

# EECS 498-004: Introduction to Natural Language Processing

**Time and Location:** Mondays and Wednesdays 10:30am - 1:30pm, online via Zoom (Zoom link is provided on Canvas and Piazza)

**Instructor:** [Lu Wang](#)

**Staff and Office Hours:**

- Prof. Lu Wang, Wednesdays 12pm - 1pm, online via Zoom (Zoom link is available on Canvas and Piazza)
- IA Yue Kuang, Thursdays 5pm - 6pm, online via Zoom (Zoom link is available on Canvas and Piazza)
- IA Ruobing Wang, Tuesdays 8pm - 9pm, online via Zoom (Zoom link is available on Canvas and Piazza)

**Discussion Forum:** [Piazza](#), sign up at [piazza.com/umich/winter2021/eecs498004](https://piazza.com/umich/winter2021/eecs498004)

---

## 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 and a research-driven course project.

### 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/](http://web.stanford.edu/~jurafsky/slp3/).
- Other Reference:
  - Jacob Eisenstein, "Introduction to Natural Language Processing", The MIT Press, 2019
  - Chris Manning and Hinrich Schütze, "Foundations of Statistical Natural Language Processing", MIT Press, 1999

### Prerequisites

This course is designed for senior undergraduate students majoring in computer science, information 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 is recommended) proficiently, and finish courses in algorithms, probability, and statistics. Linear algebra is optional, but highly recommended. It would be beneficial if the students have prior knowledge on supervised machine learning.

---

## Grading

Each assignment or report is due by the end of day on the corresponding due date (i.e. 11:59pm, EST). Canvas 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. every 24 hours). Each student has a budget of 8 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. There is no need to inform the instructors if late days are used; timestamp of the last submission on Canvas will be used for automatic grade calculation.

Grades will be determined based on assignments, project, and participation:

- Assignments (60%): four assignments, each of 15%
- Project (35%): team of 2 to 3 students, proposal (5%), reports (8%+12%, mid-term and final), presentations (7%, with 1% as bonus if selected as best project by peer students), feedback to other teams' presentations (3%)
- Participation (5%): attendance, participating in-class and Piazza discussions, etc

### Sample Writeups for Previous NLP Course Projects

- [\[report0\]](#)
  - [\[report1\]](#)
  - [\[report2\]](#)
  - [\[report3\]](#)
  - [\[report4\]](#)
  - [\[report5\]](#)
  - More sample projects from Stanford NLP course [\[link\]](#)
- 

## Schedule

### Jan 20

- Topic: Introduction, Language Models
- Slides: [\[intro\]](#) [\[6pp version\]](#) [\[Language Model\]](#) [\[6pp version\]](#)

- Reading: Ch1, Ch4.1-4.9 (2nd edition by default)
- TODO: start thinking about projects and looking for teammates

### **Jan 25 & 27**

- Topic: Language Models cont'd, Text Categorization and Evaluation, Naive Bayes, Part-of-Speech
- Reading: Ch5.1-5.5, Ch6.1-6.5

### **Feb 1 & 3**

- Topic: Part-of-Speech cont'd, HMM
- Reading: Ch6.1-6.5, Ch20

### **Feb 8 & 10**

- Topic: Machine Learning Basics, Neural Networks, Formal Grammars of English
- Reading: Ch 20, Ch7.1-7.4 in 3rd edition, Ch12.1-12.5
- Course project proposal due on Feb 12.

### **Feb 15 & 17**

- Topic: Syntactic Parsing
- Reading: Ch12.1-12.5, Ch12.7, Ch 13.4.1

### **Feb 22 & 24**

- Topic: Vector-Space Lexical Semantics, Distributional Semantics
- Reading: Ch20.7

### **Mar 1 & 3**

- Topic: Neural Language Models, Sentiment Analysis
- Reading: Ch7 and Ch21 in 3rd edition

### **Mar 8 & 10**

- Topic: Information Extraction, Question Answering
- Reading: Ch23.1-23.2

### **Mar 15 & 17**

- Topic: Summarization, Machine Translation
- Reading: Ch23.3
- Project progress report due on Mar 19.

### **Mar 22 & 24 (well-being break)**

- Topic: Dialogue Systems
- Reading: Ch24.1-24.2

### **Mar 29 & 31**

- Topic: Chatbots
- Reading: Ch24.1-24.2

### **Apr 5 & 7**

- Topic: Coreference Resolution, Discourse Analysis
- Reading: Ch25.1-25.5

### **Apr 12 & 14**

- Topic: Course Project Presentation

### **Apr 19 & 21**

- Topic: Course Project Presentation
- Project final report due on Apr 23.