

程式設計 (105-2)

作業十一

作業設計：孔令傑
國立臺灣大學資訊管理學系

繳交作業時，請至 PDOGS (<http://pdogs.ntu.im/judge/>) 為第一、二、三題各上傳一份 C++ 原始碼 (以複製貼上原始碼的方式上傳)。第三題是 bonus 加分題。每位學生都要上傳自己寫的解答。不接受紙本繳交；不接受遲交。請以英文或中文作答。

這份作業的截止時間是 **2017 年 5 月 29 日凌晨一點**。為這份作業設計測試資料並且提供解答的助教是林敬傑 (Jack Lin)。

第一題

(50 分) 很多網頁是用 HTML 這種程式語言寫的，而不純用 HTML 寫的網頁也大多都有用到 HTML 的相關技術或架構。關於 HTML 的相關資訊，大家可以自行上網搜尋¹。

舉例來說，這門課的課程網站 (<http://www.im.ntu.edu.tw/~lckung/courses/PD17/>)，也是用 HTML 寫的。從未看過 HTML 程式碼的同學，可以打開課程網站，然後按滑鼠右鍵選「檢視網頁原始碼」²，就會看到密密麻麻的程式碼。細節我們就不多談了，今天我們就跟大家介紹用 HTML 程式碼在網頁上製作表格的方式³。我們用下方的程式碼當例子：

```
<table>
  <tr>
    <td>Topic</td>
    <td>Lecture Slides</td>
    <td>Lecture Programs</td>
  </tr>
  <tr>
    <td>Introduction</td>
    <td>Slides</td>
    <td>Programs</td>
  </tr>
  <tr>
    <td>Selection & Repetition</td>
    <td>Slides</td>
    <td>Programs</td>
  </tr>
</table>
```

¹例如 <https://en.wikipedia.org/wiki/HTML>。

²如果你不是用 Windows 系統上的 Chrome，或許要使用不同的方式檢視。

³課程網頁上的表格套用了稍微進階一點的格式，讓它比較美觀，我們在這份作業就不談了。

如果你複製上方程式碼，貼到一個純文字檔裡面，然後存檔時把副檔名存成.html，接著用瀏覽器開啟，就可以看到一個（很醜的）表格了。

就算你從未學過 HTML，看著上方的程式碼，應該也能猜想它在幹嘛。成對的 `<table>` 和 `</table>` 之間建立了一個「表格環境」，接著裡面的每一對 `<tr>` 和 `</tr>` 之間是一列（row），而裡面的每一對 `<td>` 和 `</td>` 之間是一個格子。所以上方的程式碼會產生一張三列三欄（column）的表格。

在本題中，你將會得到一個依照上方結構被定義的 HTML 表格程式碼，你的任務是去解讀給定的 HTML 程式碼，並且將其中的資訊萃取出來，並做相對應的運算。在資訊科學的世界裡，這樣的工作叫做 parse 或 parsing，而你要寫的程式叫做一個 parser。當你用 Notepad、Notepad++ 等文字編輯器開啟一個 HTML 檔時，你看到的是純文字（原始碼），但當你用瀏覽器打開時，你卻是看到表格。很顯然地，在你每天使用的瀏覽器裡，就有 HTML parser 在。

輸入輸出格式

系統會提供許多筆測試資料，每筆測試資料裝在一個檔案裡。一個檔案裡面會按照題目範例的方式儲存一張 HTML 的表格，第一行一定是 `<table>`，第二行一定是 `<tr>`，第三行一定是一對 `<td>` 和 `</td>` 中間包著相對應的資訊，依此類推。表格一定是 5 個資料列、一共 3 欄，加上標頭列的話則是 4 列，而且每一列一定都有恰好 3 個格子。因此，檔案裡面一定是恰好 6 對 `<tr>` 和 `</tr>`，而每一對 `<tr>` 和 `</tr>` 之間一定是恰好 5 對 `<td>` 和 `</td>`。每一個 `<table>`、`<tr>`、`</tr>` 或 `</table>` 都會自己獨立一行；每一對 `<td>` 和 `</td>` 都會被放在同一行。換句話說，整個檔案一定是恰好 $2 + (2 + 3) \times 6 = 32$ 行。在 `<tr>` 前面一定有兩個空格，在 `<td>` 前面一定有四個空格。

在第一對 `<tr>` 與 `</tr>` 之間，記錄的是欄位名稱，一個欄位名稱一定是一個大寫英文字母，欄位名稱不會重複。在標頭列之外的 10 個資料列裡面存的都是介於 0 到 999 之間的整數。

給定一筆測試資料，請將之拆解並以表格的方式在你的程式中記錄標頭列與資料列的內容。接著請對每個欄位個別地計算該欄位所有數字的總和，然後將所有欄位依照該欄總和由大到小排序以輸出欄位名稱。如果有兩個欄位總和一樣，就先輸出字母順序在前的欄位名稱，兩個欄位名稱之間用一個空白鍵隔開。舉例來說，如果輸入是

```
<table>
  <tr>
    <td>J</td>
    <td>F</td>
    <td>D</td>
  </tr>
  <tr>
    <td>10</td>
    <td>12</td>
    <td>10</td>
  </tr>
  <tr>
    <td>1</td>
    <td>2</td>
    <td>3</td>
```

```
</tr>
<tr>
  <td>3</td>
  <td>2</td>
  <td>1</td>
</tr>
<tr>
  <td>5</td>
  <td>5</td>
  <td>5</td>
</tr>
<tr>
  <td>7</td>
  <td>7</td>
  <td>7</td>
</tr>
</table>
```

則輸出應該是

```
F D J
```

你上傳的原始碼裡應該包含什麼

你的.cpp 原始碼檔案裡面應該包含讀取測試資料、做運算，以及輸出答案的 C++ 程式碼。當然，你應該寫適當的註解。針對這個題目，你**不可以**使用上課沒有教過的方法。

評分原則

這一題的其中 50 分會根據程式運算的正確性給分。PDOGS 會編譯並執行你的程式、輸入測試資料，並檢查輸出的答案的正確性。前 40 分由 20 筆測試資料判定分數，一筆測試資料佔 2 分；後 10 分由 5 「組」測試資料判定分數，每一組裡面有若干筆測試資料，全對的話才能得到 2 分。

第二題

(50 分) 承上題，我們繼續處理 HTML 的表格。在上面的例子中，我們看到了三種「標籤」：`<table>`、`<tr>` 和 `<td>`。在 HTML 語言中，還有許許多多其他的標籤。舉例來說，在一對 `` 和 `` 中間，字會變粗體，而在一對 `<i>` 和 `</i>` 中間，字會變斜體。另外，我們可以用 `<a>` 和 `` 來插入超連結 (hyperlink)，方法是用一對 `<a>` 和 `` 包住你要加入超連結的文字，然後在 `<a>` 裡面加入 `href` 選項，然後用等號和雙引號來指示連結目標。我們在下面的例子裡展示新介紹的三種標籤：

```
<table>
  <tr>
```

```

    <td><b>Topic</b></td>
    <td><b>Lecture Slides</b>
    <td><b>Lecture Programs</b></td>
<tr>
    <td>Introduction
    <td><i>Slides</i></td>
    <td>Programs</td>
</tr>
<tr>
    <td>Selection & Repetition</td>
    <td><a href = "www.im.ntu.edu.tw">Slides</a></td>
    <td>Programs</td>
</table>

```

若你試著用瀏覽器開啟上方的程式碼，就會看到標頭列的欄位名稱是粗體、第一個資料列的 *Slides* 字樣是斜體，而二個資料列的 Slides 字樣是超連結形式，點下去會連到臺大資管系的官方網頁。

身為一個程式設計師，你應該可以想像人們寫程式有時候是會出錯的。舉例來說，很多人會漏掉應該跟 `<tr>` 配對的 `</tr>`，當然也有人會漏掉應該跟 `<td>` 配對的 `</td>`。比如說下面的程式碼：

```

<table>
  <tr>
    <td><b>Topic</b></td>
    <td><b>Lecture Slides</b></td>
    <td><b>Lecture Programs</b>
  <tr>
    <td>Introduction
    <td><i>Slides</i></td>
    <td>Programs</td>
  </tr>
  <tr>
    <td>Selection & Repetition</td>
    <td><a href = "www.im.ntu.edu.tw">Slides</a></td>
    <td>Programs</td>
</table>

```

就漏掉了兩個 `</td>` 和兩個 `</tr>`，但你如果用瀏覽器開啟，就會看到表格還是好好地呈現了。這可以被理解為你的瀏覽器的 parser 比較「強」（或者「雞婆」），它會在你出了一些小錯的時候自行修正。當然如果你出了很大的錯，例如漏掉第二個雙引號，那用編譯器打開的時候就真的會看起來有點問題了⁴。

在本題中，我們將請你擴充你在第一題寫的 parser，並且把它變成一個「貼心」（或「雞婆」）的

⁴有些時候，我們會發現有些網頁在不同的瀏覽器上瀏覽時會有不同的長相。這有很多原因，不過其中一個原因就是各瀏覽器「雞婆」的程度不一樣。如果一個網頁是完全按照世界通用的規範寫的，例如沒有漏掉任何標籤，那用各種瀏覽器打開來原則上就會一樣；但如果網頁本身沒有完全遵守規範，不同瀏覽器就可能顯示不同的結果了。

parser。首先，請讓你的 parser 可以處理 ``、`<i>` 跟 `<a>` 這幾種標籤。為了簡單起見，本題中這三種標籤內部不會包著其他任何標籤（也就是不會出現類似 `<i>abc</i>` 之類的東西）。除此之外，如果一個 `<tr>` 後面沒有相對應的 `</tr>`，你就要在下一個 `<tr>` 或 `</table>` 所在的位置上插入一個 `</tr>`；如果一個 `<td>` 後面沒有相對應的 `</td>`，你就要在下一個 `<td>` 或 `</tr>` 或 `</table>` 所在的位置上插入一個 `</td>`。都插好之後，就可以按照正常的方式做 parsing 了以及第一題要求的運算了。

說明：所謂「在某某位置插入一個什麼什麼」，只是為了說明方便而使用的一個概念。你可以選擇任何你想要的作法，並不一定要真的修改輸入的資料。

輸入輸出格式

系統會提供許多筆測試資料，每筆測試資料裝在一個檔案裡。檔案格式和第一題基本上一模一樣，但現在會出現以正確方式配對和標記的 ``、`<i>` 和 `<a>`，並且有一些 `<tr>` 和 `</td>` 可能會被漏掉。

給定一筆測試資料，請將之拆解並以表格的方式在你的程式中記錄標頭列與資料列的內容。在這一題中，請忽略所有的 ``、`<i>` 和 `<a>`，而只把資料列中的數字記錄下來，然後做如第一題般的計算與輸出。舉例來說，如果輸入是

```
<table>
  <tr>
    <td><b>J</b></td>
    <td>F
    <td>D</td>
  <tr>
    <td><i>10</i></td>
    <td>12</td>
    <td>10</td>
  </tr>
  <tr>
    <td>1</td>
    <td><a href = "...">2</a></td>
    <td>3
  <tr>
    <td>3
    <td><a href = "<td>2</td>">2</a></td>
    <td>1</td>
  </tr>
  <tr>
    <td>5
    <td>5
    <td>5
  <tr>
    <td>7</td>
    <td>7</td>
    <td>7
```

```
</table>
```

則輸出應該是

```
F D J
```

你上傳的原始碼裡應該包含什麼

你的.cpp 原始碼檔案裡面應該包含讀取測試資料、做運算，以及輸出答案的 C++ 程式碼。當然，你應該寫適當的註解。針對這個題目，你**可以**使用任何方法。

評分原則

這一題的其中 50 分會根據程式運算的正確性給分。PDOGS 會編譯並執行你的程式、輸入測試資料，並檢查輸出的答案的正確性。前 40 分由 20 筆測試資料判定分數，一筆測試資料佔 2 分；後 5 分由 5 「組」測試資料判定分數，每一組裡面有若干筆測試資料，全對的話才能得到 2 分。

第三題 (bonus)

(20 分) 承上題，現在輸入的格式又更彈性了，表格可能會有 n 列和 m 欄，而 $1 \leq n \leq 20$ 、 $1 \leq m \leq 10$ 。每一列都會有恰好 m 欄。其次，縮排不一定是 `<tr>` 兩格、`<td>` 四格，而是任何數量的空白字元都有可能。``、`<a>`、`<i>` 這幾個標籤裡面依然不會包別的標籤。

讀入資料後，請根據 ``、`<i>` 和 `<a>` 為資料列中的數字做加權運算：被粗體的數字乘以二、被斜體的數字乘以三，而被設為超連結的數字乘以五（不論它要連到哪裡去）。在每欄中將這些被乘以不同權數的數字加總後，再根據第一題的規則將欄位名稱排序後印出。

你的.cpp 原始碼檔案裡面應該包含讀取測試資料、做運算，以及輸出答案的 C++ 程式碼。當然，你應該寫適當的註解。針對這個題目，你**可以**使用任何方法。

這一題的 20 分會根據程式運算的正確性給分。PDOGS 會編譯並執行你的程式、輸入測試資料，並檢查輸出的答案的正確性。前 10 分由 5 筆測試資料判定分數，一筆測試資料佔 2 分；後 10 分由 5 「組」測試資料判定分數，每一組裡面有若干筆測試資料，全對的話才能得到 2 分。