

資料檔簡述

以二進位格式存檔，檔名如 **paris501.bin**，每列資料表示一筆紀錄，3 個欄位如下：

- [發訊者學號 putID]發訊學生的學號以 10 個字元的陣列表示。
- [發訊者學號 getID]收訊學生的學號以 10 個字元的陣列表示。
- [量化權重 weight]訊息量以浮點數 float 儲存，介於(0,1]之間的正實數。

必須遵守的原則

1. 預先不知道資料筆數，禁用固定大小的陣列，必須採用動態陣列或向量 **vector** 型別。
2. 每項任務只能分批處理檔案的資料，**嚴禁將所有資料一次載入記憶體或寫入硬碟!(雜湊表除外)**

(任務一) 外部排序 external sort

輸入：如上述的二進位檔，每列資料表示一筆紀錄

參數：外部排序所需要的緩衝區上限以資料筆數表現，一律固定為 200 筆。權重大小為 1~小數點 2 位。

步驟：

1. 以**外部合併排序方法**將互動關係資料依照[量化權重]由大到小排序，將排序結果寫成另一個二進位檔，權重相等時，則保持在**原始檔案的次序**。
2. 測量**整體的執行時間**，其中包括讀寫檔案的時間。禁止將所有資料一次載入記憶體進行內部排序，違反者視同未完成。

輸出：

1. 依照[量化權重]由大到小輸出每筆資料的 3 個欄位值至檔案，以排序檔名為 **sorted501.bin**，其檔案大小應該和原始輸入檔相等。
2. 顯示整體的執行時間於螢幕上。

(任務二) 建立雜湊表 Build Hashtable

輸入：一個權重值。

參數：讀取檔案所需要的緩衝區上限以資料筆數表現，一律固定為 200 筆。**(雜湊表並無資料筆數的限制)**

步驟：

1. 將任務一得到的檔案**依照權重做分割檔案**，檔名為該檔案以及權重*100。檔名如: **sorted501_100.bin**(權重為 1)
2. 輸入權重(輸入為 0 此任務結束，並且將步驟 1 切割出的所有檔案刪除)。
3. 以權重所對應的檔案建立雜湊表，**若發生碰撞，以 Quadratic probing 的方式處理**。若有資料無法放入雜湊表，需另存資料於陣列(Garbage)中。

輸出:

1. 將得到的雜湊表輸出，檔名為該檔案以及權重*100。檔名如: HashTable501_100.txt。
2. 若 Garbage 中有資料，要建立檔案，檔名如: Garbage501_100.txt。
3. 將相同權重的 bin 檔轉為 txt 檔方便閱讀，檔名如: sorted501_100.txt。

(任務三) 運用雜湊表搜尋資料

輸入: 學生學號。

步驟:

1. 輸入學號。(輸入為 0 結束此任務)
2. 運用任務二建立好的雜湊表得到資料在檔案的位置。
3. 直接輸出該位置的資訊，不可從檔案的頭開始找資料的所在位置。

輸出: 將符合該學號的資料全部輸出。