

CS 4120: Natural Language Processing, Fall 2020

Instructor: Felix Muzny¹ (pronunciation: "Muse-knee"; pronouns: they/them and he/him)

Contact: f.muzny@northeastern.edu

Office: Meserve 307A

TAs: Harshit Sharma, Nalin Gupta

Contact: sharma.hars@northeastern.edu, gupta.nal@northeastern.edu

Contacting course staff: we prefer that you pose all general questions on the course piazza

<https://piazza.com/northeastern/fall2020/cs4120>

Credit: 4 credits, Mondays & Wednesdays 2:50 - 4:30pm, ~~ISEC 102~~ **Room changed to Robinson Hall 109**

Zoom link for synchronous remote lectures: **linked on canvas**

Office Hours: see your calendar on Canvas; all office hours will be held on zoom, and these will be scheduled via the Zoom Meetings tool within Canvas

Many of you may be feeling apprehensive about the coming semester. This is very reasonable. If you read this syllabus and take away nothing else, remember that you can always reach out to me if you have questions, if life happens, or if you are feeling that the world isn't respecting your dignity as a human being.

Now we'll talk about language, computers, and the logistics of this course.

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 Call them "Felix" or "Professor Muzny"

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
- Text classification with single layer neural networks
- Vector semantics & word embeddings
- Part-of-speech tagging
- Viterbi algorithm (dynamic programming)
- Interactive dialog agents
- Machine translation
- Information extraction*
- Question answering systems*
- Dependency parsing*

* if time allows

Textbook & Course Configuration

1. 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. We will also link this text from the course website.
2. We will supplement this text occasionally with readings from:
 - a. [Eisenstein, Jacob. *Introduction to Natural Language Processing*. MIT Press, 2019.](#)
 - b. [Kohn, Philipp. "Neural Machine Translation." arXiv preprint arXiv:1709.07809, 2017.](#)

Websites & Technology

Make sure that you have access to all of the following websites and software:

- **Canvas:** We'll be using Canvas for some quizzes and links to homework submissions. Homework submissions will be done through Gradescope.
- **Gradescope:** Gradescope is where you will submit homework and some quizzes. You will also see your grades, feedback, and submit regrade requests via gradescope. You can find the link to Gradescope on Canvas.
- **Piazza:** This is our course discussion forum. This is where we will discuss relevant topics and answer your homework and content questions that come up outside of class. If you send us a content question via email, we'll likely ask you to post it to Piazza instead!
- **Python 3:** We'll be writing our homework coding assignments using python 3.

- **Jupyter Notebooks:** Many of the coding activities that we complete in class will be distributed as Jupyter Notebooks. You can install jupyter notebooks either by installing [Anaconda](#) or via the [command line](#).
- **IDEs:** You can develop your code using whatever your preferred IDE is. When you have coding assignments, you'll submit your solutions as .py files (make sure to run any converted-from-jupyter-notebooks-.py-files again before you turn them in)! If you installed Anaconda, it comes with Spyder, which is an IDE that can be used to write and run .py files.

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.
- **Attendance:** You are strongly encouraged to attend lecture whenever possible. We will be doing interactive activities during lecture as well as covering the material necessary for you to complete your homework and quizzes. If you are unable to attend lecture on a given day, it is your responsibility to attend office hours, consult course materials, and communicate with course staff about what you have missed.
- **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!
- **Accommodation letters:** If you have an accommodation 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.
 - If you wish to add, change, or update your pronouns in Canvas, see [this knowledge base webpage](#). To update your display name in Canvas, see [this article](#).
- **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 should be turned in on time whenever possible. All homework may be turned in up to 1 day (24 hours) late for a 30% penalty. If homework is due on Friday at 6pm, it may be turned in as late as Saturday 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 65% instead.

Once a semester, you may turn in your homework up to 24 hours late without penalty. This will be automatically applied the first time you turn in your homework late.

Quizzes may not be completed after the deadline. We've built some extra credit into the course (you'll find this on the homework), so if you miss a quiz, make sure to attempt the extra credit on the homework.

Extensions

Extensions for any work beyond the regular late policy will be given based on proactive communication with Felix. Whenever possible, this should occur at least 24 hours before the posted deadline. The sooner that you reach out, the better that Felix will be able to work with you.

Email Felix (f.muzny@northeastern.edu) with the following information:

- 1) Which assignment are you requesting an extension on & why you are requesting an extension.
- 2) When are you requesting the extension until.
- 3) What is your plan for how this extension will impact the due dates for the other assignments in this course.

You don't need to write an essay, just be sure to include the above information. This extension policy is based on our mutual understanding that living during a pandemic is difficult, we're all doing our best, and the easiest way for Felix to help you succeed in this course is proactive communication. If a situation arises that makes it impossible to reach out 24 hours before the deadline, don't panic--send Felix an email when you and we'll discuss your options.

Collaboration Policy

The work that you turn in should be your own. We encourage you to collaborate with your classmates, but remember that collaboration looks very different than working on a pair or group project.

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 (this makes it very hard to come up with your own solution afterwards); 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.

You are expected to use the internet as a place for online resources, such as documentation, not as a place to get solutions to your assignments.

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 python 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 things like "Here's the function my prof gave me to write. I need to convert this temperature in celcius to fahrenheit".
- **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 always consult the course instructional staff 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.
- **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 a 0 on the assignment in question.

The university's academic integrity policy discusses actions regarded as violations and consequences for students.

<http://www.northeastern.edu/osccr/academic-integrity>

Grading & Assignments

Category	Due Dates & Points	Grade Percentage
Homework	Due on Thursdays. There will be 4 homeworks. No HW grades will be dropped. Most homework assignments have a small amount of extra credit that you may choose to do.	45%
Quizzes	Quizzes are due on Sundays* (weeks when there was not homework due). It may be helpful for you to think of quizzes as	25%

	<p>mini take-home exams.</p> <p>Quizzes will typically focus on material covered in class that week, and will have a small number of questions based on readings and/or videos for the upcoming week. No quiz grades will be dropped.</p> <p>* Except for Quiz 8, due Wednesday, December 2nd</p> <p>The final quiz (Quiz 8) will be worth approximately 40% of your quizzes grade and will be cumulative. You can think of it as your mini final take-home exam.</p> <p>If you enrolled in this course after Quiz 1 was due and you contact Felix, we will extend this deadline for you.</p>	
Class participation	<p>Class participation is determined based on doing lecture activities, participating on piazza (taking part in discussions and helping answer your peers' questions), and by attending office hours. Expect to actively participate in at least two of these three categories to earn these points.</p> <p>If you cannot attend lecture in person or synchronously, you may still submit lecture activities.</p> <p>You will receive this grade in three installments: each about $\frac{1}{3}$ of the way through the semester.</p>	10%
Final Project	<p>Your final project will consist of four portions:</p> <ol style="list-style-type: none"> 1) proposal (due Tuesday, November 24th) 2) code & write-up (due date dependent on the state of the world in November) 3) presentation (format & due date to be determined based on the state of the world in November) 4) presentation review (due date TBD based on the state of the world in November) 	20%

Calendar

The detailed calendar is posted on the course website.

An abbreviated version (subject to change) is:

Week	Topic	Reading
1	Introduction	SLP, Chp. 2
2	Vocabularies, normalization, n-grams, language models	SLP, Chp. 3
3	Language models, text classification	SLP, Chp. 4
4	Text classification, Naive bayes, logistic regression	SLP, Chp. 5
5	Word Embeddings	SLP, Chp. 6
6	Word Embeddings	SLP, Chp. 6
7	Reading research papers, neural networks	SLP, Chp. 7; supplementary materials
8	Sequence labelling, Viterbi	SLP, Chp. 8
9	RNNs	SLP, Chp. 9
10	Ethics & NLP	Supplementary materials
11	Machine Translation	Supplementary materials
12	Machine Translation	Supplementary materials
13	Machine Translation	Supplementary materials
14	TBD based on the state of the world in November/December	

Classroom Recording

This course, or parts of this course, will be recorded for educational purposes. These recordings will be made available only to students enrolled in the course, instructor of record, and any teaching assistants assigned to the course.

If you have objections or would like to opt-out of recordings, please contact the instructor.

Only students who have arranged an accommodation with the Disability Resource Center may use mechanical or electronic transcribing, recording, or communication devices in the classroom. Students with disabilities who believe they may need such an accommodation may contact the Disabilities Resource Center.

Accommodations

It is my job to create a classroom environment that is most conducive to you learning well. If you have accommodations from the [Disability Resource Center](#), please provide your letter to me early in the semester so that I can arrange for these accommodations. If you wish to receive accommodations and do not have a letter, please visit the DRC at 20 Dodge Hall or call (617) 373-2675.

Student Names and Pronouns

We recognize that your legal information doesn't always align with how you identify. Students may update their first and middle names as well as gender marker [with the registrar](#), even if they are not your legal names or gender. Those names and gender marker are what would appear publicly in most university systems. In the absence of such updates, what we see on most university systems by default are your legal name and gender marker.

Classroom Environment

To create and preserve a classroom atmosphere that optimizes teaching and learning, all participants share a responsibility in creating a civil and non-disruptive forum for the discussion of ideas. Students are expected to conduct themselves at all times in a manner that does not disrupt teaching or learning. Your comments to others should be constructive and free from harassing statements. You are encouraged to disagree with other students and the instructor, but such disagreements need to be respectful and be based upon facts and documentation (rather than prejudices and personalities). The instructor reserves the right to interrupt conversations that deviate from these expectations. Repeated unprofessional or disrespectful conduct may result in a lower grade or more severe consequences.

Part of the learning process in this course is respectful engagement of ideas with others.

The [Code of Student Conduct can be found on the OSCCR website](#).

Title IX

Title IX of the Education Amendments of 1972 protects individuals from sex or gender-based discrimination, including discrimination based on gender-identity, in educational programs and activities that receive federal financial assistance.

Northeastern's Title IX Policy prohibits Prohibited Offenses, which are defined as sexual harassment, sexual assault, relationship or domestic violence, and stalking.

The Title IX Policy applies to the entire community, including male, female, non-binary, and transgender students, faculty and staff.

If you or someone you know has been a survivor of a Prohibited Offense, confidential support and guidance can be found through [University Health and Counseling Services](#) staff and the [Center for Spiritual Dialogue and Service](#) clergy members.

By law, those employees are not required to report allegations of sex or gender-based discrimination to the University.

Reports can be made non-confidentially to the Title IX Coordinator within the Office for Gender Equity and Compliance at: titleix@northeastern.edu and/or through NUPD (Emergency 617.373.3333; Non-Emergency 617.373.2121).

Reporting Prohibited Offenses to NUPD does NOT commit the victim/affected party to future legal action.

Faculty members are considered "responsible employees" at Northeastern University, meaning they are required to report all allegations of sex or gender-based discrimination to the Title IX Coordinator.

In case of an emergency, please call 911.

Please visit <http://www.northeastern.edu/titleix> for a complete list of reporting options and resources both on- and off-campus.

Religious Holidays

The course staff will 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 of the conflicting date to reschedule a homework or quiz due date.