CSC 401/2511: Natural Language Computing

Instructors: Frank Rudzicz & Chloé Pou-Prom
Email: csc401-2019-01@cs.toronto.edu
Lectures: MF 10h–11h, PB 250
Office hours: M 11h00–12h00, Vector Institute
Tutorials: W 10h-11h, MB 128
Web-page: `http://www.cs.toronto.edu/~frank/csc401/`
Forum: `http://piazza.com/utoronto.ca/winter2019/csc4012511`
TAs: A1: Zhewei Sun and Maryam Fallah
A2: Mohamed Abdalla and Raeid Saqur
A3: Amanjit Kainth and Jianan Chen

This course presents an introduction to natural language computing in applications such as information retrieval and extraction, intelligent web searching, speech recognition, and machine translation. These applications will involve various statistical and machine learning techniques.

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

Evaluation policies

CSC401/2511 students will be marked on three homework assignments and a final exam. The relative proportions of these marks are as follows:

Assignment with lowest mark	15%	language: Python
Assignment with median mark	20%	language: Python
Assignment with highest mark	25%	language: Python
Final exam	40%	

Graduate students enrolled in CSC2511 will have the option of undertaking a course project instead of the assignments in teams of at most two students for 60% of the course mark; all graduate students need to pass the final exam, which is worth 40% of the final mark. All code must run on the CDF machines.

Note that a 24-hour 'silence policy' will be in effect – we do not guarantee that the instructors or TAs will respond to your request within 24 hours of an assignment's due time.

Lateness

A 10% deduction is applied to late homework one minute after the due time. Thereafter, an additional 10% deduction is applied every 24 hours up to 72 hours late at which time the homework will receive a mark of zero. No exceptions will be made except in emergencies, including medical emergencies, at the instructor's discretion.

Final exam

A grade of at least D– on the final exam is required to pass the course. In other words, if you receive an F on the final exam then you automatically fail the course, regardless of your performance in the rest of the course.

Academic offenses

No unauthorized collaboration on the assignments is permitted. The work you submit must be your own. 'Collaboration' in this context includes but is not limited to sharing of source code, correction of another's source code, copying of written answers, and sharing of answers prior to submission of the work (including the final exam). Failure to observe this policy is an academic offense, carrying a penalty ranging from a zero on the homework to suspension from the University. See the academic integrity page of the University of Toronto at `http://www.artsci.utoronto.ca/osai/students`.

Readings

Required	Christopher D. Manning and Hinrich Schütze (1999) *Foundations of Statistical Natural Language Processing*, MIT Press. ***Available free at:*** `http://go.utlib.ca/cat/10552907`
Optional	Daniel Jurafsky and James H. Martin (2008) *Speech and Language Processing, 2nd ed.*, Prentice Hall.

Planned topics

- Introduction to corpus-based linguistics
- *N*-gram models and linguistic features
- Entropy and information theory
- Hidden Markov models
- Statistical machine translation
- Neural language models and word embedding
- Articulatory and acoustic phonetics
- Automatic speech recognition
- Speech synthesis
- Information retrieval
- Dialogue and chatbots

Planned course calendar

7 January	First lecture
20 January	Last day to add CSC 401
21 January	Last day to add CSC 2511
11 February	Assignment 1 due
18–22 February	Reading week – no lectures or tutorial
25 February	Last day to drop CSC 2511
8 March	Assignment 2 due
17 March	Last day to drop CSC 401
5 April	Last lecture
5 April	Assignment 3 due
5 April	Project final report due
TBD April	Final exam
	See course website for details.