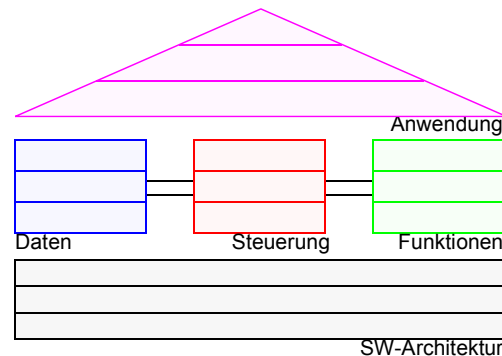


1. Einführung und Grundbegriffe

- **GBIS-Rahmen: Einordnung**



- **Miniwelt¹**

- modellhafte Abbildung eines Realitätsausschnitts
- Transaktionskonzept – ACID-Eigenschaften

- **Information – was ist das?**

- **Aufgaben eines Informationssystems?**

- Komponenten eines rechnergestützten Informationssystems
- Aufgaben und Zielsetzungen eines betrieblichen Informationssystems
- wichtige Anwendungsklassen

- **Daten in Informationssystemen**

- strukturierte und unstrukturierte Daten
- semi-strukturierte Daten (HTML, XML)
- Multimedia-Daten (VITA-Daten)

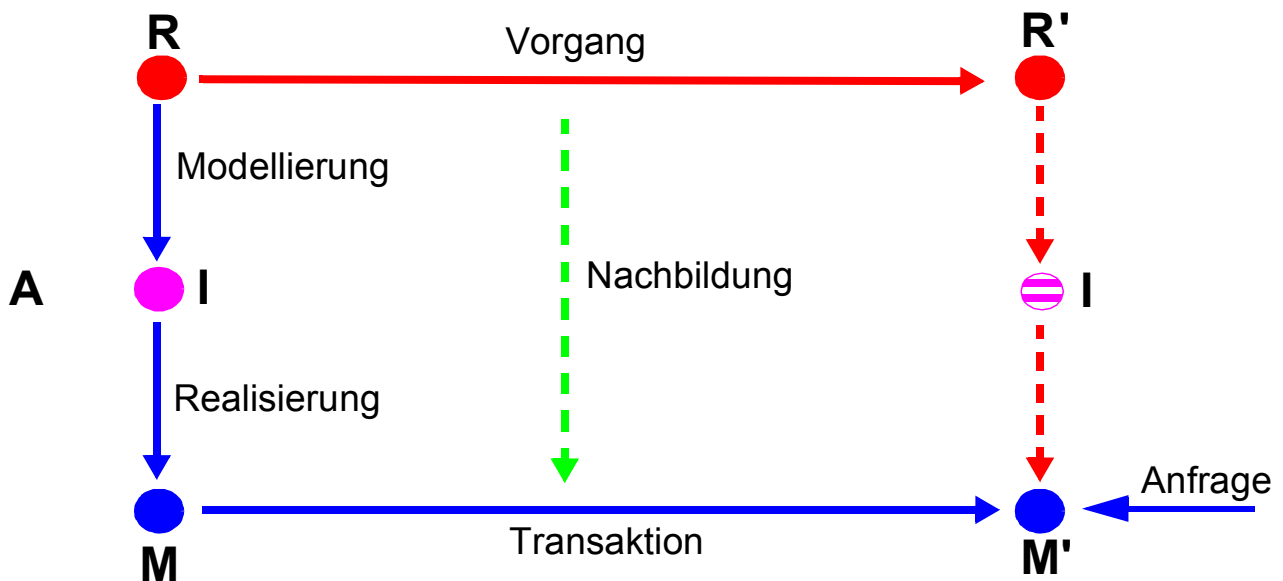
- **Beispiele für Informationssysteme**

- Informationssystem einer Universität, eines Produktionsbetriebs, einer Bank, . . .
- Straßeninformationssystem

1. Ein Datenbanksystem verwaltet Daten einer realen oder gedanklichen Anwendungswelt. Diese Daten gehen aus Informationen hervor, die stets aus den Sachverhalten und Vorgängen dieser Anwendungswelt durch gedankliche Abstraktionen (Abbilder, Modelle) gewonnen werden. Sie beziehen sich nur auf solche Aspekte des betrachteten Weltausschnitts, die für den Zweck der Anwendung relevant sind. Ein solcher Weltausschnitt wird auch als *Miniwelt* (Diskurswelt) bezeichnet.

Miniwelt – modellhafte Abbildung

- Grobe Zusammenhänge



- R:** Realitätsausschnitt (Miniwelt)
- I:** Informationsmodell
(zur Analyse und Dokumentation der Miniwelt)
- M:** DB-Modell der Miniwelt
(beschrieben durch Objekt- und Beziehungstypen, ihre Ausprägungen sowie Integritätsbedingungen usw.)
- A:** Abbildung aller relevanten Objekte und Beziehungen
 - ➔ **Abstraktionsvorgang**

- Transaktion (TA):**

- bildet Vorgang in **R** im DBS nach und
- garantiert ununterbrechbaren Übergang von **M** nach **M'**
 - ➔ **implementiert durch Folge von DB-Operationen**
- DB-Anfragen beziehen sich auf **M** bzw. **M'**

- Integritätsbedingungen:**

- Zusicherungen über **A**, **I** und **M**: $A_1: R \rightarrow I$, $A_2: I \rightarrow M$
 - ➔ **Ziel: möglichst gute Übereinstimmung von **R** und **M****
- **Idealfall:** Die DB ist zu jeder Zeit ein Abbild (Modell) der gegebenen Miniwelt

Miniwelt – modellhafte Abbildung (2)

- **Transaktionskonzept**

- führt ein neues Verarbeitungsparadigma ein
- ist Voraussetzung für die Abwicklung betrieblicher Anwendungen (*mission-critical applications*)
- erlaubt „**Vertragsrecht**“ in rechnergestützten IS zu implementieren

- **Welche Eigenschaften von Transaktionen sind zu garantieren? (ACID-Paradigma)**

- **Atomicity (Atomarität)**

- TA ist kleinste, nicht mehr weiter zerlegbare Einheit
- Entweder werden alle Änderungen der TA festgeschrieben oder gar keine („alles-oder-nichts“-Prinzip)

- **Consistency**

- TA hinterlässt einen konsistenten DB-Zustand, sonst wird sie komplett (siehe Atomarität) zurückgesetzt
- Zwischenzustände während der TA-Bearbeitung dürfen inkonsistent sein
- Endzustand muss die Integritätsbedingungen des DB-Modells erfüllen

- **Isolation**

- Nebenläufig (parallel, gleichzeitig) ausgeführte TA dürfen sich nicht gegenseitig beeinflussen
- Alle anderen parallel ausgeführten TA bzw. deren Effekte dürfen nicht sichtbar sein

- **Durability (Dauerhaftigkeit)**

- Wirkung einer erfolgreich abgeschlossenen TA bleibt dauerhaft in der DB erhalten
- TA-Verwaltung muss sicherstellen, dass dies auch nach einem Systemfehler (HW- oder System-SW) gewährleistet ist
- Wirkungen einer erfolgreich abgeschlossenen TA kann nur durch eine sog. kompensierende TA aufgehoben werden

Information – Was ist das?

- Begriff „Information“ hängt eng mit dem Begriff „Nachricht“ zusammen.

➔ Die (abstrakte) Information wird durch die (konkrete) Nachricht mitgeteilt

- **Nachrichten²**

- **Sender und Empfänger**

- Wodurch wird etwas eine Nachricht?
Mitteilungsabsicht des Senders
- Eine Nachricht ist **zweckorientiert** und will neues, zusätzliches Wissen vermitteln

- Transport durch

Raum (*Rauchzeichen, Trommeln, Telefon, ...*) und

Zeit (*Wandmalereien, Bücher, Dokumente, ...*)

- Nachrichten sind **endlich**

Beispiele:

DNS-Strang

Bientanz

Gesänge der Wale

Vogelstimmen

- Aufgabe des **Empfängers**

- **Erkennen** der Nachricht
- **Interpretation** der Nachricht

(Sender bezieht Fähigkeiten des Empfängers schon beim Erzeugen der Nachricht ein.)

-
2. Menschen kommunizieren auf äußerst komplexe Weise. Die Übermittlung von bloßen Fakten macht dabei nur den geringsten Teil aus. Weit mehr Information – zumeist emotionalen Inhalts – ist in Mimik, Tonfall, Körperhaltung oder Gestik versteckt.

Wer für all diese Signale blind ist, fühlt sich ständig „wie ein Taubstummer beim Kaffeeklatsch“. Diese Anomalie hat bereits einen Namen: Alexithymie, eine Art Gefühlsblindheit (aus dem Griechischen: a = nicht; lexis = das Lesen, die Rede; thymos = das Gemüt). Sie tritt relativ häufig auf – bei psychologischen Tests erwiesen sich 13% der Teilnehmer als alexithym (Spiegel Nr. 49/2003).

Information – Was ist das? (2)

- **Interpretation einer Nachricht³**

1. Rahmen-Nachricht

- „Ich bin eine Nachricht, entschlüssele mich, wenn du kannst!“
- „Die Rahmen-Nachricht zu verstehen heißt zu erkennen, dass ein Entschlüsselungsmechanismus benötigt wird“
(sehr leicht bei Datei-Suffix)
- manchmal schwierig: lange nicht-periodische Folge von Mustern in einer regelmäßigen geometrischen Struktur

2. Äußere Nachricht

- „Die äußere Nachricht zu verstehen heißt zu wissen, wie der Entschlüsselungsmechanismus konstruiert wird“
z. B. *Plattenspieler bauen*
jemanden holen, der Japanisch kann
- Bei Rechneranwendungen ist der Mechanismus oft vorgegeben

3. Innere Nachricht

- „Die innere Nachricht zu verstehen heißt, die Bedeutung extrahiert zu haben, die der Sender beabsichtigt hatte“
- evtl. mehrere Schichten von äußeren und inneren Nachrichten

➔ **Kann das ein Programm überhaupt?**

- **Nachrichtenarten**

Auswahl der äußeren Nachricht bzw. des Mediums

- **nichtsprachliche** Nachrichten
 - Bilder (Standbilder, Filme, Hologramme), Plastiken
 - Phonogramme (Schallplattenaufnahme, Tonbandaufnahme)

3. Hofstadter, D. R.: Gödel, Escher, Bach – ein Endlos Geflochtenes Band, Klett-Cotta-Verlag, Stuttgart, 1985, Kap. VI: „Wo die Bedeutung sitzt“.

Information – Was ist das? (3)

- **Nachrichtenarten (Fortsetzung)**

- **sprachliche** Nachrichten

- **gesprochene** Sprache
(auch Morsezeichen, Zeichensprache, Gebärdensprache)
- **geschriebene** Sprache
(optisch lesbar, taktil lesbar, akustisch lesbar)

- **Signal:**

„Der eine Nachricht übertragende zeitliche Verlauf einer physikalischen Größe“

- **Digitale Nachricht:**

Endliche Folge von **Zeichen** aus einem **Zeichenvorrat**

- **Maschinelles Verarbeiten von Nachrichten**

- **Arten**

- analog: Filterung, Verstärkung, usw.
- digital: **Manipulation von Zeichenfolgen**

- Anwendung von Interpretationsvorschrift ($N \rightarrow I$)

soll **informationstreu** sein:

- Information kann ausgefiltert werden, damit andere Information deutlich hervortritt (aus einer Folge von Zahlen die Summe bilden)
- Verbleibende Information wird nicht verfälscht (die Summe stimmt)

- Es kann niemals neue Information entstehen!

- Es kann nur die vorhandene und „verstandene“ Information **sichtbar** (erkennbar, erfassbar) gemacht werden!

➔ **Was bedeutet das bei „Datenbanken und Informationssystemen“?**
(Generierung von Berichten, Data Mining, Business Intelligence, ...)

Information – Was ist das? (4)

• Austausch von Nachrichten

Es existieren große Unterschiede, wenn sie übermittelt werden

- bei direkter zwischenmenschlicher Kommunikation
(Anreicherung durch Kontext und zusätzliche „Informationskanäle“)
- bei der Kommunikation „Mensch – Informationssystem“
(angepasste Schnittstellen, sprachliche Nachrichten, Töne und Bilder)
- bei kooperativen Informationssystemen
(abgestimmte Protokolle, gleichartiges Begriffssystem,
vorgeplante Homogenisierung durch föderierte (Datenbank-)Systeme)
- bei heterogenen Informationssystemen
(keine vorgeplanten Protokolle, verschiedenartige Begriffssysteme,
Ad-hoc-Informationsfusion,
 - „Verstehen“ über Abbildung auf gemeinsames Begriffssystem (*Ontologie*)⁴
 - Relevanzbestimmung mit Hilfe von *Recommender-Systemen* usw.)
- bei der Kommunikation „Mensch – WWW“ als Sammlung heterogener IS
(Probleme durch Missinformation (Hilfe durch *Reputationssysteme*),
Erreichbarkeit, Reasoning im Semantic Web, DAML+OIL⁵)
➔ Wieviel Informationsgehalt wird jeweils übermittelt?

• Viele Anstrengungen der KI

- zur Entwicklung „intelligenter“ und adaptiver IS-Schnittstellen
- Solche Schnittstellen werden noch auf lange Zeit nur die sprachlichen oder textlichen Anteile der „Nachrichten“ erfassen („alexithym“ bleiben)
- Wie erkennt man maschinell Gefühle (sechs Grundemotionen: Freude, Wut, Angst, Trauer, Überraschung, Ekel) in der Mimik oder die Verfassung (Körperhaltung) des Kommunikationspartners?

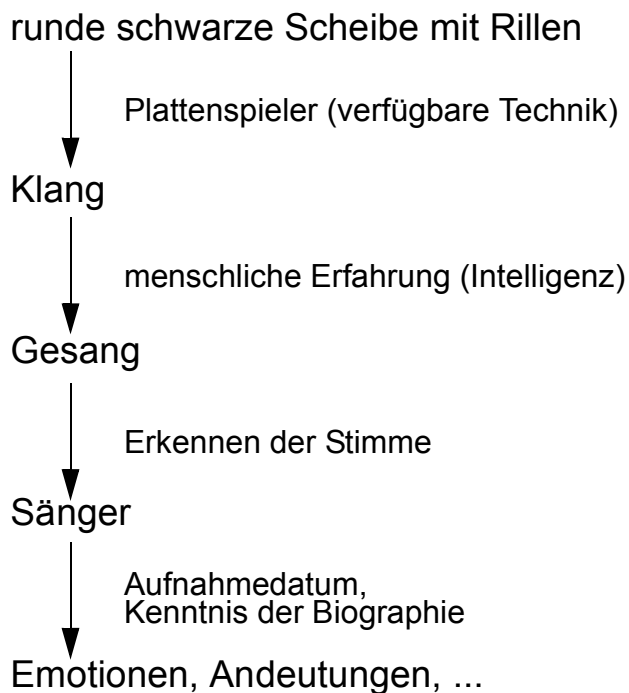
4. An ontology is a specification of a conceptualization (<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>)

5. DARPA Agent Markup Language + Ontology Inference Layer

Information – Was ist das? (5)

- **Informationsgehalt einer Nachricht (innere Nachricht)**

- Wieviel Information steckt wirklich *in* der Nachricht und wieviel kommt durch *das Hintergrundwissen* des Empfängers hinzu?
 - „Verstehen ist zu 80% Wiedererkennen“
 - Einfluss von Erziehung, Kultur, „Common Sense“ usw.
- Beispiel: **Schallplatte**



- **Wie findet man den Informationsgehalt einer Nachricht heraus?**⁶

- **„Musikbox-Theorie“:**
Nachricht enthält selbst keine Information; besteht nur aus Auslösern (Triggern), die Informationen aus einem „Abspielapparat“ abrufen
- **Gegenposition:**
„Meaning is intrinsic if intelligence is natural“ bzw. universell
- **Rolle der KI?**
(Frage nach der „Intelligenz von Maschinen“)

6. Casti, J. L.: Das Cambridge Quintett, Berlin Verlag, 1998.

Information – Definitionsversuche

- **Beobachtung**

Die Praxis der Information und Kommunikation entwickelte sich rasant, ohne dass der Informationsbegriff von einer Theorie hinreichend geklärt wurde. Es gibt keine Theorie oder gar Philosophie der Information.

- ↳ **Eine resignierende Schlussfolgerung⁷:**

Die Definition des Begriffes „Information“ ist nicht möglich. Jeder Versuch dazu setzt ähnliche Begriffe voraus, beispielsweise „Wissen“ oder „Kommunikation“. Diese Definition wäre damit zyklisch.

- **Erklärungsversuche (philosophisch, technisch):**

1. **„Information ist neben Materie und Energie etwas Drittes“**

2. **Information für Menschen über seine Umwelt:**

Information setzt den Menschen über seine Außenwelt in Kenntnis, ist also der „Stoff“, der Erkenntnis ermöglicht

3. **Informationstheorie nach Shannon:**

Statistischer Informationsbegriff (Entropie einer Nachrichtenquelle):

$$H(p_1, \dots, p_n) = - \sum_{i=1}^n p_i \cdot \log p_i$$

4. **Information und Nachricht:**

Die übermittelte Nachricht ist dann von Bedeutung, wenn wir eine Abbildung kennen, die sie mittels einer Interpretationsvorschrift α auf eine Information abbildet:

$$\mathbf{N} \xrightarrow{\alpha} \mathbf{I}$$

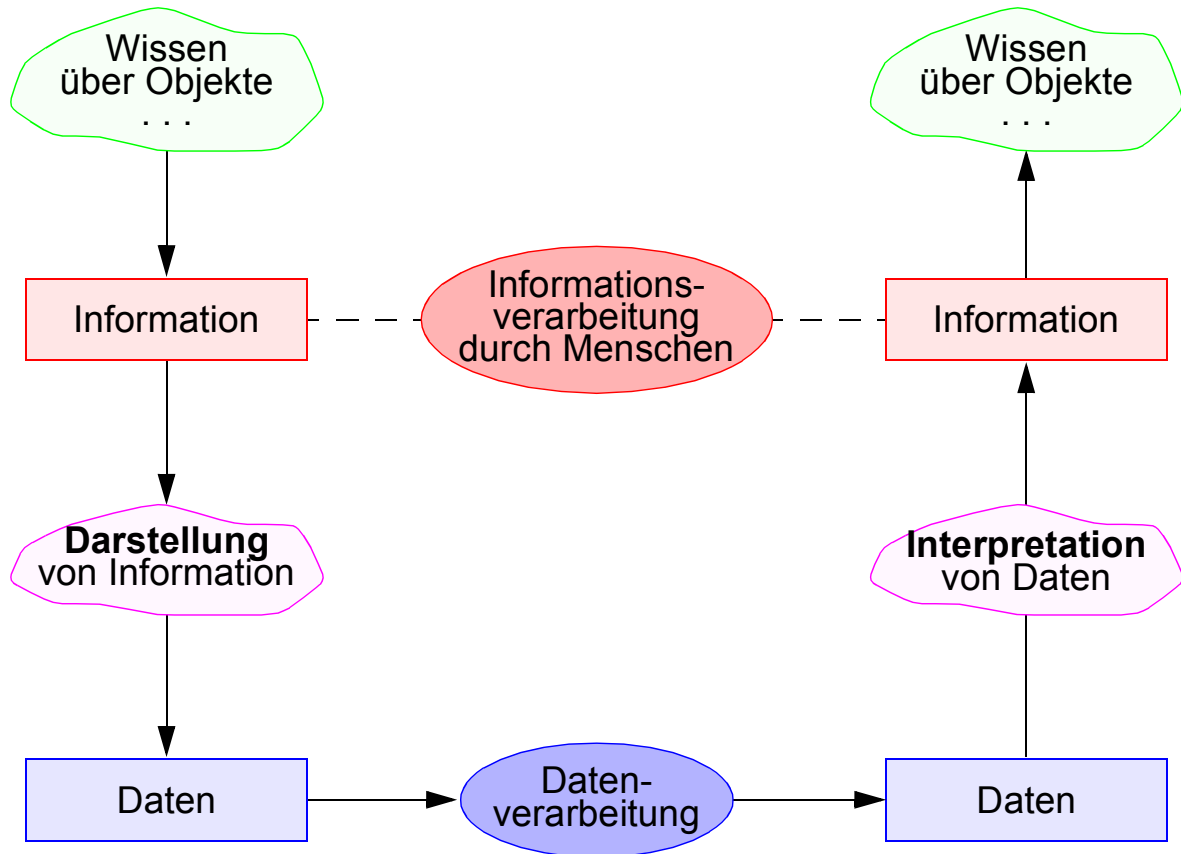
7. Bauer, F.L., Goos, G.: Informatik – Eine einführende Übersicht, 3. Auflage, Springer-Verlag, 1982

Information – Definitionsversuche (2)

- Erklärungsversuche (pragmatisch):

5. Informationsbegriff nach DIN: Erklärungsmodell

(auch für Nachrichtenaustausch zwischen Sender und Empfänger)



Information: subjektive Welt der bewerteten Daten

Daten: objektive Welt der nicht-interpretierten Daten

6. Festlegung in der BWL⁸

Information: **Angaben über Sachverhalte und Vorgänge** (Hansen)

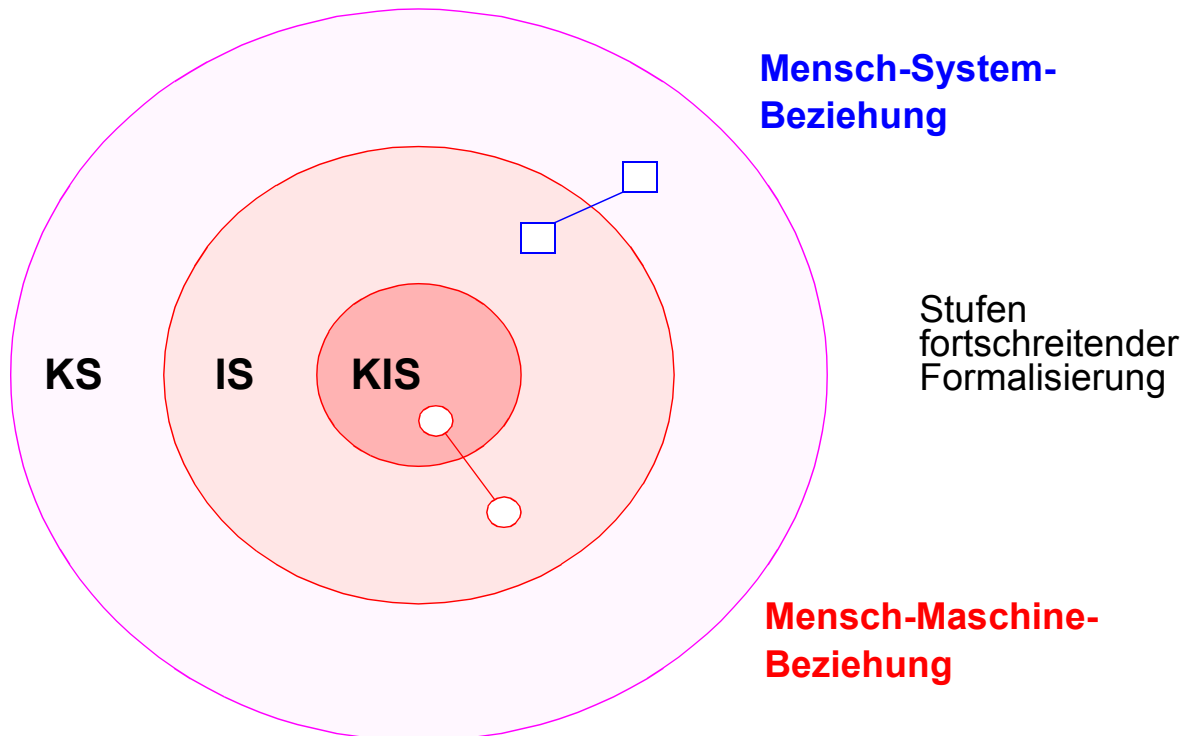
- „Dabei kann man Information im Sinne von als zweckgerichtetes Wissen zur Vorbereitung und Durchführung von Handlungen verstehen“.
- „Eine andere Sichtweise ergibt sich aus der Betrachtung der Information als Produktionsfaktor“.

8. Wirtschaftsinformatik-Lexikon, Gabler-Verlag, 1997

Was ist ein Informationssystem?

- **Charakterisierung eines IS nach M. Senko:**

“The purpose of an information system is to provide a *relatively exact, efficient, unambiguous* model of the significant resources of a real world enterprise.”



Formalisiertes Kommunikationsystem = Informationsystem

- **(Vage) Definitionen:**

Ein Informationssystem⁹ (IS) besteht aus Menschen und Maschinen, die Informationen erzeugen und/oder benutzen und die durch Kommunikationsbeziehungen miteinander verbunden sind.

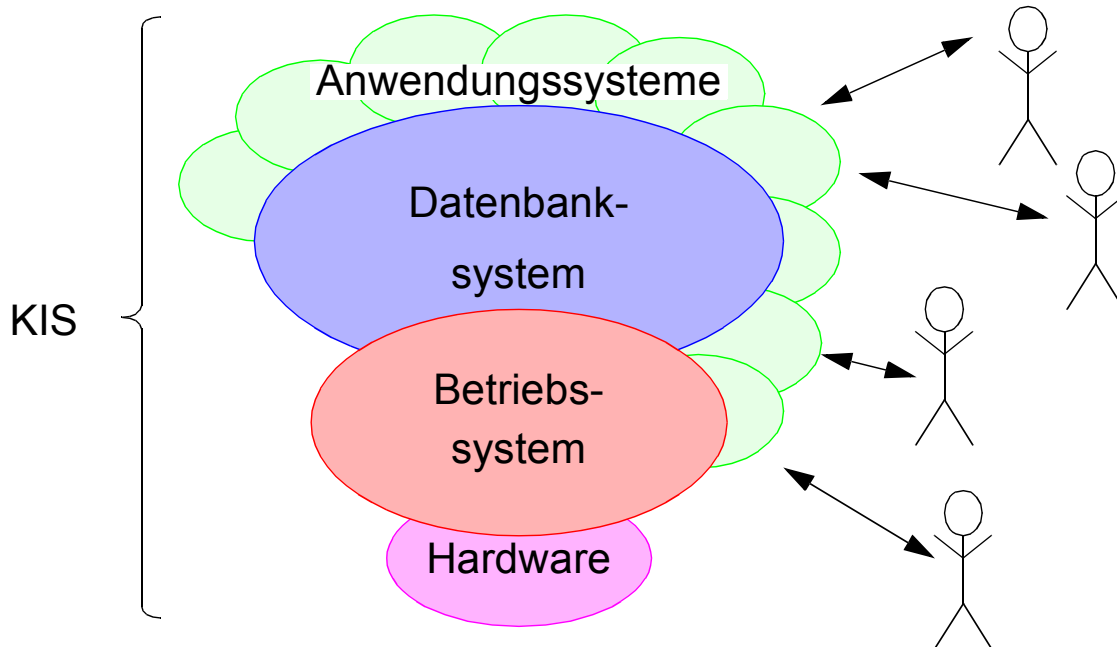
Ein betriebliches IS dient zur Abbildung der Leistungsprozesse und Austauschbeziehungen im Betrieb und zwischen dem Betrieb und seiner Umwelt.

Ein rechnergestütztes IS ist ein System, bei dem die Erfassung, Speicherung und/oder Transformation von Informationen durch den Einsatz von EDV teilweise automatisiert ist. In der betrieblichen Praxis besteht es typischerweise aus einer Menge unabhängiger Systeme, die zusammen die angestrebte Leistung erbringen (*KIS: kooperatives Informationssystem*).

9. Als „System im weiteren Sinne“ gilt (a) eine Menge von Elementen (Systembestandteilen), die (b) durch bestimmte Ordnungsbeziehungen miteinander verbunden und (c) durch klar definierte Grenzen von ihrer Umwelt geschieden sind; von „Systemen im engeren Sinne“ oder „technischem System“ spricht man, wenn sowohl die Außenwirkungen des Systems insgesamt wie auch seine Binnenstruktur (d. h. die Ordnungsbeziehungen der Systembestandteile) durch Zielfunktionen bestimmt sind (H. Wedekind).

Rechnergestützte Informationssysteme

- Aufbau von rechnergestützten Informationssystemen



Datenbanksystem (DBS): zentrale Komponente für KIS

Grobdefinition:

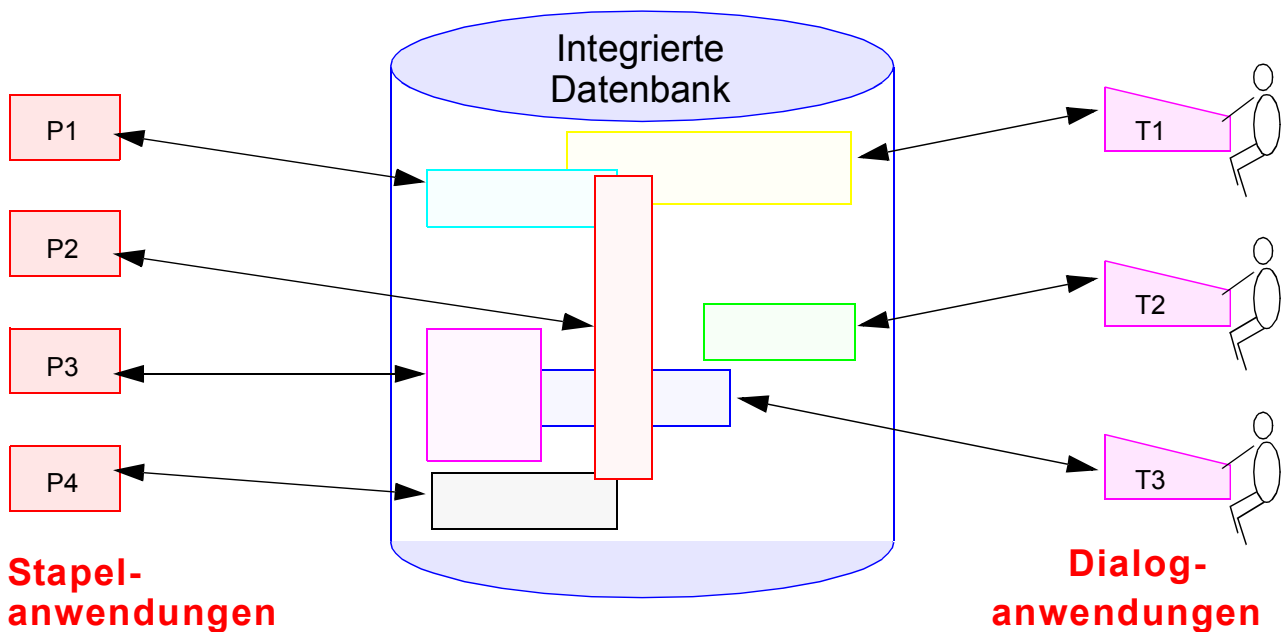
DBS = DB + Datenbankverwaltungssystem (DBVS, DBMS)

Eine **Datenbank** ist eine Sammlung gespeicherter operationaler Daten, die von den Anwendungssystemen eines bestimmten Unternehmens benötigt werden.

Ein **DBMS** ist ein standardisiertes Softwaresystem zur Definition, Verwaltung, Verarbeitung und Auswertung der DB-Daten. Es kann mittels geeigneter Parametrisierung an die speziellen Anwendungsbedürfnisse angepaßt werden.

Rechnergestützte Informationssysteme (2)

- Vereinfachte Sicht auf DB-Anwendungen (AW)



- Dialogorientierte Anwendungen:

- Transaktionsanwendungen mit Interaktion des Endbenutzers
- harte Zeitrestriktionen
- Terminal-E/A für Anforderungen und Ergebnisse

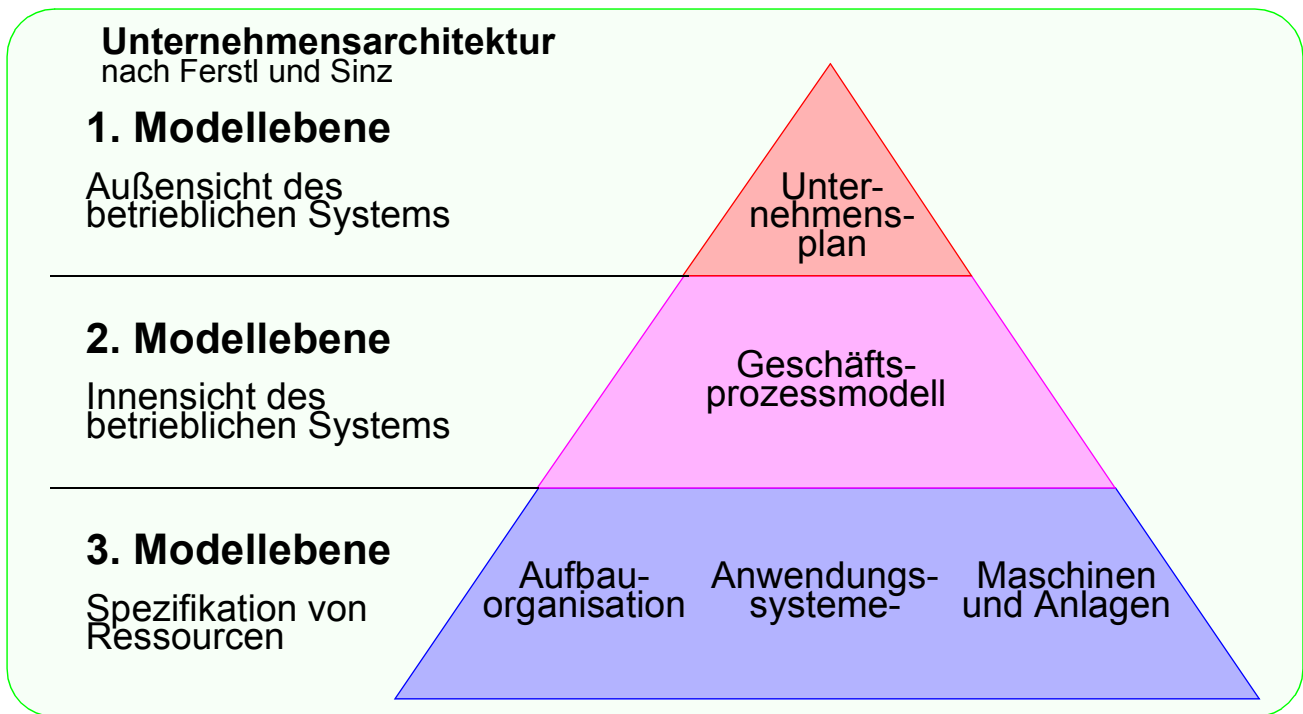
- Stapelorientierte Anwendungen:

- (Transaktions-) Anwendungen ohne Interaktion des Endbenutzers
- keine Zeitrestriktionen
- Datei-E/A für Anforderungen und Ergebnisse

Aufgaben betrieblicher Informationssysteme

• Betriebliche Informationssysteme

- zentraler und strategisch bedeutsamer Bestandteil von Unternehmen
- systematischer Aufbau und Ausrichtung an den Unternehmenszielen
- Orientierung an Unternehmensarchitektur als dreischichtiger Pyramide



• Administrative und operative Ebene als Fundament

- wird gebildet von Mitarbeitern (eingegliedert in Organisationseinheiten und Stellen einer Aufbauorganisation), Anwendungssystemen sowie Maschinen und Anlagen
- Interaktion zwischen ihnen dient der Verfolgung definierter Ziele

• Planungs- und Kontrollebene

- Interaktionen der 3. Ebene werden in Form von Geschäftsprozessmodellen formuliert
- Erreichen der Geschäftsziele wird überwacht

• Strategische Ebene

Unternehmensplan formuliert Ziele und weitere Randbedingungen, die durch Ausführung von Geschäftsprozessen erreicht bzw. eingehalten werden sollen

Aufgaben betrieblicher Informationssysteme (2)

- **Unterscheidung nach Aufgaben des**

- betrieblichen Lenkungssystems (Planung, Steuerung und Kontrolle)
- betrieblichen Leistungssystems (Administration, Disposition und Durchführung)

- **Aufgaben und Typen (Beispiel)**

- Administrationssysteme dienen der Rationalisierung und (Teil-)Automatisierung vorhandener Abläufe
- Dispositionssysteme sollen die kurzfristige betriebliche Entscheidungsfindung vereinfachen/übernehmen
- Planungssysteme unterstützen die mittel- bis langfristige Entscheidungsfindung (Erzeugung alternativer Pläne, weitreichendere Auswirkung)
- Kontrollsysteme dienen dem Erkennen von außergewöhnlichen und daher bemerkenswerten Situationen (Datenkonstellationen). Sie erhalten von Administrationssystemen Ist-Daten, um Ist-Soll-Abweichungen erkennen zu können

➔ hier: Erarbeiten der technischen Grundlagen

- **Was sind typische Aufgabenbereiche?**

Typ/Aufgabe	Produktion	Beschaffung Lagerhaltung	Vertrieb	Personal
Administration	Betriebsdatenerfassung	Lagerverwaltung	Kundenverwaltung	Personalverwaltung
Disposition		Bestelldisposition		
Kontrolle	Fertigungsleitstand		Tourenplanung	
Planung	Absatz-/Kapazitätsplanung		Marketingplanung	Personaleinsatzplanung

Web-basierte Informationssysteme

• Arten Web-basierter Informationssysteme

- Präsentation von Angeboten: Werbeplattform (keine Vertragsabschlüsse)
- „virtuelles Ladengeschäft“: Online-Vertragsabschluss
- „virtueller Marktplatz“: Anbieter und Nachfrager mit Online-Vertragsabschluss

• Das ABC des E-Business

- Aushandeln von Verträgen, Überprüfen/Gewährleisten der Gültigkeit
- Vielfältige Möglichkeiten des elektronischen Geschäftsverkehrs

	Administration	Business	Consumer
A	A2A – Beispiel: Zusammenarbeit von Behörden	A2B – Beispiel: Ausschreibungen	A2C – Beispiel: Verwaltungsservice für Bürger
B	B2A – Beispiel: Online-Handel: öffentlicher Bedarf	B2B – Beispiel: Online-Handel: Unternehmensbedarf	B2C – Beispiel: Online-Handel: privater Bedarf
C	C2A – Beispiel: Priv. Angebote an A , Übernahme ehrenamtlicher Aufgaben usw.	C2B – Beispiel: Jobbörse	C2C – Beispiel: Inserate, Versteigerungen

• Märkte

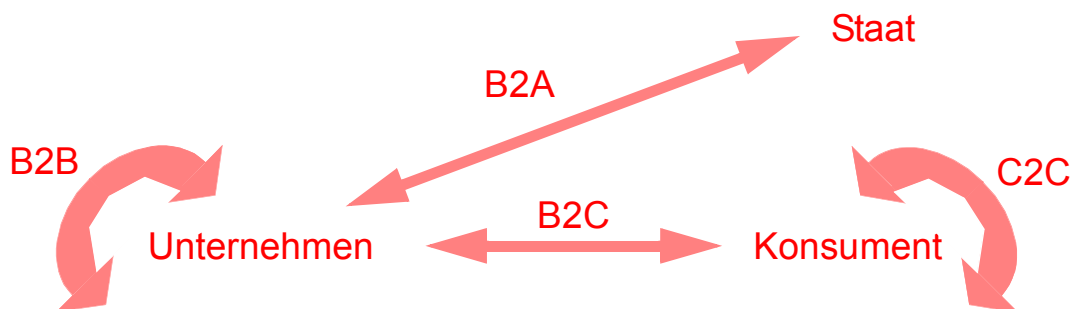
- Geschäftsabwicklung: Verkauf, Auktion, ...
- Arten: geschlossen, halb-offen, offen
- Plattformen: zentralisiert, verteilt, Peer-to-Peer (P2P: gleichrangig, autonom)

• Sicherheitsanforderungen

- Identifizierung der Marktteilnehmer (mit Reputationssystemen offline, online?)
- Nachweisbarkeit des Inhalts bzw. Gegenstand des Vertrages
- Digital Rights Management (DRM) und Urheberrecht (Gesetz von 2003)
- Abwicklung von Geschäften (Zahlungs- und Lieferungsprotokolle)

Open Buying on the Internet (OBI)

- **Rahmenwerk für B2B-Lösungen über das Internet¹⁰**
 - Architektur: Komponenten und Informationsflüsse
 - Datenformate für häufig verwendete Geschäftsvorfälle: Bestellungen, Lieferscheine, Bestätigungen
 - Transport-Protokolle
 - Sicherheitsinfrastruktur mit Public-Key-Verfahren und digitalen Unterschriften
 - Integration unterschiedlicher Zahlungsmechanismen
- **Standard definiert durch Internet Purchasing Roundtable**
 - Teilnahme vieler Firmen mit steigender Tendenz
 - Ziel ist die Implementierung und Weiterentwicklung von Standards
 - IPR ist eines von mehreren konkurrierenden Rahmenwerken
- **Akteure und Rollen beim Elektronischen Handel**



- **B2B**
 - Handel zwischen Unternehmen, Beschaffungsvorgänge entlang komplexer Prozessketten
 - Abwicklung durch Softwaresysteme
- **B2C**

Bestell- und Verkaufsprozess eines Anbieters bei großer Anzahl Kunden
- **B2A**

öffentliche Ausschreibungen, ähnlich zu B2B

10. <http://www.openbuy.org/>

Zielsetzungen für betriebliche IS

- **Anforderungen an ein Informationssystem in einem Unternehmen sind unterschiedlich, je nachdem, ob Aufgaben**

- der operierenden Ebene (Sachbearbeitung)
- der planenden Ebene (mittleres Management)
- der strategischen Ebene (Unternehmensleitung)

zu lösen sind

- **Verbesserung aller Prozesse und Aufgaben der operierenden Ebenen** des Unternehmens durch Auskunfts-, Berichts-, Buchungs-, Produktions-, Steuerungs-, Vertriebs- und Anwendungssysteme

Kennzeichen: *Verarbeitung großer Datenmengen und große Änderungshäufigkeit der Daten*

- **Unterstützung und evtl. Teilautomatisierung aller Prozesse und Aufgaben der planenden Ebene durch:**

- benutzerorientierte Bereitstellung von Informationen
- Suche und Auswertung von Daten im Dialog
- Automatisierung von Routine-Entscheidungen
- Einsatz von mathematisch-statistischen Methoden

Kennzeichen: *teilweise nicht vorhersehbarer Informationsbedarf, verdichtete Daten, kein Änderungsdienst*

- **Unterstützung der strategischen Ebene** durch Bereitstellung von Daten für einen **überwiegend nicht vorhersehbaren Informationsbedarf**

Wichtige Anwendungsklassen betrieblicher IS

• Anwendungsklassen – Terminologie

- OLTP (*On-Line Transaction Processing*)

Abwicklung von Transaktionen auf den operationalen Daten

- DW (*Data Warehouse*)

als themenorientierte, zeitlich veränderliche, nicht-flüchtige Datensammlung (vorwiegend geschichtliche Daten)

- OLAP (*On-Line Analytical Processing*):

Analyse betrieblicher Datenbestände

- viele Anwendungsfelder: Banken, Telecom, Handel, Versicherungen, ...
- Einsatz von „Grabungstechniken für Wettbewerbsvorteile“

- DSS (*Decision Support System*)

• Schlagwörter

- **Business Intelligence**: Intelligente Nutzung großer Datenbanken

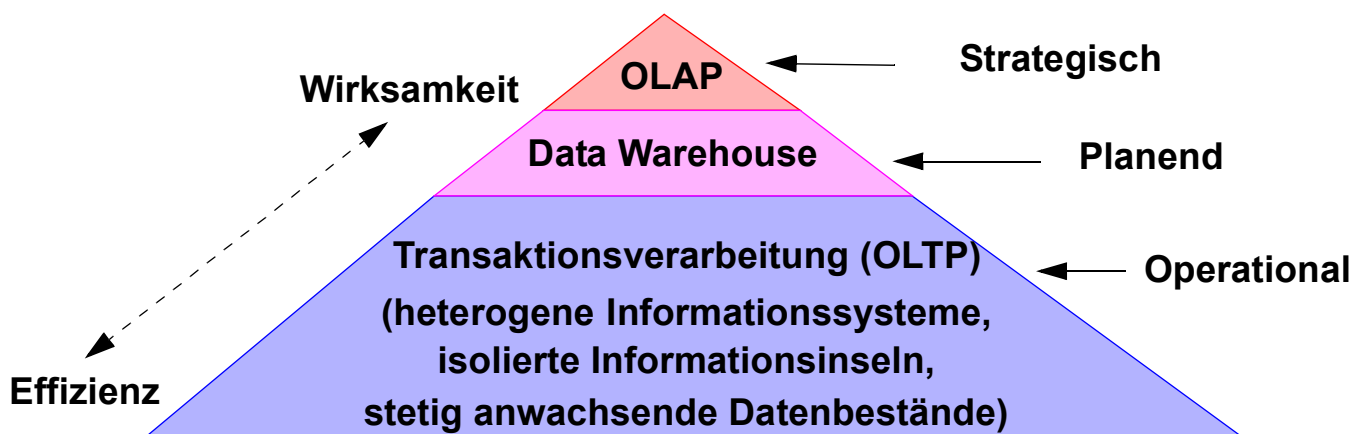
- **Data Mining**: Aufspüren von inhärenten Daten-/Informationsmustern aus großen dynamischen Datenbeständen

“In data mining applications, not only does the system define the semantics, it actually defines the queries. The user simply says ‘Go’, and the system produces what it believes to be useful answers.”

- **KDD (*Knowledge Discovery in Databases*)**, oft synonym zu Data Mining

- **CRM (*Customer Relationship Management*)**

• Informationspyramide

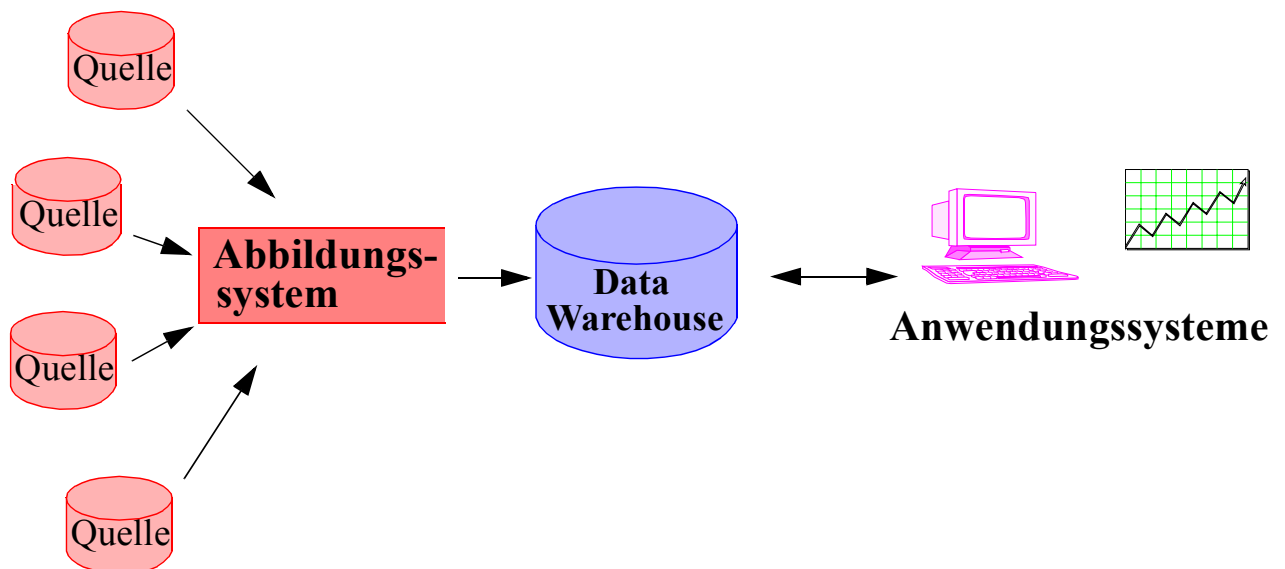


Wichtige Anwendungsklassen betrieblicher IS (2)

• DW-Anwendungen

Es sind betriebswirtschaftliche Kennzahlen, die sich als geeignete Managementunterstützung erwiesen haben, in sehr großen Datenbeständen (50 - 500 GB und mehr) abzuleiten und multidimensional aufzugliedern. Dazu ist eine Voraggregation der Daten, die in einem DW getrennt von den Daten der operativen DBS gehalten werden, aus Leistungsgründen unbedingt erforderlich. Weiterhin muss eine inkrementelle Aktualisierung dieser voraggregierten Daten aus den operationalen DBS (z. B. jede Nacht) erfolgen.

• Grobaufbau eines DW-Systems (4 Systeme)



**Datenbanksysteme
ggf. verschiedenen Typs**

• Data Mining (OLAP)

In DW oder operativen Datenbeständen sehr großer Volumina „schürfen intelligente Agenten“ selbständig nach impliziten Daten-/Informationsmustern, um bislang unbekannte Strukturen und Zusammenhänge aufzudecken. Solche für den Anwender interessanten Muster können

- Beziehungen zwischen Datensätzen oder zwischen Attributen eines Satzes
 - gewisse Regelmäßigkeiten oder Regelabweichungen in Attributwerten
- sein. Dazu ist das Erkennen von unscharfen oder probabilistischen Regeln nötig.

Daten in Informationssystemen – strukturierte Daten

- **Strukturierte oder formatierte Daten:**

(NAME = „Müller“, TAETIGKEIT = „Kalligraph“, GEBDAT = „780623“, . . .)

- maximale Länge (= endlicher Wertevorrat)
- Werte von Variablen, Feldern, Attributen; durch Namen beschrieben
- Bedeutung weitgehend vorgegeben, relativ geringer Informationsgehalt

- **Klassische Datenbanktechnik**

- formatierte Datenstrukturen, feste Satzstruktur
- Beschreibung der Objekte durch Satztyp, Attribute und Attributwerte ($S_i/A_j/AW_k$)
- jeder Attributwert AW_k wird durch Beschreibungsinformation (Metadaten) A_j und S_i in seiner Bedeutung festgelegt
- **Beispiel:** Tabelle (Relation in Tabellendarstellung)

Schema

ANGESTELLTER

Satztyp (Relation)

Ausprägungen

PNR	NAME	TAETIGKEIT	GEHALT	ALTER
496	PEINL	PFOERTNER	2100	63
497	KINZINGER	KOPIST	2800	25
498	MEYWEG	KALLIGRAPH	4500	56

➔ **DB-Schema:** Vollständige Strukturbeschreibung (strukturelle Metadaten) ist vor der Speicherung von Objekten zu spezifizieren und dem DBS bekannt zu machen

- **Art der Anfrage und Aktualisierung**

- deskriptiv (nicht-prozedural)
- mengenorientiert

Daten in Informationssystemen – unstrukturierte Daten

- **Unstrukturierte oder unformatierte Daten:**

„Er heißt Müller. Er wird in seiner Arbeitszeit überwiegend als Kalligraph eingesetzt. Geboren ist er am 23. Juni des Jahres 1978 . . .“

- beliebige Länge
- teilweise selbstbeschreibend
- Bedeutung nur schwach vorgegeben
- hoher Informationsgehalt

- **Dokumente in Information-Retrieval-Systemen**

- unformatierte Daten, keine dem IRS bekannte Dokumentstruktur
- Beschreibung der Objekte durch Dokumenttyp und Wert (D_i/W_k)
- Es gibt keine nähere Beschreibung oder Spezifikation von Struktur und Semantik, die W_k in seiner Bedeutung festlegt

➔ IRS verwaltet „lange“ Werte (z. B. Texte eines Abschnitts, Kapitels oder Buches) und stellt dafür Container (verschiedenen Typs) zur Verfügung

- **Aufgaben/Eigenschaften von IRS**

- Verwaltung von Dokumenten, Büchern, Abstracts usw.
- effiziente Suche in großen Datenmengen (typischerweise nur Retrieval)
- Ziel: „Alle relevanten und ausschließlich relevante Dokumente zu einer Anfrage sollen gefunden werden!“
- Verbesserung der Suchergebnisse durch Thesaurus oder Ontologie¹¹ (ontologische Metadaten, Begriffssystematik):
Festlegung von Begriffen und ihren Beziehungen zueinander
- Anfragesprache für Retrieval
➔ Annäherung an natürliche Sprache erwünscht

11. Ontologie ist ursprünglich eine philosophische Disziplin, neuerdings ein Modebegriff der Informatik. Meyers Enzyklopädisches Lexikon führt Ontologie als „Lehre von dem Wesen und den Eigenschaften des Seienden“. In der Informatik ist sie die „formale Spezifikation eines Gegenstandsbereichs in Form eines Begriffssystems“.

Daten in Informationssystemen – semi-strukturierte Daten

- **HTML-Dokumente im WWW**

- Aufbau des Dokumentes (syntaktische Struktur) ist festgelegt
- Formatierungsanweisungen (Tags) lassen grobe Rückschlüsse auf den Inhalt des Dokumentes zu: TITLE, HREF, ADDRESS, ...
- Inhalt des Dokumentes ist jedoch nicht weiter beschrieben; es gibt keine Metadaten, welche die Bedeutung genauer festlegen

➔ **WWW-Browser kann HTML-Dokumente aufbereiten und graphisch darstellen, ohne den Inhalt zu kennen**

- **Beispiel: HTML-Dokument, semi-strukturiert**

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http:..."
<HTML>
<HEAD>
  <TITLE>Publications 2004</TITLE>
  <META NAME="GENERATOR" CONTENT="Mozilla/3.01Gold (...) [Netscape]">
</HEAD>
<BODY BACKGROUND="http://wwwdbis.informatik.uni-kl.de/pics/paper.jpg">
<H1 ALIGN=CENTER>Publications 2004</H1>
<HR NOSHADE></P>
<H3>Last update: 2002/03/06 </H3>
<P>Härder, T., Bühmann, A.:<BR>
<AHREF="papers/HB04.IFE.html">Datenbank-Caching &ndash; Eine systematische Analyse möglicher Verfahren
</A>, in: Informatik &ndash; Forschung und Entwicklung, Band 19, Heft 1, Springer-Verlag, 2004.</P>
<ADDRESS><A HREF="mailto:wwwhaerder@informatik.uni-kl.de">wwwhaerder@informatik.uni-kl.de</A>
</ADDRESS>
</BODY>
</HTML>
```

➔ **Präsentation an Benutzerschnittstelle: für Menschen, nicht für Programme**

Daten in Informationssystemen

– semi-strukturierte Daten

- **Formatierungssprachen dienen dem Austausch von Dokumenten**

- Es gibt eine Vielzahl von Formatierungssprachen, die alle den internationalen Standard zur Textverarbeitung SGML als Meta-Sprache benutzen, um ihre Formate und Grammatik zu definieren

- **HTML**

- ist eine Sprache zur Formatierung (Strukturierung) von Dokumenten (Texten) (HyperText Markup Language, Tag Language)
- bietet eine vorgegebene Menge von Begrenzungs- und Formatierungsanweisungen (>200) mit standardisierter Bedeutung
- **Beispiel**
 - <H2>Second-Level heading </H2>
 - <P>This is a passage of text that probably belongs to the heading immediately above </P>
- vermischt **Strukturierungs- und Darstellungsaufgaben**
- kann die Suche von Dokumenten kaum unterstützen

- **DocBook**

- weitere Sprache zur Textformatierung (Software-Dokumentation)
- **Beispiel**
 - <SECT2>
 - <TITLE>Second-level heading </TITLE>
 - <PARA> This is a passage of text that certainly belongs to heading above. We know this because both are contained in the same SECT2 element.
 - </PARA>
 - </SECT2>

➔ Jede Sprache ist auf bestimmte Kategorie von Dokumenten zugeschnitten

Daten in Informationssystemen

– Multimedia-Daten

- **Bedarf**

Verwaltung großer Mengen von Multimedia-Datenobjekten auf verschiedenartigen Datenträgern (optischen Speichern, Videobändern, ...), so dass sie für möglichst viele Anwendungen auffindbar und zugreifbar (nutzbar) sind

- **Was sind Multimedia-Datenobjekte?**

Digitalisierte, im Rechner abgelegte Bilder und Tonaufnahmen sowie Texte und Graphiken; Videoaufzeichnungen, Hologramme, Radarsignale, ...

- **Ziel**

- einerseits Erhöhung der „Informations-Bandbreite“, bessere Benutzerschnittstellen
- aber auch mehr Information im System:

➔ Inhalt eines Bildes (einer Tonaufnahme, ...) ist niemals vollständig in Medien wie Text oder Graphik wiederzugeben

- **Bezeichnungen**

- **Medienobjekt** (oder Medien-Datenobjekt)
Ein Datenobjekt, das einem einzigen Medium angehört, also ein einzelnes Bild, ein Textstück
- **Multimedia-Objekt**
(Multimedia-Datenobjekt, auch „mixed-mode object“)
Aggregation (Komposition) von Medienobjekten unterschiedlichen Typs, z. B. Video (Bild + Ton)
- **Multimedia-Daten**
Sammelbegriff für Medienobjekte und Multimedia-Objekte

- **Medienobjekte**

sind aus formatierten und unformatierten Daten zusammengesetzt

Exkurs - Medienobjekte

- **Was ist eigentlich alles zu speichern?**

- **Rohdaten**

- unformatiert: lange Folge (Menge, ...) von kleinen Elementen (Bits, Buchstaben, Pixel, Linien, Energieniveaus, ...)

- **Registrierungsdaten** (Steuerungsdaten)

- obligatorisch

- erforderlich für korrekte **Interpretation und Identifikation** der Rohdaten

- **Beschreibungsdaten**: optional

- oft redundant:

- Darstellung der **Struktur** und/oder des **Inhalts** in einem anderen Medium, formatiert oder unformatiert

- **Operationen auf Medienobjekten**

- **Erzeugung** (Eingeben, Erfassen, capture)

- mit Gerät – aus Programm – aus Datei

- **Ausgeben** (Zeigen, Präsentieren, present)

- auf Gerät – an Programm – in Datei

- **Bearbeiten** (Modifizieren, Editieren)

- **Zusammensetzen**

- Erzeugen multi-medialer Objekte

- **Weitergeben** (Versenden)

- **Archivieren**

- **Auswerten** (Aggregieren, Ableiten)

- Filtern, Analysieren, Erzeugen von Beschreibungsdaten

- **Suchen** (Vergleichen)

- Mustererkennung / Ähnlichkeitssuche auf Rohdaten;
oder inhaltsorientierte Suche auf Beschreibungsdaten

Beispiel: Medienobjekt „Rasterbild“

- **Rohdaten**

- Matrix von Bildpunkten
(Pixel = Picture Element, auch „Pel“ genannt)

- **Registrierungsdaten**

- Anzahl Bits pro Pixel („Farbtiefe“, typisch: 1, 8, 24);
- Anzahl Pixel pro Zeile (Breite des Bildes);
- Anzahl der Zeilen (Höhe des Bildes);
- Art der linearen Abspeicherung
 - zeilenweise oder spaltenweise;
- Bedeutung eines Pixels:
 - Grauwert, Farbdefinition, Index einer Farbtabelle (Palette, Colormap);
- ggf. Farbtabelle mit bestimmter Anzahl von Einträgen und Länge dieser Einträge (meist 24 Bit)
- ggf. Art der Farbdefinition: RGB, IHS, YIQ, ... und anderes mehr

- **Beschreibungsdaten**

- Text, Schlüsselworte, Wissensrepräsentation
- Graphik: erkannte Linien und Flächen
- darauf aufbauend zweidimensionale Objekte wie Kreise, Rechtecke usw.

Beispiel: Medienobjekt „Rasterbild“

Abstrakte Sicht auf eine Instanz vom Datentyp IMAGE:

IMAGE

Registrierungsdaten

Höhe

Breite

Tiefe

Codierung

Farbtabelle

Länge

Tiefe

Rohdaten

Bitmap/Pixelmatrix

Beschreibungsdaten

Der neue Fußballweltmeister

Dieses Ereignis fand am 9. Juli 2006 statt

Ort des Geschehens war Berlin

...

Zusammenfassung

• Transaktionsparadigma

- macht weitreichende Zusicherungen für die Verarbeitung von DB-Daten
- ACID-Eigenschaften müssen in einer Rechnerumgebung (aufwendig) nachgebildet werden
- erlaubt die Implementierung von „Vertragsrecht“

• Information und Informationssysteme

- Daten: objektive Welt der nicht-interpretierten Daten
- Information: subjektive Welt der bewerteten Daten
- Heterogenität, Wachstum, Anforderungsvielfalt u. a. führen oft auf unabhängige IS, die zusammen als kooperatives IS die angestrebte Leistung erbringen müssen

➔ „grob“: DBS + AWS = KIS

- wichtige Anwendungsklassen für
 - operierende Ebene: OLTP
 - planende/kontrollierende Ebene: DW, OLAP
 - strategische Ebene: OLAP, DSS

• Daten in Informationssystemen

- **strukturierte** Daten
 - fest vorgegebene Satz-/Tabellenstruktur
 - Bedeutung durch Metadaten weitgehend vorgegeben
- **unstrukturierte** Daten
 - „lange“ Werte bilden Dokumente (typischerweise in IRS)
 - Bedeutung nur schwach vorgegeben: Interpretation durch benutzerdefinierte Funktionen
- **semistrukturierte** Daten
 - HTML: Vermischung von Strukturierung und Präsentation
 - Verbesserung durch XML: Datenaustausch, Interoperabilität, Datenintegration
- **Multimedia**-Daten
 - Beschreibung durch Rohdaten, Registrierungsdaten, Beschreibungsdaten
 - Verarbeitung durch medienspezifische Operationen

Beispiele für Informationssysteme

- **Informationssystem einer Universität**

Die Universitätsdatenbank ist die Sammlung aller für die Abwicklung der an einer Universität anfallenden Verwaltungsaufgaben benötigten Daten.

Eine Universität gliedert sich i. allg. in mehrere Fachbereiche, denen sowohl die Studenten als auch die Professoren zugeordnet sind.

Die Studenten belegen verschiedene Vorlesungen von Professoren und legen bei ihnen Prüfungen ab.

Typische Anwendungen sind z. B.:

Immatrikulation der Studienanfänger, Rückmeldung der Studenten, Ausfertigen von Studentenausweisen und Studienbescheinigungen, Stundenplanerstellung und Planung der Raumbelugung, Ausstellen von (Vor)diplomzeugnissen, Exmatrikulationen, Statistiken über Hörerzahlen, Raumauslastung, Prüfungsergebnisse, etc.

- **Informationssystem eines Produktionsbetriebes**

In einem Produktionsbetrieb werden Daten über die verschiedenen Abteilungen und deren Beschäftigte mit ihren Familienangehörigen gespeichert.

Die Angestellten arbeiten an verschiedenen Projekten mit. Jedes Projekt benötigt für seine Durchführung bestimmte Teile. Jedes Teil kann von Lieferanten bezogen werden. Die Projekte werden jeweils von einem Projektmanager geleitet.

Die in einem Betrieb hergestellten Endprodukte setzen sich i. allg. aus mehreren Baugruppen und Einzelteilen zusammen.

Typische Anwendungen sind z. B.:

Einstellung und Entlassung von Personal, Lohn- und Gehaltsabrechnung, Bestellung und Lieferung von Einzelteilen, Verkauf von Fertigprodukten, Lagerhaltung, Bedarfsplanung, Stücklistenauflösung, Projektplanung.

Beispiele für Informationssysteme (2)

• Informationssystem einer Bank

Eine Bank gliedert sich gewöhnlich in mehrere Zweigstellen auf. Die Angestellten der Bank gehören jeweils fest zu einer bestimmten Zweigstelle. Auch die Bankkunden sind immer einer Zweigstelle zugeordnet. Es sind Daten über die verschiedenartigen Konten der Bankkunden bereitzustellen, wie z. B. Girokonten, Sparkonten, Hypothekenkonten, Kleinkreditkonten, Wertpapierkonten, etc.

Typische Anwendungen sind z. B.:

Buchung von Zahlungsvorgängen auf den verschiedensten Konten, Einrichten und Auflösen von Konten, Kreditgewährung bzw. Bereitstellen von Daten über die Kreditwürdigkeit eines Kunden, Zinsberechnung und -verbuchung, sowie alle Vorgänge der Personalverwaltung wie z. B. Gehaltsabrechnung.

Zur Rolle rechnergestützter Informationssysteme im Bankenbereich:

“In banking, by contrast, the data actually is the inventory – the two are synonymous. In increasingly many cases, the DB transaction is the financial transaction. There are no real, tangible tokens (greenbacks) moved as a result of the monetary transfer transaction. If the data is bad, money is lost or created. There is no possibility of counting the money (bits) in order to verify the status. Fiscal responsibility dictates that creating or destroying money – even temporarily – is unacceptable.”
(Mike Burman, Bank of America)

• Straßeninformationssystem

Beschreibung

Die Straßendatenbank ist Bestandsnachweis für das Straßennetz eines Bundeslandes. Es gibt verschiedene Straßentypen (Autobahnen, Bundes-, Kreis-, Gemeindestraßen etc.) innerhalb eines Bundeslandes. Die Straßen sind aus Planungs- und Verwaltungsgründen in Abschnitte eingeteilt, die durch jeweils zwei Netzknoten, welche Abschnittsanfang und Abschnittsende markieren, definiert sind.

Organisation

Die Straßenabschnitte sind jeweils einem Bauamt zugeordnet, das für Planungsarbeiten und die geometrische Festlegung des Straßenverlaufs entlang des Abschnitts zuständig ist. Die Kosten der anfallenden Arbeiten am Straßenabschnitt trägt der Baulastträger (Gemeinde, Kreis etc.). Die Straßenabschnitte gehen durch Gemeinden. Die Gemeinden gehören zu Kreisen.

Beispiele für Informationssysteme (3)

- **Straßeninformationssystem (Fortsetzung)**

Geometrische Darstellung

Für die geometrische Festlegung der Netzknoten ist jeweils ein Bauamt zuständig. Verschiedene Bauämter können innerhalb eines Kreises für Abschnitte oder Netzknoten zuständig sein. Ein Straßenabschnitt kann mehrere Äste aufweisen (z. B. Aufteilung in 2 Einbahnstrecken). Ein Ast kann sich auch aus mehreren Abschnitten zusammensetzen. Eine Straße kann an einem Netzknoten unterbrochen sein und an einem anderen Netzknoten weiterführen. Ein Straßenabschnitt kann auf mehreren Straßen (z. B. Bundesstraße und Kreisstraße) gleichzeitig verlaufen.

Topologische Information

Zusätzlich sind jedem Straßenabschnitt Daten zugeordnet, welche den geometrischen Verlauf zwischen den begrenzenden Netzknoten festlegen (Trassierungselemente: Kreise, Geraden, Klothoiden). Die Bauwerke (Brücken, Durchlässe, Signalanlagen etc.) sind dem geometrischen Verlauf des Straßenabschnitts ebenso zugeordnet wie Fußgängerüberwege, Radwege, Gehsteige, Daten des Fahrbahnaufbaus, Höheninformation, Entwässerungsschächte etc.

Besonderheiten

Unfalldaten, Verkehrsmengen, Frostsicherheit etc. sind weitere Attribute zum Straßenabschnitt.

Zeit

Die Straßendatenbank ermöglicht die Entnahme von Spezialplanungsunterlagen (z. B. Radwege, Gehwege) aber auch regionale Vergleiche des Straßennetzes und die Entnahme statistischer Daten.

Typische Fragen

Auswahl aller Kreisstraßen im Kreis mit Breite < 5m und NN-Höhe > 500m.

Zusammenstellung aller Strecken mit Radwegen getrennt für Ortsdurchfahrt und freie Strecke.

Auswahl aller Bundesstraßenstrecken im Bauamt mit Neigungen größer als 7%.

Berechnung der befestigten Straßenfläche für alle im Jahr 1980 gebauten Bundesstraßenstrecken im Bundesland.