

資料結構 程式作業一

Word Frequency Counter

1. 作業描述

計算文章之中每個單字(不包含標點符號、空白及特殊字元)出現的次數，並輸出出現次數前 k 名的單字、其出現次數及執行時間

- I. **DO NOT** use STL. Implement the entire program by yourself.
- II. 演算法不設限，過程必須使用"array"實作
- III. 使用 C/C++ 實作
- IV. 測試平台是 GNU/Linux，gcc version 4.3.2
- V. Tip: 計算時間可以使用 `time.h` 函式或 `unix` 指令
- VI. 本作業不要求使用 `sorting`，但若需要，可自行 implement 任何一種
(助教小叮嚀：使用 `sorting` 不一定比較快)
- VII. 此處單字的定義為單字及數字

2. 輸出入及執行格式要求

I. 輸入(檔案)

文章內容，請依標點符號、空白或換行字元作為分開單字的依據

II. 執行

`./可執行檔 測試檔名 k`

III. 輸出(標準輸出)

將出現次數前 k 名的單字以(單字,次數)的方式輸出；Ex: (the,4)，每一行只輸出一個，並依照順序由多至少輸出，最後輸出總執行時間。

注意事項：

1. 英文大小寫不需區分，Ex: The 和 the 視為相同
2. 出現次數相同視為排名相同，在前 k 名內都必須輸出

Example 1

Input file:
A skunk sat on a stump. The skunk thought the stump stunk; the stump thought the skunk stunk.
Execution:
<code>./9955815_hw1 test24_hw1.txt 2</code>
Output:
(the,4)
(skunk,3)
(stump,3)
0.002ms

Example 2

Input file:
I wish to wish the wish you wish to wish, but if you wish the wish the witch wishes, I won't wish the wish you wish to wish.
Execution:
./9955815_hw1 test36_hw1.txt 3
Output:
(wish,11) (the,4) (to,3) (you,3) 0.003ms

3. 測試資料範例

請至 **e3** 平台下載作業一測試檔，test1_hw1.txt。

注意：此檔案的編碼為 **utf-8**，若標點符號出現亂碼請調整編碼

4. 繳交內容要求

程式

- I. 需繳交**程式碼**、**e3** 下載的**測試檔**與**可執行檔**
- II. 可執行檔請命名為學號_hw1、(Ex: 9955815_hw1)
- III. 程式執行時需搭配參數 (Ex: ./9955815_hw1 test1_hw1.txt 3)
- IV. 程式結果請輸出到標準輸出，並將結果紀錄到你的報告中
- V. 應具有可讀性 (Ex: 註解、變數名稱、函數名稱)
- VI. 請確保 **e3** 平台所提供的測試資料可以執行

報告規定(檔名取為學號_hw1.pdf，Ex: 9955815_hw1.pdf)

- I. 實作方法描述(Ex: 採用何種演算法、程式執行步驟)
- II. 需包含 **e3** 平台上提供的測試資料與執行結果並說明之(必須附上執行時間)
- III. 其他(例如：程式額外實作的項目，執行自行尋找的測試資料之結果)
- IV. 報告以 2 頁為限**

5. 評分標準

I. Program execution (50%)

(適用於非 DEMO 組別)

1. 正確性(20%)
2. 執行速度(30%)
 - 1~10 名，得 30%
 - 11~20 名，得 27%
 - 以此規則類推，每十名以 3%遞減

(適用於 DEMO 組別)

1. 正確性(20%)
2. 執行速度(30%)
 - 1~5 名，得 30%
 - 6~10 名，得 25%
 - 以此規則類推，每五名以 5%遞減

II. Self-test(30%)

自行測試 e3 平台上提供的測資，並記錄執行結果與時間在報告中，k=1, 5, 10

III. Report (20%)

IV. Bonus (10%)

助教將提供一份特殊測試資料，執行結果正確即可獲得 bonus 分數

6. DEMO

此作業將隨機抽選 25 份作業進行 DEMO，該 25 組的 Program execution (50%) 將在 DEMO 當天以另一份測資測試，不需要 DEMO 的組別，助教將直接替你們以其他的測資作測試，以作為 Program execution (50%)評分。

7. 繳交方式 (e3 會準時關閉，盡量避免在最後才上傳)

壓縮你所有檔案，並取名為學號_hw1.rar 或學號_hw1.zip 等(如 9955815_hw1.rar)，上傳前請先確認你的壓縮檔是否含有所有的檔案，並上傳你的壓縮檔到 e3

上傳截止時間：2012.10.24, 23:59

No late upload

8. 注意事項

相互抄襲程式內容、未繳交書面報告、上傳至錯誤修課班級、未符合繳交內容規定、程式無註解者，不予計分。