TR-82-013

Optimization of Query Processing in Distributed Database Systems

by

P. T. Chang

Institute of Information Engineering
Tamkang University, R. O. C.

J. S. Ke

Institute of Information Science
Academia Sinica
R, O. C.

August 1982

第一章　引介

分散式資料庫系統(DDBS)是計算機網路和資料庫兩大技術之結合，前者代表資訊之廣泛分佈，後者代表資訊之整合。較之傳統集中式資料庫系統，DDBS的主要優點爲：分散資料庫管理系統的負荷，縮短詢問的回覆時間，以及增强系統之可靠性。DDBS的觀念新顯引人，但同時也產生了一些異於傳統資料處理的問題，其中之一就是，詢問處理。

本文旨在以一圖形方法，稱之爲投影圖映 (Graph Projection Map)，來解決 DDBS詢問處理（以下簡稱DDBS.QP) 的問題。第二章介紹 DDBS.QP 之功能架構與 DDBS.QP 之特性，定義 DDBS.QP 問題之所在，並列舉本文中所作的一些假設條件。第三章定義詢問之圖形表示法及其類別。第四章以投影圖映來引証一測定詢問類別的方法。此測定法與其他方法最大的不同，除了非常簡單外，還有此法祗做詢問類別之測定，而不在此就只求一未必爲最佳之等效詢問，如〔BECH 81〕〔BEGO 79〕〔YUOZ 79〕〔OZYU 80〕等做法。第五章敍述求出所有等效詢問的方法。因爲由遞移律 (Law of Transitivity) 一詢問可能有多個等效詢問，不同的等效詢問由不同的子詢問組所構成，而不同的子詢問組其處理費用也可能不一。

第二章　系統架構

§2.1　DDBS.QP 之功能架構

目前討論 DDBS.QP 的文章，大多是以關係模式(Relational Model)資料庫系統爲對象。一般而言，其DDBS.QP 的方式可分爲直接結合法 (Direct Joins Approach)與半結合法 (Semi-Joins Approach)二種，這已於〔BERN81〕中討論。在此，吾人先勾劃出 DDBS.QP 之功能架構 (圖2.1)。

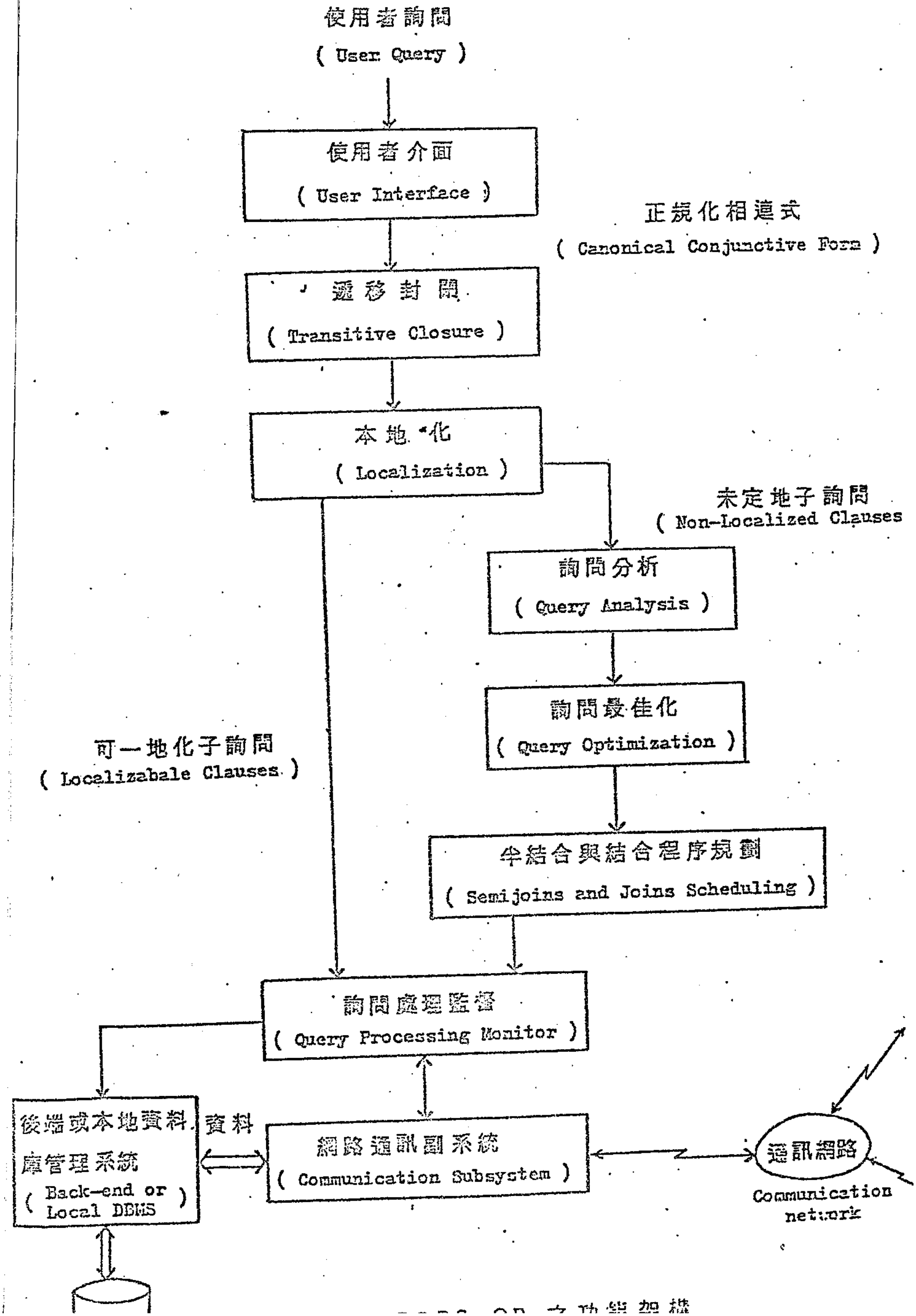
使用者詢問經由使用者介面轉譯成正規化的型式 (canonical form)，其詢問內涵(qualifications)之型式爲((A∧B∧ ...)∨(C∧D∧ ...)∨...) 其中 A,B,C,D, ... 表子詢問(query clauses)。在本文中，吾人同 [WONG 77] [CHHO 80] [BERN 81] [YLCC 82] 等文章，所討論的對象都是相連式 (conjunctive form)之子詢問，且作結合之子詢問皆爲等結合(equi-join)；如例2.1 所示。

例2.1: 設一詢問Q(TL,q), TL 爲其目標列 (Target List),

$q = (R1.A1 = R2.A2) \wedge (R2.A2 = R3.A3) \wedge (R3.A3 > 100)$,

q 爲詢問Q 之內涵。

〔定義2.1 〕遞移封閉 (Transitive Closure):對一詢問 Q(TL,q)，其相連式之內涵 q 中，若存在任二個子詢



使用者詢問
(User Query)
使用者介面
(User Interface)
正規化相連式
(Canonical Conjunctive Form)
遞移封閉
(Transitive Closure)
本地化
(Localization)
未定地子詢問
(Non-Localized Clauses
詢問分析
(Query Analysis)
詢問最佳化
(Query Optimization)
可一地化子詢問
(Localizabale Clauses)
半結合與結合程序規劃
(Semijoins and Joins Scheduling)
詢問處理監督
(Query Processing Monitor)
後端或本地資料庫管理系統
(Back-end or Local DBMS)
資料
網路通訊副系統
(Communication Subsystem)
通訊網路
Communication network

問 $(X=Y)$ 和 $(Y=Z)$，則由遞移律將產生第三個子詢問 $(X=Z)$；原來 q 中的子詢問加上所有由遞移律產生的子詢問，稱為此詢問內涵 q 之遞移封閉，記為 q^+，而 q^+ 對應之詢問為 $Q^+(TL,q^+)$。

例 2.1 之詢問 $Q(TL,q)$ 做遞移封閉後，$Q^+(TL,q^+)$，

$q^+ = (R1.A1 = R2.A2) \wedge (R2.A2 = R3.A3) \wedge (R3.A3 > 100) \wedge \underline{(R1.A1 = R3.A3)} \wedge \underline{(R1.A1 > 100)} \wedge \underline{(R2.A2 > 100)}$，劃底線的部分為由遞移律所產生的子詢問。從這裡可看出，作遞移封閉可產生所有相關的子詢問；因此，所有相關的單變數子詢問，如例 2.1 中之 $(R1.A1>100)$, $(R2.A2>100)$, $(R3.A3>100)$ 可先計算以減小將來做結合時關係檔的大小 [WONG 77]。

詢問內涵經過遞移封閉後，由系統之資料檔目錄可知那些子詢問祇參考到一地的資料，那些子詢問參考到不同地的資料。因為祇參考到一地資料的子詢問可交予該地之資料庫管理系統（local DBMS）先行處理；所以吾人將子詢問分為可一地化（ localizable ）子詢問與未定地（ non-localized ）子詢問二種。可一地化的子詢問包括所有單變數子詢問和屬於一地雙變數子詢問；未定地子詢問則為參考到不同地的雙變數子詢問。

去掉了可一地化之子詢問，吾人分析未定地之雙變

數子詢問組之詢問類別，即測定詢問是為樹狀詢問(Tree Query)抑或環狀詢問(Cyclic Query)，然後依詢問處理之策略求其最佳之詢問表示式。對此最佳之詢問表示式，再安排其半結合與結合之程序與地點，經由通訊網路分別交與各地DBMS去執行，同時在詢問地產生一監督程式以負責協調處理程序之執行。

爲明白地表示詢問內涵中子詢問所參考到的關係檔之間的關係，吾人可以一些圖形表示之。相似的表法參見[BECH 81]。

〔定義 3.1 〕詢問圖 (Query Graph)：對一詢問 Q(TL,q)，$G_Q(V_Q, E_Q)$ 表 Q(TL,q) 之詢問圖；$V_Q = \{ Ri \mid Ri$ 爲q所參考到的關係檔$\}$ 是爲節點(Vertices) 所成之集合，$E_Q = \{ (Ri,Rj) \mid i \neq j$ 且 q 中某子詢問同時參考到 Ri 和 Rj$\}$ 是爲支邊(Edges) 所成之集合。

例 3.1：設詢問 Q(TL,q)，$q = (R1.A1 = R2.A2) \wedge (R1.A1 = R3.A3) \wedge (R1.A1 = R5.A5) \wedge (R1.B1 = R6.B6) \wedge (R1.B1 = R5.B5) \wedge (R5.C5 = R6.C6) \wedge (R4.D4 = R5.D5) \wedge (R4.D4 = R6.D6) \wedge (R1.E1 = R6.E6) \wedge (R5.E5 = R6.E6) \wedge (R1.F1 = R2.F2) \wedge (R2.F2 = R3.F3)$，

此詢問之詢問圖如圖 3.1 所示；爲方便辨認，吾人在支邊上標出結合域 (join domains)名。

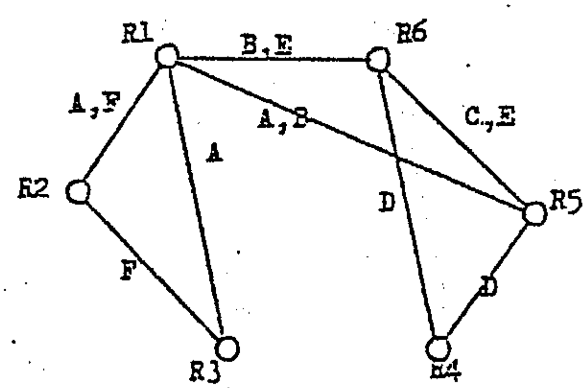


圖 3.1：例 3.1 之詢問圖 G_Q

〔定義3.2 〕結合圖(Join Graph):對一詢問Q(TL,q), $J_Q(V_J, E_J)$ 表其結合圖。$V_J=\{Ri.Ai \mid Ai$ 為結合屬性(join attribute)且 Ri.Ai 為q 所參考到$\}$為節點所成之集合,$E_J=\{(Ri.Ai,Rj.Aj) \mid i \neq j$ 且 q 中某子詢問同時參考到 Ri.Ai 和 Rj.Aj$\}$為支邊所成之集合。

例3.1 之結合圖如圖3.2 所示。

在結合圖中,相連接的部分圖形吾人稱之為結合域子圖(Join Domain Subgraph);由定義3.2 與圖3.2 可看出,每一結合域子圖是由俱有相同結合域之子詢問所形成的結合圖,如圖3.2 之 $J_Q^A, J_Q^B, \ldots, J_Q^F$ 等。

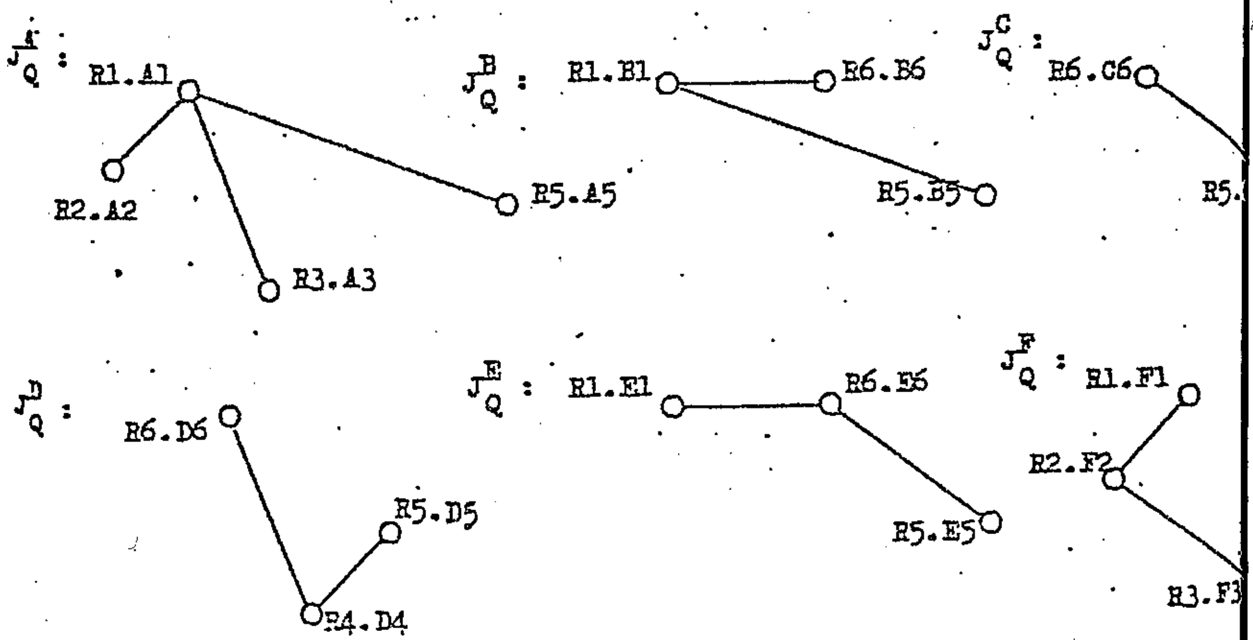


圖 3.2: 例3.1 之詢問 Q 之結合圖 J_Q,

$J_Q = J_Q^A \cup J_Q^B \cup \ldots J_Q^F$.

遞移封閉表現在結合圖上的意義即是對每一結合域子圖作完全連結；既然結合域子圖是參考到相同的結合域，故在不致令人誤解下，節點可不必再列出其結合屬性；圖 3.3 表示例3.1 作遞移封閉後之結合圖。

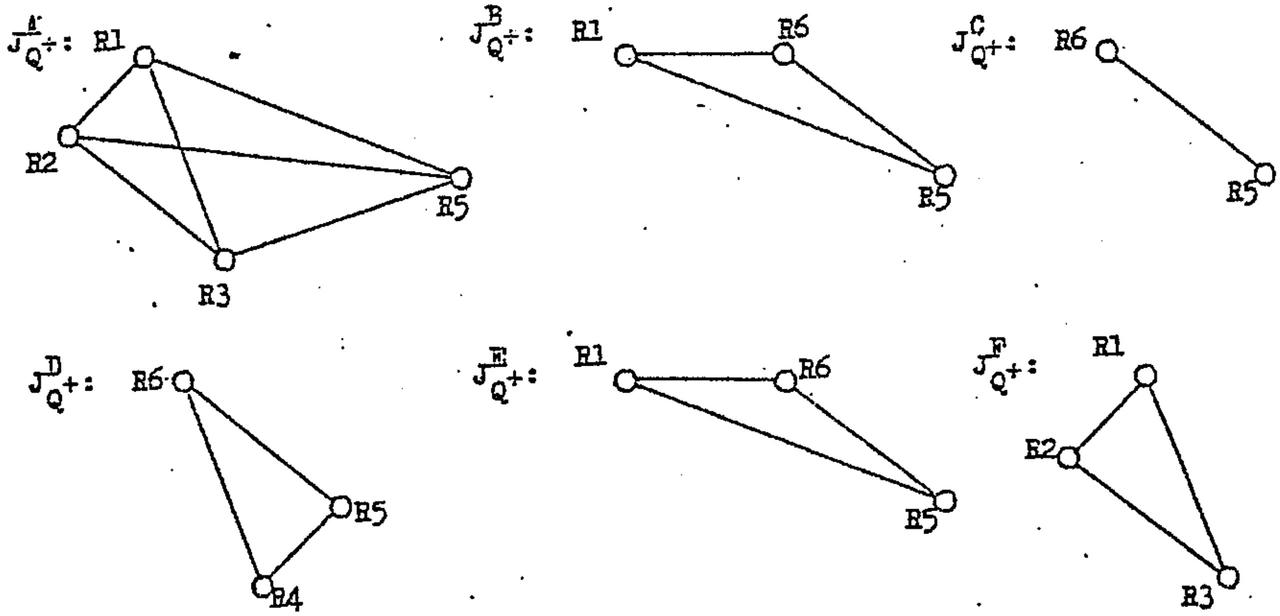


圖3.3: 例3.1 之詢問Q 作遞移封閉之結合圖 J_{Q^+}，

$J_{Q^+} = J_{Q^+}^A \cup J_{Q^+}^B \cup \ldots \cup J_{Q^+}^F$

詢問分析主要是在作詢問類別之測定。就詢問圖形而言，詢問分為兩大類，即樹狀詢問與環狀詢問[BECH 81]，分別定義如下：

〔定義3.3〕樹狀詢問 (Tree Query): 一詢問（或是其等效詢問）之詢問圖是為樹狀。

〔定義3.4 〕環狀詢問(Cyclic Query)：不屬於樹狀詢問的詢問皆爲環狀詢問。

設TQ爲樹狀詢問所成之集合，CQ爲環狀詢問所成的集合，由定義可知 $TQ \cap CQ = \emptyset$ 且 $TQ \cup CQ$ 是爲所有詢問所成的集合。

樹狀詢問與環狀詢問在DDBS.QP 作法上有很大的差異，本文主要在對樹狀詢問做一探討。

第四章　樹狀詢問之測定

吾人將詢問類別之測定轉化成圖形理論上的問題。

在未正式接觸問題之前，先介紹下面幾個定義和定理：

〔定義 4.1 〕等效詢問（ Equivalent Queries ）：若二詢問 $Q_1(TL,q_1)$, $Q_2(TL,q_2)$ ，對任一資料庫狀態，詢問所得之結果相同，則稱 Q_1, Q_2 爲等效詢問，記爲 $Q_1 \equiv Q_2$。

〔定理 4.1 〕一詢問 $Q(TL,q)$ 其遞移封閉 $Q^+(TL,q^+)$ 是唯一的且 $Q^+ \equiv Q$ 。

証明：很明顯地，由遞移律所產生的 $Q^+(TL,q^+)$ 與 $Q(TL,q)$ 是爲等效詢問，$Q^+ \equiv Q$ 。

由結合圖之定義可知，一詢問 $Q(TL,q)$ 與其結合圖 $J_Q(V_J, E_J)$ 是一對一對應；$Q^+(TL,q^+)$ 之意即是對 J_Q 之結合域子圖作完全連結；因爲任一節點集合,其完全連結圖是唯一決定，所以 $Q(TL,q)$ 之遞移封閉 $Q^+(TL,q^+)$ 是唯一的。　□得証□

〔定理 4.2 〕設二詢問 $Q_1(TL,q_1)$, $Q_2(TL,q_2)$，若 q_1, q_2 有相同的遞移封閉，${q_1}^+ = {q_2}^+$,則 $Q_1 \equiv Q_2$ 。

証明：$\because {Q_1}^+(TL,q_1^+) \equiv Q_1(TL,q_1)$, ${Q_2}^+(TL,{q_2}^+) \equiv Q_2(TL,q_2)$

又 $\because {q_1}^+ = {q_2}^+$, $\therefore {Q_1}^+(TL,{q_1}^+) \equiv Q_2^+(TL,{q_2}^+)$ ，

故 $Q_1 \equiv {Q_1}^+ \equiv {Q_2}^+ \equiv Q_2$ 。　□得証□

〔定義 4.1 〕投影圖映 (Graph Projection Map) : 設二圖形 $G_1(V_1, E_1)$, $G_2(V_2, E_2)$,投影圖映 p 是為一圖形聯集之運算子；設 $G_p(V, E)$ 為投影圖映後之圖形，則 $G_p = G_1 \xrightarrow{p} G_2$, $V=V_1 \cup V_2$, $E=E_1 \cup E_2$。

例 4.1: $G_p = G_1 \xrightarrow{p} G_2$, $G_1(V_1, E_1)$, $V_1 = \{R1, R2, R3\}$, $E_1 = \{(R1, R2), (R2, R3)\}$; $G_2(V_2, E_2)$, $V_2 = \{R_2, R_3, R_4, R_5\}$, $E_2 = \{(R2, R3), (R3, R4), (R3, R5)\}$

$G_p = G_1 \xrightarrow{p} G_2$, $G(V, E)$, $V = V_1 \cup V_2 = \{R1, R2, R3, R4, R5\}$, $E = E_1 \cup E_2 = \{(R1, R2), (R2, R3), (R3, R4), (R3, R5)\}$ 如圖4.1 所示，$G_p = G_1 \xrightarrow{p} G_2$，即為在相對應的節點上將 G_1 投影在 G_2 上所得之圖映。

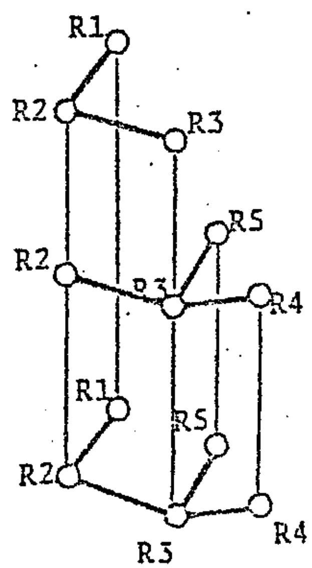


圖4.1: 例 4.1 之 $G_p = G_1 \xrightarrow{p} G_2$

設 V 表一節點集合，SP(V)，若非特別聲明，係指對 V 所形成的完全連結圖（complete graph）之一展開樹（Spanning Tree）。令 $G_i(V_i, E_i)$ 表對同一結合域之詢問組所形成的詢問圖，由定理 4.2 可知 $SP(V_i)$ 所代表的詢問與 G_i 所代表的詢問是為等效，且 $SP(V_i)$ 是為最少結合數的等效詢問。設一詢問 Q(TL,q), m 為 Q 中不同的 G_i 個數，其詢問圖 $G_Q(V_Q, E_Q)$, $V_Q=\bigcup_{i=1}^{m} V_i$, $E_Q=\bigcup_{i=1}^{m} E_i$, 也就是重覆作 $G_p= G_i \xrightarrow{P} G_p$, $i= 1,2, \ldots ,m$。於是，測定一詢問 Q 之詢問類別是否為樹狀（或環狀）的問題，吾人可視之為：是否存在一 $SP(V_Q)$ 可使得對每一結合域之詢問圖 $G_i(V_i,E_i)$ 存在 $SP(V_i) \subseteq SP(V_Q)$。這也就是重覆作 $G_p = G_i \xrightarrow{P} G_p$, $i= 1,2, \ldots,m$，一開始 $G_p= \phi$，而看最後等效詢問圖 G_p 是否為樹狀。若最後之 G_p 為樹狀則此詢問 Q 為樹狀詢問，不然則 Q 為環狀詢問。吾人已知，對一連結的簡單圖（simple connected graph），若且唯若其支線數等於節點數減一（即 $|E|=|V|-1$），則此圖形是為樹狀 [DEO 74]；由此，吾人可得下面之定理，用以測定一詢問之類別。

〔定理 4.3〕設 m 為一詢問 Q 不同的結合域子圖之個數，V_i 表第 i 個結合域之子詢問組所參考到的關係檔

所成之集合，若且唯若存在一組 $SP(V_1), SP(V_2), \ldots$ $SP(V_m)$ 其相異支邊數等於 $|V_Q| - 1$，則此詢問 Q 為樹狀詢問。

証明：因為重覆作 $G_p = SP(V_i) \xrightarrow{p} G_p$, $i = 1, \ldots, m$，一開始 $G_p = \phi$，最後之 $G_p = \bigcup_{i=1}^{m} SP(V_i)$， 故若存在一組 $SP(V_1)$, $SP(V_2), \ldots\ldots, SP(V_m)$ 其相異支邊數等於 $|V_Q| - 1$ 則 G_p 為樹狀，亦則 Q 為樹狀詢問。

反之，若 Q 為樹狀詢問, G_p 必為樹狀；於是就每一組 $SP(V_1), \ldots, SP(V_m)$，其相異支邊數不等於 $|V_Q|-1$ 之情形分小於和大於二種情形來討論：

(i) 小於 $|V_Q|-1$;

小於的情形不會發生，因為小於 $|V_Q|-1$ 之意為 V_Q 中某些節點在 V_i 中未被參考到，與 $V_Q = \bigcup_{i=1}^{m} V_i$ 相矛盾。

(ii) 大於 $|V_Q|-1$;

$G_p = \bigcup_{i=1}^{m} SP(V_i)$，若自一組 $SP(V_1), \ldots, SP(V_m)$ 相異支邊數皆大於 $|V_Q| - 1$，由樹狀圖之判別式 $|E| = |V| - 1$， 可知 G_p 不為樹狀，故 Q 為環狀詢問而非樹狀詢問。

因此，若每一組 $SP(V_1), \ldots, SP(V_m)$ 相異支

邊數都不等於 $|V_Q|-1$ 則Q 為環狀詢問，而此命題是為〃唯若〃命題之逆命題（同義命題）。

□得証□

由定理4.3 吾人可直接導出測定樹狀詢問的方法，即計算詢問Q 其m 個相異 G_i 之 $SP(V_1),\ldots,SP(V_m)$ 相異支邊數，若等於 $|V_Q|-1$ 則此詢問為樹狀詢問，一旦大於 $|V_Q|-1$ 則此詢問為環狀詢問。圖 4.2 即為例3.1 測定之情形，為方便表示計算之情形，吾人設 $|V_1|\leq|V_2|\leq\cdots|V_m|$，從 V_1 啟起，設 δ_i 表從 $SP(V_1)$ 到 $SP(V_i)$ 相異支邊數，由圖上之計算可看出 $SP(V_1)$, ..., $SP(V_m)$ 相異支邊數等於 $|V_Q|-1$；故例 3-1 之詢問為樹狀詢問。圖 4.3 是為環狀之例子。

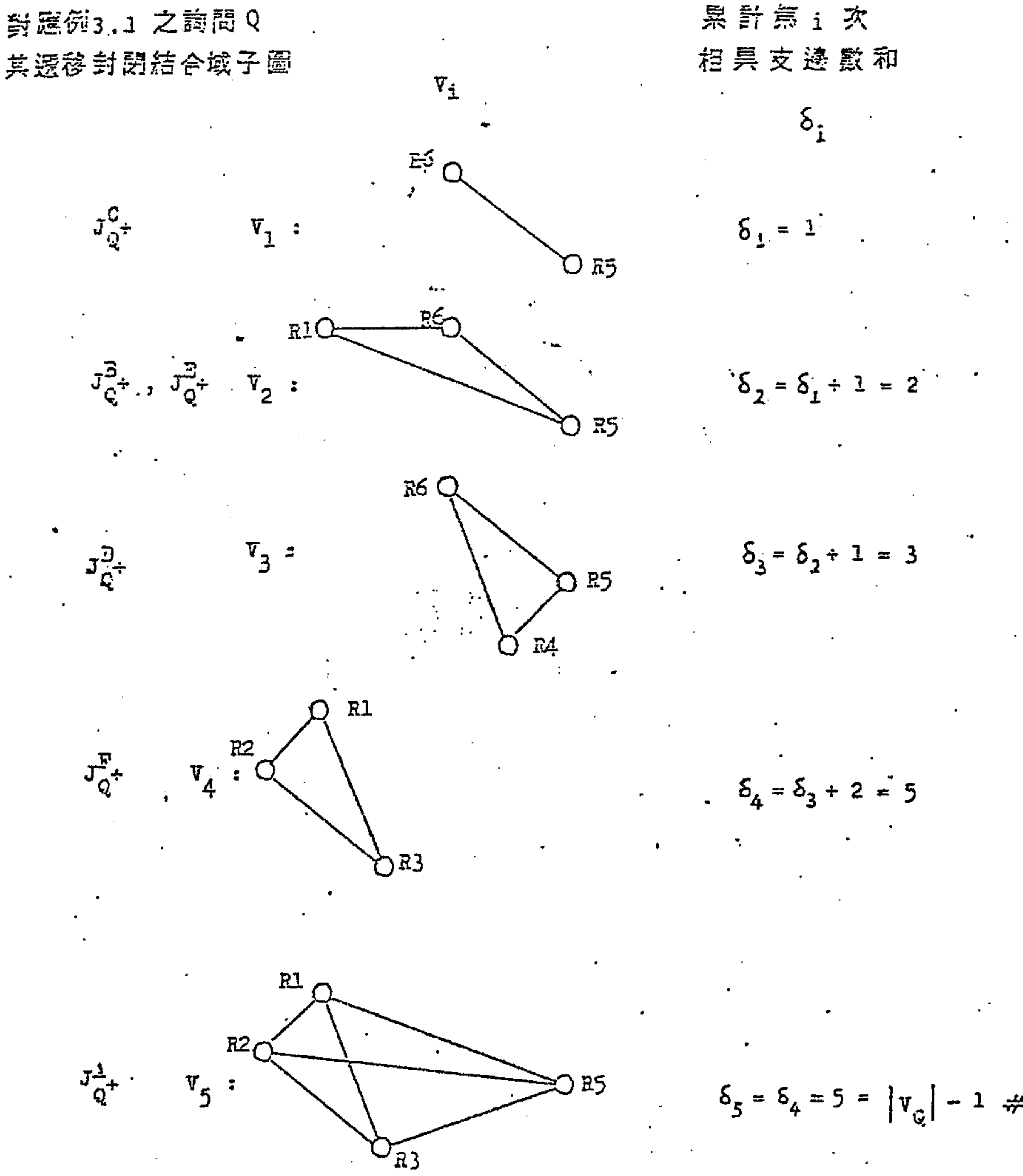
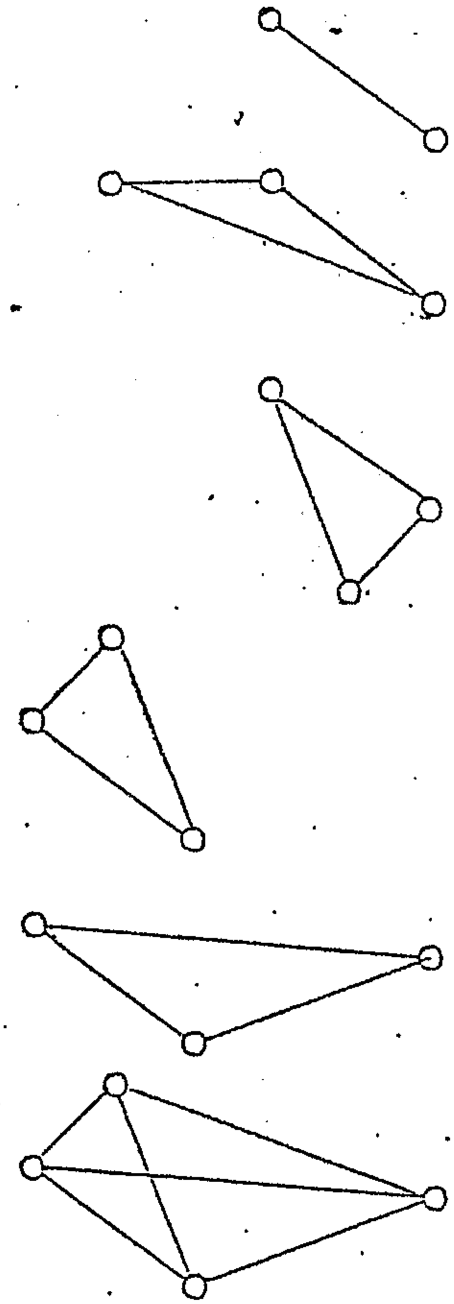


圖 4.2 $G_p = SP(V_i) \xrightarrow{P} G_p$, i=1,2, ...,5 ,
G_p最後可爲樹狀之例子(例2.1)

累計第 i 次
相異支邊數和

V_i δ_i



V_1 : $\delta_1 = 1$

V_2 : $\delta_2 = \delta_1 + 1 = 2$

V_3 : $\delta_3 = \delta_2 + 1 = 3$

V_4 : $\delta_4 = \delta_3 + 2 = 5$

V_5 : $\delta_5 = \delta_4 + 1 = 6 > |V_Q| - 1$ #

V_6 :

/* Since δ_i is monotonic increasing w.r.p. i, so, V_6 would not be necessary to be tested. */

圖4.3 $G_p = SP(V_i) \xrightarrow{P} G_p$, i=1,2, ...,6 ,
G_p最後必不爲樹狀之例子

第五章 等效詢問之最佳化

在第三章已提過等效詢問之表法可能不只一種，在本章，吾人以投影圖映可求出所有的等效詢問圖，從而求出最少層次中最長鏈的等效樹狀詢問圖。

吾人知道，對一結合域子圖 $G_i(V_i, E_i)$, $SP(V_i)$ 是為最少結合數的等效結合域子圖。對一樹狀詢問 Q(TL, q),m 為其相異結合域子圖的個數，為方便計算，設 $|V_1| \le |V_2| \le \ldots \le |V_i| \le \ldots \le |V_m|$；一開始 $G_p = \phi$, $G_p = SP(V_i) \xrightarrow{P} G_p$, i=1,...,m, 於是最後之 G_p 為樹狀之所有可能情形，即為所有的等效樹狀詢問。

因為樹狀圖是最少支邊的連結圖，所以我們在作 $G_p = SP(V_i) \xrightarrow{P} G_p$ 時，$SP(V_i)$ 支邊的選擇是以已經出現在投影圖 G_p 者為優先。下面定理 5.1 到定理 5.6 ，係討論一般化 (generalized) 第 i 次投影圖映時，$SP(V_2) \xrightarrow{P} SP(V_1)$ 之所有可能情形。分析兩集合 V_1, V_2 之間的關係可能是 $V_1 = V_2$ 或 $V_1 \neq V_2$；在 $V_1 \neq V_2$ 中又可分為 $V_1 \supset V_2$ 與 $V_1 \not\supset V_2$ 二種；在 $V_1 \not\supset V_2$ 情形中又可再分為 $V_1 \cap V_2 = V_{12}$, $|V_{12}| \le 1$

及 $V_1 \subset V_2$ 和 $V_1 \cap V_2 = V_{12}$, $|V_{12}| \geq 2$ 等三種情形。定理5.1和 5.2 係考慮 $V_1 = V_2$ 與 $V_1 \supset V_2$ 時，$SP(V_2) \xrightarrow{P} SP(V_1)$ 為樹狀之情形只有一種。定理5.3 考慮 $V_1 \cap V_2 = V_{12}$ 且 $|V_{12}| \leq 1$，$SP(V_2) \xrightarrow{P} SP(V_1)$ 為樹狀之情形有 m^{m-2} 種，$m = |V_2|$。定理 5.4 考慮 $V_1 \neq V_2$ 且 $V_1 \subset V_2$, $SP(V_2) \xrightarrow{P} SP(V_1)$ 為樹狀之情形有 $n \cdot (n+1) \cdot \ldots \cdot (m-1)$ 種；$m = |V_2|$, $n = |V_1|$。定理5.5 証明了當 $V_1 \cap V_2 = V_{12}$, $|V_{12}| \geq 2$ 時，存在 $SP(V_2) \xrightarrow{P} SP(V_1)$ 為樹狀之充分必要條件為: V_{12} 中每一點在 $SP(V_1)$ 上不經由 $V_1 - V_{12}$ 中任一點而相連結。當 V_{12} 中存在某一點與中另外一點在 $SP(V_1)$ 上的路徑 (path)須經過 $V_1 - V_{12}$ 之節點時，$SP(V_2) \xrightarrow{P} SP(V_1)$ 必不為樹狀；這是 $SP(V_2) \xrightarrow{P} SP(V_1)$ 會產生環唯一的情況。定理 5.6 係針對符合定理 5.5 之充要條件下，証明 $SP(V_2) \xrightarrow{P} SP(V_1)$ 為樹狀之情形有 $h \cdot (h+1) \ldots (m-1)$ 種, $h = |V_{12}|$, $m = |V_2|$。

〔定理 5.1 〕 V_1, V_2 為節點所成之集合, $V_1 = V_2$, 若且唯若 $SP(V_2) = SP(V_1)$ 則 $G_p = SP(V_2) \xrightarrow{P} SP(V_1)$, G_p 為樹狀是唯一決定，即 $G_p = SP(V_1)$。

証明：(i)"若"命題顯然成立。$\because SP(V_2) = SP(V_1)$, $\therefore G_p = SP(V_1) \cup SP(V_2) = SP(V_1)$。

(ii)"唯若"命題亦即是：若 $SP(V_2) \neq SP(V_1)$ 則 $G_p =$

$SP(V_2) \xrightarrow{P} SP(V_1)$, G_p 不爲樹狀。由展開樹的性質，在 $SP(V_1)$ 上加一支邊 e, $e \notin SP(V_1)$ 必定造成一環；

$\because SP(V_2) \neq SP(V_1)$ 故必存在至少一 $e \in SP(V_2)$ 且 $e \notin SP(V_1)$, $\therefore G_p = SP(V_2) \xrightarrow{P} SP(V_1) = SP(V_2) \cup SP(V_1)$，若 $SP(V_2) \neq SP(V_1)$ 則 G_p 必不爲樹狀。　　□得証□

〔定理 5.2 〕V_1, V_2 爲節點所成之集合，$V_1 \subseteq V_2$ 對一已知之 $SP(V_1)$ 必定存在唯一的 $SP(V_2)$ 使得 $SP(V_2) \subseteq SP(V_1)$; 亦即 $G_p = SP(V_2) \xrightarrow{P} SP(V_1)$, G_p 爲樹狀, $G_p = SP(V_1)$ 是唯一決定。

証明：若 $V_1 \supseteq V_2$ 則存在 $SP(V_2) \subseteq SP(V_1)$ 是很明顯的，因爲祗要先作好 $SP(V_2)$，然後投影圖映至 $G(V_1, \phi)$ 可得 $SP(V_2) \xrightarrow{P} G(V_1, \phi) = G'(V_1, SP(V_2))$, 由 $G'(V_1, SP(V_2))$ 繼續展開 $V_1 - V_2$ 諸節點，即可得 $SP(V_2) \subseteq SP(V_1)$。

於是，祗要再証明對已知一 $SP(V_1)$ 使得 $SP(V_2)$ $SP(V_1)$ 之 $SP(V_2)$ 是唯一的即可。設 $SP_1(V_2) \subseteq SP_2(V_1)$, $SP_2(V_2) \subseteq SP_1(V_1)$, 由定理 5.1, 若 $SP_1(V_2) \neq SP_2(V_2)$，則 $SP_1(V_2) \cup SP_2(V_2)$ 不爲樹狀，但因爲二子集之聯集亦爲子集，所以 $SP_1(V_2) \cup SP_2(V_2) \subseteq SP(V_1)$ 是應爲樹狀；矛盾！故 $SP_1(V_2) = SP_2(V_2)$; 即對一已知之 $SP(V_1)$, $SP(V_2) \subseteq SP(V_1)$, $SP(V_2)$ 是唯一的。　□得証□

〔定理 5.3 〕 V_1，V_2 爲節點所成之集合 $V_1 \cap V_2 = V_{12}$，$|V_{12}| = h$，$|V_2| = m$，對一已知之 $SP(V_2)$，$G_p = SP(V_2) \xrightarrow{P} SP(V_1)$，$G_p$ 爲無環 (circuit-free) 圖之情形有 m^{m-2} 種。

証明：由卡雷氏 (Cayley's) 定理 [EVEN 79]，可知 m 相異節點作展開樹有 m^{m-2} 種。$\because |V_{12}| \leq 1$，$\therefore V_1, V_2$ 之交集爲空集合或是一單一節點之集合。

(i) 若 $|V_{12}| = 0$，即 $V_{12} = \phi$，則 $SP(V_2) \cap SP(V_1) = \phi$，所以 $G_p = SP(V_2) \xrightarrow{P} SP(V_1)$ 有無環之情形爲 V_2 之圖形所決定；由卡雷氏定理知，不同的 $SP(V_2)$ 有 m^{m-2} 種，所以 $G_p = SP(V_2) \xrightarrow{P} SP(V_1)$ 爲無環圖之情形有 m^{m-2} 種。

(ii) 若 $|V_{12}| = 1$ 則 $SP(V_2) \cap SP(V_1)$ 之圖形沒有邊線存在，所以 $G_p = SP(V_2) \xrightarrow{P} SP(V_1)$ 在 V_1 上不會產生環，故 G_p 有無環之情形爲 V_2 之圖形所決定；同上，G_p 爲無環圖之情形有 m^{m-2} 種。 □得証□

〔定理 5.4 〕 V_1，V_2 爲節點所成之集合，$V_1 \subset V_2$，$|V_2| = m$，$|V_1| = n$，$m > n$，$G_p = SP(V_2) \xrightarrow{P} SP(V_1)$，$G_p$ 爲樹狀之情形有 $n \cdot (n+1) \cdots (m-1)$ 種。

証明：吾人先考慮 V_1 與 V_2 相交的部份。由定理 5.2 可知相交的部份使 G_p 爲樹狀之情形是唯一。

再考慮不相交的部分。若不相交部分祗有一節點 v_1'，則使 G_p 爲樹狀，$SP(V_2)$ 之情形爲 v_1' 與 V_1 中任一點相連結，即 $C_1^n=n$ 種情形；由此可推及不相交部分不祗一節點的情形可視爲第一次是在 n 點中選一點，第二次是在 $(n+1)$ 點中選一點，…直到不相交部分的最後一點是在 $(n+m-n-1)$ 點中選一點。故 G_p 爲樹狀之情形有 $n\cdot(n+1)\ \ldots\ (m-1)$ 種。　　□得証□

〔定理 5.5〕V_1, V_2 表節點所成之集合，$V_1 \subsetneq V_2$, $V_2 \subsetneq V_1$ 且 $V_1 \cap V_2 = V_{12}$, $|V_{12}| \geq 2$；若且唯若 V_{12} 中每一節點在 $SP(V_1)$ 上與其它 V_{12} 中最近的節點之距離(distance)等於 1，則存在至少一 $SP(V_2)$ 使得 $G_p=SP(V_2) \xrightarrow{P} SP(V_1)$ 爲樹狀。

證明：(i)"若"命題；

"若"命題的前題條件之敘述意即 V_{12} 中每一點（不經由 V_1-V_{12} 中任一點）相連結。因爲任一相連結的圖形必定存在至少一展開樹 [DEO 74]，所以吾人可作一 $SP(V_{12})$，由定理5.2 可知 $SP(V_{12}) \subset SP(V_1)$，$G_p' = SP(V_{12}) \xrightarrow{P} SP(V_1)$ 爲樹狀；由此 $SP(V_{12})$ 可繼續對 V_2-V_{12} 展開而得 $SP(V_2)$。因爲 $SP(V_{12}) \subset SP(V_2)$，所以 $SP(V_2) \cap SP(V_1)=SP(V_{12})$ 爲樹狀；且又因爲，若

且唯若一圖形爲樹狀則此圖形上任二節點之路徑是唯一決定[DEO 74]；所以，$SP(V_2) \cup SP(V_1)=G_p=SP(V_2) \xrightarrow{P} SP(V_1)$在 G_p上由此 $SP(V_2)$ 上任一點到 $SP(V_{12})$ 上任一點是唯一決定，且由 $SP(V_{12})$ 上任一點到 $SP(V_1)$ 上任一點之路徑亦是唯一決定，故 $SP(V_2)$上任一點到 $SP(V_1)$上任一點之路徑是唯一決定，所以G_p 爲樹狀。例見圖5.1a 。"若"命題得証。

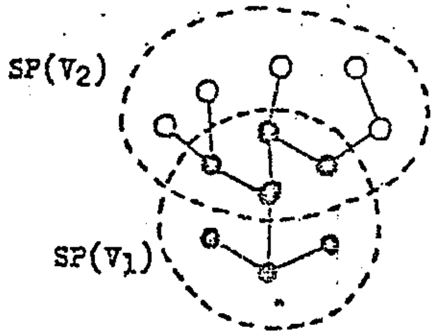
(ii)"唯若"命題；

因爲逆命題與原命題爲同義命題，故吾人祗要証明其逆命題即可。唯若命題即是：若 V_{12} 中某一節點在 $SP(V_1)$ 上與其他 V_{12} 中最近的節點之距離大於 1（"小於"不會發生，因爲 $|V_{12}|\geq 2$，兩相異點之距離不小於 1），則存在任一 $SP(V_2)$ 使得 $G_p= SP(V_2)\xrightarrow{P} SP(V_1)$ 爲非樹狀。

設 $v,v'\in V_{12}$，v在 $SP(V_1)$ 上與其他 V_{12} 中最近的節點v'之距離爲 $d_{min}>1$，此即在 $SP(V_1)$，v 到v' 的路徑 p_1 ，必經過 V_1-V_{12} 中的某些節點，$|p_1|= d_{min}>1$ ；又在G_p 上，$G_p= SP(V_2)\cup SP(V_1)$ 亦可經由 $SP(V_2)$ 到 v' 設此路徑爲 p_2 換言之， v 和v'間有兩個路徑存在，即 G_p存在一環經過v,v' 二點，如圖 5.1b所示。

若 $p_1 \neq p_2$ 則 $d_{min}=1$，與 $d_{min}>1$矛盾，$\therefore p_1=p_2$。

□得証□



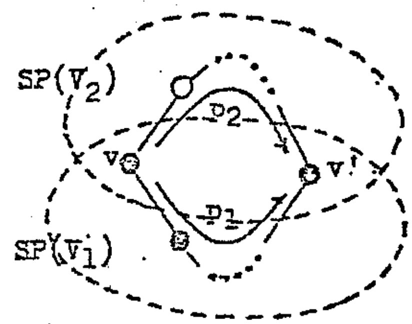


圖 5.1：

$G_p=SP(V_2)\xrightarrow{P}SP(V_1)$,黑點表 V_1 之節點，白點表 V_2 之節點；(a)G_p為樹狀，(b)G_p不為樹狀。

〔定理 5.6〕 V_1, V_2表節點所成之集合，$V_1 \subsetneq V_2$, $V_2 \subsetneq V_1$, $V_1 \cap V_2 = V_{12}$, $|V_{12}| = h$, $|V_2| = m$, $|V_1| = n$, V_{12}中每一點在 $SP(V_1)$上與其他 V_{12} 中最近的節點之距離等於 1, $G_p=SP(V_2)\xrightarrow{P}SP(V_1)$為樹狀之情形有 $h\cdot(h+1)\ldots.(m-1)$ 種。

証明：對 $SP(V_2)$ 吾人分 V_{12} 與 V_2-V_{12}兩部分來討論：

先展開 V_{12}；　已知 $V_{12}\subset V_1$ 且 $G_p=SP(V_{12})\xrightarrow{P}SP(V_1)$ 由定理5.2可知 $G_p=SP(V_{12})$ 是唯一決定。

再展開 V_2-V_{12}；此情形符合定理 5.4 故 $SP(V_2) \xrightarrow{P} SP(V_1)$ 爲樹狀之情形有 $h\cdot(h+1)\ldots(h+m-h-1) = h\cdot(h+1)\ldots.(m-1)$ 種。

□得証□

綜合以上的定理，吾人可整理出一表格（表5.1 ）以示一般化之第i 次投影圖映所有可能的情形。

現在吾人列出求所有等效樹狀詢問的步驟如下；

EQTGEN :

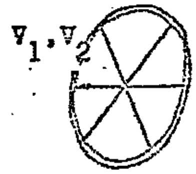
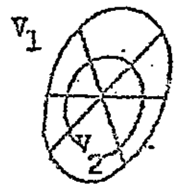
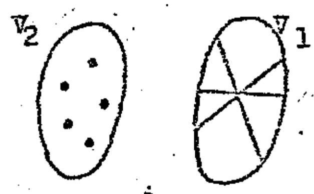
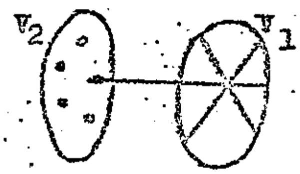
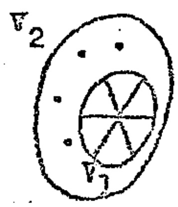
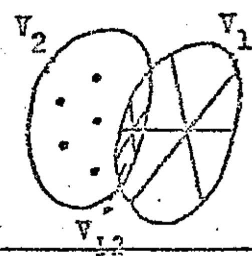
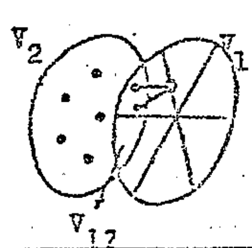
0. 合併所有參考到完全相同關係檔之結合域子圖爲一；設m 爲相異結合域子圖的個數，而其所參考到關係檔分別爲 $V_1, V_2, \ldots, V_m$; 令 $V_Q = V_1 \cup V_2 \cup \ldots V_m$。
1. $G = \phi$, $i = 1$.
2. 求出所有的 $SP(V_i)$ ，作 $G_p = SP(V_i) \xrightarrow{P} G$, G_p 爲 $SP(V_i)$ 與 G 支邊重疊最多之圖，G_p爲樹狀；在G_p支邊誌上對應 V_i之結合域名。
3. 對各個產生的G_p由每一個 V_j, $i < j \le m$, 測試是否符合定理 5.5 之環狀投影圖映的條件？

   不符合的 G_p則令$G = G_p$;繼續第 4 步；

   符合的G_p 則該分支停止。
4. $i = i+1$, 重覆作 2, 3 步； 直到 $i > m$ 或 V_Q 每一節點都已投影到。
5. 若 $i > m$ 則結束。

   若 $i \le m$ 但G_p 每一節點都已投影到則對現有產生的G_p支邊誌上每一個 V_j, $i < j \le m$,對應之結合域名，然後結束。

表 5.1： $G_p = SP(V_2) \xrightarrow{p} SP(V_1)$ 之情形

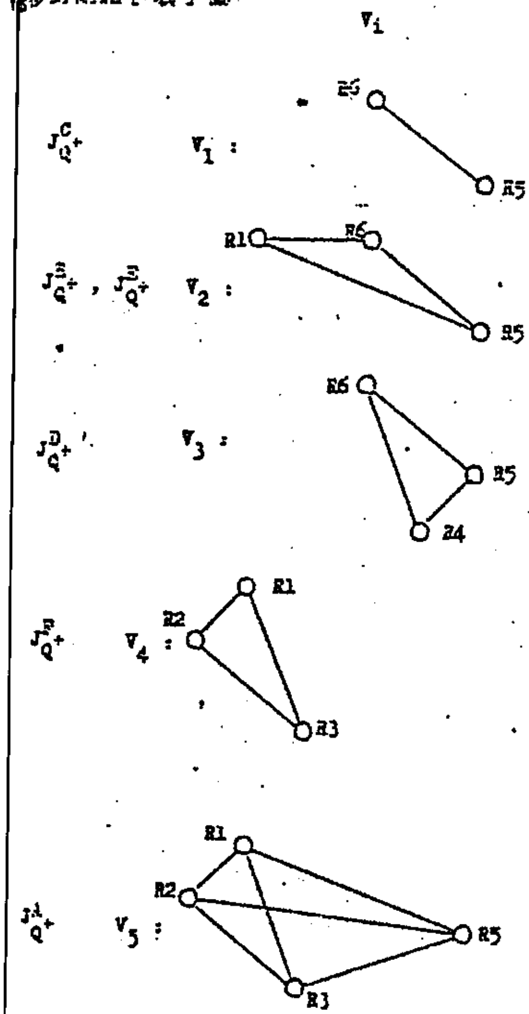
定理條件	示意圖	相異樹狀G_p之個數	註
$V_1 = V_2$		1	
$V_1 \supset V_2$		1	
$V_1 \cap V_2 = V_{12}$ $\lvert V_{12}\rvert \le 1$	 或 	m^{m-2}	$m = \lvert V_2\rvert$
$V_1 \neq V_2$ $V_1 \subset V_2$		$n\cdot(n+1)\ldots(m-1)$	$m = \lvert V_2\rvert$ $n = \lvert V_1\rvert$
$V_1 \cap V_2 = V_{12}$, $\lvert V_{12}\rvert \ge 2$ V_{12}每一點，在$SP(V_2)$上，不經$V_1 - V_{12}$之任一點而相連		$h\cdot(h+1)\ldots(m-1)$	$m = \lvert V_2\rvert$ $h = \lvert V_{12}\rvert$
$V_1 \cap V_2 = V_{12}$, $\lvert V_{12}\rvert \ge 2$ 並非V_{12}每一點在$SP(V_1)$上，不經$V_1 - V_{12}$之任一點而相連		0	

現在我們以例3.1 之詢問Q為例子，用 EQTGEN 來求出Q之所有等效樹狀詢問。如圖5.2 所示，令白點表尚未投影到的節點（關係檔），黑點表已投影到的節點。因為詢問Q中，J_Q^B, J_Q^E 參考到完全相同的關係檔，故這兩個結合域子圖可合併為一考慮，以V_2 表之。一開始$G=\phi$，如 (0)圖所示。$i=1$，作 $G_p=SP(V_1)\xrightarrow{P}G$，因為$V_1$祇參考到二個關係檔，$SP(V_1)$ 祇有1種，所以G_p祇有一個；支邊誌上對應V_1之結合域名C，如 (1)圖所示。(1)圖對每一個V_j，$1<j\leq 5$，$SP(V_1)\xrightarrow{P}(1)$均不符合定理5.5 之環狀投影的條件；故令$G=G_p$，繼續第4步。

$i=2$，$SP(V_2)$有三種，但$G_p=SP(V_2)\xrightarrow{P}G$ 為樹狀之情形，由定理5.4，可知共有2種，即(2)，(3)圖；支邊誌上對應V_2之結合域名B,E。(2)圖因為 $SP(V_5)\xrightarrow{P}(2)$ 符合定理5.5之環狀投影的條件，其情形如圖5.3 所示；於是將來V_5與 (2)圖所產生的樹狀圖之投影圖映必為環狀圖，故該分支停止。(3)圖不符合環狀 投影的條件。

$i=3$，(3)圖，$SP(V_3)\xrightarrow{P}(3)$ 為樹狀之情形有2種，即(4)，(5)圖。$i=4$，$SP(V_4)\xrightarrow{P}(4)$ 及 $SP(V_4)\xrightarrow{P}(5)$ 為樹狀之情形由定理5.3 可知各有3種，即(6)-(11)圖。

例3.1 之詢問Q
移封閉結合域子圖



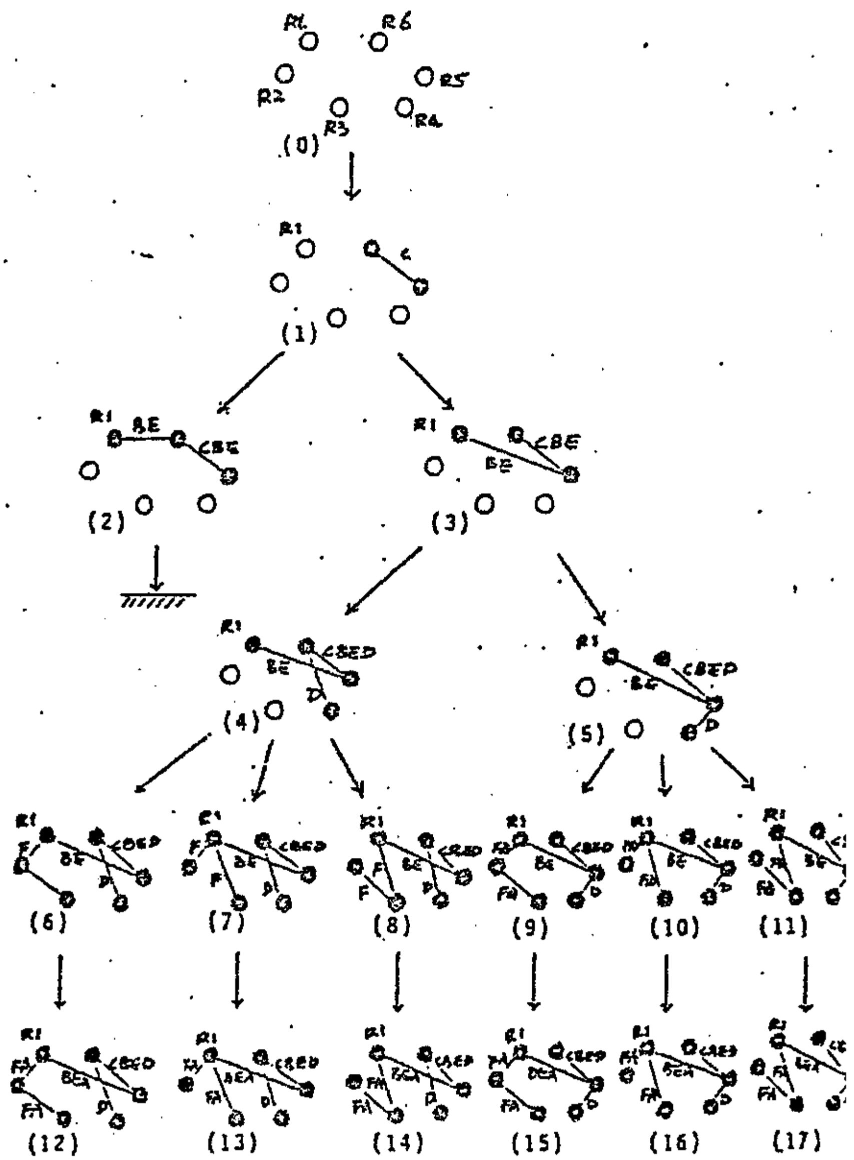
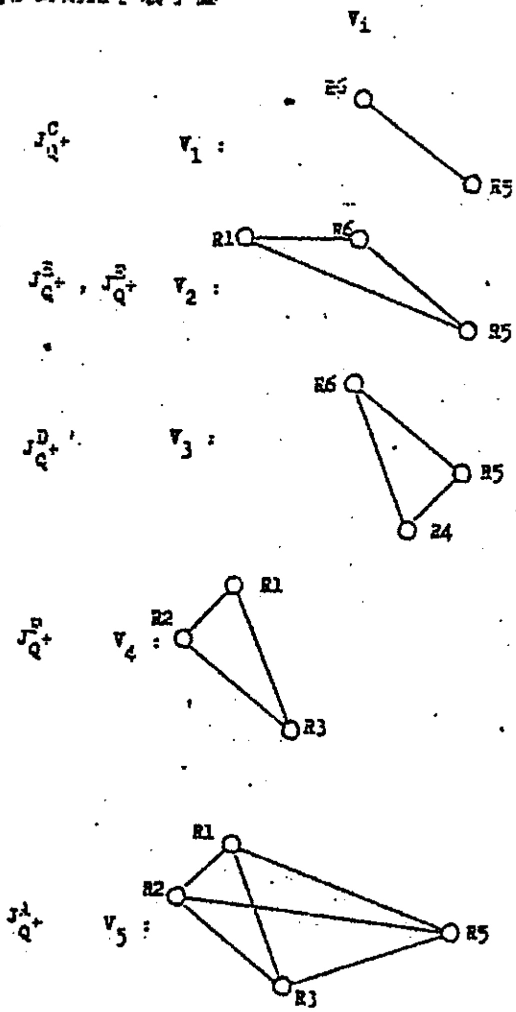


圖5.2：例 3.1 詢問Q 之所有等效樹狀詢問圖的求法，
圖 (12) – (17) 是為 Q之所有等效樹狀

例3.1 之詢問Q
移封閉結合域子圖



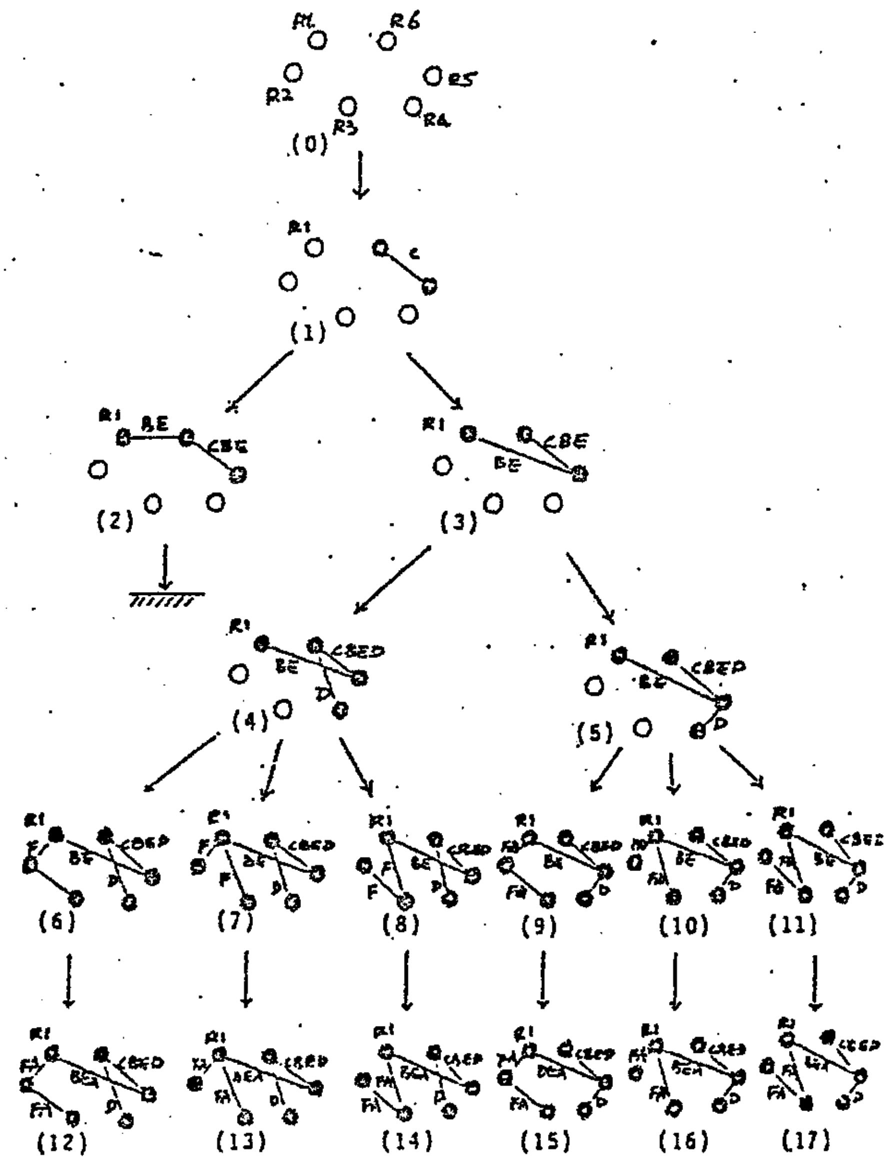


圖5.2：例 3.1 詢問Q 之所有等效樹狀詢問圖的求法，
圖 (12) – (17) 是為 Q之所有等效樹狀

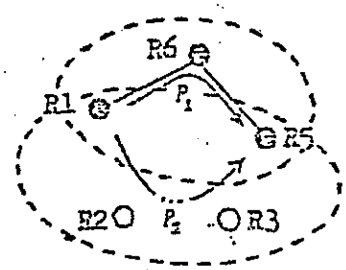


圖 5.3: $G_p=SP(V_5)\xrightarrow{P}(2)$, G_p 必爲環狀圖。

i=5,因爲 (6)-(11) 圖上已經投影到V_Q的每一節點，由定理 5.2 可知 $SP(V_5)$投影到每一 (6)-(11) 圖，爲樹狀之情形均祇有一種，分別爲 (12)-(17) 圖。於是，(12)-(17) 圖即爲 Q 之所有等效樹狀詢問圖。

設 R1 爲發詢地，則 (12)-(17) 圖中，(15)與 (17) 圖是爲最少層次中且最長鏈型之詢問圖（參見下圖）。

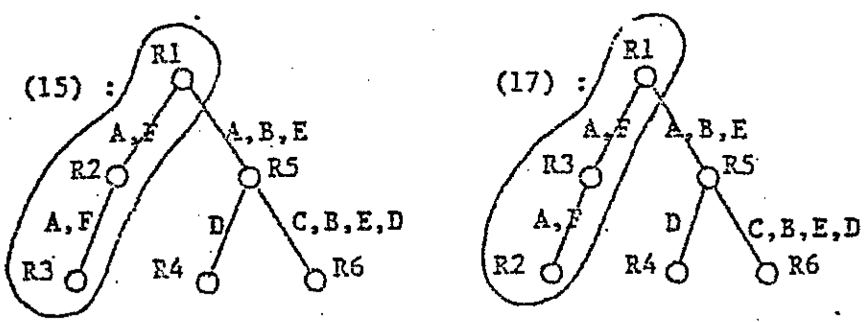


圖 5.4: 例 3.1 詢問之所有等效詢問中，最少層次且最長鏈型之詢問圖。

於是乎對上下結合程序而言，Q 之最佳等效詢問便祇要比較圖 5.4 中劃圈圈的部分；如何預估這一部分作半結合後的結果，是進一步最佳化尚待解決的問題。

第六章 結 論

本文一開始便明確地定下 DDBS.QP 之功能架構，及 DDBS.QP 核心問題之所在；也簡明地討論直接結合法與半結合法用來處理樹狀詢問與環狀詢問。

遞移封閉使一詢問轉化成最多餘（ redundant ）的唯一型式。從此最多餘的型式，吾人以投影圖映的觀念，提出一測定樹狀詢問之定理，利用該定理將使得詢問類別之測定變得很單純、快捷；更合乎理念地，本文以投影圖映可求出所有的等效詢問，從而依吾人之策略以得最佳之等效詢問；較之目前所有樹狀詢問的測定法一任意的求出〃一〃等效樹狀詢問，以之替代詢問圖非樹狀而實為樹狀詢問之詢問一顯然更合理許多。

雖然在本文中〃最少層次且最長鏈型〃之最佳等效詢問的策略仍很主觀，不過這祇是個開始，相信今後對討論最佳等效詢問，必定將以本文一些定理為基礎。

参考文獻：

[柯志昇81] "分散式資料庫系統之詢問處理"，中研院資訊所研究報告，TR-81-006.

[ADIB 80] M. Adiba et. al., "POLYPHEME: An Experience in Distributed Data Base System Design and Implementation," Distributed Data Bases, C. Delobel and W. Litwin eds., North-Holloand Pub. Co., 1980, pp.67-84.

[BECH 81] P.A. Bernstein and D.M. Chiu, "Using Semi-Joins to Solve Relational Queries," J. ACM, Vol.28, No.1, Jan. 1981, pp.25-40.

[BEGO 79] P.A. Bernstein and M. Goodman, "Full Reducers for Relational Queries Using Multi-Attribute Semijoins," Proc. Computer Network Symposium, IEEE Computer Society, 1979.

[BEGO 81] P.A. Bernstein and M. Goodman, "Power of Natural Semijoins," SIAM J. COMPUT. Vol.10, No.4, Nov. 1981, pp.751-771.

[BERN 81] P.A. Bernstein et. al., "Query Processing in a System for Distributed Databases(SDD-1)," ACM-TODS, Vol.6, No.4, Dec. 1981, pp.602-625.

[CEME 80] W. Cellary and D. Meyer, "A Multi-Query Approach to Distributed Processing in a Relational Database Management System," same Proc. of [ADIB 80] pp.99-119.

[CHHO 80] D.M. Chiu and Y.C. Ho, "A Methodology for Interpreting Tree Queries into Optimal Semi-Join Expressions," Proc. ACM-SIGMOD Conf. May 1980, pp.169-178.

[DEO 74] N. Deo, "Graph Theory with Applications to Engineering and Computer Science," Prentice-Hall.

[EVEN 79] S. Even, "Graph Algorithms," North-Holland Pub. Co.

[HEYA 79] A.R. Hevner and S.B. Yao, "Query Processing in Distributed Database Systems,"IEEE trans. Software Eng., Vol.SE-5, No.3, May 1979, pp.177-187.

[OZYU 80] M.Z. Osoyoglu and C.T. Yu, "On Identifying a Class of Database Queries That Can Be Processed Efficiently," IEEE 4th International Computer Software and Applications Conf., Oct. 1980.

[WOYO 76] E. Wong and Yossefi, "Decomposition — A Strategy for Query Processing," ACM-TODS, Vol.1, No.3, Sept. 1976, pp.223-241.

[WONG 77] E. Wong et. al., "Retrieving Dispered Data from SDD-1: A System for Distributed Databases," the second Berkerley workshop on Distributed Data Management and Computer Networks, May 1977.

[YLCC 82] C.T. Yu, K. Lam, C.C. Chang and S.K. Chang, "Promising Approach to Distributed Query Processing," to appear in Berkeley Conf. Distributed Data Base.

[YUOZ 79] C.T. Yu and M.Z. Ozsoyoglu, "An Algorithm for Tree-Query Menbership of a Distributed Query," IEEE COMPSAC, Nov. 1979.