JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Computer Science EN.601.465 / EN.601.665
Natural Language Processing
Fall, 2020 (4 credits)

Instructor

Professor Jason Eisner <jason@cs.jhu.edu>
Office hours: At flipped-classroom sessions, or by appointment
Office Zoom link: https://wse.zoom.us/my/jeisner1

Teaching Assistants

Head TA: Arya McCarthy <arya@jhu.edu>
TA: Alexandra DeLucia <aadelucia@jhu.edu>
CA: Yingqi (Halley) Ding <yding50@jhu.edu>
CA: Haoyuan Chen <chaoyua1@jhu.edu>
CA: Ayush Dalmia <adalmia1@jhu.edu>
CA: Chenyu (Heidi) Zhang <czhan105@jhu.edu>
CA: Jinhan Zhang <jzhan205@jhu.edu>

Office hours: TBA. See the class website, http://cs.jhu.edu/~jason/465.

You can reach all the course staff at once via a private post on Piazza, or via cs465-staff@cs.jhu.edu .

Meetings

Monday, Wednesday, Friday, 3:00–4:15 pm
Class Zoom link: https://wse.zoom.us/j/95850009212
Class Zoom passcode: Available from https://meetinginfo.jhu.edu/. Will also be emailed to registered and waitlisted students before the first class.

This year we'll often "flip the classroom," to make the best interactive use of our precious synchronous class meeting times. Thus, many of the class meetings will be used for Q&A, discussion, enrichment, and collaborative-problem solving. You'll be expected to watch lecture videos ahead of time, which will be announced on Piazza and posted on the class website.

Textbooks

This course does not follow any textbook. The material is covered by the video lectures and the readings that are associated with the homework projects, supplemented by in-class discussion.

However, the following textbooks are at an appropriate level. They are usually a good reference to get another written perspective on the material. The website will suggest optional readings from time to time.

- Dan Jurafsky & James Martin (2020), *Speech and Language Processing* (3rd ed.). A draft of the 3rd edition is online here.
- Jacob Eisenstein (2019), Introduction to Natural Language Processing.
- Brian Roark & Richard Sproat (2007), *Computational Approaches to Morphology and Syntax*.

- Chris Manning & Hinrich Schütze (1999), *Foundations of Statistical Natural Language Processing*. A full PDF is available via the JHU library at this link.

Online Resources

Everything you need will be linked from the class homepage, `http://cs.jhu.edu/~jason/465`.

Make sure to sign up for the Piazza site, and make sure to watch the assigned lecture videos. (And of course, come to class, do the homeworks, take the exams.)

Course Information

- **Catalog description:** 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**
  - Data Structures (601.226)
  - Partial derivatives

  The class aims to be fairly self-contained and teach you everything else you need, presenting it from an NLP perspective. That includes relevant aspects of automata (600.271), probability (553.420/620 or 553.310/311), and machine learning (601.475/675, 601.482/682, ...). So those courses are not formal prerequisites. That said, it may be helpful to have had prior exposure to those concepts.
- **Elective** (Applications)

Course Goals

This course is designed to introduce you to some of the problems and methods of natural language processing, and their relation to linguistics and statistics. At the end you should agree (I hope!) that language is subtle and interesting; feel some ownership over some of NLP's formal and statistical techniques; and be able to understand research papers in the field. (Caveat: understanding recent research papers may sometimes additional background in machine learning.)

Not to mention all this.

Specific Outcomes for this course:

- Show sensitivity to linguistic phenomena and an ability to model them with formal grammars.
- Understand and carry out proper experimental methodology for training and evaluating empirical NLP systems.
- Be able to manipulate probabilities, construct statistical models over strings and trees, and estimate parameters using supervised and unsupervised training methods.
- Be able to design, implement, and analyze NLP algorithms.

This course will address the following CSAB Criterion 3 Student Outcomes. Graduates of the program will have an ability to:

1. Analyze a complex computing problem and to apply principles of computing and other relevant disciplines to identify solutions.

2. Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline.
5. Function effectively as a member or leader of a team engaged in activities appropriate to the program's discipline.

Course Topics

Please see the class website for a list of topics and resources: http://cs.jhu.edu/~jason/465.

Course Expectations & Grading

We may revisit this breakdown as we get a sense of what is best for our online learning situation. Any changes will be announced.

- **≈ 5% participation**
- **50% homeworks** (see lateness policy)
- **15% midterm exam**
- **30% final exam**

Participation includes your contributions to an interesting and useful class discussion, whether synchronously during class sessions, or via public posts or replies on our Piazza discussion website. This includes asking questions, of course. We've found that participation has a substantial effect on the final grade.

Homeworks will be submitted via Gradescope. We expect to have about 7 homework projects, which focus on different skills you'll need to do NLP:

(1) writing linguistic grammars
(2) manipulating probability formulas
(3) implementing and properly evaluating supervised models
(4) designing and implementing combinatorial algorithms
(5) working with formal meaning representations
(6) training unsupervised models for structured prediction
(7) constructing models algebraically

We expect to have two exams—a midterm exam and a final exam. These will have some challenging questions. In some of the class sessions, you'll get to work in small groups to solve interesting problems from past exams.

Key Dates

Please see the class website: http://cs.jhu.edu/~jason/465.

Assignments & Readings

Please see the class website: http://cs.jhu.edu/~jason/465.

Since emergencies sometimes arise, I'll allow you a *total* of up to 10 late days during the term. One day = 24 consecutive hours. (Thus, if you are 12 hours late, that counts as 1 whole day.) The full policy and advice on how to use it are at http://cs.jhu.edu/~jason/465/late-policy.html.

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.

In this class (NLP), some homeworks will indicate that they do allow collaboration. For a collaborative homework, you are expected to do the work *together*, not divide it up: if you didn't work on a question, you don't deserve credit for it! Your solutions should emerge from collaborative discussions with the whole group present (whatever "present" means in online learning).

In this class (NLP), practice exam problems will be provided for you. You should not otherwise make use of homeworks or exams from previous years.

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, please notify me by email so that we can make appropriate accommodations should this affect your ability to attend class, complete assignments, or participate in assessments. The Student Health and Wellness Center is open and operational for primary care needs. If you would like to speak with a medical provider, please call 410-516-8270, and staff will determine an appropriate course of action based on your geographic location, presenting symptoms, and insurance needs. Telemedicine visits are available only to people currently in Maryland. See also https://studentaffairs.jhu.edu/student-life/student-outreach-support/absences-from-class/illness-note-policy/.
- The Johns Hopkins COVID-19 Call Center (JHCCC), which can be reached at 833-546-7546 seven days a week from 7 a.m. to 7 p.m., supports all JHU students, faculty, and staff experiencing COVID-19 symptoms. Primarily intended for those currently within driving distance of Baltimore, the JHCCC will evaluate your symptoms, order testing if needed, and conduct contact investigation for those affiliates who test positive. More information on the JHCCC and testing is on JHU's coronavirus information website.
- 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 (Shaffer 101; 410-516-4720; http://web.jhu.edu/disabilities/) to receive accommodations.
- Students who are struggling with anxiety, stress, depression or other mental health related concerns, please consider connecting with resources through the JHU Counseling Center. The Counseling Center will be providing services remotely to protect the health of students, staff, and communities. Please reach out to get connected and learn about service options based on where you are living this fall at 410-516-8278 and online at http://studentaffairs.jhu.edu/counselingcenter/.
- Student Outreach & Support will be fully operational (virtually) to help support students. Students can self-refer or refer a friend who may need extra support or help getting connected to resources. To connect with SOS, please email deanofstudents@jhu.edu, call 410-516-7857, or students can schedule to meet with a Case Manager by visiting the Student Outreach & Support website and follow "Schedule an Appointment."

Classroom Climate

As your instructor, 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 Head (Randal Burns, randal@cs.jhu.edu), the Director of Undergraduate Studies (Joanne Selinski, joanne@cs.jhu.edu), 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).

Family Accommodations Policy

You are welcome to bring a family member to class on occasional days when your responsibilities require it (for example, if emergency childcare is unavailable, or for health needs of a relative). In fact, you may see my children in class on days when their school is closed. Please be sensitive to the classroom environment, and if your family member becomes uncomfortably disruptive, you may leave the classroom and return as needed.

University Policy on Incompletes

The university recognizes that the Fall 2020 semester is surrounded with uncertainty and many students may find themselves in unexpected situations where study is difficult if not impossible. Students who are confronted with extraordinary circumstances that interfere with their ability perform their academic work may request an incomplete grade from the instructor. While approval of such a request is not automatic, it is expected that faculty will make every effort to accommodate students dealing with illness in the family and other pandemic-related hardships. The instructor and student must establish a timetable for submitting the unfinished work with a final deadline no later than the end of the third week of the Spring 2021 semester (February 12, 2021). Exceptions to this deadline require a petition from the instructor to the student's academic advising office by February 12, 2021. When entering an Incomplete grade in SIS, faculty must include a reversion grade which represents the grade the student will receive if s/he does not complete the missing work by the agreed-upon deadline.

Deadlines for Adding, Dropping and Withdrawing from Courses

Students may add a course up to September 11, 2020. They may drop courses up to October 12, 2020 provided they remain registered for a minimum of 12 credits. Between October 12 and November 13, 2020, a student may withdraw from a course with a W on their academic record. A record of the course will remain on the academic record with a W appearing in the grade column to indicate that the student registered and then withdrew from the course. For more information on these and other academic policies, see https://e-catalogue.jhu.edu/engineering/full-time-residential-programs/undergraduate-policies/academic-policies/grading-policies/

The Office of Academic Support at JHU

All programs are free to students. Please see below for specifics:

- PILOT Learning—Peer-Led Team Learning
  - Students are organized into small study teams who meet weekly to collaborate on faculty-developed problems-sets. Students work together as a team to solve problems.
  - A trained student leader acts as captain and facilitates the weekly meetings using various strategies to foster a collaborative learning environment.
  - Registration opens on August 31st at 9pm EST; registration will remain open throughout the semester if space allows.
  - Contact: Ariane Kelly <ariane.kelly@jhu.edu>
  - Instagram: @jhupilot
- Learning Den Tutoring Program - Small Group Tutoring
  - Small group, tailored tutoring of 4 students or less which is headed by one tutor. Visit the website (above) to access zoom links for drop-in sessions
  - Tutors can assist with but are not limited to:
    - Review and strengthening of subject-specific material knowledge
    - Assist with homework-like problems
    - Course-specific study skills and exam preparation
    - Contact: Kaitlin Quigley <quigley@jhu.edu>

        * Instagram: @jhulearningden
- The Study Consulting Program
  - Students work one-on-one with a study consultant to set academic goals and develop customized strategies for success. Areas addressed include but are not limited to:
    * Time management
    * Note taking and test preparation
    * Mastering large amounts of information
  - Contact: Dr. Sharleen Argamaso <sharleen.argamaso@jhu.edu>
  - Instagram: @jhustudyconsulting
- The Writing Center
  - Undergraduate and graduate students in KSAS/Whiting School of Engineering can schedule 50-min sessions with a Writing Center tutor to look over a draft of written work (up to 10 pages) or a personal statement for graduate study
  - Contact: Robert Tinkle <rtinkle1@jhu.edu>
  - Web Address: https://krieger.jhu.edu/writingcenter/