

1091540_02 說明報告

一、程式設計理念：

- 1.用 **vector** 存文件 **id** 和句子，用 **stringstream** 將句子切割成字詞，放在 **vector** 中，再用這些字詞去比對原句子，有出現的 **index** 加一，形成詞頻向量。
- 2.定義 **Data** 類別，包含當前文件 **id**、向量內容、所有文件 **id**、所有向量內容、編號，所有 **Data** 物件放進 **vector**。
- 3.用 **pthread_t** 宣告 **tid**，在 **for** 迴圈中用 **pthread_create()**產生 **thread**，**thread** 中先用 **pthread_self()**印出自己的 **tid**，接著在 **for** 迴圈中用當前向量對其他向量按照公式計算係數，再把它印出來，並用 **vector** 記錄下來，跳出 **for** 迴圈後，將所有係數加起來，除以總數，算出平均後印出，並與 **Data** 類別中的 **static** 變數比較，如果較大則更新 **static** 變數。
- 4.回到 **main** 當中，在迴圈中用 **pthread_join()**結束 **thread**，並印出關鍵文件的 **id** 和平均係數。
- 5.**main thread** 和 **child thread** 都是在開頭 **clock_t** 設定 **start** 變數，結尾設定 **end** 變數，相減後除以 **CLOCKS_PER_SEC**，再乘以 **1000** 後印出。

二、程式如何編譯：

在 **terminal** 輸入 **cd** 檔案路徑，切換到有 **prog2.cpp** 的資料夾
輸入 **g++ prog2.cpp -o prog2 -pthread**

三、程式如何操作：

輸入 **./prog2 data.txt**(或其他測試文件的檔名)
Enter 後程式自動輸出結果

四、完成部分：

1. 從命令列讀入檔名參數
2. 產生正確數量的 **pthread**，計算每個文件的關鍵文件分數，一個 **thread** 只負責一份主文件

- 3.子執行緒印出自己的 tid
- 4.計算文件詞頻向量
- 5.計算餘弦相似係數
- 6.每一個 thread 印出 CPU 時間，以 ms 為單位
- 7.主執行緒找出關鍵文件並印出它的平均餘弦相似係數

五、特殊程式庫：

<iostream>
<algorithm>
<cctype>
<cmath>
<ctime>
<fstream>
<pthread.h>
<sstream>
<string>
<vector>