

# Course Syllabus

[Jump to Today](#)

## Welcome to Applied Natural Language Processing (i256)!

Mon, Wed 10:30-12:00, 210 South Hall

Prof. [Marti Hearst](http://people.ischool.berkeley.edu/~hearst/) (<http://people.ischool.berkeley.edu/~hearst/>), [hearst@berkeley.edu](mailto:hearst@berkeley.edu)

Fall 2016

Office Hours: Wed 3-4pm, 307B South Hall

TA: [Andrea Gagliano](http://www.ischool.berkeley.edu/people/students/andreaagaglio/) (<http://www.ischool.berkeley.edu/people/students/andreaagaglio/>), [andrea.gagliano@berkeley.edu](mailto:andrea.gagliano@berkeley.edu)  
(<mailto:andrea.gagliano@berkeley.edu>)

Office Hours: Tue 3:30-4:30, South Hall Co-Lab (1st floor)

Much of the most valuable information available online today resides in textual form, but natural language is notoriously difficult to process automatically. Applied natural language processing -- also known as automated content analysis and language engineering -- can provide partial solutions.

This course will examine the state-of-the-art in applied NLP, with an emphasis on how well the algorithms work and how they can be used (or not) in applications. Today there are many ready-to-use plug-and-play software tools for NLP algorithms. For this reason, this course will emphasize getting facile with quick programs using existing tools. The intended **learning outcomes** are for students to:

- o Learn about major NLP issues and solutions
- o Become agile with NLP programming
- o Be able to assess NLP problems
- o Be able to get the gist of relevant research papers

This course will also be making use of a **different learning approach** than we use in most classes, which has been shown by hundreds of research papers to work better than the traditional lecture. This method makes use of what is variously known as active learning and peer/collaborative learning. What it means for students is:

- o Lecturing will be interspersed with active work in class, which means students must prepare for class in advance. Therefore ...
- o Students must prepare and turn in materials **before class every week**.
- o Students will be **actively engaged** during most of the class period, including **programming in class**.
- o Students will **work closely with other students** in class to improve their learning.
- o For these reasons, the class must be **taken for a grade**. No auditors, no S/U.

The course book is free online; it is [the book](http://nltk.org/book) [↗](http://nltk.org/book) (<http://nltk.org/book>) that accompanies the NLTK software, which will be working with extensively through the semester. Another terrific book is Jurafsky & Martin's [Speech and Language Processing](http://www.amazon.com/Speech-Language-Processing-2nd-Edition/dp/0131873210), [↗](http://www.amazon.com/Speech-Language-Processing-2nd-Edition/dp/0131873210) (<http://www.amazon.com/Speech-Language-Processing-2nd-Edition/dp/0131873210>), but since it is both too expensive and a bit too technical, we are not using it in this class.

The [UC Berkeley code of conduct](http://sa.berkeley.edu/code-of-conduct) (<http://sa.berkeley.edu/code-of-conduct>) is in effect in this class; you are expected to do your own work except when explicitly asked to work with others. You may consult with others but you must write your own code when that is required by an assignment. If you use code from elsewhere, you must explicitly note which pieces of code come from elsewhere and describe where the code comes from.

We are also using bcourses, which is a pretty terrific course management tool. The best way to view what is happening is via the [Modules View](#).

See the flyer for the final project poster session for 2016:

All Are Welcome!

# Applied Natural Language Processing

## Project Showcase

Wednesday, Dec. 7th  
10:30am - 12:30pm  
South Hall - Room 210

Professor: Marti Hearst  
GSI: Andrea Gagliano

### Rap Songz Generator

Natalie Goldberg, Robin Oh, Surendran Subbiah

### Robo-Joker: Your Digital Joke Assistant

Yiyi Chen, Jasper Louie, Molly Mahar

### Cyber Abuse Detection

Nihar Dalmia, Suchi Padhy, Rohit Raghavan, Nia Vivekanandan

### Deconstructing the 2016 US Presidential Election

Shirish Dhar, Boris Lo, Harman Shah Singh, Parv Sondhi

### Bulk Email Content Classification Analyzer

Shannon Hamilton, Shrestha Mohanty, Steve Trush

### Public Comments on Government Documents

Jason Danker, Proxima Dasmohapatra, Kinshuk, Emily Witt

### Terms of Service Summarization

Sai Dulla, Cristian Garay, Divya Garg, Deepa Kalpathi

### Pulitzer Prize Stylometrics

Michelle Carney, Gabe Nicholas, Sayan Sanyal, Natasha Timakova

### Course Recommendations

Fred Fan, Leon Li

### Automatically Scoring Student Responses

Avi Dixit, Elizabeth McBride

### Predicting Amazon Review Helpfulness

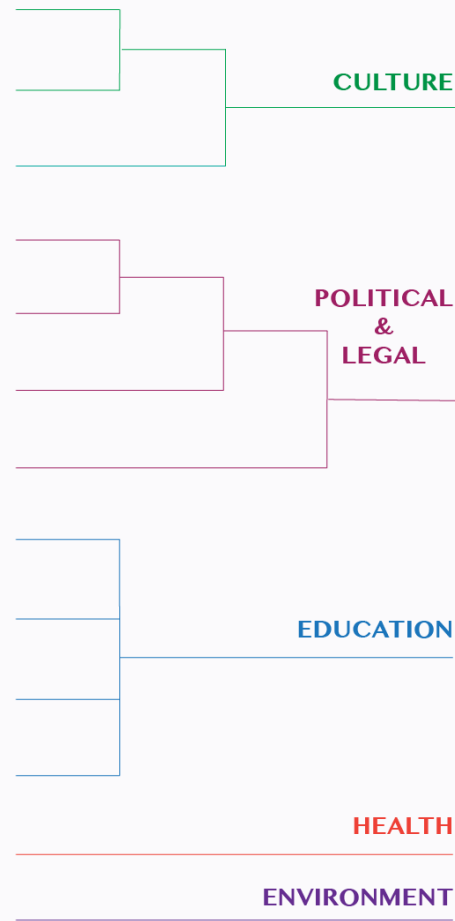
Jessica Ko, James Wei

### Predicting Drug-Drug Interactions

Sameer Bajaj, Bhuvana Bellala, Nicolas Mon, Vikram Reddy

### Yellowstone Wildlife Mapping

Guanghua Chi, Eric Hagen, Rob Kuvinka



NLP Technique: Hierarchical Agglomerative Clustering  
Lines indicate semantic similarity of projects

## Assignments Summary:

Date	Details	
Mon Aug 29, 2016	Aug 29 Preparation: Practice with ipython Notebooks	due by 9:30am
Wed Aug 31, 2016	Aug 31 Preparation: NLTK Text; Adopt a text collection	due by 9:30am
Wed Sep 7, 2016	Sept 7 Preparation: Tokenize Your Text Collection	due by 9:30am

Date	Details	
Mon Sep 12, 2016	📅 Sept 12 Prep: Create a First Look at Your Text Collection	due by 9:30am
Wed Sep 14, 2016	📅 Sept 14 Prep: Parts of Speech and Tagging	due by 9:30am
Mon Sep 19, 2016	📅 Part of Speech Tagging	due by 9:30am
	📅 Sept 19 Prep: POS Taggers	due by 9:30am
Wed Sep 21, 2016	📅 Sept 21 Prep: Practice Training a POS Tagger	due by 9:30am
Mon Sep 26, 2016	📅 Sep 26 Prep: Chunking	due by 9:30am
Wed Sep 28, 2016	📅 Sep 28 Prep: Syntactic Collocations; More on Term Weighting	due by 9:30am
Mon Oct 3, 2016	📅 Oct 3 Prep: WordNet Lexical Relations	due by 9:30am
	📅 WordNet Quiz	due by 9:30am
Wed Oct 5, 2016	📅 Oct 5 Prep: Work on your Keyphrase assignment	due by 9:30am
Mon Oct 10, 2016	📅 Keyphrase Identification Assignment	due by 9:30am
	📅 Oct 10 - 19 Preparation Information	due by 11:59pm
Wed Oct 12, 2016	📅 Oct 12 Prep: Names features	due by 9:30am
Fri Oct 14, 2016	📅 Assess Keyphrase Output	due by 11:59pm
Mon Oct 17, 2016	📅 Oct 17 Prep: Pandas Intro and Readings	due by 9:30am
Wed Oct 19, 2016	📅 Oct 19 Prep: Work on Kaggle Assignment	due by 9:30am
Mon Oct 24, 2016	📅 Oct 24 and 26 Prep: Work on Kaggle, think about final project ideas	due by 9:30am
Mon Oct 31, 2016	📅 Kaggle-based Text Classification Assignment	due by 9:30am
	📅 Oct 31 Prep: Read About Syntactic Parsing	due by 9:30am
Wed Nov 2, 2016	📅 Nov 2 Prep: Distributional Semantics readings	due by 9:30am
	📅 Review Quiz	due by 9:30am
Mon Nov 7, 2016	📅 Final Project Proposal	due by 9:30am
Mon Nov 14, 2016	📅 Nov 14 Class Prep	due by 9:30am

Date	Details	
Fri Nov 18, 2016	📅 Clustering and Distributional Semantics	due by 5pm
Wed Nov 30, 2016	📅 Nov 28 and 30 Prep: Readings and Quiz	due by 9:30am
	📅 Review Quiz	due by 9:30am
Mon Dec 12, 2016	📅 Final Project Assignment Writeup	due by 5pm