

Studiengang Information Science (Bachelor of Science) [PO 2019]

Wahlpflichtkatalog

WPs

1	Modulname Grundlagen der Sprachverarbeitung (Introduction to Natural Language Processing)
1.1	Modulkürzel 141090
1.2	Art Wahlpflicht / WP-Modul
1.3	Lehrveranstaltung(en) Grundlagen der Sprachverarbeitung
1.4	Semester 4
1.5	Modulverantwortliche(r) Prof. Dr. Margot Mieskes
1.6	Weitere Lehrende
1.7	Studiengangsniveau Bachelor
1.8	Lehrsprache Deutsch
2	Inhalt Das Seminar befasst sich mit der automatischen Analyse von Textdaten. Wir werden Texte mit linguistischen und statistischen Methoden analysieren. Schon nach kurzer Zeit können sie z.B. die durchschnittliche Satzlänge in einem Text bestimmen oder feststellen, welche Wörter häufig miteinander im selben Satz auftreten. Dabei bekommen die Studierenden eine systematische und gründliche Einführung in die Programmierung mit Python und pythonbasierten Modulen zur computerlinguistischen Textanalyse.

3	<p>Ziele</p> <p><u>Kenntnisse</u></p> <ul style="list-style-type: none"> • Die Studierenden kennen computerlinguistische Algorithmen und Methoden und können sie anwenden. • Die Studierenden kennen die entsprechenden linguistischen Hintergründe. • Die Studierenden kennen die wichtigsten Datensätze zur Entwicklung von Algorithmen. • Die Studierenden kennen Metriken, um den Erfolg der implementierten Methoden zu evaluieren. <p><u>Fertigkeiten</u></p> <ul style="list-style-type: none"> • Die Studierenden können Texte mit Python, NLTK und Textblob analysieren, indem sie Programme schreiben. • Die Studierenden können komplexere Programme und Funktionen implementieren, die auf Python und NLTK basieren. <p><u>Kompetenzen</u></p> <ul style="list-style-type: none"> • Die Studierenden erwerben die Kompetenz, Wissen aus unstrukturierten, natürlichsprachlichen Daten zu extrahieren. • Sie sind in der Lage, Projekte mit Textanalysemethoden durchzuführen. • Sie sind in der Lage natürlichsprachliche Daten mit linguistischen Informationen automatisch oder manuell anzureichern. • Sie sind in der Lage, diese Anreicherungen zu evaluieren.
4	<p>Lehr- und Lernformen</p> <p>Seminar und Übung</p>
5	<p>Arbeitsaufwand und Credit Points</p> <p>5 CP; 150 Stunden: 60 Stunden Präsenzzeit, 90 Stunden Selbststudium und Prüfungsvorbereitung</p>
6	<p>Prüfungsform, Prüfungsdauer und Prüfungsvoraussetzung</p> <p>Die Prüfungsleistung im Seminar besteht aus:</p> <ul style="list-style-type: none"> • Lösung der Übungsaufgaben im Moodle-Kurs • Kleines NLP-Programmierprojekt mit dokumentiertem Code
7	<p>Notwendige Kenntnisse</p>
8	<p>Empfohlene Kenntnisse</p> <p>Semantik I und II</p>
9	<p>Dauer, zeitliche Gliederung und Häufigkeit des Angebots</p> <p>1 Semester; jährlich im SS, 4 SWS</p>
10	<p>Verwendbarkeit des Moduls</p> <p>Bachelorstudiengang Information Science</p>
11	<p>Literatur</p> <p>Moodle-Kurs: https://lernen.h-da.de/course/view.php?id=10640</p>