Ling 317 Introduction to Computational Linguistics

Washington University in St. Louis
Dr. Ian Hollenbaugh | hian@wustl.edu
Assistant: Ariel Haberman | arielhaberman@wustl.edu

Course Info

Course Number	L44 Ling 317
Semester	Spring 2022
Time	M/W 2:30–3:50 PM
Location	Busch 202 / Zoom
Office	Umrath 127 / Zoom
General office hours (Zoom only)	M/R 1–2pm and by appointment
Ian's Zoom room	https://wustl.zoom.us/j/5274381745
Coding office hours (Zoom only)	T 2:30–4:00pm, R 3:00–4:30pm
Ariel's Zoom room	https://wustl.zoom.us/j/7715396897

*all Zoom links available on Canvas

Description

This course introduces the computational tools, both practical and theoretical, to describe and analyze natural language. We will touch upon most of the major aspects of the field of computational linguistics and natural language processing, such as analysis of raw corpora; modeling phonological, morphological, & syntactic patterns; and learnability. We will use Python 3 throughout the course. This course is an introduction to these topics, and therefore no programming experience is required. (In fact, this course is *not* meant for students with an extensive background in computer science.) Students, however, are encouraged to complete optional exercises and readings on basic programming techniques to aid in the completion of assignments and discussions.
Prerequisites: L44 Ling 170D and either L44 Ling 258 or CSE 131.

Course Goals

1. Learn practical, computational tools (e.g., Python, NLTK) to manipulate natural language data.
2. Understand the computational properties of natural language, independent of whether it is being computed by machines or humans (i.e., treating grammars as purely mathematical objects).
3. Apply these techniques to solve problems in phonology, morphology, and syntax.

Required Materials

There are no required purchases for this course. Any readings will be made available online. Additionally, all software necessary is free as well. This course assumes you have (or have access to) a working computer with a desktop operating system (macOS, Windows, Linux, etc.). For those students working primarily off a Chromebook or iOS/Android device, there are cloud options available. More details are given on Canvas.

Readings

A majority of the readings will come from the following source:

- Steven Bird et al. (2014). *Natural Language Processing with Python*. 2nd ed. URL: http://www.nltk.org/book/

The entire second edition is available online at the link above; **do not purchase** a physical copy as the printed edition is the outdated first edition.

Software

All course material that is not a PDF is distributed in a Jupyter notebook for Python 3. If you are unsure what to install, or what this means, we will go over it the first week of class. There are detailed instructions on the Canvas site as well. If you have some experience, and have Python 3 installed, you can get a head start with `pip install jupyter`.

Special COVID-19 related policies

If you are symptomatic or have been exposed to COVID-19, please do not come to class. No penalties will result from following university safety regulations. I ask only that you **email me** to let me know of any absences, and we can discuss make-up work if necessary.

Grade

The grade breakdown is shown below.

Category	Weight
Participation & Discussion*	40%
Weekly Assignments**	40%
Final Project	20%

*Includes Jupyter notebooks (graded on completion).
**Graded on correctness. The lowest grade is dropped.

Participation entails completing & uploading the in-class notebooks to Canvas once completed, and regular attendance and engagement in class discussion and on the course Piazza.

Letter grades

Letter grades are assigned based off the following scale. Numerical grades are **not** rounded.

- 100 ≥ A+ ≥ 97
- 97 > A ≥ 93
- 93 > A– ≥ 90
- 90 > B+ ≥ 87
- 87 > B ≥ 83
- 83 > B– ≥ 80
- 80 > C+ ≥ 77
- 77 > C ≥ 73
- 73 > C– ≥ 70
- 70 > D+ ≥ 67
- 67 > D ≥ 63
- 63 > D– ≥ 60

If you are taking this class pass/fail, you must receive at least a C– (70%) to pass.

If you believe there has been an error in grading, I am happy to discuss it with you. However, you must bring it up to me within one week of the graded assignment being returned to you. After this, the grade is considered final.

Schedule

The exact schedule might change as the class progresses. Please see Canvas for all readings and assignments, with due dates.

Date	Day	Week	Comment	Unit
1/19/2022	Wednesday	1		Introduction
1/24/2022	Monday	2		Introduction
1/26/2022	Wednesday	2		Chatbots
1/31/2022	Monday	3		Chatbots
2/2/2022	Wednesday	3		Regular expressions
2/7/2022	Monday	4		Regular expressions
2/9/2022	Wednesday	4		Text Processing
2/14/2022	Monday	5		Text Processing
2/16/2022	Wednesday	5		Text Processing
2/21/2022	Monday	6		Text Processing (TMNT)
2/23/2022	Wednesday	6		tf-idf
2/28/2022	Monday	7		ngram Modeling
3/2/2022	Wednesday	7		ngram Modeling
3/7/2022	Monday	8		MLE Modeling
3/9/2022	Wednesday	8		MLE Modeling

Date	Day	Week	Comment	Unit
3/14/2022	Monday	9	SPRING BREAK	**No Class**
3/16/2022	Wednesday	9	SPRING BREAK	**No Class**
3/21/2022	Monday	10		Finite State Automata
3/23/2022	Wednesday	10		Finite State Automata
3/28/2022	Monday	11	Project proposal due	Finite State Transducers
3/30/2022	Wednesday	11		Finite State Transducers
4/4/2022	Monday	12		Sentiment Analysis
4/6/2022	Wednesday	12		Pushdown Automata
4/11/2022	Monday	13		Context-free Grammars
4/13/2022	Wednesday	13		Context-free Grammars
4/18/2022	Monday	14		Presentations
4/20/2022	Wednesday	14		Presentations
4/25/2022	Monday	15		Presentations
4/27/2022	Wednesday	15	Last Day of Class	Presentations

Academic Integrity

This course adheres to the university's Academic Integrity Policy (https://studentconduct.wustl.edu/academic-integrity), and takes cheating and plagiarism very seriously. All work completed online must be done alone unless instructed otherwise, and no resources not approved by the instructor may be used during exams.

ADA Compliance

Washington University is committed to providing accommodations and/or services to students with documented disabilities. Students who are seeking support for a disability or a suspected disability should contact Disability Resources at 935-4153. Disability Resources is responsible for approving all disability-related accommodations for WU students, and students are responsible for providing faculty members with formal documentation of their approved accommodations at least two weeks prior to using those accommodations. I will accept Disability Resources VISA forms by email and personal delivery. If you have already been approved for accommodations, I request that you provide me with a copy of your VISA within the first two weeks of the semester. Please see more information at http://cornerstone.wustl.edu.

Sexual Assault Resources

The University is committed to offering reasonable academic accommodations (e.g., no contact order, course changes) to students who are victims of relationship or sexual violence, regardless of whether they seek criminal or disciplinary action. If you need to request such accommodations, please contact the Relationship and Sexual Violence Prevention Center (RSVP) at rsvpcenter@wustl.edu or 314-935-3445 to schedule an appointment with an RSVP confidential, licensed counselor. Information shared with counselors is confidential.

However, requests for accommodations will be coordinated with the appropriate University administrators and faculty. Please see more information at https://students.wustl.edu/relationship-sexual-violence-prevention-center.