University of Toronto
St. George Campus

Computer Science 485/2501
Autumn 2018

CSC 485/2501: Introduction to Computational Linguistics
Course Information

Instructor:	Gerald Penn
Lectures:	R 11–1, WI 1016
Office:	PT 396B
Tel:	(416)978-7390
Office Hours:	immediately preceding lectures 10–11, or by appointment
Email:	gpenn@teach.cs.utoronto.ca

Tutorials: F 2–3, WI 1016
(Note: some tutorial days will be used for lectures)

Teaching Assistants:

Name	Assignment
Sean Robertson	1
Nona Naderi	2
Aditya Bhargava	3

Textbooks:

Required	Jurafsky, Daniel, and Martin, James H. *Speech and Language Processing*, **2nd edition**, Pearson Prentice-Hall, 2009. Available in paper and e-book rental versions (for the latter, go to `CourseSmart.com` and search for *Jurafsky*). We'll also be referring to the draft 3rd edition: `https://web.stanford.edu/~jurafsky/slp3/`. See also the errata list for the 2nd edition: `www.cs.colorado.edu/~martin/SLP/Errata/SLP2-PH-Errata.html`.
Required	Bird, Steven, Klein, Ewan, and Loper, Edward. *Natural Language Processing with Python*, O'Reilly, 2009. Free (in HTML) with online extras at `www.nltk.org/book`.
Recommended	Mertz, David. *Text Processing in Python*. Addison-Wesley, 2003. Free ASCII version at `Gnosis.cx/TPiP`.
Optional	Allen, James. *Natural Language Understanding*, 2nd edition. Benjamin/Cummings, 1995.
Recommended	Martelli, Ravenscroft and Holden. *Python in a Nutshell*, 3rd ed., O'Reilly, 2017.

Course Web Page: `http://www.cs.toronto.edu/~gpenn/csc485/`

Evaluation: For undergraduates registered in CSC 485, there will be three homework assignments worth one third of your final mark each. Those registered for CSC 2501 must complete the three homework assignments (25% each), as well as five essays on assigned research papers ($5 \times 5 = 25\%$). There is no final examination for either course code.

- *No late homeworks will be accepted, except in case of documented medical or other emergencies.*

Policy on collaboration: No collaboration on homeworks or essays 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 a statistical and computational characterization of natural language. You will also have the chance to practice programming in Python.

Prerequisites: For undergraduates, STA247H1/STA255H1/STA257H1 and CSC209H1, but CSC324H1/CSC330H1/CSC384H1 is strongly recommended. For advice, contact the Undergraduate Office on the fourth floor of the Bahen Centre or the instructor.

Newsgroup: The course newsgroup is on the web at `https://bb-2018-09.teach.cs.toronto.edu/c/csc485`. Your teaching assistants will be monitoring it.

Tentative Syllabus:

Date	**Topic**	**Advance reading***
6–13 Sept	Intro to computational linguistics	**RP**; J&M: 1; BK&L: 1, *2.3, 4 as necessary*
13–20 Sept	Grammars and parsing	J&M: 5.0–1, 12.0–12.3.3, 12.3.7, *13.1–2*; BK&L: 8.0–8.4
21 Sept	Intro to NLTK	*BK&L: 1, 2.3, 4 as necessary*
27–28 Sept	Chart parsing	**RP**; J&M: 13.3–4; *A: 3.4, 3.6;* BK&L: 8.4 and online extras section 8.2 on chart parsing
4–12 Oct	Even More Parsing	**RP** (5th); J&M: 12.3.4–6, 15.0–3; *A: 4.1–5*; BK&L: 9
12–18 Oct	Ambiguity resolution	
18–25 Oct	Statistical attachment disambiguation	**RP**
25 Oct – 1 Nov	Lexical semantics	J&M: 19.1–4, 20.8
1–2 Nov	Word sense disambiguation	**RP**; J&M: 20.1–5
15–16 Nov	Statistical parsing	J&M: 5.2–5.5.2, 5.6, 12.4, 14.0–1, 14.3–7
16–22 Nov	Anaphoric resolution	J&M: 21.1–3
23–29 Nov	Semantic representations	J&M: 17.0–17.4.1, 17.5; BK&L: 10.0–4
30 Nov	Tutorial will meet as usual	

*J&M = Jurafsky and Martin; BK&L = Bird, Klein, and Loper; A = Allen; **RP** = research paper distributed on-line; *italics indicates optional additional reading.*

Tentative Course Calendar:

Thu, 6 September	First lecture
Thu, 13 September	Write-up 1 due (CSC 2501)
Wed, 19 September	Last day to add course (CSC 485)
Mon, 24 September	Last day to add course (CSC 2501)
Thu, 27 September	Write-up 2 due (CSC 2501)
Fri, 5 October	Assignment 1 due
Thu, 11 October	Write-up 3 due (CSC 2501)
Thu, 25 October	Write-up 4 due (CSC 2501)
Mon, 29 October	Last day to drop course (CSC 2501)
Thu, 1 November	Assignment 2 due
Mon, 5 November	Last day to drop course (CSC 485)
5–9 November	Reading Week — no lectures or tutorial
Thu, 15 November	Write-up 5 due (CSC 2501)
Thu, 29 November	Last lecture
Thu, 6 December	Assignment 3 due