資管碩二 張鑑霖 R05725034

1. 執行環境  
    Jupyter  
2. 程式語言

    Python 3  
3. 執行方式

* 解壓縮檔案，內含28.txt(文字檔)、HW1.ipynb(執行檔)、hw1\_result.txt(成果輸出檔)

#p.s.以下兩個指令在cmd執行

* pip install -U nltk

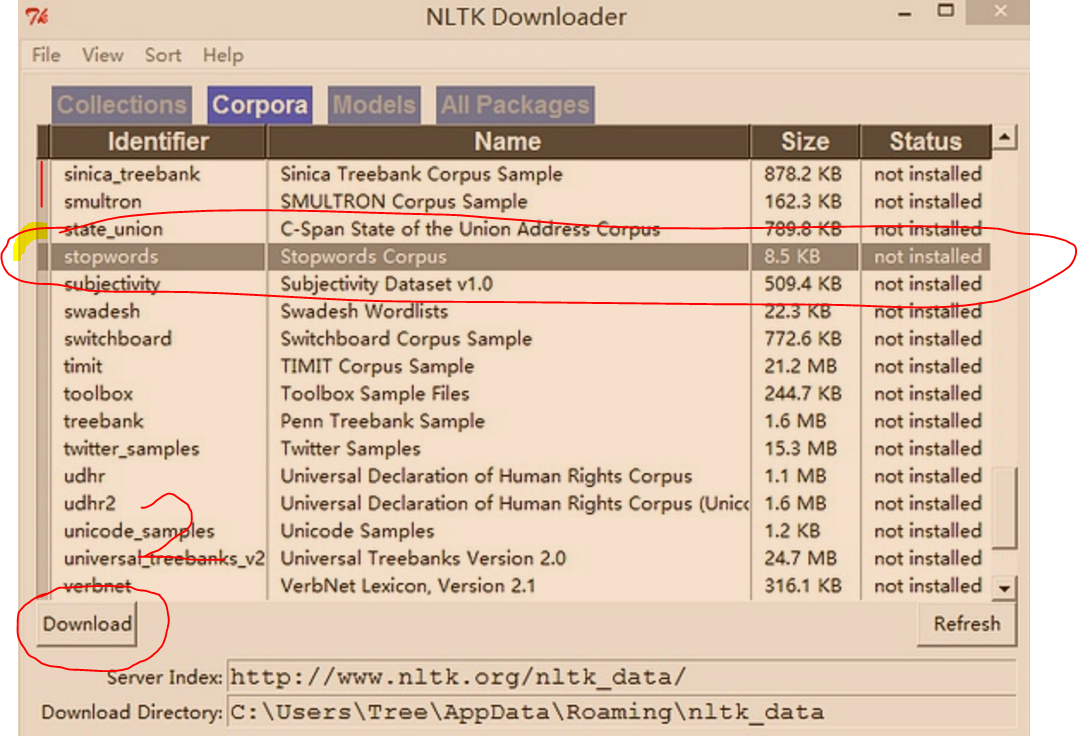
#進入python 執行import nltk、nltk.download()

* python

import nltk

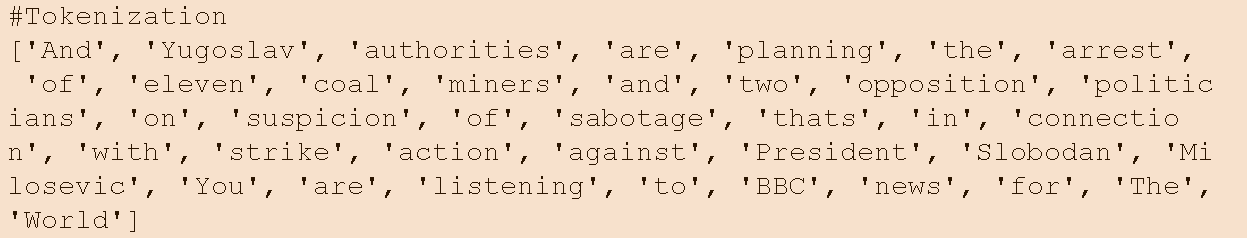
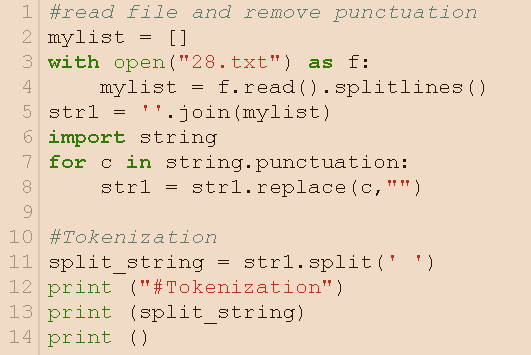
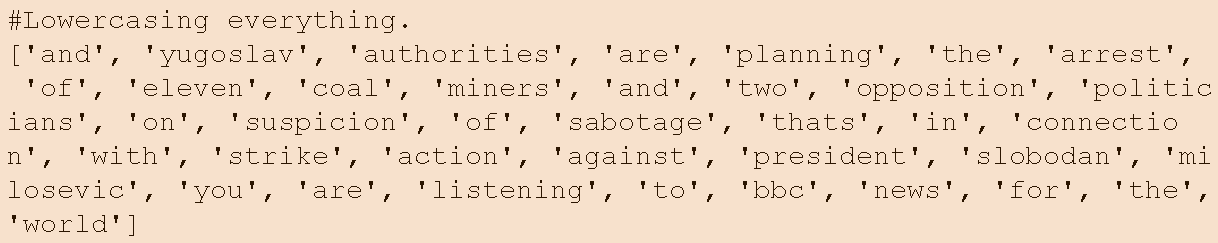
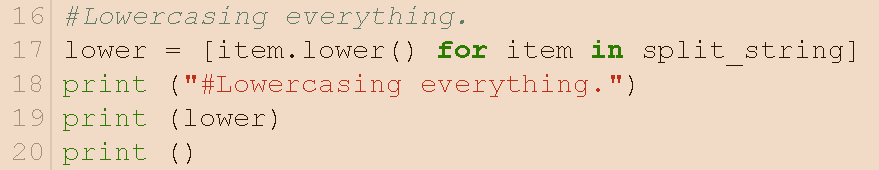
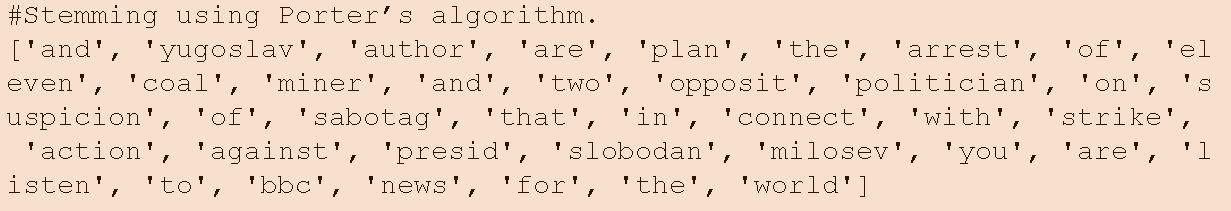
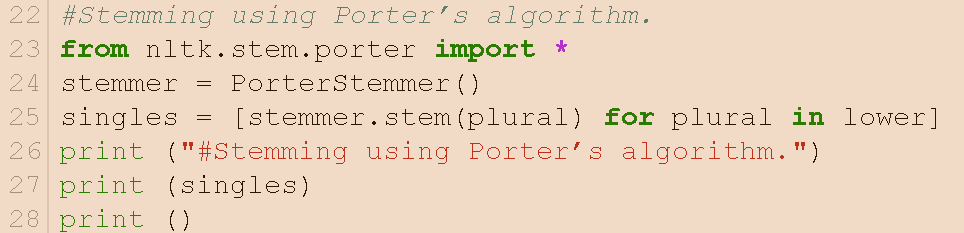
nltk.download()

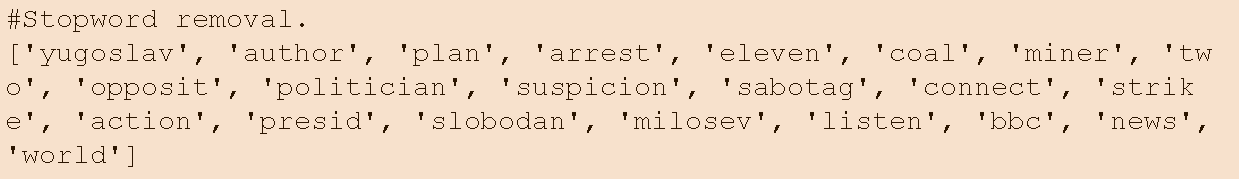
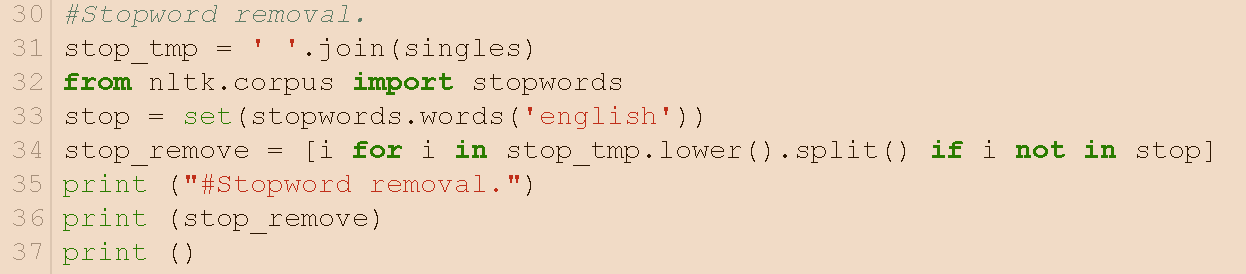
之後出現下圖，menu點到Corpora後，找到stopwords後按download



* 最後就可以於jupyter notebook上執行程式

4. 作業處理邏輯說明

1. 讀檔並去除標點符號、Tokenization  
    
2. Lowercasing everything
3. Stemming using Porter’s algorithm
4. Stopword removal (最後結果)



5.存檔輸出hw1\_result.txt，出現於同目錄下