

Report

1 執行環境

Jupyter Notebook

2 程式語言

Python 3

3 執行方式

使用 Jupyter Notebook 執行，並且把 1.txt 還有 stopwords.txt 放到跟程式碼同個資料夾直接按執行即可。

4 處理邏輯

這份程式碼在讀完檔案之後先把所有文本都轉換成小寫，接著把所有的標點符號都轉換成空格並且把所有單字依照空格切隔成一個一個的 token，接著跑一遍 porter's algorithm 處理 stemming，之後讀取 stopwords.txt 把裡面的單字依照裡面的詞去除 stopwords，最後再把檔案儲存起來。