

Course Syllabus

[Jump to Today](#)

Welcome to Applied Natural Language Processing (i256)!

Mon, Wed 10:30-12:00, 202 South Hall

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

Fall 2015

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

TA: [John Semerdjian](http://www.ischool.berkeley.edu/people/students/johnsemerdjian/) (<http://www.ischool.berkeley.edu/people/students/johnsemerdjian/>), jsemer@ischool.berkeley.edu

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 minimized in favor of 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 **extensive 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](http://courses.ischool.berkeley.edu/i256/f15/) (<http://courses.ischool.berkeley.edu/i256/f15/>).

Assignments Summary:

Date	Details	
Mon Aug 31, 2015	 Aug 31 Preparation: Practice with ipython Notebooks	due by 9:30am
Wed Sep 2, 2015	 Sept 2 Preparation: NLTK Text; Adopt a text collection	due by 9:30am

Date	Details	
Wed Sep 9, 2015	 Sept 9 Preparation: Tokenize Your Text Collection	due by 9:30am
Mon Sep 14, 2015	 Sept 14 Prep: Create a First Look at Your Text Collection	due by 9:30am
Wed Sep 16, 2015	 Sept 16 Prep: Parts of Speech and Tagging	due by 9:30am
Mon Sep 21, 2015	 Part of Speech Tagging	due by 9:30am
	 Sept 21 Prep: POS Taggers	due by 9:30am
Wed Sep 23, 2015	 Sept 23 Prep: Practice Training a POS Tagger	due by 9:30am
Mon Sep 28, 2015	 Sep 28 Prep: Chunking	due by 9:30am
Wed Sep 30, 2015	 Sep 30 Prep: Syntactic Collocations; More on Term Weighting	due by 9:30am
Mon Oct 5, 2015	 Oct 5 Prep: WordNet Lexical Relations	due by 9:30am
	 WordNet Quiz	due by 9:30am
Wed Oct 7, 2015	 Oct 7 Prep: Work on your Keyphrase assignment	due by 9:30am
Mon Oct 12, 2015	 Keyphrase Identification Assignment	due by 9:30am
	 Run Keyphrase Extraction on Mystery Text	due by 11:10am
Wed Oct 14, 2015	 Oct 14 Prep: Names features	due by 9:30am
Mon Oct 19, 2015	 Oct 19 Prep: Pandas Intro and Readings	due by 9:30am
Wed Oct 28, 2015	 Oct 26 and 28 Prep: Read About Syntactic and Semantic Parsing	due by 9:30am
	 Review, Parsing, and Logic Quiz	due by 9:30am
Fri Oct 30, 2015	 Kaggle-based Text Classification Assignment	due by 10pm
Mon Nov 2, 2015	 Nov 2 Prep: Text Clustering	due by 9:30am
Wed Nov 4, 2015	 Nov 4 Prep: Distributional Semantics readings	due by 9:30am
Mon Nov 9, 2015	 Clustering and Distributional Semantics	due by 9:30am
Mon Nov 16, 2015	 Final Project Proposal	due by 9:30am
Mon Nov 23, 2015	 Nov 23 Prep: Read About Recurrent Neural Networks	due by 9:30am

Date	Details
Mon Nov 30, 2015	 Nov 30 and Dec 2 Prep: Readings and Quiz due by 9:30am
Wed Dec 2, 2015	 Review Quiz due by 9:30am
Wed Dec 16, 2015	 Final Project Assignment Writeup due by 5pm
	 Oct 12 -21 Preparation Information
