

# 《資料處理》

## 試題評析

今年的高考統計「資料處理」的考題，維持一貫特性，涵蓋範圍非常廣泛，包含：資訊概論、電腦網路、作業系統、無線通訊技術、程式語言、資料庫與網際網路技術。今年出現一題非本科系非常難作答的問題。不過同樣是資料庫比重最重，預估一般同學能拿到40分，程度較佳同學應可有50分以上。

第一題：此題為資訊概論、電腦網路、作業系統，分為四小題：(一)資訊概論中2's補數表示法、(二)電腦網路中IP位址的分配、(三)作業系統中之虛擬記憶體技術、(四)圖形資料壓縮技術。(一)(二)應可輕鬆應答，(三)(四)較不容易作答。

第二題：此題為無線通訊技術，對於非本科系同學而言，此題應該非常陌生。

第三題：此題為程式語言，共分四小題：(一)組合語言、(二)組合語言與高階語言比較、(三)宣告式語言與程式語言比較、與(四)物件導向程式語言技術。(一)(二)應較不易答，(三)屬中等，(四)應該可輕鬆回答。

第四題：此題為資料庫，共分四小題：(一)(二)(三)均屬SQL語言查詢，(四)考資料庫技術，此題應可輕鬆得分。

第五題：此題屬資料庫與網際網路技術，共分三小題：(一)資料庫查詢語言、(二)正規化、(三)XML應用。(一)(二)應可輕鬆作答，(三)較不易回答。

一、請逐一回答下列問題：(一)在2的互補法(2's complement)中，10個位元所能表示的範圍中最大與最小的整數為何？又如何表示？(二)在目前IPv4的網路中，一個C類別的子網路其網路位址(不包含主機位址)需要多少位元表示？(三)在多工的作業系統中，虛擬記憶體的目的為何？請舉出兩個實作虛擬記憶體的方法。(四)在電腦應用中有那兩種傳統表現圖片的技術？那一種方式對圖片大小改變的應用較不適合？為什麼？(20分)

**答：**

(一)使用 2's 補數表示法，10個位元所能表示的範圍為  $(-2^9) \sim (2^9-1)$ 。

1.最大整數表示為：0111111111

2.最小整數表示為：1000000000

(二)IPv4 的網路位址規劃中，Class C 的位址是以 110 為開頭的 32 bits 表示，網路位址24bits，而主機位址8 bits。如下圖所示：



(三)電腦的可定址範圍，通常比實際所擁有的記憶體空間大。虛擬記憶體(Virtual Memory)技術，採用在硬碟中所規劃的交換空間(Swap Space)，與實際記憶體儲存空間，進行必要的資料轉移交換，來虛擬如同定址範圍這麼大的記憶體空間可使用。通常有兩個技術被採用，分別為分頁(Paging)與分段(Segmentation)技術。在多工(Multi Tasking)的電腦環境下，會有多個執行序(Process)同時進行著，每個執行序都會維護一張表，以記錄實際記憶體與硬碟中交換空間的資料對應。以分頁技術為例，即分頁對應表(Page Mapping Table, PMT)。

(四)電腦系統標示圖片的格式，一般分為兩大類：(1)失真的壓縮表示格式(如：JPEG圖片格式)。(2)非失真的壓縮表示格式(如：GIF圖片格式)。一般而言，失真的壓縮表示格式，不適合圖形改變大小的應用，因為在改變圖形大小時，又進行圖形格式的編解碼，很可能造成資料失真擴大，造成圖片觀看品質不佳的結果。

## 【參考書目】

(一)張宏亮老師，資料處理課程講義第一回，第38頁與第39頁。

(二)張宏亮老師，資料處理課程講義第一回，第69頁。

(三)一般作業系統書籍。

(四)張宏亮老師，資料處理課程講義第一回，第18頁。

二、在無線網路的領域中常提到的ISM頻帶標準，ISM指的是什麼？其所規範的三個無線區網頻帶為何？無線網路中有所謂的Ad-hoc與Infrastructure模式，請問兩者的差異為何？802.11b所定義無線網路的實體層標準中，運用了那兩種展頻技術？有何差異？又有何限制？（20分）

**答：**

(一)所謂的ISM，為Industrial Scientific Medical Bands的縮寫。規定了現今的無線網路所使用的頻帶。此頻帶是由3組不同的無線波段共同組成，包含：

- 1.提供給工業(Industrial)無線傳輸應用的902~928MHz(26MHz)波段。
- 2.提供給科學研究(Science)的2.4~2.4835GHz(83.5MHz)波段。
- 3.開放給醫療(Medical)用途的5.15~5.35GHz與5.725~5.825GHz(合計300MHz)波段。

(二)IEEE 802.X系列的協定為電機電子工程師協會(Institute of Electrical and Electronic Engineers, IEEE)所推動的標準。此標準是定義區域網路中的第一層實體層(physical layer)及第二層資料連結層(data-link layer)在網路上的存取控制。而目前IEEE 802.11b的實體層規範，主要是使用在2.4GHz的ISM頻帶，也就是提供給科學研究的波段。2.4GHz的射頻方式是依據ISM頻段以展頻技術，可不須經授權使用的規定，這個頻段使用在全世界，包含美國、歐洲、日本及台灣等主要國家都有開放。IEEE 802.11無線區域網路標準依據兩種不同的網路架構，亦規範了不一樣的功率控制方式：

- 1.所謂的Infrastructure Network指的是由存取點(Access Point, AP)所建立而成的一種網路架構。在此網路結構中，存取點AP為一個比較特殊的節點，其功能主要是將一個或多個行動節點(Mobile Node, MN)與有線網路做連結。功率控制主要是由存取點來進行，且行動節點可直接進入休眠狀態節省功率消耗。
- 2.所謂Ad-Hoc Network則是由一群獨立、可自由移動且具有無線傳送與接收能力的行動節點所組成，每個行動節點可與無線傳送範圍內的行動節點進行點對點(point-to-point)的通訊，不需AP的存在。功率控制則是由傳送端之行動節點來負責，因此傳送端必須承受比較重的負擔以確保接收端處於清醒的狀態。因此在Ad Hoc模式中功率的消耗遠比Infrastructure模式來得大。

(三)在標準中定義的實體層還包括兩種不同的展頻技術：分別為：(1)直接序列展頻技術(Direct Sequence Spread Spectrum, DSSS)(2)跳頻展頻技術(Frequency Hopping Spread Spectrum, FHSS)，針對高信賴性、正確性和安全性而設計，它們各有一套獨特的方法來傳送資料。

- 1.DSSS為目前展頻系統最常用的技術，優點主因為DSSS容易使用及可提供較高的速度。DSSS技術讓傳送及接收者都使用22MHz的頻寬，此頻寬讓設備得以支援較高的11Mbps速度。DSSS技術為了克服多點存取時所造成的Delay Spread問題，將資料信號以較高速度的位元(bit)順序作傳送，稱為Chipping Code或Processing Gain。高Processing Gain可以增加信號抵抗干擾的能力，特點為高速、範圍大、適合較固定的環境中使用。
- 2.FHSS技術是將可用的頻寬帶區域分割成好幾個頻道，它使用的窄頻載波藉著2到4階的高斯頻移調變(Gaussian Frequency Shift Keying, GFSK)不斷的改變頻率，即跳頻技術。換句話說，傳送時的頻率會在收發雙方節點都知道的範圍間，利用虛擬亂碼(pseudorandom)做任意變動。這對實體層加入了一些安全性。有心的駭客基本上沒有辦法知道接下來要切換到哪一組頻率來接收整個訊號。特點為低速、範圍較小，但適合設備移動變化大時使用且提供較佳的安全性。

**【參考書目】**

請參考無線網路通訊技術書籍與IEEE 802.11b的協定標準。

三、Assembly Language是否為第三代的程式語言？第三代程式語言有別於第二代程式語言的地方為何？在程式語言的分類上，請比較宣告式(declarative)的程式語言與傳統程序式(procedural)的程式語言的差異性。此外，物件導向式語言比傳統程序式(procedural)語言的優點在那裡？（20分）

**答：**

(一)Assembly Language(組合語言)為第二代程式語言。第三代程式語言為高階語言，例如：BASIC、PASCAL、C語言等。

- (二)1.組合語言是針對微處理器(CPU)操作的一種語言，程式的設計是以CPU為導向，不同廠牌的CPU，其組合語言的種類亦不同，所以組合語言的可攜性較低。組合語言要透過組譯器(Assembler)進行編譯，成為機器語言(0與1碼)，才能在電腦上執行。  
由於組合語言是直接對處理機CPU進行控制，因此程式執行的效率非常高，但程式碼常常過於冗長，可讀性低，較不易學習。
- 2.高階語言使用了與人類較接近的語言，符合人類的使用習慣，容易被了解且親和力高。高階語言所撰寫的程式，需要透過編譯器(Compiler)或直譯器(Interpreter)進行編譯成機器語言，才能於電腦上執行。程式執行的效率較低，但程式碼長度精簡，可讀性高，容易學習。
- (三)1.宣告式語言(Declarative programming)是以符號邏輯(symbolic logic)為基礎，在語言中只需描述運算(computation)的特性，程式的設計主要描述解決問題的推論規則和事實資料，不必管如何解決問題的細部程序。純邏輯式語言不需要控制邏輯敘述，如迴圈或選擇敘述。Relational Calculus查詢語言或邏輯語言 Prolog都為一種宣告式語言。
- 2.程序式語言(Procedural programming Language)主要包括了一步步循序的執行指令、使用變數代表的記憶體位置，與使用指定敘述改變的變數值。程序式的語言需描述成指令序列，每一指令操作一小塊資料，需清楚交代運算的細節。常見的高階語言：C、Basic、與PASCAL語言均屬程序式語言的一種。
- (四)物件導向(Object-Oriented)程式語言是以物件、屬性、方法、事件構成解決問題的語言。每一個物件(object)都有屬於自己的屬性(property)和方法(method)。屬性：物件特有的性質；方法：物件特有的行為動作。物件與物件之間是藉由方法來傳遞訊息(message)。每一個物件藉由事件(event)作出適當的反應。物件導向技術，包含三個主要特性：分別為封裝、繼承、多型，來描述物件的屬性與方法。  
物件導向程式語言的優點為：
- 1.降低軟體開發成本，尤其是軟體的維護成本。
  - 2.增加軟體的重複使用性、擴充性、可靠性、可維護性以及安全性。
  - 3.增進軟體設計之生產力，使得專案計畫能夠在預計期限內，以最低的成本完成最佳品質的軟體。

#### 【參考書目】

- (一)(二)參考一般計算機概論書籍。  
(三)張宏亮老師，資料處理講義第二回，第107頁。  
(四)張宏亮老師，資料處理講義第三回，第73頁。

四、以下兩個關連表一為學生資料表(StudentInfo)，一個課程A的資料表(CourseAInfo)，有劃底線之欄位StdId為主鍵。

StudentInfo (StdId, Name, Address, PhoneNo, Depart)

CourseAInfo (StdId, Midterm, Final, Grade, Rank)

其中StdId代表學號，Depart代表系別，Midterm代表期中考成績，Final代表期末考成績，Grade代表學期成績，Rank代表年級別。假設欄位中除了成績為整數型態外，其餘欄位皆定義為文字形態。

請回答下列問題(20分)

問題一：以SQL敘述列出以姓名排序之所有“化學系”學生的學號及姓名。

問題二：以SQL敘述列出所有“化學系”二年級學生修習課程A的學生學號及其學期成績。

問題三：以SQL敘述更改所有修習課程A中四年級的學生的學期成績，每人以加十分成績計算。

問題四：若將兩個資料表合併成為一個資料表，有何優缺點？

答：

註：“==”代表equality，“=”代表 assignment。

(一)

```
Select StdId, Name
From StudentInfo
Where Depart == '化學系'
```

Order By Name DESC;

(二)

```
Select StdId, Grade
From StudentInfo, CourseAInfo
Where StudentInfo.StdId == CourseAInfo.StdId
AND Depart == '化學系'
AND Rank == '二年級';
```

(三)

```
Update CourseAInfo
Set Grade = Grade + 10
Where Rank == '四年級';
```

(四)資料表 **StudentInfo** 記錄了所有學生的基本資料，而 **CourseAInfo** 記錄了所有選修課程A學生的年級與成績資料。兩張表都是以 **StdId** 當成主鍵。合併時會以**StdId**進行合併，成為另一張Schema為 (**StdId, Name, Address, PhoneNo, Depart, Midterm, Final, Grade, Rank**)的關連表。合併與否的優缺點，端視選修課程A的學生情況：

- 1.如果大部分的學生，都選修了課程A，若將兩張資料表合併，則會節省所需要儲存**StdId** 資料的空間。此狀況適合將兩張表合併。
- 2.若只有少部分的學生選修課程A，則會造成許多資料表中的資料 **Midterm, Final, Grade, Rank** 為空值 (NULL)。此狀況不適合將兩張資料表合併。

#### 【參考書目】

- (一) (二) 張宏亮老師，資料處理講義第二回，第88頁-90頁。
- (三) 張宏亮老師，資料處理講義第二回，第87頁。
- (四) 張宏亮老師，資料處理講義第二回，第88頁-90頁。

五、資料庫之查詢系統中除了SQL外，試舉出另外一種常用之查詢方式。關連式資料庫之正規化設計，其目的為何？以目前流行之XML格式作資料儲存，請列出並比較兩種查詢XML資料的語言。(20分)

答：

(一)資料庫查詢系統，除了使用SQL語言之外，還有另一種以Relational Calculus為基礎的查詢語言。此查詢語言為非程序式(Nonprocedural)查詢語言，是一種宣告式語言，不必一步步描述其過程，而是使用值組(tuple)或定義域(domain)變數建立查詢運算式(Query Expression)，直接宣告和定義查詢結果的關連表。此種查詢語言，一般分成兩種形式的關連式計算

- 1.值組計算(Tuple Calculus)：以值組變數(tuple variables)的公式(亦即條件運算式)描述查詢結果的關連表。例如：INGRES資料庫管理系統中的QUEL語言採此方法。
- 2.定義域計算(Domain Calculus)：使用定義域變數(domain variables) 的公式，描述查詢結果的關連表。例：查詢語言QBE(Query By Example) 即採用此方法。

(二)關連資料庫正規化，為一種標準處理程序，以決定關連表應該擁有那些屬性。其目的在建立良好結構關連表，並且在更新、刪除、新增資料時，不會造成錯誤或不一致的異常現象，例如：新增異常、刪除異常、修改異常。

- 1.新增異常(Insertion Anomaly)：意指可能只有新增值組的部分資料，而非完整的資料。
- 2.更新異常(Update Anomaly)：意指當更新指定值組的屬性時，因重複儲存資料，將需要同時更新多筆值組的資料。
- 3.刪除異常(Deletion Anomaly)：當刪除指定值組時，因其他資料與欲刪除資料是同時存在，可能導致一併被刪除。

正規化通常分為下述幾種：

- 1.一階正規化(1NF)：刪除多重值屬性(multi-value attribute)，讓關連表只擁有單元值屬性(atomic attribute)。
- 2.二階正規化(2NF)：滿足1NF，而且關連表沒有部分相依於主鍵(Primary Key)。
- 3.三階正規化(3NF)：滿足2NF，而且關連表沒有遞移相依。

4.Boyce-Code正規化(BCNF)：滿足3NF，去除不是完全功能相依於Super Key的功能相依。

5.四階正規化(4NF)：滿足BCNF，而且沒有多重值相依(Multi-valued Dependency)。

6.五階正規化(5NF)：滿足4NF，而且沒有合併相依。

(三)XML只是單純的在定義和描述資料，瀏覽XML文件內容並沒有任何意義，一個完整XML架構需要搭配相關XML技術，才能將需要的資訊顯示在使用者的面前，進行查詢資料、產生報表、建立網站內容或XML應用程式。其中查詢資料語言分為兩種：

1.XPath (XML Path Language)：XPath語言是一種XML技術的查詢語言，可以在XML文件中找出所需的節點。可以使用XPath語言來瀏覽XML文件，以便找出指定的XML元素和屬性。XPath的優點，就是可以根據您在XSL檔案的位置識別目前的內容。這樣將簡化初學者和進階使用者撰寫正確XPath運算式的過程。

2.Xquery (XMLQuery)：XQuery是一種功能強大和容易使用的XML資料查詢語言。對於不需（或不想）使用SQL或關連式結構的純XML應用程式，非常適合使用。

**【參考書目】**

1. 張宏亮老師，資料處理講義第二回，第107頁-108頁。
2. 張宏亮老師，資料處理講義第三回，第1頁與第11頁。
3. 張宏亮老師，資料處理講義第二回，第31頁-32頁。

