CSCI 3832: Natural Language Processing - Remote Edition

(changes indicated in blue)

Instructor: Felix Muzny[1] (pronunciation: "Muse-knee"; pronouns: they/them and he/him)
Contact: muzny@colorado.edu
Office: ECOT 734 (engineering center tower)

Credit: 3 credits
Mondays, Wednesdays, Fridays 2 - 2:50pm, HALE 230

Graduate Student Staff: Parth Jawale
For grading concerns, submit re-grade requests on Gradescope or reach out to both:
Parth.Jawale@colorado.edu, muzny@colorado.edu

Office Hours: calendar on Canvas, some office hours are Zoom office hours, some are Piazza office hours

Course Overview

NLP is about getting computers to perform useful and interesting tasks involving spoken and written human language. NLP is sometimes referred to as *Computational Linguistics* to emphasize the fact that involves the combination of CS methods with research insights from Linguistics (the study of human language). Practical applications of NLP include question answering, machine translation, information extraction, and interactive dialog systems (both written and spoken). Modern NLP systems rely heavily on methods involving probability, linear algebra, and calculus --- often in combination with machine learning methods.

We'll be exploring both applications and the computational methods behind them. You should be prepared to get your hands dirty in terms of the math, programming, and data that comprise the behind the scenes components of NLP systems.

Course Goals
1. Develop an understanding of the general problems that people who work on NLP study and the strategies they use to solve them.
2. Understand the role of data, machine learning, and neural networks in NLP systems.
3. Be able to implement models to solve some "standard" NLP problems.
4. Have a handful of ideas of where one might start given a new problem with NLP elements.
5. Understand some of the motivating linguistic phenomena that make NLP problems hard and why these can be hard phenomena for computers to approach.

Topics
- Words, word counting, lexicons
- Probabilistic language modeling
- Text classification with language models

---

[1] Call them "Felix" or "Professor Muzny"

- Test classification with single layer neural networks
- Vector semantics & word embeddings
- Part-of-speech tagging with Hidden Markov Models
    - Viterbi algorithm (dynamic programming)
- Dependency parsing
- Semantic roles & role-labeling
- Semantic parsing
- Information extraction
- Interactive dialog agents
- Question answering systems

Textbook & Course Configuration

Readings (textbook): We'll be using draft chapters from the 3rd Edition of *Speech and Language Processing* by Dan Jurafsky and James H. Martin. You don't need to buy the current edition, draft pdfs of the new chapters are available from the textbook website. You can also download (and print, if you desire) the entire book from the website). It is also linked from Canvas.

Canvas: We'll be using Canvas for nearly everything in the class. Readings, assignments, general course information will be posted there. Homework submissions will be done through Gradescope.

Assignments: We will have both written and programming assignments. Descriptions will be posted on canvas and work will be submitted via Gradescope.

Gradescope: Gradescope is where you will submit assignments. You will also see your grades, feedback, and submit regrade requests via gradescope. You can find the link to Gradescope on Canvas.

Classroom Environment & Expectations

- **Preparation:** When there are readings assigned, it is the expectation that you do them before the first class meeting in the following week. This course will be a great opportunity for those of you who are interested in NLP & research to start flexing those muscles, and the best way for us to go down those paths is for you to develop a solid foundation.
    - see grading table for remote teaching updates
- **Classroom environment:** It is unusually common in Computer science classes for some students to ask questions that are not really questions so much as opportunities to demonstrate knowledge of vocabulary or facts beyond the topic at hand. This can have a discouraging effect on other students who are not familiar with those terms, causing them to worry that they are less prepared to do well in the class (this is rarely the case—knowing terms outside the scope of the course is not a good predictor of success). If you find yourself wanting to make such a question or comment, please come talk to me about the topic after class or during office hours—I'm always happy to discuss tangentially related topics at those times!
- **Attendance:** I expect all students to attend lecture and to actively participate in the work that we do during this time. (Do expect to be engaging with activities during lecture.)

- **Accommodation letters:** If you have an accommodations letter, please bring it to me at your earliest convenience so that I can make sure this class is meeting your needs.
- **Name and pronouns:** If your name and pronouns are not in alignment with those listed on our class roster, please let me know either in person or via email so that I can ensure you are correctly addressed in this class.
- **Class expenses:** If obtaining any material for use in our class presents a financial hardship for you, please let me know and I will work with you to locate the resources that you need to succeed in this class.
- **Feedback:** Please don't hesitate to reach out to me if any aspect of this course or class community could be improved.

Late Policy

All homework may be turned in up to 3 days (72 hours) late for a 20% penalty. If homework is due on Friday at 6pm, it may be turned in as late as Monday at 6pm. If a student would have received a 95% had they turned their homework in on time, a late submission will earn them a 75% instead.

I strongly advise you to carefully consider the impact of -20% on your grade.

Quizzes may not be completed after the deadline.

Extensions will be given based on proactive communication with Felix. We will be going by the honor system here. The sooner that you reach out to me, the better that I will be able to accommodate your request (within reason).
Email Felix (muzny@colorado.edu) with the following information:
1) Which assignment are you requesting an extension on.
2) When are you requesting the extension unil.
3) What is your plan for how this will impact your other assignments.

~~in the case of *exceptional* circumstances accompanied by a healthcare provider letter, military activation order, obituary/memorial service notice, police/fire report, etc.~~

All requests for extensions should be directed to Felix.

Collaboration Policy

You are expected to do your work by yourself or with your partner when indicated. You are expected to use online resources as documentation, not as a place to get solutions to your assignments.

Here are three big-picture points to remember when collaborating with your classmates:

- **Strategies:** You may talk with your classmates about *general strategies* but you may not talk about *specific solutions*.
- **Explaining concepts:** You may talk with your classmates about how certain techniques work *in general* but not how to write any part (or sub-part) of the solution needed for the homework.
- **A good rule of thumb:** don't show your assignments to other people; don't look at other people's assignments; don't write code together unless the assignment explicitly states that you may work in pairs. This includes working through solutions on whiteboards as well as telling your friend verbally what you have written.

The finer-grained details:

- **Do not search for a solution online**: You may not actively search for a solution to the problem from the internet. This includes posting to sources like StackExchange, Reddit, Chegg, etc.
  - **StackExchange Clarification**: Searching for basic techniques in C++ is fine. If you want to post and ask "How do convert a float to an integer" that's fine. What you **cannot** do is post are things like "Here's the function my prof gave me to write. I need to convert this temperature in celcius to farenheit".
- **Plagiarism:** assignments **and code** that you turn in should be written entirely on your own. You should not need to consult sources beyond the class notes, posted lecture notes, examples, and resources, and python and its associated libraries' documentation.
- **Tutors:** you should <u>always consult the course instructional staff</u> if you need extra help. They are here specifically to help you! You should never have anyone else write code for you. This includes tutors, friends, strangers, friends of friends, or anyone who is not you.
  - **A second note on tutors:** our TAs are here for you. They are not paid to do your assignments, but to help you learn. If you are using a tutor that is not helping you learn, **<u>they are not helping you</u>**.
  - **A third note on tutors:** if you turn in work that you are unable to explain, this will result in an honor code violation and a loss of credit.
- **When in doubt, ask**: If you have doubts about this policy or would like to discuss specific cases, please ask the instructor.

Collaboration Policy violations will result in both a 0 on the assignment in question and a final grade reduction of one letter grade.

Grading & Assignments

Category	Due Dates & Points	Grade Percentage
Homework	Due on Thursdays @ 11:59pm Homework consists of 5 - 6 coding and written assignments and 2 - 3 in-person interviews with the course staff. See schedule below for tentative due dates.	60% ~~65%~~
Quizzes	~~Quizzes are typically due on Sundays @ 11:59pm. There will be roughly 1 quiz per week. These are primarily used to help you prepare for homework and the final exam.~~ **You must earn 100% on Quiz 0 to receive a grade for this class.** The quiz category is finished after Quiz 8 and the switch to remote.	7.5% ~~15%~~

	You must earn 100% on Remote Quiz 0 to receive a grade for this class. This will be an easy way for you to make up some of the points that you may have missed in this category.	
Exams	~~The final exam is scheduled by the university tentatively to take place on May 3rd @ 1:30 - 4pm.~~ This will be a take home exam that feels similar to a long homework covering all the material in the course. It will be due @4pm on May 3rd. Confirm this time in MyCUInfo when it is listed later in the semester.	15%
Pre-Remote Class Participation	Class participation is determined based on your participation during lecture, on piazza asking questions and helping answer your peers' questions, during office hours, and on optional assignments. (update 1/14/2020) You will be given a notification about half-way through the semester indicating what percentage of these points you are on track to receive given your participation in the first half of the course. The Pre-remote Class Participation category is finished after Lecture 25 and the switch to remote.	2.5% ~~5%~~
Pre- and post-lecture quizzes	Before and after each lecture in the remote portion of the course (week 10 - end)	12%
Remote Piazza participation	Answering the questions of your fellow classmates, posing interesting questions, etc. You will need to post at least 2x a week, beginning in week 10 to get full credit. If you have no questions about content, you'll post in the weekly discussion thread.	3%

Calendar
This calendar is a draft calendar and is subject to slight modification as the semester progresses.

Week (Monday date)	Topics	Assignments due this week (other than weekly quizzes). Tentative. Always consult Canvas/Gradescope for up-to-date times.	Textbook chapters
1/13	Words, segmentation, python & unix tools, vocabularies, Heap's law		1, 2
1/20	Language models, probability, dealing with data and sparsity	Homework 1 (written)	3.1 - 3.3, 3.4.1, 3.6
1/27	Language modelling (cont'd), n-grams, byte-pair encoding, naïve bayes	Optional homework # 1	4
2/3	Text classification, logistic regression		5
2/10	Logistic regression, SGD, word meanings and vector semantics	Homework 2 (coding)	6
2/17	Vector semantics, word2vec, part-of-speech tagging, Hidden Markov Models		6, 7
2/24	HMMs, Viterbi		8
3/2	Viterbi, RNNs, intro to dependency parsing	Homework 3 (coding)	9, 12, 15
3/9	Parsing, beam search, representing meaning		15, 17
3/16	~~Representing meaning, first-order logic, semantic analysis,~~ Named-Entity Recognition	Homework 4 (part written, part coding)	17, 23
3/23	SPRING BREAK		
3/30	NER, information extraction	~~In-person interview # 1, also known as "Homework 5" to be conducted remotely over the course of weeks 11 & 12~~	18
4/6	Machine Translation	~~Optional homework # 2~~	supplemental material
4/13	State of the Art systems, bias	Homework 5	supplemental material
4/20	NLP, ML, and bias, digital humanities	~~Homework 6 (written + coding), also known as~~	

		"final project"	
4/27	Digital humanities, review		Final take home task

Accommodation for Disabilities

If you qualify for accommodations because of a disability, please submit your accommodation letter from Disability Services to your faculty member in a timely manner so that your needs can be addressed. Disability Services determines accommodations based on documented disabilities in the academic environment. Information on requesting accommodations is located on the Disability Services website. Contact Disability Services at 303-492-8671 or dsinfo@colorado.edu for further assistance. If you have a temporary medical condition or injury, see Temporary Medical Conditions under the Students tab on the Disability Services website.

Classroom Behavior

Students and faculty each have responsibility for maintaining an appropriate learning environment. Those who fail to adhere to such behavioral standards may be subject to discipline. Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with race, color, national origin, sex, pregnancy, age, disability, creed, religion, sexual orientation, gender identity, gender expression, veteran status, political affiliation or political philosophy. For more information, see the policies on classroom behavior and the Student Code of Conduct.

Preferred Student Names and Pronouns

CU Boulder recognizes that students' legal information doesn't always align with how they identify. Students may update their preferred names and pronouns via the student portal; those preferred names and pronouns are listed on instructors' class rosters. In the absence of such updates, the name that appears on the class roster is the student's legal name.

Honor Code

All students enrolled in a University of Colorado Boulder course are responsible for knowing and adhering to the Honor Code. Violations of the policy may include: plagiarism, cheating, fabrication, lying, bribery, threat, unauthorized access to academic materials, clicker fraud, submitting the same or similar work in more than one course without permission from all course instructors involved, and aiding academic dishonesty. All incidents of academic misconduct will be reported to the Honor Code (honor@colorado.edu); 303-492-5550). Students found responsible for violating the academic integrity policy will be subject to nonacademic sanctions from the Honor Code as well as academic sanctions from the faculty member. Additional information regarding the Honor Code academic integrity policy can be found at the Honor Code Office website.

Sexual Misconduct, Discrimination, Harassment and/or Related Retaliation

The University of Colorado Boulder (CU Boulder) is committed to fostering a positive and welcoming learning, working, and living environment. CU Boulder will not tolerate acts of sexual misconduct, intimate partner abuse (including dating or domestic violence), stalking, or protected-class discrimination or harassment by members of our community. Individuals who believe they have been subject to misconduct or retaliatory actions for reporting a concern should contact the Office of Institutional Equity and Compliance (OIEC) at 303-492-2127 or cureport@colorado.edu. Information about the OIEC, university policies, anonymous reporting, and the campus resources can be found on the OIEC website.

Please know that faculty and instructors have a responsibility to inform OIEC when made aware of incidents of sexual misconduct, discrimination, harassment and/or related retaliation, to ensure that individuals impacted receive information about options for reporting and support resources.

Religious Holidays

Campus policy regarding religious observances requires that faculty make every effort to deal reasonably and fairly with all students who, because of religious obligations, have conflicts with scheduled exams, assignments or required attendance. In this class, contact the course staff at least 7 days in advance to reschedule a homework due date or an exam.

See the campus policy regarding religious observances for full details.