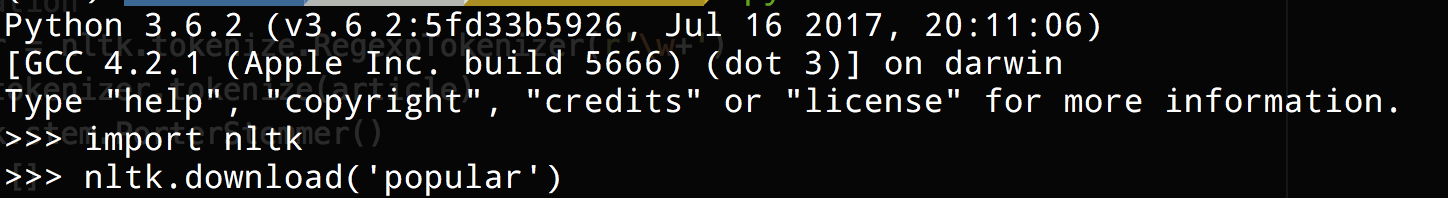
HW1 Documentation 資管碩一 r06725015 李尚恩

1. 執行環境：MacOS，Jupyter & Terminal
2. 程式語言：Python 3.6.2
3. 執行方式：

此程式僅使用到 python 的 nltk 套件，以 pip(pip3) 安裝為例做說明：

1. pip install nltk # 透過 pip 安裝 nltk 套件
2. python # 進入 python
3. import nltk  
   nltk.download(‘popular’) # 安裝 nltk 所需的 data



安裝結束後即可執行 .py 檔，在 command line 上輸入，：

python HW1.py 或 python3 HW1.py 或 python3.6 HW1.py

執行完畢後，可在同個資料夾底下看到 result.txt，而打開後就可以看到處理完的結果。

1. 作業處理邏輯說明：
2. 首先先 import 所需要使用的 python 套件，而在這次的作業中，僅需要 import nltk 套件即可。
3. 接下來，從 nltk 套件中取得英文的 stopwords 與讀入這次作業的短文。這部分我是事先把作業的檔案（28.txt）載下來，再進行 read，而不是直接讀 url 下來。
4. 再來就是進行 tokenization。因為 nltk 預設的 tokenization 會保留標點符號，因此我們透過 RegexpTokenizer 來用正規表達式設定 tokenization 的規則。同時我也先在這個段落引入 PorterStemmer 為下個階段的 stemming 做準備。
5. 對每個斷完的詞進行 stopwords 的篩選，把非 stopwords 的詞小寫化後，丟入 PorterStemmer 進行 stemming，再存入 tokens 中。
6. 最後在同個資料夾底下新增一個 result.txt，並在裡面寫入存在 tokens 中處理完的詞。