一、規格要求，違反者以零分計！

(a) 以 Dev-C++或 Code::Blocks 編譯與成功執行的 C/C++程式碼(.cpp/.c/.h/.hpp)，要有註解。

(b) 任何一部分的程式碼都不得被偵測為抄襲。

(c) 檔名限以「**DS2ex#_組別_學號_學號**」開頭，**兩人一組只限繳交一份**。

二、作業內容

整合下列任務在一個系統選單下，未整合、無法連續執行或沒有輸入防呆措施，都各扣 5 分。若影響任務執行，該任務以零分計。

資料檔簡述：

以二進位格式存檔，檔名如 pairs501.bin，每列資料表示一筆紀錄， 3 個欄位值如下：

-【發訊者學號 putID】發訊學生的學號以 10 個字元的陣列表示。

-【收訊者學號 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：(每個任務違反一項各扣 5 分)

1. 預先不知道資料筆數，禁用宣告固定大小的陣列，必須採用動態陣列或向量 vector 型別。
2. 每項任務都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！

(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：外部合併排序所需要的緩衝區上限以資料筆數表示，一律固定為 200 筆。

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，則保持在原始檔案內的次序。
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，已排序檔名為 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。

繳交項目：

- **流程圖**：上機一週前上傳至**同儕互評**，上機時修正後寫入程式說明文件。
- **程式碼**：上機三天前上傳原始碼至**作業**，程式碼首列要註解學號、姓名和系級。

(任務二) 建立主索引 primary index

輸入：只限使用任務一產生的已排序檔。

步驟：

1. 以【自認為最有效率的方法】針對【量化權重】為已排序檔建立主索引，必須能處理權重可能相等的狀況。
2. 主索引和排序檔是分開的，主索引只記錄【量化權重】及對應資料的【檔案位址】(位移量 offset)。
3. 禁止將所有資料一次全部載入記憶體，必須分批讀取檔案，違反者視同未完成！

輸出：依序顯示主索引的每筆紀錄於螢幕上，包括【量化權重】和【檔案位址】，每筆紀錄前附上一個從 1 開始的序號。

繳交項目：

- **流程圖**：上機一週前上傳至**同儕互評**，上機時修正後寫入程式說明文件。
- **程式碼**：上機三天前上傳原始碼至**作業**，程式碼首列要註解學號、姓名和系級。