CENG 463
INTRODUCTION TO NATURAL LANGUAGE PROCESSING

2023-2024 Fall

Instructor: Ayşenur Birtürk

TA: Atakan Garipler

Description: This is a foundational course on Natural Language Processing. You will learn fundamental algorithms and mathematical models for processing natural language, and how these can be used to solve practical problems.

Objective: To form a basis for Natural Language Processing and Computational Linguistics. To present applicable linguistic theory by turning theories into practical techniques with emphasis on problems for which there are widely accepted solutions. To familiarize the student with the latest advances in Natural Language Technology and applications like information extraction, machine translation, natural language interfaces, etc.

Text Book:

- Speech and Language Processing, Jurafsky and Martin, Prentice-Hall, 2008, 2.edition.

3.edition draft available at:

https://web.stanford.edu/~jurafsky/slp3

Reference Books:

- Foundations of Statistical Natural Language Processing, Manning and SchutzeMIT Press. Cambridge, MA: May 1999.
- Introduction to Information Retrieval, Manning, Raghavan, and Schutze,Cambridge University Press, 2008.
- Natural Language Understanding, J.Allen, 2.ed,Benjamin-Cummings, 1995. Prolog and

  Natural Language Analysis, F.C.N. Pereira, S.M. Shieber, CSLI, 1987.
- Natural Language Processing with Python– Analyzing Text with the NaturalLanguage Toolkit, Bird, Klein, and Loper, O'Reilly, 2009.
- Natural Language Processing for Prolog Programmers, Michael A. Covington,Prentice-Hall, 1994.

Syllabus (subject to change)

A. Fundamental Algorithms

1. Introduction
2. Regular Expressions, Text Normalization, Edit Distance
3. Hidden Markov Models
4. Spelling Correction and the Noisy Channel
5. N-gram Language Models
6. Naive Bayes and Sentiment Classification
7. Logistic Regression
8. Vector Semantics and Embeddings
9. Neural Networks and Neural Language Models
10. Sequence Labeling for Parts of Speech and Named Entities
11. RNNs and LSTMs
12. Transformers and Pretrained Language Models
13. Fine-tuning and Masked Language Models

B. Annotating Linguistic Structure

1. Context-Free Grammars and Constituency Parsing
2. Statistical Constituency Parsing
3. Dependency Parsing
4. Logical Representations of Sentence Meaning
5. Computational Semantics and Semantic Parsing
6. Relation and Event Extraction
7. Time and Temporal Reasoning
8. Word Senses and WordNet
9. Semantic Role Labeling and Argument Structure
10. Lexicons for Sentiment, Affect, and Connotation
11. Coreference Resolution
12. Discourse Coherence
13. Phonetics

C. Some NLP Applications :
Machine Translation, Question Answering and Information Retrieval, Chatbots and Dialogue Systems, Automatic Speech Recognition and Text-to-Speech

Lecture Slides:

https://web.stanford.edu/~jurafsky/NLPCourseraSlides.html

https://web.stanford.edu/~jurafsky/slp3

Plus additional slides

Grading (tentative):

Midterm Exam: 30%

Final Exam: 40%

Take Home Assignments (Programming/Written HWs): 30%

Bonus: Attendance %5

Exam Conduct (subject to change)

- Midterm and Final exams will be closed book and face to face (F2F).

Policy

- The take-home exam assignments that are designated as individual assignments must be completed individually. Copying from others, either from other students or off the internet is strictly forbidden and will surely constitute grounds for failure.
- In case of cheating in one of the take-home exams, all of the take-home exams will be graded as 0 and furthermore the case will be passed to the Student Discipline Committee.

Communication

All announcements will be made in ODTU Class.