

1. 執行環境：jupyter notebook
2. 程式語言：python 3.11.9
3. 執行方式：
  - `nlTK`：用於文本處理（例如停用詞去除、詞幹提取）。
  - `numpy`：進行數學運算和矩陣操作。
  - `os`：處理檔案與目錄操作。
  - `math`：進行數學運算。
4. 作業邏輯說明：程式的邏輯分為幾個部分。第一部分是數據加載與前處理，從 `IRTM` 資料夾中讀取所有文檔，按文件名排序後加載內容。文本清理過程包括移除特殊字符和標點符號、去除停用詞和進行詞幹提取，以減少冗餘並規範化詞彙表現。

第二部分是特徵生成。首先計算每篇文檔的 `TF` 和文件夾的 `DF`。然後，基於 `DF` 計算每個詞的 `IDF` 值，進一步生成每篇文檔的 `TF-IDF` 向量，並對向量進行正則化，便於後續的相似度計算。

第三部分是 `HAC` 的核心部分。程式使用餘弦相似度公式計算文檔之間的相似度矩陣，並且每次合併相似度最高的兩個 `cluster`。在 `clusters` 數目達到 20、13 和 8 時，分別保存結果到對應的文件中。

最後部分是結果保存，按照文檔編號升序保存每個聚類的內容，`clusters` 之間以空行分隔。