

CSC 401/2511: Natural Language Computing

Instructors: Annie En-Shiun Lee, Raeid Saqur, and Zining Zhu
Email: csc401-2023-01@cs.toronto.edu
Lectures: **M** 10h-11h (LEC0101), 11h-12h (LEC0201), EM 001.
W 10h-11h (LEC0101) AH 100, 11h-12h (LEC0201)EM 001.
Tutorials: **F** 10h-11h (LEC0101), 11h-12h (LEC0201); EM 001.
Web-page: <http://www.cs.toronto.edu/~raeidsaqur/csc401/>
Forum: Piazza
Office hours: TBD

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, CSC 311 (or equivalent) are 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 teaching 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

The final exam will be a timed 3-hour test. A grade of 50% or higher on the final exam is required to pass the course. In other words, if you receive a grade lower than 50% on the final exam then your final grade in the course will be no higher than 47%, regardless of your performance in the rest of the course.

Academic offences

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 offence, 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 <https://www.academicintegrity.utoronto.ca/>.

Readings

- Optional Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. MIT press, 2016 *Deep Learning*, 1st ed.. **Available at:** <https://www.deeplearningbook.org/>
- Optional 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 (2021) *Speech and Language Processing*, 3rd ed.. **Available free at:** <https://web.stanford.edu/~jurafsky/slp3/>

See website for additional readings.

Planned topics

1. Introduction to corpus-based linguistics
2. N -gram models and linguistic features
3. Entropy and information theory
4. Neural language models and word embedding
5. Machine translation (statistical and neural)
6. Hidden Markov models
7. Natural language understanding
8. Automatic speech recognition
9. Information retrieval, Commonsense Q/A
10. Interpretability and Large Language Models

Planned course calendar

See Academic dates & deadlines for undergraduate students and Sessional dates for graduate students for any changes.

9 January	First lecture
17 January	Last day to add CSC 2511
23 January	Last day to add CSC 401
10 February	Assignment 1 due
20 February	Last day to drop CSC 2511
20–24 February	Reading week – no lectures or tutorial
10 March	Assignment 2 due
14 March	Last day to drop CSC 401
8 April	Last lecture
8 April	Assignment 3 due
8 April	Project final report due
TBD April	Final exam
	See course website for details.