

CMPE 561 NATURAL LANGUAGE PROCESSING SYLLABUS

Instructor: Tunga Güngör (E-mail: gungort@boun.edu.tr, Room: ETA 34)

Course Description:

There has been a striking growth in text data such as web pages, news articles, e-mail messages, social media data, and scientific publications in the recent years. Developing tools for processing and utilizing this huge amount of textual information is getting increasingly important. This introductory course will cover techniques for processing and making sense of text data written in natural (human) language. We will examine the core tasks in natural language processing, including morphological analysis, language modeling, syntactic analysis, probabilistic parsing, and semantical interpretation. We will also explore how these techniques can be used in several applications.

Prerequisites:

- Background in Artificial Intelligence

Text Books:

- Speech and Language Processing, D.Jurafsky, J.H.Martin, **2nd & 3rd Editions**, Pearson-Prentice Hall, 2009/2018.
- (*Supplementary*) Foundations of Statistical Natural Language Processing, C.D.Manning, H.Schütze, MIT Press, 2002.

Reference Books:

- Handbook of Natural Language Processing, N.Indurkha, F.J.Damerau (eds), Chapman & Hall, 2010.
- Introduction to Natural Language Processing, J.Eisenstein, MIT Press, 2019
- Natural Language Processing in Action: Understanding, analyzing, and generating text with Python, H.Lane, H.Hapke, C.Howard, Manning Pub., 2019

Lecture Hours and Rooms:

Tuesday 14:00-17:00 Online

Course Schedule (subject to change):

Introduction
Basic Text Processing
Morphological Analysis
Paper presentations (Basic text processing)
N-gram Language Models
Smoothing
Naive Bayes Classification
Logistic Regression Classification
Lexical Semantics
Word Embeddings
Word Classes and Part-of-Speech Tagging
Hidden Markov Models
Paper presentations (LMs, Word semantics, POS tagging)
Grammar Formalisms and Treebanks
Syntactic Parsing with CFGs
Shallow Semantic Parsing
Statistical Parsing and Probabilistic CFGs
Dependency Parsing
Paper presentations (Syntactic & semantic parsing)
Semantic Representation
Computational Semantics
Research project presentations

Evaluation (subject to change):

Midterm	: 15%
Presentations	: 20%
Research Project	: 15%
Application Project (2)	: 30% (2*15%)
Final	: 20%

(continued on next page)

Notes:

- Attendance for the exams, submitting the projects, and attendance for the presentations are required. Otherwise, you will fail the course, regardless of the grades obtained in other parts of the course.
- Two application projects will be assigned. In the scope of these projects, systems related to NLP tasks will be developed.
- A research project about an NLP topic will be prepared. A project report will be written and the project will be presented in the class.
- Paper presentations will be held throughout the semester. Given a topic, you will select papers from an NLP journal and present it in the class.
- You can follow the announcements via the university's Moodle system (<https://moodle.boun.edu.tr>). (Students registered to the course are automatically added to the course in Moodle.)
- The 2nd edition of the textbook is available at the book store. Both the 2nd and 3rd editions are available on the web. You can ask the instructor for the reference books.
- Please read the section "graduate courses" in the web page [General Information for Students](#). This page explains the course policy, the grading system, and information about the assignments and projects.