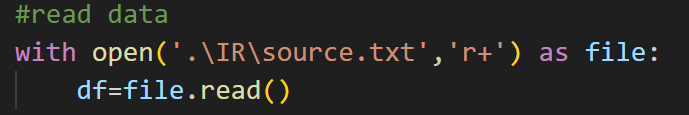
1. 執行環境: VS code
2. 程式語言: Python 3.11.5
3. 執行方式:
   1. 將source.txt與stopword.txt放在名為IR的資料夾中
   2. 在powershell中輸入pip install nltk，以安裝nltk套件
4. 作業處理邏輯
   1. 讀入檔案，命名為df



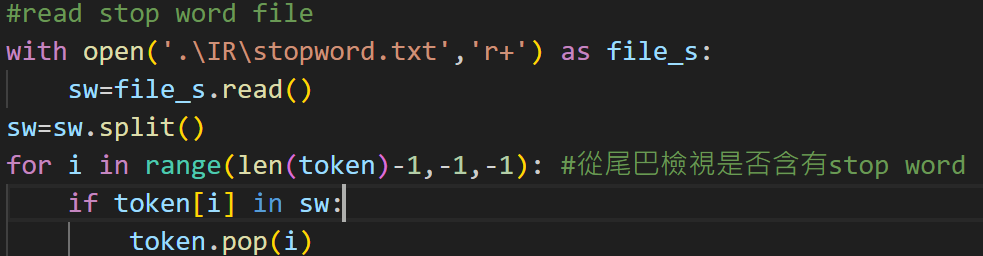
* 1. 將檔案中的大寫字母轉為小寫



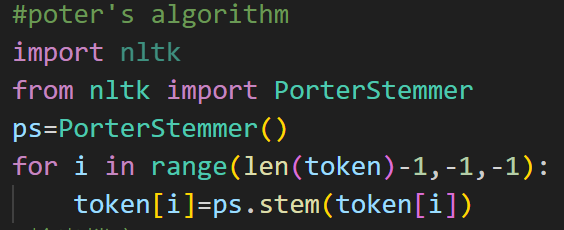
* 1. 建立名為token的空list(後續若指這個list，會在token後面標註)與這時存放字串的sub變數。使用loop逐一跑過df中每個字元，以空格為分割基礎，若字串中含有標點符號或是\n等描述文件格式的內容就忽略。最後將字元放入token(list)當中，token(list)的每個成員就為tokenized token。



* 1. 讀進stopword.txt，此檔案中存放stop word。利用split()函式將stopword.txt中的字串分割成token，存放在名為sw的list中。若token(list)中的成員在sw(list)中有出現，就刪除此成員。



* 1. 使用nltk套件執行porter’s algorithm，利用loop將每個token中的成員轉換成單一形式



* 1. 在相同資料夾中建立一個名為result.txt的檔案，並逐一寫入已經清理好的token(list)內容。

