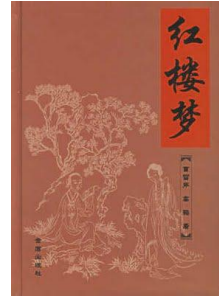


《紅樓夢》是一部古典長篇章回小說，人物甚多：賈寶玉、林黛玉、薛寶釵、王熙鳳、史湘雲、賈探春、賈惜春等，情節十分精彩，是中國四大名著之一。在本作業中，你要設法統計一下《紅樓夢》書中的字頻及人名出現的數量等資訊。



我們已在課程網站上提供一個《紅樓夢》的電子檔 The_Dream_of_the_Red_Chamber.txt，請你自行下載，並自行檢查一下。

任務 A：字頻統計

請寫一個程式 program_A，先讀入《紅樓夢》的電子檔(一律存在 C:/The_Dream_of_the_Red_Chamber.txt)，再累計《紅樓夢》中每個中文字出現的次數，不是中文字(如標點符號、英文、阿拉伯數字、特殊符號)則略過不計。接著依出現次數由多至少，將書中所有的單字排序。最後產生的輸出檔案(一律存在 C:/OUTPUT_A.TXT)內容如下 (這是教育部的統計，你的結果將不一樣，但需列出書中所有的單字)。另外，最後還要輸出這本書共有多少個(=n)中文字，並顯示你的程式執行時總共花費多少時間 T(n)。

字頻序號	字	出現頻次
1	的	32739
2	不	24362
3	一	22524
4	我	19414
5	是	17698
6	人	17638
7	有	14346
:	:	:

任務 B：人名統計

請寫一個程式 program_B，先讀入《紅樓夢》的電子檔(一律存在 C:/The_Dream_of_the_Red_Chamber.txt)，再找出《紅樓夢》中「人名」出現次數最多的 200 個人名，依出現次數，由多至少列出前 200 名。最後產生的輸出檔案(一律存在 C:/OUTPUT_B.TXT)內容如下 (這是一個範例，你的結果將不一樣)。另外，最後還要顯示你的程式執行時總共花費多少時間 T(n)。

人名序號	人名	出現頻次
1	賈寶玉	2273
2	林黛玉	1436

3	薛寶釵	1252
4	王熙鳳	941
5	史湘雲	769
6	賈探春	176
7	賈惜春	143
:	:	:

注意：此處的「人名」，其字數為 2~4，指的是含「姓」及「名」的全名。

這個作業要求你完成下列多元的目標：(請從 1~10 逐項寫在你的作業報告中)

- (1) 設法搜集網路上對字頻及人名統計的介紹，已有那些分析的結果了？有沒有人已寫出程式了？請你自行整理並重新說明清楚，也記得列出這些資訊的出處。(你如何搜集及撰寫？)
- (2) 撰寫一份任務 A 的虛擬碼(pseudo-code)出來並仔細說明你的演算法。(你如何撰寫？)
- (3) 撰寫 program_A 的程式出來。請提供你的**原始碼**、**執行檔**及**如何執行你的程式的說明**。原始碼中需有**註解輔助說明**。(你如何寫程式？)
- (4) 試分析 program_A 的 time complexity。請附上答案及說明。(你如何分析？)
- (5) 請做一些實驗，設法繪出 program_A 的不同 n 值(可使用字數不同的小說)範圍的 T(n)時間成長統計圖表，並請加以說明成長的函數看來大約為何。注意：請詳細說明你所使用之機器軟硬體規格以及你為何選擇這樣的規格。(你如何估算？)
- (6) 撰寫一份任務 B 的虛擬碼(pseudo-code)出來並仔細說明你的演算法。你如何找出「人名」？這部份你用了甚麼辦法？
- (7) 撰寫 program_B 的程式出來。請提供你的**原始碼**、**執行檔**及**如何執行你的程式的說明**。原始碼中需有**註解輔助說明**。(你如何寫程式？)
- (8) 試分析 program_B 的 time complexity。請附上答案及說明。(你如何分析？)
- (9) 請做一些實驗，設法繪出 program_B 的不同 n 值(可使用字數不同的小說)範圍的 T(n)時間成長統計圖表，並請加以說明成長的函數看來大約為何。注意：請詳細說明你所使用之機器軟硬體規格以及你為何選擇這樣的規格。(你如何估算？)
- (10) 請說明你做此作業所碰到的一些狀況及困難。
- (11) 注意：此作業**要求快速的計算，愈快的程式愈好**，而且要求你的程式所使用的記憶體(含程式本身)最多為 1G bytes，你需說明你如何估算記憶體所佔用的大小。
- (12) 本作業沒有規定的部份，你可自由發揮。

請將上述任務之**作業報告**、**原始碼**、**執行檔**及**如何執行你的程式的說明**用 e-mail 交給陳志宏助教(e-mail：honhonzon@gmail.com)。另外請提供**你的連絡電話**及**e-mail**，以備不時之需。

本課程允許遲交作業(三天內)，遲交的話，一天內約扣 3 分，二天內約扣 6 分，三天內約扣 9 分。

提醒：程式及報告如有抄襲狀況(不管被抄或抄別人)，一律依學校規定處理。