University of Toronto
St. George Campus

Computer Science 401/2511
Spring 2008

CSC 401/2511: Natural Language Computing
Course Information

Instructor: Gerald Penn
Lectures: WF 12–1, RW 110 (see web-page for exceptions)
Office: PT 396B
Tel: (416)978-7390
Office Hours: WF 1–2, or by appointment
Email: gpenn@cdf.utoronto.ca

Tutorials: M 12–1, RW 110 (see web-page for exceptions)

Teaching Assistants:

Name	Assignment
Chris Parisien	1
Xiaodan Zhu	2
Siavash Kazemian	3

Textbooks:

Required	C. Manning & H. Schuetze, *Foundations of Statistical Natural Language Processing*, MIT, 1999.
Optional	D. Jurafsky & J. Martin, *Speech and Language Processing*, Prentice Hall, 2000.
Recommended	A. Martelli, *Python in a Nutshell*, O'Reilly, 2003.
Optional	M. Lutz, D. Ascher, *Learning Python, 2nd ed.*, O'Reilly, 2003.

Course Web Page: `http://www.cs.toronto.edu/~gpenn/csc401/`

Evaluation: There will be three homework assignments (20% each), and a final exam (40%).

- To pass this class, you must pass (D- or higher) the final exam.
- No late homeworks will be accepted, except in case of documented medical or other emergencies.

Policy on collaboration: No collaboration on homeworks is permitted. The work you submit must be your own. Failure to observe this policy is an academic offense, carrying a penalty ranging from a zero on the homework to suspension from the university.

Course Goals: This course is an introduction to the application of statistical and computational methods to the processing of natural language text and speech. You will also learn the scripting language, Python.

Prerequisites: CSC 207 or 209 or 228, and STA 247 or 255 or 257, and a CGPA of 3.0 or a CSC subject POSt. MAT 223 or 240 is strongly recommended.

Newsgroup: The course newsgroup is on the web at `http://csc.cdf.toronto.edu/bb/YaBB.pl?board=CSC401H1S`. Your teaching assistants will be monitoring it.

Tentative Syllabus:

- Introduction to Corpus-based Linguistics
- Text Categorisation
- N-gram Models
- Markov Models
- Automatic Speech Recognition
- Part-of-Speech Tagging
- Information Retrieval
- Text Summarisation
- Statistical Machine Translation

Tentative Course Calendar:

Mon, 7 January	First lecture
Sun, 20 January	Last day to add course (CSC 401)
Fri, 18 January	Last day to add course (CSC 2511)
Mon, 11 February	Assignment 1 due
18–22 February	Reading Week — no lectures or tutorial
Fri, 29 February	Last day to drop course (CSC 2511)
Sun, 9 March	Last day to drop course (CSC 401)
Mon, 10 March	Assignment 2 due
Mon, 7 April	Assignment 3 due
Fri, 11 April	Last lecture
TBA	Final Exam