

JOUR 5020 Syllabus

109-2 社會科學程式設計 Programming for Social Science

本門課主要的授課對象為對應用計算機方法於政治傳播與社會議題有興趣的學生，相較於演算法與數值運算，將著重在社會科學的文字運算、網絡結構、人口變項、地理資訊等跟資料特性相關的運算。如果你未來想要用這類議題作為碩士論文或未來想到國外修讀計算社會科學碩士學位者，將是最合適的選修人選。或者你曾經選修過經濟系、共教中心、新聞所等R語言相關課程但卻尚未學過Python程式語言的同學，更是你接續學習第二種語言的最好機會。

本門課相較於一般程式設計課程會更重視文字的處理和實際的應用，可作為經濟系「資料科學與社會研究」Lv 3的跨域應用課程。如果已修過其他校內Python相關課程請勿選修，避免浪費學習資源，如孔令傑老師的商管程式設計、或者國發所的Python程式語言與文字探勘等課程，均著重在不同的應用面向，非常值得學習。因此，本門課選課的優先順序為新聞所 > 社科院、法學院、生傳系 > 文學院 > 其他科系。

Announce

1. 開學第一週前如尚未加選成功者，請務必在上課前(2/21 23:59)填好此問卷，當天填不計(<https://forms.gle/XqGDAkRUMr5v4ZAbA>)。
2. 開學第一週請上Datacamp申請帳號，並將帳號所用之E-mail，在02/26 12:00前填寫至<https://forms.gle/HSsbhuoY3Meo7eiu7>，以利開通半年免費權限。

Basic Information

INSTRUCTOR: 謝吉隆(新聞所副教授)
TA: 吳亭霓 hihidia2175@gmail.com
劉哲愷 b07305051@ntu.edu.tw
TIME: MON 2, 3, 4
ROOM: 新聞所R103

COURSE MATERIALS

- <https://github.com/P4CSS/PSS>。本課程使用github分享上課所用的程式碼，同學可在上課期間用google colab直接觀看，也可clone整份github repository至個人電腦上，用Anaconda + Jupyter Lab執行。
- 程式開發環境。請安裝Anaconda Python 3.X版(安裝Anaconda的方法 [Anaconda 3 安裝教學及說明](#)。Python 開發環境介紹- Anaconda 3 完整安裝說明及步驟| by Coding Lab | AI for K-12)。
- Think Python 2e [Think Python 2e – Green Tea Press](#)。看英文應該會比較容易懂，但 [《Think Python 2e》最新版中文翻譯 – 《Think Python 2e》中譯本1.0 文档](#)

COURSE WEBSITE: NTU COOL <https://cool.ntu.edu.tw/>

Q&A: www.facebook.com/groups/ntupss/ 請盡量在此發問，在NTU COOL發問不見得能夠即時回覆。請勿貼上可成功執行作業的程式碼，會被刪除貼文。

Hybrid face-to-face/Asynchronous learning

本學期將採用混成教學，總授課時數包含實體課程共12週26小時、線上教材共39小時、週末工作坊6小時。

- 實體課程：除第一週與報告週次外，自第二週起縮短課堂時間為兩小時(10:20~12:10)，主要活動為程式概念講解、Lab操作、作業說明等。
- 線上非同步教學：每週約2~3小時程式實作教學影片，隨著老師講解，從零開始按步驟寫程式，估計約3~5小時可以消化完畢，依每個學生的情形不同。實體課堂教學
- 業界教學：本學期將有業界參與教學，業界會提供等同於Hahow上價值約兩千多元的課程內容教學，非常實用。但因業者無法於平日授課，**目前規劃將教學統一集中至週末一天進行，如果無法接受週末另外上課的同學(僅一週)，請勿選修。**

Grading

- 缺席 -3%：請一律由[學生請假系統](#)請假，本課程會不定期在實體面對面授課課程利用小考來提醒同學重要的概念，如經查未到課未假者，酌扣學期總成績3分。
- 作業 40%：平均每週會出一個作業共10回，為期一週；
逾期**三日內(週三晚上23:59)**補交分數乘以0.6、逾期**三日**後將不接受補交。
- 線上影片 5%：NTU COOL平台會紀錄影片觀看長度，將依照觀看比例給分。
- Datacamp 3 modules: 5%

- 
- 課堂練習10%:每次課堂教學後都會有課堂練習共10次。
在授課當天23:59前完成者得1分, 隔天完成者得0.5分。
 - 專題 40%:包含期末海報展評審20%、期末報告20%。
未參加海報展者一律只能獲得20%。

Tentative Calendar

Week	Date	In-Class	Online
1	2/22	Course overview Python Basic: Counting (variable and flow control), keyness of text	DataCamp: Introduction to Python 4h
2	3/01	228 補假	DataCamp: Intermediate Python 4h
3	3/08	Python Basic: List and Dictionary	DataCamp: Data Manipulation with Pandas 4h Assignment #1
4	3/15	Python Basic: for-loop and if-else	Assignment #2
5	3/22	Reading and manipulating data with Pandas	Assignment #3
6	3/29	Web Scraping: Overview	Assignment #4
7	4/05	春假	Assignment #5
	4/11	Workshop-Web scraper and data flow with GCP	
8	4/12	不上課, 調整為週末工作坊共六小時	
9	4/19	不上課, 調整為週末工作坊共六小時	
10	4/26	Web Scraping	Assignment #6
11	5/03	Text processing I: Tokenization and Collocation	Assignment #7
12	5/10	Text processing II: Keyness and word embeddings	Assignment #8
13	5/17	Text processing III: Topic modeling and text classification	Assignment #9
14	5/24	Text Processing IV: Word Embedding and Network	Assignment #10
15	5/31	ML: Unsupervised learning	Assignment #11
16	6/07	ML: Supervised learning DL: Using simpleTransformer	

17	6/14	端午節	
18	6/21	期末成果發表與課程聯合海報展	

Projects

期末專案需要以線上新聞、社群媒介內容(Twitter或Youtube Comments)作為主題。主題將限制在政治傳播、兩岸關係、意識形態上的應用(如股價、電影票房不包含在此類)，分數著重在文本分析對字詞清理、整理的深度。

繳交內容

請繳交上傳至Google Drive [PSS1092 Final Project](#)。該網址中有Sample Folder指示要上傳的內容，包含：

1. 程式碼與資料
2. 期末報告簡報(需於本學期最後一週報告，至少需完成資料爬取、資料清理和基本描述性統計)
3. 期末專題報告
 - a. 摘要僅需說明主題、主要分析結果含方法(<300 words)
 - b. 著重說明資料清理流程與資料取捨
 - c. 造表說明分析結果(例如比較不同群集法、主題建模下的主題差異;或比較不同演算法的分類成效)

報告方式

使用Google Slide, 預先將簡報連結貼至專題總表 [PSS1092-Final](#)。請注意簡報的分享狀態是否已經公開。報告時間為10分鐘，講評2分鐘，請著重在問題意識、資料的使用和資料分布、用圖表呈現結果並加以描述。最多為15張簡報，超過20張簡報(含標題與大綱)將酌量扣分。

繳交期限

1. 擬於期末考後的週一用Google Slide進行期末報告(6/21或6/28需討論)
2. **期末專題報告繳交時間(7/2 23:59)**。如需延後專題報告繳交時間，請規劃繳交時間，並經全組組員回信表示同意後將該信件寄給老師。然而，無論是否延長，均須在指定的期末報告日進行報告進度，並在規劃繳交時間前一併上傳Slide與期末報告，並再次進行結果報告。

3. 應屆畢業生請務必集中火力完成專題，不要影響到個人的畢業時程。

專題要求

- 對文本內容初步視覺化(例如詞頻)
- 清楚的文字清理流程和說明
- 使用非監督式學習探勘主題
 - a. 就群集結果反覆選擇適合的k,
 - b. 對整體群集結果的視覺化、對每個群集做關鍵字的視覺化,
 - c. 用文字討論非監督式學習的結果，並對每個群進行主題命名，說明關鍵字。
 - d. 使用gensim的topic modeling套件做主題建模(必要條件)
- 使用監督式學習
 - a. 必須評估過各種分類演算法、包含深度學習分類演算法，並利用ROC Curve以及Classification report做模型的比較說明。
 - b. 對於imbalance data(Negative和Positive數量相去甚遠)的資料，必須就不同的模型說明Recall和Precision的差異。
 - c. 過去認為Precision-Recall Curve (PR Curve)比較能夠呈現Imbalance Data的特性。請為你們挑選的模型繪製PR Curve比較其與ROC Curve的差異(必要條件)。對於imbalance資料的問題，請參考這篇論文 Rauchfleisch, A., & Kaiser, J. (2020). The False positive problem of automatic bot detection in social science research. PloS one, 15(10), e0241045.

可用資料

如果你們對於以下題目感興趣，可向助教登記使用我們研究室的資料。然而，不得將該資料以其他方式發表，僅限與本研究室成員共同發表。

- 未標記的涉中、涉美Youtube comments資料集(應用非監督式學習分析疫情前後對中、美態度的影響)
- 標記未完的涉中、涉美Youtube comments classification(目前資料標記正負項不均衡，需要補充協助標記資料至少3000則，並檢驗二人標記的一致性)

容易取得資料的題目

- 新聞不當標題分析
- 歌詞分析
- PTT、DCard文本探勘

PSS1092期末專題題目

1. 社群輿論議題的此消彼長關係探究－以#China相關Tweet為例
2. 疫情下的標籤：比較「武漢肺炎」和「新冠肺炎」。
3. 愛情與婚姻兩性「公平」論述
比較愛情版和婚姻版關於「公平」的文字論述
4. 疫情下的政治人物網軍－以「陳時中為例」
結果從時間特性發現侯友宜的網軍
5. 待補
6. 待補

參考題目

1. PTT網軍大風吹：網軍的人口統計特徵
2. 「誰是業配雯？」Dcard業配文偵測
3. 爬文PTT八卦版留言之情緒分析
4. 成為Airbnb房東，開始賺取收入
5. 從網路論壇對美食外送平台市場的觀察-以 Ptt、Dcard 為例
6. 媒體國教政策立場分析：以關鍵評論網與親子天下教育專欄為例
7. KKBOX排行榜熱門歌曲文本探勘
8. 經典該不該讀？佛教如來宗 vs 中台佛寺
9. 男女歌詞用詞分析與差異
10. 親愛的，你以為的電影姿態，不是電影姿態
11. 劇本裡的千層套路 - 從情感分析談電影
12. 射擊類遊戲steam評論分群
13. 透過python文字探勘分析地方政府新聞公告-以台北市、台中市、台南市、屏東縣、台東縣為例
14. 草根 vs. 菁英：公民參與與媒體社論政策關注差異
15. 中國外交部用詞分析：以耿爽為例，以及與台灣外交部之用詞比較
16. Dcard總統候選人網路聲量分析
17. 2020選舉臉書廣告內容分析
18. Twitter資料探勘：環保少女與香港示威者討論熱度比較
19. 文字探勘 X PTT上的性別歧視言論
20. 「是男？是女？」Dcard卡民性別辨識
21. 分手後我有話要說：分析PTT男女版上近十年的文章內容

Q&A

Policies

Q: 如果我已經上過校內類似商管程式設計等用Python作為程式語言或已經具有相當Python能力了, 那我還適合選修嗎?

只要你修過任何Python程式設計課程, 你都不太適合選修本門課。雖然本門課有些內容和其他使用Python的程式課程有所不同, 但本門課有過半屬於基礎教學, 而台大校內各個上Python的老師一定會涵蓋這些基本功夫。如果有心往資料科學界前進者, 應該避免重複上基礎課程。對於要拿到授權碼的加選者, 必定要沒修過Python相關程式課程始得加選。

Skill Q&A

Q: 如何安裝Python的開發環境Anaconda?

請上網查詢Anaconda的安裝方法, 如[Anaconda 3 安裝教學及說明](#), [Python 開發環境介紹- Anaconda 3 完整安裝說明及步驟](#) by Coding Lab | AI for K-12

Q: 如何用colab讀取github中的檔案?

請上網查詢「github colab」會找到 [Using Google Colab with github](#)

Related resources

Courses

- Leeds - [Broadening PIED3702 Introduction to Programming for Social Science](#)

Books

- [Programming with Python for Social Scientists](#)

Rundown

Week 1 Rundown

IN-CLASS

- Slide - [PSS Introduction](#)
- Slide - [Py01 Counting](#)
- <https://github.com/P4CSS/PSS>
- https://colab.research.google.com/github/P4CSS/PSS/blob/master/P01_counting.ipynb

NEXT DUE

- 下下週開始 3/8 上課時間為10:20~12:10
- 安裝好Anaconda with Python 3.x
- [DataCamp- Introduction to Python](#) (Due: 03/07 23:59)
- [DataCamp - Intermediate Python](#) (Due: 03/07 23:59)

Week 2 Rundown

AT-HOME

- Do practice [\[Datacamp\] Intermediate Python](#)
- Slide-[Py01 Counting](#)
- Code-https://colab.research.google.com/github/P4CSS/PSS/blob/master/P01_counting.ipynb
- Watching [py01_1 Counting](#)
- Watching [Py01_2 Counting more](#)
- Watching [Py01_3 Counting words](#)

NEXT-DUE

- 下週開始 3/8 上課時間為10:20~12:10
- 安裝好Anaconda with Python 3.x See [Anaconda 3 安裝教學及說明. Python 開發環境介紹- Anaconda 3 完整安裝說明及步驟| by Coding Lab | AI for K-12](#)
- Submitting screenshot of datacamp [Datacamp - Complete Screenshot - Introduction to Python](#) (Due: 03/07 23:59)

- Submitting screenshot of datacamp [Datacamp - Complete Screenshot - Intermediate Python](#) (Due: 03/07 23:59)

Week 3 03/08 Rundown

IN-CLASS

- [\[Slide\] Py02 Basic](#)
 - [P02 basic blank.ipynb](#)
- [P03 list dictionary blank.ipynb](#)
- Submitting [Week03 Practice](#). No later than 13:00 today.

AT-HOME

- [AS#1 Python Basic](#)
- [Datacamp - Complete Screenshot - Data manipulation with pandas](#)
- Video Review
 - [\[Video\] Py02 Python Basic](#)
 - [\[Video\] Py03_1 List](#)
 - [\[Video\] Py03_2 Dictionary](#)
 - [\[Video\] Py03_3 Accessing Youbike data](#)
- (Options) [PEP8 Python 編碼規範手冊](#)

NEXT-DUE

- [AS#1 Python Basic](#) (Due: 3/14 23:59)
- [Datacamp - Complete Screenshot - Data manipulation with pandas](#) (Due: 3/14 23:59)

Week 4 03/15 Rundown if-else

IN-CLASS

- [\[Slide\] Py04 flow control](#)
- [\[Code\] Py04 for if else.ipynb](#)

AT-HOME

- [\[Video\] Py04_1 Condition and flow control: An introduction + Case study](#)
- [\[Video\] Py04_2 if+for: Accessing and cleaning AQI data](#)
- [\[Video\] py04_3 Categorizing and rescaling data with if+for](#)

- [\[Video\] Py04 4 if+for: Detecting prominent items](#)
- [\[Video\] Py04 5 Sorting by if+for+list](#)
- [Week04 Practice](#)
- [AS#2 for-loop and if-else](#)

NEXT-DUE

- 03/19 Fri 23:59 (No Delay) [Week04 Practice](#)
- 03/21 Sun 23:59 (Last chance 03/24 Wed 23:59) [AS#2 for-loop and if-else](#)

Week 5 03/22 Rundown Reading data and pandas

IN-CLASS

- [\[Slide\] Py05 Read CSV and JSON](#)
- [\[Code\] P05 read_csv json.ipynb](#)
- [\[Slide\] Py06 Pandas and data summarization](#)
- [\[Code\] P06 pandas.ipynb](#)

AT-HOME

- [\[Video\] Py06 Pandas01 Reading data](#)
- [\[Video\] Py06 Pandas02 Accessing Dataframe](#)
- [\[Video\] Py06 Pandas03 Creating Variables](#)
- [\[Video\] Py06 Pandas04 Groupby and Sorting](#)
- [\[Video\] Py06 Pandas05 Series to df rent591](#)
- [AS#3 Data summarization with Pandas](#)
- [AS03 Description](#) 作業AS03説明

NEXT-DUE

- AS#3 Due 3/28 23:59 (Late but no later than 3/31 23:59)

Week 6 03/29 Rundown crawler

IN-CLASS

- [\[Slide\] Py07 Crawler](#)
- [\[Slide\] HTML Tutorials, CSS selector, and XPath](#)
- [\[Code\] P07p_get_web_json.ipynb](#) (github)
- [\[Code\] P07p_get_web_json.ipynb](#) (gdrive)

AT-HOME

- [AS#4 Web Scraping](#)
- [Week06 Practice \(Follow the video\)](#)
- [\[Video\] Py07_1 web request and response](#)
- [\[Video\] Py07_2 Chrome DevTools](#)
- [\[Video\] Py07_3 104.com Scraping](#)

NEXT-DUE

- 04/05逢春假不上課, 但04/11週末09:30要參加工作坊
- AS#4 and Week06 Practice Due 04/08 23:59 (No later than 04/08 23:59)

Week 7 04/05 change to 04/11

IN-CLASS

-

AT-HOME

-

NEXT-DUE

-

Week 10 04/26 Rundown

IN-CLASS

- [\[Slide\] Py07 Crawler](#)
- [\[Slide\] HTML xpath and css selector](#)
- [\[Code\] P08p_PTTcrawler.ipynb](#)
- [\[Code\] P09 Scraping with POST](#)

AT-HOME

- Assignment [AS#5 Web Scraping II](#), before 5月2日
- [Week 8 Practice \(Follow videos\)](#), before 5月2日
- [\[Video\] Py07_4 HTMLParser01 Getting and Parsing HTML](#)
- [\[Video\] Py07_4 htmlParser02 Finding elements](#)

- [\[Video\] Py07_4 htmlParser03 ErrorHandling](#)
- [\[Video\] Py07_4 htmlParser04 get more links](#)
- [\[Video\] Py07_4 htmlParser05 getting posts](#)

Week 11 05/03 Rundown

IN-CLASS

- [TM Tutorials](#)
- [\(Code\) TM01_tokenization.ipynb](#)
- [\[Code\] TM04_Chinese Processing.ipynb](#)
- [\[Video\] TM00 Text mining overview](#)
- [\[Video\] TM01_1 Term frequency: Tokenization \(43 mins\)](#)
- [\[Video\] TM01_2 Term frequency: stopword and punctuation removals](#)
- [\[Video\] TM01_2 Term frequency: stopword and punctuation removals](#)
- [\[Video\] TM01_3 Term frequency: Stemming and Lemmatization](#)
- [TM04 Chinese segmentation](#)
- [AS#6 Tokenization and Term Frequency](#)

Week 12 05/10 Rundown - Collocation

Colab	github
TM02_collocation.ipynb	TM02_collocation.ipynb
AS07p_Collocation.ipynb	AS07p_Collocation.ipynb
pttpost_20210509_n178.dat	pttpost_20210509_n178.dat

- [\[Slide\] Collocation](#)
- [\[Tutorial\] Collocation](#)
- [\[Video\] TM02_1 Collocation Introduction](#)
- [\[Video\] TM02_2 Collocation: Frequency-based](#)
- [\[Video\] TM02_3 Collocation: Collocation with distance](#)
- [\[Video\] TM02_4 Collocation by Chi-Square](#)
- [\[Video\] TM02_5 Collocation by Mutual Information](#)
- [AS#7 Collocation for commentor network](#)

Week 13 05/17 Rundown

- **Course Live Recording:** Watch [PSS-0517-Live-Part1-POS](#) with [\[Slide\] TM03 POS](#) and [TM03 POS Tagging.ipynb](#)
- **Course Live Recording:** Watch [PSS-0517-Live-Part2-Word-Embedding](#) with [\[Slide\] TM06 Representation & Embedding](#) and [TM06_embeddings.ipynb](#) and [TM06_embedding_chinese.ipynb](#)
- **Step-by-Step Tutorial:** Watch Video [\[Video\] TM06-Word Embedding \(EN\)](#) with [TM06_embeddings.ipynb](#) to learn how to use and train word2vec models
- **Step-by-Step Tutorial:** Watch Video [\[Video\] TM06 Word Embedding \(Chi\)](#) with [TM06_embedding_chinese.ipynb](#) to learn how to do dimension reduction (PCA and t-SNE) and a better visualization library Bokeh
- **Assignment :** Take a look at [\[Slide\] TM06 Representation & Embedding](#) and watch the end of [PSS-0517-Live-Part2-Word-Embedding](#) to do [AS#8 Gender Embeddings](#)

Week 14 05/24 Rundown: Clustering

1. **Course Live:** Student sharing, scraping twitter and using youtube data api
2. **Video tutorial**
 - a. Watch [\[Video\] pyTM04-IR-Search-Ranking](#) with code [TM04 Search](#) and slide [TM04 Search](#), to know how to extract important information from each document
 - b. Watch [\[Video\] pyTM06-Clustering](#) to learn the clustering concep with slide [TM06 Clustering](#)
 - c. Watch [\[Video\] pyTM06-Clustering-Coding](#) with code [TM06 Clustering](#)
3. **Assignment** Watch [\[Video\] Assignment-Clustering Introduction](#) to know how to do [AS09p-Clustering-Youtube-Comments](#)

Week 15 05/31 Rundown

- **Course Live Recording.** Will introduce the project and learning materials of the week

會議鏈結<https://ntucc.webex.com/ntucc/j.php?MTID=m4c637667bda9a2e1b2c58af7289d0def>
會議號: 184 257 5536
密碼: 1092

- **Video Tutorials:** Using Code [TM07_doc_classification.ipynb](#) with Slide [\(py\)ML01 Supervised Learning](#), [data/sentiment.csv](#) and following videos.
 - a. [\[Video\] pyML01 Supervised learning overview](#) will introduce the basic concept and flow of supervised learning
 - b. [\[Video\] pyML02 Models](#) will briefly introduce several classical models including Naive Bayes classifier, support vector machine (SVM), tree-based models, and logistic regression.
 - c. [\[Video\] pyML03 Coding](#) will run and introduce the code [TM07_doc_classification.ipynb](#)
 - d. [\[Video\] pyML04 ROC and confusion matrix](#) will introduce the confusion matrix and the idea of ROC curve
- **Submitting Practice after watching Video** [Week15 Practice - Doc Classification](#)

在本次教學中，我們使用了三種feature representation and selection的方法，包含tfidf, chi-square和embedding。在影片中，我們可以得知，當老師將chi-square的SelectBest開到1000個features的時候，似乎可以得到比其他model都還要好的f1-score。所以，請你嘗試用ROC Curve和classification_report比較看看，RandomForest, logisticRegression和SVC在Chi-Square Feature Selection下，k=100, k=1000, k=2000時ROC圖型、AUC和Classification_report分別為多少，哪一個Model在k為多少的狀況下是最好的？

請直接寫在 [TM07_doc_classification.ipynb](#) 最後的練習中並上傳。

- **Assignment:** [AS#10 Document Classification \(I\)](#)

Week 16 Rundown

- **LIVE: 10:20~11:00**

會議鏈結
<https://ntucc.webex.com/ntucc/j.php?MTID=m4c637667bda9a2e1b2c58af7289d0def>
會議號: 184 257 5536
密碼: 1092
- **At-home**
 - Watching [\[Video\] pyDL-Transformer01-introduction](#) to have some basic understandings of Deep learning

- Watching [\[Video\] pyDL-Transformer02-Coding](#) to know how to use simpletransformers package to run document-classification and clustering tasks
- **Assignment AS11**
 - Watching <https://youtu.be/yXEMJmgv31E> to get AS11 overview with colab <https://colab.research.google.com/drive/1ibv9Z3BtFvHMd5WkhGRRLOYgxWj32JAO?usp=sharing>
 - Submitting your link or ipynb (after download) to [AS#11 Document Classification \(II\)](#)

Learning path

Py01 Counting

Py02 Basic

Py03 List and dictionary

Py04 for-loop & if-else

Assignments

Doc-Classification (I)

Document Classification方法可用來預測物件的分類。然而，訓練資料是非常重要的。就像現在的人工智慧，確實真的是人工所產生的智慧，必須要有品質好的標記才能幫助機器訓練出好的模型。為了避免各位未來在做分類任務時有眼高手低的問題，也要能夠跟協作者、團隊或外包的標記者進行溝通，這個作業將邀請各位執行一個標記任務，每個人標記300篇留言。我們將把這週的標記結果作為下週作業的輸入資料。

在資料夾[USRelated-Run3-n28652](#)中，我們已把同學每三人一組分成五組，每組是同樣的一份300筆資料。這些資料都是用關鍵字所篩選出來的涉美Youtube留言，只是我們不知道這些留言是否與對美立場相關（例如有人提及「現在歐美都很進步了」，顯然就與對美的態



度無關, 只是意指普遍的西方我家); 對於有立場的留言, 也不容易從文本中去判斷留言的「親美、仇美與中立」態度。

這種標記方法是立場標記慣用的手法, 必須要區分出與立場相關和非立場相關的文本, 再進一步區分是哪一種立場。透過每組三人的標記, 我們應該可以比對標記的一致性。

每個Google sheet的檔名有列出各組同學的學號, 每個同學都負責自己學號的那個Sheet。每個檔案的最後一個sheet有標記的參考範例, 但為了不影響同學標記, 我們舉親中／反中態度為例。請謹慎標記, 也不要參考其他同學的標記。

請務必準時標記完畢。我們會在下週上課前(6/6 23:59)後合併各位的標記結果。

Doc-Classification (II)