

## OS Prog\_2

學號:1091433 姓名:鍾彥彬

- 程式設計理念

利用多執行緒讓一個 **main thread** 讀完檔案後可以叫出相對應數量的 **thread** 做同步計算

1. 首先用命令列的方式讀檔，把檔名從 **argv** 裡讀出來，接著確認每一行裡面的字串是文件編號或是文件內容
2. 接著分割文件裡面的字串，讓他變成很多個單字分別放進 **vector** 陣列裡面，順便在此時建立一個字典，存放有出現過的字跟這個文件裡單字出現過的次數，這樣之後才能計算詞頻，並全部存進事先創建好的 **structure data** 裡面，**data** 擔任了包裝傳輸資料的功用。
3. 前面的資料處理完成之後開始呼叫 **thread**，用 **pthread\_create()** 讓 **thread** 執行 **print\_message\_function()**，並把 **data** 傳進去利用題目提供的公式計算出詞頻向量，以及平均向量。
4. 當每個 **Thread** 計算完之後，**main thread** 就會開始比對各個文件的平均向量誰是最大的，以及判斷有沒有兩個文件有相同平均向量，這時就接著判斷將編號比較小的當作 **keyDoc** 輸出。
5. **CPU time** 計算:

先在 **main thread** 跟 **thread** 的頭尾各用 **clock()** 取得當下的時間，接著在最後用 **CPU time** 公式算出頭尾相差了多少時間/**c++**內建常數 **CLOCKS\_PER\_SEC** 得出 **CPU time**。

- 編譯方式

參數: 文件檔名.txt

File: 1091433.cpp

Compile: \$ g++ -o 1091433 1091433.cpp -Wall -Werror -lpthread

Execute: \$ ./1091433 filename.txt

- 使用函式庫

```
#include<iostream>
#include<fstream>
#include<string>
#include<vector>
#include <unistd.h>
#include <sys/types.h>
#include <errno.h>
#include <stdio.h>
#include <stdlib.h>
```

```
#include <pthread.h>
#include <string.h>
#include <cmath>
#include <iomanip>
#include <time.h>
#include <sstream>
```

- 完成部分  
完成作業內容所有要求。