

Modul: Text Analytics

Inhalt:

Das Modul behandelt Methoden zur computergestützten Analyse von Texten. Die Themen reichen vom Information Retrieval (Suchmaschinen, Anfragesprachen, Indexierung) über die Verfahren der Computerlinguistik (Kollokationsanalyse, Sprachmodelle, Wortart-Tagging, Disambiguierung) bis zu fortgeschrittenen Methoden im Text Mining (Dokumentklassifikation und -clustering, Informationsextraktion, Plagiaterkennung). Dabei werden sowohl algorithmische Grundlagen als auch Anwendungen, etwas im Bereich des Wissensmanagements und der Intra- und Internetsuche, behandelt.

Die Vorlesung wird durch ein Praktikum begleitet. Dieses Praktikum vertieft die im Halbkurs gelernten Methoden durch praktische Umsetzung. In Gruppen wird ein komplexeres Problem des Text Mining, aufbauend auf existierenden Frameworks, gelöst. Voraussetzung für die Teilnahme sind gute Kenntnisse in Java.

Qualifikationsziele:

Studierende erlangen die Fähigkeit, Analyseverfahren für textuelle Daten zu entwerfen und einzuschätzen. Sie lernen die grundlegenden Verfahren zur Suche in Texten, zur computerlinguistischen Aufbereitung von Dokumenten und zum Management und zur Analyse großer Dokumentsammlungen kennen.

Voraussetzungen für die Teilnahme am Modul:

Vordiplom, Kenntnisse in Algorithmen und Datenstrukturen, gute Kenntnisse in der Programmierung mit Java

Lehrveranstaltungen	SWS	SP und Beschreibung der Arbeitsleistung auf deren Grundlage die SP vergeben werden	
VL + Praktikum	3 + 3	8 SP : Prüfung und erfolgreiche Teilnahme am Praktikum	
Prüfung (Prüfungsform, Umfang/Dauer, SP)	Mündliche Prüfung		
Häufigkeit des Angebotes	Etwa in jedem zweiten Sommersemester		
Dauer des Moduls	1 Semester		