

The University of Western Ontario
London, Canada

Department of Computer Science
CS 9660 — Computational Linguistics

Course Outline — Winter 2017

Course Description

This course is a graduate level introduction to Computational Linguistics. We will look at four main aspects of Computational Linguistics: the word (morphology, finite state automata, finite state transducers, part-of-speech tagging, n-grams, Hidden Markov Models, lexical semantics), the sentence (grammars, parsing, and semantics), discourse, and tools (Word2Vec and NLTK).

Objectives

Students will be exposed to topics that are current and important for doing research in Computational Linguistics. It is not expected that students will become experts in Computational Linguistics.

Prerequisites

Either a Computer Science background or a Linguistics background.

Instructor

Dr. Robert E. Mercer
MC 28-A-2
661-2111 Ext 86893
E-Mail: mercer@csd.uwo.ca

Textbook, Lecture Notes

Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (2nd ed.)

Daniel Jurafsky and James H. Martin

Slides of the lectures will be available on the course website (see below).

Course Website

<http://moodle.gaul.csd.uwo.ca>

Course Content

The course is divided into a two-hour class and a one-hour class.

The two hour class follows a typical lecture structure. Its content consists of material found in the textbook. There will be an introductory lecture and then 13 lectures covering the following topics, each lasting approximately two hours.

- Topic 1. Finite state automata
- Topic 2. Finite state transducers
- Topic 3. N-grams - Word sequences
- Topic 4. POS tagging and Hidden Markov Models
- Topic 5. Grammars
- Topic 6. Parsing
- Topic 7. More parsing (including statistical parsing)
- Topic 8. Semantics and Computational Semantics
- Topic 9. Lexical Semantics and Computational Lexical Semantics
- Topic 10. Discourse
- Topic 11. Information Extraction and Question Answering

Students are responsible for the readings on the Moodle.

The one-hour classes will be used to introduce software tools and to build software.

Project/Paper

The student will choose between a computer related project and a research paper.

The computer project will typically be an analysis of text data or the development of a software tool that processes text data. Another possibility is to analyze the ability of a software tool or application, comparing its ability to what humans can do or comparing it to tools that have been developed with a significant amount of human input. The outcome of the project will be the analysis or software together with a report that explains the analysis or software plus a short review of the papers that were read to motivate and/or give the background for the project. The report would be approximately 10 pages long. Of course, if appropriate, the analysis and report can be combined into a single document of about 15-20 pages.

The research paper will typically be a summary and analysis of a set of papers concerning a Computational Linguistics related topic. The paper would be approximately 15-20 pages long.

Presentation

The student will give a presentation to the class about his/her project/analysis/research paper. The presentation is to be no more than 25 minutes, followed by 5 minutes for questions. If a project/analysis is done the presentation should consist of the background material used for the project/analysis (even if it repeats some of the material already seen in lectures) and either the final results or the current state of the project/analysis together with what needs to still be accomplished. Part of the presentation should provide some discussion of why the results are what they are and if work still needs to be done, a speculative discussion of what is expected in the results and why.

The presentation will be judged on how well you present your topic (not on the results or current state of the work). Specifically, are the slides appropriate, and is the topic explained well for the audience.

Student Evaluation

Project/Paper: 65%

Presentation: 25%

Class participation: 10%

Project/Paper Schedule

The presentation will be in the week following classes (11 April 2017 and 13 April 2017).

The project/paper is due on 30 April 2017.