**106-1 資訊檢索與文字探勘導論 Programming Assignment 1 Report**

**R06725032 資管碩一 吳定寰**

1. 執行環境

(.ipynb) Jupyter notebook

(.py) cmd

1. 程式語言與版本

Python 3.6

1. 執行方式及安裝套件
   1. pip install nltk
   2. 執行方式：

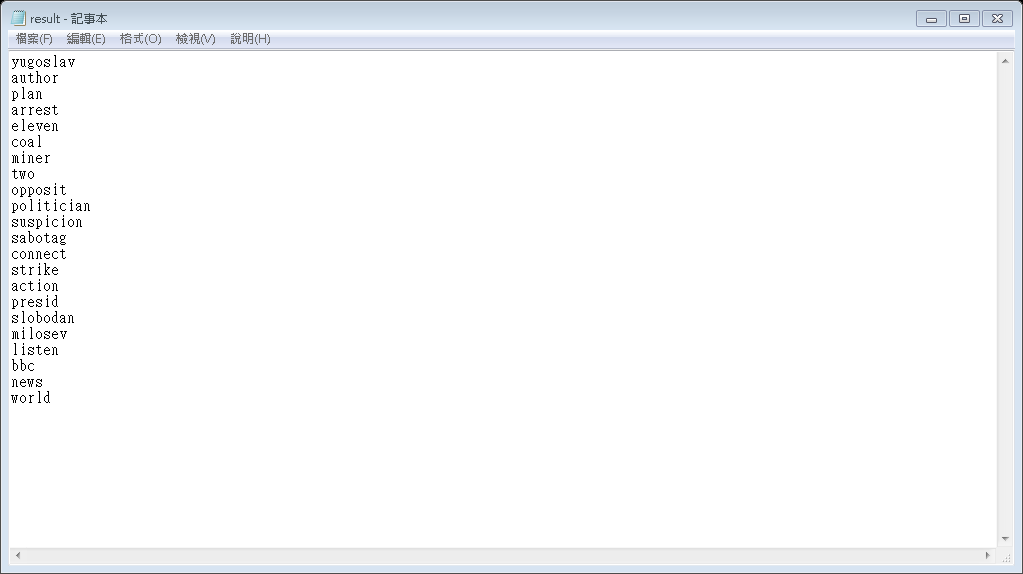
直接執行.py檔或在Jupyter內run cell即可，會在同資料夾內輸出result.txt檔，執行方式如下圖所示。



在Jupyter notebook開啟textProcessing.ipynb檔並執行



同資料夾內會輸出result.txt檔



輸出之結果

1. 作業處理邏輯說明

開啟news.txt檔進行全小寫化、去跳行符號並依空白字元進行初步分割後，使用正則表示式留下符合規則的token，接著使用nltk套件的PorterStemmer，最後再藉由讀入stopword.txt比對來去除stopwords，最後輸出結果。