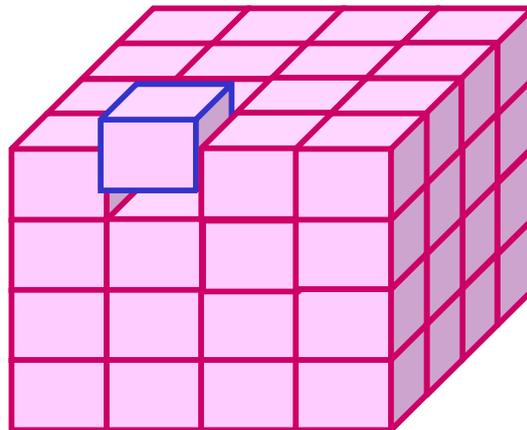


Data Warehousing

Sommersemester 2020

Prof. Dr. E. Rahm

Universität Leipzig
Institut für Informatik



http://dbs.uni-leipzig.de/stud/ss_2020/dwh



DBS-Module Master

■ Master Data Science (neu)

- 10-INF-DS01 – Skalierbare Datenbanktechnologien 1 (Pflichtmodul, 10 LP)
- 10-INF-DS101 – Skalierbare Datenbanktechnologien 2 (5 LP)
- 10-INF-DS102 – Big Data Praktikum (5 LP)
- 10-INF-DS103 – Praktikum Data Warehousing und Data Mining (5 LP)
- 10-INF-DS301 – Aktuelle Trends in Data Science (5 LP)
- 10-INF-DS02 – Masterseminar Data Science (5 LP)

■ Master Informatik

- 10-202-2215 – Moderne Datenbanktechnologien (kleines Modul)
- 10-202-2216 – *Moderne Datenbanktechnologien (großes Modul)*
- 10-202-2213 – Anwendungsbezogene Datenbankkonzepte (kleines Modul)
- 10-202-2214 – *Anwendungsbezogene Datenbankkonzepte (großes Modul)*
 - Seminarmodul
- 10-202-2011 – Masterseminar Informatik
 - Masterarbeit



DBS-Module Bachelor

- 10-201-2211 – Datenbanksysteme 1
- 10-201-2212 – Datenbanksysteme 2
- 10-201-2210 – Datenbankpraktikum
- 10-201-2224 – Realisierung von Informationssystemen
- Seminarmodul
- 10-201-2010 – Bachelorseminar Informatik
- Bachelorarbeit



Mapping: Module – Lehrveranstaltungen SS20

- Master-Module **SDBT2**, Anwendungsbez. DBK (5 LP),
Bachelor-Modul **Realisierung von IS**
 - Data Warehousing
 - NoSQL-Datenbanken
- Master-Modul **Anwendungsbez. DBK (10 LP)**
zusätzliche Belegung von
 - Big-Data-Praktikum (zulassungsbegrenzt) oder
 - Aktuelle Trends in Data Science (Ringvorlesung mit prakt. Übungen)
- Bachelorseminar / Masterseminar
 - Vortrag über laufende Bachelor/Masterarbeit
 - extern an Uni-Außenstelle in Zingst (25.-29.5.) oder an Uni



Masterstudium Data Science

- neuer Studiengang seit SS20
- Hauptinhalte
 - Skalierbares Datenmanagement („Big Data“), mind. 20 LP
 - Datenanalyse / Machine Learning, mind. 20 LP (davon mind. 10 LP Statistik)
 - Ergänzungs-/Anwendungsmodule
 - Anwendungsorientierung (Praktika, z.B. Bioinformatik/Digital Humanities)

1. Semester 2. Semester 3. Semester 4. Semester

Skalierbare Datenbanktechnologien 1		Skalierbares Datenmanagement	Skalierbares Datenmanagement	Vertiefung	Mastersem. Data Science (5 LP)
Datenanalyse		Datenanalyse		Vertiefung	Masterarbeit (25 LP)
Ergänzung	Ergänzung	Ergänzung	Anwendung/Ergänzung	Anwendung/Ergänzung	



Beispielbelegungen Data Science

Wintersemester		Sommersemester		Wintersemester	4. Semester
Skalierbare Datenbanktechnologien 1		SDBT2	Big Data Praktikum	Fortgeschr. Methoden Information Retrieval	Masterseminar (5 LP)
Künstl. neuronale Netze u. maschin. Lernen		Multivariate Statistik und Data Mining		Wissens- und Content Management	Masterarbeit (25 LP)
IT-Sicherheit	Prakt. Data Wareh./ Mining	Textdatenbanken	Aktuelle Trends in DS	Verfahren und Anwendungen in den Digital Humanities	

Skalierbare Datenbanktechnologien 1		SDBT2	Big Data Praktikum	Text Mining	Masterseminar (5 LP)
Statistisches Lernen		Multivariate Statistik und Data Mining		Künstl. neuronale Netze u. maschin. Lernen	Masterarbeit (25 LP)
Sequenzanalyse und Genomik		Grundl. komplexer Systeme	Aktuelle Trends in DS	Visualisierung	



Masterstudium Informatik

Sem	5 LP	5LP	10 LP	10 LP
1.	Kernmodul I	Kernmodul II	Vertiefungsmodul I	Ergänzungsfach
2.	Kernmodul III	Seminarmodul	Vertiefungsmodul II	Ergänzungsfach
3.	Vertiefungsmodul III		Vertiefungsmodul IV	Fakultätsinterne Schlüsselqualifikation
4.	Masterseminar	Masterarbeit		

- Schwerpunkt Big Data ausgelaufen
- individuelle Schwerpunktbildung, zB. Datenmanagement/DB



Master Informatik mit DB-Profil

Sem	5 LP	5LP	10 LP	10 LP
1.	Kernmodul I	Kernmodul II	Moderne DB-Technologien	Ergänzungsfach
2.	Vertiefungsmodul III		Anwendungsspezifische DB-Konzepte	Ergänzungsfach
3.	Moderne DB-Technologien	Seminarmodul	Vertiefungsmodul IV	Fakultätsinterne Schlüsselqualifikation
4.	Masterseminar	Masterarbeit		

DB-Modul

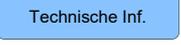
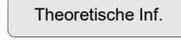
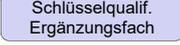
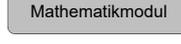
- *Moderne DBT und Anwendungsspez. DB-Konzepte* können in verschiedenen Semestern in großer und kleiner Ausführung belegt werden
 - falls disjunkte Untersetzung vorliegt



Bachelor Informatik mit DB-Profil

Sem.	5 LP	5 LP	5 LP	5 LP	5 LP	5 LP
1	Modellierung und Programmierung 1	Algorithmen u. Datenstrukturen 1	Technische Informatik 1	Logik	Analysis	
2	Modellierung und Programmierung 2	Algorithmen u. Datenstrukturen 2	TI 2 + HW-Praktikum	Java-Praktikum	Lineare Algebra	
3	Datenbanksysteme 1	Software-Technik	Kommunikationssysteme	Automaten und Sprachen	Diskrete Strukturen	Wahrscheinlichkeitstheorie
4	Datenbanksysteme 2	Softwaretechnik-Praktikum	Kernmodul 2	Berechenbarkeit	DB-Praktikum	
5	Realisierung v IS	Seminarmodul	Vertiefungsmodul		Ergänzungsfach	
6	Kernmodul 4	Bachelorseminar	Bachelorarbeit		Schlüsselqualifikation	

Legende:

 Praktische Inf.	 Technische Inf.	 Wahlmodule
 DBS-Modul	 Theoretische Inf.	 Schlüsselqualif. Ergänzungsfach
	 Mathematikmodul	



DBS-Lehrveranstaltungen

Logo	Name	Typ	SWS	Sem.
	Datenbanksysteme 1	Einführung	2+1	WS
	Datenbanksysteme 2	Einführung	2+1	SS
	Implementierung von DBS 1	Vertiefung	2	WS
	Implementierung von DBS 2	Vertiefung	2	SS
	Mehrrechner-DBS	Vertiefung	2	WS
	Data Warehousing	Vertiefung	2	SS
	Datenintegration	Vertiefung	2	WS
	Cloud Data Management	Vertiefung	2	SS
	NoSQL-DB	Vertiefung	2	SS

Name	Typ	Sem.
DB-Praktikum	Praktikum	SS
Data-Warehouse-Praktikum	Praktikum	WS
Big Data Praktikum	Praktikum	SS
Problemseminar	Seminar	WS
Bachelorseminar	Seminar	SS+WS
Masterseminar	Seminar	SS+WS



Leistungsbewertung

- Prüfungsklausur von 60 Minuten
 - voraussichtl. Juli 2020
 - überprüft konzeptionelles Wissen + Anwendungsfälle
- Klausurerfolg durch
 - Beherrschen der Vorlesungsinhalte
 - Nutzung der Online-Übungen
 - vertiefendes Literaturstudium



Online-Übungen

- LOTS (Leipzig Online Test System), <http://lots.uni-leipzig.de>
 - Kennung zum Beitritt in Übungsgruppe s. DW-Webseite

The screenshot shows the homepage of the Leipzig Online-Test-System (LOTS). The header features the LOTS logo on the left, the text 'Leipzig Online-Test-System' in the center, and the affiliation 'UNIVERSITÄT LEIPZIG Fakultät für Mathematik und Informatik Institut für Informatik Abteilung Datenbanken' on the right. A navigation bar includes links for 'Home', 'Registrierung', and 'Impressum'. The main content area is divided into several sections: a 'Login' form with fields for 'Username:' and 'Passwort:', a 'Gast Login' section with explanatory text and a 'als Gast einloggen' button, and a 'System Info' box showing '# Benutzer: 31' and a clock icon with the date '05.02.2007 12:15:21'. A 'News' section at the bottom contains the text 'Bitte loggen Sie sich ein, um personalisierte News lesen zu können.'



Lernziele der Vorlesung

- **Überblicks- und Detailkenntnisse zum Aufbau, Betrieb, Funktionsweise und Einsatz von Data Warehouses**
 - Kenntnis von Architekturalternativen
 - Techniken zur Datenintegration / Datenbereinigung
 - Techniken zur Sicherstellung einer hohen Performanz
- **Modellierung von Data Warehouses**
- **Überblick zu Analyseverfahren**
 - OLAP, SQL-Erweiterungen
 - Data Mining
- **sachkundige Beurteilung von kommerziell verfügbaren Data-Warehouse-Lösungen**



Vorläufige Vorlesungsübersicht

1. Einführung

2. Architektur von Data Warehouse-Systemen

3. Modellierung / Operationen

- MOLAP, ROLAP, hybride Ansätze
- SQL-Erweiterungen, Operatoren

4. Metadaten und ETL (Extraktion, Transformation, Laden)

- Metadaten-Integration, Schema Matching
- Datenbereinigung (Data Cleaning) / Instanz-Matching

5. Performance-Techniken

- Indexstrukturen
- Materialisierte Sichten
- Nutzung paralleler Datenbanken

6. Data Mining-Verfahren

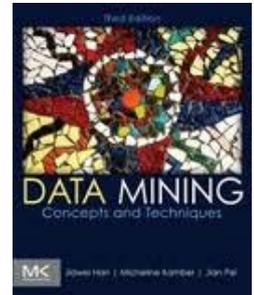
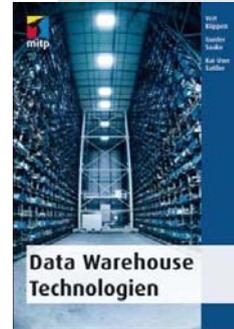
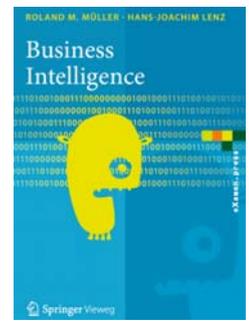
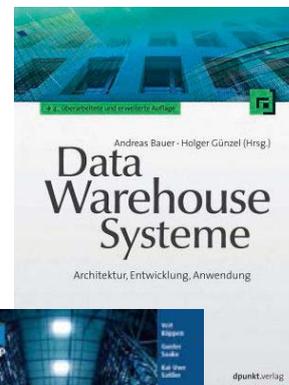
- Clusterung
- Assoziationsregeln
- Deep Learning



Literatur

■ Buchauswahl

- Bauer/Günzel (Hrsg.): *Data Warehouse Systeme*. 4. Auflage, dpunkt 2013
- Köppen/Saake/Sattler: *Data Warehouse Technologien*. 2. Auflage, Mitp 2014
- Müller/Lenz: *Business Intelligence*. Springer 2013
- Han / Kamber / Pei: *Data Mining: Concepts and Techniques*. 3rd edition, Morgan Kaufmann 2011



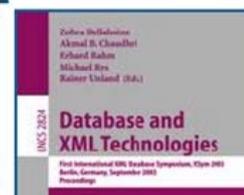
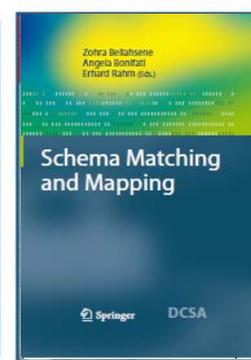
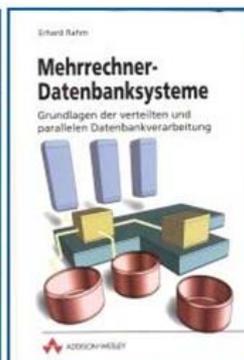
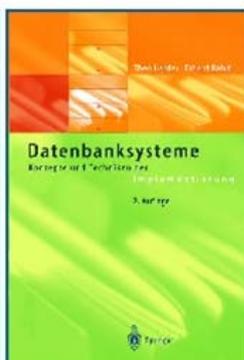
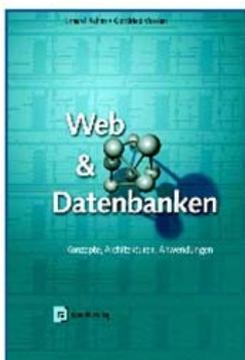
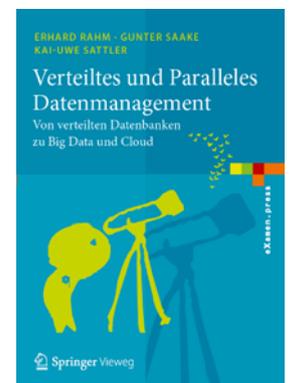
■ Eigene Forschungsarbeiten: <http://dbs.uni-leipzig.de>

- Schema Matching (COMA++)
- Data Cleaning / Instanz-Matching (FAMER)



Lehrstuhl Datenbanksysteme

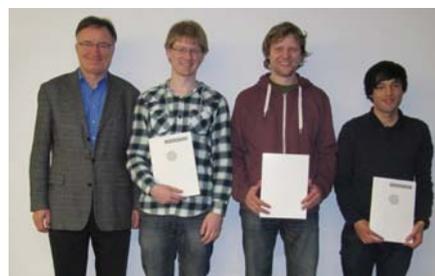
- seit 1994 am Institut für Informatik
- umfangreiches Lehrangebot
 - Vorlesungen, Praktika, Seminare
 - Online-Übungssystem LOTS
 - Eigene Lehrbücher: MRDBS, IDBS, ...



Oberseminare an Uni-Außenstelle Zingst – seit 2001



Auszeichnung von Top-Student(inn)en - seit 2008



Forschung


Abteilung Datenbanken Leipzig
 am Institut für Informatik

UNIVERSITÄT LEIPZIG
 Hilfe | Registrieren

Suchen

Startseite login

- Inhalte**
- ▶ Mitarbeiter
 - ▼ **Forschung**
 - Publikationen
 - ▶ Projekte
 - Prototypes
 - Jahresberichte
 - Kooperationen
 - ▶ Promotionen
 - Colloquia
 - ▶ Conferences
 - ▶ Studium
 - ▶ Service

Forschung

Publications (2020, 2019, 2018, 2017, 2016, 2015, 2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007, ...)

FAst Multi-source Entity Resolution system (FAMER)

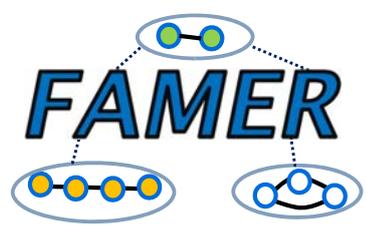
Benchmark datasets for entity resolution

Graph-based data analysis (**GRADOOP**)

Privacy-Preserving Record Linkage for Big Data

Big Data Center ScaDS Dresden/Leipzig

Entity Matching for Big Data (Dedoop), LOD Link Discovery



KI-Zentrum ScaDS.AI Dresden/Leipzig

- Center for Scalable Data AnalyticS and Artificial Intelligence scads.ai
 - eines der vom Bund geförderten nationalen KI-Zentren
 - Erweiterung des seit 2014 bestehenden Big-Data-Kompetenzzentren
 - Ausbau u.a. mit 4 neuen KI-Professuren sowie einer Graduiertenschule

■ Forschungsthemen:

