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 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.Indurkhya, F.J.Damerau (eds), Chapman & Hall, 2010.
- Natural Language Processing, E.Kumar, I K International Publishing House, 2011.
- Natural Language Processing for Online Applications : Text Retrieval, Extraction and Categorization, P.Jackson, I. Moulinier, John Benjamins, 2007.
- Natural Language Processing with Python, S.Bird, E.Klein, E.Loper, O'Reilly Media, 2009.
- Natural Language Processing and Text Mining, A.Kao, S.R.Poteet (eds), Springer, 2007.

Lecture Hours and Rooms:
Tuesday 14:00-17:00 BM A5

Course Schedule (subject to change):
Introduction
Regular Expressions
Basic Text Processing
Morphological Analysis
Finite State Transducers
N-gram Language Models
Smoothing
Naive Bayes Classification
Logistic Regression Classification
Lexical Semantics
Word Embeddings
Neural Language Models
Word Classes and Part-of-Speech Tagging
Hidden Markov Models
Sequence Processing with RNNs
Grammar Formalisms and Treebanks
Syntactic Parsing with CFGs
Statistical Parsing and Probabilistic CFGs
Shallow Semantic Parsing
Paper presentations

Evaluation (subject to change):

Midterm	: % 25
Application Project	: % 20
Research Project	: % 20
Final	: % 35

Notes:

- Midterm will be held on 12.11.2019.
- The midterm and final exams will be "closed books and notes".
- An application project will be assigned. In the scope of the project, a system related to an NLP task will be developed.
- A research project about an NLP topic/paper will be prepared. A project report will be written and the project will be presented in the class.
- You can follow the announcements via the university's Moodle system (https://moodle.boun.edu.tr).
- The textbook (2nd Edition) is available at the book store. You can consult 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.