

CS 6120/4120: Natural Language Processing

Time and Location: Mondays from 6:00 pm - 9:00 pm in Shillman Hall 415

Instructor: [Lu Wang](#), Office Rm 911, 177 Huntington Ave.

Staff and Office Hours:

- Prof. Lu Wang, Mondays 4:30pm - 5:30pm or by appointment, Rm 911, 177 Huntington Ave.
- Vishruth Krishna Prasad (Vish) (email: krishnaprasad.v@husky.neu.edu), Tuesdays 3-4pm, 132H Nightingale
- Apoorva Kasoju (email: kasoju.a@husky.neu.edu), Wednesdays 2-3pm, 132H Nightingale
- All regular office hours (OH) start from the week of Sep 17th, 2018. To attend OH at 177 Huntington Ave., you'll need to put down your name on Piazza beforehand (by 4pm each Monday), and then bring a photo ID (e.g. husky card) with you and check in at the front desk.

Discussion Forum: [Piazza](#), sign up at piazza.com/northeastern/fa2018/cs6120

Important Announcement

[9/5/2018] We will have a quiz with 20 **simple** True or False questions (relevant to probability, statistics, and linear algebra) in the first lecture (9/10/2018). This quiz will be graded, but will not be counted in your final score if you're enrolled in CS6120/CS4120. The purpose of this quiz is to indicate the expected background of students. 80% of the questions should be easy to answer. If you find yourself struggling with this quiz, it's possible that you need to catch up on the background or it may be preferable to take one or two preliminary courses. For students previously do not take any algorithm course (CS 5800 or CS 7800, see Prerequisites), an 80% or above is required to enroll in this course.

Course Description (and Syllabus)

This course aims to introduce fundamental tasks in natural language processing, and its recent advances based on machine learning algorithms (e.g., neural networks) and applications for interdisciplinary subjects (e.g., computational social science). The course materials are mostly delivered as lectures, and accompanied with reading materials. The students will be evaluated based on assignments, a **research-driven** course project, and an open-book final exam.

Please find the syllabus here: [\[Link\]](#)

Textbooks and Reference

- Main Textbook:
 - Dan Jurafsky and James H. Martin, "Speech and Language Processing, 2nd Edition", Prentice Hall, 2009.
 - Third edition draft is available at web.stanford.edu/~jurafsky/slp3/.
- Other Reference:

- Chris Manning and Hinrich Schutze, "Foundations of Statistical Natural Language Processing", MIT Press, 1999
- Since many natural language processing problems are driven by machine learning techniques nowadays, we also highly encourage you to read machine learning textbooks:
 - Christopher M. Bishop, "Pattern Recognition and Machine Learning", Springer, 2006.
 - Tom Mitchell, "Machine Learning", McGraw Hill, 1997.

Prerequisites

This course is designed for graduate students and senior undergraduate students majoring in computer science, linguistics, and other related areas. Students who take this course are expected to be able to write code in some programming languages (e.g., Python, Java, or C/C++) proficiently, and finish courses in algorithms (CS 5800 or CS 7800), multivariable calculus, probability, and statistics. Linear algebra is optional, but highly recommended.

Grading

Blackboard is used for electronic submission. Assignment or report turned in late will be charged 20 points (out of 100 points) off for each late day (i.e. 24 hours). Each student has a budget of 5 days throughout the semester before a late penalty is applied. You may want to use it wisely, e.g. save for emergencies. Each 24 hours or part thereof that a submission is late uses up one full late day. Late days are not applicable to final presentation. Each group member is charged with the same number of late days, if any, for their submission.

Grades will be determined based on two assignments, eight in-class tests, one course project, one open-book exam, and participation:

- Assignments (30%): two assignments, each of 15%
- Quiz (5%): eight quick in-class tests, each of 1%; three with lowest grades will be dropped, and NO make-up
- Project (25%): team of 2 to 3 students, proposal (3%), reports (7%+10%), final presentation (5%, with 1% as bonus if selected as best project by peer students)
- Exam (35%): open-book
- Participation (5%): classes (participating in-class discussions, etc), Piazza (answering questions, sharing notes, etc)

Assignments

- [\[Assignment 1\]](#)
- [\[Assignment 2\]](#)

Sample Writeups for Project

- [\[Project Proposal\]](#)
- [\[Project Progress Report\]](#)
- [\[Project Final Report\]](#)
- Sample projects from Stanford NLP course [\[link\]](#)

Schedule

September 10

- Topic: Introduction, Language Models
- Slides: [\[intro\]](#) [\[6pp version\]](#) [\[Language Model\]](#) [\[6pp version\]](#)
- Reading: Ch1, Ch4.1-4.9
- TODO: start thinking about projects and looking for teammates

September 17 (quiz 1)

- Topic: LM cont'd, Naive Bayes, Text Categorization and Evaluation, Part-of-Speech Tagging, Sequence Labeling, Hidden Markov Models
- Slides: [\[NB\]](#) [\[6pp version\]](#) [\[POS Tagging\]](#) [\[6pp version\]](#)
- Reading: Ch5.1-5.5, Ch6.1-6.5

September 24 (quiz 2)

- Topic: HMM cont'd, Word Sense Disambiguation
- Reading: Ch20

October 1 (quiz 3)

- Topic: Formal Grammars of English, Syntactic Parsing, Dependency Parsing
- Reading: Ch12.1-12.5
- Course project proposal due on Oct 1 at 11:59pm.

October 8 (NO CLASS, Columbus Day)

- Assignment 1 is due on Oct 10 at 11:59pm.

October 15 (quiz 4)

- Topic: Semantics, Vector-Space Lexical Semantics, Information Extraction
- Reading: Ch13.1-13.4.1, Ch20.7

October 22 (quiz 5)

- Topic: Neural Networks, Neural Language Models, Sentiment Analysis, Opinion Mining
- Reading: Ch22.1-22.2, Ch23

October 29 (quiz 6)

- Topic: Fairness and Bias, Course Project Feedback
- Reading: TBD
- Course project progress report due on Oct 31 at 11:59pm.

November 5 (quiz 7)

- Topic: Text Summarization, Discourse Analysis
- Reading: Ch23.3-23.7, Ch21
- Assignment 2 is due on Nov 7 at 11:59pm.

November 12 (NO CLASS, Veterans Day)**November 19 (quiz 8)**

- Topic: Question Answering, Dialogue Systems and Chatbots, Machine Translation
- Reading: Ch23.1-23.2, Ch24, Ch25.1-25.5

November 26

- Topic: Course Project Presentation

December 3

- Topic: final exam, time: 6pm-8pm, location: Shillman Hall 415
 - Project final report due on December 10.
-

Academic Integrity

This course follows the [Northeastern University Academic Integrity Policy](http://www.ccs.neu.edu/home/luwang/courses/cs6120_fa2018/cs6120_fa2018.html). All students in this course are expected to abide by the Academic Integrity Policy. Any work submitted by a student in this course for academic credit should be the student's own work. Collaborations are allowed only if explicitly permitted. Violations of the rules (e.g. cheating, fabrication, plagiarism) will be reported.