

題號： 141

國立臺灣大學 108 學年度碩士班招生考試試題

科目： 資訊科學與新聞

題號：141

節次： 4

共 1 頁之第 1 頁

1. 假設現在你收到數萬則 PTT 的貼文資料與該貼文的回文資料如下。請 (1) 設計一內含多個表格的關聯式資料庫 (以 E-R Diagram 來描繪) 紀錄下述資料。其次, 說明你如何利用正規化的技術來降低資料的重複性 (Redundancy)? 又如何利用 primary key 與 foreign key 來確保資料的完整性? (2) 此時, 你打算建立網民的 Profile 來偵測其實否為網軍, 當你要查詢所有資料中的個體分別有貼文過多少次、回文過多少次、曾經出現在多少個 ID 時, 就你前述的資料表設計, 你要如何設計 SQL? 而你認為這樣的網軍定義是否適切? 為什麼? (15%+15%)

作者 x09xxx13 (x09xxx13) 看板 Gossiping

標題[問卦] 請問這是什麼機器

時間 Fri Jan 11 15:54:41 2018

各位週末下午不想上班的宅宅好 宅宅我啊 昨天路上看到一台很詭異的機器 請問有人看過嗎

-- ※ 發信站: 批踢踢實業坊(ptt.cc), 來自: 36.111.147.208 ※ 文章網址:

<https://www.ptt.cc/bbs/Gossiping/M.154711112.A.FB6.html>

→ hsnxxxxxn:這世界上只有 10 種人 懂二進位的和不懂二進位的 10/27 00:13

-- ※ 發信站: 批踢踢實業坊(ptt.cc), 來自: 223.124.73.41 ※ 文章網址:

<https://www.ptt.cc/bbs/Gossiping/M.1547283914.A.39A.html>

噓 hate0504: 五樓知道 01/11 15:55

推 okbong1: hate0504 的數學老師 XD DDDDDDDDDDDDDDDDDDDDDDDDDDDDD 01/11 15:55

2. 網路上的新聞經常在刊出後會被新聞編輯從後台再次修改, 比如傷亡人數更新, 或者因為被網友抗議而修改其一些標題或內容。無論新聞編輯的動機為何, 顯然在不同時間所看到的網路新聞標題或內文可能是不一樣的。你現在被要求設計一個爬蟲程式, 想大量抓取新聞資料來偵測標題或內容在上線後被更動的情形, 請問 (1) 你計畫如何儲存該資料, 以便於你在日後能夠比較有哪些新聞標題曾經被修改過? 這些資料筆數可能相當龐大, 爬蟲也要爬很久, 請考慮時間、空間與計算複雜度等因素, 試問, 你打算用哪種檔案或資料庫儲存之? 為什麼? 就你所選的格式, 你會設計哪些資料欄位來便於你後續比較新聞的差異? 又 (2) 在你這樣的設計架構下, 你今天要寫一個函式比較某一則新聞的不同版本有哪些差異字時, 你所設計的演算法為何? 計算複雜度為何? (15%+15%)
3. 請問你過去印象最深刻的一則資料新聞報導為何? 他可能運用到哪些資料或技術? 你認為資料可能包含哪些欄位? (10%)
4. 你現在設計好了一則資料新聞, 你希望能夠知道使用者的閱覽行為, 我們可以把它稱為使用者評估或者使用者經驗評估。請問, 如果是你, 你知道有哪些方式可以紀錄、調查各種類型的使用者經驗。除描述方法之外, 請對其優點與限制進行簡略說明。(15%)
5. 排序: 現有一資料如下, 請你用一種你熟悉的程式語言, 直接把該筆資料用你認為適合的變數型態/資料結構打在程式碼內, 並指向一個變數。請簡單設計一個排序函式, 可以根據某一行進行排序所有的資料, 例如根據從小到大的 M96 數值調換資料列的先後順序。(15%)

country	M95	M96	M97
Albania	5	5	5
Algeria	3	3	3
Argentina	2	2	2

試題隨卷繳回