

IRTM PA4 report

B11705058 資管三 廖振翔

程式語言: python 3.11.5

執行方式: 在 “./data/” 放入 docs，再來要先 “pip install nltk numpy”，在執行後程式就會開始跑了，跑完後會把結果 8,13,20.txt 輸出在同個目錄下

作業處理邏輯說明:

1. Preprocessing
2. Function of TF DF IDF cosine similarity
3. implementation of (priority queue) Maxheap
4. clustering(complete-link)with priority queue
5. output result