Computational Linguistics I
CMSC 723/LING 723/INST 735

Jordan Boyd-Graber and Naomi Feldman

jbg@umiacs.umd.edu
nhf@umiacs.umd.edu

Syllabus, Fall 2024

Course Webpage:
https://users.umiacs.umd.edu/~jbg/teaching/CMSC_723/index.html

Description and Goals

Computational linguistics has two main areas: how to build technology that does useful things with human language (this area is usually referred to as "natural language processing", NLP, or sometimes "human language technology"), and how to improve our scientific understanding of how language works using computational methods and models. We will be looking at both, with an emphasis on conceptual understanding, looking at data, and understanding the ways in which the computational study of language differs from other areas in which computational approaches are being used.

Computers have made it possible, even easy, to collect vast amounts of text from a wide variety of sources. It is not always clear, however, how to use those data and how to extract useful information from data. This problem is faced in a tremendous range of scholarly, government, business, medical, and scientific applications. The purpose of this course is to teach some of the best and most general approaches to get the most out of these data.

Learning Objectives

By the end of the course, students should understand and implement the typical annotations of the natural language processing pipeline: segmenting characters into words or word pieces, representing those components and larger linguistic structures with distributed representations, and then using those representations for downstream generation or other tasks. Moreover, students should have the foundation to implement new algorithms and approaches for natural language processing tasks.

Course Discussion List

This term we will be using Piazza for class discussion. The system is highly catered to getting you help fast and efficiently from classmates, the TA, and the faculty. Rather than emailing questions to the teaching staff, we encourage you to post your questions on Piazza. If you have any problems or feedback for the developers, email team@piazza.com.

Find our class signup link at: https://piazza.com/umd/spring2021/cmsc723

Expect the Unexpected

Although every effort has been made to be complete and accurate, unforeseen circumstances arising during the semester could require the adjustment of any material given here. Consequently, given due notice to students, the instructors reserve the right to change any information on this syllabus or in other course materials. If you have concerns about any changes please discuss them with one of the instructors.

Required Background

Mathematical maturity: We will work extensively with probability and mathematical functions such as logarithms and differentiation. You should be comfortable manipulating these concepts algebraically.

- Linear algebra. We are both huge fans of 3Blue1Brown. These are videos that provide incredibly intuitive explanations for mathematical ideas, particularly for people who think visually. Even for people who are already comfortable with the mechanics of linear algebra, westrongly recommend the 3B1B 3B1B series on linear algebra, because the explanation of core concepts is incredibly helpful in understanding what's going on with neural networks. And those lead quite nicely into their really nice videos introducing the intuitive fundamentals of deep learning in a similarly visual and intuitive way.
- Probability and statistics. You're assumed to have the basics.

We will make extensive use of the Python programming language. It is assumed that you know or will quickly learn how to program in Python. Apart from a quick introduction, there will be no introduction to this skill-set.

- The NLTK book is a nice place to go for useful background and Python basics using NLP examples.

The computer-based aspects of this course will be oriented toward Unix-like operating systems (Linux, OS X). It may be possible to complete the course using other operating systems, but you will be responsible for troubleshooting any issues you encounter.

- Ken Church's venerable Unix for Poets is still one of the nicest concise introductions to fundamental and relevant Unix commands.

Grading

Components of the final grade are as follows:

	Percentage
Homework	25%
Exam	35%
Final Project	25%
Quizzes	5%
Participation	10%
	100%

It is possible to earn extra credit by going above and beyond the expectations of the assignment.

Letter grades will be assigned as follows:

- 98.00+ A+
- 92.50-97.99 A
- 89.50-92.49 A-
- 86.50-89.49 B+
- 82.50-86.49 B
- 79.50-82.49 B-
- 76.50-79.49 C+
- 72.50-76.49 C
- 69.50-72.49 C-
- 59.50-69.49 D
- 0.00-59.49 F

We reserve the right to curve up the threshold (i.e. a lower point value may result in a higher grade), but we will not curve down (i.e., a higher point value will not result in a lower grade). The thresholds will be placed uniformly for the entire class.

Incomplete as a final Grade

We will not issue an 'incomplete' as a course grade except for serious, valid reasons, generally in the category of serious emergencies. If you are having problems of any kind, please *talk to me as soon as possible*. In the event that a medical issue interferes with any major class

requirement, such as the midterm or final project, you are required to let us know in advance or as quickly as can reasonably be expected, and to provide documentation signed by a health care professional.

Assignments

There will be a number of homework assignments (with different numbers of points). Together, they are worth a large percentage of your final grade. Assignments are designed to help you learn the material, so please use them for that! You are allowed, and even encouraged, to collaborate with others (as many people as you'd like), but you must turn in your own assignment. For example, you could work together in a group, but each person must write up their solutions individually. Everything you submit must be your own. Directly from your fingers. Copying and pasting from another student will be considered plagiarism and will be handled according to the university's academic integrity policies.

Recognizing the distinction between cheating and cooperation is very important. If you simply copy someone else's solution, you are cheating. If you let someone else copy your solution, you are cheating. If someone dictates a solution to you, you are cheating. Everything you hand in must be in your own words, and based on your own understanding of the solution. If someone helps you understand the problem during a high-level discussion, you are not cheating. If you work collaboratively with explicit permission from the instructor, you are not cheating. We strongly encourage students to help one another understand the material presented in class, in the readings, and general issues relevant to the assignments. Any student who is caught cheating will be given an F in the course and referred to the Office of Student Conduct. Please don't take that chance -- if you're having trouble understanding the material, or if you need some help clarifying what is ok to do and what is not, please let us know and we will be more than happy to help (if you ask your questions on Piazza, we will warn you when you're getting close to the "danger zone").

The above applies not just to homework coding solutions but also to your written material. Please review appropriate citation policy.

Assignments are due at 11:59pm Eastern on the day indicated. Late policy: each person has seven free late days to be used, no questions asked, during the course. Late assignments beyond the provided late days will not be graded. The grade of the lowest homework assignment will be thrown out.

Midterm

This will be an in-class paper exam, and it will not involve programming. You will have access to a page of hand-written notes and can use a calculator but not access to notes or Internet.

Quizzes

We will have unannounced quizzes throughout the semester. This is to ensure students are keeping up with the course material. We will drop the lowest quiz grades. Because the quizzes

are frequent and many of the quiz grades will be dropped, there will not be the ability to make up missed quizzes. With very rare exceptions, most students get full credit for the 5% of the grade that is made up by quizzes (also the questions are quite easy!).

Final

This will be structured as a significant, long-term team project that will definitely involve programming. It typically involves an open research problem that we will give you. This project will have three stages:

- an initial proposal, explaining what you aim to do
- a first-step deliverable (due three quarters into the semester) to make sure students are grappling with the problem, e.g. a transformation of the data, performing a preprocessing step, or creating a baseline for comparison
- a final project report writeup

Grading is based on the writeup, so it is extremely important that you devote significant time and attention to quality when writing up the project; don't leave the writing to the last minute.

Class Participation

Each class is critical to your learning experience, and we expect you to come to class prepared (having read all assigned readings, ready to engage). We also expect active participation, not passive reception of the material. Your energy in contributing to class discussions and hands-on exercises will make this class an enjoyable experience for all of us. Class participation contributes to your total grade (see above).

We will also be using the online learning platform Piazza. You can get credit for participation by answering and asking useful questions on that platform. Ideally you should be participating both online and in class, however.

Academic Honesty

The University of Maryland, College Park has a nationally recognized Code of Academic Integrity, administered by the Student Honor Council. This Code sets standards for academic integrity at Maryland for all undergraduate and graduate students. As a student you are responsible for upholding these standards for this course. It is very important for you to be aware of the consequences of cheating, fabrication, facilitation, and plagiarism. For more information on the Code of Academic Integrity or the Student Honor Council, please visit http://www.shc.umd.edu.

The Honor Code and Honor Pledge prohibit students from cheating on exams, plagiarizing papers, submitting the same paper for credit in two courses without authorization, buying papers, submitting fraudulent documents, and forging signatures. We expect you to follow the academic integrity policy but we are exempting the class from the requirement of hand-writing and signing the honor pledge.

Absences for Religious Holidays and Interviews

Please send the course staff a private Piazza message with a list of all holidays you observe during the semester by the end of the first week of classes, so they can be taken into account in the course schedule.

Per university policy, job interviews are not considered excused absences. We'll try to be flexible, but don't schedule interviews during final project presentations or midterm exam.

Accessibility and Disability Services

See https://www.counseling.umd.edu/ads for official information. Students with a documented disability should inform course staff via a private Piazza post within the add-drop period if academic accommodations will be needed. We will plan together how accommodations will be implemented throughout the semester. To obtain the required Accommodation Letter, please contact Accessibility and Disability Service (ADS) at 301-314-7682 or adsfrontdesk@umd.edu.

Mental Health

Let's face it: grad school can be really hard. Sometimes students don't know that they need help, or they somehow know they're in trouble but they don't know what to do about it. What's really important for you to know is that at a big university like this one, you don't need to cope with it alone. There are many people on this campus who know how to help students in all kinds of circumstances. It's their job. Some resources you can take advantage of include the Counseling Center, in the Shoemaker Building, 301-314-7651, and Mental Health Services, in the Health Center, 301-314-8106; the Office of Student Affairs, 301-314-8430, is another place you can connect with to find help of various kinds.
If you are concerned about the behavior of another student, and in particular if you are worried that they might pose a threat to themselves or others, see this page for students concerned about another student.

Names and Pronouns

Many people might go by a name in daily life that is different from their legal name. In this classroom, we seek to refer to people by the names that they go by. Pronouns can be a way to affirm someone's gender identity, but they can also be unrelated to a person's identity. They are simply a public way in which people are referred to in place of their name (e.g. "he" or "she" or "they" or "ze" or something else). In this classroom, you are invited (if you want to) to share what pronouns you go by, and we seek to refer to people using the pronouns that they share. The pronouns someone indicates are not necessarily indicative of their gender identity. Visit trans.umd.edu to learn more.

Because this is a large class and we want to learn your names, please bring a name card to class that identifies you as you'd wish to be addressed. Particularly for Chinese students, because Pinyin doesn't communicate tones, please write (for example) 王家衛 as Wáng Jiā Wèi or Wang2 Jia1Wei4. For longer names, feel free to denote syllable breaks (e.g., Go-pal-krish-na Vish-wa-nath).

Course Feedback

We welcome your suggestions for improving this class, so please don't hesitate to share your thoughts during the semester, e.g. as a private Piazza note! You will also be asked to give feedback using the CourseEvalUM system at the end of the semester.

Use of student work

Your completed work may be used by us in this or subsequent semesters for educational purposes. Before making such use of your work, We will either get your written permission, or render the work anonymous by removing all your personal identification from the material.

Basic Needs Securities

Any student who has difficulty affording groceries or accessing sufficient food to eat every day, or who lacks a safe and stable place to live, and believes this may affect their performance in this course, is encouraged to use the resources listed below for support. Students are better served and supported when such circumstances are shared with the professor. Please consider sharing your situation with your professor who may be able to assist you in finding the appropriate resources.

- Campus Pantry: Alleviates food insecurity and provides a safe space to distribute emergency food to current UMD students. The Campus Pantry is located in the Health Center, Heilsa Room 0143 (Ground Floor), and is open each Friday during the semester from 9 a.m. - 5 p.m. Individual appointments are also available. Contact 301.314.8054 or campuspantry@umd.edu. For information see http://campuspantry.umd.edu/.
- Fostering Terp Success: Provides a safe and supportive campus network for students who were or are in foster care, who are homeless or at risk of being homeless, and who are without a supportive family system. For information see https://umd.edu/fostering-terp-success.
- Student Crisis Fund: For students who have an unexpected critical situation and need immediate financial support. Students will be asked for basic information to describe their circumstances of the emergency need and what other sources of funds are available. For more information, visit http://www.crisisfund.umd.edu/gethelp.html.