Natural Language Processing (CS322)
Instructor: Jack Hessel (jhessel@carleton.edu)
Where/When: CMC 301; 2a (M/W 9:50-11AM, F 9:40-10:40AM)
Office Hours: CMC 324, Tuesday 11:30AM-12:30PM/Friday 3PM-4:30PM, by request
Class Website: `https://jmhessel.github.io/CS322-SP19/`

Description:

Natural languages (e.g., Chinese, English, etc.) enable humans to communicate, but, for the better part of history, computers have been left out of the conversation. Enabling machines to understand language is the goal of Natural Language Processing (NLP), and achieving this goal (or coming close) offers immense promise in fields like human-computer interaction, computational social science, and medicine (among others). However — language understanding is quite difficult, as spoken/written language often encodes complex factors beyond literal meaning; in fact, NLP is so hard that it is sometimes called "AI Complete," i.e., if you could build a machine that truly understands language, you could build a machine with intellectual capacity equal to a human.

This course will cover several topics in Natural Language Processing, with a particular focus on statistical methods that learn patterns automatically from corpora (in contrast to methods that rely on hand-designed features and rules). Topics will include language modeling, supervised learning with bag-of-words inputs, lexical/vector semantics, feed-forward neural networks (for language modeling), and recurrent neural networks (for sequence tagging).

Textbook + Readings:

We will be using the 3rd edition of Jurafsky and Martin's "Speech and Language Processing" which is available online: `https://web.stanford.edu/~jurafsky/slp3/`. While the 3rd edition is still a work-in-progress (e.g., some chapters are missing, and sometimes citations are broken) it will serve as the primary source of our reading and problem sets. Additional readings will be linked on the course website.

Grading:

Grades will be assigned according to the following percentages

Homeworks	40%
Midterm Exam	20%
Final Project	40%
Participation	±5%

Note that this class will have a final project in lieu of a final exam.

Assignments: There will be either four or five homework assignments (I haven't quite decided yet). Each assignment will have a written portion and a coding portion. The written portion will consist of a number of short-answer questions, mostly drawn from J+M; the solutions to these questions should be typeset in latex, and submitted as a PDF. The coding portion will consist of implementing some of the concepts covered in class, and testing out how those ideas perform on real corpora; (even though its name may suggest otherwise) the coding portion of homework assignments may also involve a writeup. Because python3 has emerged as the *Lingua franca* NLP language, all assignments must be submitted in python3. Depending on the assignment, starting code may (or may not) be provided.

Collaboration is encouraged! In fact: you can work with up to one person for any of the homework assignments (it may be the same person each time, or a different person each time, etc.). If you do work in pairs, only one of you needs to submit the assignment, though please indicate who you worked with. Finally, if you used any external sources (e.g., stackoverflow posts, etc.) please indicate as such.

A few caveats: 1) if you work in groups larger than yourself (and your potential partner) the work you submit must be your own. This means that you should *not* work directly from notes taken at larger collaborative sessions, or copy code that is not yours. 2) If prompted, you should be able to verbally explain

any written or programming assignment you submit, i.e., even if your partner and you split up some of the work, you are both responsible for being able to explain all of the answers.

In summary, unless otherwise specified, I expect a zip file containing the following to be submitted for each assignment on Moodle:

written_writeup.pdf	your latex typeset answers to the written portion
coding_writeup.pdf	your latex typeset answers to the coding portion
code.zip	a zip file of python3 code + README with run instructions
collaboration.txt	who you worked with (if anyone)
sources.txt	list of sources you used

Assignments are due at 11:59PM the day indicated on the course website. Late assignments can receive a maximal 50% score if submitted within 24 hours of the due date. Beyond that, assignments will not be accepted. If you're going to have trouble meeting a deadline, let me know as soon as possible, and we'll work something out (though I reserve the right to be stringent!).

For some assignments, there may be two due dates: one for the written assignments, and one for the coding assignments. If you are turning in just the written assignment, you should turn in written_writeup.pdf, collaboration.txt, and sources.txt. If you are turning in just the coding assignment, you should turn in coding_writeup.pdf, code.zip, collaboration.txt, and sources.txt.

Here are some tips for the homework:

- ***Start early.*** Some of the assignments could take 10-15 hours, depending on your mathematical background! Spread over 1-2 weeks, this equates to about an hour of work per day. However — spread over a single day (e.g., the day before the assignment is due) this equates to a bad time. At least read-over the assignment the day it comes out!
- ***Don't be afraid to ask for help!*** Be it from me, from your classmates on piazza, or from other on-campus resources, it is okay to struggle with the problems! If you think you're falling behind or having other serious problems; please let me know early! I am on your team, and want to help you get as much out of this class as possible.

Group Project: Throughout the course, you will be working on a group research project with 3-5 other students. If you have particular people you'd like to work with, I will be circulating a partner preference form early in the term; if you don't submit a preference (or don't know anyone in the class yet), not to worry! I will assign you to a group.

Your project will be focused on one of several real NLP tasks that researchers are trying to solve. There are five deliverables for the group project (with due dates throughout the term), each comprising the following percent of the total grade for the project:

proposal	10%
progress report	10%
update presentation (5 minutes)	10%
final presentation	20%
writeup/final code	50%

Starting around Week 7-8 (depending on the number of groups) groups will present 5 minute updates at the start of class to let everyone know what they are working on and what progress they have made. The final presentations will span the final two days of the course.

More information about projects (e.g. potential topics, etc.) will be made available on the course website.
Midterm Exam: The midterm exam is on **May 13th;** it will cover all content covered in-class up to the point of the exam.

Support: This class will cover difficult concepts, and I expect that some of the assignments will be hard (especially if some of the mathematical concepts are new to you). In addition to office hours, there will be discussions on Piazza (e.g., questions about reading, questions about assignments, etc.). In addition to myself answering questions on this forum, **you are required to make at least one post each week.** This post can be a question, an answer to a question, a contribution to a discussion, or an idea you had inspired by the reading. I will be relying on Piazza to shape what topics we cover in class, e.g., if something isn't clear from the reading, we will be spending more time on it. So — if you are having trouble understanding, please do post on Piazza!

Participation: I expect everyone to attend every class, and engage in discussions (both online and off). Thus, part of this grade is determined from attendance/in-class participation, and the other half is based off of the online contributions on Piazza (see "Support"). If you're going to miss class, please let me know beforehand.

Writing Center: The Writing Center, located in 420 4th Libe, has peer writing consultants who can work with you during any stage of the writing process (brainstorming to final proofreading). Hours and more information can be found on the writing center website. You can reserve specific times for conferences in 420 4th Libe by using their online appointment system. Walk-ins are welcome, though writers with appointments have priority.

If you are a second language writer and believe you might benefit from working individually with a writing consultant on a regular basis this term, email Renata Fitzpatrick, Multilingual Writing Coordinator, call her at x5998, or stop by her office in 420D 4th Libe. She can arrange once- or twice-a-week meetings between you and a specific writing consultant throughout the term.

Academic Honesty: (from the Carleton handbook:) It is assumed that a student is the author of all course work (quizzes, problem sets, online contributions, tests, papers, lab work, etc.) that he/she submits, whether for a grade or not, and that the work has not been submitted for credit in another class without the instructor's permission. Images, ideas, data, audio clips, or phrases borrowed from others should be fully identified by standard procedures for making such acknowledgment. All permitted collaboration with others must still be acknowledged... At Carleton College, an act of academic dishonesty is therefore regarded as conflicting with the work and purpose of the entire College and not merely as a private matter between the student and an instructor; all cases involving such dishonesty are referred for appropriate action to the Academic Standing Committee (ASC) via the Associate Dean of Students or the Associate Dean of the College.

A personal note: I take academic honesty *extremely* seriously. I have zero tolerance for cheating of any sort because it is a waste of my time and yours. If you feel that you may have violated the academic honesty policy (accidentally or otherwise) it is better for everyone that you let me know. I should also emphasize that collaboration is strongly encouraged: if you cite your sources, write up your own solutions, and do not work from notes you took at collaborative sessions, collaboration can be extremely valuable! If you have any questions, talk to me as soon as possible.

Public Speaking Support: Speech coaching is a student-staffed resource designed to assist you with class presentations, comps talks, and other speech-related events. Your coach can assist you with speech and communication skills including clarity, organization, articulation, projection, body language, eye contact, and effective use of aids (e.g., notes, PowerPoint, Keynote, etc.). Depending on your goals, your coach can also work with you on the content of the presentation: organization, voice, clarity, and, ultimately, persuasive impact. For more information, visit `go.carleton.edu/speakeasy`.

Accommodations for Students with Disabilities: Carleton College is committed to providing equitable access to learning opportunities for all students. The Disability Services office (Burton Hall 03) is the campus office that collaborates with students who have disabilities to provide and/or arrange reasonable accommodations. If you have, or think you may have, a disability (e.g., mental health, attentional, learning, autism spectrum disorders, chronic health, traumatic brain injury and concussions, sensory, or physical), please contact Chris Dallager, Director of Disability Services, by calling 507-222-5250 or sending an email

to cdallager@carleton.edu to arrange a confidential discussion regarding equitable access and reasonable accommodations.

Assistive Technology: The Assistive Technology program brings together academic and technological resources to complement student classroom and computing needs, particularly in support of students with physical or learning disabilities. Accessibility features include text-to-speech (Kurzweil), speech-to-text (Dragon) software, and audio recording Smartpens. If you would like to know more, contact aztechs@carleton.edu or visit go.carleton.edu/aztech.

Acknowledgment: Some of the materials in this course (and even in this syllabus!) are borrowed with permission from Blake Howald.