

CPSC 477/577 Natural Language Processing

INSTRUCTOR

[Dragomir Radev](http://www.cs.yale.edu/homes/radev/nlp.html)

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 READING

Speech and Language Processing
Daniel Jurafsky and James Martin
Second Edition, 2009
Prentice Hall
ISBN-13: 978-0131873216
ISBN-10: 0131873210
<https://web.stanford.edu/~jurafsky/slp3/>

WEEKLY READING WORKLOAD

30-50 pages of the textbook

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

- Project 0: The Natural Language Toolkit
- Project 1: Language Modeling, Hidden Markov Models, Part of Speech Tagging
- Project 2: Syntax, Dependency Parsing
- Project 3: Question Answering, Deep Learning
- Project 4: Machine Translation
- Project 5: Semantic Parsing and SEQ2SEQ

- Written psets, midterm and 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, Brown Clustering, Information Extraction, Syntax and Parsing, Context-Free Grammars, CKY Parsing, the Penn Treebank, Parsing Evaluation, Lexicalized Parsing, Dependency Syntax, Dependency Parsing, Features and Unification, Mildly Context-Sensitive Grammars, Tree-Adjoining Grammars, Combinatory Categorical Grammars, Noun sequence parsing, Prepositional Phrase Attachment

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

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

4. PRAGMATICS, DISCOURSE, DIALOGUE, APPLICATIONS OF NLP

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

5. TEXT CLASSIFICATION, KERNEL METHODS, DISTRIBUTED REPRESENTATIONS, DEEP LEARNING

Text classification, Vector Classification, Linear Models, Perceptron, Support Vector Machines, Kernel Methods, Feature Selection, Text clustering, Word Embeddings, word2vec, Deep Neural Networks, Sentence Representations

ACADEMIC HONESTY

Unless otherwise specified in an assignment all submitted work must be your own, original work. Any excerpts, statements, or phrases from the work of others must be clearly identified as a quotation, and a proper citation provided. Any violation of the University's policies on Academic and Professional Integrity may result in serious penalties, which might range from failing an assignment, to failing a course, to being expelled from the program.

Violations of academic and professional integrity will be reported to Student Affairs. Consequences impacting assignment or course grades are determined by the faculty instructor; additional sanctions may be imposed.