

關聯式模式的資料運算

關聯式代數 (Relational Algebra)

【定義】是一種較低階的、程序性的、規範性之抽象的查詢語言，它是來描述如何產生查詢結果的步驟。我們可以想像成「演算法」，亦即描述解決問題的步驟。

關聯式資料庫的 SQL 語法是以「關聯式代數」作為它的理論基礎，而在「關聯式資料模型」中，根據E.F.Codd所提出的「關聯式代數」有八種基本運算子。如下表所示：

運算子		意義	運算子		意義
基本運算子	σ	限制(Restrict)	比較運算子	$>$	大於
	π	投影(Project)		\geq	大於等於
	\cup	聯集(Union)		$<$	小於
	\times	卡氏積(Cartesian Product)		\leq	小於等於
	$-$	差集(Difference)		$=$	等於
非基本運算子	\bowtie	合併(Join)	邏輯運算子	\neq	不等於
	\cap	交集(Intersection)		\neg	NO
	\div	除法(Division)		\wedge	AND
			\vee	OR	

關聯式代數的「基本運算子」

【定義】是指不能由其他「運算子」導出的運算子

【運算子種類】

- ①限制(Restrict) , 代表符號： σ
- ②投影(Project) , 代表符號： π
- ③聯集(Union) , 代表符號： \cup
- ④卡氏積(Cartesian Product) , 代表符號： \times
- ⑤差集(Difference) , 代表符號： $-$

以上五種運算子所成的集合稱為「完整集合」(Complete set)。

關聯式代數的「非基本運算子」

【定義】是指可以由「基本運算子」導出的運算子。

【運算子種類】

① 合併(Join) ，代表符號： \bowtie

② 交集(Intersection) ，代表符號： \cap

③ 除法(Division) ，代表符號： \div

限制(Restrict)

【定義】

- 1.限制又稱為選擇操作(Select Operation)
- 2.「選擇運算子」含有兩個參數
 - (1)選取條件 (Predicate) P
 - (2)關聯表名稱 R
- 3.從 R 中選出符合條件 P 的值組

3.從 R 中選出符合條件 P 的值組

是指在關聯 R 中選擇滿足條件P 的所有值組。

【代表符號】 σ (唸成sigma)

【關聯式代數】 $\sigma_p(R)$

【概念圖】

從關聯表 R 中選取符合條件 (Predicate) P 的值組。其結果為原關聯表R記錄的「水平」子集合。如下圖所示：

R	
A	B
a1	b1
a2	b2
a3	b3
a4	b4

P

=

$\sigma_p(R)$	
A	B
a1	b1
a3	b3

【例如】 $\sigma_{\text{身高}<170 \text{ AND } \text{體重}<60}$ (學生資料表)

則 (1) 選取條件 (Predicate) P \rightarrow 身高<170 AND 體重<60

(2) 關聯表名稱 R \rightarrow 學生資料表

【實例1】

請利用限制(Restrict)來查詢學生資料表

	學號	姓名	性別	身高	體重
#1	S0001	張三	男	175	75
#2	S0002	李四	男	169	65
#3	S0003	王五	男	172	80
#4	S0004	林六	女	158	45
#5	S0005	陳靜	女	163	50

Q1：請問 $\sigma_{\text{體重} > 70}$ (學生資料表)=?

【Q1解答】

	學號	姓名	性別	身高	體重
#1	S0001	張三	男	175	75
#2	S0003	王五	男	172	80

【實例2】

請利用限制(Restrict)來查詢學生資料表

	學號	姓名	性別	身高	體重
#1	S0001	張三	男	175	75
#2	S0002	李四	男	169	65
#3	S0003	王五	男	172	80
#4	S0004	林六	女	158	45
#5	S0005	陳靜	女	163	50

Q2：請利用關聯代數來表示

身高小於170公分及體重小於60公斤的學生記錄

	學號	姓名	性別	身高	體重
#1	S0004	林六	女	158	45
#2	S0005	陳靜	女	163	50

【Q2解答】 $\sigma_{\text{身高}<170 \text{ AND } \text{體重}<60}$ (學生資料表)

分析

註1： $\sigma_{\text{身高}<170 \text{ AND } \text{體重}<60}$ (學生資料表) 可以用別兩種方式表示

(1) $\sigma_{\text{身高}<170} (\sigma_{\text{體重}<60} (\text{學生資料表}))$

(2) $\sigma_{\text{體重}<60} (\sigma_{\text{身高}<170} (\text{學生資料表}))$

註2：

運算所產生的結果關聯，其值組的數目會少於或等於原有關聯的值組數目。

投影(Project)

【定義】

是指從**關聯R**上的**投影**，亦即從關聯R中**選擇出許多「欄位」**後，
再**重新組成一個新的關聯**。

【代表符號】 π (唸成pai)

【關聯式代數】 $\pi_A(R)$ ，其中：**A** 為 R 中的**屬性欄位**。

【概念圖】

從**關聯表 R**中選取想要的**欄位**。其結果為**原關聯表R**記錄的「**垂直**」**子集合**。如下圖所示：

R			$\pi_{\text{欄位}}(R)$	
A	B	C	A	C
a1	b1	c1	a1	c1
a2	b2	c2	a2	c2
a3	b3	c3	a3	c3
a4	b4	c4	a4	c4

=

【實例1】

請利用投影(Project)來查詢學生資料表

	學號	姓名	性別	身高	體重
#1	S0001	張三	男	175	75
#2	S0002	李四	男	169	65
#3	S0003	王五	男	172	80
#4	S0004	林六	女	158	45
#5	S0005	陳靜	女	163	50

請問 $\pi_{\text{學號,身高,體重}}$ (學生資料表)=?

【解答】

	學號	身高	體重
#1	S0001	175	75
#2	S0002	169	65
#3	S0003	172	80
#4	S0004	158	45
#5	S0005	163	50

【實例2】

請利用「關聯式代數」來撰寫下列的查詢

學生資料表

學號	姓名	性別	科系
S0001	張三	男	資管系
S0002	李四	男	資工系
S0003	王五	男	資管系
S0004	李崑	女	企管系
S0005	李安	女	資工系

請找出「資管系」學生的學號、姓名？

【解答】：

$\pi_{\text{學號,姓名}}(\sigma_{\text{科系}='資管系'}(\text{學生資料表}))$

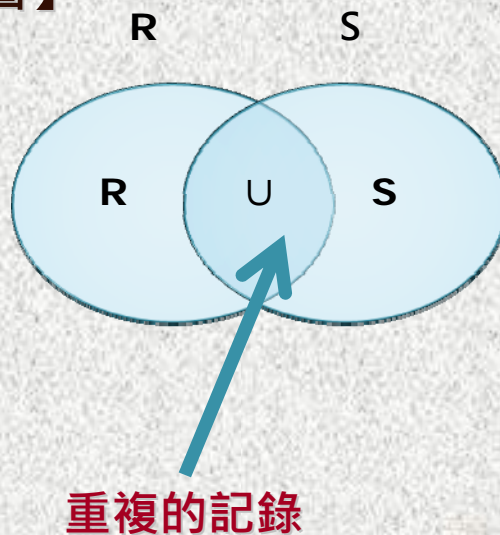
聯集(Union)

【定義】

是指**關聯表R**與**關聯表S**作「聯集」時，則會重新組合成一個**新的關聯表**，而**新的關聯表**中的記錄為原來兩關聯表的所有記錄，**若有重複的記錄**，則只會出現一次。

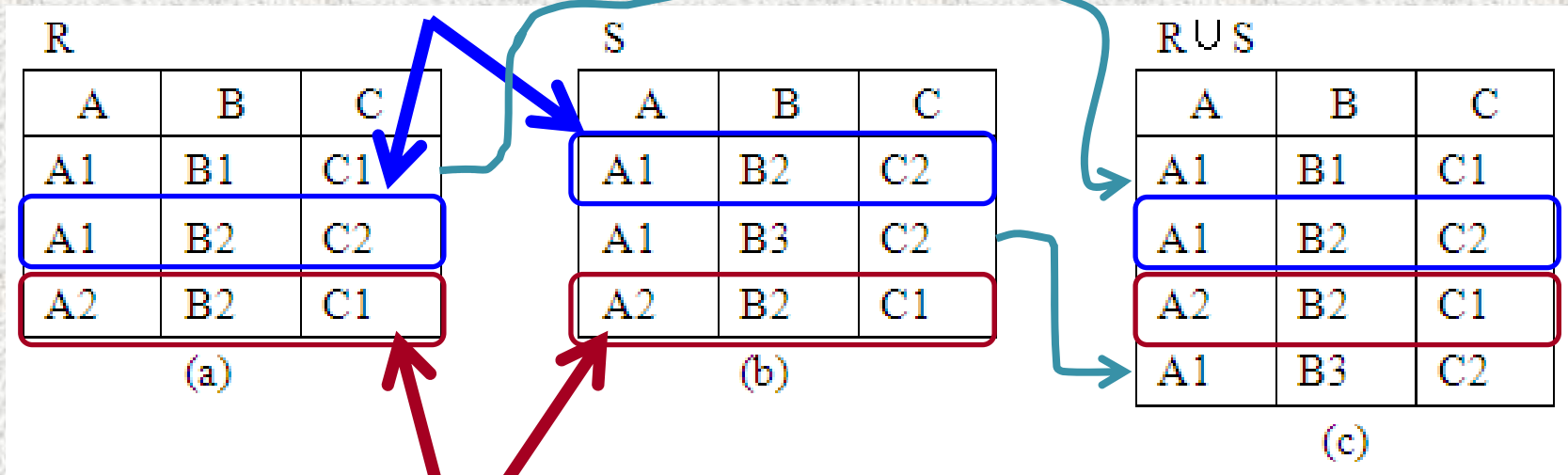
【關聯式代數】 $R \cup S$

【概念圖】



【概念分析】

重複的記錄(只會出現一次)



重複的記錄(只會出現一次)

【實例】

請利用「聯集(Union)」來查詢學生資料表

R 資料表

	學號	姓名	性別	身高	體重
#1	S0001	張三	男	175	75
#2	S0002	李四	男	169	65
#3	S0004	林六	女	158	45

S 資料表

	學號	姓名	性別	身高	體重
#1	S0001	張三	男	175	75
#2	S0005	陳靜	女	163	50

重複的記錄

只會出現一次

請問RUS=?

【解答】

	學號	姓名	性別	身高	體重
#1	S0001	張三	男	175	75
#2	S0002	李四	男	169	65
#3	S0004	林六	女	158	45
#4	S0005	陳靜	女	163	50

卡氏積(Cartesian Product)

【定義】是指將兩關聯表R與S的記錄利用集合運算中的乘積運算形成新的關聯表。

【作法】

1. 關聯R和關聯S分別為n欄和m欄，其卡氏積是一個(n+m)欄的值組的集合，其前n欄是關聯R的，後m欄是關聯S的。
2. 若關聯R有X個值組，關聯S有Y個值組，則關聯R和關聯S的卡氏積有X*Y個值組。

【關聯式代數】 $R \times S$

【概念分析】

(1) 假如關聯R中 $n=3$ (欄)，關聯S中 $m=3$ (欄)，在笛卡爾乘積之後，變成一個 $(n+m)$ 欄，所以 $R \times S$ 共有6欄。

(2) 假如關聯R中 $X=3$ 個值組，關聯S中 $Y=3$ 個值組，則關聯R和關聯S在笛卡爾乘積有 $X*Y$ 個值組，所以 $R \times S$ 共有9筆值組。

$n=3$ (欄)

R		
A	B	C
A1	B1	C1
A1	B2	C2
A2	B2	C1

X=3筆

(a)

$m=3$ (欄)

S		
A	B	C
A1	B2	C2
A1	B3	C2
A2	B2	C1

Y=3筆

(b)

$n+m=3+3=6$ (欄)

$R \times S =$

$R \times S$					
A	B	C	A	B	C
A1	B1	C1	A1	B2	C2
A1	B1	C1	A1	B3	C2
A1	B1	C1	A2	B2	C1
A1	B2	C2	A1	B2	C2
A1	B2	C2	A1	B3	C2
A2	B2	C1	A1	B2	C2
A2	B2	C1	A1	B3	C2
A2	B2	C1	A2	B2	C1

(c)

$X*Y=3*3=9$ 筆

【實例】

請利用卡氏積(Cartesian Product)來查詢學生資料表

R 資料表

	學號
#1	S0001
#2	S0002
#3	S0003

S 資料表

	課號
#1	VB101
#2	DB203

請問 $R \times S = ?$

【解答】

	學號	課號
#1	S0001	VB101
#2	S0001	DB203
#3	S0002	VB101
#4	S0002	DB203
#5	S0003	VB101
#6	S0003	DB203

說明：若關聯R有X筆值組(記錄)，關聯S有Y筆值組(記錄)，則

$R \times S$ 共有 $X * Y$ 筆記錄。

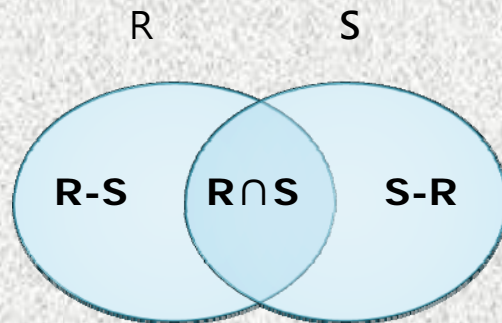
差集(Difference)

【定義】

是指**關聯 R**差集**關聯 S**之後的結果，則為**關聯 R**減掉**RS**兩關聯共同的值組。

【關聯式代數】 $R - S$

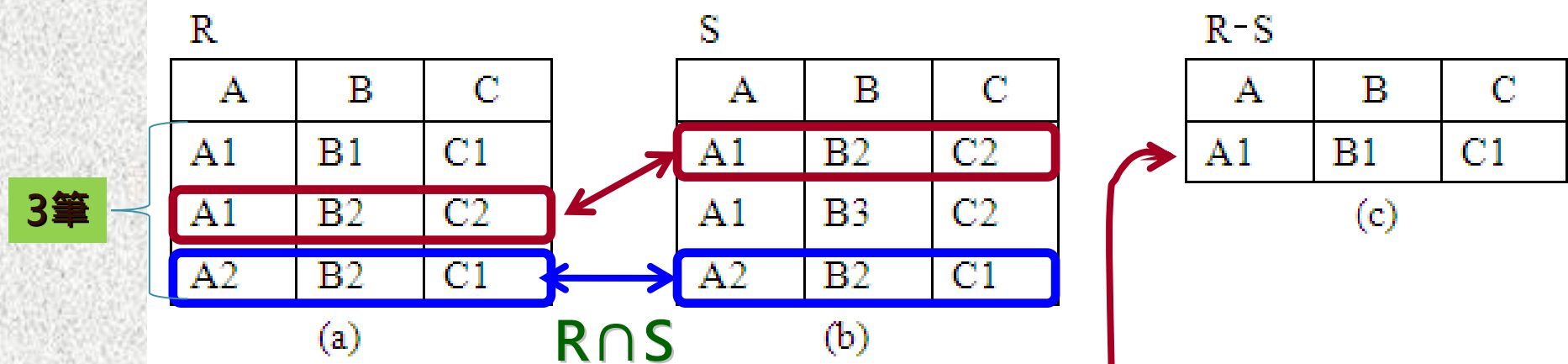
【概念圖】



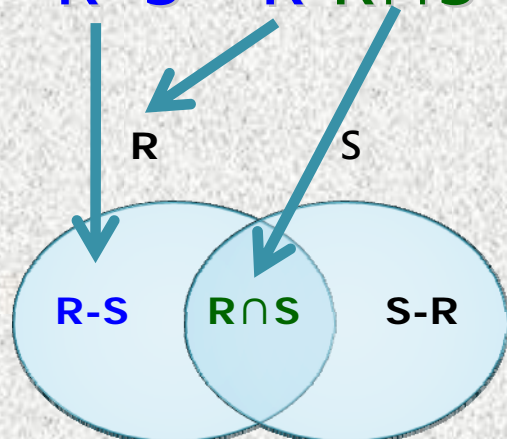
註：R-S 代表：屬於R，但不屬於S
S-R 代表：屬於S，但不屬於R

【概念分析】

$R-S$ 代表：屬於R，但不屬於S，亦即 $R-R \cap S$



$$R-S = R - R \cap S = 3 \text{ 筆} - 2 \text{ 筆} = 1 \text{ 筆}$$



【實例】

請利用差集(Difference)來查詢學生資料表

R 資料表

	學號	姓名	性別	身高	體重
#1	S0001	張三	男	175	75
#2	S0002	李四	男	169	65
#3	S0004	林六	女	158	45

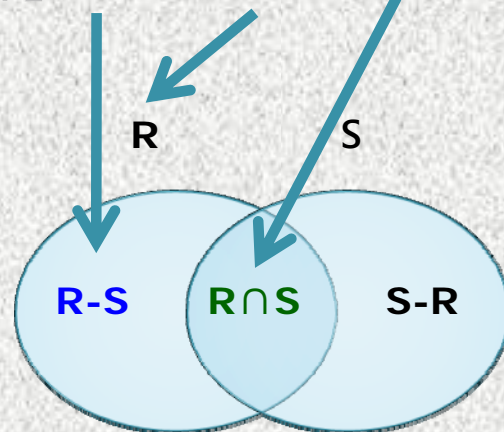
S 資料表

	學號	姓名	性別	身高	體重
#1	S0001	張三	男	175	75
#2	S0005	陳靜	女	163	50

1筆相同(亦即 $R \cap S$)

請問 $R-S=?$

【解答】 $R-S = R - R \cap S = 3筆 - 1筆 = 2筆$



	學號	姓名	性別	身高	體重
#1	S0002	李四	男	169	65
#2	S0004	林六	女	158	45

【練習 1】

若 $A=\{1,2,3,4\}$ ， $B=\{3,4,5,6\}$ ，則 $A-B$

請問 $A - B = ?$

【解答】

$A-B=\{1,2\}$

【練習 2】

若 $A=\{1,2,3,4\}$ ， $B=\{3,4,5,6\}$ ，則 $B-A$

請問 $B-A=?$

【解答】

$B-A=\{5,6\}$

合併(Join)

【定義】

是指將兩關聯表R與S依合併條件合併成一個新的關聯表 R_3 ，假設P為合併條件，以 $R \bowtie_p S$ 表示此合併運算。

【作法】從兩個關聯的「卡氏積」中選取屬性間滿足一定條件的值組。

【關聯式代數】 $R \bowtie_p S$

自然合併(Natural Join)

【定義】

自然合併(Natural Join)又稱為內部合併(Inner Join)，它必須在左右兩邊的關聯中找到對應值組才行，而Outer join則無此規定。

一般的結合(Join)都是屬於此種方法。

【概念分析】

R	A	B	C
	A1	B1	C1
	A2	B2	C1
	A3	B3	C2

(a)

S	C	D	E
	C1	D1	E1
	C2	D2	E2

(b)

$R \bowtie_p S$

R.A	R.B	R.C	S.D	S.E
A1	B1	C1	D1	E1
A2	B2	C1	D1	E1
A3	B3	C2	D2	E2

重複欄位，
只會出現一次

(c)

【作法】透過SELECT指令FROM部分的Natural Join。

【例如】 From A Natural Join B

【實例】

學生			班級			
	學號	姓名	班級代號	班級代號	人數	導師
#1	S0001	張三	1	1	60	李春雄
#2	S0002	李四	2	2	55	李碩安
#3	S0003	王五	5	3	50	王靜旻
#4	S0004	林六	NULL	4	50	葉小宏
				5	45	陳靜華

請問 學生(\bowtie 學生.班級代號=班級.班級代號)班級=?

【解答】

	學生.學號	學生.姓名	學生.班級代號	班級.人數	班級.導師
#1	S0001	張三	1	60	李春雄
#2	S0002	李四	2	55	李碩安
#3	S0003	王五	5	45	陳靜華

National Join 只出現一次

\bowtie

交集(Intersection)

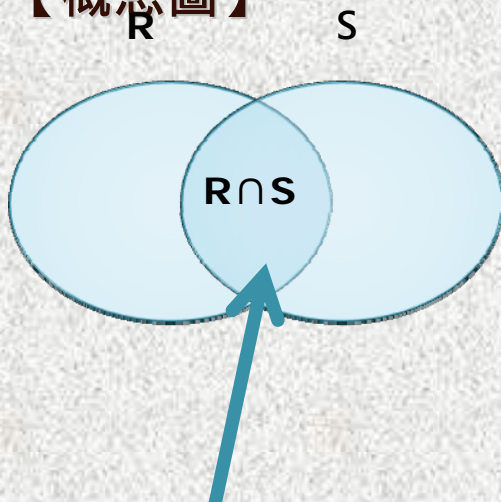
【定義】

是指關聯 R 與關聯 S 作「交集」時，則將原來在兩個關聯式中**都有出現的值組(記錄)**組合在一起成為**新的關聯式**。

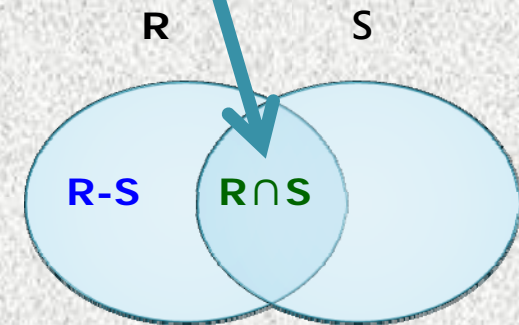
【關聯式代數】 $R \cap S$ 代表既屬於 R 又屬於 S 的值組組成。

關聯的「交集」可以用「差集」來表示，即 $R \cap S = R - (R - S)$

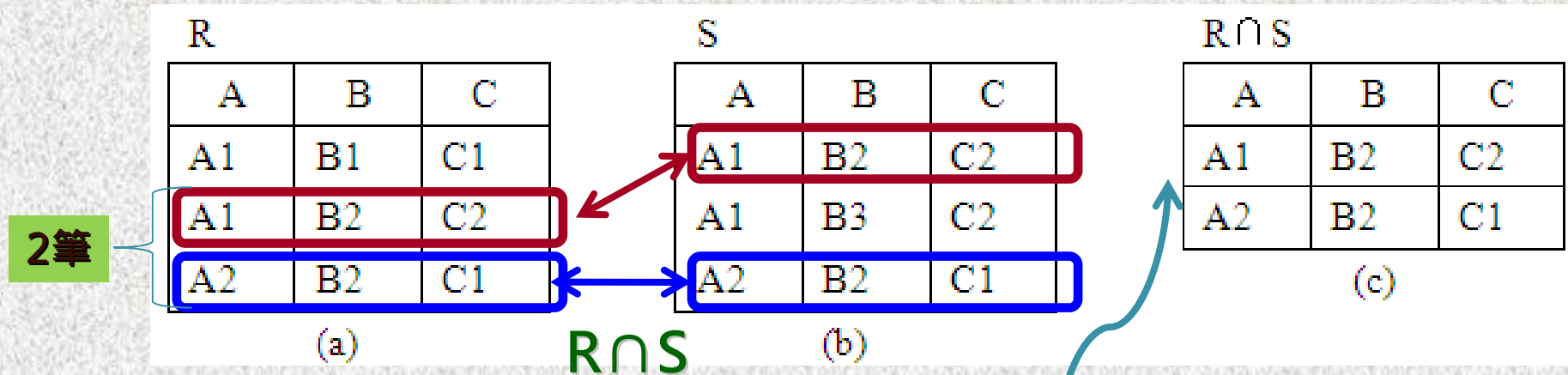
【概念圖】



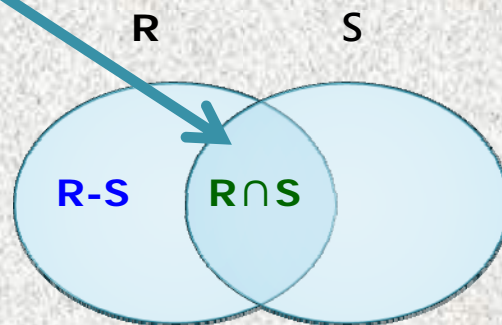
共同出現的記錄



【概念分析】



$$R \cap S = R - (R - S) = 3 \text{筆} - 1 \text{筆} = 2 \text{筆}$$



【實例】

請利用交集(Intersection)來查詢學生資料表

R 資料表

	學號	姓名	性別	身高	體重
#1	S0001	張三	男	175	75
#2	S0002	李四	男	169	65
#3	S0004	林六	女	158	45

S 資料表

	學號	姓名	性別	身高	體重
#1	S0001	張三	男	175	75
#2	S0005	陳靜	女	163	50

1筆共同出現的記錄
(亦即 $R \cap S$)

請問 $R \cap S = ?$

【解答】

	學號	姓名	性別	身高	體重
#1	S0001	張三	男	175	75

【練習】

若 $A=\{1,2,3,4\}$ ， $B=\{3,4,5,6\}$ ，則 $A \cap B$

請問 $A \cap B = ?$

【解答】

$A \cap B = \{3,4\}$

除法(Division)

【定義】

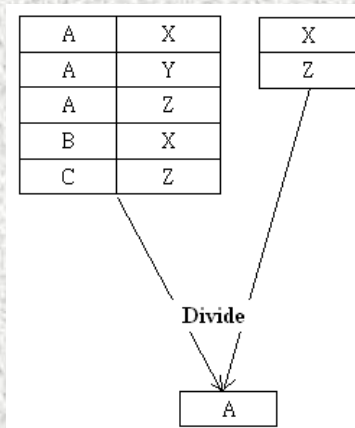
此種運算如同數學上的除法一般，有二個運算元：**第一個關聯表R**當作「**被除表格**」，**第二個關聯表S**當作「**除表格**」。

其中「被除表格」的屬性必須比「除表格」中的任何屬性中的值域都要與「被除表格」中的某屬性之**值域相符合**。

【關聯式代數】 $R \div S$

【概念圖】

關聯 R 與關聯 S 作「除法」運算時，只作用在兩個關聯中**相同的部份**。



【簡易作法】

關聯R「除以」關聯S時，則分兩個步驟來處理

步驟1：

檢查關聯R中的每一列資料，若有包含關聯S中的某一列時，則將該列資料取出

步驟2：

將步驟1取出資料列，再刪掉關聯S之資料行

【實例分析】

假設現在有關聯R與關聯S，如下圖所示：

關聯R

A	X
A	Y
A	Z
B	X
C	Z

關聯S

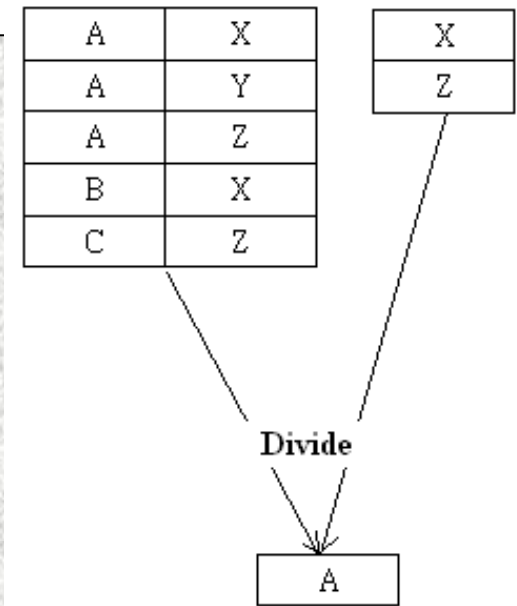
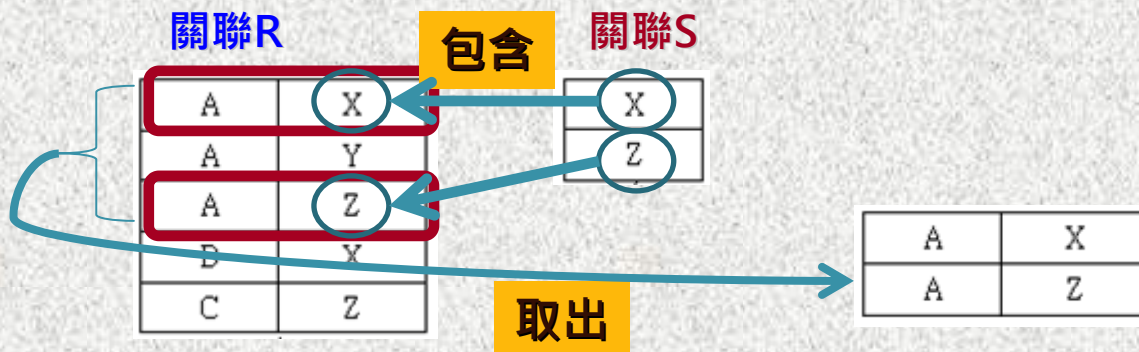
X
Z

現在欲作 $R \div S$ 時，則其執行過程為何？

【解答】

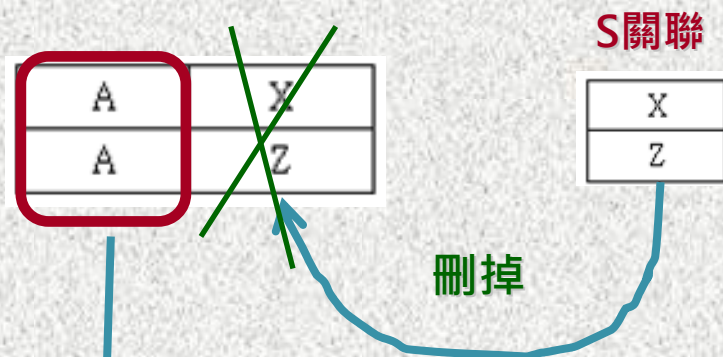
步驟1：

檢查關聯R中的每一列資料，若有包含關聯S中的每一列時，則將該列資料取出



步驟2：

將步驟1取出資料列，再刪掉S關聯之資料行



刪除後的結果
(重複時，只出現一次)

The diagram shows the result of the deletion: a single box containing the letter 'A'.

A
