

INTRODUCTION TO APPLIED COMPUTATIONAL LINGUISTICS

SPRING 2022

AUGUST 23, 2022

Sections: IIS 4/6511

Time: Tues/Thurs 11:20 am - 12:45 pm

Place: 407 Fedex Institute of Technology

Instructor: Andrew Olney

Email: aolney@memphis.edu

Office: FedEx Institute of Technology, Room 403e

Office Hours: Monday 2:30 pm - 3:30 pm

Phone: 678-5032

Web: <https://olney.ai>

Description

From the catalog:

Understand how to explore, model, and reason about linguistic data. Identify categories of linguistic problems and apply and extend existing methods to address these problems. The course is intended for students with little or no programming experience from any major. Programming in the course will focus on libraries in Python rather than core algorithms (as in COMP 7/8780).

You should take this class if:

- You want to understand NLP algorithms *eventually*
- You want to do things with NLP *today*

Our focus will be on using existing NLP tools to solve problems. The course format is minimal lecture and maximal hands-on problem solving in class.

Objectives

- Explain how computational linguistics can be applied across different fields.
- Apply computational practices to quantify linguistic phenomena.
- Analyze linguistic problems and applicable methods.
- Identify the impacts of computational linguistics on society.

Methods and Activities

- Read and prepare questions before class
- Brief class lectures

- In-class labs and activities
- Out-of-class homework
- Independent projects

Materials

Laptop You must bring a laptop with 10GB disk space and a WiFi connection to complete class exercises.

Book Bird, S., Klein, E., & Loper, E. (2019). *Natural language processing with Python*. Available at: <http://www.nltk.org/book/>

Software The following software resources, which we will install in class, will also be used:

JupyterLab <https://jupyter.org/> and **Blockly extension**

spaCy <https://course.spacy.io/>

scikit-learn <https://scikit-learn.org/stable/tutorial/index.html>

gensim <https://radimrehurek.com/gensim/>

Keras <https://keras.io/>

Additional resources Links to background readings will be linked in lecture slides if they are open access or available through the library and posted on eCourseware otherwise.

Reference resources Resources we will not directly use in class but which you may wish to refer are:

- Jurafsky, D. & Martin, J.H. (2000; 2014; in preparation). *Speech and language processing*. Read at: <https://web.stanford.edu/~jurafsky/slp3/>
- Manning, C. D. & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT press. Available from various sources.

The above are classic books used in NLP courses. Both are fairly technical, but they are great resources if you want to go deeper on a topic. They are in the U of M library system.

- Bender, E. M. (2013). *Linguistic fundamentals for natural language processing: 100 essentials from morphology and syntax*. Morgan & Claypool Publishers. Slides at: <https://faculty.washington.edu/ebender/papers/100things.pdf>
- Bender, E. M., & Lascarides, A. (2019). *Linguistic fundamentals for natural language processing II: 100 essentials from semantics and pragmatics*. Morgan & Claypool Publishers. Slides at: <https://faculty.washington.edu/ebender/papers/Bender-ACL2018-tutorial.pdf>

These cover basic aspects of linguistics in a crash-course, NLP oriented way. You might refer to them to better understand linguistic phenomena. Both books are in the U of M library system as electronic resources. We will be making use of the slide versions as selected readings.

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. New York: Springer. Read at: <http://faculty.marshall.usc.edu/gareth-james/ISL/>
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press. Preview at: <https://goo.gl/px2Zw3>

These books cover machine learning rather than NLP. The first (ISL) is a beginner book, and Murphy's book is more advanced. Again, this is for your reference only – use these if you want to dig deeper.

- [David Bamman's Applied Natural Language Processing class](#)

This class is out of the School of Information at Berkeley, and it has a more interdisciplinary orientation to NLP than most courses. I've found it to be an extremely useful model of an applied NLP course, and I have integrated some of the ideas and datasets into this course. Because Bamman's class assumes that students are already proficient at programming in Python, both his treatment of topics and programming exercises are too technical for IIS 4/6511. In many respects, IIS 4/6511 can be viewed as a suitable preparation for Bamman's course, so you may find it useful to refer to his materials or attempt his exercises at the conclusion of this course.

Data resources The following websites may be useful for coming up with project data:

- [NLTK corpora](#). Very easy to use. Some key NLP datasets are included, but the coverage of general text is limited.
- [Linguistic Data Consortium](#). Has a tremendous amount of data. The IIS has a license, so you can access some of it for free if you are on campus (see link below).
- [Kaggle](#). Kaggle is a competition website that shares many datasets, including text datasets. Jupyter notebooks are used in competitions, so you can also find example code for working with these datasets.
- [IIS dataset links](#). A large collection of links to data resources, along with tips for working with and cleaning data.

Grading

Grades will be on a +/- letter scale (e.g. A+/A/A-) and calculated according to the following percentages.

Class exercises (10%) In-class exercises, or labs, will be performed in groups.

Homework (40%) Homework will cover similar content as in-class labs but must be accomplished individually.

Projects (2 * 25%) Specifications for the projects will be outlined in the third week of class. **Students enrolled in 6511 will be required to submit a conference quality final project.**

Unless you are explicitly asked to work with others, you should do your own work. Talking through a problem with someone else is always OK, but you should never share code on an independent assignment. If you use code you found on the Internet, you must explicitly mark it and where it came from.

Assignments will be submitted using eCourseware (<https://elearn.memphis.edu>). You may sign into eCourseware from the Quick Links option on the U of M homepage. Then choose the link for this course to gain access to dropboxes for submitting assignments.

Attendance & Make-up Policy

Since in-class exercises constitute a significant portion of your grade, tardiness and missing classes will negatively impact your grade.

Work is expected to be turned in on time unless arrangements have been made prior to the due date (including University-sanctioned activities or legitimate religious observances).

No credit will be given for late work.

Extensions are given only under extreme circumstances and with prior permission of instructor.

Attempts to falsify attendance records is a violation of the University's Code of Student Rights and Responsibilities. Falsification includes entering attendance information for someone else, having someone else report that you attended when you did not, and using a fabricated story in an attempt to obtain an excused absence.

Electronic communications

Course announcements as well as consultation with the instructor may occur via e-mail. You are required to activate

your university e-mail account in order to be a student at the University. The University considers this account to be your official university e-mail address and will use it to disseminate information. You must either check your university e-mail account regularly or forward your university e-mail to a personal e-mail account that you will check regularly. After you have established your university e-mail account, you can use iAM, the University's identity management service, to forward your university e-mail to a personal e-mail account.

Office Hours

If you have any difficulties during the course, let me know immediately. Office hours are the perfect opportunity to discuss teaching and research issues. Do you want to have more information on a certain topic? Do you have difficulties finding a research question or writing a paper? Do you feel uncomfortable about assignments? Are you looking for work in a research lab? Do you have suggestions or comments? See me during office hours or send me an email.

Diversity in the Classroom

Diversity means the fair representation of all groups of individuals, the inclusion of minority perspectives and voices, and the appreciation of different cultural and socioeconomic group practices. We aspire to foster and maintain an atmosphere that is free from discrimination, harassment, exploitation, or intimidation. Academic courses will aim at providing opportunities for students to discuss issues of diversity including, but not limited to, ethnicity, gender, disability and sexual orientation as they can be related to course content. The University of Memphis has adopted policies prohibiting discrimination based upon race, sex, disability, or sexual orientation.

If you feel that you have experienced discrimination based on culture, disability, ethnicity, gender, generation, sexual orientation, national origin, privilege, race, and different views on religion, please contact the Office for Institutional Equity at the Administration Building, Room 156 (901-678-2713). To make a report, you may fill out an online form at <https://www.memphis.edu/report/submit-a-report/index.php>.

Special Accommodations for Disabilities

If you have a disability that interferes with completion of any coursework (including tests) or difficulty in accessing any course materials, (1) notify the instructor privately during the first two weeks of the course and (2) contact Disability Resources for Students (DRS) located in 110 Wilder Tower and at 678-2880. DRS offers a comprehensive program of services and academic accommodations designed to provide access and opportunity to students with disabilities. The instructor will work with you and DRS to determine how best to adapt course materials or instruction.

Academic Integrity

Students are expected to behave in accordance with the university's Code of Student Rights and Responsibilities, found here: <https://www.memphis.edu/osa/pdfs/csrr.pdf>. Plagiarism, cheating, and other forms of academic dishonesty will not be tolerated. Students engaging in academic dishonesty will receive a 0 on the associated assignment and may be reported to the chair of the department or the University's Office of Student Accountability and the Academic Integrity Committee

Classroom Misconduct

Students are expected to behave in accordance with the university's Code of Student Rights and Responsibilities, found here: <http://www.memphis.edu/saos/pdfs/csrr.pdf>. Disruptive behavior, use of an electronic or other noise- or light-emitting device that disturbs others, and excessive use of electronic devices for text messaging, telephone, or video-based conversations during instructional time in the classroom will not be tolerated. Students engaging in disruptive behaviors or general conduct that violates the rules and regulations of the university may be removed or excluded from the classroom. The instructor may report classroom misconduct to the Department Chair or the University's Office of Student Accountability.

Sexual Misconduct

All faculty, administrators, and most University staff are mandatory reporters. According to the University's Title IX policy and federal law, I must report potential incidents of sexual misconduct (harassment, assault, dating violence, domestic violence, and stalking) to the Office for Institutional Equity. If you tell me about (or if I become aware of) sexual misconduct, I will reach out to Office for Institutional Equity for assistance. For more information, contact the Office for Institutional Equity located in the Administration Building, Room 156 (901-678-2713). To make a report of sexual misconduct, you may fill out an online form at <http://www.memphis.edu/report>. To read the University's Sexual Misconduct policy, follow this link: <https://memphis.policytech.com/dotNet/documents/?docid=465&public=true>.

Student Health

As a student, you may experience a range of issues that can cause barriers to learning, such as physical health problems, strained relationships, increased anxiety, alcohol and drug problems, feeling down, difficulty concentrating, and lack of motivation. These mental and physical health concerns or stressful events may lead to diminished academic performance and may reduce your ability to participate in daily activities. The University of Memphis has a range of confidential mental and physical health services available on campus to assist you, including the Psychological Services Center in the Psychology Building, Room 126 (901-678-2147); The University Counseling Center at 214 Wilder Tower (901-678-2068); and The University Student Health Center at 200 Hudson Health Center (901-678-2287).

Syllabus Changes

The instructor reserves the right to make changes as necessary to this syllabus. If changes are necessitated during the term of the course, the instructor will immediately notify students of such changes both by individual email communication and posting both notification and nature of change(s) on the course website.

Course Schedule

Class	Topics
1	Course introduction and software install
2	Using JupyterLab and Blockly for programming and Markdown
3	Preprocessing text
4	Length-based metrics
5	Case studies 1
6	Distribution-based metrics
7	Vectorization and weighting
8	Case studies 2
9	Web scraping
10	Single word transformations
11	Multi-word transformations
12	Case studies 3
13	Latent variable vectorization
14	Linear regression
15	Logistic regression
16	Case studies 4
17	Causality in experiments
18	Decision trees
19	Random forest
20	Case studies 5
21	Case studies 6
22	Topic modeling
23	Ethics in NLP
24	Natural language generation
25	Speech recognition
26	Case studies 7
