



Computer Science 601.465/665
Natural Language Processing
Fall 2019 (3 credits)

Instructor

Professor Kevin Duh, kevinduh@cs.jhu.edu, <http://cs.jhu.edu/~kevinduh>
Office: Stieff Building / 810 Wyman Park Drive
Office hours: by appointment, contact on Piazza

Teaching Assistants and Course Assistants

TA: Arya McCarthy (Head TA), Suzanna Sia
CA: Xiang Lisa Li, Coleman Haley, Devin Hill, Craig Guo
Office hours: Mon, 1000-1100; Tue, 1330-1430; Wed, 1500-1600; Thur, 1845-2000
; Fri, 1300-1400; Fri, 1600-1700

Meetings

Monday, Wednesday, Friday, 11:00–11:50 am, Shaffer 301

Textbook

[Jurafsky & Martin, Speech and Language Processing, 2nd ed.](#) (P98.J87 2009 in Science Ref section on C-Level)

Online Resources

Course website: <https://kevinduh.github.io/nlp-course/>
Please log in to Piazza for all materials related to this course.

Course Information

This course is an in-depth overview of techniques for processing human language. How should linguistic structure and meaning be represented? What algorithms can recover them from text? And crucially, how can we build statistical models to choose among the many legal answers? The course covers methods for trees (parsing and semantic interpretation), sequences (finite-state transduction such as morphology), and words (sense and phrase induction), with applications to practical engineering tasks such as information retrieval and extraction, text classification, part-of-speech tagging, speech recognition and machine translation. There are a number of structured but challenging programming assignments. [Applications] **Prerequisites:** 600/601.226 data structures

Course Goals

- Show sensitivity to linguistic phenomena and an ability to model them with formal grammars. [program outcomes (a),(c*), (i),(j)]
- Understand and carry out proper experimental methodology for training and evaluating empirical NLP systems. [program outcomes (b),(c),(e*)]
- Be able to manipulate probabilities, construct statistical models over strings and trees, and estimate parameters using supervised and unsupervised training methods. [program outcomes (a),(i),(j*)]
- Be able to design, implement, and analyze NLP algorithms. [program outcomes (a),(c),(d),(i*), (j)]

Program Outcomes

- (a) An ability to apply knowledge of computing and mathematics appropriate to the discipline.
- (b) An ability to analyze a problem, and identify and define the computing requirements appropriate to its solution.
- (c) An ability to design, implement, and evaluate a computer-based system, process, component, or program to meet desired needs.
- (d) An ability to function effectively on teams to accomplish a common goal.
- (e) An understanding of professional, ethical, legal, security, and social issues and responsibilities.
- (i) An ability to use current techniques, skills, and tools necessary for computing practice.
- (j) An ability to apply mathematical foundations, algorithmic principles, and computer science theory in the modeling and design of computer-based systems in a way that demonstrates comprehension of the tradeoffs involved in design choices.

Course Topics

The course is organized into 11 lecture modules (below) and 5 research talks.

1. Introduction: Applications, Modeling Grammaticality
2. Language Models
3. Text Classification
4. Linguistics 101
5. Tree Parsing
6. Neural Networks
7. Sequence Tagging
8. Topic Models
9. Finite State Transducers
10. Semantics
11. Structured Prediction

Course Expectations & Grading

- 6 homework assignments: 60% (10% each) → these are non-trivial: if you can do these proficiently, then you have mastered the material
- participation: 5% → in class and on Piazza
- midterm exam: 15% (October 16, in-class)
- final exam: 20%
- Late homework policy: It's important to get homework done on time so that you can follow the subsequent lectures. We understand that emergencies do occur, so you are allowed up to 10 late days throughout the term. They are only intended to cover situations where you would ordinarily ask for an extension. Rather than ask me, just use a late day: I don't want to be in the position of deciding which excuses are worthy and whose aren't. If you run out of late days, we'll have to give you zeroes. But it is still to your advantage to turn in all homework to get feedback.

Key Dates

Midterm Exam: Wednesday, October 16, 11:00-11:50AM

Final Exam: Thursday, December 19, 9 AM-12 PM

Assignments & Readings

See course website (<https://kevinduh.github.io/nlp-course/lectures>) for assignments and suggested readings

Ethics

The strength of the university depends on academic and personal integrity. In this course, you must be honest and truthful, abiding by the *Computer Science Academic Integrity Policy*:

Cheating is wrong. Cheating hurts our community by undermining academic integrity, creating mistrust, and fostering unfair competition. The university will punish cheaters with failure on an assignment, failure in a course, permanent transcript notation, suspension, and/or expulsion. Offenses may be reported to medical, law or other professional or graduate schools when a cheater applies.

Violations can include cheating on exams, plagiarism, reuse of assignments without permission, improper use of the Internet and electronic devices, unauthorized collaboration, alteration of graded assignments, forgery and falsification, lying, facilitating academic dishonesty, and unfair competition. Ignorance of these rules is not an excuse.

Academic honesty is required in all work you submit to be graded. Except where the instructor specifies group work, you must solve all homework and programming assignments without the help of others. For example, you must not look at anyone else's solutions (including program code) to your homework

problems. However, you may discuss assignment specifications (not solutions) with others to be sure you understand what is required by the assignment.

If your instructor permits using fragments of source code from outside sources, such as your textbook or on-line resources, you must properly cite the source. Not citing it constitutes plagiarism. Similarly, your group projects must list everyone who participated.

Falsifying program output or results is prohibited.

Your instructor is free to override parts of this policy for particular assignments. To protect yourself: (1) Ask the instructor if you are not sure what is permissible. (2) Seek help from the instructor, TA or CAs, as you are always encouraged to do, rather than from other students. (3) Cite any questionable sources of help you may have received.

On every exam, you will sign the following pledge: "I agree to complete this exam without unauthorized assistance from any person, materials or device. [Signed and dated]". Your course instructors will let you know where to find copies of old exams, if they are available.

Report any violations you witness to the instructor.

You can find more information about university misconduct policies on the web at these sites:

- For undergraduates: <http://e-catalog.jhu.edu/undergrad-students/student-life-policies/>
- For graduate students: <http://e-catalog.jhu.edu/grad-students/graduate-specific-policies/>

Personal Wellbeing

- If you are sick, in particular with an illness that may be contagious, notify me by email but do not come to class. Rather, visit the Health and Wellness: 1 East 31 Street, 410-516-8270. See also <http://studentaffairs.jhu.edu/student-life/support-and-assistance/absences-from-class/illness-note-policy/>
- All students with disabilities who require accommodations for this course should contact me at their earliest convenience to discuss their specific needs. If you have a documented disability, you must be registered with the JHU Office for Student Disability Services (385 Garland Hall; 410-516-4720; <http://web.jhu.edu/disabilities/>) to receive accommodations.
- If you are struggling with anxiety, stress, depression or other mental health related concerns, please consider visiting the JHU Counseling Center. If you are concerned about a friend, please encourage that person to seek out our services. The Counseling Center is located at 3003 North Charles Street in Suite S-200 and

can be reached at 410-516-8278 and online at <http://studentaffairs.jhu.edu/counselingcenter/>

Classroom Climate

I am committed to creating a classroom environment that values the diversity of experiences and perspectives that all students bring. Everyone here has the right to be treated with dignity and respect. I believe fostering an inclusive climate is important because research and my experience show that students who interact with peers who are different from themselves learn new things and experience tangible educational outcomes. Please join me in creating a welcoming and vibrant classroom climate. Note that you should expect to be challenged intellectually by me, the TAs, and your peers, and at times this may feel uncomfortable. Indeed, it can be helpful to be pushed sometimes in order to learn and grow. But at no time in this learning process should someone be singled out or treated unequally on the basis of any seen or unseen part of their identity.

If you ever have concerns in this course about harassment, discrimination, or any unequal treatment, or if you seek accommodations or resources, I invite you to share directly with me or the TAs. I promise that we will take your communication seriously and to seek mutually acceptable resolutions and accommodations. Reporting will never impact your course grade. You may also share concerns with the department chair, the director of undergraduate studies, the Assistant Dean for Diversity and Inclusion (Darlene Saporu, dsaporu@jhu.edu), or the Office of Institutional Equity (oie@jhu.edu). In handling reports, people will protect your privacy as much as possible, but faculty and staff are required to officially report information for some cases (e.g. sexual harassment).