



(19)中華民國智慧財產局

(12)發明說明書公開本

(11)公開編號：TW 201232292 A1

(43)公開日：中華民國 101 (2012) 年 08 月 01 日

(21)申請案號：100103067

(22)申請日：中華民國 100 (2011) 年 01 月 27 日

(51)Int. Cl. : **G06F17/20 (2006.01)**

(71)申請人：鴻海精密工業股份有限公司 (中華民國) HON HAI PRECISION INDUSTRY CO., LTD. (TW)

新北市土城區自由街 2 號

(72)發明人：李忠一 LEE, CHUNG I (TW)；葉建發 YEH, CHIEN FA (TW)；盧秋樺 LU, CHIU HUA (TW)；盧俊錡 LU, GEN CHI (TW)

申請實體審查：無 申請專利範圍項數：10 項 圖式數：9 共 22 頁

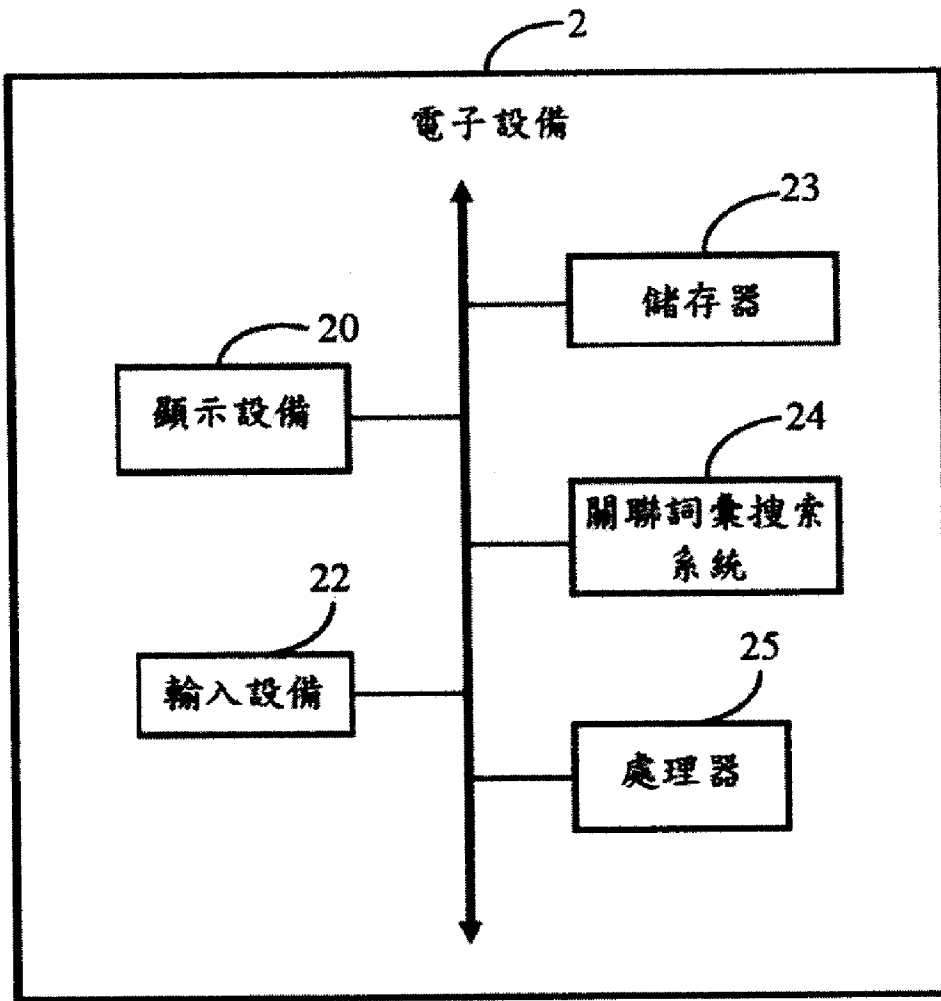
(54)名稱

關聯詞彙搜索系統及方法

SYSTEM AND METHOD FOR SEARCHING INDIRECT TERMS

(57)摘要

一種關聯詞彙搜索系統及方法，該方法包括步驟：計算詞彙與詞彙之間的直接關係強度；計算詞彙與詞彙之間的關係強度比例；根據詞彙與詞彙之間的直接關係強度及關係強度比例，計算詞彙與詞彙之間間接關係強度；根據詞彙與詞彙之間間接關係強度，確定間接關聯詞彙。利用本發明可以獲取詞彙的間接關聯詞彙。



- 2：電子設備
- 20：顯示設備
- 22：輸入設備
- 23：儲存器
- 24：關聯詞彙搜索系統
- 25：處理器

六、發明說明：

【發明所屬之技術領域】

[0001] 本發明涉及一種關聯詞彙搜索系統及方法。

【先前技術】

[0002] 在自然語言處理（Natural Language Processing，NLP）技術中，搜尋詞彙與詞彙之間的關係強度是相當重要的方面。透過詞彙與詞彙之間的關係強度，傳統的自然語言處理技術可以獲取一個詞彙的關聯詞彙，並透過這些關聯詞彙延伸擴展出該詞彙的相關詞彙。

[0003] 但是，傳統的自然語言處理技術中只是透過詞彙與詞彙之間的直接關係強度，獲取一個詞彙的直接關聯詞彙（如近義詞和同義詞等），卻不能獲取與該詞彙有間接關係的詞彙。例如，對於詞彙“棒球”而言，假設詞彙“棒球”與詞彙“運動”存在直接關係，而詞彙“運動”又與詞彙“籃球”存在直接關係，傳統的自然語言處理技術只能找到與詞彙“棒球”有直接關聯的詞彙，如“運動”等，而對於與“棒球”有間接關係的詞彙，如“籃球”等，卻無法獲取。

【發明內容】

[0004] 鑒於以上內容，有必要提供一種關聯詞彙搜索系統及方法，其可獲取詞彙的間接關聯詞彙，從而在詞彙與詞彙之間的直接關係中找出間接的、隱性的關係，用以加強輔佐詞彙與詞彙之間的直接關係。

[0005] 一種關聯詞彙搜索系統，應用於電子設備中，該系統包括：

- [0006] 第一計算模組，用於計算詞彙與詞彙之間的直接關係強度 $R_{i, j}$ ，得到一個直接關係強度矩陣 R ；
- [0007] 第二計算模組，用於計算詞彙與詞彙之間的關係強度比例 $P_{i, j}$ ，得到一個關係強度比例矩陣 P ；
- [0008] 第三計算模組，用於根據詞彙與詞彙之間的直接關係強度 $R_{i, j}$ 及關係強度比例 $P_{i, j}$ ，計算詞彙與詞彙之間間接關係強度 $R'_{i, j}$ ；及
- [0009] 關聯詞彙確定模組，用於根據詞彙與詞彙之間間接關係強度，確定間接關聯詞彙。
- [0010] 一種關聯詞彙搜索方法，運行於電子設備中，該方法包括如下步驟：
- [0011] 計算詞彙與詞彙之間的直接關係強度 $R_{i, j}$ ，得到一個直接關係強度矩陣 R ；
- [0012] 計算詞彙與詞彙之間的關係強度比例 $P_{i, j}$ ，得到一個關係強度比例矩陣 P ；
- [0013] 根據詞彙與詞彙之間的直接關係強度 $R_{i, j}$ 及關係強度比例 $P_{i, j}$ ，計算詞彙與詞彙之間間接關係強度 $R'_{i, j}$ ；及
- [0014] 根據詞彙與詞彙之間間接關係強度，確定間接關聯詞彙。
- [0015] 前述方法可以由電子設備（如電腦）執行，其中該電子設備具有附帶了圖形用戶介面（GUI）的顯示螢幕、一個或多個處理器、儲存器以及儲存在儲存器中用於執行這

些方法的一個或多個模組、程式或指令集。在某些實施例中，該電子設備提供了包括無線通信在內的多種功能。

[0016] 用於執行前述方法的指令可以包含在被配置成由一個或多個處理器執行的電腦程式產品中。

[0017] 相較於習知技術，所述的關聯詞彙搜索系統及方法，其可獲取詞彙的間接關聯詞彙，從而在詞彙與詞彙之間的直接關係中找出間接的、隱性的關係，用以加強輔佐詞彙與詞彙之間的直接關係，提高了使用者使用檢索系統（如自然語言處理搜索引擎）的效率。

【實施方式】

[0018] 參閱圖1所示，係本發明電子設備的結構示意圖。在本實施例中，所述電子設備（如伺服器）2包括透過資料匯流排相連的顯示設備20、輸入設備22、儲存器23、關聯詞彙搜索系統24和處理器25。可以理解，在其他實施例中，所述關聯詞彙搜索系統24也可以設置於其他計算裝置，如PDA（Personal Digital Assistant，個人數位助理）。

[0019] 所述關聯詞彙搜索系統24用於獲取詞彙的間接關聯詞彙，從而在詞彙與詞彙之間的直接關係中找出間接的、隱性的關係，具體過程以下描述。

[0020] 所述儲存器23用於儲存所述關聯詞彙搜索系統24的程式碼等資料。所述顯示設備20和輸入設備22用做電子設備2的輸入輸出設備。

[0021] 在本實施例中，所述關聯詞彙搜索系統24可以被分割成一個或多個模組，所述一個或多個模組被儲存在所述儲存器23中並被配置成由一個或多個處理器（本實施例為一個處理器25）執行，以完成本發明。例如，參閱圖2所示，所述關聯詞彙搜索系統24被分割成第一計算模組201、第二計算模組202、第三計算模組203和關聯詞彙確定模組204。本發明所稱的模組是完成一特定功能的程式段，比程式更適合於描述軟體在電子設備2中的執行過程。

[0022] 參閱圖3所示，係本發明關聯詞彙搜索方法的較佳實施例的流程圖。

[0023] 步驟S1，第一計算模組201計算詞彙與詞彙之間的直接關係強度 $R_{i, j}$ ，得到一個直接關係強度圖（參閱圖4所示）。其中，所述詞彙儲存於儲存器23中。可以理解，詞彙與詞彙之間的直接關係強度與詞彙的排列順序有關。例如，參閱圖4所示，詞彙“Term₁”與詞彙“Term₂”的直接關係強度為2，而詞彙“Term₂”與詞彙“Term₁”的直接關係強度為1。

[0024] 在其他實施例中，所述直接關係強度圖可以用矩陣R表示（參閱圖5所示，以下稱為“直接關係強度矩陣R”）。參閱圖5所示， $R_{i, j}$ 表示詞彙“Term_i”與詞彙“Term_j”的直接關係強度（Relation(term_i, term_j））。

[0025] 步驟S2，第二計算模組202計算詞彙與詞彙之間的關係強度比例 $P_{i, j}$ ，得到一個可轉移的關係強度比例矩陣P（或簡稱為“關係強度比例矩陣P”）。

[0026] 在第一實施例中，所述第二計算模組202透過計算詞彙與詞彙之間的條件機率，以獲取詞彙與詞彙之間的關係強度比例 $P_{i,j}$ 。參閱圖6所示，矩陣 P 中的每個元素 $P_{i,j}$ 表示詞彙與詞彙之間的條件機率 $P((Term_i \cap Term_j) | Term_i)$ ， $P_{i,i}$ 表示詞彙 $Term_i$ 單獨出現的機率。舉例而言，假設詞彙A出現次數為100次，當詞彙A出現時詞彙B出現30次，則條件機率 $P(A \cap B | A) = 0.3$ ，或稱之詞彙A可轉移至詞彙B的關係強度比例為30%。

[0027] 在其他實施例中，第二計算模組202也可以透過其他方法來獲取詞彙與詞彙之間的關係強度比例 $P_{i,j}$ ，得到一個可轉移的關係強度比例矩陣 P' （參閱圖7所示）。舉例而言，假設詞彙A與詞彙B的直接關係強度為100，而詞彙B除了與詞彙A有關係以外，還跟詞彙C有直接關係強度300，與其他詞彙無關係（ $A \rightarrow B \rightarrow C$ ）。因此，所有跟詞彙B有關的詞彙關係強度總和為400，其中詞彙A佔100，詞彙C佔300。由此可知，詞彙B與詞彙C的關係強度比例為0.75（或稱詞彙B可轉移至詞彙C的關係強度比例為0.75）。進一步地，可用此方法推知詞彙A與詞彙C的間接關係強度為 $100 \times 0.75 = 75$ 。根據該方法，可以得到圖7所示的詞彙與詞彙之間的關係強度比例。

[0028] 步驟S3，第三計算模組203根據詞彙與詞彙之間的直接關係強度 $R_{i,j}$ 及關係強度比例 $P_{i,j}$ ，計算出詞彙與詞彙之間的間接關係強度 $R'_{i,j}$ 。計算公式如下：

[0029]

$$R'_{i,j} = \sum_{k=1}^n R_{i,k} \times P_{k,j}, k \neq i, k$$

[0030] 其中， n 為詞彙的數量（即圖4中的節點數），以圖4為例來說， $n = 7$ ，詞彙 Term_1 與其他詞彙的間接關係強度參閱圖8所示，所有詞彙之間的間接關係強度參閱圖9所示。

[0031] 步驟S4，關聯詞彙確定模組204根據詞彙與詞彙之間的間接關係強度，確定間接關聯詞彙。在本實施例中，關聯詞彙確定模組204選取間接關係強度大於或等於預設值（如1.0）的詞彙作為一個詞彙的間接關聯詞彙。舉例而言，參閱圖9所示，依據間接關係強度由大到小的順序，詞彙 Term_1 的間接關聯詞彙有： Term_7 、 Term_4 、 Term_3 、 Term_5 。

[0032] 在本實施例中，只說明挖掘第二層的間接關係（即隱性關係），例如，根據 $A \rightarrow B \rightarrow C$ ，可推知 $A \rightarrow C$ 。實際上，在其他實施例中，該方法可以應用到更多層次的間接關係挖掘，例如詞彙A與詞彙B有關聯，詞彙B與詞彙C常一起出現，詞彙C與詞彙D常一起出現，則可推知詞彙A與詞彙D也有間接關聯，即根據 $A \rightarrow B \rightarrow C \rightarrow D$ ，可推知 $A \rightarrow D$ 。

[0033] 最後應說明的是，以上實施例僅用以說明本發明的技術方案而非限制，儘管參照較佳實施例對本發明進行了詳細說明，本領域的普通技術人員應當理解，可以對本發明的技術方案進行修改或等同替換，而不脫離本發明技術方案的精神和範圍。

【圖式簡單說明】

[0034] 圖1係本發明電子設備的結構示意圖。

- [0035] 圖2係關聯詞彙搜索系統的功能模組圖。
- [0036] 圖3係本發明關聯詞彙搜索方法的較佳實施例的流程圖。
- [0037] 圖4係本發明直接關係強度示意圖。
- [0038] 圖5係圖4所述直接關係強度示意圖對應的矩陣。
- [0039] 圖6和圖7係可轉移的關係強度比例矩陣示意圖。
- [0040] 圖8係本發明間接關係強度示意圖。
- [0041] 圖9係本發明間接關係強度矩陣示意圖。

【主要元件符號說明】

- [0042] 電子設備：2
- [0043] 顯示設備：20
- [0044] 輸入設備：22
- [0045] 儲存器：23
- [0046] 關聯詞彙搜索系統：24
- [0047] 處理器：25
- [0048] 第一計算模組：201
- [0049] 第二計算模組：202
- [0050] 第三計算模組：203
- [0051] 關聯詞彙確定模組：204



專利案號：100103067



日期：100年01月27日

發明專利說明書

※申請案號：100103067

※IPC分類：

G06F 17/20 (2006.01)

※申請日：100.1.27

一、發明名稱：

關聯詞彙搜索系統及方法

System and Method for Searching Indirect Terms

二、中文發明摘要：

一種關聯詞彙搜索系統及方法，該方法包括步驟：計算詞彙與詞彙之間的直接關係強度；計算詞彙與詞彙之間的關係強度比例；根據詞彙與詞彙之間的直接關係強度及關係強度比例，計算詞彙與詞彙之間間接關係強度；根據詞彙與詞彙之間間接關係強度，確定間接關聯詞彙。利用本發明可以獲取詞彙的間接關聯詞彙。

三、英文發明摘要：

The present invention provides a system and method for searching indirect terms. The system is configured for calculating a direct relation between every two terms, calculating a relation coefficient between every two terms, calculating an indirect relation between every two terms according to the direct relation and the relation coefficient between every two terms, and determining indirect terms of each term according to the indirect relation between every two terms. The present invention can obtain indirect terms of each term automatically.

七、申請專利範圍：

- 1 . 一種關聯詞彙搜索系統，應用於電子設備中，其中，該系統包括：

第一計算模組，用於計算詞彙與詞彙之間的直接關係強度 $R_{i,j}$ ，得到一個直接關係強度矩陣 R ；

第二計算模組，用於計算詞彙與詞彙之間的關係強度比例 $P_{i,j}$ ，得到一個關係強度比例矩陣 P ；

第三計算模組，用於根據詞彙與詞彙之間的直接關係強度 $R_{i,j}$ 及關係強度比例 $P_{i,j}$ ，計算詞彙與詞彙之間間接關係強度 $R'_{i,j}$ ；及

關聯詞彙確定模組，用於根據詞彙與詞彙之間間接關係強度，確定間接關聯詞彙。

- 2 . 如申請專利範圍第1項所述之關聯詞彙搜索系統，其中，所述第二計算模組透過計算詞彙與詞彙之間的條件機率，以獲取詞彙與詞彙之間的關係強度比例 $P_{i,j}$ 。

- 3 . 如申請專利範圍第1項所述之關聯詞彙搜索系統，其中，所述第三計算模組計算詞彙與詞彙之間間接關係強度 $R'_{i,j}$ 的公式為：

， n 為詞彙的數量。

$$R'_{i,j} = \sum_{k=1}^n R_{i,k} \times P_{k,j}, k \neq i, k$$

量。

- 4 . 如申請專利範圍第1項所述之關聯詞彙搜索系統，其中，所述關聯詞彙確定模組選取間接關係強度大於或等於預設值的詞彙作為一個詞彙的間接關聯詞彙。
- 5 . 如申請專利範圍第4項所述之關聯詞彙搜索系統，其中，

所述預設值為1.0。

6. 一種關聯詞彙搜索方法，運行於電子設備中，該方法包括如下步驟：

計算詞彙與詞彙之間的直接關係強度 $R_{i,j}$ ，得到一個直接關係強度矩陣R；

計算詞彙與詞彙之間的關係強度比例 $P_{i,j}$ ，得到一個關係強度比例矩陣P；

根據詞彙與詞彙之間的直接關係強度 $R_{i,j}$ 及關係強度比例 $P_{i,j}$ ，計算詞彙與詞彙之間間接關係強度 $R'_{i,j}$ ；
及

根據詞彙與詞彙之間間接關係強度，確定間接關聯詞彙。

7. 如申請專利範圍第6項所述之關聯詞彙搜索方法，其中，所述詞彙與詞彙之間的關係強度比例 $P_{i,j}$ 透過計算詞彙與詞彙之間的條件機率獲取。

8. 如申請專利範圍第6項所述之關聯詞彙搜索方法，其中，所述詞彙與詞彙之間間接關係強度 $R'_{i,j}$ 根據以下公式獲取，

$$R'_{i,j} = \sum_{k=1}^n R_{i,k} \times P_{k,j}, k \neq i, k$$

，n為詞

彙的數量。

9. 如申請專利範圍第6項所述之關聯詞彙搜索方法，其中，所述根據詞彙與詞彙之間間接關係強度，確定間接關聯詞彙的步驟包括：選取間接關係強度大於或等於預設值的詞彙作為一個詞彙的間接關聯詞彙。
10. 如申請專利範圍第9項所述之關聯詞彙搜索方法，其中，所述預設值為1.0。

八、圖式：

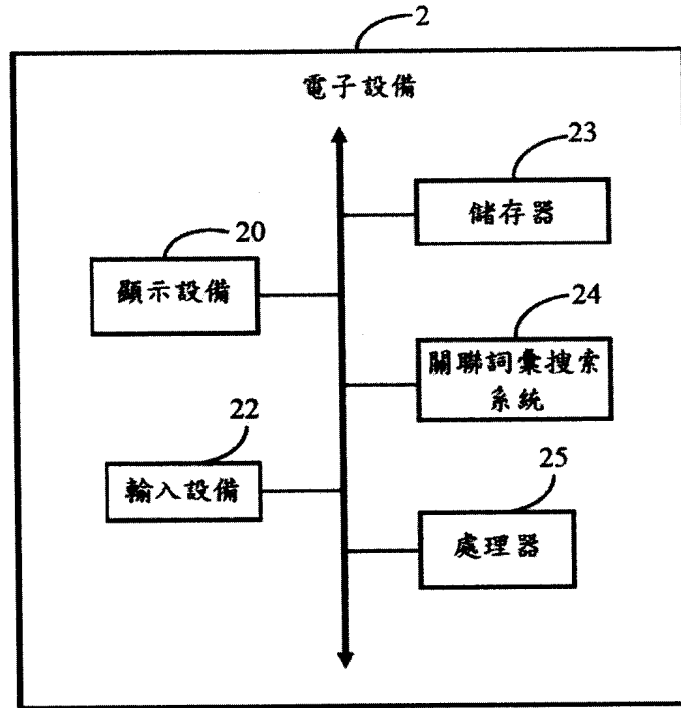


圖1

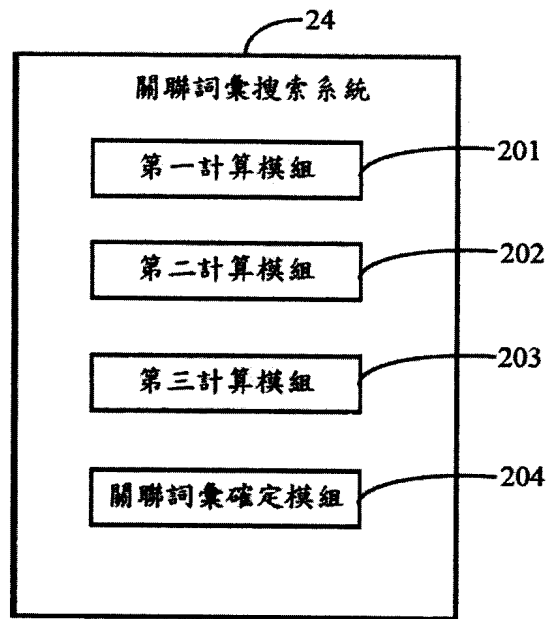


圖2

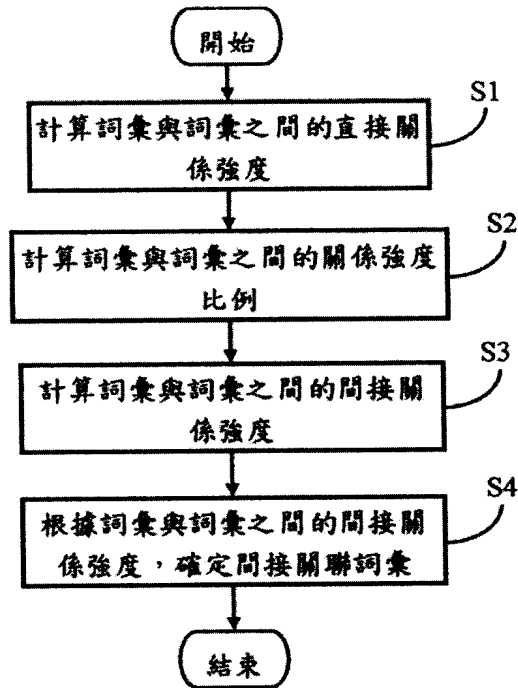


圖3

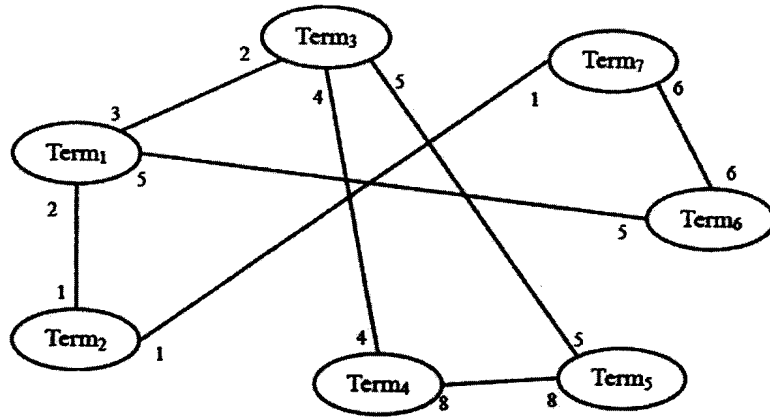


圖4

$$R = \begin{matrix} & \begin{matrix} \text{Term}_1 \\ \text{Term}_2 \\ \text{Term}_3 \\ \text{Term}_4 \\ \text{Term}_5 \\ \text{Term}_6 \\ \text{Term}_7 \end{matrix} \\ \begin{matrix} \text{Term}_1 \\ \text{Term}_2 \\ \text{Term}_3 \\ \text{Term}_4 \\ \text{Term}_5 \\ \text{Term}_6 \\ \text{Term}_7 \end{matrix} & \begin{pmatrix} 0 & 2 & 3 & 0 & 0 & 5 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 2 & 0 & 0 & 4 & 5 & 0 & 0 \\ 0 & 0 & 4 & 0 & 8 & 0 & 0 \\ 0 & 0 & 5 & 8 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 & 0 & 6 \\ 0 & 1 & 0 & 0 & 0 & 6 & 0 \end{pmatrix} \end{matrix}$$

圖5

$$P = \begin{matrix} & \begin{matrix} \text{Term}_1 \\ \text{Term}_2 \\ \text{Term}_3 \\ \text{Term}_4 \\ \text{Term}_5 \\ \text{Term}_6 \\ \text{Term}_7 \end{matrix} \\ \begin{matrix} \text{Term}_1 \\ \text{Term}_2 \\ \text{Term}_3 \\ \text{Term}_4 \\ \text{Term}_5 \\ \text{Term}_6 \\ \text{Term}_7 \end{matrix} & \begin{pmatrix} 0.2 & 0.3 & 0.3 & 0.0 & 0.1 & 0.0 & 0.0 \\ 0.2 & 0.1 & 0.1 & 0.5 & 0.0 & 0.0 & 0.1 \\ 0.1 & 0.3 & 0.0 & 0.4 & 0.1 & 0.0 & 0.1 \\ 0.0 & 0.2 & 0.6 & 0.0 & 0.0 & 0.1 & 0.1 \\ 0.1 & 0.0 & 0.0 & 0.0 & 0.0 & 0.9 & 0.0 \\ 0.0 & 0.0 & 0.3 & 0.0 & 0.2 & 0.0 & 0.5 \\ 0.0 & 0.4 & 0.0 & 0.1 & 0.0 & 0.4 & 0.1 \end{pmatrix} \end{matrix}$$

圖6

$$P' = \begin{matrix} & \begin{matrix} \text{Term}_1 \\ \text{Term}_2 \\ \text{Term}_3 \\ \text{Term}_4 \\ \text{Term}_5 \\ \text{Term}_6 \\ \text{Term}_7 \end{matrix} \\ \begin{matrix} \text{Term}_1 \\ \text{Term}_2 \\ \text{Term}_3 \\ \text{Term}_4 \\ \text{Term}_5 \\ \text{Term}_6 \\ \text{Term}_7 \end{matrix} & \begin{pmatrix} 0.0 & 0.2 & 0.3 & 0.0 & 0.0 & 0.5 & 0.0 \\ 0.5 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 \\ 0.18 & 0.0 & 0.00 & 0.36 & 0.45 & 0.00 & 0.00 \\ 0.0 & 0.0 & 0.33 & 0.0 & 0.67 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.38 & 0.62 & 0.0 & 0.0 & 0.0 \\ 0.45 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.55 \\ 0.0 & 0.14 & 0.0 & 0.0 & 0.0 & 0.86 & 0.0 \end{pmatrix} \end{matrix}$$

圖7

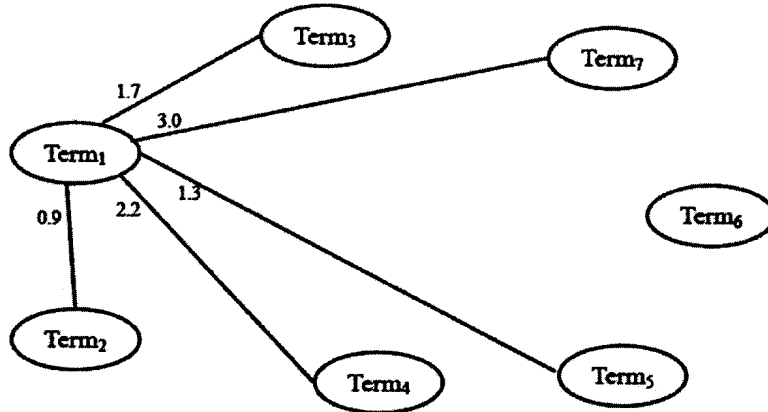


圖8

$$R' = \begin{matrix} & \begin{matrix} \text{Term}_1 \\ \text{Term}_2 \\ \text{Term}_3 \\ \text{Term}_4 \\ \text{Term}_5 \\ \text{Term}_6 \\ \text{Term}_7 \end{matrix} \\ \begin{matrix} \text{Term}_1 \\ \text{Term}_2 \\ \text{Term}_3 \\ \text{Term}_4 \\ \text{Term}_5 \\ \text{Term}_6 \\ \text{Term}_7 \end{matrix} & \begin{pmatrix} 0.0 & 0.9 & 1.7 & 2.2 & 1.3 & 0.0 & 3.0 \\ 0.0 & 0.0 & 0.3 & 0.1 & 0.1 & 0.4 & 0.0 \\ 0.5 & 1.4 & 0.0 & 0.0 & 0.2 & 4.9 & 0.4 \\ 1.2 & 1.2 & 0.0 & 0.0 & 0.0 & 0.1 & 0.1 \\ 0.1 & 0.0 & 0.0 & 0.0 & 0.4 & 7.2 & 0.4 \\ 0.0 & 3.9 & 1.5 & 0.6 & 0.5 & 0.0 & 0.0 \\ 0.2 & 0.0 & 1.9 & 0.5 & 1.2 & 0.0 & 0.0 \end{pmatrix} \end{matrix}$$

圖9

四、指定代表圖：

(一)本案指定代表圖為：圖(1)

(二)本代表圖之元件符號簡單說明：

電子設備：2

顯示設備：20

輸入設備：22

儲存器：23

關聯詞彙搜索系統：24

處理器：25

五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

Intellectual
Property
Office