Prof. Dr.-Ing. Stefan Deßloch

AG Heterogene Informationssysteme

Zi. 36/329, Tel.: 0631-205-3275 E-Mail: dessloch@cs.uni-kl.de

http://wwwlgis.informatik.uni-kl.de/cms/his/staff/dessloch/

# Informationssysteme

Sommersemester 2009

TU Kaiserslautern
Fachbereich Informatik
Postfach 3049
67653 Kaiserslautern

#### **Vorlesung:**

Dienstag, 15.30 - 17.00 Uhr, Raum 46-220

und

Donnerstag, 10.00 - 11.30 Uhr, Raum 46-220

Beginn: 23.4.2009

<u>Übung:</u>

n. V., s. Aushang





# Was sind Informations systeme (IS)?

#### Computergestützte Programmsysteme,

- die Informationen erfassen, dauerhaft speichern, verarbeiten, verändern, analysieren, bereit stellen, anzeigen

#### Betriebliche Informationssysteme

- spiegeln Geschäftsmodell eines Unternehmens wider
- organisieren und unterstützen Arbeitsabläufe
- integrieren eine Vielzahl von Datenquellen

#### Web-basierte Informationssysteme

- stellen Informationsdienste über Web-Schnittstellen zur Verfügung
- unterstützen die Abwicklung von Geschäftsvorgängen über das Internet (E-Business)

#### • Beispiele

- Universitätsinformationssystem verwaltet Daten über Fachbereiche, Studenten, Professoren, Vorlesungen und Prüfungen, um typische Verwaltungsaufgaben zu unterstützen.
- Produktionsbetrieb speichert Daten über Abteilungen, Mitarbeiter, Produkte,
   Projekte und Lieferanten, um Vorgänge wie Gehaltsabrechnung, Fertigung und Versand von Produkten, Lagerhaltung abzuwickeln
- Eine Bank verwaltet Daten über Filialen, Kunden und Konten zur Durchführung von Zahlungsverkehrs, Kundenverwaltung, Geldanlage, Zinsberechnung, etc.
- Ein Internet-Auktionshaus speichert Daten zu Produkten, Käufern, Verkäufern und Auktionen und unterstützt die Suche nach in Auktionen angebotenen Waren, den Informationsbedarf bzgl. der Vertrauenswürdigkeit der Verkäufer und das Durchführen der eigentlichen Auktionen.





# Rolle von Datenbanksystemen bei der Entwicklung von IS

#### • Vielfältige, wiederkehrende Anforderungen und Aufgaben

- Vermeidung von Datenredundanz und Inkonsistenzen
- Flexible und mächtige Zugriffs- bzw. Anfragemöglichkeiten
- Verwaltung großer persistenter Datenmengen (TB bis PB)
- Unterstützung von Mehrbenutzerbetrieb
- Zuverlässigkeit und Fehlertoleranz
- Überwachung von Integritätsbedingungen
- Zugriffskontrolle

#### ⇒ Informationsorienter Entwurf von IS

- anwendungs-/funktionsübergreifende Datenmodellierung

# ⇒ Einsatz von Datenbanksystemen zur Übernahme von kritischen Datenverwaltungsaufgaben

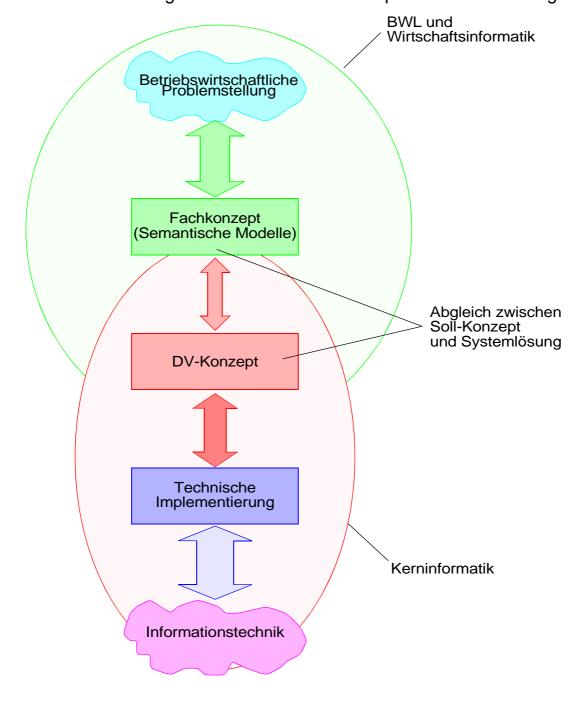
- Zuverlässige Verwaltung großer Datenmengen
- Effizienter Datenzugriff
- Deklarative Anfragesprachen (SQL)
- Transaktionskonzept
- Zugriffsprivilegien für Benutzer, Rollen





# **Entwicklung von Informationssystemen**

- Sicht auf den Themenbereich (nach Scheer)
  - Was ist die Gesamtaufgabe?
  - BWL-Anforderungen an die Weiterentwicklung der Informationstechnik
  - IT eröffnet neue Möglichkeiten für BWL-Konzepte und -Anwendungen<sup>1</sup>



<sup>1.</sup> Die Abenteuer von heute finden in den Computern und den Laboratorien statt. (Herbert Spencer)





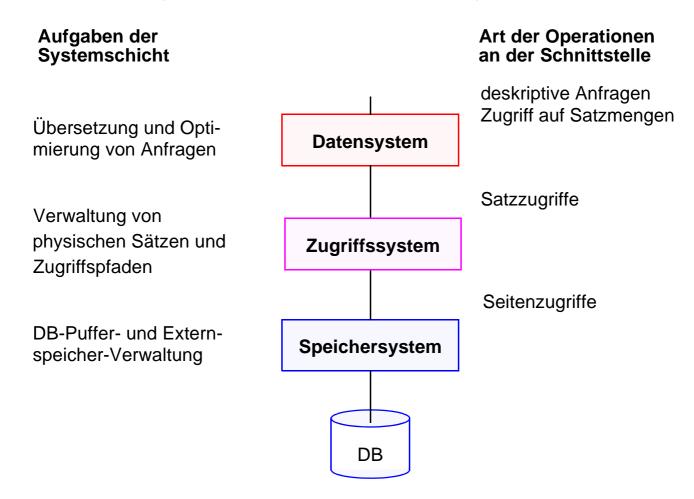
# **DBS als Kern von Informationssystemen**

#### Kern der Vorlesung

- Modellierung, Verwaltung und Sicherung von strukturierten Daten (Sätze)
- Es gibt verschiedenartige Datenmodelle und die sie realisierenden DBS
  - relational und objekt-relational (RDBS/ORDBS auf Basis von SQL)
  - hierarchisch (DBS nach dem Hierarchiemodell)
  - netzwerkartig (DBS nach dem Codasyl-Standard)
  - objektorientiert (OODBS)
  - Verarbeitung von semi- oder unstrukturierten Daten

#### • Bezugspunkt ist ein Schichtenmodell

- Erklärung der Abläufe im DBS
- Verfeinerung der Sichtweise in späteren Vorlesungen



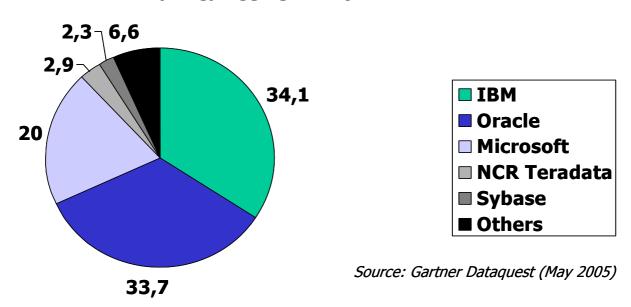




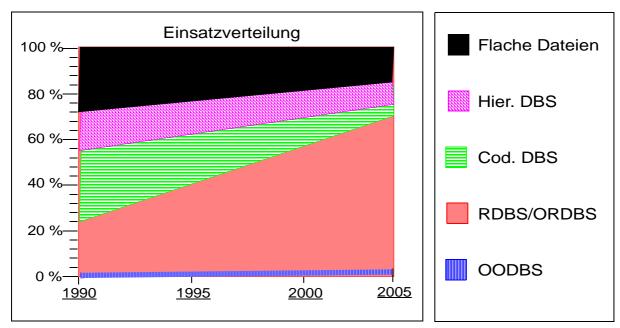
# **DBS als Kern von Informationssystemen (2)**

Wie sieht der Herstellermarkt aus?

#### Marktanteile in %



Verteilung von DBS und Dateien



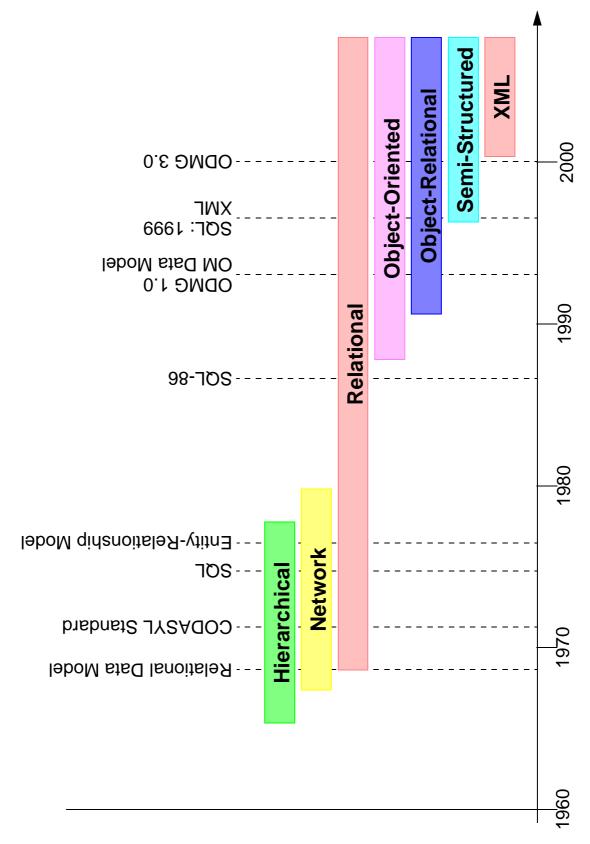
## Künftige DBS

- Aufstellung berücksichtigt nur strukturierte Daten. 85% der weltweit verfügbaren Daten aber sind semi- oder unstrukturiert (Internet, wiss. Aufzeichnungen und Experimente usw.)
  - SQL-XML-DBS, XML-SQL-DBS, native XML-DBS





# **Evolution und Historie**

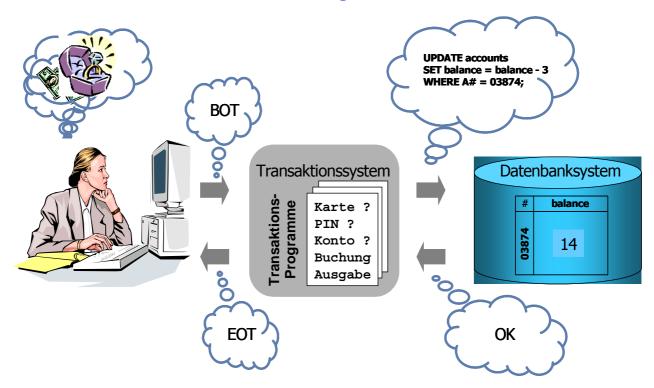




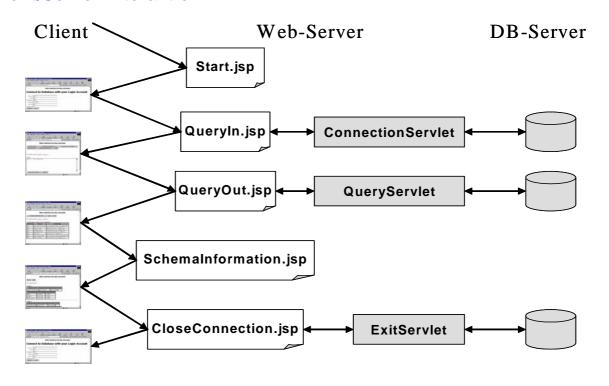


# Wie hängt das alles zusammen?

### • Klassische Transaktionsverarbeitung



#### • Client/Server-Interaktion



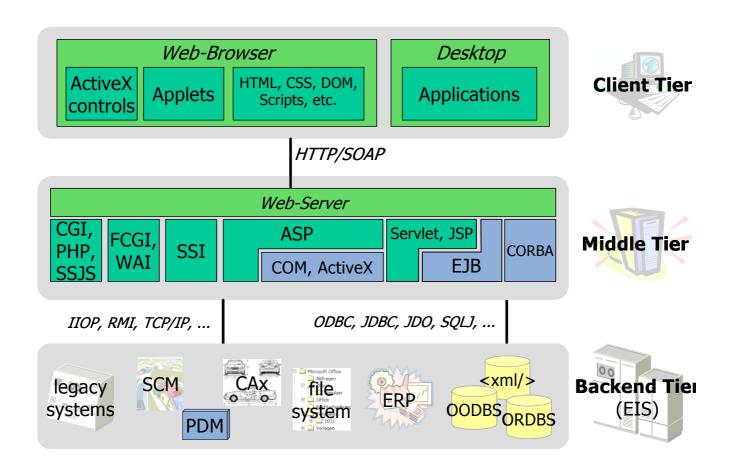




# Wie hängt das alles zusammen? (2)

#### Architektur von Web-Applikationen

- klassische Schichtenarchitektur (3-tier)
- Zusammenstellung von wichtigen Sprachen, Protokollen, Komponenten usw.
- Client/Server- und Web-Applikationen werden nicht in dieser Vorlesung behandelt
- Sie sind Gegenstand von Vertiefungsvorlesungen und Praktika







# Ziele der Vorlesung

- Vermittlung von Grundlagen- und Methodenwissen<sup>2</sup> sowie Erwerb von Fähigkeiten und Fertigkeiten im Bereich Datenbanken und Informationssysteme:
  - Nutzung von Informations- und Datenmodellen, insbes.
    - Entity/Relationship-Modell und Erweiterungen
    - Relationenmodell
  - Entwurf, Aufbau und Wartung von Datenbanken
    - Modellierung von Miniwelten und Entwicklung von DB-Schemata
    - Einsatz von SQL als DB-Standard-Sprache
  - Sicherung der Abläufe in Datenbanken
    - Transaktionskonzept mit ACID-Eigenschaften
    - Mehrbenutzerbetrieb, semantische Integrität, Fehlerbehandlung
  - Verwaltung und Handhabung semi- und unstrukturierter Daten/Dokumente
    - Markup-Sprachen, Web-Zugriff
    - XML-Datenmodell und XQuery
  - Geschäftsprozessmodellierung und Abbildung auf Workflows
- Voraussetzungen für Übernahme von Tätigkeiten:
  - Entwicklung von betrieblichen Anwendungs- und Informationssystemen, insbesondere DB-gestützte Anwendungen
  - Planung und Realisierung von Geschäftsprozessen
  - Systemverantwortlicher für Datenbanksysteme, insbes. Unternehmens-, Datenbank-, Anwendungs- und Datensicherungsadministrator

<sup>2.</sup> Grundlagenwissen ist hochgradig allgemeingültig und nicht von bestimmten Methoden abhängig. Die Halbwertzeit ist sehr hoch. Methodenwissen muss ständig an die aktuelle Entwicklung angepasst werden. In der Informatik haben sich die entscheidenden Methoden alle 8-10 Jahre erheblich geändert. Werkzeugwissen ist methodenabhängig. Werkzeuge haben in der Informatik oft nur eine Lebensdauer von 2-3 Jahren.





# ÜBERSICHT (vorl.)

# 0. Übersicht und Motivation

- Was sind Informationssysteme?
- Rolle von Datenbanksystemen

# 1. Einführung und Grundbegriffe

- Miniwelt modellhafte Abbildung
- Information was ist das?
- Aufgaben eines Informationssystems (IS)
- Daten in Informationssystemen
- Beispiele für Informationssysteme

# 2. E/A-Architektur und Zugriff

- E/A-Architektur von Informationssystemen
- Einsatz von Speicherhierarchien
- Datenstrukturen auf Externspeichern
- B-Bäume und B\*-Bäume
- Informationssuche bei strukturierten Daten

#### 3. Informationsmodelle

- Vorgehensweise bei DB-Entwurf und -Modellierung
- Entity-Relationship-Modell (ERM)
- Erweiterungen, insbes. Abstraktionskonzepte
  - Generalisierung
  - Aggregation und Assoziation





# ÜBERSICHT (2)

# 4. Grundlagen des Relationenmodells

- Konzepte des Relationenmodells (RM)
- Relationenalgebra
  - Operationen
  - Algebraische Optimierung
- Abbildung ERM → RM

### 5. Die Standardsprache SQL (SQL92)

- Anfragesprache
- Datenmanipulation
- Datendefinition
- Abbildung von Beziehungen (Referentielle Integrität)
- Wartung von Beziehungen (Referentielle Aktionen)

# 6. Transaktionskonzept

- Überblick über ACID-Eigenschaften
- Gewährleisten von Atomarität
- Sicherung der Datenintegrität
- Serialisierbarkeit und Synchronisation (Concurrency Control)
- Synchronisationsprotokolle: Überblick
- Aufgaben und Ziele der Fehlerbehandlung





# ÜBERSICHT (3)

#### 7. Semistrukturierte Daten

- XML-Einführung
- Datenmodellierung mit XML
- Neue Sprachkonzepte
- Anfrageverarbeitung mit XPath und XQuery (Überblick)

#### 8. Unstrukturierte Daten

- Informationssuche bei strukturierten und unstrukturierten Daten
- Information Retrieval, Suche und Dokumentenzugriff im Web

# 9. Modellierung und Verwaltung von Geschäftsprozessen

- Anforderungen
- Geschäftsprozessmodellierung, Modellierungsaspekte und -sprachen
- Petri-Netze und Workflows
- Workflow-Management





#### LITERATURLISTE

- Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval, Addison-Wesley, 1999
- Baumgarten, B.: Petri-Netze Grundlagen und Anwendungen, 2. Auflage, Spektrum Akademischer Verlag, 1996
- Elmasri, R., Navathe, S.: Grundlagen von Datenbanksystemen, 3., überarbeitete Auflage, Pearson Studium, 2005
- Faeskorn-Woyke, H., Bertelsmeier, B., Riemer, P., Bauer, E.: Datenbanksysteme Theorie und Praxis mit SQL2003, Oracle und MySQL, Pearson Studium, 2007
- Kemper, A., Eickler, A.: Datenbanksysteme Eine Einführung, 6. aktualisierte und erweiterte Auflage, Oldenbourg Wissenschaftsverlag, 2006
- Kemper, A, Wimmer, M.: Übungsbuch Datenbanksysteme, Oldenbourg Wissenschaftsverlag, 2006
- Korth, H.F., Silberschatz, A.: Database System Concepts, 3rd Edition, McGraw-Hill Book Comp., New York, 1997
- Kudraß, T. (Hrsg.): Taschenbuch Datenbanken, Fachbuchverlag Leipzig, 2007
- Lausen, G.: Datenbanken Grundlagen und XML-Technologien, Elsevier, München, 2005
- Pernul, G., Unland, R.: Datenbanken im Unternehmen: Analyse, Modellbildung und Einsatz, Oldenbourg-Verlag, 2001
- Saake, G., Sattler, K.-U., Heuer, A.: Datenbanken Konzepte und Sprachen, 3. Auflage, mitp, 2008
- Vossen, G.: Datenmodelle, Datenbanksprachen und Datenbankmanagementsysteme, 5. Auflage, Oldenbourg Wissenschaftsverlag, 2008

#### **ZEITSCHRIFTEN:**

Transactions on Database Systems, ACM Publikation (vierteljährlich)

Information Systems Pergamon Press (6-mal jährlich)

The VLDB Journal (vierteljährlich)

Informatik - Forschung und Entwicklung (vierteljährlich)

#### TAGUNGSBÄNDE:

SIGMOD Tagungsband, jährliche Konferenz der ACM Special Interest Group on

Management of Data

VLDB Tagungsband, jährliche Konferenz "Very Large Data Bases"

IEEE Tagungsband, jährliche Konferenz "Int. Conf. on Data Engineering"

GI Tagungsbände der Tagungen der Gesellschaft für Informatik, Tagungen inner-

halb des Fachausschusses 2.5 Informationssysteme





# Zum Nachdenken – Wie viele Informationen gibt es auf der Welt?<sup>3</sup>

#### Was ist überhaupt Information?

- schwierige Abgrenzung, individuelle Bewertung
- Was wird aufgezeichnet, was lohnt sich aufzuheben?
- Redundante Information (Exemplare eines Buchs, Kopien eines Films) wird hier ausgeschlossen
- Nicht alle Daten, die weltweit durch Sensoren, Mess- und Aufnahmegeräten usw. (Experimente, Überwachung, ...) erzeugt werden, führen zu relevanten Informationen (in unserem Sinne)

#### Annahme

- Um alle relevanten Informationen aufzuheben, genügt die Speicherung von wenigen Tausend PBytes<sup>4</sup>
- Die Produktion von Magnetplatten und -bändern hat diese Speicherkapazität bereits um das Jahr 2000 erreicht

#### • Behauptung: In wenigen Jahren

- sind wir in der Lage, "alles" aufzuheben,
   d. h., keine Information muss weggeworfen werden
- werden die "typischen" Informationen nur noch von Rechnern aufbewahrt, gesucht und aufbereitet; der Mensch sieht weder die Daten, noch kennt er den Aufbewahrungsort und die genauen Ableitungsverfahren.

3. http://www.lesk.com/mlesk/ksg97/ksg.html

4. 1 Gigabyte (GByte) = 1.000 Megabytes = 10<sup>9</sup> Bytes

1 Terabyte (TByte) = 1.000 Gigabytes 1 Petabyte (PByte) = 1.000 Terabytes 1 Exabyte (EByte) = 1.000 Petabytes 1 Zettabyte (ZByte) = 1.000 Exabytes 1 Yottabyte (YByte) = 1.000 Zettabytes





### Schlussfolgerungen

- Es wird genug Platten- und Bandspeicher geben, um alles zu speichern, was alle Menschen schreiben, sagen, tun oder photographieren.
  - Für das Schreiben gilt dies bereits heute
  - In einigen Jahren trifft das auch für die restlichen Informationen zu
  - Wie lange wird es noch dauern, bis alle VITA-Dokumente eines Menschen als Lebensgeschichte aufgezeichnet werden können?
- Rechner speichern und verwalten Informationen besser und effektiver als Menschen
  - Nach dem Jahr 2000 werden viele Platten und Kommunikationsverbindungen direkt Informationen aus Rechner-zu-Rechner- und nicht mehr (nur) aus Mensch-zu-Mensch-Kommunikation speichern
  - Wie lange wird es noch dauern, bis der Mensch die meiste gespeicherte Information gar nicht mehr zu sehen bekommt?
  - Wir müssen lernen, wie alles automatisch ausgewertet werden kann und was bei unserer knappen Zeit unserer besonderen Aufmerksamkeit bedarf.

### • Künftige Entwicklung

- Heute konzentriert man sich bei den "Digitalen Bibliotheken" auf die Eingabe: auf das Scanning, Komprimieren und OCR von Informationen.
- Morgen wird anstelle der Eingabe die "relevante Auswahl" die wesentliche Rolle spielen: Selektion, Suche und Qualitätsbewertung von Informationen
- ⇒ Wir können eine reale "World Encyclopedia" mit einem echten "planetary memory for all mankind" aufbauen, wie H.G. Wells bereits 1938 in seinem Buch "World Brain" geschrieben hat!





# Betriebliche Informationssysteme<sup>5</sup>

Betriebliche Informationssysteme spiegeln die Geschäftsmodelle von Unternehmen wider und dienen dazu, deren Arbeitsabläufe zu organisieren und zu unterstützen. Darum sind sie wichtig.

Betriebliche Informationssysteme sind stark datenbankbasierte Anwendungen, oft mit sehr vielen Benutzern (Tausende und mehr). Es sind transaktionsverarbeitende Systeme, d.h., sie erbringen ihre Leistung in vielen, kleinen Schritten für die gleichzeitig zugreifenden Benutzer. Dabei müssen sie die Integrität der Daten gewährleisten sowie hohen Durchsatz und kurze Antwortzeiten schaffen. Betriebliche Informationssysteme laufen heutzutage typischerweise auf einem leistungsfähigen Server und präsentieren sich dem Benutzer mit einer grafischen Oberfläche als Client/Server-Systeme, zunehmend auch via Internet. Sie sind aber nicht nur Dialogsysteme, sondern benötigen meist auch einen Batch, der Massenverarbeitung effizient außerhalb des Dialogs abwickelt.

Ein **Data Warehouse** ist ein spezieller Aspekt betrieblicher Informationssysteme. Es akkumuliert und verdichtet die Daten aus den operativen Transaktionssystemen zu dem Zweck, die **Geschäftsentwicklung zu analysieren** und darauf aufbauend Entscheidungen zur Steuerung des Unternehmens zu treffen.

Die Software für betriebliche Informationssysteme ist meist sehr komplex und umfangreich (einige Hunderttausend bis Millionen Zeilen Programmcode), ihre Entwicklung verursacht erheblichen Aufwand und kostet viel Geld. Und sie lebt lange (zehn bis zwanzig Jahre und mehr), weshalb ihre Wartungskosten erheblich sind und die der Erstentwicklung oft übersteigen. Es lohnt sich folglich, Software von vornherein wartungsfreundlich und erweiterbar zu konstruieren.

Ohne betriebliche Informationssysteme könnten Unternehmen ihr Geschäft nicht betreiben. Das gilt in der industriellen Produktion ebenso wie im Dienstleistungssektor, ganz besonders jedoch für jene Unternehmen, die im Kern reine Informationsverarbeiter sind, z.B. Banken, Versicherungen, Touristikanbieter. Bei allen wirken die IT-Systeme tief auf die Unternehmensorganisation ein. Betriebliche Informationssysteme dienen dem Kerngeschäft von Unternehmen und damit Investitionsgüter von strategischer Bedeutung.

Diese Beschreibung der Aufgaben von betrieblichen Informationssystemen und von E-Business ist dem Buch "sd&m - Ein Software- und Beratungshaus stellt sich vor" (mit unwesentlichen Änderungen) entnommen (sd&m: software design & management, München, 2001).





#### **E-Business**

E-Business ist die Nutzung des Internets zu geschäftlichen Zwecken aller Art. Das fängt an mit der E-Mail und dem Herunterladen von Software per File-Transfer, geht weiter mit der Darstellung des Unternehmens, seiner Produkte und seiner Stellenangebote im World Wide Web sowie mit dem Intranet als elektronischem schwarzem Brett und Werkzeug des Wissensmanagements und schließt E-Commerce mit ein.

Bei E-Commerce fließt Geld, denn es geht um Handel, also den Abschluss und die Abwicklung von Kaufverträgen. Dabei werden Varianten unterschieden, je nachdem, wer mit wen handelt: ein Unternehmen mit seinen Endkunden (Business-to-Consumer, B2C), Unternehmen untereinander (Business-to-Business, B2B) oder Endkunden direkt miteinander über Börsen und Auktionen (Consumerto-Consumer, C2C). Es lohnt sich zudem, nach dem "Aggregatzustand" der gehandelten Produkte zu unterscheiden, d.h. zwischen dem Handel mit materiellen Produkten und jenem mit Informationen; denn die Geschäftsmodelle sind völlig verschieden. Bei materiellen Gütern kommt es entscheidend auf die Logistik an, also auf den Versand. Für unempfindliche Artikel (z.B. Bücher, Kleidung) genügt ein Paketdienst, für verderbliche und zeitkritische Ware (Blumen, die zur Geburtstagsfeier eintreffen sollen) muss ein spezieller Lieferservice eingerichtet werden. Für den reinen Informationshandel (z. B. mit Musik oder dem kostenpflichtigen Nachschlagen in einem E-Lexikon) ist das Internet der ideale Transportweg, nur zahlt dafür niemand gerne. Geschäftlich interessant dagegen ist Information, hinter der eine Dienstleistung steht (z. B. das Buchen eines Fluges, verbunden mit dem Kauf eines Tickets, oder der Abschluss einer Versicherung).

E-Business braucht starke Softwaresysteme. Es sind komplexe Systeme, denn es genügt nicht, sich mit einer gut gestalteten Web-Oberfläche dem Benutzer zu präsentieren – werblich ansprechend, um ihn zu gewinnen, ergonomisch, um ihn nicht zu verlieren. Dahinter muss mehr stehen: eine flexible Anwendung, die sich schnell an geänderte Geschäftsprozesse anpassen lässt, und eine gehaltvolle Datenbank. Aber das ist noch nicht alles. Das Internet erfordert erhöhte Sicherheit und Bezahlung in neuen Formen. Tausende von Anwendern greifen gleichzeitig zu und erwarten rasche Reaktion. Weiterhin ist Anwendungsintegration von zentraler Bedeutung: Vorhandene (Legacy-) Systeme müssen über das Internet zugänglich gemacht und Anwendungen verschiedener Unternehmen darüber direkt verbunden werden. Kurzum, Software-Engineering für komplexe Systeme ist gefragt: @business @ngineering.





# E-Business (2)

E-Business-Systeme sind nichts anderes als betriebliche Informationssysteme, allerdings mit besonderen Anforderungen: Erstens sind es eigentlich überbetriebliche Informationssysteme, denn sie verbinden über ein Unternehmen hinausgehend Mitarbeiter, Lieferanten und Kunden und werden vor allem von Menschen genutzt, die nicht Angestellte des betreibenden Unternehmens sind. Diese Anwender sind unbekannt, d. h., man kann sie beim Entwurf des Systems nicht befragen, man kann sie nicht schulen, ihnen keine Anweisungen erteilen, und man muss mit Ungeschicklichkeit oder gar böswilligen Attacken rechnen. Dennoch kommt es gerade auf sie an, denn es sind (potenzielle) Kunden (im B2C-Geschäft) oder Mitarbeiter von guten Geschäftspartnern (B2B). Die Ergonomie der Benutzeroberfläche ist also in bisher nicht bekannter Weise erfolgskritisch. Ebenso sind es die Organisation der Geschäftsprozesse, die nun über das eigene Unternehmen hinausreichen, sowie bestimmte Sicherheitsmaßnahmen (Firewall, Verschlüsselung, elektronisches Bezahlen).

Zweitens muss man beim Bauen und Integrieren von E-Business-Systemen eine noch nie da gewesene Komplexität der Technologie beherrschen. Man muss sich mit der Programmierung der Web-Oberfläche auskennen (HTML, XML, Java-Applets etc.), Netzprotokolle (z.B. HTTP) und Web-Server einzusetzen verstehen, Anwendungs-programme in Java schreiben und unter der Transaktionskontrolle von Application-Servern zum Laufen bringen, Standard-Internet-Anwendungen (z.B. Intershop) sowie vorhandene (Legacy-) Systeme (CICS, Cobol, DB2) integrieren. Damit nicht genug: Systeme für E-Mail, Verschlüsselung, Zahlung und anderes müssen eingebunden werden. Schließlich ist das Ganze so zu integrieren, zu testen und zu tunen, dass es korrekt, zuverlässig und schnell läuft. Die zu verknüpfende Technologie reicht vom Mainframe über Client/Server-Systeme zum Internet.

Weitere Unterstützung: Ein **Webservice** ist eine mit einem Uniform Resource Identifier (URI) eindeutig identifizierbare Software-Anwendung, deren Schnittstellen als XML-Artefakte definiert, beschrieben und gefunden werden können. Ein Webservice unterstützt die direkte Interaktion mit anderen Software-Systemen unter Verwendung XML-basierter Nachrichten durch den Austausch über internetbasierte Protokolle.

Ein Beispiel ist die Interaktion zwischen Fluggesellschaften und Reisebüros. Die Fluggesellschaften stellen Möglichkeiten zum Nachschlagen bzw. Buchen von Flügen über einen Webservice bereit. Die Reisebüros bieten auf ihrem Web-Auftritt Flüge verschiedener Fluggesellschaften an, von denen die Reisebüros zur Laufzeit über UDDI erfahren. Der Kunde kann so beim Reisebüro zentral Preise und Termine verschiedener Flüge vergleichen und direkt buchen (http://de.wikipedia.org/wiki/Webservice).



