

IR Programming Assignment#1

B06406009 資管三 陳佩如

執行環境：

Python3 / Anaconda Jupyter Notebook

要先下載 nltk

步驟一：

打開文字檔 (IR.txt) 讀入檔案，存成變數text

步驟二：

把所有字母lowcast，轉成小寫

步驟三：

tokenization，切字串 //用nltk提供的stop word

步驟四：

Stopword removal，把不必要的符號、字母丟掉（自訂）

步驟五：

把 List 存成 string，開個名叫 “the result of document 1.txt” 的檔案寫入。