

CSC 401/2511: Natural Language Computing
Course Information

Instructor: Gerald Penn
Lectures: WF 12–1, SS 2118 (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, SS 2118 (see web-page for exceptions)
Teaching Assistants:

Name	Assignment
Cosmin Munteanu	1
Xiaodan Zhu	2
Frank Rudzicz	3

Textbooks:

Required C. Manning & H. Schuetze, *Foundations of Statistical Natural Language Processing*, MIT, 1999.
Optional D. Jurafsky & J. Martin, *Speech and Language Processing, 2nd ed.*, Prentice Hall, 2008.
Recommended A. Martelli, *Python in a Nutshell*, 2nd ed., O'Reilly, 2006.
Optional M. Lutz, *Learning Python, 3rd ed.*, O'Reilly, 2007.

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.

No student is permitted to discuss the final exam with any other student until the instructor or TAs make the solutions publicly available.

Failure to observe either of these policies 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, 5 January	First lecture
Fri, 16 January	Last day to add course (CSC 2511)
Sun, 18 January	Last day to add course (CSC 401)
Mon, 9 February	Assignment 1 due
16–20 February	Reading Week — no lectures or tutorial
Fri, 27 February	Last day to drop course (CSC 2511)
Sun, 8 March	Last day to drop course (CSC 401)
Mon, 9 March	Assignment 2 due
Mon, 6 April	Assignment 3 due
Wed, 8 April	Last lecture
TBA	Final Exam