

Spring, 2021

Course: Python and NLP

Time: 9:10 – 21:00, Friday mornings

Instructor: Jiun-Shiung Wu, Ph.D.

Contact: wu.jiunshiung@gmail.com or Ext. 21307

Office: Rm 271, College of Humanities

Office hours: 9:10-12:00, Tuesdays

- Course description

This course attempts to equip students with knowledge and ability to conduct research on computational linguistics with the help of a famous python module NLTK (Natural Language Tool Kit). NLTK implements many useful tools and resources so that users can concentrate on important issues they are interested in and do not need to worry about the details of implementing those tools and loading and formatting corpora, etc.

Computational linguistic tools are also important to teachers who teach Chinese. For example, if a Chinese teacher wants to find out the linguistic features and common mistakes and interaction of any type in TCSL compositions, a corpus study of students' compositions is inevitable. It will help a great deal if a Chinese teacher has the knowledge and ability to use computational linguistic tools to assist his/her teaching and maybe research.

The goal of the course is two-fold. First, students will acquaint themselves with the tools provided by NLTK to play with the linguistic resources incorporated in NLTK and then to conduct his/her own research. Second, students can learn the issues covered in NLTK and be inspired to go beyond and explore other ways to conduct computational linguistic research.

Although the title of this course is “Python and NLP”, those who do not have any background in Python are enthusiastically encouraged to try. Those with no previous knowledge of Python will pick up sufficient information about Python such that he/she can follow the class instructions, finish the homework and conduct a meaningful research on computational linguistics.

- Requirements

1. homework: a few, 40%
2. Final paper: a research paper, 40%
3. Class participation: 10%
4. Attendance: 10%

- Textbook

Steven Bird, Ewan Klein, and Edward Loper. 2020. *Natural Language Processing with Python*, on line: <http://www.nltk.org/book/>

- Tentative syllabus (subject to change)

Week	Date	Contents
1	2/26	1. going through syllabus 2. Chap. 0: Preface 3. Chap. 1: Language Processing and Python
2	3/5	Chap. 1 (Cont'd) Chap. 2 Accessing Text Corpora and Lexical Resources
3	3/12	Chap. 2 (Cont'd) Chap. 3 Processing raw text
4	3/19	Chap. 3 (Cont'd) Chap. 4 Writing structured programs HW 01 distributed
5	3/26	Chap. 4 (Cont'd) Chap. 5 Categorizing and Tagging Words HW01 Due
6	4/1	校際活動，停課
7	4/8	Chap. 5 (Cont'd) Chap. 6 Learning to classify texts
8	4/15	Chap. 6 (Cont'd) HW02 Distributed
9	4/22	Chap. 7 Extracting information from text HW02 Due
10	4/29	Chap. 7 (Cont'd)
11	5/7	Chap. 8 Analyzing sentence structure
12	5/14	Chap. 8 (Cont'd) HW 03 Distributed
13	5/21	Chap. 9 Building feature-based grammar HW03 Due
14	5/28	Chap. 9 (Cont'd)
15	6/4	Chap. 10 Analyzing the meaning of sentences
16	6/11	Chap. 11 Managing Linguistic Data
17	6/18	Term paper due Term paper presentation
18	6/25	Final Exam Week, no class

