

《資料處理》

試題評析

本次試題的第一題至三題，為資料處理的常見題型，同學們應該要早已有所準備。而四、五題之前半段皆較單純，但後半段考的是較深入的技術或語法，同學們回答上可能較為辛苦。故估計一般同學應可在60分上下，程度較佳者可拿到70分。

一、以2個位元組 (byte) 的長度來編碼，最多可以編出多少不同的碼？若要表示正負整數，以最顯著位元0表示正數，1表示負數，其餘以一般二進位表示的編碼方式有何不妥？2的補數法中，2個位元所能表示的最大數是多少？霍夫曼碼 (Huffman code) 技術常被用來作為符號的編碼，請問其編碼方式與ASCII碼有何不同之處？(20分)

答：

(一) 2 bytes = 16 bits,

16 bits 共可編出 2^{16} 個不同的碼。

(二) 最顯著位元0表示正數，1表示負數，則缺點為零亦分為正負零，採用不同編碼。

(三) 2補數法中，2進位數值之第一位為符號位元，0為正數，1為負數。將二進位正數值之第一位元改為1，其餘位元以其2補數表示，則為其負數。16個位元所能表示的範圍為 -2^{15} 至 $2^{15}-1$ ，故最大數為 $2^{15}-1$ 。

(四) 霍夫曼碼的編碼方式不同於ASCII之處為，ASCII的編碼方式下，每個符號所佔用的編碼長度皆相同；但霍夫曼碼則是依照符號出現機率，在對資訊源符號逐一編碼條件下 (The symbols be coded one at a time)，最佳之編碼方法。出現頻率高之符號佔用長度較短，頻率低之符號則較長，故可用以達到無失真資料壓縮的目的。

二、假設一關聯式資料庫中存在有三個資料表，分別為StudentInfo、CourseInfo以及GradeInfo。

各表格的欄位如下：StudentInfo (stuID, stuName, address, phoneNo)、CourseInfo (courseID, courseName, teacherID) 及 GradeInfo (stuID, courseID, year, scores)，其中除了scores為整數的資料形態外，其餘皆為文字的資料形態，假設(1)沒有相同名字的學生，(2)可能有多個學生住在同一個地方 (相同address)，(3)每個地方都有一支連絡電話 (phone)。請回答以下的問題：(20分)

(一) 請寫出SQL語法及其相對應之關連代數式 (relational algebra) 以列印出學生學號 (stuID) 及其修課名稱 (courseName)。

(二) 請問StudentInfo是否滿足BCNF或3NF的設計，為什麼？請詳述理由。

答：

(一)

1. SQL:

```
SELECT  stuID, courseName
FROM    GradeInfo, CourseInfo
WHERE   GradeInfo.courseID = CourseInfo.courseID
```

2. Relational Algebra:

$$\pi_{stuID, courseName} (GradeInfo * CourseInfo)$$

(二) StudentInfo已滿足BCNF及3NF的設計。

StudentInfo中，僅有一個決定因素stuID，其可決定所有其他屬性(stuName, address, phoneNo)；故StudentInfo的候選鍵即為stuID。

因此，StudentInfo關連中無任何遞移相依出現(滿足3NF)，且唯一之決定因素stuID即為主鍵(滿足BCNF)，故此關連同時滿足3NF以及BCNF。



三、何謂B⁺-tree？請描述其特性。B⁺-tree與B-tree有何差異性？假設一個B⁺-tree中，每個節點最多可存4個搜尋鍵（key），請問在這種有2層的B⁺-tree中，最多可以記錄多少搜尋鍵？有3層的B⁺-tree中，最多又可以記錄多少搜尋鍵？（20分）

答：

(一)B⁺-tree

一棵Order為m的B-tree是一棵m-way搜尋樹，空樹也是B-tree。一棵B-tree若不是空樹，則：

1. 樹根至少有兩個子節點一個鍵值。
2. 除了樹根之外，所有內部節點至少有 $\lceil m/2 \rceil$ 個子節點 ($\lceil m/2 \rceil - 1$ 個鍵值)，最多有m個子節點 (m-1 個鍵值)。
3. 所有樹葉節點皆在同一階層 (Level)。
4. 所有樹葉節點均在同一層，亦即從樹根節點到樹葉節點所經之路徑長度均相同。
5. 樹的高度增加只發生於樹根節點一分為二時。
6. 在B⁺-tree中，只有葉節點的索引記錄才含有資料指標，其餘的節點中僅含有單純的索引值

(二)B-tree與B⁺-tree最大的不同點，在於B-tree的所有節點皆為相同結構，包含索引值以及記錄指標，而B⁺-tree的內部節點 (Internal node) 不含資料指標，資料指標一律僅存在於樹葉節點 (leaf node)。

因此B⁺-tree每個節點能存放的索引數目會比較多，相對地樹狀結構的層級數目就少一點，減少了資料區塊存取數目，所以在實際的應用上，仍以B⁺-tree較為多見。

(三)每一節點可儲存4個搜尋鍵，代表其可以儲存5個指向下一階層節點的Pointer。故各層可儲存之資料如下：

Level	Node	Key	Pointer
Root(Level 1)	1	4	5
Level 2	5	20	25
Level 3	25	100	

因此，2層的B⁺-tree可儲存20個搜尋鍵，3層的B⁺-tree可儲存100個搜尋鍵。

四、HTML的超連結 (Hyperlink) 可以將一段文字或一個圖片連結到另一份文件，請問在HTML檔中需要對該段文字或圖片作怎麼樣的處理，以達到超連結的效果？相對於HTML的超連結，在XML中需要如何描述才能達到相同的效果？此外，XML中XLink所定義的連結比HTML的超連結多提供了那些功能？在XML中，若要對文件的位置點做標示，需使用那個標準語法來達成？（20分）

答：

(一) HTML的超連結，使用標準<a>語法，使用方式如下：

`連結文字或圖片`

其中，URL為超連結之目的地。

例1：`國立台灣大學`

例2：``

例1為文字超連結，例2為圖片超連結。

(二) XML可採用Xlink方式達到超連結之目的，其描述標籤可由使用者自行決定，而不需為<a>。

例如：`<homepage xlink:type="simple"`

`xlink:href="http://www.ntu.edu.tw">國立台灣大學</homepage>`

(三)

屬性	屬性值	說明	XLink較HTML多出之功能
type	simple	簡單形式連結	Y
	extended	延伸形式連結	
href	URL	超連結的URL	N
role		啟動相關應用程式並做出回應	Y
title		瀏覽者了解該鏈結的意義並做出回應	N
show	new	開啟新視窗來讀取該鏈結的目標	N

【台北】台北市開封街一段2號8樓·(02)2331-8268

【台中】台中市東區復興路四段231-3號1樓·04-22298699

【高雄】高雄市新興區中山一路308號8樓·07-235-8996

【另有淡水·三峽·中壢·逢甲·東海·中技·台南】

高上高普特考

www.get.com.tw/goldensun

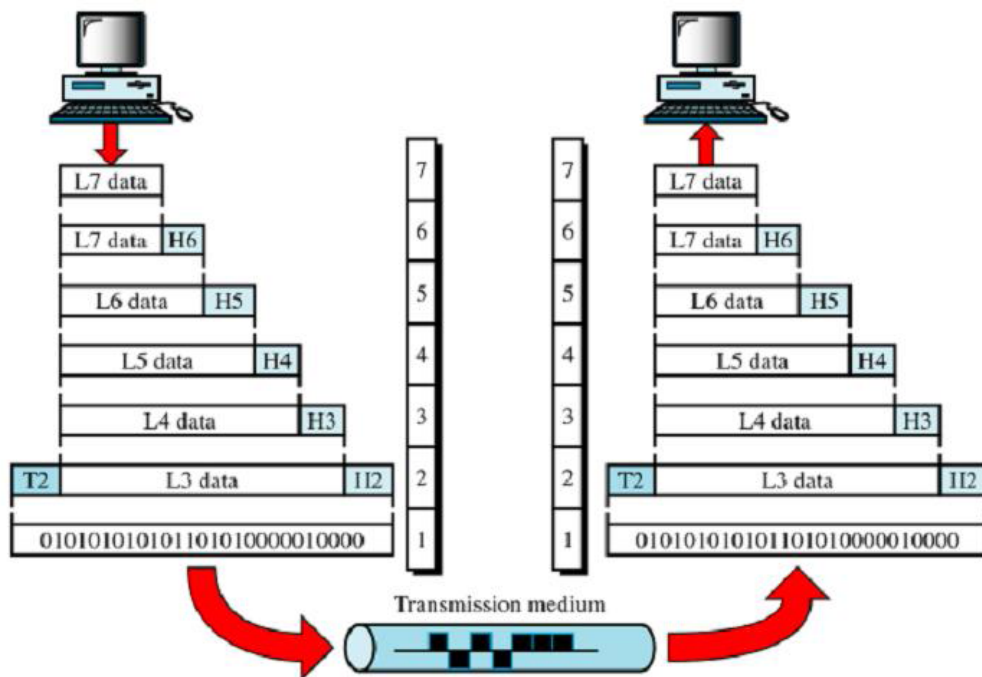
	replaced	以鏈結的目標替換目前的網頁	
	embed	以鏈結的目標替換鏈結處	
	undefined		
actuate	onLoad	當應用程式解讀到Xlink時，會自動載入並啟動該鏈結	Y
	onRequest	必須等到使用者的要求才啟動該鏈結	
	undefined		

(四) 若要標示XML文件位置點，可使用標準語法「locator」。

五、請描述IP網路中各層次資料封裝的關係為何？位址解析（address resolution）之應用時機為何？透過甚麼協定以完成位址解析？網路遮罩（netmask）的作用是什麼？請舉例說明。最後請說明IP層提供可靠性（reliability）的做法為何？（20分）

答：

(一)以ISO/OSI網路架構為例：傳送端將資料由上層向下依序封裝上該層所需標頭(Header)，傳送至目的端後，再由下往上依序解開資料封裝。如下圖例：



(二) Address Resolution的使用時機與目的為，將網路IP位址，轉換為實際硬體的MAC位置。

(三) Address Resolution 一般透過ARP協定(Address Resolution Protocol)。

(四)由於傳統區分Class A~E的IP分級，有浪費網址等缺點，故使用網路遮罩（Network Mask），卸除IP 位址固定隸屬於網址分級，依情況再細分為子網路，並用Network Mask將IP位址分割為網路位元、子網路位元及主機位元。例如：子網路遮罩255.255.255.0，表示IP位址的前三個數字相同者，為同一個子網路。

(五) IP層的可靠性，可靠IP標頭(IP Header)中的8 bits的Type of Services定義不同等級優先權來達到。

其中參數位置代表的意義如下：

位置	參數名稱
第1~3 bits	Precedence
第4 bits	Delay
第5 bits	Throughput
第6 bits	Reliability

【台北】台北市開封街一段2號8樓·(02)2331-8268

【台中】台中市東區復興路四段231-3號1樓·04-22298699

【高雄】高雄市新興區中山一路308號8樓·07-235-8996

【另有淡水·三峽·中壢·逢甲·東海·中技·台南】

高上高普特考

www.get.com.tw/goldensun

第7 bits	Cost
第8 bits	Reserved

Reliability的位元，用以定義IP 封包對於路徑可靠度的要求。High Reliability代表路由器應儘量選擇可靠度較高的路徑來傳送此封包。

若應用程式希望在傳送過程中儘量減少IP封包遺失，會將Reliability參數設為1。一般的IP封包皆設為0。

