

Master of Science Medizininformatik

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-202-2307	Wahlpflicht

Modultitel	NLP, IR, and the Web Archive
Modultitel (englisch)	NLP, IR, and the Web Archive
Empfohlen für:	1./3. Semester
Verantwortlich	Automatische Sprachverarbeitung / Text Mining und Retrieval
Dauer	1 Semester
Modulturnus	jedes Wintersemester
Lehrformen	<ul style="list-style-type: none"> • Seminar "NLP, IR, and the Web Archive" (2 SWS) = 30 h Präsenzzeit und 70 h Selbststudium = 100 h • Übung "NLP, IR, and the Web Archive" (1 SWS) = 15 h Präsenzzeit und 35 h Selbststudium = 50 h
Arbeitsaufwand	5 LP = 150 Arbeitsstunden (Workload)
Verwendbarkeit	<ul style="list-style-type: none"> • M.Sc. Digital Humanities • Seminarmodul im M.Sc. Informatik • M.Sc. Medizininformatik
Ziele	<p>Das Internet und das Web wachsen in exponentiellem Tempo und bilden die Datengrundlage für unzählige intelligente Anwendungen des Natural Language Processings (NLP) und des Information Retrievals (IR). Datenquellen dieser Größenordnung erfordern skalierbare verteilte Architekturen für die kognitive Analyse, die die Geschäftsentscheidungen vieler Branchen prägen. Darüber hinaus wurde das Deep Learning in den Mainstream dieser Forschungsfelder vorangetrieben und ist nun dank Tools wie TensorFlow, PyTorch und Docker für Forscher und Unternehmen gleichermaßen zugänglich. Die Webis-Forschungsgruppe betreibt einen großen, modernen Hochleistungsrechner (https://webis.de/facilities.html). Dort wird ein mehr als zehn Petabyte großer Webausschnitt für Forschungszwecke vorgehalten, der mehr als die letzten zehn Jahre des Webs widerspiegelt – ein einmaliger Forschungsgegenstand.</p> <p>Nach der aktiven Teilnahme am Modul sind die Studierenden je nach gewähltem Seminarthema in der Lage, wissenschaftliche Arbeiten zu reproduzieren, skalierbare Algorithmen des NLP und des IR auf großen Datenmengen anzuwenden, Methoden des Maschinellen Lernens und insbesondere des Deep Learnings anzuwenden, ein algorithmisches Forschungsproblem zu erfassen und in Wort und Schrift zu erklären.</p>
Inhalt	<p>Die Studenten erhalten eine anwendungsorientierte Ausbildung in NLP, IR, Big Data und Deep Learning Frameworks, lösen Aufgaben, reproduzieren aktuelle wissenschaftliche Arbeiten und untersuchen eigenverantwortlich interessante Forschungsfragen. Dank der Größe des vorhandenen Clusters und den Kompetenzen der Webis-Gruppe in den Bereichen NLP, IR, und Big Data, bietet dieses Seminar ein außergewöhnliches Ausbildungsniveau.</p> <ul style="list-style-type: none"> - Lehrsprache: englisch oder deutsch - Prüfungssprache: englisch oder deutsch

Die Festlegung hierzu erfolgt vor der Moduleinschreibung auf elektronischem Weg (Vorlesungsverzeichnis) oder zu Beginn der Veranstaltung durch den Dozenten/die Dozentin.

Teilnahmevoraussetzungen Vorkenntnisse in Natural Language Processing, Information Retrieval oder Web Data Mining.

Literaturangabe Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

Vergabe von Leistungspunkten Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

Prüfungsleistungen und -vorleistungen

Modulprüfung: Präsentation (30 Min.) mit schriftlicher Ausarbeitung (8 Wochen), mit Wichtung: 1	
	Seminar "NLP, IR, and the Web Archive" (2SWS)
	Übung "NLP, IR, and the Web Archive" (1SWS)