

## 資料結構

### 期中考

#### 程式設計：由五小題任選四小題

1. 以下為王維的「山居秋暝」詩句，但多了些重複的贅字：

空空空山山新雨後後，天天氣晚晚晚晚來秋。  
 明月月月月松間間照，清清清泉石石上上流流。  
 竹竹喧歸歸歸浣女女，蓮蓮動下下下漁舟舟。  
 隨隨意春春芳歇歇，，王王孫孫自自可可留。

若詩句是以 UTF-8 編碼型式存入檔案，撰寫程式讀入詩句使用 `unique` 泛型函式去除重複的贅字後印出。

2. 若要使以下的程式運作，請設計缺少的程式：

```
string foo = "be right back in 10 mins" ;

istringstream istr(foo) ;
istream_iterator<string> iter(istr) ;

vector<string> bar ;

// 僅儲存英文單字
copy_if( iter , istream_iterator<string>() , back_inserter(bar) , Eng_Word() ) ;

// 依英文單字數量排序，由字少到多排序
stable_sort( bar.begin() , bar.end() , By_Length() ) ;

// 列印出：be in back mins right
copy(bar.begin(),bar.end(),ostream_iterator<string>(cout," " ) ) ;

cout << endl ;
```

3. 讀入李清照的「聲聲慢」詞檔，在每一列的右側印出已重複出現的中文字與其出現的次數，重複字不包含標點符號，有關中文字的辨別請見第四題的說明。例如：

尋尋覓覓，冷冷清清，	尋 [2] 覓 [2] 冷 [2] 清 [2]
淒淒慘慘戚戚。	淒 [2] 慘 [2] 戚 [2]
乍暖還寒時候，最難將息。	
三杯兩盞殘淡酒，怎敵他、晚來風急。	
雁過也，正傷心，卻是舊時相識。	時 [2]
滿地黃花堆積。憔悴損，如今有誰堪摘。	
守著窗兒，獨自怎生得黑。	怎 [2]
梧桐更兼細雨，到黃昏點點滴滴。	黃 [2] 點 [2] 滴 [2]
這次第，怎一個愁字了得！	怎 [3] 得 [2]

4. 某資料檔包含 unicode 碼各漢字的筆畫資料，若某列的第二欄為 kTotalStrokes，則其前一欄為此字在 unicode 的編碼位置，以 U+xxxxxxx，U+ 之後的位置數字以十六進位表示，第三欄的數字為此字的筆畫數，例如：

```
U+4E00 kTotalStrokes 1
```

U+4E00 代表中文字：「一」，筆畫數為 1 畫。在 unicode 編碼的順序中，漢字是介於 [U+4E00,U+9FA5]，以十進位表示則是介於 [19968,40869]，在此範圍的漢字包含了傳統中文字與大陸地區所使用的簡體字，但並不包含用於漢字句子間的標點符號。請撰寫程式讀入此資料檔，將所有漢字的筆畫資料存於適當的(複)映射型別，然後作以下的題目：

- (A) 將所有中文字依筆畫序印出來，每列印出 25 個字，且最後依數量由多到少印出筆畫與其個數的統計資料。以下為部份輸出內容：

```
1 畫：10
一 | 丶 | 丿 | ...
...
3 畫：98
万 | 丈 | 三 | 上 | 下 | 丌 | 与 | 个 | 丫 | ...
...
```

以下為筆畫與其個數的統計資料：

```
12 畫：1957 個
11 畫：1859 個
13 畫：1741 個
10 畫：1711 個
...
2 畫：44 個
27 畫：44 個
28 畫：26 個
1 畫：10 個
29 畫：10 個
30 畫：10 個
...
```

- (B) 讀入一個成語檔案，請依各成語的筆畫數由筆畫少到筆畫多的方式印出各個成語。為便於印證起見，在輸出的成語後請印出各個字的筆畫數，輸出型式如下：

```
...
13 畫：
愚公移山：13|4|11|3
勢如破竹：13|6|10|6
塞翁失馬：13|10|5|10

15 畫：
暴虎馮河：15|8|12|8

16 畫：
燃膏繼晷：16|14|20|12
...
```