JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Computer Science EN.601.465 / EN.601.665
Natural Language Processing
Fall 2023 (4 credits)

Instructor

Professor Jason Eisner (he/him) <jason@cs.jhu.edu>
Office hours: After class in the classroom, or by appointment
Office location: Hackerman 324C, or https://wse.zoom.us/my/jeisner1

Teaching Assistants

Head TA: Zhichu (Brian) Lu <zlu39@jhu.edu>
TA: Li (Leo) Du <ldu10@jhu.edu>
CA: Xinyuan (Henry) Li <xli257@exchange.johnshopkins.edu>
CA: Kevin Kim <kkim170@jhu.edu>
CA: Sungwon Kim <skim434@jhu.edu>
CA: Camden Shultz <cshultz3@jhu.edu>
CA: Shaobo Liang <sliang24@jhu.edu>

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

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

Meetings

Mon/Wed/Fri 3:00–4:15 pm, Hodson 110
Tue 6:00–7:30 pm, Hodson 210 (recitation)

This class is in the "flexible time slot" MWF 3-4:30. Please keep the entire slot open. Class will usually end at 4, followed by office hours in the classroom from 4-4:30 (stick around to get your money's worth!). However, class will sometimes run till 4:15 in order to keep up with the syllabus. I'll try to give advance notice of these "long classes," which among other things make up for no-class days when I'm out of town.

We also run a once-per-week recitation led by the prof or the TAs. This session is normally held in the Tuesday slot, and will usually focus on solving problems together. It is meant as an efficient and cooperative way to study for an hour: it reinforces the past week's class material without adding to your homework load. Thanks to recitation, you won't be startled by the exam style—the discussion problems are taken from past exams and are generally interesting.

Occasionally we may swap lecture and recitation slots.

We'll occasionally ask you to watch some video lessons outside of class. (Similar to a required reading. These video lessons were created during a previous attempt to "flip the classroom" for Covid virtual learning.)

If you miss a lecture, please email <eisner@jhu.edu> to get access to a video recording. Watch it promptly so you can follow subsequent lectures. Lecture recordings are also available for review upon request, and will be made available before each exam.

If you click the CC button, automatically generated closed captions will show up on the video. You can also view the full transcript and use it to navigate to a specific part of the video. (The video lessons have manual captions. The lecture recordings have automatic captions; these aren't perfect, but they might help you understand my words better.)

Recitations will generally not be recorded, but the materials from recitation are available on the class website.

Textbooks

This course does not follow any textbook. The material is covered by the lectures and videos and the reading handouts 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*. PDFs of the full book and the individual chapters are 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.

Go there now! 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)
  - Python
  - Basic familiarity with partial derivatives, matrix multiplication, and probabilities

  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 require additional background in machine learning.)

In the end, I hope that NLP, like all good courses, stretches your mind and leaves you with new ways of thinking. 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

Your work will be weighted as follows (although we are considering a fancier scheme):

- **50% homeworks**, equally weighted (see lateness policy)
- **15% midterm exam**
- **30% final exam**
- **$\approx$ 5% participation**

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 neural models for structured prediction
(7) constructing models algebraically **or** extracting answers from large language models

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

Since some questions are hard or are graded harshly, we use a curve to pull the grades back up. If your grade was already high in absolute terms, it will stay high (so if everyone gets 100, everyone gets an A+, not a C).

Extra credit is added only after the curve is determined. Thus, extra credit is not required to do well, although it can help make up for low grades elsewhere.

Late Homework Policy

Since emergencies sometimes arise, I'll grant extensions on request, in multiples of 24 hours, up to a *total* of up to 10 "late days" during the term. The full policy and advice on how to use it are at http://cs.jhu.edu/~jason/465/late-policy.html.

Audit Policy

If you register for the course as an auditor (or switch to this status), please let me know that you are auditing. You're still expected to attend the lectures and recitations and to participate in them like any other student. (That's what it means for the course to appear on your transcript with an AU grade.) If you are not able to sustain this level of commitment, you should drop the class.

You are also welcome to do any homeworks or exams that are useful to your learning. We will grade them and give you the usual feedback.

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.

The material below is standard for CS Department syllabi as of Fall 2023, except for portions in this color.

Integrity

The strength of the university depends on academic and personal integrity. You must be honest and truthful in this course, abiding by the CS department's Academic Integrity Code:

> 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 participating.

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 also contact the Office of the Dean of Student Life at 410-516-8208 or via email at <studentconduct@jhu.edu>.

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 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 Director of Graduate Studies (Scott Smith, <scott@cs.jhu.edu>), the WSE 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).

Personal Wellbeing

Illness. If you are sick, in particular with an illness that may be contagious, please notify me by email but do not come to class. I will provide you with a live Zoom link and/or a video recording of the lecture. We can also make arrangements to take exams remotely if necessary.

You may also contact Student Outreach and Support at <studentoutreach@jhu.edu> for help managing your coursework during an illness, with medical leave of absence being an option for a longer illness. For your medical needs, see the Student Health and Wellness Center.

Mental Health. JHU has several resources to support students. Many students struggle at times with stress, anxiety, or depression. The Counseling Center has many resources available to students. In an crisis, call the Behavioral Health Crisis Support Team at 410-516-4600, or call 911. For less urgent matters, undergraduates should contact Student Outreach and Support (410-516-7857, <studentoutreach@jhu.edu>), and graduates should contact the relevant dean: Megan Barrett (WSE) or Renee Eastwood (KSAS).

Disabilities. Johns Hopkins University values diversity and inclusion. We are committed to providing welcoming, equitable, and accessible educational experiences for all students. Students with disabilities (including those with psychological conditions, medical conditions, and temporary disabilities) can request accommodations for this course by providing an Accommodation Letter issued by Student Disability Services (SDS). For further information or to start the process of requesting accommodations, please contact Student Disability Services (Shaffer Hall 101, 410-516-4720, <studentdisabilityservices@jhu.edu>). Please request accommodations for this course as early as possible to provide time for effective communication and arrangements.

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). Please be sensitive to the classroom environment, and if your family member becomes uncomfortably disruptive, you may leave the classroom and return as needed.