Cmpe 561 Natural Language Processing

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 language modeling, syntactic analysis, semantic interpretation, and discourse analysis. We will also explore how these techniques can be used in applications such as information extraction, question answering, summarization, and sentiment analysis.

Instructor: Arzucan Özgür

Course Objectives:
- Learn the basic principles underlying natural language processing.
- Learn techniques and tools for developing natural language processing systems.
- Learn about the state of the art and open research problems in natural language processing.

Web site: Course content will be available on Moodle (https://moodle.boun.edu.tr). You can log in using your boun email account.

Textbook:
Daniel Jurafsky and James H. Martin, *Speech and Language Processing*, Second Edition, Prentice-Hall, 2008. http://www.cs.colorado.edu/~martin/slp.html

Supplementary Textbook:
Christopher D. Manning and Hinrich Schütze, *Foundations of Statistical Natural Language Processing*, MIT Press, 1999. http://nlp.stanford.edu/fsnlp/

Tentative List of Topics:
- Introduction
- Basic Text Processing
- N-gram Language Models
- Word Classes and Part-of-Speech Tagging
- Hidden Markov and Maximum Entropy Models
- Grammar Formalisms and Treebanks
- Parsing with Context Free Grammars
- Statistical Parsing and Probabilistic Context Free Grammars
- Lexical Semantics and Word Sense Disambiguation
- Semantic Role Labeling and Semantic Parsing
- Information Extraction
- Question Answering and Summarization
- Sentiment Analysis

Course Requirements:
The lectures will take place on Thursdays between 14:00-17:00. You are encouraged to attend and actively participate in the lectures.

The homework assignments will involve some programming where you will implement and test some of the techniques that we cover in class. You can use any programming language of your choice such as Java, Python, Perl, and etc.

The midterm exam will consist of problems covering the material in the lectures.

Each student (individually or in pairs) will be responsible for designing and completing a research project that demonstrates the ability to use concepts from the class. Each team will choose a project topic by selecting a recent scientific paper from an NLP conference or journal. The project will involve addressing the limitations of the existing work, and extending or replicating it in some way. The teams will give short project presentations in front of the class and submit a project report written in the format of a scientific paper in the end of the semester.

Grading:
- Midterm Exam: 20%
- 4 Homework Assignments: 40%
- Project: 35%
- Class Participation: 5%