1. 執行環境

Mac Command Shell

2. 程式語言

Python(2.7.10)

3. 執行方式

I. 安裝 pip 套件

II. 安裝 nltk 模組

```
~/Desktop/106-1-IR/hw/hw1 sudo pip install nltk
```

III. 執行 hw1.py

(先於某目錄下解壓縮，並於該目錄下輸入以下指令)

```
~/Desktop/106-1-IR/hw/hw1 python ./hw1.py
```

4. 作業處理邏輯說明

主要分成七個步驟(function)

i. GetFile: 取得檔案內容

ii. Tokenize: 用','或' '或'.'作為分隔符號，將原有檔案內容切分為陣列

iii. LowerCase: 用內建 map 方法的每一個字串變為小寫

iv. Stem: 導入 porter module，將傳入的陣列中的每一個字串 stemming 後傳出

v. StopWordRemove: 將傳入的陣列過濾 stopList.txt 檔案的字串後傳出

stopList.txt 檔案來源(http://ir.dcs.gla.ac.uk/resources/linguistic_utils/stop_words)

vi. SpecialCharRemove：將傳入的陣列的每一字串，去除特殊字元後再以陣列形式傳出

vii. ContentSave: 將傳入的陣列用換行字元('\n')結合，並在同目錄下輸出目標檔案：result.txt

七個步驟依序執行即可產出結果：result.txt