

Inhaltsverzeichnis: Kapitel 5

- ◆ 5.1 Betriebliche Informationssysteme
 - 5.1.1 Management Informationssysteme
 - 5.1.2 Online-DB, Information Retrieval
 - 5.1.3 CAD-Systeme
- ◆ 5.2 Hypertext- und Multimediasysteme
 - 5.2.1 Hypertextmodell
 - 5.2.2 Multimediasysteme

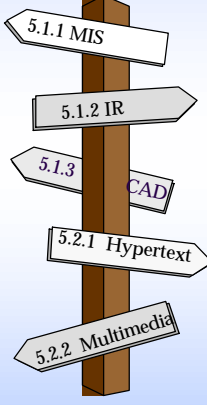



Bild 5 -1 Datenbank- und Informationssysteme (c) 10-99 Prof. Dr. F. Laux

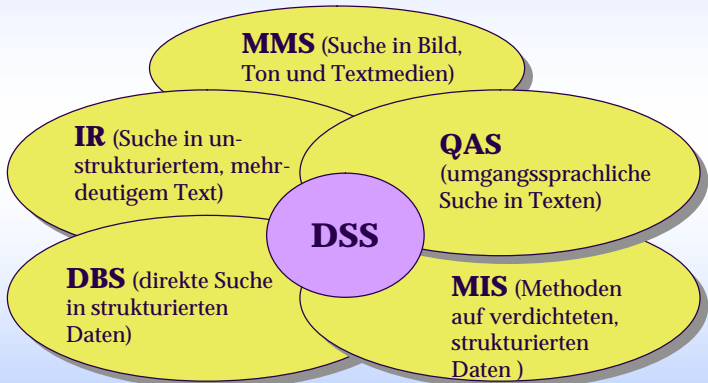
Die wichtigsten Anwendungen für Datenbanken finden sich bei betrieblichen Informationssystemen. Dazu gehören operative Systeme wie Auftrags- und Lagerverwaltung, Produktionsplanung, CAD (Computer Aided Design) und Management Information Systeme wie z.B. Prognosesysteme, DSS (Decision Support Systems) und IR (Information Retrieval).

Neuere Anwendungsbereiche finden sich in Form von Hypertext und Multimediasystemen, wobei sich hier vor allem objektorientierte Datenbanken anbieten. (warum?)



5 Klassifizierung von Informationssystemen

- ◆ **Definition:** Ein DB-basiertes System zur benutzerfreundl. Aufbereitung von Informationen (Daten & Bedeutung) heißt **Informationssystem**
- ◆ **Typisierung:**



- ◆ **Zielsetzung:** Befriedigung der Info-Bedürfnisse der Benutzer

Bild 5 -2
Datenbank- und Informationssysteme
(c) 10-99 Prof. Dr. F. Laux

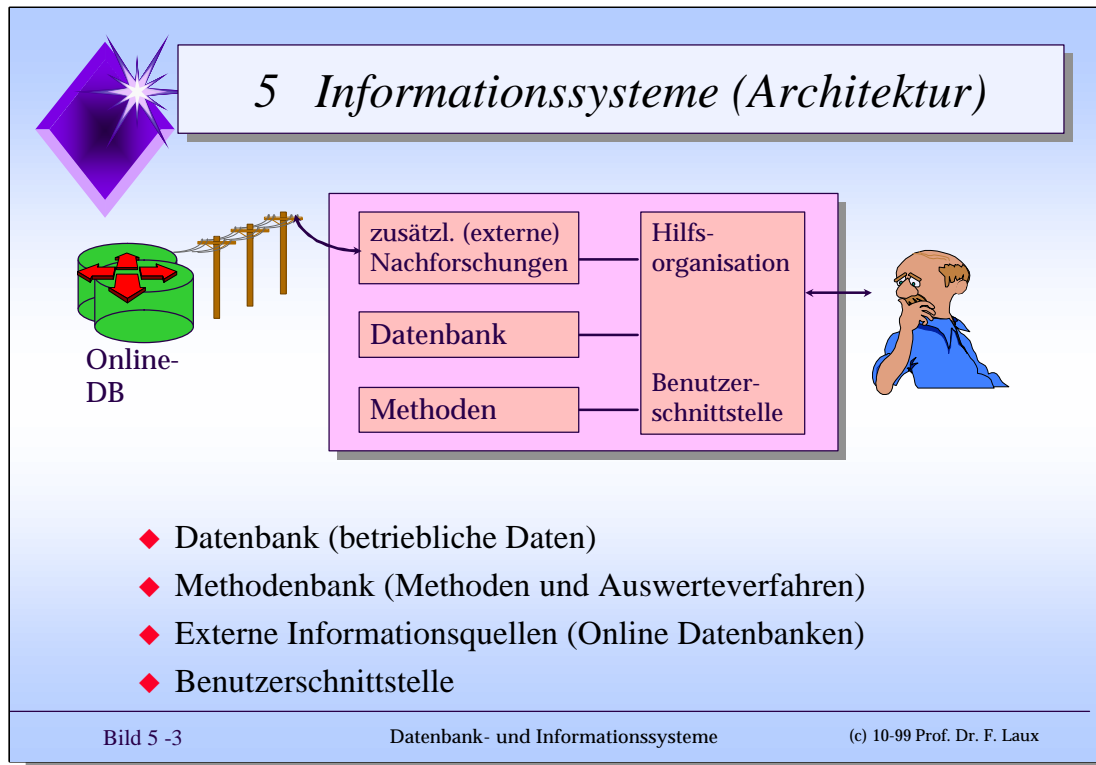
Es gibt sicherlich mehrere Möglichkeiten Informationssystem zu typisieren. Wir klassifizieren sie bezüglich ihrer Datenstruktur und ihrer Recherchemethoden.

Bei *Datenbanksystemen (DBS)* werden Informationen durch unmittelbare Suche in strukturierten Daten gewonnen.

Management Informationssysteme (MIS) durchsuchen dieselben Daten wie DBS, verdichten und gruppieren die Suchergebnisse zu Aggregaten wie Summen, Durchschnitte etc, um zu weitergehenden Erkenntnissen zu gelangen. Zu dieser Gruppe zählen auch Analysensysteme (*OLAP (Online Analytical Processing)*), die mit anwendungsspezifischen Verfahren (z.B. Zeitreihenanalyse) die Daten untersuchen und Prognosen liefern. In jüngster Zeit sind sogenannte *Data Warehouses* mit *Data Mining* (navigierender Datenzugriff) in Mode gekommen. Diese Systeme versprechen dem Benutzer beim Durchlaufen (Navigation) des Datenbestandes verborgene Zusammenhänge durch geeignete Darstellung (z.B. mehrdimensionale Datenwürfel) zum Vorschein zu bringen. Im Gegensatz zu traditionellen Informationssystemen muß bei Data Mining keine präzise Anfrage in Form einer Query an das System gerichtet werden. Die Informationen ergeben sich aus dem Weg auf dem die Daten durchlaufen wurden.

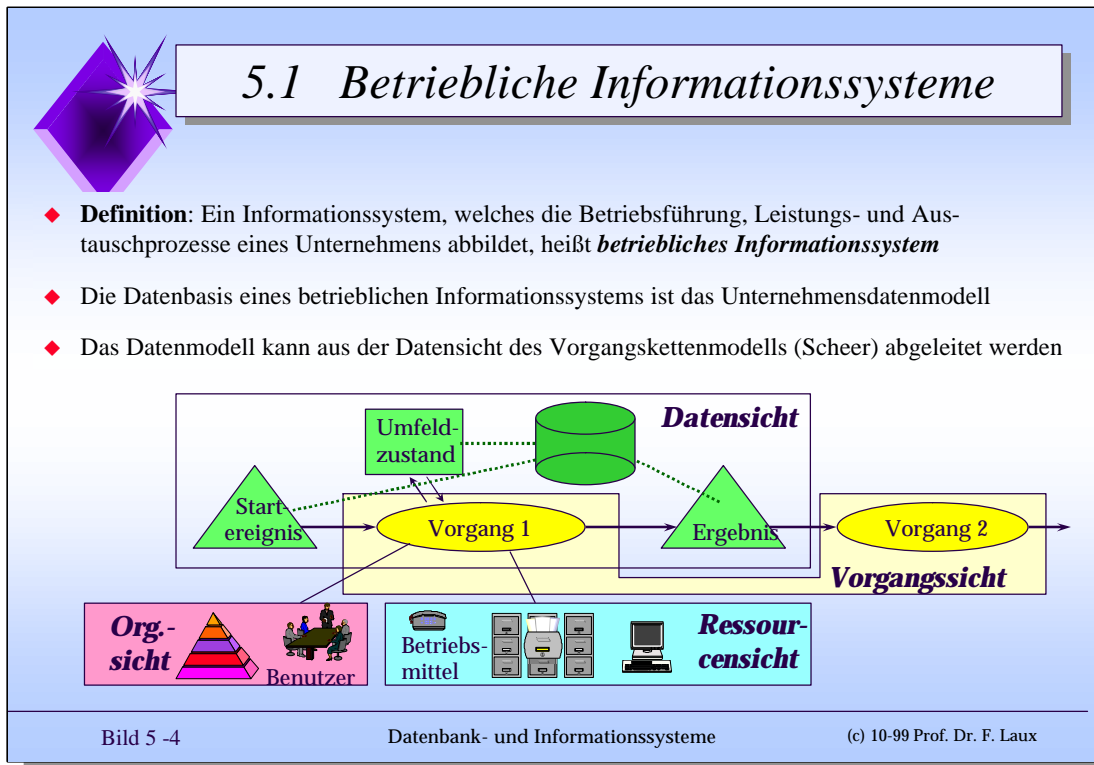
Information Retrieval (IR) Systeme und auch natürlichsprachige Recherchesysteme (*QAS, Question and Answer Systems*) verwenden gewöhnliche Textquellen (Prosa, Tabellen, etc.) zu ihrer Informationsbeschaffung. Dabei sind Mehrdeutigkeiten, Schreibvarianten, Kontextabhängigkeit etc. zu berücksichtigen. Approximative Suchverfahren sind dazu notwendig. Bei IR Systemen wird mit einer künstlichen Recherchesprache gearbeitet, während bei QAS Abfragen in Umgangssprache formuliert werden und das System diese dann interpretieren muss. Multimediasysteme (*MMS*) dehnen die Informationen auch auf Ton- und Bildträger aus. Dazu sind aufwendige Sprach-, Musik- und Bilderkennungsalgorithmen erforderlich.

Bei entscheidungsunterstützenden Systemen (*DSS, Decision Support Systems*) finden sich alle Kombinationen der vorgenannten Systeme.



Die Basis von Informationssystemen bilden Datensammlungen (betriebliche Datenbanken, externe Daten). Diese Datenquellen verwenden verschiedene Strukturen und Semantiken für ihre Informationen. Dies erschwert eine gemeinsame Nutzung, denn die Daten müssen aufbereitet und vergleichbar gemacht werden. Erst dann können sie nach verschiedenen Methoden ausgewertet werden. Die verfügbaren Methoden sind häufig mit einer Datenbank assoziiert, also nicht global einsetzbar. Deshalb ist die Datenkonversion und Aufbereitung eine wichtige Voraussetzung. Dazu gehört auch das Bereinigen inkonsistenter Daten, die Konvertierung in gleiche Dimensionen, Standardwerte (Default-Werte) und die Verdichtung zu adäquaten Einheiten.

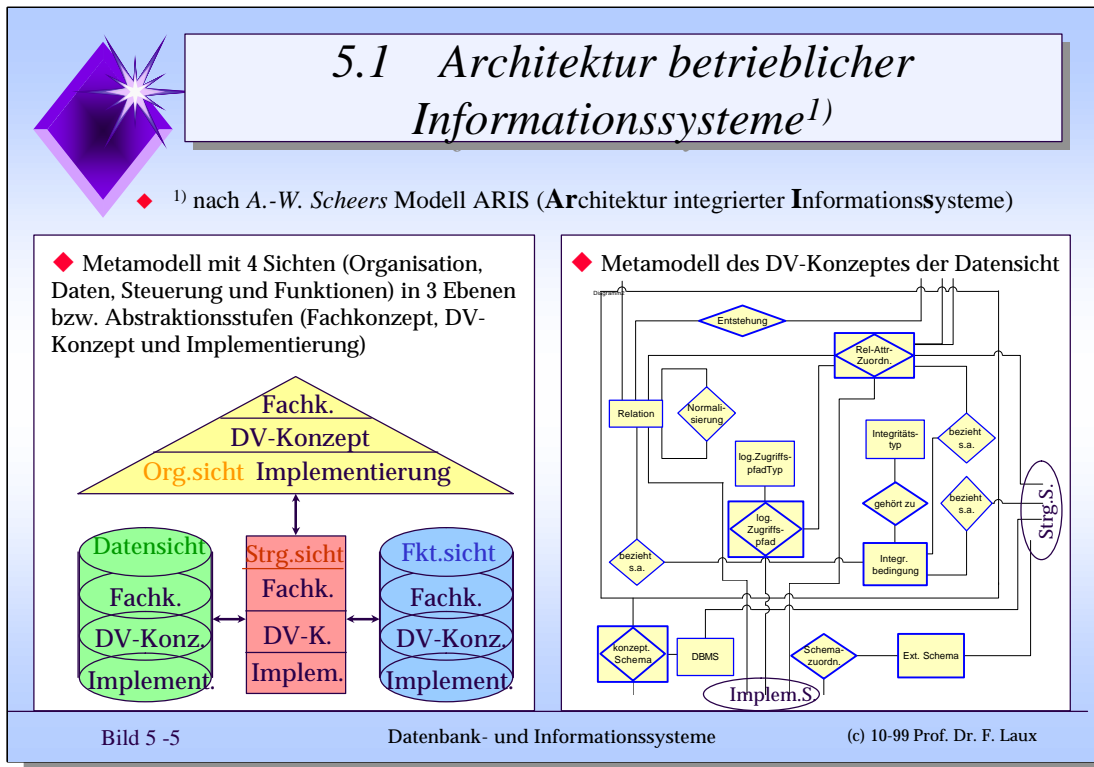
Als Auswertemethoden kommen meist mathematische Verfahren (Statistik, Zeitreihenanalyse, Extrapolationsverfahren) zum Einsatz, um quantifizierbare Analysen und Prognosen durchzuführen. Für die Spezifikation der Auswertungen und die Präsentation der Ergebnisse sind besonders graphische Benutzerschnittstellen geeignet. Sie gestatten eine einfache Handhabung und eine optisch eindrucksvolle Präsentation.



Das hier dargestellte Modell eines betrieblichen Informationssystems ist in vier Sichten gegliedert. Das Datenmodell wird aus den betrieblichen Vorgängen (Geschäftsprozesse) und den dabei zu verarbeiteten Daten hergeleitet.

Das Vorgangskettenmodell versteht eine Unternehmung als eine Menge von Vorgangsketten, welche die Geschäftsprozesse abwickeln. Eine Vorgangskette ist eine lineare Folge von Ereignissen und Vorgängen. Dabei wechseln sich Ereignis und Vorgang ständig ab. Eine Vorgangskette beginnt mit einem Ereignis und endet mit einem Ergebnis (auch ein Ereignis).

Vorgänge operieren auf Daten. Ein Vorgang wird durch eine Organisation gestützt, welche die Vorgangskette definiert und realisiert. Technisch wird ein Vorgang durch den Einsatz von Betriebsmitteln (z.B. Datenbanken) umgesetzt.



Wir präsentieren hier ein Metamodell für die Architektur integrierter Informationssysteme (ARIS, entwickelt von Prof. Scheer (TH Aachen)). Im Modell werden die Elemente eines integrierten, betrieblichen Informationssystems definiert und ihre Struktur dargestellt. Es verwendet drei Abstraktionsebenen (Fachkonzept, DV-Konzept und Implementierung) für jede der vier Sichten auf die Architektur.

Als Darstellungsmittel wird das Entity-Relationship-Modell verwendet. Im Beispiel oben wird das Fachkonzept der Datensicht dargestellt. Es enthält die Objekte *DBMS*, *konzeptionelles Schema*, *Relation*, *Integritätsbedingung* und *-Typ*, *Relation - Attribut Zuordnung*, etc. Dies bedeutet, die Datensicht des Fachkonzepts benutzt diese Objekte als Beschreibungsmittel in der dargestellten Art und Weise.

Was läßt sich aus diesem Modell herauslesen?


Zum Beispiel,

- daß sich Integritätsbedingungen immer auf Relationen beziehen (leider fehlen die Beziehungskardinalitäten im Metamodell),
- oder wenn eine Relation normalisiert wird, entstehen wieder Relationen.
- Einer Integritätsbedingung ist immer ein Typ zugeordnet.

All dies läßt sich aus dem Modell entnehmen.

Wozu dienen all diese Modelle und Architekturen?

Sie sollen die Komplexität reduzieren, indem sie Abstraktionen und Teilaspekte (Sichten) anbieten. Dadurch wird das Verständnis dieser Systeme erleichtert. Durch die Strukturierung fällt es auch leichter, solche System zu entwickeln, da bei der Analyse die gleichen Strukturierungshilfen verwendet werden können, um fachliche Objekte (business objects) zu finden.



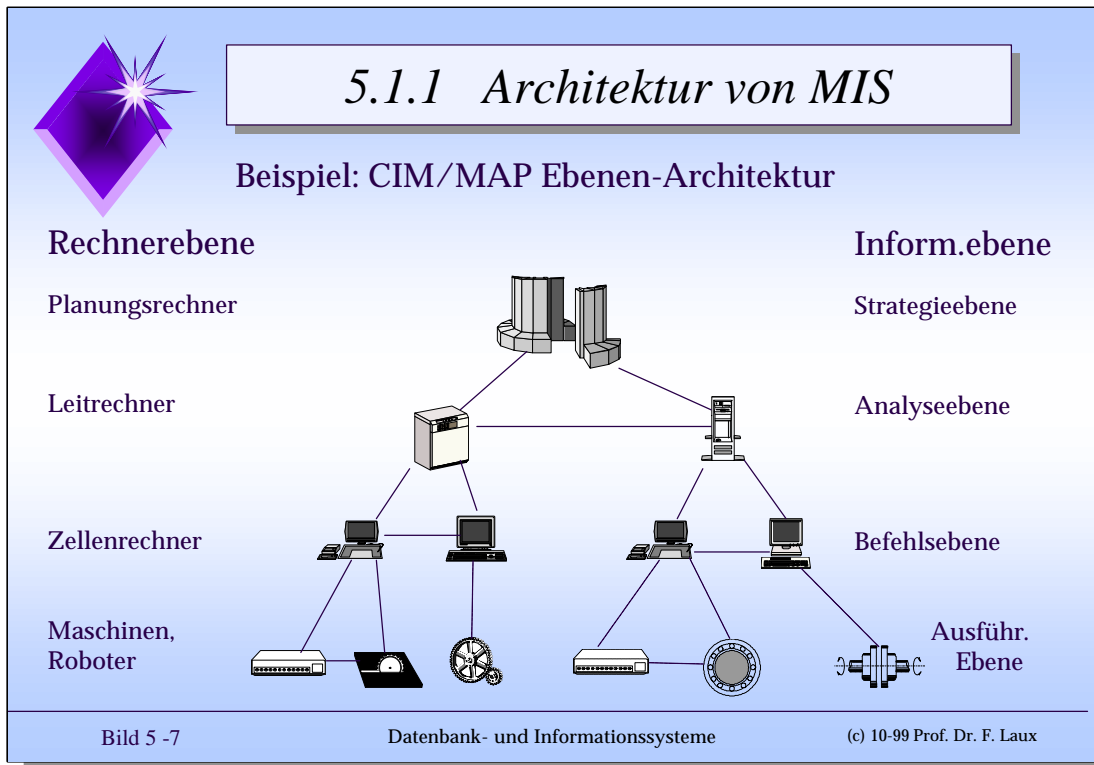
5.1.1 Management Informationssysteme

- ◆ Ebenen des Management und Abfragetypen (nach J. Martin)
 - **Einfache Abfragen** (Daten) für Routinevorgänge
 - ▲ Antwort ist direkt als *Datum* in der DB vorhanden
 - ▲ kann vollautomatisiert werden (z.B. BDE, Auftrags- und Lagerverwalt.)
 - **Planung und Prognose** (Daten & Methoden)
 - ▲ Datenauswertung mit Hilfe e. Modells (*Information*)
 - ▲ Vorgänge, die Denken und Erfahrung voraussetzen
 - ▲ können durch Computerunterstützung teilautomatisiert werden
 - **Strategische Anfragen und Planung**
 - ▲ Auswertung eigener und externer Information (*Nachforschungen*)
 - ▲ Bewertung der Ergebnisse (intellektuelle Tätigkeit)
 - ▲ nachprüfbare Schlußfolgerungen (KI-Systeme)

Bild 5 -6 Datenbank- und Informationssysteme (c) 10-99 Prof. Dr. F. Laux

Informationssysteme lassen sich nach J. Martin - etwas vereinfacht - auch nach der Komplexität ihrer Anfragen und der Verwendung der Ergebnisse klassifizieren:


- SQL-Queries gehören nach dieser Einteilung zu den einfachen Abfragen, die auf operativer Ebene (Sachbearbeiter) eingesetzt werden.
- Werden für die Auswertung Modelle (z.B. Zeitreihen) eingesetzt, so gehören sie zu den Planungs- und Prognoseanfragen der mittleren Managements.
- Vorstandsbereich und Top Management befassen sich mit strategischen Anfragen, die zusätzlich durch externe Informationen gestützt werden müssen.



Die zuvor behandelte Abfrage- und Managementebenen spiegeln sich gelegentlich auch in der Hierarchie des Rechnernetzwerkes einer Unternehmung wider. Die dargestellte Struktur entspricht der bei General Motors und Boeing entwickelten Architektur für MAP/TOP Protokolle. (MAP = Manufacturer Automation Protocol, TOP = Technical Office Protocol)

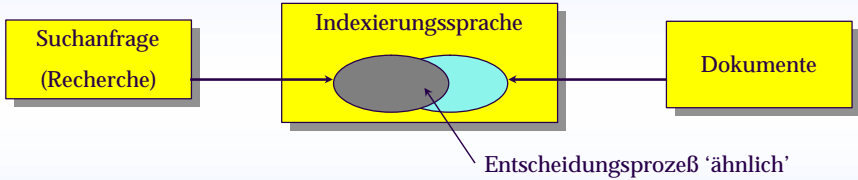
An der Konzernspitze stehen *Planungsrechner* für strategische Aufgaben. Die Daten dafür beziehen diese Systeme von *Leitrechnern*, welche den Workflow steuern und Daten der *Zellenrechner* verdichten und auswerten. Die *Zellenrechner* wiederum geben *Maschinen und Robotern* Befehle, die diese in Echtzeit ausführen.

Die Vernetzung erfolgt hierarchisch und zwischen Rechnern der gleichen Ebene.



5.1.2 Information Retrieval (IR) Systeme

IR := Suche, Zugriff und Darstellung von Dokumenten
(Informationen) in Textform



- ◆ Indexierungssprache := Menge der möglichen Schlagwörter (Deskriptoren)
- ◆ Annahme: Jedes Dokument kann durch Indexliste (Vektor) 'erschlossen' werden
- ◆ Kontrollierter Index := fest vorgegebene Deskriptorenliste (Thesaurus)
- ◆ Unkontrollierter Index := Deskriptoren stammen aus dem Dokument selbst

Bild 5 -8 Datenbank- und Informationssysteme (c) 10-99 Prof. Dr. F. Laux

Information Retrieval Systeme unterscheiden sich von den bisher besprochenen Datenbanksystemen vor allem durch ihre spezifische Datenstruktur und ihre *approximativen Queries*. Die Suche in „unstrukturiertem“, *mehrdeutigem Text* stellt ganz andere Anforderungen als die exakte Suche in formatierten Daten:

formatierte Daten	Textdokumente
<ul style="list-style-type: none"> • Daten sind semantisch präzise festgelegt • Abfragen nach Daten sind exakt formulierbar 	<p>die freien Formulierungen in Textdokumenten sind verschieden interpretierbar.</p> <p>eine exakte Suche in kontextabhängigen Texten liefert keine zufriedenstellenden Ergebnisse, da das gewünschte Stichwort in einem relevanten Text nicht notwendigerweise vorkommen muß</p>

Deskriptoren und Thesauri stellen eine Möglichkeit dar, den Inhalt eines Textes zu erschließen und ihn Abfragen zugänglich zu machen.

Deskriptoren beschreiben den Inhalt eines Dokuments in einer festgelegten Weise, so daß durch eine Abfrage, welche ihre Suchbegriffe einem *Thesaurus* (Fachwörterbuch) entnimmt, ein definierter Bezug zwischen Dokument und Anfrage hergestellt werden kann. Je mehr Suchbegriffe mit den Deskriptoren eines Textes übereinstimmen, desto zutreffender ist dieser für die Abfrage. Mit Hilfe eines Thesaurus läßt sich auch ein Maß (Metrik) der *Ähnlichkeit* von Begriffen definieren.

Diese spezifischen Anforderungen zeigen, daß für Textdatenbanken ein neues Datenmodell notwendig wird. Wir stellen im folgenden ein Modell für das Information Retrieval und zur Erschließung von Texten vor.




5.1.2 Indexierung

- ◆ Vor Aufnahme eines Dokuments in ein IR-System ist eine **Indexierung** erforderlich
- ◆ Indexierung := formale und inhaltliche Beschreibung eines Dokuments; dazu gehören:
 - formale Angaben (z.B. Autor, Verlag, Datum) (= **alphabetischer Katalog**)
 - inhaltliche Beschreibung mit Hilfe von Schlagwörtern (Deskriptoren) (= **systematischer Katalog**)
- ◆ Deskriptoren entstammen
 - dem Dokument selbst, oder (= **unkontrollierte Indexsprache**)
 - einem Fachwörterbuch (Thesaurus) (= **kontrollierte Indexsprache**)
- ◆ Durchführung der Indexierung auf Basis einer **Textanalyse**

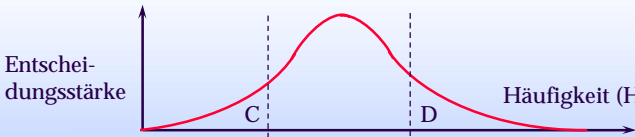
Bild 5 -9 Datenbank- und Informationssysteme (c) 10-99 Prof. Dr. F. Laux

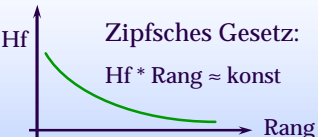
Zur Erfassung des Inhaltes eines Dokuments muß eine *Indexierung* durchgeführt werden. Diese besteht aus *formalen Angaben* (Autor, Datum, etc.) und einer *inhaltlichen Beschreibung* durch relevante Stichwörter, sogenannte *Deskriptoren*. Wird die *Textanalyse* maschinell vorgenommen, so stammen die Deskriptoren normalerweise aus dem Dokument selbst. Bei einer manuellen Analyse entstammen die Stichwörter meist einem Thesaurus.



5.1.2 Textanalyse

- ◆ Durchführung der Textanalyse
 - *manuell* durch Fachexperten
 - *automatisch* mit Hilfe eines Computers
- ◆ automatische Textanalyse basiert auf einer Häufigkeitsanalyse
 - Füllwörter durch Stoppliste ausblenden
 - seltene (unbedeutende) Begriffe eliminieren
 - Wörter aus Text mit hoher Entscheidungsstärke herausfiltern (Deskriptoren)
[Gute Deskriptoren sind meist Begriffe mittlerer Häufigkeit]





Zipfsches Gesetz:
 $Hf * Rang \approx konst$

- ◆ Schwierigkeit: Bestimmung der Schwellwerte C und D

Bild 5 -10
Datenbank- und Informationssysteme
(c) 10-99 Prof. Dr. F. Laux

Die automatische Textanalyse basiert auf Häufigkeitsuntersuchungen.

Füll- und Bindewörter, die häufig vorkommen, werden als unbedeutend ignoriert, ebenso sehr seltene Begriffe. Übrig bleiben Begriffe mittlerer Häufigkeit, von denen man annimmt, daß sie eine hohe *Aussagekraft* (*Entscheidungsstärke*) haben.

Werden alle Wörter eines Textes nach ihrer Häufigkeit (Hf) in Form einer Rangliste angeordnet, so kann man feststellen, daß $Hf * Rang$ konstant sind (Zipfsches Gesetz). Dies läßt sich zu folgender Formel präzisieren:

Sei x ein Wort in einem Text, dann gilt:

$$Hf(x) \approx Hf(\text{häufigstes Wort}) / Rang(x)$$


Daraus ergeben sich die folgende Schlußfolgerungen:

Wenn N die Häufigkeit des Wortes mit Rang 1 (häufigstes Wort) ist, dann ist N das Vokabular (Anzahl verschiedener Worte) in einem Text. Die Anzahl T der Wörter dieses Textes läßt sich abschätzen zu:

$$T \approx N + N/2 + N/3 + \dots + 1.$$

Beispiel: Text mit 50 verschiedenen Wörtern ist ca. $225 \approx 50 + 25 + 16.67 + 12.5 + \dots + 1$ Worte lang. Dabei ist 50 die Häufigkeit von Rang 1; 25 die Häufigkeit von Rang 2 und der letzte Rang (50) hat die Häufigkeit 1.

Siehe hierzu auch die Studie von G.R. Turner: Relationship Between Vocabulary, Text Length and Zipf's Law; <http://www.btinternet.com/~g.r.turner/ZipfDoc.htm>.




5.1.2 Online-Datenbanken

- ◆ Definition
 - Online: aktive Verbindung zur externen Datenbank
 - Online-Datenbank: von einer externen Organisation (Provider) betriebene Datenbank
- ◆ Typisierung von Online-Datenbanken
 - Text-DB (Volltext-, Bibliographische DB)
 - Fakten-DB (Statist. DB (Zeitreihen), Fakten-, Modell-DB)
 - integr. DB ("Text+Fakten+Tabellen", Hypertext- u. Multimedia)
- ◆ sonstige Typisierung
 - nach Speichertechnik (satzorientiert, nach Datentypen)
 - klassisch (Referenz-DB, Quell-DB)


Bild 5 -11
Datenbank- und Informationssysteme
(c) 10-99 Prof. Dr. F. Laux

Online Datenbank ist die Bezeichnung für Datensammlungen (Texte, Tabellen, neuerdings auch Audio- und Videoinformationen), die von einer externen Organisation (bezogen auf den eigenen Betrieb) eingerichtet, gepflegt und betrieben wird. Online Datenbanken sind i.d.R. kostenpflichtig. Der Zugang wird über *öffentliche Datenübertragungsnetze* insbesondere über das *Internet* hergestellt. Der Begriff "Online" bezeichnet in diesem Zusammenhang eine aktive Verbindung zur Datenbank während einer Abfrage. Damit soll die Aktualität der Daten unterstrichen werden. Viele dieser Datenbanken stammen aus der Zeit vor dem WWW, so daß gelegentlich noch eine zeichenorientierte Benutzeroberfläche (z.B. Telnet) mit einer kryptischen Abfragesprache benutzt werden muß. Zunehmend bedient man sich jedoch der Web-Browser als Benutzerschnittstelle.

Die oben angegebene *Typisierung* ist inhaltsbezogen. Darüber hinaus sind auch Typisierung nach Speichertechnik und die klassische Einteilung in Referenz- und Quelldatenbanken üblich. *Referenzdatenbanken* enthalten nur Verweise auf Dokumente; *Quelldatenbanken* hingegen enthalten die Dokumente selbst.



5.1.2 Beispiel einer Index-Recherche in einer Textdatenbank

- ◆ **Kontrollierte Indexsprache:** {Lärm, Umwelteinfluß, Bewertungsstudie, Wohnumwelt, Hörschaden, Kfz, Ozon, Luftverschmutzung}
- ◆ **Dokument:** Studie über Umweltbelastung durch Lärm
- ◆ **Doku-Index:** (1,1,1,0,1,1,0,1) (Vektor)
- ◆ **Suchanfrage:** Welche Hörschäden werden durch Lärm verursacht ? 
- ◆ **Suchindex:** (1,0,1,1,1,0,0,0)

◆ Suche nach Dokumenten, die eine hohe Übereinstimmung mit dem Suchindex aufweisen. => Korrelation für o.g. Dokument = 3 (Indexwörter)

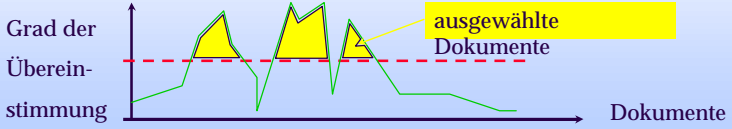


Bild 5 -12
Datenbank- und Informationssysteme
(c) 10-99 Prof. Dr. F. Laux

Das Beispiel soll die Mechanismen einer Indexsuche verdeutlichen.

Es sei eine kontrollierte Indexsprache (z.B. Thesaurus) gegeben. Die Indexsprache kann zur besseren Vorstellung als geordnete Liste betrachtet werden. Die Dokumente seien entsprechend indiziert.

Wir suchen eine *Studie über Umweltbelastung durch Lärm*. Aus dieser umgangssprachlichen Anfrage wird ein Abfragevektor (Suchindex) gebildet, indem die Anfrage mit relevanten Stichwörtern aus unserer Indexierungssprache assoziiert werden. Der Index (Such- oder Doku-Index) ist ein Vektor, bei dem jede Vektorkoordinate auf einen Deskriptor der Indexsprache zeigt. Eine "1" bedeutet, daß der Deskriptor zu Anfrage (bzw. zum Dokument) gehört, eine "0" zeigt an, daß er nicht darin vorkommt. Dieser Suchvektor wird mit den Deskriptoren der Dokumente verglichen. Dokumente welche eine Übereinstimmung aufweisen, die oberhalb eines wählbaren Schwellwertes liegen, werden als Ergebnis der Abfrage ausgegeben.



5.1.2 Volltextoperatoren


- ◆ logischen Verknüpfungsoperatoren (AND, OR, NOT, XOR)
- ◆ zusätzliche **Textoperatoren**
 - <a> SAME (<a> und im gleichen Absatz)
 - <a> NOT SAME (nicht im selben Absatz)
 - <a> WITH (im gleichen Satz)
 - <a> ADJ (<a> und hintereinander (adjacent))
 Beispiel: Suchanfrage (Information ADJ Retrieval) OR IR\$ findet 'Information Retrieval', 'IR', 'IR-System', etc., jedoch nicht 'Information'
 - <a> NEAR (folgt <a> innerhalb e. best. Abstands)
- ◆ **Vergleich, Ähnlichkeit**
 - Maskierung ("Joker", "Wildcards") (?, \$ (belieb. Zeichenf.), % (Ziffern))
 - Signatur (Bit-Vektor) (quantifizierbare Ähnlichkeit, Thesaurus)

Bild 5 -13
Datenbank- und Informationssysteme
(c) 10-99 Prof. Dr. F. Laux

Volltextabfragen erfordern besondere Operatoren für Textsuche und -vergleich.

Neben den logischen Operatoren, wie sie von strukturierten Datenbanken bekannt sind, werden Textoperatoren angeboten, die auf Textbausteinen (Satz, Absatz, Abschnitt, Kapitel, etc.) wirken. Innerhalb solcher Textbereiche können Lokationen bestimmt und Abstände bzw. Relationen zwischen Begriffen festgelegt werden. Eine besondere Notation wird auch für Wortvergleiche und Begriffsähnlichkeit benötigt, da die Suche normalerweise mit unscharfen Vergleichen arbeitet. Beispielsweise kann festgelegt werden, ob Groß- und Kleinschreibung relevant sind, ob exakt ("") gesucht werden soll, ob bestimmte Wörter ausgeschlossen werden sollen (-) oder ob ein bestimmtes Wort vorkommen muß (+).

Sie auch URL: http://www.altavista.digital.com/av/content/help_simple.htm für die Möglichkeiten der AltaVista Suchmaschine.



5.1.3 CAD-Systeme

- ◆ **Charakteristika des Datenmodells**
 - komplexe Datentypen (Strukturen) und Beziehungen, typspezifische Operationen
 - idealer Anwendungsbereich für objektorientierte Datenbanken
- ◆ **Präsentation**
 - grafische Darstellung der Objekte (Konstruktionselemente) und der Beziehungen (z.B. räumliche Position)
 - verschiedene Sichten auf die Konstruktion (z.B. Schaltung, Layout, Funktionslogik, zeitlicher Signalfluß)
- ◆ **Besonderheiten der DB-Funktionen**
 - "lange" Transaktionen, geschachtelte Transaktionen
 - Versionen


Bild 5 -14 Datenbank- und Informationssysteme (c) 10-99 Prof. Dr. F. Laux

Neben Textdatenbanken ist Computer Aided Design (CAD) ein weiteres Beispiel für eine Nichtstandard-Datenbankanwendung. CAD Daten sind so komplex strukturiert, daß klassische Datenbanken (hierarchisch, Netzwerk oder relational) hierfür nicht geeignet sind.

Welche Anforderungen stellen CAD-Systeme an eine Datenbank?

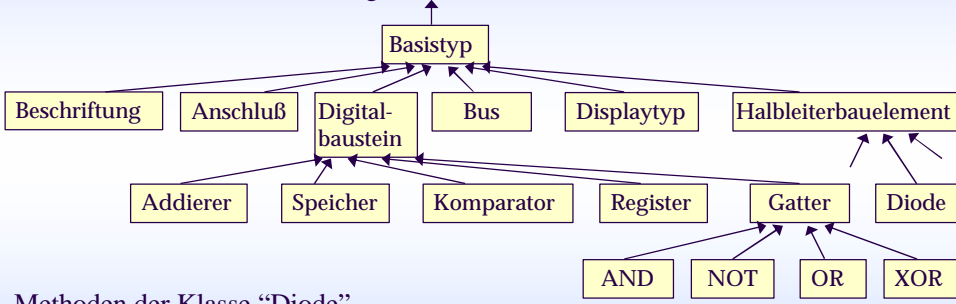
- Es müssen komplexe und benutzerdefinierte Datentypen (z.B. Zahnrad) und typspezifische Operationen (z.B. rotieren) verarbeitet werden
- Die Darstellung der Daten für den Benutzer muß graphisch (z.B. technische Zeichnung) erfolgen
- Unterstützung für lange und geschachtelte Transaktionen (z.B. Änderung der Motorkonstruktion) ist notwendig
- Datenbankobjekte werden in verschiedenen Versionen benötigt (die Versionsverwaltung muss mehrere Varianten speichern und pflegen)

Diese Anforderungen deuten darauf hin, daß objektorientierte Datenbanken dafür besonders geeignet sind. Wir wollen dies am Beispiel des Entwurfs von elektronischen Schaltungen näher untersuchen.



5.1.3 Beispiel eines CAD Datenmodells

◆ Bauteilehierarchie im Schaltungsentwurf (Klassenhierarchie)



```

graph TD
    Basistyp --> Beschriftung
    Basistyp --> Anschluß
    Basistyp --> Digitalbaustein
    Basistyp --> Bus
    Basistyp --> Displaytyp
    Basistyp --> Halbleiterbauelement
    Digitalbaustein --> Addierer
    Digitalbaustein --> Speicher
    Digitalbaustein --> Komparator
    Digitalbaustein --> Register
    Digitalbaustein --> Gatter
    Digitalbaustein --> Diode
    Halbleiterbauelement --> AND
    Halbleiterbauelement --> NOT
    Halbleiterbauelement --> OR
    Halbleiterbauelement --> XOR
  
```

◆ Methoden der Klasse "Diode"

- Modell ändern (U/I-Kennlinie, Sperrschichtkapazität C/U)
- Parameter ändern (Temperaturkoeffizient, Ptot, Pnenn, maxSperrspannung, etc.)
- Diode zeichnen, positionieren, verbinden, ...
- Strom berechnen (U), Spannung berechnen (I), Zeitanalyse, ...

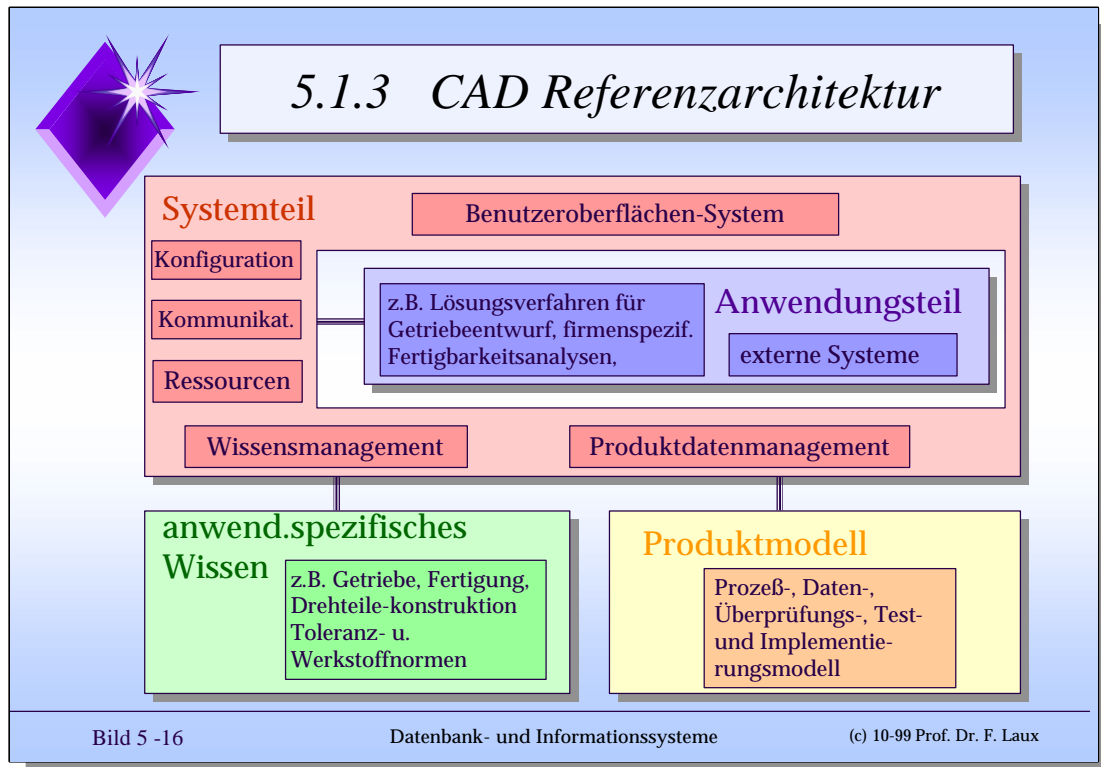
Bild 5 -15
Datenbank- und Informationssysteme
(c) 10-99 Prof. Dr. F. Laux

Betrachten wir zunächst die Bauteile der Elektronik. Sie fügen sich leicht in ein hierarchisches Klassifizierungsschema. In vielen Fällen lassen sich Funktionen vererben. Beispielsweise haben alle Digitalbausteine der TTL Serie 7400 die gleichen Stromversorgungs-, In- und Outputspezifikationen, unabhängig davon, ob es sich um Register, Addierer, AND oder OR-Gatter handelt.

Jedes Bauteil besitzt spezifische Eigenschaften (Kennlinie, Parameter, etc.) und Funktionalität (Stromfluß durch eine Diode, Signalverstärkung eines Transistors, etc.), aber auch allgemeine. Die allgemeinen Eigenschaften lassen sich im Basistyp unterbringen, damit sie an alle Bauteile vererbt werden. Dazu gehört zum Beispiel:

- jedem Bauteil ist ein Symbol zugeordnet, das gezeichnet werden kann
- die Bauteile müssen sich zu einer Schaltung verbinden lassen.

Wie die Beziehungen der Bauteile innerhalb einer Schaltung aussehen oder ob ein Schaltungslayout kreuzungsfrei ist, sind Fragestellungen, die mit traditionellen Datenbankstrukturen nur unzulänglich beantwortet werden können.



Einem CAD System liegt auf Grund seiner Gesamtkomplexität ein Architekturmodell mit vielen Facetten zugrunde. Die verschiedenen Bereiche lassen sich nur schwer isolieren (dies wird durch die verzahnten Schnittstellen deutlich), trotzdem können vier Bereiche identifiziert werden:


Der **Systemteil** enthält die Benutzeroberfläche und alle Konfigurations-, Steuerungs- und Kommunikationsfunktionen. Er besitzt Schnittstellen zu allen übrigen Teilsystemen.

Der **Anwendungsteil** umfaßt anwendungsspezifische Analyse- und Lösungsverfahren für Entwurfsaufgaben.

Zum **Produktmodell** gehören Modelle für Produktionsverfahren, Überprüfungs- und Testmodelle.

Bauteilewissen und Normen bilden das **anwendungsspezifisches Wissen**. Beispiele hierfür sind Werkstoff- und Toleranznormen.

Das gezeigte Referenzmodell ist in *O. Abeln, CAD-Referenzmodell* (siehe Literaturliste zu Kap. 5) detailliert beschrieben.



5.2.1 Hypertext-Systeme

- ◆ Definition
 - Ein Hypertextsystem ist ein Datenbanksystem, in dem Wissen in *nichtlinearer* Form dargestellt und manipuliert werden kann
- ◆ Grundkonzept
 - Die Struktur eines Hypertextes ist netzartig (Knoten und Kanten (Links))
 - Knoten := Informationseinheiten (Text, Graphik, AV-Information)
 - Link := Verknüpfungen zweier Knoten, Navigation durch "Verfolgen" von Links
- ◆ Normaler Text

... Hypertext ist eine nichtlineare¹⁾ Form der Wissensrepräsentation²⁾

.....

1) Hypertext realisiert n:m (n,m > 0) Beziehungen
2) Wissen ist Erfahrungsgut
- ◆ Hypertext

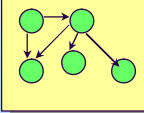
... Hypertext ist eine nichtlineare Form der Wissensrepräsentation

.....

→ Knoten

→ Link

Hypertext realisiert n:m (n,m > 0) Beziehungen



Wissen ist Erfahrungsgut, ...

Bild 5 -17 Datenbank- und Informationssysteme (c) 10-99 Prof. Dr. F. Laux

Normaler (linearer) Text kann in Texteinheiten (z.B. Sätze, Abschnitte, Paragraphen, Kapitel, Fußnoten, etc.) gegliedert werden. Wenn diese Einheiten zu einer netzartigen Struktur verbunden werden und entlang dieser Netzstruktur durchsucht werden können, spricht man von einem Hypertext.

Mathematisch gesehen ist ein Hypertext ein *Graph* mit Informationseinheiten (z.B. Text, Bild, etc.) als *Knoten* und *Kanten* als Verbindungen zwischen den Knoten. Die Kanten eines Hypertextes werden als Links bezeichnet, weil sie als Verweis auf weitere Informationen verwendet werden. Hypertext wird technisch mit Hilfe einer Beschreibungssprache (z.B. HTML (HyperText Markup Language), XML (Extensible Markup Language)) erstellt. Dabei werden einzelne Informationseinheiten definiert und ihr Erscheinungsbild beschrieben. XML ist im Gegensatz zu HTML durch den Benutzer erweiterbar, indem neue Informationseinheiten mit Hilfe der DTD (document type definition) definiert werden, d.h. DTD beschreibt die Struktur von Dokumenten (Metainformation). Um den Hypertext sichtbar zu machen, werden Hypertext Viewer verwendet. XML und HTML als eine Ausprägung davon sind für den Einsatz in Netzwerken geeignet, da die Informationseinheiten an beliebigen Orten im Internet gespeichert sein können. Ein Verweis (Link) auf die gewünschte Information reicht aus, um dorthin 'navigieren' zu können. Diesen Umstand machen sich die HTML-basierten Internetbrowser (Hypertext-Viewer) zunutze, um bequem durch das World Wide Web zu 'surfen'.

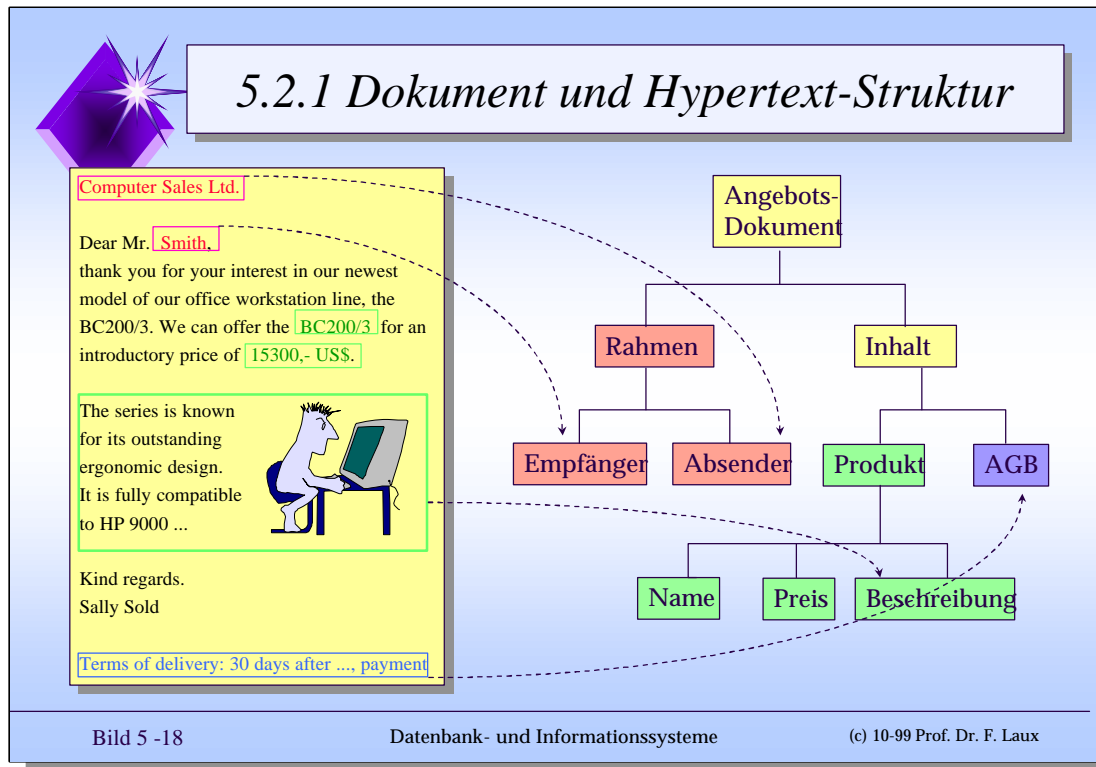


Bild 5 -18

Datenbank- und Informationssysteme

(c) 10-99 Prof. Dr. F. Laux

In diesem Beispiel wird gezeigt, wie ein Angebot als Hypertext strukturiert werden kann. Die Informationseinheiten (Knoten) können wiederum eine Struktur besitzen. Ein Knoten kann natürlich nicht nur Text, sondern auch Bild, Ton oder multimediale Elemente enthalten (z.B. der Knoten *Beschreibung* in unserem Beispiel). Für Angebote im Internet ist eine derartige Strukturierung schon aus Gründen der Übersicht notwendig.


Das vorangegangene Beispiel als XML-Dokument:

(Syntax farblich hervorgehoben und hierarchische Struktur durch Einrücken visualisiert)

```

<?xml version="1.0"?>
<Dokument Typ="Angebot">
  <Rahmen>
    <Empfaenger>
      Dear
      <Anrede>
        Mr.
      </Anrede>
      <Name>
        Smith
      </Name>
    </Empfaenger>
    <Absender>
      ...
    </Absender>
  </Rahmen>
  <Inhalt>
    thank you for your interest in our newest model of our office
    workstation line, the BC200/3. We can offer the
    <Produkt>
      <Name>
        BC200/3
      </Name>
      for an introductory price of
      <Preis>
        <Betrag>
          15300
        </Betrag>
        <Waehrung>
          US$
        </Waehrung>
      </Preis>
      <Beschreibung>
        The series is known for its outstanding ergonomic design. It
        is fully compatible to HP 9000 ...
      </Beschreibung>
    </Produkt>
    Kind regards.
    Sally Sold
    <AGB>
      Terms of delivery: 30 days after ..., payment ...
    </AGB>
  </Inhalt>
</Dokument>

```



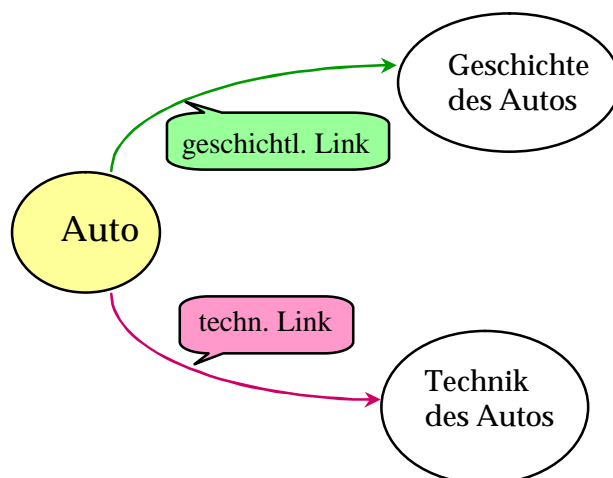
5.2.1 Elemente eines Hypertextes

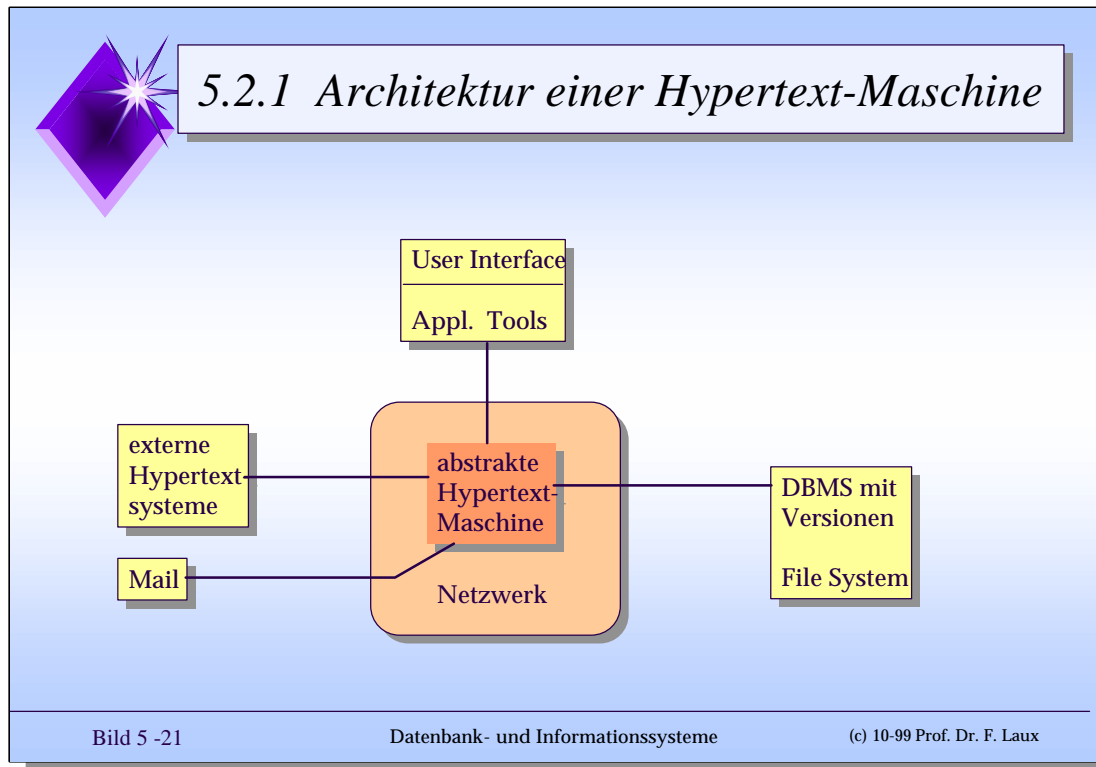
- ◆ Informationstyp
 - Klassifizierung (Datentyp) der Knoten und Links
 - TypId -> Typname, Typbeschreibung, Datum
- ◆ Knoten
 - Informationseinheit als Grundelement des Hypertextes
 - KnotenId -> TypId, Datum, Autor, Inhalt, Deskriptoren (Attribute)
- ◆ Link
 - gerichtete Verbindung zwischen 2 Knoten
 - LinkId -> TypId, Datum, Autor, Ausgangsknoten, Zielknoten, Deskriptoren (Attribute)
- ◆ Attribut
 - Zusatzinfo zu Knoten und Links, um Inhalt zu beschreiben ("erschließen")

Bild 5 -20
Datenbank- und Informationssysteme
(c) 10-99 Prof. Dr. F. Laux


Die formalen Elemente eines Hypertextes sind Knoten und Links. Beide enthalten Beschreibungen (Metainformationen), die Navigations und Darstellungshilfen für Browser darstellen, sowie formale Angaben über Autor, Datum, etc.. Leider wurde mit HTML (Hypertext Markup Language) die Metainformation nicht vom Inhalt getrennt. Dies erfordert spezielle HTML-Editoren und erschwert die Darstellung und Lesbarkeit der Information. Wenn ein neuer Informationstyp oder Deskriptorattribute hinzukommen, müssen Textquellen (Dokumente) und eventuell auch Programme geändert werden, um diese wirksam werden zu lassen. Soll z.B. eine Überschrift anders dargestellt werden, ist das Textdokument (HTML-Text) zu ändern.

Ein Link kann mit Hilfe eines Linkattributs (Metainformation) kategorisiert werden können.





Zentrum und physischer Ort einer abstrakten Hypertextmaschine ist das Netz. Diese paradoxe Aussage macht deutlich, dass es keine Lokalisierung eines Hypertextsystems geben kann. Es ist im Netzwerk verteilt. Das System ist mit vielen verschiedenen Ressourcen gekoppelt. Es kann durchaus vorkommen, daß die Grenzen der lokalen und der Netzressourcen verschwimmen. Umso wichtiger werden deshalb Sicherheitsmechanismen, die Schutz gegen alle unrechtmäßigen Aktivitäten im Internet bieten.



5.2.1 Funktionen einer Hypertext-Maschine


- ◆ Manipulation der Grundelemente
 - Funktionen zur Manipulation von Typen, Attributen, Knoten und Links
- ◆ Navigation
 - Aktiviere Knoten, Ermittle KnotenId,
 - Ermittle/verfolge ausgehende/eingehende Links
 - Ermittle Ziel/Quellknoten
- ◆ Filter
 - Definiere/entferne Knoten- und Linkfilter
- ◆ Orientierung
 - Setze Marke ("Lesezeichen")
 - Ermittle Historie ("Faden der Ariadne")
 - Guided Tour

Bild 5 -22 Datenbank- und Informationssysteme (c) 10-99 Prof. Dr. F. Laux

Die **Manipulation** der Hypertextelemente ist die Hauptfunktion für die Autoren eines Hypertextes.

Die **Navigation** („Internet klicken“, „surfen“) ist die wichtigste Tätigkeit eines Hypertextbenutzers.

Ein weiterer Schwerpunkt für Hypertextsysteme bildet die **Orientierung**. Da Hypertexte als ein Netzwerk von Informationen organisiert ist, sind Anfang und Ende einer Recherche unbestimmt. Das macht die Orientierung schwierig. Filter können die Netzstruktur im Prinzip dadurch vereinfachen, daß nur die für eine Recherche relevanten Links angezeigt werden. Allerdings, wer weiß schon im Voraus, welche Links (Klassifikation?) relevant sind? Lesezeichen (Bookmarks) helfen beim Wiederauffinden bereits besuchter Knoten. Die meisten Browser merken sich die Links einer Sitzung, so daß man jederzeit den Weg zurück findet (*Faden der Ariadne*). Einige Online-Anbieter helfen dem Benutzer bei bestimmten Themen mit einer *Guided Tour*.



5.2.1 Hypertextspezifische Probleme

- ◆ **Delinearisieren von Texten**
 - Texte sollen in logische Informationseinheiten zerlegt werden (Kapitel, Absatz, ...)
- ◆ **Generieren “linearer” Texte**
 - Widerspruch zum Hypertextsystem
 - Algorithmen für Zirkelbezug
- ◆ **“Lost in Hyperspace”**
 - zu viele Links (“Spagettitext”)
 - Gegenmaßnahmen: Orientierungskarte (Fisheye View), Lesezeichen, Historie, Guided Tour, Filter
- ◆ **Lernphase**
 - Information entsteht durch Suchprozeß
 - Startpunkt vom Leser bestimmt, Ziel für das System undefiniert

Bild 5 -23


Datenbank- und Informationssysteme

(c) 10-99 Prof. Dr. F. Laux

Häufig liegen Informationen bereits in linearer Form vor, wenn sie ins Internet aufgenommen werden sollen. Die Aufteilung und Delinearisierung eines Textes ist selten automatisierbar. Bücher eignen sich durch ihre hierarchische Gliederung (Kapitel, Paragraphen, Sätze) für eine automatische Bearbeitung. Allerdings werden dann die besonderen Möglichkeiten eines Hypertextes (z.B. die Netzstruktur) nicht genutzt.

Wenn man Hypertexte wieder auf einem Druckmedium darstellen will, müssen alle zuvor mühsam erstellten Informationseinheiten wieder linearisiert werden. Auch das ist meistens nicht ohne manuelles Zutun möglich. Beispielsweise können Schleifen auftreten, die aufgelöst werden müssen.

Auch die besten Orientierungshilfen nützen wenig, wenn die Links 'wie Spagettis' gesetzt werden. Dieses Herumirren im WWW nennt sich dann 'lost in hyperspace' oder 'surfen'?



5.2.2 Multimedia-Systeme

◆ **Definition:**
Ein System, das die **Integration** von Informationen aus **verschiedenen Medien** (Stand- und Bewegtbild, Ton und Text) realisiert.

◆ **Charakteristika**

- Einheitliche Benutzeroberfläche für alle Medien
- transparente Integration der Multimedia-Geräte
- semantisches Wissen über den Inhalt der verschiedenen Medien
- synchrone Aufnahme und Wiedergabe von Information
- Umsetzung von einem Medium in ein anderes


Bild 5 -24

Datenbank- und Informationssysteme

(c) 10-99 Prof. Dr. F. Laux

Multimediasysteme unterscheiden sich von Hypertextsystemen weniger durch die hinzukommenden Datentypen (Medien) als durch ihre Integration. Eine Integration setzt normierte Formate voraus, nur so kann eine transparente Integration der Information unter einer Benutzeroberfläche erfolgen. Ein weiterer Anspruch ist das semantische Wissen über den Inhalt der verschiedenen Medien. Damit kann erreicht werden, daß bei der Suche nach einer Vogelart nicht nur ein Text und das zugehörige Bild gefunden wird, sondern auch ein Video mit ganz anderem Grundthema z.B. ein Reisebericht, weil darin zufällig auch ein Vogel zu sehen oder zu hören ist.

Auf der technischen Seite ist die Synchronisation bei der Aufnahme und Wiedergabe von verschiedenen Medien erforderlich. Eine neue Dimension für alle Rechner, die kein Echtzeit-Betriebssystem haben. Schließlich ist auch die Umsetzung von einem Medium in ein anderes wünschenswert. Eine Vision ist die Generierung eines Videos aus einen vorhandenen Text. Damit könnten Romane automatisch mit (virtuellen) Schauspielern 'verfilmt' werden .



5.2.2 Anforderungen an Multimedia-DBS


- ◆ MM-Datentypen
 - z.B. Text, Bild, Graphik, Ton
- ◆ komplexe MM-Objekte
 - z.B. Luftbild, Satellitenbild, Panorama, etc.
- ◆ Inhaltsadressierung und -suche
 - z.B. auf Basis einer Personenbeschreibung
 - z.B. Recherche nach Pressefotos (Kohl u. Mitterrand im Palais. d. Champs E.)
- ◆ Techniken des Information Retrieval
 - Approximative Suche eines Themas (Siehe 5.1.2)
- ◆ Konvertierung der Medien
- ◆ synchrone Verarbeitung
- ◆ lange Transaktionen, Versionen

Bild 5 -25 Datenbank- und Informationssysteme (c) 10-99 Prof. Dr. F. Laux

Alle technischen und inhaltlichen Anforderungen sind auf dieser Folie nochmals zusammengefaßt.

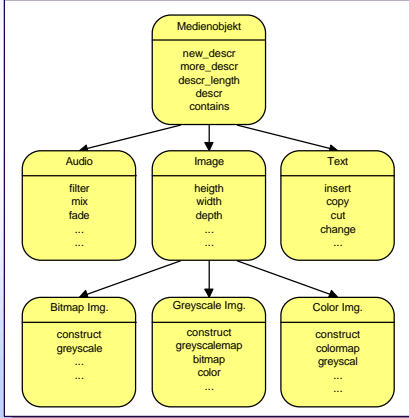
Einige der oben genannten Anforderungen sind für spezielle Gebiete und Aufgaben schon realisiert:

- Multimediaobjekte (☞ sind in objektorientierten Datenbanken zu finden)
- automatische Analyse von Luftbildern und Satellitenaufnahmen (☞ sind für Geheimdienste, Geologen in speziellen Systemen realisiert)
- automatisches Scannen und Auswerten von Funk und Telefon (☞ wird von Geheimdiensten und Staatsschutz eingesetzt)
- Erkennung einfacher Muster oder Personenkontrolle mit Hilfe von KI-Methoden (☞ realisiert für Bar-Code Scanner, Prototypen für komplexe Objekte z. B. für Sicherheitskontrollen)
- synchrone Verarbeitung verschiedener Medien (☞ wird in Film- und Fernsehstudios eingesetzt)



5.2.2 MM-Datenstruktur

- ◆ Klassenhierarchie für Medienobjekte
- ◆ Ein Objekt wird kontextunabhängig gespeichert -> durch Angabe des Codes ist Konvertierungen in Zielkontext möglich

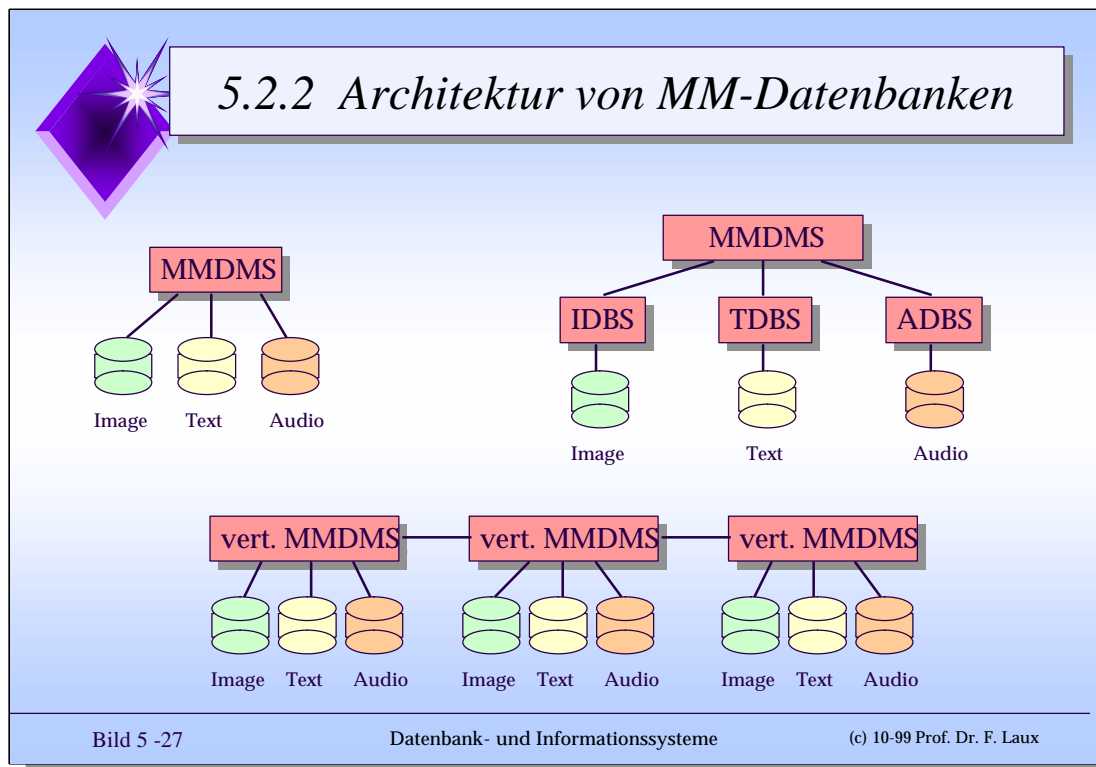


```
typedef struct
{
  int height;          /* Bildhöhe */
  int width;           /* Bildbreite */
  int depth;           /* max. Farbtiefe */
  float aspect_ratio; /* l/b-Verhältnis */
  enum encoding {TIFF,WMF,PCX,JPEG,BMP};
  int colormap_length; /* # Farbtabelle */
  int colormap_depth; /* Farbtabelle tiefe */
  struct list /* zweidimens. Liste */
  {
    /* mit Farbtabelle */
    colormap *map;
  };
  bit *pixelmatrix; /* Bitmap */
} color_image;
```


Bild 5 -26
Datenbank- und Informationssysteme
(c) 10-99 Prof. Dr. F. Laux

Aus objektorientierter Sicht muß für die verschiedenen Multimediaobjekte eine Klassenhierarchie entwickelt werden. Die Klassenhierarchie muß erweiterbar sein. Deshalb kann auch weder Oracle V.7 noch Access 95 als Multimediadatenbank bezeichnet werden, da die verfügbaren Datentypen nicht erweiterbar sind. Neuere Versionen von Oracle, Universal Database DB/2 u.a. können als sogenannte 'objektrelationale' DBS eigene Datentypen definieren indem sie Strukturen aus vorhandenen Typen aufbauen.

Neben der statischen Struktur sind natürlich auch die Operationen (Methoden) für diese Objekte zu realisieren. Neueste Entwicklungen bei objektrelationalen Datenbanken wie Oracle V.8i, Sybase und Universal DB/2 V.6 verwenden funktionale Erweiterungen, die zum Teil von Fremdherstellern geliefert werden, um z.B. Bilder, Graphiken und Ton zu manipulieren. Diese Erweiterungen werden ähnlich wie Plug-Ins zur Datenbank gebunden. Oracle nennt seine Erweiterung *Data Cartridges* (<http://www.oracle.com/st/>), Informix seine *DataBlades* (http://www.informix.com/informix/products/new_plo/plo1.htm) und IBM UDB/2 seine *Extenders* (<http://www.software.ibm.com/data/db2/udb/>).



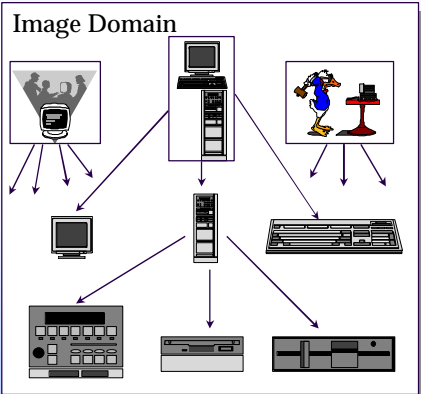
Stand der Technik war bis 1997 getrennte Speicher für die drei Grundmedien. Begünstigt durch bessere Kompressionstechniken (z.B. Wavelets), neue Standards (JPEG, MPEG, MP3) und höhere Speicherkapazität (DVD, Digital Versatile Disc) werden in Zukunft alle Datentypen gemeinsam und digital gespeichert. Auch bei multimedialen Datenbanken dürfte das Ziel ein verteiltes, transparentes Management (MMDMS, Multimedia Data Management System) sein. Dazu sind allerdings echtzeitfähige (realtime) Protokolle (z.B. ATM Protokoll (Asynchronous Transfer Mode Protocol, siehe auch URL: <http://www.franken.fh-heilbronn.de/Fachhochschule/Einrichtungen/Rechenzentrum/Netze/atm.html>) notwendig, um die Synchronisation der Medien sicherzustellen.



5.2.2 Query in Multimedia-Systemen

- ◆ Ein Problem bei MM-Systemen ist, den Inhalt von Ton, Bildern und Videos zu erfassen und zu interpretieren. Dazu sind z.B. komplexe Bildanalysen erforderlich.
- ◆ Beispiel: MM-Abfrage in Dokumenten (Angeboten, Datenblättern, Werbematerial)

Image Domain



```

FIND DOCUMENTS VERSION LAST
WHERE Document.Date > 1.1.1996 AND
(*Sender.Name = "Olivetti" OR
*Product_Info CONTAINS "Olivetti") AND
(*Product_Descr. CONTAINS "Personal
Computer%" OR IMAGE
*Product_Picture MATCHES (DOMAIN
Computer_Pictures OBJECTS
IBM_like_PC WITH AT LEAST 1
Floppy_Drive ))
AND (*Address.Country = "Italy" OR TEXT
CONTAINS "Italy")
AND WITH *Company_Logo;
          
```

Bild 5 -28
Datenbank- und Informationssysteme
(c) 10-99 Prof. Dr. F. Laux

Abschließend wird ein Beispiel für eine multimediale Abfrage vorgestellt.

Die Abfrage ist wie SQL deskriptiv. Bemerkenswert ist die Einschränkung der Query auf die letzte (aktuelle) Version der Objekte und dass sich die Abfrage auch auf den Bildinhalt erstreckt :

```

....
IMAGE *Product_Picture MATCHES
(DOMAIN Computer_Pictures OBJECTS
IBM_like_PC WITH AT LEAST 1 Floppy_Drive)
....
[DB-Objekte/Attribute sind durch * gekennzeichnet]
  
```

Das Objekt `*Product_Picture` wird mit einer Klassifikation (DOMAIN) von `IBM_like_PC` verglichen, dabei wird vom Objektinhalt (Bildinhalt) verlangt, daß wenigstens 1 Diskettenlaufwerk enthalten ist. D.h., das Bild zeigt einen IBM-ähnlichen PC mit mindestens einem Diskettenlaufwerk.



Literatur zu Kapitel 5

- ◆ A.-W. Scheer: Architektur integrierter Informationssysteme, Springer V. 1991
- ◆ O. Abeln (Hrsg.): CAD-Referenzmodell, Teubner V.1995
- ◆ Salton, McGill: Information Retrieval, McGraw Hill 1987
- ◆ Meyer-Wegener: Multimedia-Datenbanken, Teubner V. 1991
- ◆ Goldfarb, Prescod: The XML Handbook, Prentice Hall 1998

Bild 5 -29

Datenbank- und Informationssysteme

(c) 10-99 Prof. Dr. F. Laux

Die o.g. Literatur befindet sich in unserer Bibliothek. Wer sich für HTML als Beispiel einer Hypertextsprache interessiert, findet eine fast lineare (traditionelle) Einführung unter der URL:
<http://www.ncsa.uiuc.edu/General/Internet/WWW/HTMLPrimer.html>

Online Referenzen zu XML finden sich unter:

W3C XML Recommendation, <http://www.w3.org/TR/REC-xml>

XML Schema Working Draft in Development,

<http://www.w3.org/TR/xmlschema-1> (Structures),

<http://www.w3.org/TR/xmlschema-2> (Data-Types)

Allgemeine XML Ressourcen:

<http://www.xml.com>

M. Johnson: "XML for the absolute beginner", online in JavaWorld 4-1999

<http://www.javaworld.com>

R. Pfeiffer: XML Tutorial in 3 Teilen (Overview, XML Documents, XML und Java),

<http://www.ibm.com/xml>