

發明專利說明書

(本說明書格式、順序及粗體字，請勿任意更動，※記號部分請勿填寫)

※ 申請案號： 96144552

※ 申請日期： 96.11.23

※IPC 分類： G06Q 10/00 (2006.01)

一、發明名稱：(中文/英文)

G06F 17/30 (2006.01)

於網路社群中搜尋知識擁有者之方法與系統

二、申請人：(共 1 人)

姓名或名稱：(中文/英文)

財團法人資訊工業策進會

代表人：(中文/英文)

吳靜雄

住居所或營業所地址：(中文/英文)

(106)台北市和平東路2段106號11F

國 籍：(中文/英文)

中華民國

三、發明人：(共 4 人)

姓 名：(中文/英文)

1. 黃義焜

2. 謝文泰

3. 王俊程

4. 何榮昌

國 籍：(中文/英文)

1~4. 中華民國

四、聲明事項：

☐ 主張專利法第二十二條第二項☐第一款或☐第二款規定之事實，其事實發生日期為： 年 月 日。

☐ 申請前已向下列國家（地區）申請專利：

【格式請依：受理國家（地區）、申請日、申請案號 順序註記】

☐ 有主張專利法第二十七條第一項國際優先權：

☐ 無主張專利法第二十七條第一項國際優先權：

☐ 主張專利法第二十九條第一項國內優先權：

【格式請依：申請日、申請案號 順序註記】

☐ 主張專利法第三十條生物材料：

☐ 須寄存生物材料者：

國內生物材料 【格式請依：寄存機構、日期、號碼 順序註記】

國外生物材料 【格式請依：寄存國家、機構、日期、號碼 順序註記】

☐ 不須寄存生物材料者：

所屬技術領域中具有通常知識者易於獲得時，不須寄存。

九、發明說明：

【發明所屬之技術領域】

本發明是有關於一種網路社群分析方法，特別是指一種於網路社群中搜尋知識擁有者之方法。

【先前技術】

在知識爆炸的時代，每分每秒在世界各角落不斷有新的知識產生，這些新、舊知識透過網際網路被分散在世界各地的人們討論、分享，各種生活、教育、科技、休閒等各領域的知識分享網、討論區也因應而生且蓬勃發展。以雅虎知識家-繁體中文網站來說，由於使用族群廣大，每一個問題平均有 2.8 個回答，平均每一小時會形成 450 個知識，整個資料庫有 900 萬個知識。由此可知，蘊藏在網際網路、網友間的知識資源豐富，透過網際網路進行知識搜尋或網友間的提問/解答，儼然成了現代人找尋問題解答的必要手段之一。

雖然在例如雅虎知識家的網路平台上提出問題，大多可以獲得回應，但由於該平台是利用網友評價制度來衡量回應者的專業程度，人為的主觀成分較強而容易遭到刻意操弄，擁有高積分者不見得真的是「專家」。此外，現實上，回應者為了賺取積分而到處留言回答、回答不夠專業的情況仍時有所聞，因此就提問者來說，相對產生的缺點就在於：真正的專家不知在何處，提問者只能「被動」等待回應，因此即便等不到滿意的答案也無計可施，或者即便獲得初步的回覆，也無法主動聯繫回答者做進一步討教。

為了更積極善用網際網路、網友間的知識資源，勢必有必要尋找另一機制，使知識交換成功的機會提高。

【發明內容】

因此，本發明之目的，即在提供一種運用社群分析方式於網路社群中搜尋知識擁有者之方法。

本發明之另一目的，在於提供一種運用社群分析方式於網路社群中搜尋知識擁有者之系統。

於是，本發明於網路社群中搜尋知識擁有者之系統，連結一網路平台資料庫，針對一提問者 X 所提出之一問題，配合該網路平台資料庫執行於網路社群中搜尋知識擁有者之方法；該系統包含一關鍵字擷取模組、一與該關鍵字擷取模組及該網路平台資料庫連接的搜尋模組、一與該搜尋模組連接的歷史分析模組，及一推薦模組。該方法包含以下步驟：

(A) 關鍵字擷取模組針對該問題擷取出至少一關鍵字。

(B) 搜尋模組以該至少一關鍵字做為主題，在該網路平台資料庫搜尋曾經參與討論該主題的網友 $Y_1 \sim Y_n$ 。

(C) 歷史分析模組分析該等網友 $Y_1 \sim Y_n$ 在該網路平台曾有過的問答互動歷史，並依據該歷史求出每一網友之至少一種指標值。

(D) 推薦模組依據該至少一種指標值的其中一種或全部，挑選出至少一該主題之知識擁有者，並推薦給該提問者 X。

【實施方式】

有關本發明之前述及其他技術內容、特點與功效，在以下配合參考圖式之一個較佳實施例的詳細說明中，將可清楚的呈現。

在本發明被詳細描述之前，要注意的是，在以下的說明內容中，類似的元件是以相同的編號來表示。

參閱圖 1，本發明於網路社群中搜尋知識擁有者之系統 1 的較佳實施例與網路社群所在之網路平台 100 的資料庫 2 連結，且針對一提問者 X 提出之一問題 10，從該網路平台 100 的眾多使用者中搜尋出知識擁有者。前述網路平台 100 是指各網站的知識討論區、經驗交流區，或專門提供知識交換的網站，例如雅虎知識家等；所謂網路社群即是指該網路平台 100 的使用者。

於網路社群中搜尋知識擁有者之系統 1 包含一關鍵字擷取模組 3、一與該關鍵字擷取模組 3 及網路平台資料庫 2 連接之搜尋模組 4、一與該搜尋模組 4 連接的歷史分析模組 5，及一與該歷史分析模組 5 連接的推薦模組 6。

請同時參閱圖 1、2，當一使用者在該網路平台 100 提出問題 10—舉例來說「如何用 Java 寫一支 socket 程式？」，該系統 1 便開始執行搜尋知識擁有者之方法，步驟如下：

步驟 S₁—關鍵字擷取模組 3 針對該被提出之問題，擷取出至少一關鍵字，以前述所舉例子來說，可擷取出「Java」、「socket」、「程式」三個關鍵字，並依據該等關鍵字

的概念或相關詞擴張形成一關鍵字群組。

該步驟中，形成關鍵字群組的方法，可以透過將在該網路平台資料庫 2 中經常同時出現的字彙建立連結，相當於建立字彙的相關性網絡。詳言之，由「Java」進一步找到經常同時出現的相關詞彙包括「J2EE、J2SE」，由「Socket」進一步找到「TCP、Port」，由「程式」進一步找到「API、工具」。因此，針對該被提出之問題，關鍵字擷取模組 3 形成關鍵字群組包含以下關鍵字：Java、J2EE、J2SE、Socket、TCP、Port、程式、API、工具。

步驟 S_2 —搜尋模組 4 以上述關鍵字群組做為主題，在該網路平台資料庫 2 找尋曾經參與討論該主題的使用者（以下稱「網友」） $Y_1 \sim Y_n$ 。

步驟 S_3 —歷史分析模組 5 分析該等網友 $Y_1 \sim Y_n$ 在該網路平台 100 曾有過的問答互動歷史，並依據該歷史求出每一網友之至少一種指標值。本實施例針對每一網友以計算四種指標值舉例說明，該等指標值的計算方式詳述於下文，但不以此為限，也可以只計算其中一種或兩、三種或更多種指標值。

i. 中心指標值 (centrality)

中心指標值的計算方式，首先將曾經針對該主題有過問與答之互動關係的網友連線，建立如圖 3 所示之網友 $Y_1 \sim Y_n$ 的網絡關係。接著，計算每一網友的總連線數，且任二網友之間的連線不重複計算（以圖 3 來說， Y_1 總連線數 1、 Y_5 總連線數 5...），並依該總連線數求得到一中心指標值

。在本實施例，每一網友的中心指標值等於該網友的總連線數與其可能之最高連線數（以圖 3 來說， $n=8$ ，每一網友的最高連線數等於 $n-1$ ，也就是 7）的比值。

一般來說，當其中一網友具有較高的中心指標值，代表該網友針對該主題與其他網友有過較多的討論互動，其中狀況包括：曾經提問獲得多人回答，或者針對該主題回答過較多人的問題。因此，利用每一網友的中心指標值，可推測網友在該主題討論的群組中是否扮演較重要的角色。

ii. 有效值 (efficiency)

有效值的計算方式，乃利用 Burt's 演算法求出。Burt's 演算法的理論基礎，在於計算社會資源的關鍵位置——結構洞 (structural holes)；佔據該關鍵位置者，一般被認為容易獲取有利的資訊，經常扮演溝通角色，並且具有較高的獲取非重複性資源的機會，甚至可獲取控制利益。

利用 Burt's 演算法計算一網路社群中每個人的有效值的公式如下：

$$\text{有效值} = \sum_j \left[1 - \sum_q p_{iq} m_{jq} \right], q \neq i, j$$

其中， i 代表一網絡行為人（計算標的）， j 代表其他網絡行為人， q 則代表該網絡中除了 i 與 j 之外的其他網絡行為人；上述公式是以網絡行為人 i 來逐一計算與每一 j 的接觸，並進行加總，在計算該網絡行為人 i 與一網絡行為人 j 的接觸關係時，又逐一針對每一網絡行為人 q 進行 p_{iq} 值

、 m_{jq} 值的計算；其中 p_{iq} 代表行為人 i 與 q 的連接關係投入程度，而 m_{jq} 是指與接觸 j 與接觸 q 連結所帶來的邊際優勢(Maginal Strength)。當有效值等於 1，代表此人較可能位於結構洞位置。

以圖 3 來說，曾參與該主題討論之網友 $Y_1 \sim Y_n$ 所構成的網絡中，有效值較高者，較有可能取得多方資源，因此該項指標值也列入知識擁有者之推薦名單的考量。

iii. 距離指標值

請參閱圖 1、2、4，距離指標值的計算包含以下步驟：

步驟 S_{31} —在該網路平台資料庫 2 中找到任何能建立「問題的提問者 X 」與「網友 $Y_1 \sim Y_n$ 」間之連線關係的「網友 $Z_1 \sim Z_k$ 」，並記錄每個連線的連線次數。如圖 5 所示，提問者 X 透過網友 Z_1 與網友 Y_4 搭上關係；透過網友 Z_1 、 Z_2 與網友 Y_3 搭上關係；...依此類推。

步驟 S_{32} —利用該步驟 S_{31} 所紀錄連線次數推算每一連線代表的個別距離值；在本實施例，個別距離值等於連線次數的倒數。例如提問者 X 與網友 Z_1 曾有過三次的問答關係，網友 Z_1 與網友 Y_4 有過二次問答關係，則提問者 X 與網友 Z_1 距離值為 $1/3$ ，約 0.33，網友 Z_1 與網友 Y_4 距離值為 $1/2$ ，也就是 0.5。

步驟 S_{33} —利用逐段加總該個別距離值之方式，計算每一網友 $Z_1 \sim Z_k$ 、 $Y_1 \sim Y_n$ 與該提問者 X 之間的距離指標值。承步驟 S_{32} 舉例，提問者 X 與網友 Y_4 的距離指標值為 0.33 與 0.5 的加總，等於 0.83，其餘依此類推。

本實施例求出之距離指標值，代表該提問者 X 與網友 $Z_1 \sim Z_k$ 、 $Y_1 \sim Y_n$ 的密切程度，距離值越低代表兩者之間問答互動頻繁，關係越密切，通常可推論：與提問者 X 之間的距離指標值低者，是較可能回覆該提問者 X 所提出問題的人，因此在針對該主題提供知識擁有者的推薦名單時，距離指標值應列入考量。

iv. 結構相似值 (structure equivalence)

結構相似值的計算方式，乃利用 *concor* 演算法求得。*concor* 演算法是利用多次迭代計算的方式——計算全部節點的連結型態之兩兩差異性，通常可以 Euclidean Distance 計算方式，得到一數值，其結果為一矩陣；針對此矩陣，應用統計方法 *concor* (convergence of iterated correlation analysis) 進行分類，找出與發問者結構類似的節點，並以樹形圖表達各個位置之間的結構對等性程度。數值越高代表該網友與該提問者 X 在該社群網絡中的結構位置越相似，值得作為搜尋該主題之知識擁有者的參數。前述 Euclidean Distance 的計算方式，是假設有 A 、 B 、 C 、 D 、 E 五點；若為算出 node A 、 B 的差異，可以把 A 到 C 的連結數減去 B 到 C 的連結數，加以平方。同樣針對 D 、 E 作計算。全體的值加起來，取其平方根，即為 A 與 B 的連結差異性之數值。

步驟 S_4 —推薦模組 6 依據步驟 S_3 所求出之每一網友 $Y_1 \sim Y_n$ 的各種指標值，挑選出至少一該主題之知識擁有者，並推薦給該提問者 X 。舉例來說，系統 1 可向提問者 X 提

供如下所示的推薦名單：針對「如何用 Java 寫一支 socket 程式？」之主題，曾有頻繁互動者為一網友 Y_5 、 Y_4 、 Y_8 ；可能獲得多方資源者為一網友 Y_1 、 Y_2 、 Y_3 、 Y_7 ；較可能回應您所提出之問題者為一網友 Y_4 ；在該主題的討論社群中，與您有類似地位者為一網友 Y_6 。

值得一提的是，前述中心指標值及有效值之運算，屬於社群網絡中絕對關係的運算，單純針對在該網路平台 100 中針對該主題曾進行討論者進行分析；而距離指標值、結構相似值則是將提問者 X 帶入網絡，使提問者 X 在該網路平台 100 上的互動歷史，或說該提問者 X 在網絡中的地位，也影響指標值的運算結果。藉由前述多種指標，可獲得全方位且較客觀的綜合性推薦名單，提問者 X 可依據該推薦名單，視情況自行評估是否與特定網友進行積極聯繫，如此一來，提問者可主動詢問知識擁有者，不再只是被動等待回應。

此外，推薦名單的形式當然不以上述實施例為限，推薦模組 6 可以僅採用該等指標值中較具特徵者（例如數值特別顯著而具有較重大意義者）作為挑選知識擁有者之依據，甚至可另以一綜合公式將該等指標值作綜合運算，求出單一推薦人選或列出推薦人選的排名。

歸納上述，本發明於網路社群中搜尋知識擁有者之方法與系統 1，針對在該網路平台 100 中曾針對該問題相關主題參與討論之網友，分析其互動歷史，並計算出各種各具代表意義的指標值，作為系統 1 推薦知識擁有者之依據，

充份利用了隱藏在社群中豐富的互動資料，對提問者 X 而言，提供相當大的助益，進而增加知識交換成功的機會，確實可達到本發明之目的。

惟以上所述者，僅為本發明之較佳實施例而已，當不能以此限定本發明實施之範圍，即大凡依本發明申請專利範圍及發明說明內容所作之簡單的等效變化與修飾，皆仍屬本發明專利涵蓋之範圍內。

【圖式簡單說明】

圖 1 是一方塊圖，說明本發明於網路社群中搜尋知識擁有者之系統的較佳實施例的架構；

圖 2 是一流程圖，說明本發明於網路社群中搜尋知識擁有者之方法的較佳實施例；

圖 3 是一網絡示意圖，說明曾經針對該主題有過問與答之互動關係的網友 $Y_1 \sim Y_n$ 的網絡關係；

圖 4 是一流程圖，說明計算距離指標值的步驟；及

圖 5 是一網絡示意圖，說明提問者 X 透過網友 $Z_1 \sim Z_k$ 與網友 $Y_1 \sim Y_n$ 之間形成的網絡關係。

【主要元件符號說明】

1 …………… 於網路社群中搜尋知識擁有者之系統	4 …………… 搜尋模組
10 …………… 被提出之問題	5 …………… 歷史分析模組
100 …………… 網路平台	6 …………… 推薦模組
2 …………… 網路平台資料庫	$S_1 \sim S_4$ … 步驟
3 …………… 關鍵字擷取模組	$S_{31} \sim S_{33}$ … 步驟

五、中文發明摘要：

一種於網路社群中搜尋知識擁有者之方法，配合一網路平台資料庫執行，並包含步驟：(A) 針對該問題擷取出至少一關鍵字。(B) 以該至少一關鍵字做為主題，在該網路平台資料庫搜尋曾經參與討論該主題的網友 $Y_1 \sim Y_n$ 。(C) 分析該等網友 $Y_1 \sim Y_n$ 在該網路平台曾有過的問答互動歷史，並依據該歷史求出每一網友之至少一種指標值。(D) 依據該至少一種指標值的其中一種或全部，挑選出至少一該主題之知識擁有者，並推薦給該提問者 X。本發明善用隱藏在社群中豐富的互動資料，進行角色分析，提供有用的推薦名單，可大幅提高知識交換成功機會。

六、英文發明摘要：

十、申請專利範圍：

1. 一種於網路社群中搜尋知識擁有者之方法，針對一提問者 X 所提出之一問題，配合一網路平台資料庫執行；該方法包含以下步驟：

(A) 針對該問題擷取出至少一關鍵字；

(B) 以該至少一關鍵字做為主題，在該網路平台資料庫搜尋曾經參與討論該主題的網友 $Y_1 \sim Y_n$ ；

(C) 分析該等網友 $Y_1 \sim Y_n$ 在該網路平台曾有過的問答互動歷史，並依據該歷史求出每一網友之至少一種指標值；及

(D) 依據該至少一種指標值的其中一種或全部，挑選出至少一該主題之知識擁有者，並推薦給該提問者 X。

2. 依據申請專利範圍第 1 項所述之於網路社群中搜尋知識擁有者之方法，其中，該步驟 (A) 更將所擷取出之關鍵字，依其概念或相關詞擴張形成一關鍵字群組，該步驟 (B) 即以該關鍵字群組作為主題搜尋網友。
3. 依據申請專利範圍第 1 項所述之於網路社群中搜尋知識擁有者之方法，其中，該步驟 (C) 將曾有過問與答之互動關係的網友連線，建立該網友 $Y_1 \sim Y_n$ 的網絡關係，利用該網絡關係進行指標值運算。
4. 依據申請專利範圍第 3 項所述之於網路社群中搜尋知識擁有者之方法，其中，該步驟 (C) 針對每一網友 $Y_1 \sim Y_n$ 個別計算他與其他網友的總連線數，其中任二網友之間

的連線不重複計算，並依該總連線數推算一中心指標（centrality）值；該中心指標值為本步驟所求得之其中一種指標值。

5. 依據申請專利範圍第 4 項所述之於網路社群中搜尋知識擁有者之方法，其中，該中心指標值等於該網友的總連線數與其可能之最高連線數的比值。
6. 依據申請專利範圍第 1 項所述之於網路社群中搜尋知識擁有者之方法，其中，該步驟（C）利用 Burt's 演算法計算每一網友的有效值（efficiency），該有效值為本步驟所求出之其中一種指標值。
7. 依據申請專利範圍第 1 項所述之於網路社群中搜尋知識擁有者之方法，其中，該步驟（C）還在該網路平台資料庫中找到任何能使該提問者 X 與網友 $Y_1 \sim Y_n$ 搭上問答互動關係的網友 $Z_1 \sim Z_k$ ，並針對網友 $Y_1 \sim Y_n$ 、 $Z_1 \sim Z_k$ 與提問者 X 之間曾有過的問與答之互動關係者進行連線，藉此建立該網友 $Y_1 \sim Y_n$ 、 $Z_1 \sim Z_k$ 與提問者 X 之網絡關係。
8. 依據申請專利範圍第 7 項所述之於網路社群中搜尋知識擁有者之方法，其中，該步驟（C）記錄每個連線的連線次數，藉該連線次數推算個別距離值，並逐段加總該個別距離值而推算出每一網友 $Z_1 \sim Z_k$ 、 $Y_1 \sim Y_n$ 與該提問者 X 之間的距離指標值；該距離指標值為本步驟所求出之其中一種指標值。
9. 依據申請專利範圍第 8 項所述之於網路社群中搜尋知識擁有者之方法，其中，該個別距離值等於連線次數的倒

數。

10. 依據申請專利範圍第 7 項所述之於網路社群中搜尋知識擁有者之方法，其中，該步驟 (C) 利用 concor 演算法計算每一網友的結構相似值 (structure equivalence)；該結構相似值為本步驟所求得其中一種指標值。

11. 一種於網路社群中搜尋知識擁有者之系統，連結一網路平台資料庫；該系統包含：

一關鍵字擷取模組，針對一由一提問者 X 提出之問題擷取出至少一關鍵字；

一搜尋模組，與該關鍵字擷取模組及該網路平台資料庫連接，以該至少一關鍵字做為主題在該網路平台搜尋曾經參與討論該主題的網友 $Y_1 \sim Y_n$ ；

一歷史分析模組，與該搜尋模組連接，分析該等網友 $Y_1 \sim Y_n$ 在該網路平台曾有過的問答互動歷史，並依據該歷史求出每一網友之至少一種指標值；及

一推薦模組，依據該至少一種指標值的其中一種或全部，挑選出至少一該主題之知識擁有者，並推薦給該提問者 X。

12. 依據申請專利範圍第 11 項所述之於網路社群中搜尋知識擁有者之系統，其中，該關鍵字擷取模組更將所擷取出之關鍵字，依其概念或相關詞擴張形成一關鍵字群組，且該搜尋模組即以該關鍵字群組作為主題找尋網友。

13. 依據申請專利範圍第 11 項所述之於網路社群中搜尋知識擁有者之系統，其中，該歷史分析模組將曾有過問與答

之互動關係的網友連線，建立該網友 $Y_1 \sim Y_n$ 的網絡關係，利用該網絡關係進行指標值運算。

14. 依據申請專利範圍第 13 項所述之於網路社群中搜尋知識擁有者之系統，其中，該歷史分析模組針對每一網友 $Y_1 \sim Y_n$ 個別計算他與其他網友的總連線數，其中任二網友之間的連線不重複計算，並依該總連線數推算一中心指標（centrality）值；該中心指標值為歷史分析模組所求出之其中一種指標值。

15. 依據申請專利範圍第 14 項所述之於網路社群中搜尋知識擁有者之系統，其中，該中心指標值等於該網友的總連線數與其可能之最高連線數的比值。

16. 依據申請專利範圍第 11 項所述之於網路社群中搜尋知識擁有者之系統，其中，該歷史分析模組利用 Burt's 演算法計算每一網友的有效值（efficiency），該有效值為該歷史分析模組所求出之其中一種指標值。

17. 依據申請專利範圍第 11 項所述之於網路社群中搜尋知識擁有者之系統，其中，該歷史分析模組還在該網路平台資料庫中找到任何能使該提問者 X 與網友 $Y_1 \sim Y_n$ 搭上問答互動關係的網友 $Z_1 \sim Z_k$ ，並針對網友 $Y_1 \sim Y_n$ 、 $Z_1 \sim Z_k$ 與提問者 X 之間曾有過的問與答之互動關係者進行連線，藉此建立該網友 $Y_1 \sim Y_n$ 、 $Z_1 \sim Z_k$ 與提問者 X 之網絡關係。

18. 依據申請專利範圍第 17 項所述之於網路社群中搜尋知識擁有者之系統，其中，該歷史分析模組記錄每個連線的

連線次數，藉該連線次數推算個別距離值，並逐段加總該個別距離值而推算出每一網友 $Z_1 \sim Z_k$ 、 $Y_1 \sim Y_n$ 與該提問者 X 之間的距離指標值；該距離指標值為該歷史分析模組所求出之其中一種指標值。

19. 依據申請專利範圍第 18 項所述之於網路社群中搜尋知識擁有者之系統，其中，該個別距離值等於連線次數的倒數。

20. 依據申請專利範圍第 17 項所述之於網路社群中搜尋知識擁有者之系統，其中，該歷史分析模組利用 concor 演算法計算每一網友的結構相似值 (structure equivalence)；該結構相似值為該歷史分析模組所求得其中一種指標值。

十一、圖式：

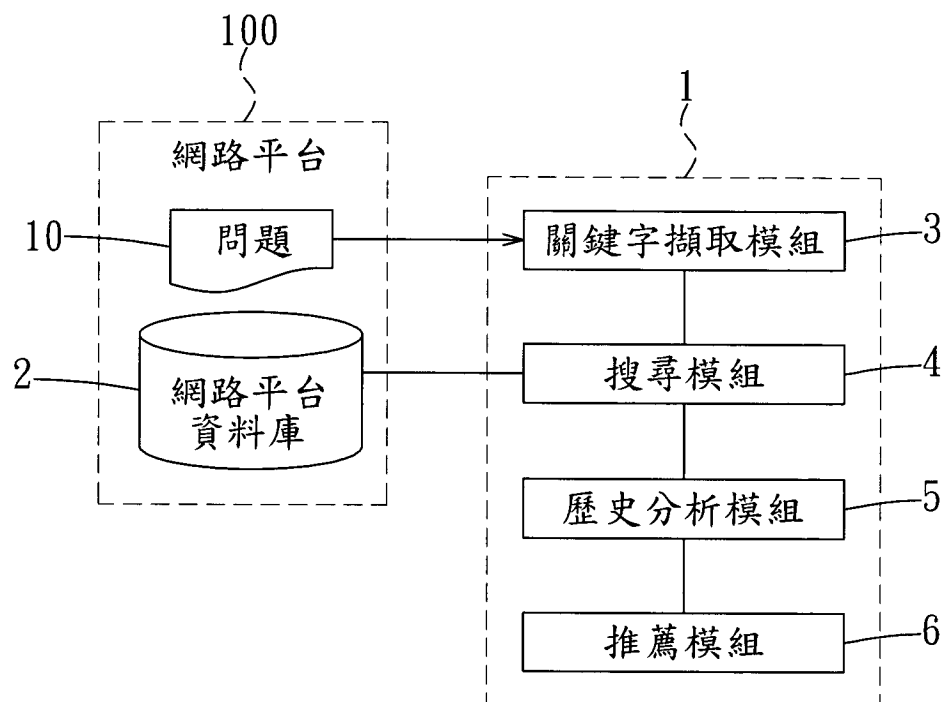


圖 1

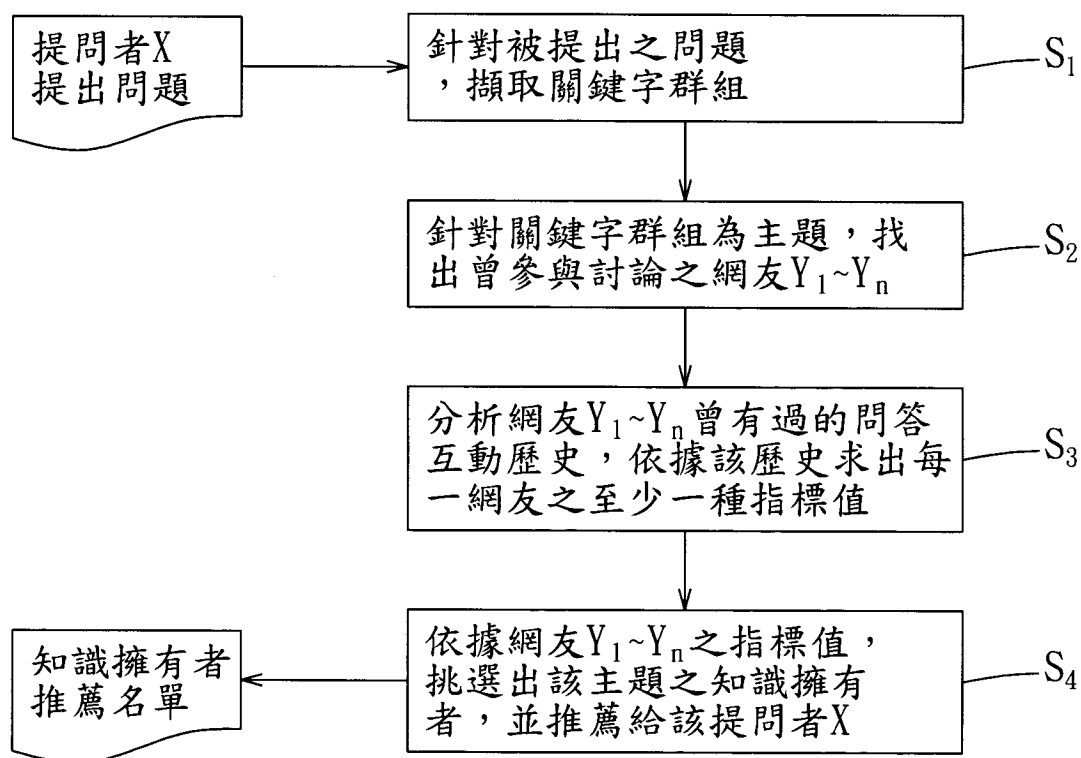


圖 2

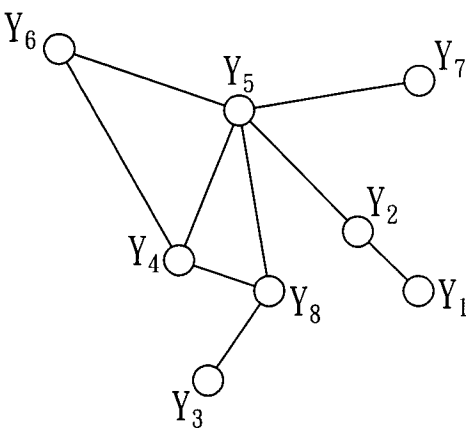


圖 3

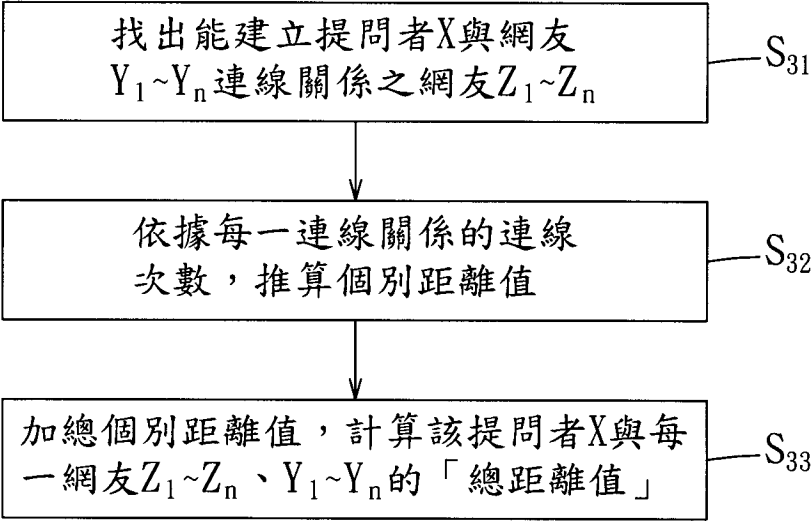


圖 4

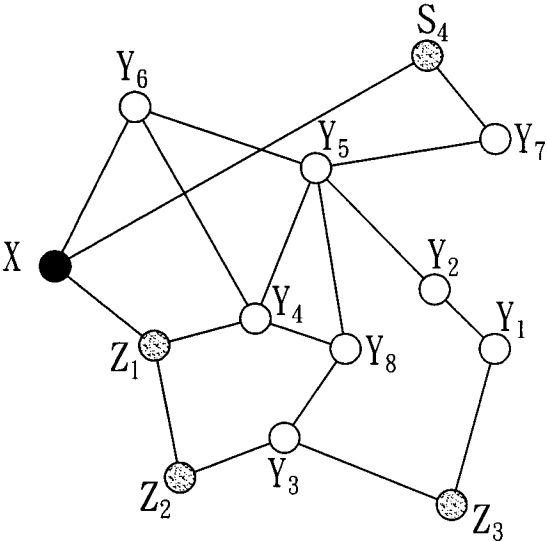


圖 5

七、指定代表圖：

(一)本案指定代表圖為：第 (2) 圖。

(二)本代表圖之元件符號簡單說明：

S₁~ S₄ …… 步驟

八、本案若有化學式時，請揭示最能顯示發明特徵的化學式：