1. 執行環境：jupyer notebook
2. 程式語言：python 3.13.0
3. 執行方式：學生自己是透過 anaconda 跑 jupyter notebook 的，檔案類型也是 ipynb 檔，然後總共 import 四個套件分別是 os 用於讀檔案，re 用於去除數字 terms，math 用於進行數學運算以及 defaultdict 用於儲存 dictionary。
4. 作業邏輯說明：總共分成三個區塊：第一個區塊是引用作業一的 tokenization 方法，並創建一個計算 df 的函式呼叫 tokenize 並進行各個 terms 在不同 doc 中出現的次數並排序輸出 dictionary 最後存檔。第二個區塊是用於計算 tf-idf 的，前面步驟一樣讀檔案並計算然後根據檔案名稱區分每篇文章，對每篇文章都去計算他們各自的 tf 以及 idf 最後相乘起來在創建他們的 vector_file 直到所有文章都完成。第三個區塊是創建計算餘弦相似度的函式，先用一個空 array 去儲存抓出來的 vector 值 ,然後先分別對分母跟分子去計算各自的向量長度跟內積，接著判斷是否為正交，否則根據相似度的公式去計算最後的相似度。