Syllabus LING131A

Introduction to NLP with Python

Description

This is an introductory graduate-level course on the computer processing of natural language text with the Python programming language. Python has quickly become the most popular programming language in Natural Language Processing (NLP) since it first came into existence because it has built-in data structures that allow natural language text to be manipulated with ease and elegance. Python has a relatively short learning curve compared with other high-level programming languages such as Java and beginners in Python can build up their programming skills fairly quickly. In addition, a large number of Python modules (such as the NLTK) already exist for language processing purposes so that linguistically oriented Python programmers can start to write practically useful programs within a relatively short period of time. Students are discouraged, however, from becoming overly reliant on third-party modules so that they could write code optimized to their own specific needs. The key to being successful in the course is to get your hands dirty and write a lot of code. By taking this course you have shown a commitment to become proficient in programming. If you have never written any code before you may need to adapt to a new learning style that is practice-oriented rather than reading-oriented.

Textbook

The book used in this course is *Natural Language Processing with Python*, by Steven Bird, Evan Klein and Edward Loper. 2009. O'Reilly Media, Inc. It is available online at `http://www.nltk.org/book/`.

Learning objectives

- Provide an introduction of the fundamental data structures and programming constructs in Python.
- Provide an introduction of the basic programming techniques and linguistic resources that contribute to solving natural language problems.
- Help students acquire solid Python programming skills by manipulating natural language data.

- Provide an introduction of the fundamental techniques used in lexical, morphological, syntactic, and semantic analysis.

Prerequisites

You can take this class if one of the following holds:

- You are a CL Masters student.
- You are an undergraduate student and you have taken LING100a and CS11a (or equivalents).
- You are a graduate student and you have taken CS11a (or equivalent).

Basically, to do well in this course it helps if you have some programming background and know some linguistics. When in doubt, contact the instructor.

Topics

The NLP topics discussed in class include the following: corpus processing, word lists, WordNet, tokenization, regular expressions, Unicode, part-of-speech tagging, the vector space model, classification, trees.

A more precise class schedule will be published on the website at `http://www.cs.brandeis.edu/~ling131/`.