

106年公務人員特種考試警察人員、一般警察
人員考試及106年特種考試交通事業鐵路
人員、退除役軍人轉任公務人員考試試題

代號：20250

全一張
(正面)

考試別：一般警察人員考試

等別：二等考試

類科別：刑事警察人員犯罪分析組

科目：資料探勘技術(包括資料庫管理與運用、線上交易處理【OLTP】、資料倉儲【Data Warehouse】、資料探勘【Data Mining】)

考試時間：2 小時

座號：_____

※注意：(一)禁止使用電子計算器。

(二)不必抄題，作答時請將試題題號及答案依照順序寫在試卷上，於本試題上作答者，不予計分。

一、全球 3C 公司是一間行銷全世界的歷史悠久的電腦、通訊與消費電子產品公司，它的資料庫記錄了 1950-2017 年的每個月 200 種產品、在全世界各地各城市的銷售數量與金額。請針對下列每一項的資料倉儲操作先給定義，再以上述背景為例來說明清楚該操作內涵：(1)向下探究(drill-down)、(2)向上擷取(roll-up)、(3)切片(slice)、(4)切塊(dice)、(5)轉軸(pivot)。(25 分)

二、假設關聯資料庫的表格 $R(A,B,C,D)$ 有四個屬性 A,B,C,D ，其功能相依 (Functional Dependency) 為 $\{A,B,D\} \rightarrow C$ 與 $\{C,D\} \rightarrow B$

(一)請以屬性封閉性 (Closure) 的概念，找出 R 的所有候選鍵 (Candidate Key)，請說明清楚你使用功能相依的每一個推演步驟。(6 分)

(二)請證明 R 滿足第二正規化 (2NF) 或反證 R 不滿足 2NF。(3 分)

(三)請證明 R 滿足第三正規化 (3NF) 或反證 R 不滿足 3NF。(4 分)

(四)請證明 R 滿足 Boyce-Codd 正規化 (BCNF) 或反證 R 不滿足 BCNF。(4 分)

(五)若 R 不滿足上述任一正規化，請將其切割，以求滿足之，你必須說明切割後的各表格為何滿足你所聲稱的正規化。(6 分)

三、假設 XYZ 公司的商品包含 A、B、C、D、E、F、G 等多種，由其交易資料庫記錄，我們得知表一的資料，假設我們要求的最小支持度 (Support) 為 0.2、最小信心水準 (Confidence Level) 為 0.7。請(一)找出所有包含一個商品的高頻項目集合 (Frequent Item-sets)，及其個別的支持度 (7 分)、(二)找出所有包含兩個商品的高頻項目集合，及其個別的支持度 (12 分)、(三)找出所有包含三個商品的高頻項目集合，及其個別的支持度 (4 分)、(四)列出至少兩條包含三個商品的強 (Strong) 關聯規則 (Association Rules) 及其個別規則的信心水準 (8 分)。

表一

交易編號	銷售商品
A10001	A、B、C
A10002	A、D
A10003	D、E
A10004	A、B、C
A10005	C、F
A10006	F
A10007	D、F
A10008	A、B、C、G
A10009	E、G
A10010	A、B

(請接背面)

106年公務人員特種考試警察人員、一般警察人員考試及106年特種考試交通事業鐵路人員、退除役軍人轉任公務人員考試試題

代號：20250

全一張
(背面)

考試別：一般警察人員考試

等別：二等考試

類科別：刑事警察人員犯罪分析組

科目：資料探勘技術（包括資料庫管理與運用、線上交易處理【OLTP】、資料倉儲【Data Warehouse】、資料探勘【Data Mining】）

四、假設 U 大學之學生資料庫有下列表二的 12 筆資料，其 Status（學籍狀態）、Major（主修）、Birth_Place（出生地）、GPA（平均成績）四個屬性值之部分的概念樹個別如圖 1,2,3,4（註：圖 4 的[2.0-2.9]表示其數值介於 2.0 至 2.9 間，也可為 2.9 或 2.0。其餘[3.0-3.4]等概念類推）。請利用這些資料，以概念樹導向方式，找出研究生（Graduate）與大學生（Undergraduate）的區別規則（Discrimination Rule），注意：請列出你的推導過程，且最後的規則應為最抽象、精簡，並加上機率。（21 分）

表二

Student_ID	Status	Major	Birth_Place	GPA
1001	M.A.	History	臺北市	3.5
1002	Junior	Math	新北市	3.7
1011	Junior	Liberal_arts	桃園市	2.6
1012	M.S.	Physics	臺南市	3.9
1021	Ph.D.	Math	溫哥華市	3.3
1022	Sophomore	Chemistry	高雄市	2.7
1031	Senior	Computing	基隆市	3.5
1032	Ph.D.	Biology	大阪市	3.4
1041	Sophomore	Music	臺東市	3.0
1042	Ph.D.	Computing	基隆市	3.8
1051	M.S.	Statistics	東京市	3.2
1052	Freshman	Literature	花蓮市	3.9

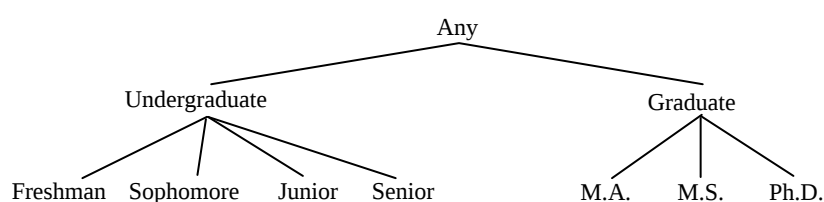


圖 1 學籍狀態概念樹

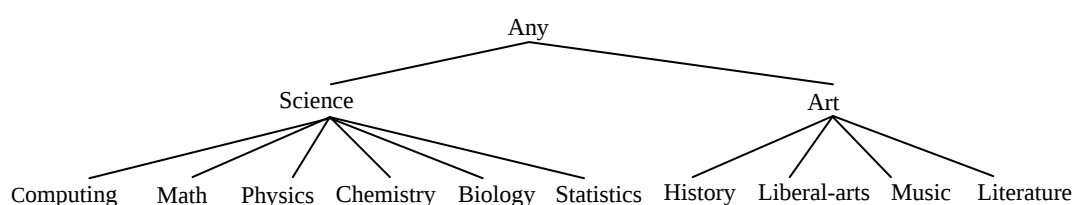


圖 2 主修概念樹

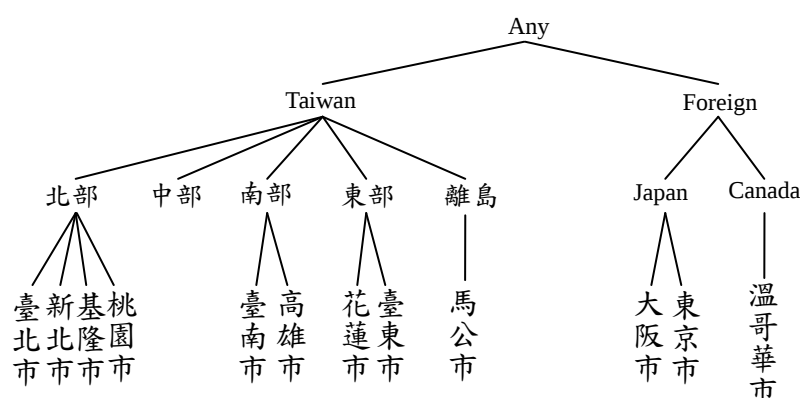


圖 3 出生地概念樹

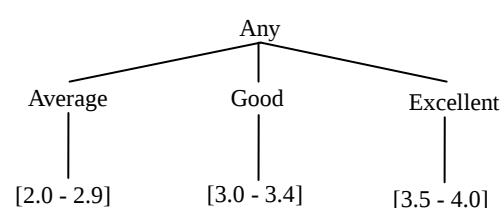


圖 4 平均成績概念樹