

CPSC 477/577 Natural Language Processing

INSTRUCTOR

[Dragomir Radev](#)

SHORT DESCRIPTION

Linguistic, mathematical, and computational fundamentals of natural language processing (NLP).

Topics include part of speech tagging, Hidden Markov models, syntax and parsing, lexical semantics, compositional semantics, machine translation, text classification, discourse and dialogue processing. Additional topics such as sentiment analysis, text generation, and deep learning for NLP.

PRINCIPAL READINGS

Introduction to Natural Language Processing

Jacob Eisenstein

First Edition, October 2019

MIT Press

ISBN: 9780262042840

<https://github.com/jacobeisenstein/gt-nlp-class/blob/master/notes/eisenstein-nlp-notes.pdf>

Speech and Language Processing

Daniel Jurafsky and James Martin

Third Edition, 2019

Prentice Hall

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

WEEKLY READING WORKLOAD

Approximately 50 pages of the textbooks

TYPE OF INSTRUCTION

Lecture

WEEKLY MEETINGS

Two lectures of 75 minutes each.

DISTRIBUTION REQUIREMENTS

This course satisfies the Quantitative Reasoning (QR) requirement.

SAMPLE COURSE ASSIGNMENTS

- Assignment 0: The Natural Language Toolkit
- Assignment 1: Language Modeling, Hidden Markov Models, Part of Speech Tagging
- Assignment 2: Syntax, Dependency Parsing (or alternative)

- Assignment 3: Written problem set
- Assignment 4: Neural Machine Translation
- Assignment 5: Neural Semantic Parsing and NLP for database access
- Midterm
- Final exam

MAIN GOALS OF THE COURSE

1. Learn the basic principles and theoretical issues underlying natural language processing
2. Understand why language processing is hard
3. Learn techniques and tools used to build practical, robust systems that can understand text and communicate with users in one or more languages
4. Understand the limitations of these techniques and tools
5. Gain insight into some open research problems in natural language

PREREQUISITES

(CPSC 202 and CPSC 223) OR "permission of the instructor". All programming assignments are in Python.

DETAILED SYLLABUS

1. BACKGROUND, INTRODUCTION, LINGUISTICS, NLP TASKS

Class logistics, Why is NLP hard, Methods used in NLP, Mathematical and probabilistic background, Linguistic background, Python libraries for NLP, NLP resources, Word distributions, NLP tasks, Preprocessing

2. LANGUAGE MODELING, PART OF SPEECH TAGGING, HIDDEN MARKOV MODELS, SYNTAX AND PARSING, INFORMATION EXTRACTION

Language Modeling, Noisy Channel, Hidden Markov Models, The Viterbi Algorithm, Statistical Part of Speech Tagging, Syntax and Parsing, Context-Free Grammars, CKY Parsing, the Penn Treebank, Parsing Evaluation, Dependency Syntax, Dependency Parsing, Features and Unification, Tree-Adjoining Grammars, Combinatory Categorical Grammars, Noun sequence parsing

3. LEXICAL SEMANTICS, VECTOR SEMANTICS, COMPOSITIONAL SEMANTICS, KNOWLEDGE REPRESENTATION, SEMANTIC PARSING

Text Similarity, Stemming, WordNet, Word Similarity, Vector Semantics, Dimensionality Reduction, Representing Meaning, First Order Logic, Inference, Semantic Parsing, Abstract Meaning Representation, Sentiment Analysis

4. PRAGMATICS, DISCOURSE, DIALOGUE, APPLICATIONS OF NLP

Question Answering, Text Summarization, Text Generation, Discourse Analysis, Dialogue Systems, Machine Translation, Syntax-based Machine Translation

5. TEXT CLASSIFICATION, KERNEL METHODS, DISTRIBUTED REPRESENTATIONS

Text Classification, Vector Classification, Linear Models, Text clustering

6. NEURAL NETWORKS

Perceptron, Word Embeddings, word2vec, Deep Neural Networks, Sentence Representations, Neural approaches to question answering, parsing, machine translation, summarization, etc., Transformers, BERT.