CMPT 413 - D100 Computational Linguistics

Instructor(s): Anoop Sarkar

Calendar Objective/Description:
This course examines the theoretical and applied problems of constructing and modeling systems, which aim to extract and represent the meaning of natural language sentences or of whole discourses, but drawing on contributions from the fields of linguistics, cognitive psychology, artificial intelligence and computing science.

Instructor's Objectives:
Imagine a world where you can pick up a phone and talk in English, while at the other end of the line your words are spoken in Chinese. Imagine a computer animated representation of yourself speaking fluently what you have written in an email. Imagine automatically uncovering protein/drug interactions in petabytes of medical abstracts. Imagine feeding a computer an ancient script that no living person can read, then listening as the computer reads aloud in this dead language. Natural Language Processing is the automatic analysis of human languages such as English, Korean, and thousands of others analyzed by computer algorithms that can make these applications possible. Unlike artificially created programming languages where the structure and meaning of programs is easy to encode, human languages provide an interesting challenge, both in terms of its analysis and the learning of language from observations.

Prerequisites:
completion of nine units in Computing Science upper division courses or, in exceptional cases, permission of the instructor.

Topics:
- Language models
- Edit distance
- Supervised machine learning for NLP
- Sequence labeling
- Unsupervised machine learning for NLP
- Machine translation
- Parsing and semantics

Grading:
Final Project: 30%. 1 Midterm: 18%. 5 Homeworks: 52%.
Students must attain an overall passing grade on the weighted average of exams in the course in order to obtain a clear pass (C- or better).

Recommended Books:
Natural Language Processing with Python, Steven Bird, Ewan Klein, and Edward Loper, O'Reilly Media, 2009, 9780596516499, see: http://www.nltk.org/book - Can be downloaded

Reference Books:
Foundations of Statistical Natural Language Processing, Christopher Manning and Hinrich Schutze, MIT Press, 1999, 9780262133609

Academic Honesty Statement:
Academic honesty plays a key role in our efforts to maintain a high standard of academic excellence and integrity. Students are advised that ALL acts of intellectual dishonesty will be handled in accordance with the SFU Academic Honesty and Student Conduct
Policies (http://www.sfu.ca/policies/gazette/student.html).