Overview

This course is designed to be the first of a two course sequence that exposes linguistic graduate students to core methodologies and techniques in computational linguistics. This course will introduce you to fundamental programming techniques at a level that will allow you to begin building NLP (Natural Language Processing) applications as well as owing the abilities to perform various corpus processing tasks. This course will be taught in English and assumes only minimal background with computers, no programming skills or knowledge are required.

The programming language we will use is Python (and a bit Linux commands). It is noted that although assignments, lectures and projects will be mostly text and linguistically-oriented, the course provides sufficient 'general' programming background for students from any discipline who are interested in learning programming for their works.

Course title Introduction to computational linguistics with Python = Python programming for linguists = Programming fundamentals for computational linguistics

Teaching assistants Yu-Yun (Taco) [yuyun.unita@gmail.com] and Meng-Xian (Simon) [simon.xian@gmail.com].

Syllabus (Schedule subject to change)

Week	Date	Topic (2 hours)	Lab (1 hour)
1	11/09	Orientation	
2	18/09	Introduction to Python and Language Processing; Installation, First program	
3	25/09	Variables, expressions and statements; Working with corpora	
4	02/10	Function (I) Working with corpora	
5	09/10	Data flow (conditionals and recursion); Working with lexical resources	
6	16/10	Functions (II); Working with lexical resources	
7	23/10	Data flow (iteration); Text/corpus processing and case study	
8	30/10	Data types (strings) Text/corpus processing and case study	
9	06/11	Data types and data structures (lists, tuples); ; Text/corpus processing and case study	
10	13/11	Data structures (dictionaries) Text/corpus processing and case study with JSON	
11	20/11	File input and output; Scientific Plotting(`matplotlib`) Excursion: python for fun (VPython)	
12	27/11	Modules and Packages; Web APIs Big data and Mash-up with Python (I); Proposal discussion	
13	04/12	Regular expressions; Big data and Mash-up with Python (II)	
14	11/12	More regular expressions; Proposal presentation	
15	18/12	[Paris:official leave] Object-oriented programming (I) Python libraries and Web applications	
16	25/12	Object-oriented programming (II) Python libraries and Web applications	
17	01/01	[National Holiday]	
18	08/01	Term project **presentation and report** due (on `Github`)	

Class Activities and Requirements

- You are expected to complete weekly readings and take-home assignments, and actively participate in both class and online activities. A course website with NTU-CEIBA is already set up for this course. Check it out at `https://ceiba.ntu.edu.tw/1021python`. Additionally, the project page will be available soon at `http://lope.linguistics.ntu.edu.tw/LOPEN`.

- Each class will be divided into two sessions: lecture and discussion (two hours), and in-class Lab session (one hour). Lab session will give participants hands-on experience in writing Python programs for language processing.

- You will be given in-class and homework assignments weekly, which are due at the end of their due date (midnight). Late submission will incur penalty (30% of the total possible marks for that task). All assignments must be submitted online. It is no longer possible to submit assignments via email or in person. Instructions on how to submit your homework assignments will be given by our TAs next week.

- You are expected to work with others in conducting a semester long term project on the selected topics to be announced in the class. We will jointly come up with a plan for the projects. Basically, we will be working on the development of `Linguistic Corpus Search and Sketch Engine` (in Chinese) and an `APP for Personality/Emotion Recognition` (in English). So you should start thinking about it as early as possible. Project samples and instruction guides will be accessible soon on the course web page.

- Grading is based on: one (mid-term) exam for 10%. 60% of the grade will be based on participation and exercises. The remaining 30% will be based on their performance on a project, including the effort going into its development and the quality of their written and oral presentation of it during the last week of class.

- There will be no single textbook for this course. However, we will mainly refer to two learning materials: [1] "How to Think Like a Computer Scientist", [1] and [2] "Natural Language Processing with Python — Analyzing Text with the Natural Language Toolkit" [2]. These two books are available for free on the web. Additionally, we will rely heavily on additional self-learning resources such as `http://www.codecademy.com` for preview homework.

  For your interest, other references are listed as follows.

References

[1] Hetland, Magnus Lie. *Beginning Python: From Novice to Professional*. Apress.

[2] Johnson, Mark. 2009. *Essential Python for Corpus Linguistics*. Blackwell, New York.

[3] Lutz, Mark. 2008. *Learning Python*. O'REILLY.

[4] Merz, David. 2003. *Text Processing in Python*. Addison-Wesley, Reading, USA.

[5] Perkins, Jacob. 2010. *Python Text Processing with NLTK 2.0 Cookbook*. PACKT Publishing.

[6] Sampson, Geoffrey and Diana McCartthy. 2004. *Corpus Linguistics: Readings in a Widening Discipline.* Continuum.

[1] (`http://openbookproject.net/thinkcs/python/english2e/`, 2nd, 2012)
[2] (`http://nltk.org/book/`)(2009)