CMPT 413-3 SECT Computational Linguistics

WHERE LECTURE TIME TBA ... EXAM TIME TBA Instructor: A. Sarkar

OBJECTIVE/DESCRIPTION:

Computational Linguistics is the study of human language from a computational perspective. This course will examine algorithms used in the automatic analysis or production of language. Along with formal models of language, we will also study the engineering of natural language processing software. Many of the concepts, methods and algorithms studied are related to topics like formal linguistics, information retrieval, data mining from text, and bioinformatics. The course will also provide an introduction to programming with Perl. Assignments will develop expertise in Perl and provide exposure to commonly used datasets.

TOPICS:

- Introduction to linguistics and formal language theory
- Finite-state transducers: application to word morphology
- Introduction to probability and information theory
- Minimum Edit Distance
- Speech recognition: hidden markov models
- Document classification and clustering
- Word classes and part of speech tagging
- Mathematical linguistics and the Chomsky Hierarchy
- Feature structures and unification
- Machine translation
- Semantics
- Word-sense disambiguation and Wordnet
- Discourse and dialog models
- Natural Language Generation

GRADING:

Assignments (40%), Midterm (20%), Final Examination (40%).

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).

TEXTBOOKS:

- Speech and Language Processing, Daniel Jurafsky and James Martin, Prentice-Hall, 2000

RECOMMENDED:

- Programming Perl, Larry Wall, Jon Orwant, and Tom Christiansen, O'Reilly, 2000

REFERENCES:

- Foundations of Statistical Natural Language Processing, Christopher Manning and Hinrich Schutze, MIT Press, 1999
- Introduction to the Theory of Computation, Mike Sipser, PWS Publishing Co, 1996

PREREQUISITES/COREQUISITES:

Completion of nine credits in Computing Science upper division courses, or permission of the instructor.

Distributed: October 14, 2003