

1. 執行環境 Mac terminal

2. 程式語言 Python 3.8.5

3. 執行方式

下載套件：nltk

pip install nltk

python3

import nltk

nltk.download('punkt')

nltk.download('stopwords')

執行指令：

python3 pa1.py data.txt

```
Last login: Sat Oct  3 20:25:03 on ttys000
[lalami@wangpeilindeMacBook-Pro hw1 % python3
Python 3.8.5 (default, Jul 21 2020, 10:48:26)
[Clang 11.0.3 (clang-1103.0.32.62)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> import nltk
>>> nltk.download('punkt')
[nltk_data] Downloading package punkt to /Users/lalami/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
True
>>> nltk.download('stopwords')
[nltk_data] Downloading package stopwords to
[nltk_data]   /Users/lalami/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
True
>>>
zsh: suspended  python3
[lalami@wangpeilindeMacBook-Pro hw1 % python3 pa1.py data.txt
```

結束後會在同個資料夾產生result.txt 的文件

4. 作業處理邏輯說明：

以檔名為參數，讀檔後用 replace 移除標點符號和句點，用 split 作 tokenize，用 lower()變小寫，再用 PorterStemmer function 作 Stemming，最後把不在 stopword（使用 nltk.corpus）的結果寫入檔案。