

Kommunikation und Datenhaltung

Sommersemester 2009

IPD – *Systeme der*
Informationsverwaltung



Frank Eichinger, Mirco Stern
Institut für Programmstrukturen und Datenorganisation (IPD)

Prof. Dr. Martina Zitterbart
Institut für Telematik



- Prof. Dr. Martina Zitterbart
 - Gebäude 20.20 [Rechenzentrum], Raum 359
 - Sprechstunde: Montag 11:00 – 12:00 (alle 2 Wochen)
 - E-Mail: zit@tm.uka.de
- Frank Eichinger
 - Gebäude 50.34 [Informatikgebäude am Fasanengarten], Raum 340
 - Sprechstunde: Dienstag 14:00 – 15:00 Uhr
 - E-Mail: eichinger@ipd.uka.de
- Mirco Stern
 - Gebäude 50.34 [Informatikgebäude am Fasanengarten], Raum 340
 - Sprechstunde: Dienstag 14:00 – 15:00 Uhr
 - E-Mail: mirco.stern@ipd.uni-karlsruhe.de
- Martin Röhrich
 - Gebäude 20.20 [Rechenzentrum], Raum 355
 - E-Mail: roehricht@tm.uka.de
- Clemens Heidinger
 - Gebäude 50.34 [Informatikgebäude am Fasanengarten], Raum 336
 - E-Mail: heidinger@ipd.uni-karlsruhe.de

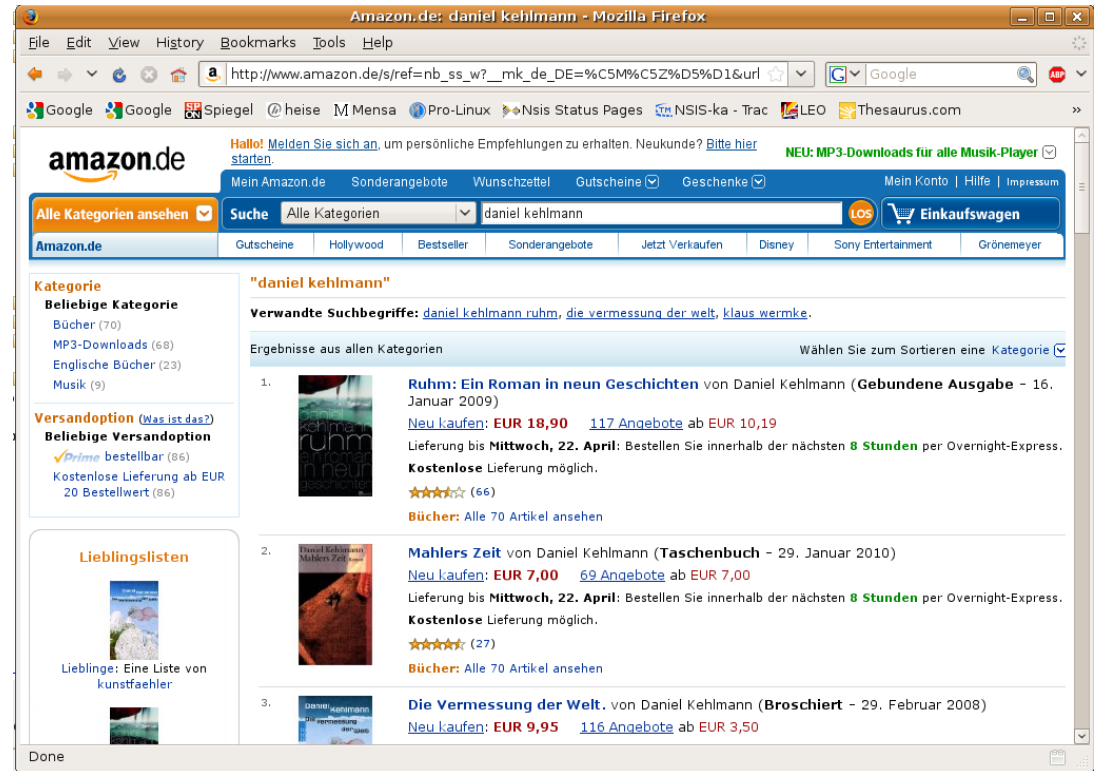
- Zentraler Gegenstand der Vorlesung:
Informationsaustausch
- Überwindung von
 - Raum und
 - Zeit
- Raum
 - Datenübertragung
- Zeit
 - Datenspeicherung

Verteilte Informationssysteme

- Definition „Verteilte Informationssysteme“
 - Informatik-Systeme, die sich dem Informationsaustausch in Raum und Zeit widmen.

Zusammenhang „K“ und „D“?

- Kernbestandteile der meisten komplexen vernetzten Systeme der Informatik
- Beispiel Amazon
 - Online-Marktplatz
 - Oracle RAC DBMS, Linux, 25 TB Daten
 - Umsatz 2008: \$52,8 Mio. pro Tag; \$35000 pro Minute



⇒ Enges Zusammenspiel von „K“ und „D“ vorteilhaft

- Erst „K“ und „D“ in Kombination ermöglichen die meisten Dienstangebote

Was erwartet Sie heute?

- Organisatorisches zur Vorlesung und Übung
- Vorstellung K&D-Team
- Datenhaltungsteil: Einstieg

Ablauf der Vorlesung

- Donnerstag, 23. April 2009:
 - Organisatorisches
 - Einführung Datenhaltung

- Datenhaltungsteil: Donnerstags 11:30 – 13:00 Uhr

- Kommunikationsteil: Freitags 09:45 – 11:15 Uhr

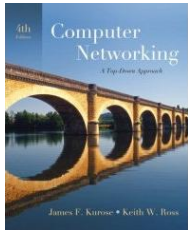
- (Aber: Tausch in den ersten 2 Wochen:
24.04.: D, 30.04.: K, 01.05.: Tag der Arbeit)

- Web-Seite der Vorlesung und Übung
 - Aktuelle Informationen
 - Alle Vorlesungsunterlagen
 - Alle Übungsunterlagen
 - ▶ Übungsblätter, Musterlösungen

<http://www.ipd.uni-karlsruhe.de/~komdat/>

Organisation der Übung

- Übungstermin ist Montags, 15:45 – 17:15 Uhr im Gaede-Hörsaal
- Übungsblätter werden eine Woche vor der jeweiligen Übung als PDF freigeschaltet
- Voraussichtliche(!) Übungstermine:
 - 25.05.2009 1. D-Übung
 - 08.06.2009 1. K-Übung
 - 15.06.2009 2. D-Übung
 - 22.06.2009 2. K-Übung
 - 29.06.2009 3. D-Übung
 - 13.07.2009 3. K-Übung
 - 20.07.2009 4. D-Übung
- An einigen Übungsterminen normale Vorlesung (siehe Ankündigungen auf der Webseite)
- Klausurtermin Kommunikation und Datenhaltung: 30. Juli 2009 um 14:00 Uhr



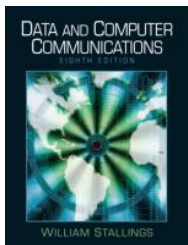
- J.F. Kurose, K.W. Ross: **Computer Networking – A Top-Down Approach**, Addison-Wesley, 4. Auflage, 2007
ISBN-13: 978-0321497703

- Auch auf Deutsch erhältlich



- F. Halsall: **Computer Networking and the Internet. With Internet and Multimedia Applications**, Addison-Wesley, 5. Auflage, 2005; ISBN-13: 978-0321263582

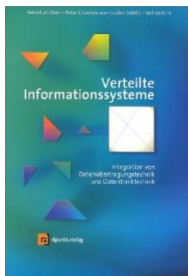
- Sehr gut zum Nachschlagen



- W. Stallings: **Data and Computer Communications**, Prentice Hall, 8. Auflage, 2006
ISBN-13: 978-0132433105

- S. Abeck, P. C. Lockemann, J. Seitz, J. Schiller: **Verteilte Informationssysteme**, dpunkt-Verlag, 1. Auflage, 2002
ISBN-13: 978-3898641883

- Früher praktisch Skript zur Vorlesung, K-Teil vieles noch gut passend, D-Teil eher mehr Änderungen

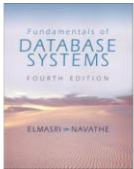




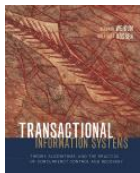
- Andreas Heuer, Kai-Uwe Sattler, Gunther Saake:
Datenbanken – Konzepte und Sprachen
3. Aufl., mitp-Verlag, Bonn, 2007



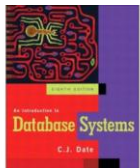
- Alfons Kemper, André Eickler:
Datenbanksysteme. Eine Einführung
6. Aufl., Oldenbourg Verlag, 2006



- R. Elmasri, S.B. Navathe:
Fundamentals of Database Systems
4. Auflage, Benjamin/Cummings, 2000



- Gerhard Weikum, Gottfried Vossen:
Transactional Information Systems
Morgan Kaufmann, 2002



- C.J. Date:
An Introduction to Database Systems
8. Auflage, Addison-Wesley, Reading, 2003

Was erwartet Sie heute?

- Organisatorisches zur Vorlesung und Übung
- Vorstellung K&D-Team
- Datenhaltungsteil: Einstieg

- **Professoren**

- Prof. Dr. Hannes Hartenstein (seit 2003)
- Prof. Dr. Martina Zitterbart (seit 2001)
- Prof. Dr. Wilfried Juling (seit 1998)
- Prof. Dr. Sebastian Abeck (seit 1996)
- em. Prof. Dr. Dr.h.c.mult. Gerhard Krüger (seit 1971)



- **Mitarbeiter**

- ca. 40 wissenschaftliche Mitarbeiter
- technische Mitarbeiter und Sekretärinnen/Verwaltungsangestellte

- **Studierende**

- ca. 35 Hiwis
- ca. 250 mündliche Prüfungen in Telematik etc. pro Jahr
- über 40 Diplomarbeiten pro Jahr

WS

<p>Advanced Web Applications</p>	<p>Multimedia-kommunikation</p>	<p>Hochleistungs-kommunikation</p>	<p>Drahtlose Sensor-Aktor-Netze</p>	<p>Netzwerk- und IT-Sicherheitsmanagement</p>	<p>Verkehrs-telematik</p>	<p>Ubiquitäre Informations-technologien</p>	<p>Web Engineering</p>
----------------------------------	---------------------------------	------------------------------------	-------------------------------------	---	---------------------------	---	------------------------

SS

<p>Advanced Web Applications (AWA)</p>	<p>Next Generation Internet</p>	<p>Mobilkommunikation</p>	<p>Netzicherheit – Architekturen & Protokolle</p>	<p>Simulation von Rechnernetzen</p>	<p>Parallelrechner und Parallelprogrammierung</p>
--	---------------------------------	---------------------------	---	-------------------------------------	---

WS

<p>Telematik</p>	<p>Telematik für Informationswirte</p>
------------------	--

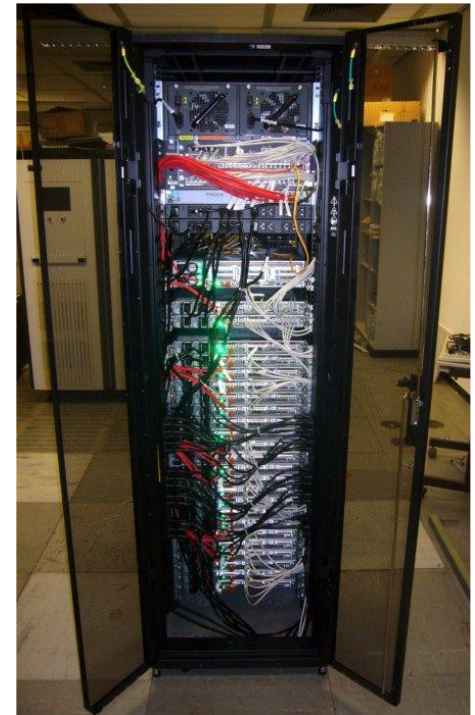
SS

<p>Kommunikation und Datenhaltung</p>



- Next Generation Networking
 - Quality-of-Service
 - Routing
 - Selbstorganisation
 - Sicherheit
 - Peer-to-Peer
- Mobilkommunikation
 - Multi-hop Ad-hoc Netze
 - Multicast Overlays
 - Mobilitätsmanagement
 - Diensteterminnung
 - Auto-Konfiguration
- Ambient Technologies
 - Sensornetze
 - Dienste-Beschreibung und -Erkennung
 - Sicherheit

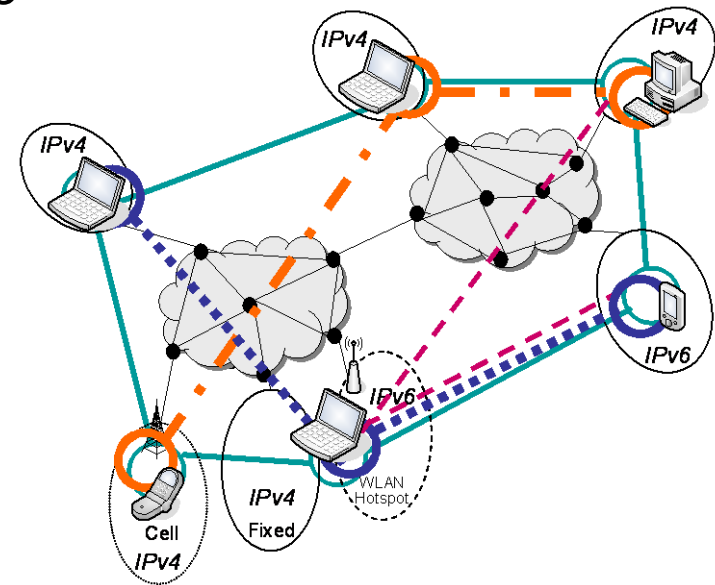
- Studien- und Experimentierplattform für das Internet der Zukunft
 - Gefördert durch BMBF
 - Förderzeitraum: 01.10.2008 – 30.09.2011
 - über 170 Knoten an 6 Standorten:
TU Berlin, Uni Kaiserslautern,
TU Darmstadt, Uni Würzburg,
TU München, KIT
- Ziele
 - Aktive Erforschung der Grundlagen für ein zukünftiges Internet
 - Aufbau einer deutschlandweiten Experimentierplattform
- Arbeitspakete
 - Virtuelle Netze
 - Angriffserkennung
 - Dienstkomposition





- EU Project FP7 (2008–2009)
 - 37 Projektpartner, Koordinator Ericsson
- Neue Netzarchitekturen für das Internet der Zukunft
 - Horizont 2020
 - Mobilität, Dienstgüte, Sicherheit von Anfang an integrieren
 - Grundthemen
 - ▶ „Lasst 1000 Netze blühen“
 - ▶ Lass die Netze sich selbst verwalten/managen
 - ▶ Lass einen Netzpfad eine aktive Einheit sein
 - ▶ Lass Netze Informations-zentrisch sein
- Arbeitspakete
 - Business Innovation, Regulation and Dissemination – BIRD
 - **New Architectural Principles and Concepts – NewAPC**
 - **Network Virtualisation – VNET**
 - In-Network Management – INM
 - Forwarding and Multiplexing for Generic Paths – ForMux
 - Network of Information – NetInf

- Definition von Konzepten und Mechanismen zur Erstellung von Spontanen Virtuellen Netzen
 - Bei Bedarf erzeugte Overlay-Netze mit Unterstützung für
 - ▶ Gruppenkommunikation
 - ▶ Dienstgüteunterstützung
 - ▶ Heterogene (z.B. Festnetz und zellulare) Netze
- Anwendungsbeispiele
 - Realzeitspiele
 - Video-Streaming
- Forschungsschwerpunkte am Institut für Telematik
 - Underlay-Abstraktion
 - Gruppenkommunikation
 - Messung und Charakterisierung von Spielerverkehr



- „Zuverlässige Informationsbereitstellung in energiebewußten ubiquitären Systemen“
 - Untersucht Energieeffizienz, Sicherheit und Robustheit in drahtlosen Sensornetzen
 - Gefördert von der Landesstiftung BW
 - <http://www.zeus-bw-fit.de/>
- Kooperation mit Datenbänklern (Prof. Böhm), Theoretikern (Prof. P.H. Schmitt), Algorithmikern (Prof. Wagner), OS-lern (Prof. Bellosa)
- **Zuverlässige** Informationsbereitstellung?
 - Sicher, gegen Angreifer?
 - Robust, gegenüber Umwelteinflüssen?



- „Sichere und flexible Grenz- und Liegenschaftsüberwachung durch drahtlose Sensornetze“
 - In Kooperation mit dem **Bundesamt für Sicherheit in der Informationstechnik (BSI)**
 - Überwachung der „grünen Grenzen“
 - Prototyp mit 250 Knoten wird entwickelt



- Falls Sie über die Lehrveranstaltungen hinaus Interesse haben, sich mit dem Fachgebiet vertraut zu machen, wie wäre es denn als

- Hiwi
- Studienarbeiter
- Diplomarbeiter
- ... oder als aktiver Teilnehmer an einer Arbeitsgemeinschaft?



- Sowohl die Mitarbeiter als auch Prof. Zitterbart stehen Ihnen hierzu gerne als Ansprechpartner zur Verfügung
- Schauen Sie doch einfach mal am Institut vorbei!
 - Informatikgebäude am Schloss (Geb. 20.20), 3. Stock

- Prof. Dr. Klemens Böhm
 - Leiter Lehrstuhl für Systeme der Informationsverwaltung am Institut für Programmstrukturen und Datenorganisation (IPD)
- Prof. em. Dr. Dr. h. c. Peter Lockemann
- Dr. Erik Buchmann
- Frank Eichinger
- Mirco Stern
- Sowie 14 weitere Mitarbeiter:
Lehrstuhl-Webseite: <http://dbis.ipd.uka.de/>



Praxisorient.
Spezialvorl.

Praxis der
Unternehmensberatung

Projektmanagement

Praxis des
Lösungsvertriebs

Praktika
WS/SS

Datenbank-
praktikum

Praktikum
Verteilte Datenhaltung

Praktikum
Data Warehousing und Mining

Vertiefungs-
vorl. SS

Datenschutz und Privatheit
in vernetzten Informationssystemen

Datenbankimplementierung
und -Tuning

Die digitale
Bibliothek

Vertiefungs-
vorl. WS

Moving Objects
Databases

Workflow-Management
Systeme

Informationsintegration
und Web-Portale

Datenbankeinsatz

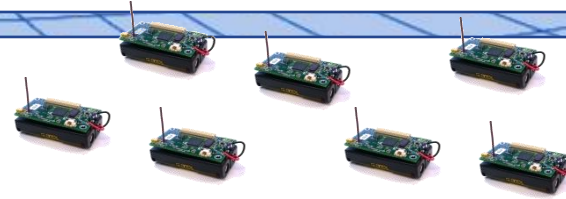
Verteilte
Datenhaltung

Data Warehousing
und Mining

Kernvorl.
SS

Kommunikation und Datenhaltung

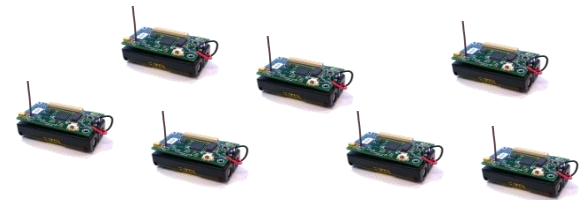
- Selektivitätsschätzung (Anfrageoptimierung)
- Datenanalyse – Data Mining/ Graph Mining
- Anfragebearbeitung in Sensornetzen
- Räumlich-/ zeitliche Anfragen an Sensornetze
- Privacy in vernetzten Informationssystemen
- Workflowmanagement-Systeme
- Information Retrieval
- ...



- **Sensornetze**
 - Überwachung von Umgebungen im industriellen und wissenschaftlichen Bereich
 - Hochgradig verteilte Systeme, geringe Ressourcen, insbes. Batterie
- **Problem: Interaktion mit dem Netz/ Datenbeschaffung**
 - Probleme der Programmierung eingebetteter und verteilter Systeme
- **Deklarative Anfragen (→ SQL) können hier Abhilfe schaffen**
 - Spezifikation des Anfrage-**Ergebnisses** („WAS“)
 - Keine Angaben zur **Beschaffung** der gefragten Daten („WIE“)
 - Beispiel: `SELECT AVG(temp) FROM Sensors`
- **Wesentliche Fragestellung: Energie-effiziente Bearbeitung von Anfragen**
 - Grundlage: Verteilung datenreduzierender Operatoren im Netz
 - Enger Zusammenhang von Kommunikation und Datenhaltung
 - ▶ Routingstrukturen beeinflussen die Effizienz/ optimale Ausführungsorte von Operatoren
 - ▶ Oft schichtenübergreifendes Design, um Netzwerkpakete zusammenzufassen
- **Schwerpunkt: komplexe Anfragen**
 - Join-Anfragen
 - Nicht-selektive Anfragen

- Beispiel: Anfrage mit Verbund (*JOIN*)
„Gibt mir die Luftfeuchtwerte aller **Sensor-Paare**, die die gleiche Temperatur gemessen haben.“

```
SELECT S1.humidity, S2.humidity  
FROM Sensors S1, Sensors S2  
WHERE S1.temp = S2.temp  
AND S1.nodeID > S2.nodeID
```



- Herausforderungen:
 - Wie können komplexe Anfragen bearbeitet werden?
(mögliche Methoden für Algebraoperatoren)
 - Wo ist der energieeffizienteste Ort für die Anfrageverarbeitung?
(Anfrageoptimierung: Finden eines optimalen Anfrageplans)
 - Welche Informationen werden benötigt, um diesen Ort zu bestimmen?
(in der Regel Statistiken, auf deren Basis die Kosten alternativer Anfragepläne geschätzt werden)

- Data Mining
 - Ausgangslage, z.B. in Unternehmen
 - ▶ In Datenbanken sammeln sich große Mengen an Daten an
 - Ziel von Data Mining
 - ▶ Aufdecken neuer und nützlicher Informationen
 - ▶ Lernen aus historischen Daten und Treffen von Vorhersagen
 - Aktueller Trend: Integration von Data Mining Techniken in Datenbank Management Systeme
- Workflow Management Systeme
 - Informationssysteme zur Steuerung von Geschäftsprozessen
 - Integrieren verteilte Prozesse zu einem Geschäftsprozess (Service Orientierte Architekturen, SOA)
 - Sammeln Historien von Workflow-Ausführungen in Datenbanken, die mit Data Mining analysiert werden können

- Forschungsprojekt mit 10 Partnern, gefördert durch das Bundesministerium für Wirtschaft und Technologie
- Ziel:
 - Bessere Auslastung von LKW und Verlagerung auf die Schiene
- Abbildung von Logistik-Prozessen in Workflow Management System
- Analyse der Prozesse mit Data Mining, insbes. Graph Mining
 - Welche Teilprozesse (Subgraphen) von Transport-Prozessen führen zu gestiegenen Kosten oder Verspätungen?
 - Wie sehen typische Transporte einer bestimmten Relation aus?
 - ▶ Wie können diese optimiert werden?
 - Welche Ursachen haben Störungen?



- Zunehmende IT-Durchdringung des Alltags
(verteilte Informationssysteme!)
 - Verteilte Datenbanken, mobile Systeme, Sensornetze, Ubiquitous Computing

⇒ Leichtes Zusammenführen von Daten, Austausch, etc.

- Problem: aktuelle Datenschutzmaßnahmen unwirksam
 - Verlust der Privatheit durch Intransparenz oder als Nebenwirkung neuer Anwendungen
- Idee: Web 2.0 für Datenschutz
 - Nutzer kommentieren Bedrohungen für Privatsphäre in Folksonomy URLs, geografische Positionen, RFID-Funklabels
 - Austausch von Erkenntnissen, vergleichbar mit Wikipedia, del.icio.us, Flickr, YouTube etc.

Ziele:

- Einfaches Erzeugen, kein techn. Fachwissen
- Transparenz, Offenlegung von Verstößen
- Gesellschaftliches Interesse, gesellschaftliche Standards