Dr. N. Ritter
AG Datenbanken und Informationssysteme
Zi. 36/338, Tel.: 0631-205-3264
E-Mail: ritter@informatik.uni-kl.de
http://wwwdbis.informatik.uni-kl.de/

Datenbanksysteme I

Wintersemester 2000/2001

Universität Kaiserslautern
Fachbereich Informatik
Postfach 3049
67653 Kaiserslautern

Vorlesung:

Ort: 46 - 110

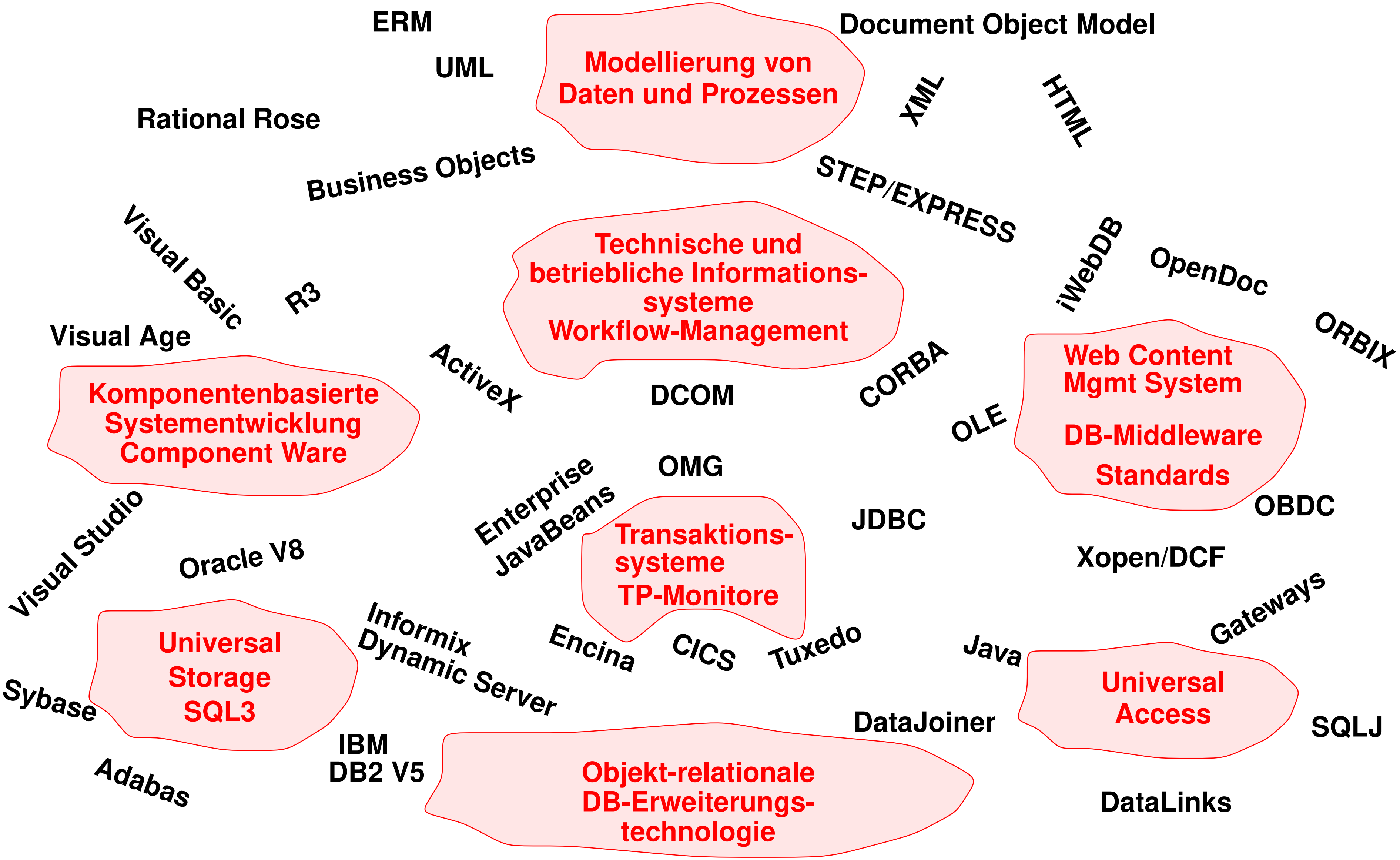
Zeit: Mo., 10.00 - 11.30 Uhr
und
Ort: 46 - 110

Zeit: Mi., 11.45 - 13.15 Uhr

Übung:

Ort: 36 - 265

Zeit: Do, 15.30 - 17.00 Uhr


ERM
UML
Rational Rose
Modellierung von
Daten und Prozessen
Document Object Model
XML
HTML
Business Objects
STEP/EXPRESS
Visual Basic
R3
Technische und
betriebliche Informations-
systeme
Workflow-Management
iWebDB
OpenDoc
ORBIX
Visual Age
Komponentenbasierte
Systementwicklung
Component Ware
ActiveX
DCOM
CORBA
OLE
Web Content
Mgmt System
DB-Middleware
Standards
OMG
Enterprise
JavaBeans
OBDC
JDBC
Transaktions-
systeme
TP-Monitore
Xopen/DCF
Visual Studio
Oracle V8
Gateways
Universal
Storage
SQL3
Informix
Dynamic Server
Encina
CICS
Tuxedo
Java
Universal
Access
Sybase
SQLJ
DataJoiner
IBM
DB2 V5
Objekt-relationale
DB-Erweiterungs-
technologie
Adabas
DataLinks

0 - 2

Forschungsthemen der AG Datenbanken und Informationssysteme

2000
Web-basiertes Lehren und Lernen,
iWebDB

1999
Erfahrungs-DB
Entwicklung großer SW-Systeme

1998
DB-Middleware,
Anwendungsintegration mit WFs

1997
Erweiterungstechnologie für ORDBMS

1996
Workflow-Management
Web und Datenbanken
Mehr-Server-Architekturen

1995
Objekt-Relationale DBS
RITA - IS für technische Anwendungen

1994
Multi-User KRISYS
Standardisierung
STEP/EXPRESS/SDAI

1993
Kooperation und Koordination
CONCORD
Entwicklung von Frameworks

1992
PHOENIX
Mehrrechner-DBS

1991
Client/Server-Architekturen
Workstation/Server-DBS
neue Transaktionskonzepte

1990
Parallele u. Verteilte DBS
Multimedia-DBS

1989
KRISYS
Wissenbankverwaltungs-systeme

1988
PRIMA
DB-Kern-Architektur

1987
TECHMO : CAD-Entwurfsunterstützung
MAD-Modell

1986
XPS-Kopplung
RDA

1985
Transaktionsysteme (HLTS) DB-Sharing-Systeme

1980
relationale und netzwerkartige Datenmodelle und DBS
- Leistungsanalyse
- ACID-Eigenschaften
- TP-Monitore

Ziele

- **Vermittlung von Grundlagen- und Methodenwissen[1] im Bereich betrieblicher Informationssysteme; Erwerb von Fähigkeiten und Fertigkeiten**
  - in der Nutzung von Informations- und Datenmodellen, insbesondere
    - Entity/Relationship-Modell und Erweiterungen
    - Relationenmodell und SQL
    - objektorientierte und objekt-relationale Datenmodelle
  - in der Modellierung von anwendungsbezogenen Realitätsausschnitten (Miniwelten, Diskursbereiche)
  - in der Programmierung von DB-Anwendungen
  - im Entwerfen, Aufbauen und Warten von Datenbanken

- **Voraussetzungen für Übernahme von Tätigkeiten:**
  - Entwicklung von datenbankgestützten Anwendungen
  - Nutzung von Datenbanken unter Verwendung von (interaktiven) Datenbanksprachen
  - Systemverantwortlicher für Datenbanksysteme, insbesondere Unternehmens-, Datenbank-, Anwendungs- und Datensicherungsadministrator

1.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

1. Einordnung von DBS

- Betriebliche Informationssysteme (IS)
- Strukturierte, semi-strukturierte und unstrukturierte Daten
- Klassen von DB-Anwendungen

2. Anforderungen und Beschreibungsmodelle

- Nachteile von Dateisystemen
- Anforderungen an DBS
- Aufbau von DBS
- Beschreibungsmodelle
  (Fünf-Schichten-Modell, Drei-Ebenen-Architektur nach ANSI/SPARC)

3. Informations- und Datenmodelle

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

4. Grundlagen des Relationalen Datenmodells

- Konzepte des Relationenmodells
- Abbildung ERM --> RM
- Relationenalgebra
  - Operationen
  - Algebraische Optimierung

ÜBERSICHT (2)

5. Die Standardsprache SQL (SQL2)

- Anfragesprache
- Datenmanipulation
- Datendefinition
- Abbildung von Beziehungen
- Wartung von Beziehungen

6. Anwendungsprogrammier-Schnittstelle

- Schemaevolution
- Sichtenkonzept
- Kopplung mit einer Wirtssprache
- Eingebettetes / Dynamisches SQL
- WWW-Anbindung von Datenbanken

7. Transaktionskonzept

- Das Transaktionsparadigma
- Synchronisation von Transaktionen
- Konsistenzebenen
- Commit-Protokolle

8. Integritätskontrolle und aktives Verhalten

- Semantische Integritätskontrolle
- Regelverarbeitung in DBS
- Trigger-Konzept von SQL
- Definition und Ausführung von ECA-Regeln

ÜBERSICHT (3)

9. Datenschutz und Zugriffskontrolle

- Technische Probleme des Datenschutzes
- Zugriffskontrolle in SQL
- Sicherheitsprobleme in statistischen DBs

10. Logischer DB-Entwurf

- Konzeptioneller DB-Entwurf
- Normalformenlehre (1NF, 2NF, 3NF, 4NF)
- Synthese von Relationen

11. Einführung in objekt-relationale Datenmodelle (SQL3)

- Abstrakte Datentypen
- Typhierarchien, Subtypen, . . .
- Erweiterbarkeit

12. . . .

- . . .

LITERATURLISTE

Bernstein, P.A., Hadzilacos, V., Goodman, N.: Concurrency Control and Recovery in Database Systems, Addison-Wesley Publ. Comp., 1987.

Chamberlin, D.: DB2 Universal Database, Addison-Wesley Publ. Comp., 1999

Date, C.J.: An Introduction to Database Systems, Addison-Wesley Publ. Comp., Reading, Mass., 7th Edition, 2000

Dreßler, H.: Datenstrukturentwurf – Vom Faktenchaos zur Anwenderdatenbank, Oldenbourg-Verlag, München, 1995

Guluzan, P., Pelzer, T.: SQL-99 Complete, Really, R&D Books Miller Freeman, Inc., Lawrence, Kansas, 1999

Heuer, A., Saake, G.: Datenbanken – Konzepte und Sprachen, 2. Auflage, Int. Thompson Publ. Comp., 2000

Kemper, A., Eickler, A.: Datenbanksysteme – Eine Einführung, 3. Auflage, Oldenbourg-Verlag, 1999

Ramakrishnan, R.: Database Management Systems, McGraw-Hill, Boston, 1998.

Vossen, G.: Datenbankmodelle, Datenbanksprachen und Datenbankmanagementsysteme, 4. Auflage, Oldenbourg-Verlag, München, 2000

ZEITSCHRIFTEN:

TODS 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 innerhalb des Fachausschusses 2.5 Informationssysteme

und viele weitere Konferenzreihen

Datenstrukturen ?

- **Bisher bekannte Datenstrukturen**
  - Felder (Reihungen, Arrays, ...)
  - Verbunde (Tupel, Sätze, Records, ...)
  - Listen
  - Graphen
  - Bäume

  → bisher im **Hauptspeicher**,
  d. h. Bestand nur für die Dauer einer Programmausführung
  („transiente“ Daten)

- **hier nun neue Aspekte:**
  Nutzung von Hintergrundspeicher und neuen Operationen
  - **Persistenz:**
    Werte bleiben über Programmende, Sitzungsende,
    Betriebssystem-Uptime, Rechnereinschaltung, ...
    hinaus erhalten
  - andere Arten des Zugriffs: Lese- und Schreiboperationen,
    in vorgegebenen Einheiten von Blöcken und Sätzen

    → Strukturen und zugehörige Algorithmen (Suchen, Sortieren),
    die im Hauptspeicher optimal sind, sind es auf Sekundärspeicher
    nicht unbedingt!
  - gezielter, wertabhängiger Zugriff auch auf sehr große Datenmengen
  - umfangreiche Attributwerte (z. B. Bilder, Texte)

Datenmodelle ?

- **Mengen von Konstruktoren**

  zur (abstrakten) Erzeugung von Datenstrukturen
  mit darauf definierten Operatoren

  z. B. Tabellen (Relationen): Mengen von gleichartig strukturierten Tupeln

```
CREATE TABLE STUDENT
        (MATRIKELNUMMER        INTEGER,
        NACHNAME               VARCHAR(40),
        FBNUMMER               INTEGER,
        GEBURTSDATUM           CHAR(8),
        .... )

CREATE TABLE FACHBEREICH
        (FBNUMMER              INTEGER,
        FBNAME                 VARCHAR(20),
        DEKAN                  PROF,
        .... )
```

  nicht nur ein einzelner Satz, sondern Menge

- **Wichtige Rolle von Beziehungen**

  in einigen Datenmodellen auch sehr spezielle Arten von **Beziehungen** zwischen Sätzen und/oder Tupeln: Hierarchien, Zusammensetzungen, funktionale Zuordnungen u.v.a.m.

- **Modellbegriff**

  vorgegebene Menge von (sprachlichen) Ausdrucksmitteln, mit denen Diskursbereich beschrieben (erfaßt) werden muß

  → auch Programmiersprachen und Betriebssysteme haben ein „Datenmodell“

Datenbanken ?

- **große Datenmengen:**
  - auch, aber **nicht immer** entscheidend

- **Datenunabhängigkeit (der Anwendungen):**
  - Benutzung der Daten, ohne Details der systemtechnischen Realisierung zu kennen (abstraktes „Datenmodell“, z. B. Tabellen)
  - einfache Handhabung der Daten, mächtige Auswertungsoperationen

- **Offenheit der Daten für neue Anwendungen (Anwendungsneutralität der Speicherung)**
  - symmetrische Organisationsformen (keine Bevorzugung einer Verarbeitungs- und Auswertungsrichtung)
  - explizite Darstellung der Annahmen (nicht in den Anwendungsprogrammen verstecken)
  - Redundanzfreiheit aus der Sicht der Anwendung:
keine wiederholte Speicherung in unterschiedlicher Form
für verschiedene Anwendungen
  - Konsistenzüberwachung durch das Datenbanksystem

Datenbanken ? (2)

- **Transaktionskonzept mit Garantie von ACID-Eigenschaften**
  - Atomizität (*atomicity*)
  - Konsistenz (*consistency*)
  - Isolation (*isolated execution*)
  - Dauerhaftigkeit (*durability*)

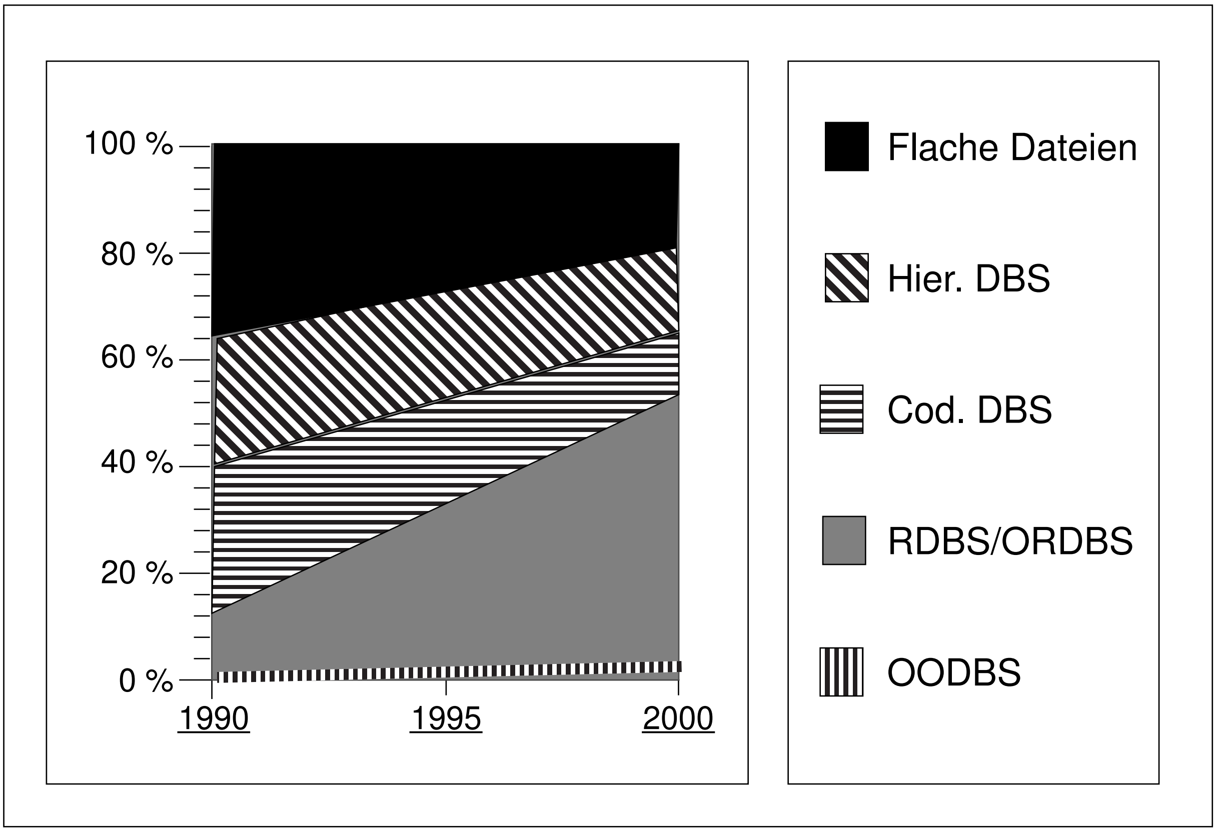
- **Ausfallsicherheit**
  - Aufzeichnung redundanter Daten im Normalbetrieb
  - Replikation von Datenstrukturen
  - automatische Reparatur der Datenbestände nach Programm-, System- und Gerätefehlern
    - Rückgängigmachen unvollständiger Transaktionen, so daß sie wiederholt werden können
    - Wiederherstellen der Ergebnisse vollständiger Transaktionen, so daß sie nicht wiederholt werden müssen

- **Mehrbenutzerbetrieb**
  - gleichzeitiger (zeitlich eng verzahnter) Zugriff verschiedener Anwendungen und Benutzer auf gemeinsame Daten
  - Synchronisation, d. h. Vermeidung von Fehlern in der wechselseitigen Beeinflussung
  - Kooperation über gemeinsame Daten mit hohem Aktualitätsgrad

Verteilung von DBS und Dateien

- **Es gibt verschiedenartige Datenmodelle und die sie realisierenden DBS**
  - relational und objekt-relational (RDBS/ORDBS)
  - hierarchisch (DBS nach dem Hierarchiemodell)
  - netzwerkartig (DBS nach dem Codasyl-Standard)
  - objektorientiert (OODBS)

	1990	1995	2000
OODBS	1 %	2 %	3 %
RDBS/ORDBS	9 %	25 %	50 %
Cod. DBS	30 %	22 %	12 %
Hier. DBS	25 %	20 %	15 %
Flache Dateien	35 %	31 %	20 %
Gesamt	100 %	100 %	100 %



Wie viele Informationen gibt es auf der Welt?[1]

- **Was ist überhaupt Information?**
  - z. T. 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[2]
  - Die Produktion von Magnetplatten und -bändern erreicht diese Speicherkapazität um das Jahr 2000

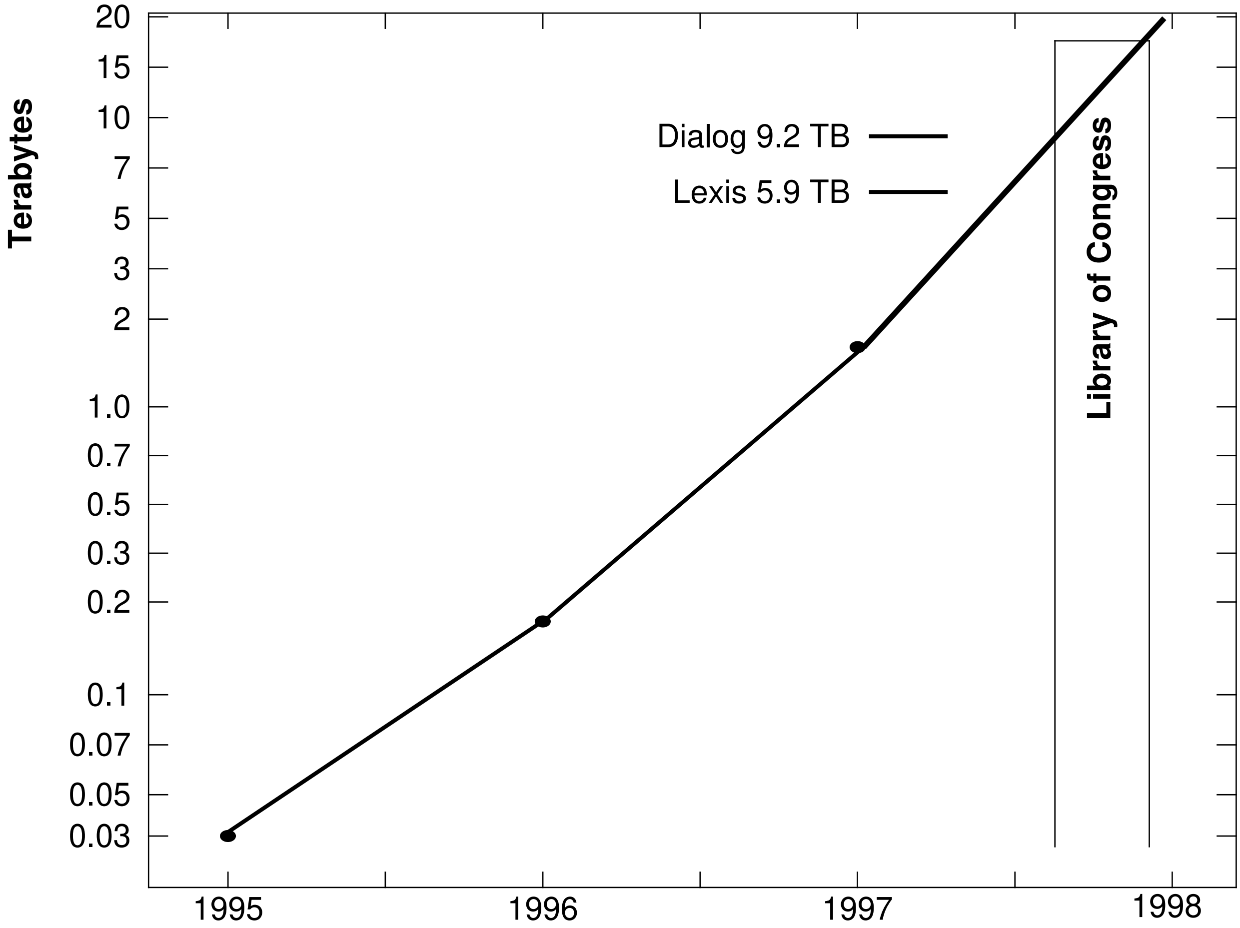
- **Behauptung:**

  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.

---

1. http://www.lesk.com/mlesk/ksg97/ksg.html
2. 1 Gigabyte (GByte) = 1,000 Megabytes = 10^9 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 Yotabyte (YByte) = 1,000 Zettabytes

Web-Größe



nur Text-Dateien (Ascii-Daten)

Vergleichsmaß ist Library of Congress (LC): 20 Mio Bücher * 1 MByte ~ 20 TBytes
Dialog: http://www.dialog.com/
Lexis: http://www.lexis-nexis.com/

- **Wachstum des Web** (Schätzung 1998)
  - Faktor 10 pro Jahr
  - Wie lange kann es so weiter gehen?
  - Anzahl der Benutzer: Faktor 20?
    heute: ~ 100 Mio → 2 Mrd?
  - Wieviel mehr Daten (Text) als heute stellt der durchschnittliche Benutzer online ins Web: Faktor 20?

→ 8000 TBytes ASCII-Daten im Web?

Andere Informationsarten

- **LC hat neben gedrucktem Text weitere Informationsquellen, die weit mehr Speicherplatz benötigen:**
  - 13 Mio Photos ergeben, selbst komprimiert auf 1 MBytes JPG, noch 13 TBytes
  - 4 Mio Landkarten lassen sich auf ~ 200 TBytes einscannen
  - 0.5 Mio Filme mit jeweils 1 GByte (komprimiert, meist nicht in Farbe und voller Länge) benötigen 500 TBytes
  - 3.5 Mio Tonaufzeichnungen, jeweils auf einer Audio-CD, erreichen fast 2,000 TBytes.

  → Gesamter Umfang der LC-Informationen: ~ 3 PBytes

- **Vergleich Web – LC**
  - LC enthält vorwiegend publiziertes Material
  - Im Web stehen bereits mehr englischsprachige Texte, die in den letzten 18 Monaten geschrieben wurden, als in der LC
  - Viele URLs enthalten Informationen, die man auch in einer großen Bibliothek finden kann (~ 28 %)
  - Inhalt dieser Dokumente (Lexis-Nexis und Dialog) ist jedoch besser aufbereitet (Hypertext) und einfach zu durchsuchen

- **Erinnerung**
  - Vor Gutenberg (~ 1450) gab es in ganz Europa nur ~ 30,000 Bücher, fast ausschließlich Bibeln und Bibelkommentare[1]
  - Um das Jahr 1500 gab es über 9 Mio Bücher über alle möglichen Themen. Flugschriften und sonstige Druckwerke beeinflußten Politik, Religion, Wissenschft und Literatur

  → Der *Information Highway* wird unsere Kultur ebenso dramatisch verändern wie Gutenbergs Druckerpresse die Welt des Mittelalters

1. James Burke: „In dieser Welt war jegliche Erfahrung persönlich: Die Horizonte waren eng, die Gemeinschaft nach innen gekehrt. Die Außenwelt kannte man nur vom Hörensagen.“

Menschliches Gedächtnis

- **Wieviel Information kann das menschliche Gehirn aufnehmen?**
  - 200 MBytes an Information
    T.K. Laudauer: „How much do people remember? Some estimates of the quantity of learned information in long-term memory“, Cognitive Science 10 (4) pp. 477-493 (Oct-Dec 1986).
    - Studie zur Aufnahme- und Vergessensrate sowie zur Nutzung bei normalen Tätigkeiten
    - Bei 10^{12}-10^{14} Neuronen kann man davon ausgehen, daß das Gehirn 10^{3}-10^{5} Neuronen für jedes Bit des Gedächtnisses enthält
    - Der größere Teil des Gehirns wird für die Wahrnehmung, Bewegungskontrolle usw. genutzt

  → Even if only 1 % of the brain is devoted to memory, it looks like your head accepts considerable storage inefficiency in order to be able to make effective use of the information (Landauer).

- **Mit der Landauer'schen „Schätzung“ und den ~ 6 Mrd Menschen auf Erden erhält man als gesamtes Speichervermögen aller menschlichen Gedächtnisse 1,200 PBytes.**

  → Rechner können heute alles digital speichern, was alle Menschen wissen oder an was sie sich erinnern können

- **Diese Schätzungen sind ungenau. Jedoch wäre es auch bei einem anderen „Aufnahmemodell“ kein Problem, die Informationsmenge eines Menschen auf eine Laptop-Platte zu speichern.**
  - Menschliche Aufnahmerate beim Langzeitgedächtnis: 1 Byte/sec nach Landauer
  - Gesamtmenge in einem menschlichen Leben: 2 GBytes (2 Mrd Sek.)

Menschliches Gedächtnis (2)

- **Selbst wenn der Mensch sein Leben lang kein Wort vergessen würde, ist die Speicherung der Informationsmenge heute schon machbar (auf einer Platte)**

  Bilanz (durchschnittl. Zahlen für US-Bürger/Jahr)

	Stunden	Wort/Minute	Wörter/Jahr	MBytes
TV	1578	120	11 Mio	50
Film	12	120	-	-
Lesen	354	300	6.4 Mio	32

→ ~ 6 GBytes von ASCII-Daten (Text) in 75 Jahren

Die automatische Erfassung dieser Informationen (Spracherkennung, und OCR oder elektronische Bücher/Zeitungen und ASCII-Skripte von TV-Sendungen) ist in naher Zukunft mit tragbaren Geräten möglich. Was bedeutet das beispielsweise für Schüler/Studenten?
Wie weit gehen diese Möglichkeiten? (Hörsaal mit Laptop-/Notebooks)

→ Das gilt nicht für Ton, Bild und Bewegtbild

- **Aber in 20 Jahren?**

  Speicherbedarf für eine Person (nach Jim Gray)

Datentyp	**Datenrate (Bytes/ Sek.)**	**benötigter Speicher pro Stunde und Tag**	**benötiger Speicher für eine Lebenszeit**
gelesener Text	50	200 KB; 2-10 MB	60-300 GBytes
gesproch.Text @120 wpm	12	43 K; 0.5 MB	15 GBytes
Sprache (komprimiert)	1,000	3.6 MB; 40 MB	1.2 TBytes
Bewegtbild (komprimiert)	500,000	2GB; 20 GB	1 PByte

→ Aufzeichnung der gesamten Lebensgeschichte wird möglich!

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, 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!