Course: Natural language processing

2025/26, Spring semester

Lecturer: Prof. Dr. Marko Robnik-Šikonja
Assistants: Assoc. Prof. Dr. Slavko Žitnik, Assist. Aleš Žagar

Course objectives: Learn about the theory and main practical approaches in natural language processing and understanding. Use modern large language models and statistical techniques for language processing.

Student's obligations:

- five web quizzes
- project
- written exam

Grading

The practical work involves using natural language processing tools and large language models. It is graded through a project that must be completed on time. The projects are done in groups of three students. The project topics are set at the start of the semester. The results shall be described in a paper and publicly presented to the class.
The exam is a written test. The preconditions for the written exam are successfully passing the quizzes and the project. During the written exam, students may use one double-sided A4 sheet of paper. In case of any doubts about the score of a project, quizzes, or written exams, the oral exam is obligatory. The final grade is the sum of the project scores and the written exam score. The contribution of each part to the final grade and the conditions to pass the exam are listed below:

Work	**% of total**	**subject to**
Five quizzes	0%	≥ 50% altogether
The project	50%	≥ 50%
Written exam	50%	≥ 50%

The grades are valid in the current and next year. The students who did not pass the exam in this time have to redo the quizzes and the project.

Syllabus

The syllabus is based on a selection of modern natural learning techniques and their practical use. The lectures introduce the main tasks and techniques, explain their operation, and provide their theoretical background. The knowledge gained during practical sessions and seminars is applied to practical tasks using open-source tools. Students investigate and solve assignments based on real-world research and industrial problems, primarily in English and Slovene.

1. Introduction to natural language processing: motivation, language understanding, ambiguity, traditional, statistical, and neural approaches.
2. Text preprocessing: normalization, string similarity, lemmatization, regular expressions, and grammars.
3. Text similarity: measures, clustering approaches, language networks, and graphs.
4. Language resources: corpora, datasets, dictionaries, thesauri, semantic databases, WordNet.
5. Text representation: sparse and dense embeddings, cosine distances, language models.
6. Deep neural networks for text: recurrent neural networks, CNNs for text, transformers.
7. Neural embeddings: word2vec, fastText, BERT, cross-lingual embeddings, sentence, and document embeddings.
8. Large pretrained language models: BERT, GPT, and T5 families, multimodal models.
9. Improving LLMs: instruction tuning, preference alignment, and test-time computation.
10. Inference with LLMs: prompt engineering, retrieval augmented generation, function calling, and agentic LLMs.
11. Linguistic tasks: part-of-speech tagging, dependency parsing, named entity recognition, word sense disambiguation.
12. Affective computing: sentiment, emotions, etc.
13. Popular tasks: text summarization, question answering, machine translation, and paraphrasing.
14. Reasoning tasks: natural language inference, commonsense reasoning, and logical reasoning.
15. Advanced topics: long context in LLMs, alternatives to transformers.

Literature (all freely available)**:**

1. Jurafsky, David, and Martin, James H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition, 3rd edition draft, 2026.*
   This is the primary course literature; available on the authors' webpages
2. Simon J.D. Prince: *Understanding Deep Learning. MIT Press, 2023* *https://udlbook.github.io/udlbook/* This recent book intuitively explains many important concepts and techniques in deep learning.