

Module title		Abbreviation
Natural Language Processing		10-I-NLP-222-m01
Module coordinator		Module offered by
holder of the Chair of Computer Science XII		Institute of Computer Science
ECTS	Method of grading	Only after succ. compl. of module(s)
5	numerical grade	--
Duration	Module level	Other prerequisites
1 semester	undergraduate	--
Contents		
Introduction to Text Mining and Natural Language Processing; Traditional computational representations of text data (bag-of-words) and text preprocessing (sentence splitting, tokenization, morphological normalization, stemming); Corpus linguistics and lexical association measures (ngram frequencies, co-occurrences, collocations and terminology extraction); Syntactic analysis: Part-of-Speech tagging and chunking (with Hidden Markov Models and Conditional Random Fields), parsing (Probabilistic Context Free Grammars and parsers); Distributional semantics and latent text representations: distributional hypothesis, Latent Semantic Analysis (LSA), word embeddings; Light introduction to (modern) deep learning-based NLP: embeddings, convolutional and recurrent networks, Transformers. NLP Applications: text classification tasks (e.g., document classification, sentiment analysis) vs. token classification tasks (e.g., information extraction - named entity recognition) vs. text generation tasks (e.g. machine translation and text summarization).		
Intended learning outcomes		
Students will obtain broad theoretical and practical knowledge of the typical methods and algorithms in the field of text mining and natural language processing. They will be able to solve practical problems with the obtained knowledge: analyze the text data for the task at hand, choose the appropriate representation for their texts as well as the appropriate (machine learning for NLP) model to solve the task. They will have gained rich practical experience implementing solutions for a wide range of common NLP tasks and applications.		
Courses (type, number of weekly contact hours, language — if other than German)		
V (2) + Ü (2)		
Method of assessment (type, scope, language — if other than German, examination offered — if not every semester, information on whether module is creditable for bonus)		
written examination (approx. 60 to 120 minutes). If announced by the lecturer at the beginning of the course, the written examination may be replaced by an oral examination of one candidate each (approx. 20 minutes) or an oral examination in groups of 2 candidates (approx. 15 minutes per candidate). Language of assessment: German and/or English creditable for bonus		
Allocation of places		
--		
Additional information		
--		
Workload		
150 h		
Teaching cycle		
--		
Referred to in LPO I (examination regulations for teaching-degree programmes)		
--		
Module appears in		
Bachelor' degree (1 major) Mathematical Data Science (2022)		

Bachelor' degree (1 major) Artificial Intelligence and Data Science (2022)
Bachelor' degree (1 major) Artificial Intelligence and Data Science (2023)