Prog 2

1091528 江品萱

1. 設計理念:

使用迴圈創建需要的thread，有M個文件，就會產生M個子執行緒。

進行詞頻向量相關計算。

1. 程式如何編譯及執行

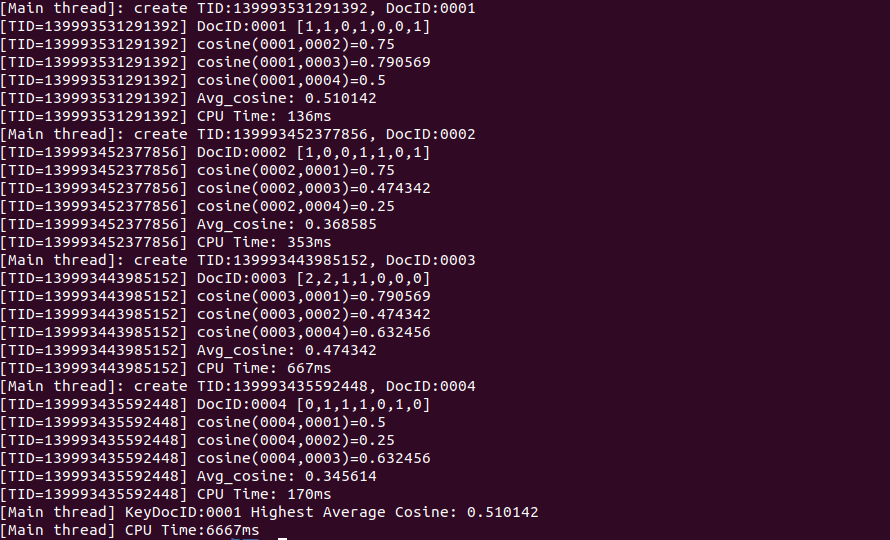
用g++編譯



1. 操作:

使用老師給的test case: 1102-prog2\_data\_linux.txt 進行測試

下圖為執行結果:



1. 程式碼解釋:

下圖為主行程緒與子行程緒共用的資料。

thdata宣告地的物件將用來儲存每個文件各自的資料:

thread\_no:文件編號

message:文件內容

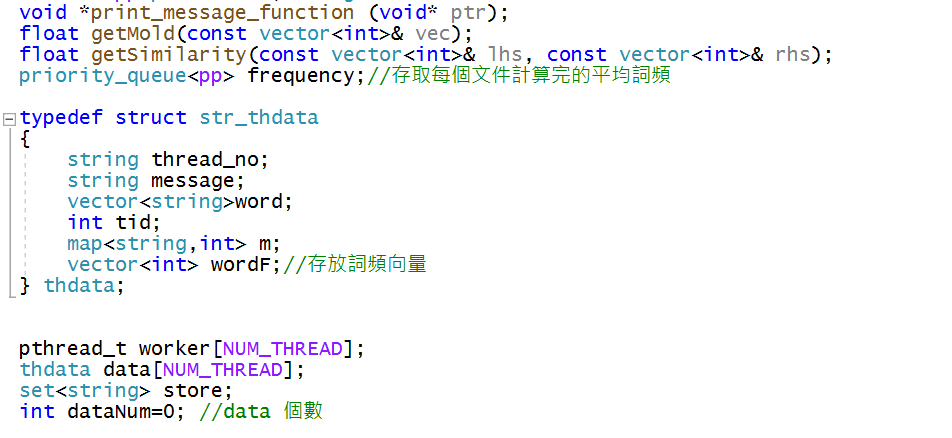
word:儲存有效文件內容

tid:第幾個文件

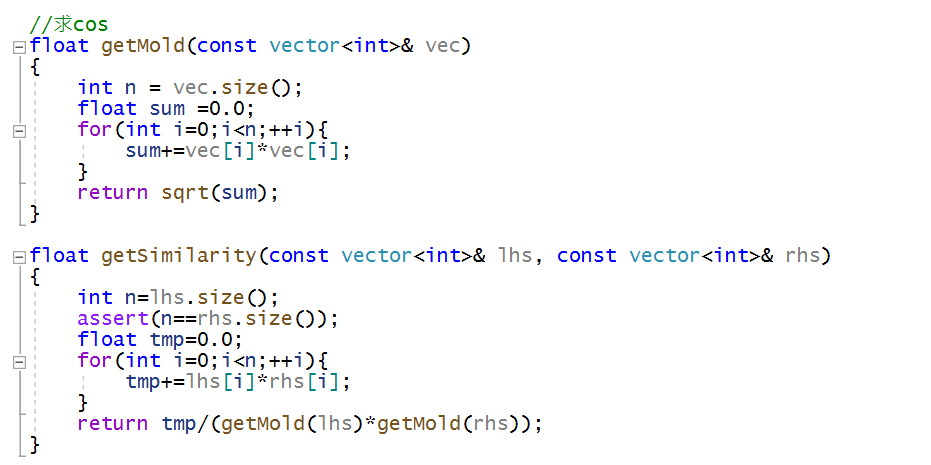
m:計算詞頻向量

wordF:存放詞頻向量

set<string> store:存取所有文建會出現的詞，使用set儲存不會存到重複的詞



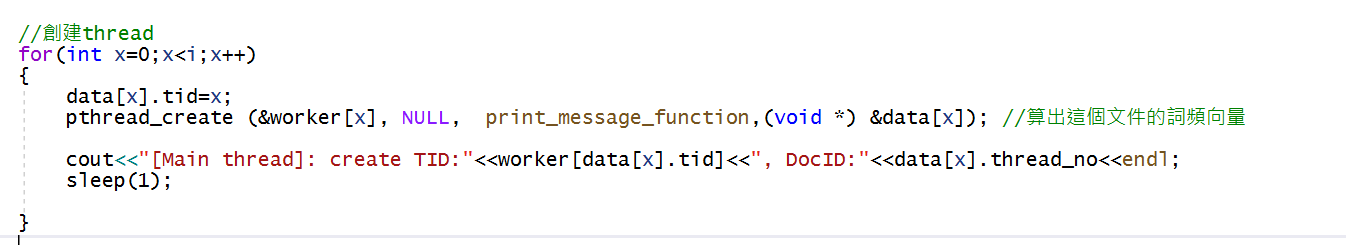
求餘弦相似係數程式碼部分:



1. 完成部分
2. 從命令列讀入檔名參數。本項滿分10分。(完成)



ii. 能產生正確數量的 pthread。例如如果有15份文件，就要產生15個thread，分別對每個文件計算它的關鍵文件分數。一個thread只負責一份主文件。本項滿分20分。(完成)



iii. 子執行緒可以印出本身的tid。本項滿分10分。(完成)

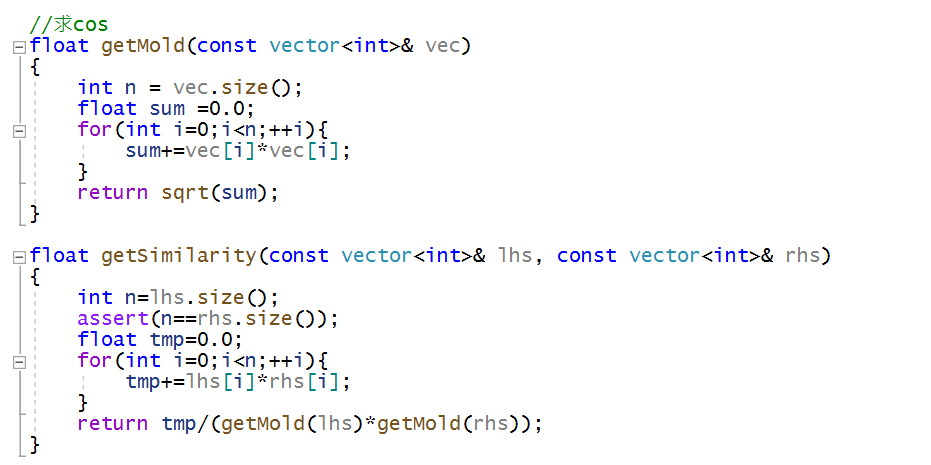


iv. 正確計算出文件的詞頻向量。本項滿分20分。(完成)

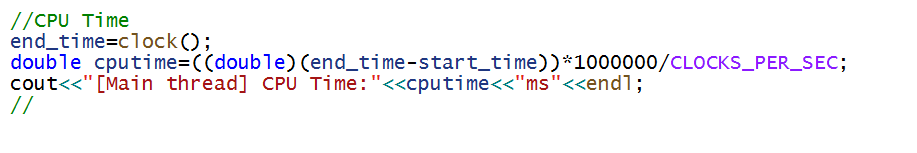


v. 正確進行餘弦相似係數計算。程式碼中不可以使用任何套件或函式庫，

必須有完整的程式碼。本項滿分20分。(完成)



vi. 每一個 thread 都印出執行過程所用的總共CPU時間，以ms為單位。本項滿分20分。注意，是CPU時間。(完成)







vii. 主執行緒找出關鍵文件並印出它的平均餘弦相似係數。本項滿分20分(完成)

