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。