

課程名稱：Python 與自然語言處理 (二)

學期：110學年度第二學期

上課時間：星期五早上9:10-12:00

授課教師：吳俊雄博士

連絡方式：wu.jiunshiung{at}gmail.com

留辦時間：待定

課程描述

本門課程介紹以下主題：(一) 如何將詞以數據的方式呈現，如，獨熱向量、詞袋、TF-IDF等；(二) 文本語意分析，如，LSA、SVD、PCA、LDiA等，與文本-主題相關技術，並介紹文本相似性的計算；(三) 深度學習，神經網路，如，Word2vec、CNN、RNN、LSTM、Seq2Seq等。同時也介紹這些技術在自然語言處理中的應用。本課程預設了選修同學具語言學背景，最主要的目的在訓練語言所同學了解這些技術，培養應用這些技術的能力，配合本所學生語言分析的專長，進行語料前處理及標註，並結合這些演算法，讓本所同學具備應用自然語言處理演算法進行研究的能力。另外，也引導語言學、語言教學同學從對語言的深刻認識來對自然語言處理技術進行反思。

教材

Lane, Hobson, et al. 2019. *Natural Language Processing in Action*. Shelter Island, NY: Manning.

成績

1. 作業：數次，30%
2. 專題：三個，60%
3. 出席及上課參與：10%
4. 因為本課程預設選課學生具備語言學知識，凡非本所學生選修，均需加做語言學作業，總成績百分比分配如下：
 - 語言學作業：數次，40%
 - NLP作業：數次(與上面相同)，20%
 - 專題：三個，30%
 - 出席及上課參與：10%

進度(暫訂，依實際上課情況修正)

週次	日期	上課內容
1	2/18	1. 課程大綱講解 2. Chap 02, Building your vocabulary
2	2/25	1. Chap02, Building your vocabulary (Cont'd) 2. Chap 03, Math with words (TF-IDF vectors)
3	3/4	1. Chap 03, Math with words (TF-IDF vectors) (Cont'd) 2. Chap04, Finding meaning in word count (semantic analysis)
4	3/11	1. Chap 04, Finding meaning in word count (semantic analysis) (Cont'd)
5	3/18	1. Chap 04, Finding meaning in word count (semantic analysis) (Cont'd) 2. Chap 05, Baby steps with neural networks
6	3/25	1. Chap 05, Baby steps with neural networks (Cont'd) 2. Chap 06, Reasoning with word vectors 3. Project 1 due
7	4/1	春假・放假一天
8	4/8	1. Chap 06, Reasoning with word vectors (Cont'd) 2. Chap07, CNN
9	4/15	1. Chap 07, CNN (Cont'd)
10	4/22	1. Chap08, RNN
11	4/29	1. Chap 08, RNN (Cont'd)
12	5/6	1. Chap 09, LSTM 2. Project 2 due
13	5/13	1. Chap 09, LSTM (Cont'd)
14	5/20	1. Chap 10, Seq2Seq
15	5/27	1. Chap 10, Seq2Seq
16	6/3	端午節・放假一天
17	6/10	TBA
18	6/17	1. Project 3 due 2. Projects (all three) presentations