

Hw4 資管四 B06705048 王佩琳

1. 執行環境： mac
2. 執行: python3 pa4.py 輸出 8.txt 13.txt 20.txt
3. (C.npy 和 doc_vector.npy 需存在)
4. 邏輯：延續作業二將 documents 轉成 tf-idf vector 存成 doc_vector.npy，和儲存 document similarity matrix C.npy，初始化 I 和 A，這裏的 index 是實際編號-1。再按照課堂投影片邏輯寫出 H A C 演算法，使用 complete link 的方式計算 cluster similarity，merge 包含 i 和包含 m 的 cluster 後，透過判斷 cluster j 和 cluster i 及 cluster j 和 cluster m 的 similarity，取較小值（complete link->最遠兩點），更新 C[i][j]，再依照找到的 A 去 merge，merge 到 K 為 20, 13, 8 時寫檔輸出。