Georg-August-Universität Göttingen **Module B.Inf.1250: Deep Learning for Natural Language Processing**	9 C 4 WLH

Learning outcome, core skills: The course seeks to enable students to solve a wide range of applied problems in Natural Language Processing. After successfully completing the course, the participants should be able to: • Explain state-of-the-art methods to tackle NLP sub-problems, such as text representation, information extraction, text mining, language modeling, and similarity detection • Determine the conceptual requirements of specific NLP tasks • Assess the strengths and limitations of state-of-the-art NLP approaches • Devise solutions for complex, interdisciplinary NLP problems by implementing and adapting suitable algorithms and data structures • Evaluate NLP methods and systems quantitatively and qualitatively	**Workload:** Attendance time: 56 h Self-study time: 214 h

Course: B.Inf.1250.Lec **Lecture Deep Learning for Natural Language Processing** (Lecture) *Contents*: The lecture will cover the following topics: Foundational NLP • Text representation (words, sentences, paragraphs, documents) • Text processing, stopwords, regular expressions, tokenization, stemming, lemmatization • Bag-of-Words, weighting schemes (e.g., tf-idf), information retrieval • Minimum edit distance • Language models, N-grams, perplexity, smoothing • Word sense, lexical databases, distance measures • Word embeddings (sparse and dense vector representation) • Vector representation • Evaluation and metrics Deep Learning • Neural Networks • Feed-Forward Networks • Activation functions, cost function, gradient descent, regularization • Backpropagation • Neural Language Models, RNN (and improvements) • Vanishing Gradients • Seq2Seq • Attention • Transformers, self-attention • Pre-training and post-training (e.g., supervised fine-tuning, reinforcement learning with human feedback, direct preference optimization) • Large language models and related topics (e.g., adaptation, prompting, reasoning)	2 WLH

Applications • Lexical databases, lexical semantics • Word sense disambiguation, semantic similarity • Part-of-speech tagging, parsing • Word similarity, word dissimilarity, distance measures • Text classification • Sentiment analysis/evaluation • Named entity recognition, information extraction, relation extraction • Questioning and answering, chatbots • Text generation and summarization • Machine translation Please visit www.gipplab.org/teaching for details on this course.	
Course: B.Inf.1250.Lab **Practical Course Deep Learning for Natural Language Processing** (Practical course) *Contents*: In the practical course, students work on applied research projects (teamwork is possible) that address complex NLP downstream tasks and subtasks, such as: • Word sense disambiguation and similarity • Document and sentence classification • Named entity recognition • Question and answering systems • Text generation and summarization • Paraphrase generation and detection • Sentiment analysis • Part-of-speech tagging • Machine translation Applications that participants can address in their projects include but are not limited to: • Plagiarism and paraphrase detection • Social media analysis • Fake news identification and classification • Detection of political opinions • Identification of opinion polarity • Online harassment and bias identification systems • Sentiment analysis in social media • Question and answering systems • Semantic evaluation Invited speakers may present selected advanced topics in NLP during the lecture and/or tutorial sessions. Using the programming language Python is mandatory. Please visit www.gipplab.org/teaching for details on this course.	2 WLH
Examination: Written test (90 min.) and Project submission; in case of 15 or fewer participants: oral exam (approx. 20 min.) and project presentation (approx. 20 min.)	9 C

B.Inf.1250.Mp: Lecture Deep Learning for Natural Language Processing

Examination requirements:

Examination for the lecture (40% of the final grade)

- Knowledge of major NLP tasks, sub-tasks, and applications
- Ability to explain state-of-the-art methods to address NLP tasks, such as text representation, information extraction, text mining, language modeling, and similarity detection
- Ability to analyze the conceptual requirements of specific NLP tasks
- Ability to compare the suitability of state-of-the-art NLP approaches for specific tasks
- Ability to evaluate NLP methods and systems quantitatively and qualitatively

Examination for the practical course (60% of the final grade)

- Ability to analyze the conceptual requirements of specific NLP problems
- Ability to determine the conceptual requirements of specific IR and NLP problems
- Ability to compare the suitability of algorithms and data structures for specific NLP problems
- Ability to devise solutions for complex, interdisciplinary NLP tasks by implementing and adapting suitable algorithms and data structures.
- Ability to evaluate NLP methods and systems quantitatively and qualitatively

The examination for the lecture and the practical course must be completed successfully in the same semester. A repeated examination always encompasses both components.

Admission requirements:

none

Recommended previous knowledge:

This is an advanced course primarily intended for master’s students. Advanced bachelor's students can participate in the course if they possess the following recommended previous knowledge:

Advanced knowledge of Python is required to complete the course. Experience with numpy, sckit-learn, pandas, and other libraries in the SciPy ecosystem is beneficial. At the University of Göttingen's computer science department, the courses B.Inf.1101: Grundlagen der Informatik und Programmierung and B.Inf.1842: Programmieren für Data Scientists: Python provide a good foundation for this course.

Knowledge of neural networks is strongly recommended to participate in this course. Participants should be familiar with basic neural network architectures, hidden layers, activation functions, derivatives, classification, training and test strategies, precision, recall, backpropagation, gradients, and other foundational topics in machine

	learning and artificial neural networks. We strongly recommend completing at least two of the following courses prior or concurrently to this course to obtain the knowledge required for this course: • B.Inf.1236 Machine Learning or equivalent • B.Inf.1237: Deep Learning for Computer Vision or equivalent • B.Inf.1248: Language as Data or equivalent
Language: English	**Person responsible for module:** Prof. Dr. Béla Gipp PD Dr. Terry Lima Ruas
Course frequency: each summer semester	**Duration:** 1 semester[s]
Number of repeat examinations permitted: twice	**Recommended semester:** Bachelor: 4 - 6; Master: 1 - 3
Maximum number of students: 50	

Additional notes and regulations:

The course provides a good foundation for a bachelor's or master's thesis in our group. Visit https://gipplab.org/students-corner/graduation-projects for our current theses proposals.

The module B.Inf.1250 may not be taken if the module M.Inf.2202 has already been completed.