Natural Language Processing (NLP) is the study of the computational treatment of natural language—the words we use everyday. NLP draws on research in Linguistics, Theoretical Computer Science, Mathematics and Statistics, Psychology, Artificial Intelligence, Machine Learning, with broad applications to many other fields. This course introduces students to a variety of NLP methods available for reasoning about text in computational systems. We will focus on major algorithms used in NLP for various applications (e.g., part-of-speech tagging, parsing, machine translation), on the linguistic phenomena those algorithms attempt to model, and on the people who interpret and utter the language. Students will implement a variety of algorithms for different linguistic aspects (e.g., syntax, semantics) and also understand the creation of ground truth data through linguistically annotating data on which those algorithms depend.

Broadly speaking, the course has six major parts which are interleaved through most discussions:

1. Linguistic, mathematical, and computational background
2. Computational models of morphology, syntax, semantics, discourse, pragmatics
3. Core NLP technology: parsing, part of speech tagging, text generation, semantic analysis, etc.
4. Applications: text classification, sentiment analysis, text summarization, question answering, machine translation, information extraction, etc.
5. Human and social factors in language understanding and generation, including sociolinguistics, style, and data annotation
6. Deep neural networks and related techniques, such as autoencoders, recurrent networks, convolutional networks, transformers, attention, etc.
This NLP course has the following major learning goals:

1. Learn the basic principles, algorithms, and theoretical issues underlying natural language processing
2. Learn computational techniques and tools used to develop practical, robust systems that can communicate with users in one or more
3. Gain insight into many open research problems in natural language
4. Develop advanced software development skills for complex technical problems
5. Become familiar with recent deep learning frameworks

TEXTBOOKS AND OTHER MATERIALS:

The syllabus includes five textbooks, most of which are available online for free through the university. The inclusion of multiple textbooks is intended to allow students to have a complementary second (or third) description of an algorithm or topic. In general, we will use Dan Jurafsky and James Martin, Speech and Language Processing (3rd ed. draft) for most class activities and refer to the other materials only for specific topics; however students are encouraged to read other sources based on their interests.

- [SLP3] Dan Jurafsky and James Martin, Speech and Language Processing (3rd ed. draft) [Available here].
- [SLP2] Dan Jurafsky and James Martin, Speech and Language Processing (2nd ed., 2009)

In addition to the textbooks, each week will have an associated list of recent research papers, which are used five times during the semester for reading responses that are submitted as homework

GRADING:

34% - Homework Assignments
- 2% Intro assignment (Regular expressions)
- 8% Text Classification
- 8% Word Vectors
- 8% Parsing
• 8% Deep Learning
20% - Exams
  • 20% Midterm
40% - Course project
  • 5% Project Proposal
  • 10% Halfway Update
  • 5% Blog Post
  • 20% Final report
6% - Readings
  • 1% for each weekly paper readings (must submit 6; can’t submit more than 6)

Grading follows the traditional scale.
97-100: A+
93-97: A
90-93: A-
87-90: B+
83-87: B
80-83: B-
77-80: C+
73-77: C
70-73: C-
67-70: D+
63-67: D
60-63: D-
Below 60: F

There is no curve for this class. That said, usually the majority of students have received As. If you are concerned about your grade, please come see the instructors early (not the last week of classes) while there is still time to improve your performance. No extra credit or make-up assignments are allowed (e.g., no retroactive work after the end of the semester to bump up grade). The instructor reserves the right to raise letter grades at the borderline between two on the basis of extra circumstances (but will never lower grades).

CLASS COMMUNICATION AND PIAZZA POLICY:
All official course announcements will be sent through Canvas. Should any scheduling change (e.g., setting the date for the midterm), you will be responsible for monitoring this email. We will also announce the same things at the beginning of class. Please be on time to hear course announcements.

All other questions will use Piazza, accessed through Canvas. Questions will be replied to as quickly as possible, though we only guarantee a response time of 24 hours Monday-Friday and 48 hours on weekends. No questions should be submitted by email, unless it is concerning a personal matter.
Code should be kept to a minimum for questions on Piazza, or (preferably) avoided entirely. You are encouraged to discuss things at a high level or seek in-person guidance for code debugging. For questions that require more than 10 sentences to explain, we encourage you to ask in the office hours.

What does Participation Mean in the Zoom era?

Normally, participation is a bit easier to measure in class, but given that we're doing everything virtually and for some folks, asynchronous, we'll try measuring participation broadly to allow everyone to participate, even if they can't make it to class or don't want to speak up. Following are a few ideas on how to participate. We'll also have a running Piazza thread where we can pitch ideas for participation.

Participation ideas and suggestions

- Holding a multi-group work time for course projects
- Leading Break-time Calisthenics
- Sharing new class-related papers and resources on Piazza
- Asking questions in class
- Asking questions on Piazza
- Answering questions on Piazza (very helpful!)

Pandemic Time Considerations

The Fall 2020 semester is certainly going to be an unusual one. I will do all I can to help create a positive learning environment that works well regardless of location, bandwidth, and other factors in my control. However, the instructor fully recognizes that health-related and personal life events can (and often should) take priority over course work in some circumstances. If you find yourself in such circumstances, please reach out and we'll find a way for you to complete the course. I will typically not require extensive evidence of this but will ask for some evidence that is not of a sensitive nature (we will discuss this on a case by case basis). If something has become chronic or systemic, please do reach out early and we'll work together. If something is urgent, go take care of it now and we'll discuss later and work together on it; life is more than just this class!

I also recognize that this semester will likely be more stressful than past, so I encourage anyone experiencing such feelings to reach out to the CAPS folks (details below).
Where and How We Meet (with Backup Plans)

This class is fully virtual in the W21 semester. Class will be held synchronously at the official meeting time. The official Zoom link is on Canvas to join. The lecture will be interactive and students are encouraged to attend. In addition, the lecture will be recorded and made available within 24 hours after the lecture period for students who are unable to attend.

If during the lecture, the professor’s network connection drops or his video freezes, don’t give up on the meeting! Hang out for at least 5 minutes while he reboots his computer or switches to his cell phone for Internet connectivity. Thanks!

If in the event that virtual class disaster ensues, your instructor will record whatever remaining part of the lecture is left and upload this video to Canvas for asynchronous viewing.

How you are encouraged to use Zoom in class

- Have your camera on as much as possible during class meetings (but mute your microphone as much as possible, too)
- I strongly encourage you to keep your camera on for community-building purposes (but will not require it).
- Check out your background when you are on camera. Remove or cover items you don't want others to see
- Set up a profile photo for your account (for when your camera is off)
- Pets are welcomed! 🐱
- Test your audio and video - use headphones if necessary
- Ask questions! Audio questions are preferred but chat questions are great too
  - If you see a question in the chat go by and I haven’t gotten to it, feel free to let me know (even if it’s not your question!). Sometimes I will miss these so your help is much appreciated.

Other general rules for Zoom use in this class:

- You must always use your UMich credentials/account to access this course. If you have another Zoom account, it will not work
  - See this page for information on how to log into your U-M Zoom account (Links to an external site.)
  - If you have already used your UM email to create a non-UM Zoom account, please see this page for information on how to migrate your account. (Links to an external site.) (Note that this can take several hours, so please do it well before the first class meeting.)
- Only connect using a Zoom client, preferably on your laptop/desktop (better than phone or tablet). NEVER on a web browser
- Please keep your Zoom client updated to get the newest features!
● Change your display name in Wolverine Access (“Preferred” name in “Campus Personal Information”)

What to do if you are on a low-bandwidth connection?

● Leave video off when you don't need it
● Turn off HD video (in the Zoom app video settings)
● When you screen share, only do so for as long as necessary
● Use collaborative (e.g., Google) documents rather than screen sharing
● Mute your audio when not speaking (do this anyway!)

To improve your overall Zoom performance, consider asking others at your location to limit their high-bandwidth activities while you need to be online for a course or required meeting.

Avoid running other data-intensive programs during Zoom meetings, such as streaming video or music or other web sites with dynamic content. (Why are you watching streaming video during class?)

(These tips come from our colleagues at Cornell. See their doc for more details)

WEEKLY SYLLABUS:

Each week will cover a set of related topics and have associated readings from different sources. The following is a week-by-week guide and are subject to change based on course scheduling. Any updates will be posted to Canvas.

Week 1: January 20th

● Topics: Welcome to NLP
  ○ Introduction
  ○ Regular Expressions
  ○ Text Classification (Part 1): Naive Bayes

● Readings:
  ○ SLP2 Chapter 1
  ○ SLP3 Section 2.1, SLP3 Chapter 6
  ○ MS Chapter 2 and 3

Week 2: January 27th

● Topics: Supervised Text Classification
  ○ Text Classification (Part 2)
  ○ Logistic Regression
Neural Networks

Readings:
- SLP3 Chapters 7
- MS Chapter 16
- (Optional) SLP3 Chapter 18
- (Optional) G Chapter 2 (very math intensive)

Week 3: Feb 3rd

Topics: Language Models
- Language Models
- Recurrent Neural Networks

Readings:
- SLP3 Chapters 4, 8
- MS Chapter 6
- G Chapter 9, 14G

Week 4: Feb 10th

Topics: Word Vectors
- Latent Semantic Analysis
- Word Vectors (word2vec + GloVe)
- Morphology

Readings:
- SLP3 Chapters 15, 16
- G Chapter 10

Week 5: February 17th

Topics: Sequence Labeling
- Hidden Markov Models
- Part of Speech Tagging
- LSTM Networks

Readings:
- SLP3 Chapters 9, 10
- G Chapter 15
- MS Chapter 9 and 10
- (optional: G Chapter 19)

Week 6: February 24th (No class; Spring-Well-being Break)

Topics: ————
Week 7: March 3rd

- **Topics:** Syntax
  - Context-free parsing
  - Dependency Parsing
- **Readings:**
  - SLP3 Chapters 11-14
  - MS Chapters 11 and 12

Week 8: March 10th

- **Topics:** Unsupervised Learning
  - Brown Clustering
  - EM Algorithm
  - Topic Modeling
- **Readings:**
  - SLP3 Section 16.4 (Brown clustering)
  - MS Section 14.2

Week 9: March 17th

- **Topics:** Finishing Topic Modeling + new Neural Networks
  - Gibbs Sampling
  - Elmo, Bert, GPT-2
  - Machine translation
- **Readings:**
  - SLP2 Chapter 25
Week 10: March 24th

- **Topics:** Learning From Text
  - Information Extraction
  - Natural Language Inference
  - Question Answering
- **Readings:**
  - SLP3 Chapters 21, 28
  - SLP2 Chapter 22

Week 11: March 31st

- **Topics:** Discourse
  - Chat bots
  - Coreference Resolution
- **Readings:**
  - SLP3 Chapter 29, 30

Week 12: April 7th

- **Topics:** Lexical Semantics
  - Word Sense Disambiguation
  - Entity Linking
  - Semantic Roles + Frames
  - Diachronic Change
- **Readings:**
  - SLP3 Chapter 17, 22
  - MS Chapter 7

Week 13: April 14st

- **Topics:** Human Understanding of Language
  - Computational Sociolinguistics
  - Digital Humanities
  - Social NLP
- **Readings:**
  - All on Canvas
Week 14: April 21st

- **Topics:** Data Acquisition
  - Annotation
  - Crowdsourcing
  - Construction of Truth
  - Hypothesis Testing
  - Algorithmic Bias

- **Readings:**
  - PS chapter 6

**PREREQUISITES:**

SI 330 or 507/508 or EECS 281 or PhD Graduate Standing

Students should also have a good understanding of the following topics, though they can be learned throughout the duration of the course:

- Linear algebra: vectors and matrices
- Probabilities: random variables, discrete and continuous distributions, Bayes’ theorem
- Programming: using a UNIX environment

You’ll need to know how to write some programs on your own from scratch and how to extend an already-developed code-base to finish your assignment.

**ASSIGNMENTS:**

Students will have a programming-based assignment on a roughly bi-weekly basis. Assignments will be done in the python programming language, though students are allowed to use other languages with instructor permission. Python is strongly preferred for the NLP course due to the availability of deep learning libraries and other NLP-related packages. However, Java and R are also possible.
The full list of assignments and deliverables is as follows:

<table>
<thead>
<tr>
<th>Week</th>
<th>Assigned</th>
<th>Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HW0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>HW1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>HW0</td>
</tr>
<tr>
<td>4</td>
<td>HW2</td>
<td>HW1</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Project Proposal</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Relaxing (Wellness Break)</td>
<td>Relaxing</td>
</tr>
<tr>
<td>8</td>
<td>HW3</td>
<td>HW2</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>Project Update</td>
</tr>
<tr>
<td>11</td>
<td>HW4</td>
<td>HW3</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>Midterm</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>HW4</td>
</tr>
<tr>
<td>Finals Week</td>
<td></td>
<td>Project Report + Blog Post</td>
</tr>
</tbody>
</table>

**READINGS:**

Students will submit weekly responses to the readings. The first is a reading response to one of the research papers for that week. In this response, a student summarizes the paper and then comments on and critiques the paper’s contributions. The responses are intended to (i) teach critical reading of academic papers (ii) expose students to the latest advancements in the field, and (iii) provide a model of the research process students can use in their own course projects. A detailed template for students to follow in these responses is provided through Canvas. All responses are expected to follow this template.
Students are expected to submit 8 of each response throughout the semester, for 1% of their grade each. Responses are graded as either a 0 or 1, with no regrades. The primary reasons for getting a 0 are not meeting the length requirement (it’s one page) or writing overly vague responses. If for whatever reason you receive a 0, you can submit another response in a later week (if any are left) and get the points.

LATE POLICY:

Homework is due by the posted date and time. Students have three total late days they can use and no late homework after those days are used up is accepted without prior approval. For planned conflicts (e.g., conference travel, work interviews, etc.), students can request and extension by emailing the GSI at least 24 hours in advance of the deadline. Requests made under 24 hours will only be considered under exceptional circumstances (e.g., broken laptop, severe illness). Submitting at 12:05 will consume a whole late day, which is just as frustrating for us to see as it is for you. Please submit on time.

ACADEMIC HONESTY:

Unless otherwise specified in an assignment all submitted work must be your own, original work. Any excerpts, statements, or phrases from the work of others must be clearly identified as a quotation, and a proper citation provided. The instructors actively check for plagiarism and have zero tolerance. If caught, you will receive a zero for that portion of the assignment (or the whole assignment), depending on the scope of the plagiarism. All cases will be referred to the Office of Academic Integrity, regardless of their scope. Please (please) do not plagiarize; we hate finding these cases. If you are stressed and tempted to submit other’s work, please come talk to us first.

Re-using worked examples of significant portions of your own code from another class or code from other people from places like online tutorials, Kaggle examples, or Stack Overflow will be treated as plagiarism. If you are in doubt, please contact one of the instructors to check.

Any violation of the University’s policies on Academic and Professional Integrity may result in serious penalties, which might range from failing an assignment, to failing a course, to being expelled from the program.

Violations of academic and professional integrity will be reported to Student Affairs. Consequences impacting assignment or course grades are determined by the faculty instructor; additional sanctions may be imposed.

STUDENT MENTAL HEALTH AND WELLBEING:
The University of Michigan is committed to advancing the mental health and wellbeing of its students. If you or someone you know is feeling overwhelmed, depressed, and/or in need of support, services are available. For help, contact Counseling and Psychological Services (CAPS) at (734) 764-8312 and https://caps.umich.edu during and after hours, on weekends and holidays, or through its counselors physically located in schools on both North and Central Campus.

You may also consult University Health Service (UHS) at (734) 764-8320 and https://www.uhs.umich.edu/mentalhealthsvcs, or for alcohol or drug concerns, see www.uhs.umich.edu/aodresources.

For a listing of other mental health resources available on and off campus, visit: http://umich.edu/~mhealth/

INFORMATION FOR STUDENTS WITH DISABILITIES:

If you think you need an accommodation for a disability, please let me know at your earliest convenience. Some aspects of this course, the assignments, the in-class activities, and the way we teach may be modified to facilitate your participation and progress.

As soon as you make me aware of your needs, we can work with the Office of Services for Students with Disabilities (SSD) to help us determine appropriate accommodations.

SSD (734-763-3000; http://ssd.umich.edu) typically recommends accommodations through a Verified Individualized Services and Accommodations (VISA) form.

I will treat any information that you provide in as confidential a manner as possible.