

Informationssysteme

**Grundvorlesung Informatik
Sommersemester 2004
Universität des Saarlandes, Saarbrücken**

Dr. Ralf Schenkel und Prof. Dr. Gerhard Weikum
 {schenkel, weikum}@mpi-sb.mpg.de
<http://www.mpi-sb.mpg.de/units/ag5/teaching/ss04/is.htm>

Informationssysteme SS2004 1-1

Organisation

- **Vorlesung:** Di 9-11 und Do 9-11 in 27/001
 Sprechstunde Dr. Schenkel:
 Sprechstunde Prof. Weikum: Di 14-15 in 36/326 oder n.V. (ggf. e-mail)
- **Übungsgruppen:** Mi und Fr
 Übungsgruppenleiter:
 Übungsgruppenleiter: Christian Zimmer (czimmer@mpi-sb.mpg.de)
 Sergej Sizov (sizov@mpi-sb.mpg.de)
 Erster Übungsgruppentermin: in der zweiten Semesterwoche
- **Leistungsprüfung:**
 - erfolgreiche Teilnahme an zwei von drei Klausuren:
 1) Sa 12.6., 2) Sa 24.7., 3) Mitte Oktober
 - erfolgreiche Bearbeitung der praktischen Übungen
 (Teamarbeit in Dreiergruppen möglich)
 - Präsentation mindestens einer Übungslösung in der Übungsgruppe

Informationssysteme SS2004 1-2

Gliederung

1. Einführung und Überblick: Anwendungen, Systeme, Prinzipien
2. Vektorraummodell für Suchmaschinen
3. Automatische Klassifikation von Dokumenten
4. Linkanalyse für Autoritäts-Ranking
5. Relationenmodell und algebraorientierte Anfragesprachen
6. Logikorientierte Anfragesprachen
7. Datenbanksprache SQL
8. Anwendungsentwicklung mit SQL und JDBC
9. Integritätssicherung mit SQL
10. Objektorientierte und objekt-relationale Datenmodelle
11. Relationale Entwurfstheorie
12. Datenbankentwurf mit UML
13. Prozessmodellierung mit Statecharts
14. Datenspeicherung, Indexstrukturen, Anfrageauswertung
15. Transaktionsverwaltung: Concurrency-Control
16. Transaktionsverwaltung: Recovery
17. OLAP und Data-Warehouses
18. Daten-Mining: Klassifikation, Assoziationsregeln
19. Semistrukturierte Daten und XML-Anfragesprachen

Teil I:
Suchmaschinen

Teil II:
Datenbankschnittstellen

Teil III: Entwurf
von Datenbanken
und Anwendungen

Teil IV:
Implementierungskonzepte
von DBS

Teil V:
Informationsdienste

1-3

Informationssysteme SS2004 1-3

Kapitel 1: Einführung und Übersicht – Anwendungen, Systeme, Prinzipien

- 1.1 Anwendungen**
- 1.2 Systemarchitekturen**
- 1.3 Grundprinzipien von Datenbanksystemen**
- 1.4 Grundprinzipien von Suchmaschinen**

Informationssysteme SS2004 1-4

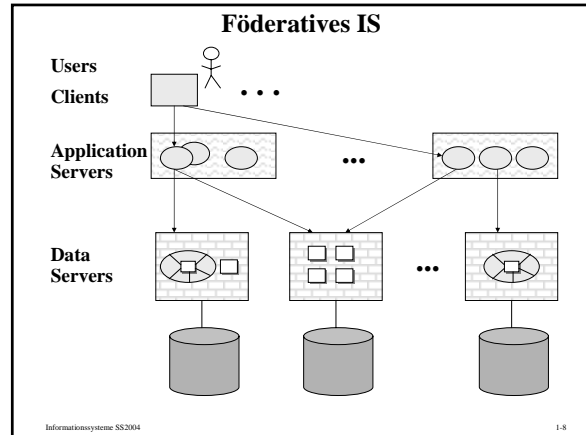
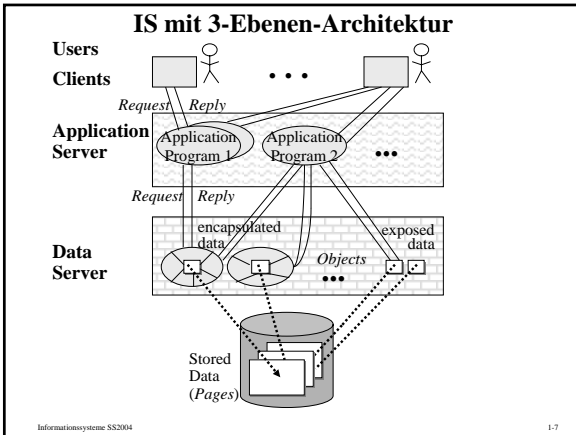
Warum sind Datenbanken und IS so wichtig ?

Informationssysteme SS2004 1-5

Technische Anforderungen an IS

- Textsuche mit nach Relevanz geordneten Trefferranglisten
- Automatische Organisation von Dokumenten
- Verwaltung von (Deep-)Web-Daten, strukturierten Datenbanken und semistrukturierten (XML-)Daten
- Komfortable GUIs (Graphical User Interfaces)
- deklarative Anfragesprachen
- Zuverlässige Verwaltung sehr großer, persistenter Daten (> 5 TB)
- Hohe Verfügbarkeit (7 x 24) und langfristige Archivierung
- Gute - vorhersagbare und garantierbare - Leistung:
 hoher Durchsatz, kurze Antwortzeiten („Quality of Service“)
- Konsistenz verteilter Daten
- Integrierter Zugriff auf heterogene Daten
- Daten-Mining nach Korrelationen, Regeln, Klassifikationen, etc.
- Aktive Regeln und Prozesskoordination
- Komplexe Datentypen (Landkarten, Satellitenbilder, Himmelskarten, ...)
- Ähnlichkeitssuche auf Bildern, wissenschaftlichen Daten, etc.
- Integration mit Web-Applikation und Web-Services

Informationssysteme SS2004 1-6



Grundtugenden von DBS

- **Effiziente Verwaltung großer, persistenter Datenmengen mit Optimierung der Sekundärspeicherzugriffe**
- **Programm-Daten-Unabhängigkeit: Kapselung der Speicherungsstrukturen, so daß Optimierungen transparent für Anwendungsprogramme möglich sind**
- **Gewährleistung der Datenkonsistenz durch das DBS**
- **Datenbankänderungen innerhalb von Transaktionen: atomar, konsistenzhaltend, isoliert, dauerhaft**

Informationssysteme SS2004 1-9

Grundtugenden von Suchmaschinen

- **Kompakte Repräsentation von Textdokumenten und Multimedia-Daten als Feature-Vektoren**
- **Ähnlichkeitssuche mit Resultats-Ranking nach Relevanz**
- **Algebraische Analysen und statistische Lernverfahren zur Organisation und Bewertung von Dokumenten**

Informationssysteme SS2004 1-10

Gliederung

1. Einführung und Überblick: Anwendungen, Systeme, Prinzipien
2. Vektorraummodell für Suchmaschinen
3. Automatische Klassifikation von Dokumenten
4. Linkanalyse für Autoritäts-Ranking
5. Relationenmodell und algebraorientierte Anfragesprachen
6. Logikorientierte Anfragesprachen
7. Datenbanksprache SQL
8. Anwendungsentwicklung mit SQL und JDBC
9. Integritätssicherung mit SQL
10. Objektorientierte und objekt-relationale Datenmodelle
11. Relationale Entwurfstheorie
12. Datenbankentwurf mit UML
13. Prozessmodellierung mit Statecharts
14. Datenspeicherung, Indexstrukturen, Anfrageauswertung
15. Transaktionsverwaltung: Concurrency-Control
16. Transaktionsverwaltung: Recovery
17. OLAP und Data-Warehouses
18. Daten-Mining: Klassifikation, Assoziationsregeln
19. Semistrukturierte Daten und XML-Anfragesprachen

Teil I: Suchmaschinen

Teil II: Datenbankschnittstellen

Teil III: Entwurf von Datenbanken und Anwendungen

Teil IV: Implementierungskonzepte von DBS

Teil V: Informationsdienste

2, 3, 4

5, 6, 7, 8, 9, 10, 11, 12, 13

14, 15, 16

17, 18, 19

Informationssysteme SS2004 1-11