objektes. Der Nutzer formuliert dabei konkrete Anfragen in einer Anfragesprache. Die Anfragen können sich entweder auf Metadaten der jeweiligen Multimedia-Objekte oder auf den Inhalt der Objekte selbst beziehen. Es bestehen jedoch Unterschiede zu Abfragen in konventionellen Datenbanken. In konventionellen Datenbanken basieren Abfragen auf vorhersagbaren Ausdrücken und numerischen Feldern bzw. Attributen. Diese existieren bei den Anfragen auf Metadaten von Multimedia-Objekten zwar auch, bei der inhaltsbezogenen Form der Suche reichen diese jedoch oft nicht aus. Abhängig von der Art des Multimediaobjekts können z. B. Stimminformationen bei Tonaufnahmen oder bestimmte Szenen bei Bewegtbildern in die Abfrage eingeschlossen werden (vgl. KOSHAFIAN 96, S. 330). Die Existenz von visuellen Abfragen durch Beispiele, mögliche Gewichtungen sowie Spielräume bei den Abfragen sind weitere für Multimedia-Abfragesprachen gewünschte Eigenschaften. Bei der Ausgabe der Ergebnisse sollte ein Vergleich der Gewichtungen in der Anfrage mit den Gewichtungen der einzelnen Attribute der gefundenen Objekte erfolgen und anschließend eine Darstellung in einer entsprechende Rangfolge erfolgen (vgl. KOSHAFIAN 96, S. 333). Eine weitere wichtige Funktionalität sind implementierte Feedback-Mechanismen. Durch automatische Auswertungen des Nutzerverhaltens soll das Suchverfahren iterativ verbessert werden. So erstellt das System z. B. neue, optimierte Suchanfragen oder ändert die Gewichtungen. Diese Änderungen können natürlich auch vom Nutzer direkt vorgenommen werden.

Während bei der indirekten Suche die Metadaten über ein bestimmtes Objekt der entscheidende Faktor sind, spielen bei der inhaltsbezogenen Suchen in Multimedia-Objekten die äußeren Attribute keine Rolle. Hier wird der Inhalt des jeweiligen Objektes direkt auf entsprechende Übereinstimmungen mit der Anfrage untersucht (vgl. KOSHAFIAN 96, S. 353). Abhängig vom Medientyp existieren dafür verschieden Methoden. In Textdateien wird nach Ausdrücken gesucht die in der Anfrage formuliert wurden. In Graphiken kann z. B. nach einzelnen Objekten oder Farbverläufen gesucht werden oder versucht werden bestimmte Muster zu erkennen (vgl. DUNCKLEY 03, S. 362ff.). Häufig anzutreffen sind sogenannte OCR-Systeme, die Muster in als Graphik gespeicherten Textdokumenten identifizieren (vgl. KOSHAFIAN 96, S. 354). Auch der Inhalt von Tonaufnahmen kann direkt durchsucht werden. Zu nennen sind hier Techniken wie Stimm- oder Spracherkennung (vgl. DUNCKLEY 03, S. 408). Da Bewegtbilder aus Sequenzen von einzelnen Bildern sowie den dazugehörigen Audiodaten bestehen können hier die gleichen Techniken wie bei der Suche in Graphiken bzw. Tonaufnahmen angewendet werden (vgl. KOSHAFIAN 96, S. 354; DUNCKLEY 03, S. 414f.).

3.7 Multimedia-Speichermanagement

Wie in allen Bereichen der Computer-Technologie hat auch im Bereich der Speicher in den letzten Jahren eine rasante Entwicklung stattgefunden. Die verfügbaren Kapazitäten sind dabei in dem Maße gestiegen wie Zugriffszeiten und Kosten gesunken sind (vgl. KOSHAFIAN 96, S. 501). Trotzdem erfordern Ansammlungen von speicherintensiven Multimedia-Objekten spezielle Maßnahmen des Speichermanagements (vgl. DUNCKLEY 03, S. 402), da aus Kosten- und Kapazitätsgründen sowie aufgrund von technischen Restriktionen nicht alle vorhanden Daten auf einem magnetischen Speicher mit Direktzugriff gelagert werden können (vgl. KOSHAFIAN 96, S. 453). Im Folgenden soll auf einige solcher Konzepte kurz eingegangen werden.

Bei der RAID-Technologie werden die Daten auf verschiedenen kleineren magnetischen Platten gespeichert die insgesamt als eine große Platte organisiert sind (vgl.

SCHNEIDER 00, S. 510). Im Hinblick auf die für Abfragen benötigte Bandbreite ergeben sich keine Nachteile. Demgegenüber werden jedoch die Kosten gesenkt und die Verlässlichkeit des Speichers erhöht (vgl. KOSHAFIAN 96, S. 456).

Eine weitere Möglichkeit ist das Auslagern von Daten auf optische Speichermedien. Die Vorteile sind hier eine nahezu unbeschränkte Kapazität bei gleichzeitig geringen Kosten. Neben der relativ langen Zugriffszeit muß aber auch die Fehleranfälligkeit als Nachteil genannt werden (vgl. KOSHAFIAN 96, S. 461).

Die oben genannten Unterschiede der Speicherarten führten zum Konzept des hierarchischen Speichermanagements (vgl. Abbildung 2). Die Gesamtspeicherkapazität wird dabei in verschiedene Ebenen eingeteilt. Das System versucht nun Muster in der Nutzung des Datenbank-Systems zu erkennen (vgl. DUNCKLEY 03, S.403) und speichert die Objekte in einer bestimmten Ebene (vgl. KOSHAFIAN 96, S. 488). Während auf der einen Seite Speicherarten für häufig genutzte Daten mit sehr niedrigen Zugriffszeiten aber auch Kapazitäten stehen werden sehr selten benötigte Objekte in Offline-Speicher ausgelagert. So soll ein Ausgleich zwischen effektiver Nutzung und Kostenintensität gefunden werden (vgl. CHUNG 96, S. 61ff.).

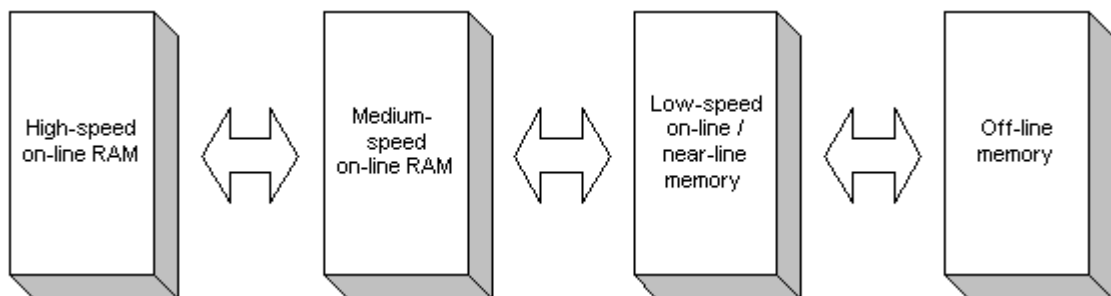


Abbildung 2: Hierarchisches Speichermanagement (in Anlehnung an KHOSHAFIAN, S. 489)

4. Document Warehouse

Mit dem Konzept des Data Warehouse hat sich in den letzten Jahren eine Möglichkeit zur Unterstützung von MUS äußerst erfolgreich etabliert (vgl. SCN 01, S.33ff.). Mit dieser Technik lassen sich jedoch nur strukturierte Daten speichern und verarbeiten. Ein Großteil der in Unternehmen anfallenden Daten liegt jedoch in schwach- oder unstrukturierter Form, z.B. als Textdokument, vor. Um auch diese für Business Intelligence nutzbar zu machen wird mit dem Konzept des Document Warehouse versucht die erfolgreichen Ideen des Data Warehouse speziell auf Dokumente anzuwenden (vgl. BANGE 02, S. 17). In den folgenden Kapiteln wird zunächst der Begriff des Document Warehouse definiert und von verwandten Konzepten abgegrenzt (Kapitel 4.1). Anschließend wird in Kapitel 4.2 auf die einzelnen Komponenten eines Document Warehouse bzw. eines Document-Warehouse-Systems näher eingegangen.

4.1 Definition und Abgrenzung

Nach BANGE (vgl. 02, S. 115) ist ein Document Warehouse eine integrierte, zeitbezogene und dauerhafte Sammlung von Dokumenten zur Entscheidungsunterstützung des Managements. Integration meint hier sowohl die physische Integration von Daten aus verschiedenartigen Quellen als auch die inhaltliche Integration durch Vereinheitlichung von Dokumententypen, -formaten und -metadaten sowie die einheitliche Repräsentation (vgl. BANGE 02, S. 115). Dabei ist jedes gespeicherte Dokument zeitbezogen. Hier interessiert nicht nur das Erstellungs- oder Speicherdatum sondern vor allem der zeitliche Bezug des Inhaltes (vgl. BANGE 02, S. 116). Außerdem werden Daten grundsätzlich langfristig gespeichert um z. B. aus historischen Daten auf aktuelle Ereignisse schließen zu können. Ähnlich argumentiert SULLIVAN (vgl. 01, S. 13). Er charakterisiert ein Document Warehouse durch die folgenden Eigenschaften:

- Es existiert keine einheitliche Dokumentstruktur bzw. kein einheitlicher Dokumenttyp

- Es existieren verschiedenartige Quellen für Dokumente
- Metadaten über Dokumente werden automatisch erzeugt und gesondert gespeichert
- Semantisch in Verbindung stehende Dokumente werden miteinander integriert

Zusammenfassend bezeichnet er das Document Warehouse als reichhaltige Sammlungen von miteinander verbundenen Führungsinformationen (vgl. Abbildung 3).

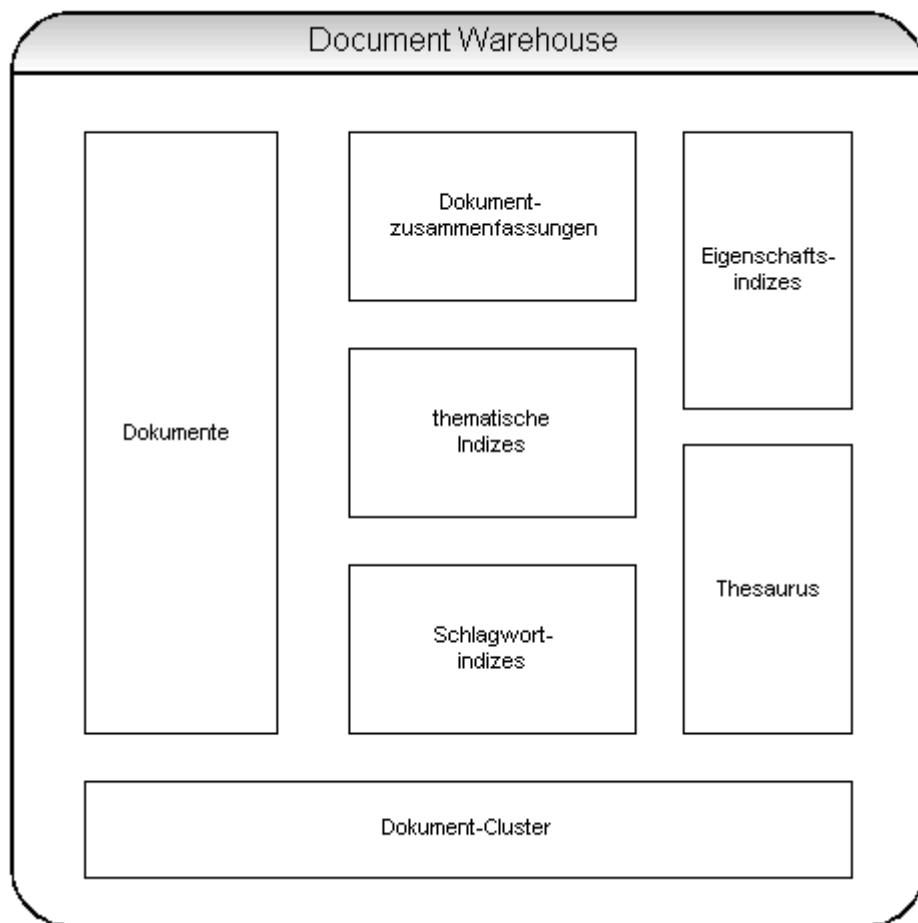


Abbildung 3: Document Warehouse als Sammlung von Führungsinformationen (in Anlehnung an SULLIVAN 01, S. 25)

BANGE (vgl. 02, S. 116) unterscheidet außerdem grundsätzlich zwischen dem Document Warehouse im engeren Sinne und einem Document-Warehouse-System. Für ihn ist das Document Warehouse die zentrale Speicherkomponente eines komplexen

Document-Warehouse-Systems welches zusätzlich vor- und nachgelagerte Komponenten enthält.

Aufgrund der Entstehungsgeschichte ergeben sich gemeinsame Eigenschaften, von Document Warehouse und Data Warehouse. Beispiele hierfür sind die Zielgruppe, die Integration verschiedener Datenquellen oder die Zeitbezogenheit der gespeicherten Informationen (vgl. SULLIVAN 01, S. 11; GLUCHWOSKI 97). Die beiden Konzepte unterscheiden sich hinsichtlich der gespeicherten Daten. Während das Data Warehouse strukturierte Daten speichert und verarbeitet, werden Document Warehouses für schwach strukturierte Daten, im speziellen Dokumente, entwickelt.

Klassische Dokumentenmanagementsysteme (DMS) sind ebenfalls von Document Warehouses zu unterscheiden. Bestandteile der DMS sind Archivierungskomponente, Vorgangs- oder Workflowunterstützung und Recherchemöglichkeiten (vgl. STAHLKNECHT 99, S. 440). Die Dokumente sind hier, im Gegensatz zum Document Warehouse, meist gleichartig und interner Herkunft, treten dafür aber in großer Zahl auf. Außerdem erfolgt im Normalfall keine oder eine nur rudimentäre semantische Anreicherung der Dokumente mit Metadaten. Beispielanwendungen sind das Belegwesen bei Banken oder Versicherungen. Zweck der DMS ist also nicht die Entscheidungsunterstützung, sondern die effiziente Abwicklung eines bestimmten betrieblichen Prozesses (vgl. BANGE 02, S. 113f.). Es bestehen somit Unterschiede sowohl bei Art und Menge der gespeicherten Dokumente, der Zielgruppe als auch beim Zweck von DMS und Document Warehouse.

4.2 Architektur eines Document-Warehouse-Systems

Nach SULLIVAN (vgl. 01, S. 123) besteht ein Document-Warehouse-System aus folgenden Schlüsselkomponenten:

- Dokumentenquellen
- Textverarbeitungsserver
- Speichermöglichkeiten

- Metadatenbanken
- Benutzerprofilen
- Komponenten zur Integration mit dem Data Warehouse

Große Ähnlichkeit mit den eben genannten Schlüsselkomponenten hat die Architektur eines Document-Warehouse-Systems nach BANGE (vgl. 02, S. 117; Abbildung 4).

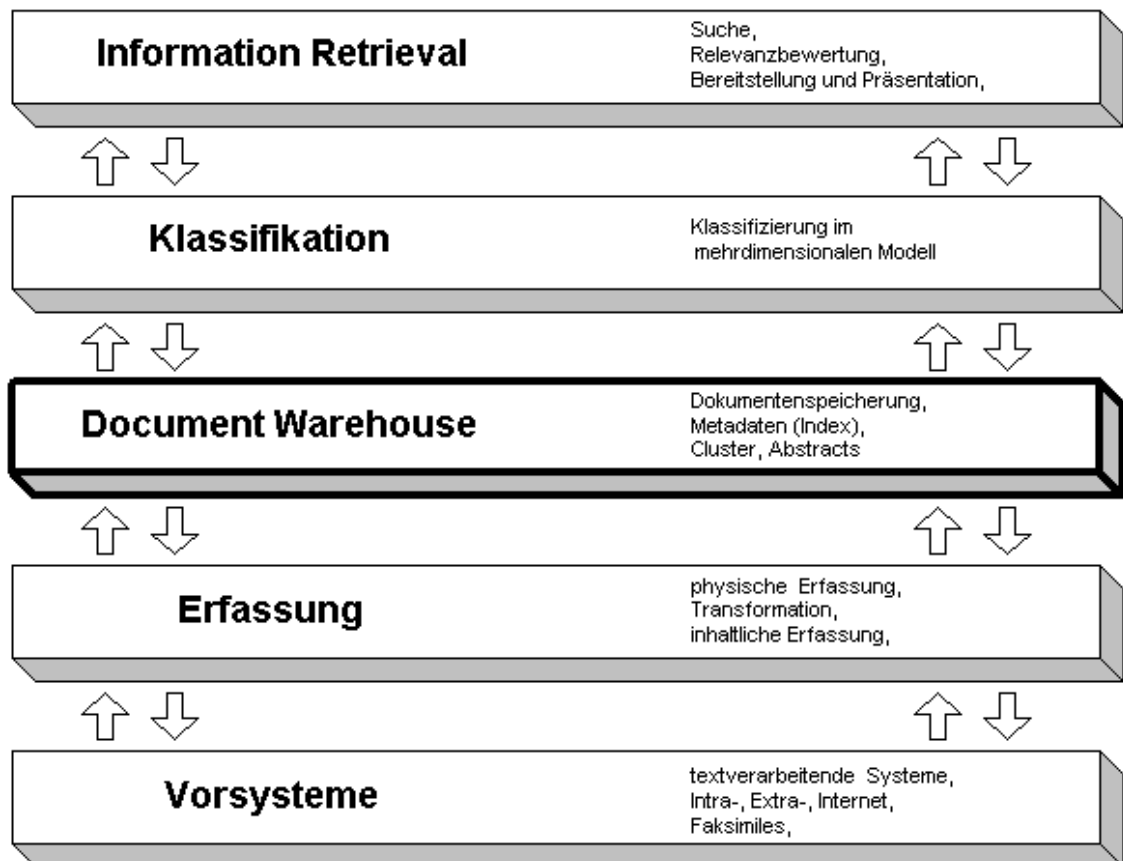


Abbildung 4: Document-Warehouse-Systemkomponenten und Prozess
(in Anlehnung an BANGE 02, S. 117)

Im Folgenden wird auf die einzelnen Komponenten der Architekturen nach SULLIVAN (vgl. 01, S. 123) und BANGE (vgl. 02, S. 117) näher eingegangen. Da das Hauptaugenmerk dieser Arbeit auf der Speicherung schwach strukturierter Daten liegt, also dem Document Warehouse im engeren Sinne, wird daher auf eine nähere Beschreibung der Komponente Benutzerprofile sowie auf Ausführungen zum Thema Benutzerschnittstelle eines Document Warehouse verzichtet.

4.2.1 Dokumentenquellen

Eines der Hauptmerkmale eines Document Warehouse sind die verschiedenen Dokumentenquellen (vgl. BANGE 02, S. 119) und die deren daraus resultierenden verschiedenen Dokumentstrukturen (siehe 3.1). Grundsätzlich kommen eine große Anzahl von sowohl internen als auch externen Quellen zur Versorgung des Document Warehouse mit Daten in Frage. Zu nennen sind hier beispielsweise interne Dokumentenmanagementsysteme oder externe Internetquellen (vgl. SULLIVAN 01, S. 124), welche unterschiedlich gut für die Nutzung in Document Warehouses geeignet sind. Interne Quellen bieten verlässliche Informationen in bekannter Sprache und Format. Eine Nutzung kann demzufolge größtenteils problemlos erfolgen. Die Heterogenität sowohl in Verlässlichkeit, Inhalt, Form oder Sprache externer Quellen ist dagegen oft problematisch und erfordert bestimmte Maßnahmen der Aufbereitung und Redaktion. Beispiele für solche Quellen sind allgemeine Informationsseiten im World Wide Web (WWW), spezialisierte kommerzielle Informationsdienste oder Informationsseiten öffentlicher Einrichtungen (vgl. SULLIVAN 01, S. 87).

4.2.2 Textverarbeitungsserver

Textverarbeitungsserver spielen eine zentrale Rolle in Document Warehouses (vgl. SULLIVAN 01, S. 136). Sie sind verantwortlich für das Beschaffen, Analysieren und Verteilen von Dokumenten.

Geeignete Werkzeuge für die automatische Beschaffung von Daten sind Crawler und Agenten. Crawler sind zum Einsatz in Hypertextstrukturen wie dem WWW geeignet. Ausgehend von einem Wurzeldokument mit bestimmten Inhalt können durch Analyse der Hyperlinks Dokumente mit ähnlichen Inhalten gefunden werden (vgl. SULLIVAN 01, S. 137). Da die Semantik der Hyperlinks jedoch keine Rolle spielt, sondern nur die Syntax der Adresse des verlinkten Dokuments können aber auch völlig unsinnige Ergebnisse zu Stande kommen (vgl. WALDBURGER 01, S. 18). Anders dagegen arbeiten sogenannte Agenten. Sie erfüllen eine ganz spezielle Aufgabe, z.B. das Finden

von Informationen zu einem bestimmten Themengebiet (vgl. WALDBURGER 01, S. 19). Im Gegensatz zu Crawlern können Agenten während ihrer Ausführung mit Nutzern und/oder anderen Agenten interagieren (vgl. SULLIVAN 01, S. 137). Sie besitzen also ein gewisse Intelligenz und sind somit für Dokumentensammlungen zu bestimmten Themen besser geeignet als Crawler.

Es existieren eine Reihe von Werkzeugen zur Analyse der gefundenen Dokumente (vgl. BANGE 02, S. 121f.). Kernoperationen sind z.B. das Umwandeln von Dokumente in bestimmte Formate als Voraussetzung für verschiedene Techniken der Indizierung oder Metadatengewinnung. Weiterhin zu nennen sind das automatische Bilden von Dokument-Zusammenfassungen oder das Gruppieren von Dokumenten in Clustern (vgl. SULLIVAN 01, S. 140). Erst nach Anwendung der genannten Techniken existiert eine effiziente Repräsentation des Dokuments für das Document Warehouse. Es ist nun soweit aufbereitet, dass es zur Entscheidungsunterstützung herangezogen werden kann.

Die Verteilung von Dokumenten im Document Warehouse steht für das differenzierte Speichern bestimmter Daten. So werden Metadaten auf der einen und die eigentlichen Dokumente auf der anderen Seite meist in verschiedenen getrennten Datenbanksystemen gespeichert.

4.2.3 Speichermöglichkeiten

Es bieten sich unterschiedliche Möglichkeiten an die verschiedenen Daten in einem Document Warehouse zu speichern.

Datenbanken sind zwar traditionell nicht für die Speicherung von Text optimiert, durch die Entwicklung in den letzten Jahren (vgl. Kapitel 2) kann man aber inzwischen bei den meisten modernen Datenbankmanagementsystemen von einer ausreichenden Textunterstützung sprechen (vgl. SULLIVAN 01, S. 141). Die verschiedenen Datenbanktypen eignen sich in unterschiedlicher Art für die Verwendung in Document Warehouses (vgl. SULLIVAN 01, S. 142ff.). Für objektorientierte Datenbanken

sprechen neben der Klassenhierarchie auch die Möglichkeit der einfachen Definition von komplexen Dokumentstrukturen und die Unterstützung von LOBs. Relationale Datenbanken unterstützen dagegen besonders gut Indizierungstechniken und das Speichern von Metadaten. Im Unterschied zu den beiden schon genannten Typen sind Textbanken, wie etwa Lotus Notes, speziell auf die Speicherung von Dokumenten ausgerichtet (vgl. BANGE 02, S. 123). Hier sind meist auch Werkzeuge zur Suche und/oder Indizierung bereits implementiert (vgl. SULLIVAN 01, S: 143).

Alternative Möglichkeiten zur Speicherung in Datenbanken sind Datei-Systeme und verteilte Speicherung. Datei-Systeme bieten eine einfache Möglichkeit zur Speicherung einer großen Anzahl unterschiedlichster Dokumente. Auf diese können aber nur einige Basisoperationen, wie etwa das Löschen oder Kopieren, angewendet werden (vgl. SULLIVAN 01, S. 125f.). Der Funktionsumfang von Datenbanksystemen wird bei weitem nicht erreicht. Durch verteilte Speicherung wird dagegen versucht die Größe eines Document Warehouse zu begrenzen. Im Document Warehouse direkt werden hierbei nur Metadaten und Verweise auf Dokumente gespeichert. Die eigentlichen Dokumente verbleiben dabei in ihren Quellen außerhalb (vgl. SULLIVAN 01, S. 141).

4.2.4 Metadatenbanken

Neben den Dokumenten selbst müssen auch Informationen über diese Dokumente sowie aus diesen gewonnene Daten (vgl. Kapitel 3.2.2) gespeichert werden (vgl. BANGE 02, S. 140). Diese Metadaten werden in gesonderten Datenbanken gespeichert (vgl. Kapitel 3.2.1). Metadaten lassen sich dabei in verschiedene Kategorien einteilen.

Dokument-Metadaten beziehen sich auf einfache Attribute des Dokumentes wie z.B. Titel, Sprache oder Autor. SULLIVAN (vgl. 01, S. 145) empfiehlt auf verbreitete Standards wie etwa den Dublin Core Metadatenstandard zurückzugreifen. Vorteilhaft wirkt sich das vor allem bei der Integration mit anderen Systemen aus.

Such- und Retrieval-Metadaten beschreiben dagegen wie und wann das Dokument in das Document Warehouse gelangt ist. Beschrieben werden außerdem Quellen und eventuelle verschiedene Versionen des Dokumentes. Für Nutzer die auch an Aktualität und Verlässlichkeit der von ihnen genutzten Dokumenten interessiert sind spielt diese Art von Metadaten eine entscheidende Rolle (vgl. SULLIVAN 01, S. 145f.).

Das Ergebnis von verschiedenen Text-Mining-Operationen wird in sogenannten *Text-Mining-Metadaten* gespeichert. Dazu gehören Stichwörter, Themengebiete, Zusammenfassungen oder Beziehungen zwischen bestimmten Objekten in einem Dokument. Hinzu kommen Informationen über die Häufigkeit bestimmter Stichwörter oder die Relevanz eines bestimmten Themengebietes bezogen auf ein Dokument (vgl. SULLIVAN 01, S. 147). Genutzt werden diese Metadaten somit ähnlich wie Indizes in relationalen Datenbanken.

Darüber hinaus existieren Metadaten welche die Speicherung eines Dokumentes betreffen (*Speicher-Metadaten*). Sie beschreiben wie das Dokument behandelt wird wenn es sich im schon im Document Warehouse befindet (vgl. SULLIVAN 01, S. 148f.). Es wird dabei z. B. festgelegt inwieweit ein Originaldokument gespeichert wird oder nur dessen Zusammenfassung oder Übersetzung. Eine andere mögliche Fragestellung wäre ob ein Dokument im Document Warehouse verbleibt oder ausgelagert wird und nur noch eine Referenz gespeichert wird.

4.2.5 Komponenten zur Integration mit dem Data Warehouse

In Unternehmen existieren Document Warehouses oft neben Data Warehouses, da jedes Konzept für sich allein Fragen oft nicht ausreichend genug beantwortet. So können durch Analysen im Data Warehouse aufgeworfene Probleme vielleicht nicht durch weitere Zahlen erklärt werden. Hilfe könnten hier für das spezielle Problem relevante Text-Dokumente bieten (vgl. Abbildung 5). Andererseits könnten durch Dokumente gewonnen Erkenntnisse möglicherweise durch Analysen im Data Warehouse überprüft werden (vgl. SULLIVAN 01, S. 155). Eine Integration beider Konzepte bietet also

einen bedeutenden Mehrwert. Diese kann z. B. durch die Verknüpfung von strukturierten Datenbeständen im Data Warehouse mit relevanten schwach strukturierten Dokumenten im Document Warehouse erfolgen.

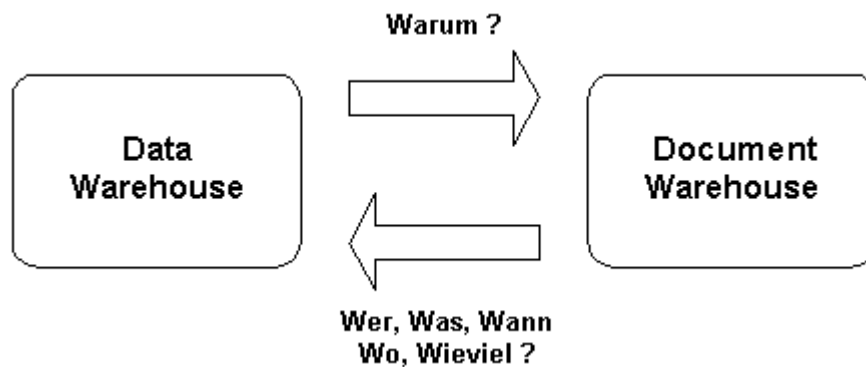


Abbildung 5: Analyse als Zusammenspiel von Data Warehouse und Document Warehouse (in Anlehnung an SULLIVAN 01, S.77)

5. Eignung der Konzepte für die Nutzung in MUS

Die vorgestellten Konzepte sollen nun, anhand der in Kapitel 1 erarbeiteten Anforderungen, auf ihre Eignung für MUS untersucht werden. Zunächst wird dabei in Kapitel 5.1 auf die mögliche Nutzung von Multimedia-Datenbanken für MUS eingegangen, bevor dann das Konzept des Document Warehouse und dessen Eignung untersucht wird (Kapitel 5.2). Am Ende folgt ein abschließendes vergleichendes Fazit der beiden Konzepte (Kapitel 5.3).

5.1 Nutzung von Multimedia-Datenbanken für MUS

Die in Kapitel 2.1 geforderten Eigenschaften an klassische Datenbanken werden von Multimedia-Datenbanken grundsätzlich erfüllt (vgl. Kapitel 3.2). Eine Eignung als MUS-Datenbasis ist jedoch nur bedingt gegeben, da Multimedia-Datenbanken nicht speziell für die Nutzung in MUS entwickelt wurden. Zweck dieser Datenbanken ist

grundsätzlich die Speicherung einer großen Anzahl von Datenobjekten in den unterschiedlichsten Datenformaten (vgl. Kapitel 3.2). Gerade diese Fähigkeit ist für die Datenbasis eines MUS wichtig, da Datenobjekte in den unterschiedlichsten Formaten bzw. aus den unterschiedlichsten Quellen gespeichert und verarbeitet werden müssen. Andere Forderungen an MUS-Datenbasen werden dagegen nicht erfüllt, da die Multimedia-Datenbanken nicht auf die Zielgruppe der MUS ausgerichtet sind. So ist zwar ein Metadatenkonzept vorhanden (vgl. Kapitel 3.5), dessen Ausprägung erreicht aber nicht den erforderlichen Grad.

5.2 Nutzung von Document Warehouses für MUS

Laut Definition wurde das Document-Warehouse-Konzept direkt für die Nutzung in MUS entwickelt. Darum werden auch fast alle Anforderungen, sowohl an klassische Datenbanken als auch an MUS-Datenbasen erfüllt. Ein Einsatz von Document Warehouses in MUS kann als problemlos erfolgen. Kritisch muß jedoch die Spezialisierung auf Text-Dokumente betrachtet werden, die zusätzliche Systeme zur Speicherung und Verarbeitung anderer Datenformate notwendig macht. Die sich daraus ergebende separate Speicherung verschiedenartiger Daten ist mit Problemen verbunden (vgl. MERTENS 99, S. 405ff.). Zwar existieren Ansätze zur Integration des Document Warehouse mit anderen Speichersystemen (vgl. Kapitel 4.2.5), die Forderung nach einer integrierten MUS-Datenbasis zur Speicherung sämtlicher zur Managementunterstützung benötigten Daten ist jedoch nicht erfüllt.

5.3 Vergleichendes Fazit

Ein Vergleich der Konzepte Multimedia-Datenbank und Document Warehouse im Hinblick auf ihre Eignung für die Speicherung von schwach strukturierten Daten für die Verarbeitung in MUS ist nur bedingt möglich. Der Hauptgrund hierfür ist die unterschiedliche Ausrichtung bzw. Zielgruppe der Konzepte.

Multimedia-Datenbanken sind nicht direkt für die Verwendung in MUS konzipiert (vgl. Kapitel 5.1). Trotz der daraus resultierenden unzureichenden Funktionalität ist das Konzept als MUS-Datenbasis, als integrierte Speicherkomponente eines komplexen MUS, gerade wegen der Möglichkeit der Speicherung der unterschiedlichsten Datentypen, geeignet.

Das Document Warehouse dagegen, speziell für die Führungsunterstützung konzipiert, kann direkt als MUS eingesetzt werden, da es neben der Speicherkomponente an sich auch der Speicherung vor- und nachgelagerte Komponenten enthält. Erst diese Komponenten gewährleisten die für MUS geforderten Funktionalitäten.

Tabelle 2 gibt abschließend einen Überblick über die Erfüllung der Anforderungen an MUS-Datenbasen durch Multimedia-Datenbanken und Document Warehouse.

Anforderungen	Multimedia-Datenbanken	Document Warehouse
Klassische Datenbankfunktionalität	erfüllt	erfüllt
Unterschiedliche Datenformate	erfüllt	nur unterschiedliche Textformate
Große Objekte	erfüllt	tlw. erfüllt
Direkte Recherche in Objekten	erfüllt	erfüllt
Komplexes Metadatenkonzept	nicht ausreichend erfüllt	erfüllt
Beliebige Verknüpfbarkeit	erfüllt	erfüllt

Tabelle 2: Multimedia-Datenbanken und Document Warehouse im Vergleich

Quellenverzeichnis

- [BANGE 02] Bange, C.: Integration von Data und Document Warehouses – Verknüpfung strukturierter und unstrukturierter Daten in entscheidungsunterstützten Informationssystemen durch den Knowledge Integrator; Diss., Würzburg 2002.
- [BANGE 03] Bange, C.: Business Intelligence – Systeme und Anwendungen. [http://www.competence-site.de/bisysteme.nsf/91590929723655DBC1256CFB00614A49/\\$File/bi_grundlagen.pdf](http://www.competence-site.de/bisysteme.nsf/91590929723655DBC1256CFB00614A49/$File/bi_grundlagen.pdf), Abruf am 2003-12-10.
- [CHUNG 96] Chung, S.M.: Multimedia Information Storage and Management; Kluwer Academic Publishers, Norwell 1996.
- [DUNCKLEY 03] Dunckley, L.: Multimedia Databases – An Object-Relational Approach; Pearson Education, London u.a. 2003.
- [FALOUTOS 98] Faloutsos, C.: Searching Multimedia Databases by Content; Kluwer Academic Publishers, Norwell 1998.
- [FERSTL 01] Ferstl, O. K.; Sinz E. J.: „Grundlagen der Wirtschaftsinformatik“; Bd. 1, 4. Auflage, Oldenbourg Wissenschaftsverlag, München, 2001.
- [GLUCHOWSKI 97] Gluchowski, P.: GI Informatik Lexikon – Data Warehouse; <http://www.gi-ev.de/informatik/lexikon/inf-lex-data-warehouse.shtml>, Abruf am 2003-12-10.
- [KOSHAFIAN 96] Khoshafian, S./Baker, A.B.: MultiMedia and Imaging Databases, Morgan Kaufmann Publishers, San Francisco 1996.
- [MATHIESSEN 00] Matthiessen, G./Unterstein, M.: „Relationale Datenbanken und SQL“, 2. Auflage, Addison-Wesley Verlag, München, 2000.

- [MERTENS 99] Mertens, P.: „Integration interner, externer, qualitativer und quantitativer Daten auf dem Weg zum Aktiven MIS“, in *Wirtschaftsinformatik* 41. Jg. 1999 (5), S. 405-415
- [MERTENS 03] Mertens, P.: „Die Wirtschaftsinformatik auf dem Weg zur Unternehmensspitze – alte und neue Herausforderungen und Lösungsansätze“, in Uhr, W./Esswein, W./Schoop, E.: *Wirtschaftsinformatik 2003/Band 1*, 1. Auflage, Physica-Verlag, Heidelberg, 2003, S. 49-74.
- [MEYER 91] Meyer-Wegener, K.: *Multimedia-Datenbanken*, Teubner, Stuttgart, 1991.
- [MEYER 02] Meyer-Wegener, K.: *Multimedia-Datenbanken – Kapitel 1 – Einführung und Übersicht.*; Vorlesungsskript, <http://www.dvs.informatik.uni-kl.de/courses/MMDB/WS2003/Vorlesungsunterlagen/Einfuehrung.full.pdf>, Abruf am 2003-12-12.
- [SCHNEIDER 00] Schneider, U./Werner, D.: *Taschenbuch der Informatik; Fachbuchverlag Leipzig im Carl Hanser Verlag, München/Wien 2000.*
- [SCN 01] SCN Education B.V. (Eds): *Data Warehousing – The Ultimate Guide to Bulding Corporate Business Intelligence*, Vieweg, Braunschweig/Wiesbaden, 2001.
- [STAHLKNECHT 99] Stahlknecht, P./Hasenkamp, U.: *Einführung in die Wirtschaftsinformatik; 9. Auflage*, Springer-Verlag, Berlin/Heidelberg/New York, 1999.
- [SUBRAHMANIAN 98] Subrahmanian, V.S.: *Principles of Multimedia Database Systems*, Morgan Kaufmann Publishers, San Francisco 1998.
- [SULLIVAN 01] Sullivan, D.: *Document Warehousing and Text Mining*; John Wiley & Sons Inc., New York u. a., 2001.
- [UHR 99] Uhr, W.: „Externe Daten in Management-Support-Systemen; in: *Wirtschaftsinformatik* 41. Jg 1999 (5), S. 403.

[WALDBURGER 01]

Waldburger, M.: „Information Retrieval und das Web – Grundlagen & Problematik“, interuniversitäres Seminar, Datenbanktechnologie für das Web, Universität Konstanz, Universität Zürich, 2001.

Selbständigkeitserklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen wörtlich oder sinngemäß übernommenen Gedanken sind als solche kenntlich gemacht.

Ich erkläre ferner, dass ich die vorliegende Arbeit an keiner anderen Stelle als Prüfungsarbeit eingereicht habe oder einreichen werde.

Ort, Datum

Unterschrift