

Modulbezeichnung	Datenintegration								
Modulverantwortliche(r)	Prof. Dr. S. Maneth								
Modulart	Pflicht/Wahl <input checked="" type="checkbox"/> Wahlpflicht <input type="checkbox"/>								
Spezialisierungsbereich									
Dauer des Moduls	1 Semester								
Kreditpunkte	6 CP								
Arbeitsaufwand	<table> <tr> <td>Berechnung des Workloads</td> <td></td> </tr> <tr> <td>Präsenz</td> <td>56 h</td> </tr> <tr> <td>Übungsbetrieb/Prüfungsvorbereitung</td> <td>124 h</td> </tr> <tr> <td>Summe</td> <td>180 h</td> </tr> </table>	Berechnung des Workloads		Präsenz	56 h	Übungsbetrieb/Prüfungsvorbereitung	124 h	Summe	180 h
Berechnung des Workloads									
Präsenz	56 h								
Übungsbetrieb/Prüfungsvorbereitung	124 h								
Summe	180 h								
Turnus des Moduls	i.d.R. alle 2 Semester								
Voraussetzung für die Teilnahme	Keine <input type="checkbox"/> Folgende Inhaltliche Voraussetzungen: Datenbankgrundlagen								
Lehr- und Lernformen	Seminar <input type="checkbox"/> Vorlesung <input checked="" type="checkbox"/> Tutorium <input checked="" type="checkbox"/> Praktikum <input type="checkbox"/> Projekt <input type="checkbox"/>								
Lernziele	<ul style="list-style-type: none"> • In der Terminologie des Fachgebietes Datenbanken kommunizieren können, Teilprobleme und Methodiken des Gebietes unterscheiden und klassifizieren können. • Methoden der Schemaabbildungen zwischen relationalen Datenquellen verstehen und anwenden können. • Möglichkeiten zur Informationsextraktion (z.B. aus Web Seiten) verstehen und bewerten können. • Mechanismen zum Umgang mit fehlerhaften Daten verstehen und anwenden können. • Datenprovenienz innerhalb von relationalen Datenbanken modellieren können. 								
Lerninhalte	<ul style="list-style-type: none"> • Datenbankkonzepte: Integritätsbedingungen, konjunktive Anfragen (kA), Datalog Programme • Anfrageäquivalenz und Enthaltensein: kAs, Vereinigungen von kAs, kAs mit Negation, Multimengen und Aggregation • Anfrageauswertung auf Views: Bucket und MiniCon Algorithmus, Inversregel Algorithmus, Updates auf Views • Ontologie-Basierter Datenzugriff: Ontologiesprachen, DL-LITE, Rewriting versus Unfolding • Abbildungen von Schemata: Global-as-View (GAV), Local-as-View (LAV), Global-and-Local-as-View (GLAV), Tupelgenerierende Abhängigkeiten • Anfrageauswertung mit/ohne Bedingungen: GAV, LAV, und GLAV settings, Auswertung vs. Rewriting • Wrappers: manuelle und lernbasierte Konstruktion, Web-Wrappers, Systeme: Stalker, Poly und Lito • Umgang mit Unsicherheit und fehlerhaften Daten: zuverlässige Antworten, probabilistische Datenbanken, probabilistische Schemaabbildungen • Datenprovenienz: Annotationen und Datenabhaengigkeitsgraphen, der Provenienzhilbring 								
Prüfungsformen	i.d.R. Übungsaufgaben und Fachgespräch oder mündliche Prüfung								
Literatur	AnHai Doan, Alon Halevy, Zachary Ives: Principles of Data Integration, Morgan Kaufmann, 2012 Dan Such, Dan Olteanu, Christopher Re, Christoph Koch: Probabilistic Databases, Morgan & Claypool Publishers, 2011								