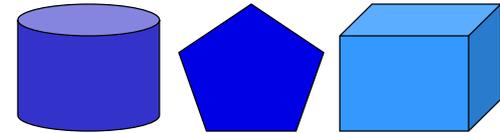


Datenintegration

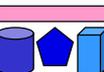
Datenintegration



Kapitel 0: Organisatorisches

Dr. Anika Groß
Sommersemester 2016

Universität Leipzig
Institut für Informatik
<http://dbs.uni-leipzig.de>



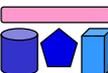
Organisatorisches

- Termin: donnerstags, 9.15 – 10.45 Uhr, HS19
- Webseite mit allen wichtigen Infos: http://dbs.uni-leipzig.de/study/ss_2016/di

The screenshot shows the website for the Department of Databases at Leipzig University. The header includes the department name and a search bar. A navigation bar shows 'Startseite > Studium > SS 2016'. On the left, a sidebar menu lists various categories like 'Mitarbeiter', 'Forschung', 'Studium', and 'Klausurtermine'. The main content area is titled 'Datenintegration' and lists the instructor 'Dr. Anika Groß'. It provides details about the lecture: it is held on Thursdays, starting on 07.04.2016 at 09:15 in HS 19, and is registered via 'Almaweb'. A 'Teilnehmerkreis' section identifies students from Informatics and other faculties. An 'Übersicht' section describes the lecture's focus on distributed data integration. A 'Skript / Gliederung' table shows the course structure.

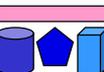
Kapitel	Thema
0	Organisation/Inhalt

- Prüfung
 - 60-minütige Klausur zum Ende des Semesters
 - Konzeptionelles Wissen + Anwendungsfälle



DBS-Module

- Master-Studium
 - 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)
 - 10-202-2011 – Masterseminar Informatik / Seminarmodul
- Bachelor-Studium
 - 10-201-2211 – Datenbanksysteme 1
 - 10-201-2212 – Datenbanksysteme 2
 - 10-201-2210 – Datenbankpraktikum
 - 10-201-2224 – Realisierung von Informationssystemen
 - 10-201-2010 – Bachelorseminar Informatik / Seminarmodul
- <http://dbs.uni-leipzig.de/de/Module/Lehrangebot>



Bachelor DBS-Profil

Legende:

Inf-Pflichtmodul

MI-Modul

DBS-Modul

Mathematikmodul

Schlüsselqualif.
Ergänzungsfach

Algorithmen und
Datenstrukturen 1
(5 LP)

Algorithmen und
Datenstrukturen 2
(5 LP)

**DBS 1
(5 LP)**

**DBS 2
(5 LP)**

**Realisierung
von IS
(5 LP)**

Kernmodul 5
(5 LP)

Modellierung und
Programmierung 1
(5 LP)

Modellierung und
Programmierung 2
(5 LP)

Softwaretechnik
(5 LP)

Softwarepraktikum
(5 LP)

**Seminarmodul
(5 LP)**

**Bachelor-
seminar
(5 LP)**

Technische
Informatik 1
(5 LP)

Technische
Informatik 2 inkl.
Hardwarepraktikum
(5 LP)

Kommunikations-
systeme
(5 LP)

Kernmodul (5 LP)

Vertiefungsmodul
(10 LP)

**Bachelor-
arbeit
(10 LP)**

Logik
(5 LP)

Java-Praktikum
(5 LP)

Automaten und
Sprachen
(5 LP)

Berechenbarkeit
(5 LP)

Analysis 1
(10 LP)

Lineare Algebra 1
(10 LP)

Diskrete Strukturen
(5 LP)

**DB-Praktikum
(5 LP)**

Ergänzungsfach
(10 LP)

Fakultäts-
übergreifende
Schlüssel-
qualifikation
(10 LP)

Wahrscheinlich-
keitstheorie
(5 LP)

Ergänzungsfach
(5LP)

1. Semester

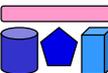
2. Semester

3. Semester

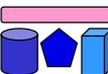
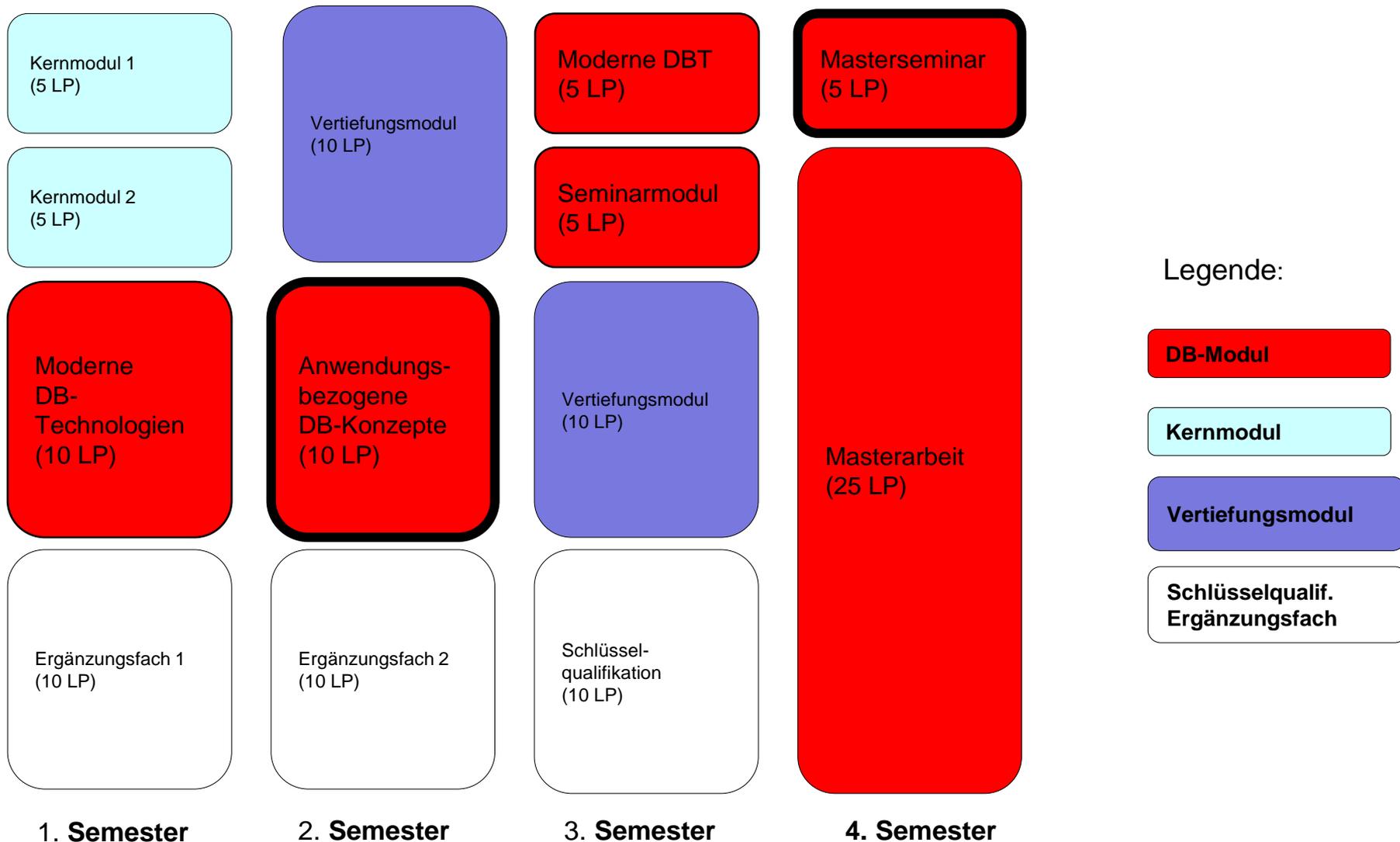
4. Semester

5. Semester

6. Semester

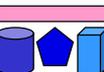


Masterstudium DBS-Profil



Mapping Module – Lehrveranstaltungen SS 2016

- Master-Module “Anwendungsbezogene Datenbankkonzepte” (zwei bzw. drei Veranstaltungen für kleines bzw. großes Modul):
 - Data Warehousing
 - Datenintegration
 - Big Data Praktikum
 - ggf. DBS 2
- Bachelor-Modul “Realisierung von Informationssystemen”
 - Data Warehousing
 - Datenintegration
- Relationales Datenbankpraktikum
- Bachelorseminar / Masterseminar
 - Vortrag über laufende Bachelor/Masterarbeit
 - Termin: tba



Online Übungen

- **LOTS** (Leipzig Online Test System), <http://lots.uni-leipzig.de>



LOTS Leipzig Online-Test-System

Home Training SQL-Training XQuery Registrierung Impressum Hilfe

Login

Username:

Passwort:

[Passwort vergessen?](#)

Willkommen bei LOTS

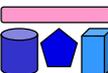
LOTS ist ein in der Abteilung Datenbanken der Universität Leipzig entwickeltes E-Learning System. LOTS wird primär im Übungsbetrieb des Informatikstudien-ganges eingesetzt, steht aber auch allen weiteren Interessierten zur Verfügung. Einige sächsische Schulen nutzen das SQL-Training-Modul von LOTS bereits im Informatikunterricht.

Gleich ausprobieren als Gast...

- [Übungsaufgaben](#)
- [SQL Übungen](#)
- [XQuery Übungen](#)

Der Funktionsumfang von LOTS steht Ihnen jetzt schon in Teilen zur Verfügung. Sie können sofort einige [Übungsaufgaben](#) lösen sowie praktische [SQL-](#) oder [XQuery-Übungen](#) durchführen.

Falls Sie Hilfe benötigen, können Sie die [Online-Hilfe](#) verwenden oder uns eine [Frage per E-Mail](#) stellen. Diese E-Mail-Adresse finden Sie übrigens am Ende jeder Seite.



Forschung

- <http://dbs.uni-leipzig.de/de/research>

The screenshot shows the website for the Department of Databases at Leipzig University. The header features the department name in orange and the institute name in grey. A search bar is present. A dark blue navigation bar contains a German flag, a UK flag, and the text 'Startseite'. The main content area is divided into a left sidebar and a main right section. The sidebar has three sections: 'Inhalte' with a tree view of navigation items, 'Mitarbeiter Menu' with an 'Interneta' item, and 'Neue Publikationen' with a list item. The main section has a 'Forschung' header, a set of tabs (Anzeigen, translation, Bearbeiten, Gliederung, Versionen, diff), and a list of research topics with external links.

Abteilung Datenbanken Leipzig
am Institut für Informatik

Suchen

Startseite

Inhalte

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

Mitarbeiter Menu

- ▶ Interneta

Neue Publikationen

- Semi-Automatic Identification of Counterfeit Offers in Online Shopping

Forschung

Anzeigen translation Bearbeiten Gliederung Versionen diff

Publications (2016, 2015, 2014, 2013, 2012, 2011, 2010, 2009, 2008, 2007, 2006, 2005, ...)

Big Data Center ScaDS Dresden/Leipzig

Graph-based data analysis (GRADOOP, BIIIG)

Entity Matching for Big Data (Dedoop)

LOD Link Discovery

Evolution of Semantic Annotations (ELISA)

Schema and Ontology Matching (COMA++, GOMMA), Ontology Merging (ATOM)

Semantic ontolgy matching (STROMA mapping enrichment, SemRep repository)

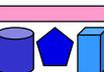
Evolution of ontologies and mappings (Schema Evolution bibliography)

Web Data Integration (WDI) Lab

Object Matching / Entity Resolution

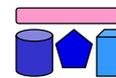
Bibliometric Analysis

Current prototypes and implementations



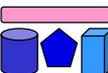
Forschungsgebiete / Projekte

- Big Data Zentrum ScaDS Dresden/Leipzig seit Okt'2014
 - Themen u.a. Cloud Data Management
 - Privacy preserving record linkage
 - Entity Matching for Big Data (Dedoop)
- Graph-based data analysis (BIIG, Gradoop)
- Datenintegration
 - Schema and Ontology Matching (COMA++, GOMMA, STROMA, SemRep)
 - Ontology Merging (ATOM)
 - Object Matching / Entity Resolution
 - Evolution of ontologies and mappings
 - DFG-Projekt: “ELISA - Evolution of Semantic Annotations”
 - DFG Projekt: “Learning-based Link Discovery”
 - BMPF-Projekt: “Leipzig Health Atlas“ u.a. Kooperation mit IMISE, IZBI und dem LIFE-Projekt...



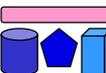
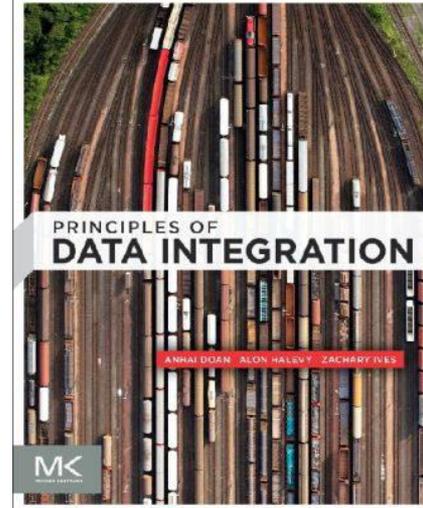
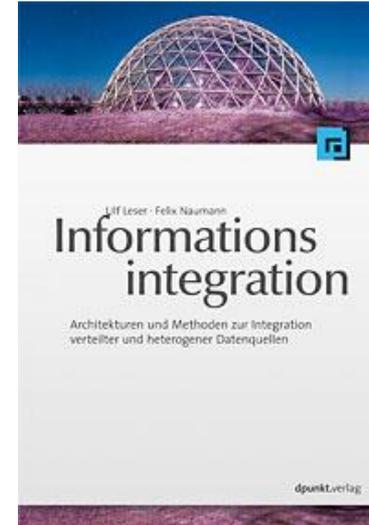
Abschlussarbeiten / SHK

- Wir suchen ständig interessierte Studenten für
 - SHK/WHK Tätigkeiten
 - Bachelor- oder Masterarbeiten
- zur Mitarbeit an aktuellen Forschungsthemen des Lehrstuhls, sowie gemeinsamen Themen mit dem Big Data Zentrum ScaDS Dresden/Leipzig
 - Auswahl mögl. Themen: <http://dbs.uni-leipzig.de/de/theses>
- Kooperationen mit (regionalen) Unternehmen, z.B.
- Anfragen per Formular oder persönlich
 - http://dbs.uni-leipzig.de/de/service/anfrage_abschlussarbeit



Literatur

- Ulf Leser, Felix Naumann:
„Informationsintegration“,
dpunkt.verlag, Oktober 2006
- AnHai Doan, Alon Halevy, Zachary Ives:
“Principles of Data Integration”,
Morgan Kaufmann, July 2012
- Hinweise auf wissenschaftliche Arbeiten in den einzelnen Kapiteln
- Vorlesung ausgearbeitet von Prof. Dr. Andreas Thor (HfT Leipzig)
- Vorlesungsskript basiert auf Skripten zur Vorlesung
“Informationsintegration” von
 - Prof. Dr. Ulf Leser
 - Professur “Knowledge Management in Bioinformatics”
 - Humboldt-Universität zu Berlin
 - Prof. Dr. Felix Naumann
 - Fachgebiet “Informationssysteme”
 - Hasso-Plattner-Institut, Potsdam



Inhaltsverzeichnis

1. Einführung
 - Begriffsdefinition
 - Beispiele und Anwendungsgebiete
 - Integration am Beispiel
2. Verteilung, Autonomie und Heterogenität
 - Verteilung
 - Autonomie
 - Heterogenität
3. Eigenschaften von Integrationssystemen
 - Bzgl. Verteilung, Autonomie und Heterogenität
 - Weitere Kriterien
 - Materialisiert vs. virtuell
4. Integrationsarchitekturen
 - Data Warehouse
 - Föderierte Datenbanken
 - Mediator-Wrapper-Ansatz
 - Peer-Data-Management-Systeme
 - Architekturen vs. Integrationsaufgaben
5. Anfrageverarbeitung
 - SchemaSQL
 - Global-as-View (GaV), Local-as-View (LaV)
 - Quellen mit beschränkten Anfragemöglichkeiten
6. Schemamanagement
 - Schemaintegration
 - Schema Matching
 - Schema Mapping
7. Datenfusion Datenqualität
 - Duplikaterkennung
 - Konfliktbehandlung

