

## Natural Language Processing Syllabus

### General Information

Natural Language Processing (NLP) is a rapidly developing field that aims to allow machines to break down and interpret human language. It combines the power of linguistics and computer science and takes advantage of machine learning techniques to learn the rules and structure of language and build intelligent systems that can understand, analyze and generate natural language text.

This course is intended as a theoretical and methodological introduction to Natural Language Processing. It will focus on the basic discussion and analysis of various NLP tasks, algorithms for effectively solving these problems, and methods for evaluating their performance. There will be a focus on statistical and neural-network learning algorithms that train on (annotated) text corpora to automatically acquire the knowledge needed to perform the task.

### Topics

- Overview of natural language processing: NLP tasks in syntax and semantics. NLP applications such as information extraction, question answering and machine translation. The challenge of variety and ambiguity of language, and the role of machine (deep) learning in NLP.
- Common NLP tasks, including language modeling, Part-of-Speech Tagging and sequence labeling, syntactic and semantic parsing, semantic analysis, such as lexical semantic representation learning, word sense disambiguation, text classification, sentiment analysis.
- High-level NLP applications
  - information extraction including entity, relation and event extraction, entity linking and coreference resolution.
  - summarization and natural language generation
  - machine translation
  - question answering and natural language understanding.
- Deep learning for NLP: basic neural networks for NLP

### Prerequisites

Prerequisite courses include data structures, algorithms, machine learning and python programming. Students should have basic knowledge of formal language. Basic skills of deep learning related programming, e.g., *Pytorch*, *Tensorflow*, are highly preferred as the assignments and final project will be based on deep neural networks.

### Learning Objectives

A student who successfully completes this class should:

- be familiar with a breadth of concepts and tasks in natural language processing.
- gain a foundational understanding about the methods and evaluation metrics for various natural language processing tasks.
- learn how to evaluate the strengths and weaknesses of various NLP technologies and frameworks.
- gain practical experience in the NLP toolkits available.
- gain basic skills for conducting NLP research, including reading and analyzing research papers, analyzing results, and how to improve the approaches.

## **Grading**

The course will involve multiple paper presentations, 3-4 homework assignments, and a final research project:

- **Presentation:** Each student will present 2-4 papers for a specific topic in NLP.
- **Homework Assignments:** Students need to finish 3-4 homework assignments individually, while each assignment will be related to the particular NLP tasks and algorithms discussed in lectures.
- **Final Project:** Students will form groups of at most three to work on a final project. The final project will be research-oriented and related to natural language processing tasks. Students are expected to generate a paper draft based on the research project.