

發明專利說明書 200529009

(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：93130562

※申請日期：2004年10月8日

※IPC分類：

G06F17/00

一、發明名稱：(中文/英文)

使用超級單元用於搜尋處理之系統及方法

SYSTEMS AND METHODS FOR SEARCH PROCESSING USING
SUPERUNITS

二、申請人：(共1人)

姓名或名稱：(中文/英文)

美商·雅虎股份有限公司

Yahoo! Inc.

代表人：(中文/英文)

卡拉翰麥克

CALLAHAN, MICHAEL

住居所或營業所地址：(中文/英文)

美國加州日光山谷第一大道 701 號

701 First Avenue, Sunnyvale, CA 94089, U.S.A.

國籍：(中文/英文)

美國/USA

三、發明人：(共3人)

姓名：(中文/英文)

1.卡普宣楊/KAPUR, SHYARN

2.帕瑞克金納區/PARIKH, JIGNASHU

3.喬西迪帕/JOSHI, DEEPA

國 籍：(中文/英文)

1. 印度/INDIA
2. 印度/INDIA
3. 印度/INDIA

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：

1. 美國；2003年10月9日；60/510,220
2. 美國；2004年3月9日；10/797,614

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。

玖、發明說明：

【發明所屬之技術領域】

本發明係有關於網路與網際網路搜尋與介面系統，且尤其關於提供增強搜尋功能之搜尋系統。

【先前技術】

隨著網際網路及透過全球資訊網(網路)可用於一使用者之大量網頁與媒體內容的降臨，提供使用者流線型方法以自網路過濾與獲得想要的資訊之需求已經產生。搜尋系統與過程已經研發以符合使用者需求，以獲得想要的資訊。此技術的範例可透過雅虎、Google與其他網站存取。一般來說，一使用者輸入一查詢，且一搜尋過程傳回一或多個關於該查詢之連結(於搜尋網路的情況中)、文件及/或參照(在不同搜尋語言資料庫的情況中)。傳回的連結對該使用者真正尋找的可能極為相關，或可能完全不相關。該查詢的結果的相關性可能部份為輸入的實際查詢之一功能，以及所使用的搜尋系統(基本收集系統)的穩固性。相關性可能主要由一使用者決定或主要由一使用者可能尋找者決定。

使用者輸入之查詢一般由一或多個字所做成。舉例來說，「hawaii」係一查詢，「new york city」也是，且「new york city law enforcement」也是。就這一點而論，查詢整體上對人類大腦並沒有整體性。換句話說，人類不會自然的根據查詢思考。它們是人工結構，部分加上查詢搜尋引擎或查閱圖書館目錄的需求。人類也不會自然的只根據單

字思考。人類根據的思考是自然概念。舉例來說，「hawaii」與「new york city」是根據字數測量長度的截然不同查詢，但是對人類來說，它們共用一個重要的特徵：它們各由一個概念所做成。相反的，一個人會將查詢「new york city law enforcement」視為基本上完全不同者，由於它是由兩個不同的概念所做成：「new york city」與「law enforcement」。

人類也根據概念之間的邏輯關係思考。舉例來說，「law enforcement」與「police」是相關的概念，由於警察是實施法律的重要代理者；輸入這些概念其中之一之一使用者可能對關於另一概念的網站有興趣，即使那些網站不含有該使用者剛好輸入的特定字或詞。由於此思考模式，藉由輸入一或多個自然概念的自然建立查詢，人類不只是一串可變長度的單字，且查詢一般不包含使用者可能會察覺的所有相關概念。同樣的，該使用者的意圖並不一定反應於查詢的個別單字中。例如，「law enforcement」是一個概念，而個別的單字「law」與「enforcement」並不個別傳達當單字結合時的相同使用者意圖。

任何主要搜尋提供者的目前技術，如 MSN、Google 或任何其他主要搜尋引擎網站，不適用人類建立它們的相同方法了解查詢。例如，目前的搜尋引擎一般搜尋使用者輸入的確實字或詞，並非搜尋使用者實際在腦海中的基本自然概念或相關概念。也許最重要的原因使避免搜尋提供者識別一使用者的意圖與提供最佳的搜尋結果與內容。

如可見到的，需要改良的搜尋與介面技術，幫助提供更符合一使用者可能有興趣的實際概念及一較佳使用者經驗的結果。

本發明主張 2003 年 10 月 9 日申請之美國臨時申請案第 60/510,220 號，標題為「使用單元叢集用於搜尋處理之系統與方法」之權益，其揭示在此為所有目的併入參照。

本發明係有關於 2003 年 11 月 12 日申請之相同受讓人之同時申請中的申請案第 10/713,576 號，標題為「用以自搜尋查詢產生概念單元之系統與方法」，及關於 2003 年 4 月 4 日申請之相同受讓人之同時申請中的臨時申請案第 60/460,222 號，標題為「萬用搜尋介面系統與方法」。為所有目的，於此分別揭示上述申請案以併入參照。

【發明內容】

本發明之具體實施例提供用以處理搜尋查詢之系統與方法，包含分析已接收查詢，以提供所見資訊之一較複雜了解。藉由將查詢分析為單元，並定義該單元間各種關係，如根據一起出現於查詢中之單元樣式，自一組查詢產生一概念網路。從該概念網路，可偵測不同單元間的各種相似性，且具有某些相同識別特性的單元可能群組化為超級單元。對於各超級單元來說，有一對應簽章定義該群組之識別特性。一查詢可藉由識別成分單元、決定某些或全部成本單元的超級單元成員、及使用該資訊訂定對該查詢之一回應加以處理。

根據本發明之一方面，提供用以自使用者搜尋查詢產

生超級單元之一電腦實施方法。許多先前查詢呈現為一概念網路，該概念網路包含單元與定義於該單元對間之關係，其中各關係具有一相關邊界加權。辨識一超級單元種子；該超級單元種子具有至少一成員單元，其中各成員單元係概念網路之複數個單元其中之一。為該超級單元種子定義一簽章。該簽章包含一或多個簽章單元，且各簽章單元於該概念網路中具有與一最小數量成員單元之一關係。該超級單元種子接著藉由加入來自該概念網路之一或多個新成員單元加以擴充，其中各新成員單元根據該簽章滿足一匹配規則。該簽章根據該已擴充超級單元種子修改。重複該擴充與修改步驟，直到滿足一收斂規則為止，且一旦滿足該收斂規則，形成一最終超級單元與一最終簽章。接著為該最終超級單元之各成員單元儲存超級單元成員資訊，且可用於回應後續查詢。舉例來說，該超級單元成員資訊可為該最終超級單元之各成員單元包含一成員加權，在此一成員加權基於該概念網路中該成員單元與該最終簽章之簽章單元間之關係。

根據本發明之另一方面，用以自使用者搜尋查詢產生超級單元之一系統包含一概念網路建立器模組、一超級單元種子模組、一超級單元建立器模組、及一儲存模組。該概念網路建立器模組用以自一組先前查詢產生一概念網路；該概念網路包含單元與定義於單元對間之關係，其中各關係具有一相關邊界加權。該超級單元種子模組用以識別包含至少一成員單元之一超級單元種子，其中各成員單

元係該概念網路之單元其中之一。該超級單元建立器模組用以配置建構超級單元與以該超級單元種子開始之簽章。各超級單元包含複數個成員單元，且各簽章關於該超級單元其中之一。各簽章包含一或多個簽章單元，在此各簽章單元於該概念網路中具有與該相關超級單元之一最小數量成員單元之一關係。該儲存模組用以配置為該成員單元儲存超級單元成員資訊；該超級單元成員資訊由該超級單元建立器模組提供。在某些具體實施例中，該系統亦包含一查詢回應模組，耦合至該儲存模組且用以配置接收一目前查詢。該查詢回應模組將該目前查詢分析為一或多個成分單元，為一或多個成分單元自該儲存模組取出該超級單元成員資訊，及至少部分根據該已取出超級單元成員資訊訂定對目前查詢之一回應。

下列詳細說明與附圖將提供本發明之特性與優點之較佳理解。

【實施方式】

I. 簡介

A. 網路實施方式

第 1 圖舉例說明根據本發明一具體實施例包含一客戶端系統 20 之一資訊取出與通訊網路 10 之一大致簡介。在電腦網路 10 中，客戶端系統 20 透過網際網路 40 或其他通訊網路，如透過任何區域網路 (LAN) 或廣域網路 (WAN) 連接，耦合至任何數量的伺服器系統 50_1 至 50_N 。如將在此所描述者，根據本發明，客戶端系統 20 係用以配置與任何

伺服器系統 50_1 至 50_N 通訊，如以存取、取出及顯示媒體內容與其他資訊，例如網頁。

第 1 圖中所示的系統中的幾個元件包含在此不需要詳細解釋的傳統已知元件。舉例來說，客戶端系統 20 可包含一桌上型個人電腦、工作站、膝上型、個人數位助理 (PDA)、行動電話、或任何 WAP 啟動裝置或可直接或間接處理至網際網路的介面之任何其他計算裝置。客戶端系統 20 一般執行一瀏覽程式，例如 Microsoft 的 Internet Explorer™ 瀏覽器、Netscape Navigator™ 瀏覽器、Mozilla™ 瀏覽器、Opera™ 瀏覽器、或一行動電話、PDA 或其他無線裝置等等之情況中之一 WAP 啟動瀏覽器，允許客戶端系統 20 之一使用者透過網際網路 40 自伺服器系統 50_1 至 50_N 存取、處理與檢視有效之資訊與頁面。客戶端系統 20 一般亦包含一或多個使用者介面裝置 22，例如一鍵盤、一滑鼠、觸控式螢幕、筆等等，用以與該瀏覽器於一顯示器(如螢幕畫面、LCD 顯示器等等)上提供之一圖形使用者介面 (GUI) 互動，結合伺服器系統 50_1 至 50_N 或其他伺服器提供之頁面、表單或其他資訊。本發明適用於使用網際網路，指的是網路之一特定全球內部網路。然而，應了解可使用其他網路取代，或附加於網際網路上，例如一內部網路、一外部網路、一虛擬私人網路 (VPN)、一非 TCP/IP 型網路、任何 LAN 或 WAN 等等。

根據一具體實施例，客戶端系統 20 與其所有組件為操作人員，可使用包含電腦程式碼使用一中央處理單元執行

之一應用程式，例如一 Intel Pentium™ 處理器、AMD Athlon™ 處理器等等或多處理器。用以操作與規劃客戶端系統 20 如在此說明通訊、處理與顯示資料與媒體內容之電腦程式碼最好下載與儲存於一硬碟上，但是整個程式碼或其部分亦可能儲存於已知之任何其他揮發性或非揮發性記憶體媒體或裝置中，例如一 ROM 或 RAM，或提供於可儲存程式碼之任何媒體上，例如一光碟 (CD) 媒體、一數位萬用碟 (DVD) 媒體、一軟碟等等。此外，該整個程式碼或其部分可自一軟體來源傳送與下載，如透過網際網路自伺服器系統 50₁ 至 50_N 其中之一至客戶端系統 20，或使用任何通訊媒體與通訊協定 (如 TCP/IP、HTTP、HTTPS、乙太網路、或其他傳統媒體與通訊協定) 透過任何其他網路連接 (如外部網路、VPN、LAN、或其他傳統網路) 傳送。

應了解用以實施本發明各方面之電腦程式碼可為 C、C++、HTML、XML、Java、JavaScript 等等程式碼，或任何其他適用程式語言 (如 VBScript)、或可執行於客戶端系統 20 上或編譯以執行於客戶端系統 20 上之任何其他適用程式設計語言。在某些具體實施例中，沒有程式碼下載至客戶端系統 20，且必要的程式碼由一伺服器執行，或執行已經存在於客戶端系統 20 之程式碼。

B. 搜尋系統

第 2 圖舉例說明根據本發明一具體實施例用以通訊媒體內容之另一資訊取出與通訊網路 110。如圖所示，網路 110 包含客戶端系統 120、一或多個內容伺服器系統 150、

及一搜尋伺服器系統 160。在網路 110 中，客戶端系統 120 以可通訊之方式透過網際網路 140 或其他通訊網路耦合至伺服器系統 150 與 160。如上所討論者，客戶端系統 120 與其組件用以配置透過網際網路 140 或其他通訊網路與伺服器系統 150 與 160 及其他伺服器系統通訊。

1. 客戶端系統

根據一具體實施例，執行於客戶端系統 120 上之一客戶端應用程式(呈現為模組 125)包含用以控制客戶端系統 120 與其組件以與伺服器系統 150 與 160 通訊及處理與顯示自此接收之資料內容之指令。客戶端應用程式 125 最好自一軟體來源傳送與下載至客戶端系統 120，例如一遠端伺服器系統(如伺服器系統 150、伺服器系統 160 或其他遠端伺服器系統)，雖然客戶端應用程式模組 125 可提供於任何軟體儲存媒體上，例如一軟碟、CD、DVD 等等，如上方所討論般。舉例來說，在一方面，客戶端應用程式模組 125 可於包含各種控制項之一 HTML 包覆器中透過網際網路 140 提供至客戶端系統 120，舉例來說，嵌入 JavaScript 或 ActiveX 控制項，用以操控資料與執行各種物件、框架與視窗中之資料。

此外，客戶端應用程式模組 125 包含用以處理資料與媒體內容之各種軟體模組，例如用以處理搜尋查詢與搜尋結果資料之一特定搜尋模組 126，用以執行文字與資料框架與作用中視窗中之資料與媒體內容之一使用者介面模組 127，如瀏覽器視窗與對話方塊，及用以處理介面及與執行

於客戶端 120 上之各種應用程式接合及通訊之一應用程式介面模組 128。最好用以配置根據本發明各方面處理介面之應用程式介面模組 128 為執行於客戶端系統 120 上之各種應用程式之範例包含各種電子郵件應用程式、立即訊息(IM)應用程式、瀏覽器應用程式、文件管理應用程式及其他。此外，介面模組 127 可包含一瀏覽器，例如用於配置客戶端系統 120 上之一預設瀏覽器或一不同瀏覽器。在某些具體實施例中，客戶端應用程式模組 125 提供一萬用搜尋介面之特性，如描述於上方參照之暫時申請案第 60/460,222 號所述。

2. 搜尋伺服器系統

根據一具體實施例，搜尋伺服器系統 160 用以配置提供搜尋結果資料與媒體內容至客戶端系統 120，且內容伺服器系統 150 用以配置提供資料與媒體內容至客戶端系統 120，例如網頁，以回應搜尋伺服器系統 160 提供之搜尋結果頁面中選擇之連結。在某些變化中，搜尋伺服器系統 160 傳回內容，以及或取代連結及/或參照至內容。搜尋伺服器系統 160 亦最好用以配置以下方描述之查詢登錄檔形式記錄使用者查詢活動。

一具體實施例中之搜尋伺服器系統 160 參照集頁面、連結至頁面、代表索引頁面內容之資料等等之各種頁面索引 170。頁面索引可由各種收集技術產生，包含自動網路爬蟲、蜘蛛等等，以及用以分類與排名一階層性結構中網頁之手動或半自動分類演算法與介面。這些技術可實

施於產生一頁面索引 170 並使其可用於搜尋伺服器系統 160 之搜尋伺服器系統 160 上或一分離系統(未顯示)中。

頁面索引 170 中之一項目 162 包含一搜尋項目、至呈現該項目之一頁面之一連結(或其他編碼識別器)及該頁面之一內容識別器。該內容識別器可用於為可能具有不同意亦於不同內容中之搜尋項目群組化類似結果。舉例來說，該搜尋項目「Java」可指 Java 電腦語言，指印尼 Java 島、或指咖啡(常常白話稱為 Java)。一頁面之內容識別器有利的指哪些內容是可以用的。一頁面連結可結合於多個內容識別器，所以相同頁面(或其連結)可顯示於多個內容中。內容識別器最好在使用者執行相關搜尋時由系統自動結合於頁面連結；然而，該識別器亦可由一群一或多個索引編輯器手動修改且結合於連結。在此方式中，從許多搜尋收集之知識可回饋至該系統中，以定義及重新定義內容以使該已顯示搜尋結果對該要求使用者更有價值且更有用。

搜尋伺服器系統 160 用以配置提供資料，回應至從一客戶端系統接收之各種搜尋要求，尤其是來自搜尋模組 126。舉例來說，搜尋伺服器系統 160 可以用以配置處理與排名關於一給定查詢(如根據邏輯相關性之一組合，如由該查詢中搜尋項目中出現樣式所測量者；內容識別器；頁面贊助者等等)之網頁之搜尋相關演算法加以規劃。根據本發明之具體實施例，這些演算法包含用於概念分析之演算法。

例如，本發明之某些具體實施例分析搜尋查詢及/或結果，並群組化內容中之結果，以供顯示於該使用者電腦 120

上。舉例來說，為了回應該搜尋項目「Java」，搜尋伺服器系統 160 之某些具體實施例傳回分成三類(或更多，如果識別其他內容)內容或文字之搜尋結果：電腦語言 Java、島嶼 Java、及咖啡 Java。該系統可用以配置於結合於各內容提供之連結組中顯示結果，或該系統可只顯示該內容(具有足夠資訊以對該使用者區分內容)，而沒有任何連結允許該使用者選擇想要的內容以顯示相關連結。舉例來說，在雅虎網路系統中，一組內容可以具有來自該搜尋索引至頁面之一組連結、結合贊助者匹配之連結、結合目錄匹配之連結及結合 Inside Yahoo!(IY)匹配之連結之各內容加以顯示。

除了具有不明確意義的字或詞，例如「Java」，本發明之某些具體實施例用以配置為不一定不明確的搜尋項目將結果群組化成內容。一範例係為該搜尋項目「Hawaii」傳回之結果。其本身之項目「Hawaii」可能並非不明確；然而，為此一項目傳回之結果之角色可能非常廣，關於討論或僅提到 Hawaii 的每個網站。為了提供更有用的結果給該使用者，本發明之系統最好藉由併入該結果真正相關的知識將搜尋結果歸納至內容中。舉例來說，對 Hawaii 來說，該系統可以各種內容群組傳回結果，例如「Hawaii: travel」、「Hawaii: climate」、「Hawaii: geography」、「Hawaii: culture」等等。此內容識別器(「travel」、「climate」等等)可儲存於如上所述之頁面索引項目 162 中。

應了解在此描述之搜尋系統係為舉例說明，且該變化與修改係為可能的。該內容伺服器與搜尋伺服器系統可為

單一組織之部分，如一分散式伺服器系統，例如由雅虎公司提供給使用者者，或其可為不同組織之部分。各伺服器系統一般包含至少一伺服器與一相關資料庫系統，且可包含多個伺服器與相關資料庫系統，且雖然顯示為單一方塊，可為以地理上的方式分散。舉例來說，一搜尋伺服器系統之所有伺服器可互相接近放置(如於位於單一大樓或校園中放置之一伺服器群中)，或其可於互相遠離之位置分散(如一或多個伺服器位於城市 A 中且一或多個伺服器位於城市 B 中)。因此，如在此所使用者，一「伺服器系統」一般包含一或多個邏輯及/或實際連接區域或跨一或多個地理位置分散之伺服器；該詞「伺服器」與「伺服器系統」可互相交換使用。

該搜尋伺服器系統可配置具有一或多個頁面索引與演算法，用以存取該頁面索引與提供搜尋結果給使用者，以回應自客戶端系統接收之搜尋查詢。該搜尋伺服器系統可產生該頁面索引本身、自另一來源(如一分離伺服器系統)接收頁面索引、或自另一來源接收頁面索引並執行其另外處理(如新增或更新該內容識別器)。

C. 概念網路與超級單元

在一具體實施例中，搜尋伺服器系統 160 上之演算法執行搜尋項目之概念分析，以提供更相關的結果給使用者。舉例來說，對該搜尋詞「New York City」來說，最可能的是該使用者有興趣關於紐約市(城市或區域)的網站，而非紐約州中的任何其他城市。同樣的，對於「New York

City law enforcement」來說，最有可能的是該使用者有興趣關於紐約市中之法律施行(如工作部門)。然而，大部分傳統的搜尋引擎只會使用個別的詞「New」、「York」、「City」、「law」及「enforcement」搜尋，不管該詞出現在該搜尋詞中的順序為何。其他傳統搜尋引擎可能會嘗試尋找也出現在一索引中該搜尋詞中最長的子字串。舉例來說，若該索引含有「New York」、「New York City」及「New York City law」但沒有「New York City law enforcement」，該搜尋引擎會使用「New York City law」與「enforcement」搜尋，這不一定是該使用者想要的，且不太可能會產生最佳結果。

搜尋伺服器系統 160 最好用以於一查詢，例如「New York City law enforcement」，中偵測該概念「New York City」與「law enforcement」，且為這兩個概念傳回結果。在某些具體實施例中，搜尋伺服器系統 160 使用呈現於一查詢中的搜尋詞的順序，以識別其成分概念。舉例來說，使用「New York City law enforcement」做為該搜尋詞，該系統藉由散列定址法識別「New York City」與「law enforcement」為該搜尋詞中之兩個概念，並為這些概念傳回結果。相同的結果會為「law enforcement in New York City」傳回。然而，對「city law enforcement in New York」來說，根據該概念「law enforcement」與「New York」與「city」或「city law enforcement」與「New York」，可能會傳回不同結果。同樣的，「enforcement of law in New

York City」可能會被識別為包含該概念「New York City」、「law」與「enforcement」。因此，概念的順序不像做成一概念的詞的順序般重要。在某些具體實施例中，概念包含於頁面索引中(如作為項目及/或內容識別項)或可實施一個別概念索引。應注意「law enforcement」可被視為與「enforcement of law」相同，也可視為不同，根據該內容而定。在某些具體實施例中，一查詢中之概念最好由至包含一已知概念(或「單元」)清單之一單元字典 172 之參照加以偵測。

單元字典 172 最好根據某個數量(最好是一大的數量，如至少幾十萬)的先前查詢由一概念發掘過程加以產生。將在下方說明範例之概念發掘包含該查詢之分析，以產生一概念網路，且可由搜尋伺服器 160 或由另一伺服器(未顯示)加以執行。

如在此所使用者，該詞「概念網路」囊括概念間關係之任何呈現方式。舉例來說，第 3 圖係為一小數量概念之一概念網路 300 之一圖形呈現方式。各概念或單元(如「New」、「York」、「New York City」等等)係該網路之一「節點」(如節點 302)，且由呈現概念間各種關係之「邊界」(如邊界 304)連接至其他節點。一概念網路可抓取各種關係。在第 3 圖中所示之具體實施例中，該關係包含擴充(「ext」)、結合(「assoc」)、及替代(「alt」)；其他關係亦可附加或取代在此所描述者外抓取。

在此所使用之一「擴充」係當序連兩個單元所獲得知

字串亦為一單元時存在之兩個單元間之一關係。舉例來說，序連單元「new york」與「city」所獲得之字串係「new york city」，其亦為一單元。該擴充關係顯示於第 3 圖中為一「T」結合，具有橫木連接擴充結合之兩個單元(如「new york」與「city」)，且枝幹連接至該擴充單元(如「new york city」)。

在此所使用之一「結合」係一起出現於查詢中之兩個單元間存在之一關係。舉例來說，第 3 圖顯示該單元「hotels」係單元「new york」與「new york city」之一結合。相關單元對在此亦稱為「鄰居」，而一單元之「鄰域」係其鄰居組。為了建立單元間之一結合，可要求共同出現之一最小頻率。應注意結合相關之單元不需要互相相鄰出現於查詢中，且序連相關單元所獲得之字串不必要為一單元。(若是，則存在一擴充關係。因此，一擴充關係可視為一結合之一特殊種類。)

一第一單元之一「替代」係相同表示法之一不同型式(可為一較佳、校正、或其他變化形式)；舉例來說，第 3 圖顯示「motel」與「hotel」可替代。替代的其他範例包含「brittany spears」與「britney spears」(不同拼法)，或「belgian」與「belgium」(不同說法)。在一組替代單元間，可指定一者為「較佳」，如根據出現之頻率；舉例來說，「britney spears」(該受歡迎歌手之名字之正確拼法)可為錯誤拼法替代之一較佳替代，例如「brirtany spears」。在此描述之具體實施例為不管大小寫，且僅字首大小寫不同

的詞(如「Belgium」與「belgium」)指的是相同單元；其他實施例可根據大小寫區別單元，且可識別僅字首大小寫不同的單元為替代品。

在某些具體實施例中，該概念網路中之邊界可被指派加權(未顯示於第3圖中)，即代表不同關係之相關強度之數值。舉例來說，一第一單元與一結合單元間之邊界加權可根據含有亦含有相關單元之第一單元之所有查詢之分數，或根據含有亦含有另一單元之一單元之所有查詢之分數。加權最好反應相關強度；因此，加權可以任何想要的方式正規化。應了解第3圖僅為舉例說明，且亦可使用其他關係以及不同單元或概念間之連接或關係之其他呈現方式；在此使用之詞「概念網路」囊括其他呈現方式。

在本發明具體實施例中，呈現於該概念網路中之關係亦包含「超級單元」中各種單元之成員。在此所使用之「超級單元」該詞指的是具有一識別共同特徵之一組單元。該已識別共同特徵(可包含多個元件)由可用於決定是否另一單元屬於該超級單元中之超級單元之一「簽章」加以呈現。在某些具體實施例中，該簽章亦用於根據該單元特徵與該簽章特徵間之相似度為各成員單元決定一「成員加權」。可定義一閾值成員加權，且該超級成員可僅包含其成員加權超過此閾值之單元。

舉例來說，一超級單元可由城市(如「New York City」、「San Francisco」、「Chicago」等等)做成，且其簽章可包含常常出現於關於一城市名稱之查詢中某數量其他單

元(如「hotel」、「museum」、「mayor」、「jobs」等等)。可藉由比較其對該簽名之結合評估一新單元以決定是否其為一城市(如該超級單元之一成員)。作為另一範例，另一超級單元可由互相替代之單元做成(如「britney spears」、「brittany spears」、「britney speers」等等)，且其簽章可包含與該歌手名字相關之單元(如「photos」、「mp3」、「tour」等等)以及一「編輯距離」參數，指示拼字之相似性。會排除具有相似結合但有一較大編輯距離之一單元(如「barbra streisand」或「celine dion」，而包含 Britney Spears 的其他拼字錯誤。用以從查詢產生超級單元與簽章之特殊技巧描述於下。如同單元之其他關係，各種單元之超級單元簽章與超級單元成員資訊(如成員加權)可儲存於單元字典 172 中。

在某些具體實施例中，一超級單元之簽章之不同元件可被指派不同加權。該加權最好選擇以反映特徵化該超級單元之不同簽章元件之相關有效性。

搜尋伺服器 160 最好使用超級單元資訊以回應查詢，如藉由決定一查詢中之單元屬於何超級單元，並將該查詢之單元與這些超級單元之簽章比較，以決定該使用者最可能的意圖是什麼。搜尋伺服器 160 可使用關於使用者可能意圖之資訊，如以組織該搜尋結果、建議相關搜尋等等。這些搜尋伺服器 160 之特性將在下方第 III 部分中描述。

II. 概念分析系統

第 4 圖係根據本發明一具體實施例用以執行概念發掘

或概念分析之一系統 400 之一方塊圖，包含超級單元之產生。一或多個查詢登錄檔 402(或實際查詢)由一查詢處理引擎(亦稱為一查詢引擎)404 接收，其產生一單元字典 406。查詢引擎 404 可為搜尋伺服器系統 160 的組件(第 2 圖)或與搜尋伺服器系統 160 通訊的不同系統。在一具體實施例中，查詢引擎 404 包含一概念網路(CN)建立器 410、一超級單元種子模組 412、及一超級單元建立器 414。CN 建立器 410 分析查詢登錄檔 402 之內容，並產生包含單元、單元間關係(如擴充、結合、及替代)、及該關係之邊界加權之一概念網路 408。超級單元種子模組 412 選擇性地藉由概念網路 408 之分析，產生單元之初步分類至超級單元(在此稱為「種子」)中。超級單元建立器 414 使用超級單元種子模組 412 所提供之種子處理概念網路 408，以產生許多超級單元。該單元與其關係，包含超級單元，抓取於一單元字典 406 中。

單元字典 406 可以任何格式實施，並儲存於任何適當儲存媒體上，包含磁碟或磁帶，光學儲存媒體，例如光碟(CD)等等。單元字典 406 之內容最好包含該單元，以及關於各單元之額外資訊，例如關係(如擴充、結合、替代)與 CN 建立器 410 與超級單元成員(如成員加權)由超級單元建立器 414 產生決定所提供之統計資料(如邊界加權)。單元字典 406 亦可包含關於該超級單元本身之資訊，例如關於一超級單元之簽章之參數。儲存於單元字典 406 中之資訊可由一搜尋伺服器(如第 2 圖之搜尋伺服器 160)使用，以

回應後續查詢。

一查詢登錄檔 402(或一實際查詢)可透過網際網路或透過各種網路連接自各種來源接收，如 LAN、WAN、直接連結、分散媒體(如 CD、DVD、軟碟)等等。來源之範例包含搜尋伺服器系統 160(第 2 圖)、或搜尋伺服器之一分散式網路中之多個搜尋伺服器 160、及一或多個內容伺服器 150。查詢登錄檔來源一般關於相同組織或實體，如雅虎伺服器，但不一定要。該查詢登錄檔(亦稱為查詢登錄)由查詢引擎 404 使用統計方法加以處理，例如可用於資訊理論或概念中者，例如共同資訊。在某些具體實施例中，使用每日查詢登錄，雖然可使用想要的不同時間週期之登錄，如小時、星期等等。查詢登錄一般包含由使用者傳送之實際查詢(如文字字串)，且亦可為某些或所有查詢包含額外資訊(在此稱為「元資訊」)，例如查詢使用者之地理位置、時間戳、客戶端系統之 IP 位址、網路餅乾、客戶端類型(如瀏覽器類型)等等。舉例來說，查詢登錄項目可格式化為 <query_string, meta-information> 或 <count, query_string>，在此「count」代表出現頻率。(頻率可正規化，且可不要，視需要而定。)

A. 概念網路建立器

CN 建立器 410 處理該查詢登錄 402，以產生概念網路 408。在較佳具體實施例中，CN 建立器 410 使用一查詢中搜尋詞之順序，以識別做成該查詢之一或多個單元。舉例來說，一單元可為一字(如「Java」)或時常互相相鄰出現

的一群字(如「new york city」)。該單元對應至該概念網路中之節點(概念)。

CN 建立器 410 亦分析該單元，以偵測關係，例如擴充(可根據一字或單元有時候後方跟著另一字或單元，有時候又沒有加以偵測)、結合(可根據單元對出現之頻率加以偵測)、及替代(可根據「編輯距離」加以偵測，即需要轉換一單元至另一單元之印刷上變更之數量)。用以識別單元與單元間關係(包含結合、擴充、及替代)之特殊技巧詳細說明於上方參照之申請案第 10/713,576 號中。應了解除了在此描述者之外，或可替代在此描述者，CN 建立器 410 亦可實施其他技巧，以產生概念網路 408。

概念網路 408 之一呈現方式可儲存於單元字典 406 中。在某些具體實施例中，此呈現方式包含該單元以及各單元之關係與加權組。各種資料壓縮技術可用於呈現此資訊於單元字典 406 中。

B. 超級單元種子模組

超級單元種子模組 412 產生超級單元可自此建構之一或多個種子。如在此所使用者，一「種子」可為單一單元或具有一或多個共同特性之一單元清單。超級單元種子模組 412 可使用各種技巧以產生種子。現在將說明此技巧之四個範例：(1)概念網路 408 之分析；(2)至外部來源之參照；(3)使用者行為之分析；及(4)該搜尋語言資料庫中文文件之分析。亦應了解單一單元可使用為一種子，且超級單元種子模組 412 可僅從概念網路 408 選擇某數量之單元欲

作為種子(如根據出現頻率、鄰域大小、或某些其他規則)。

1. 根據概念網路之種子(叢集與支系)

在一具體實施例中，超級單元種子模組 412 使用概念網路 408 執行該查詢之進一步分析，以建立可使用為種子之叢集(即相關單元群)。在此具體實施例中，叢集藉由識別具有相似鄰域(如相關單元組)之不同單元(該叢集之「成員」)字單元產生。該叢集可為超級單元產生使用為種子；如將所見到者，該叢集本身亦可為超級單元。

舉例來說，考量一情況，其中使用者搜尋關於其喜愛音樂表演者之資訊。一般來說，這些使用者會建構一查詢，包含該表演者之名字(如「Avril Lavigne」或「Celine Dion」或「Matchbox Twenty」)，以及某些其他字反映看到的資訊類型，例如「lyrics」、「mp3」、「guitar tabs」、「discography」等等；這些其他字為相鄰單元，傾向於與不同表演者之名稱一起出現。根據相似相鄰單元之出現，超級單元種子模組 412 將該表演者名稱分類至一叢集中。

尤其是，第 5 圖係可由超級單元種子模組 412 執行，以產生來自一概念網路 408 之叢集之一過程 500 之一流程圖。在步驟 502，該概念網路 408 提供至超級單元種子模組 412。

於步驟 503，選擇用以形成一叢集之一基本單元。在某些具體實施例中，該概念網路中之每個單元可使用為一基本單元。在其他具體實施例中，基本單元可被限制，如限制為出現至少某最小頻率之單元。應了解可藉由使用不

同基本單元重複過程 500 建立任何數量之叢集。

於步驟 504，選擇概念網路 408 中之另一單元為一候選單元，用以包含於具有該基本單元之一叢集中。一循環程序可使用以選擇所有單元對，或選擇可限制至符合特定規則之單元。舉例來說，在一具體實施例中，概念網路 408 包含一特殊單元與各種相鄰單元之結合。在此具體實施例中，步驟 504 包含比較該基本單元與一第二單元之鄰域，以決定重複程度；若太小，該第二單元不會變成一候選單元。在此具體實施例中，考量的單元選擇簡化為以一基本單元 B 開始、尋找一相鄰單元 A、接著尋找亦為 A 之一鄰居之一第三單元 C。比較單元 B 與 C 之鄰域決定是否選擇單元 C 為一候選單元，以包含於具有單元 B 之一叢集中。此程序降低可能候選單元組至具有至少一鄰居與該基本單元相同之那些單元。

在步驟 506，識別與該基本單元與該候選單元一起出現之相鄰單元範例。如，若「Avril Lavigne」與「matchbox twenty」為該基本單元與候選單元，相同的相鄰單元可能包含「lyrics」、「discography」等等。與基本與候選單元一起出現之相鄰單元在此稱為「已匹配」單元。

在步驟 508，根據該相鄰單元計算用於該候選單元之一叢集加權，包含該已匹配單元。此叢集加權係之該候選單元與該基本單元間相似性之一測量；可以各種方式計算。現在將說明用以計算叢集加權之五個適當演算法範例；熟知該項技藝人士應了解亦可使用其他演算法。

範例 1：一演算法考量已匹配單元之數量為一相似性測量。單元 u_1 與 u_2 之叢集加權定義為：

$$W(u_1, u_2) = N_C / N_T, \quad (1)$$

其中 N_C 係已匹配單元之數量，而 N_T 係單元 u_1 的相鄰單元總數與 u_2 的相鄰單元總數中較大者。

可能有變化。舉例來說， N_T 可定義為兩個總數的較小者（而非較大者），或為兩個總數的平均。

範例 2：一第二演算法考量每個已匹配單元之頻率（且因此一相鄰單元為一單元有多重要）。單元 u_1 與 u_2 之叢集加權定義為：

$$W(u_1, u_2) = F_M / F_T, \quad (2)$$

其中 F_M 係透過所有已匹配單元 s_i ，單元 s_i 與單元 u_1 一起出現的頻率及單元 s_i 與單元 u_2 一起出現的頻率的總合；而 F_T 係不論已匹配與否，透過所有相鄰單元之相同頻率總合。

範例 3：相關頻率係另一重要性測量，其中在該已匹配單元與單元 u_1 與 u_2 一起出現的相關頻率不同之情況中附加一懲罰（遞減加權）。在此範例中， $R1i$ 與 $R2i$ 分別定義為相鄰單元 s_i 與單元 u_1 與 u_2 的相關頻率。該叢集加權給定為：

$$W(u_1, u_2) = \sum_{s_i} \frac{1}{1 - R1i} * (1 - P * |R1i - R2i|) \quad (3)$$

其中該總合透過已匹配單元 s_i ，而 P 為加權相關頻率差異之一懲罰因素。該 P 值可相異；於一具體實施例中，

$P=2$ 。

範例 4：以頻率的遞減順序(排名)比較相鄰單元係測量重要性的另一方法。類似於範例 3，附加一懲罰至該已匹配單元之排名之任何差異。各已匹配單元 s_i 指派二排名 $Q1i$ 與 $Q2i$ ，分別表示其與單元 u_1 與 u_2 的排名。該叢集加權給定為：

$$W(u_1, u_2) = \sum_{s_i} [M - |Q1i - Q2i|] \quad (4)$$

其中 M 係單元 u_1 之相鄰單元總數與單元 u_2 之相鄰單元總數較小者，而該總合透過已匹配單元 s_i 。

範例 5：不像之前的演算法，此演算法考量一相鄰單元之鑑別度。任何單元 u 的「結合性」可藉由比較該單元與一查詢中之一或多個其他單元(可為任何單元)一起出現的頻率(f_u)與該單元單獨出現於一查詢中之頻率(f_q)加以定義。在一測量中，結合性由 $\rho(u) = f_u / f_q$ 給定。

此結合性測量可結合上述討論之相關頻率概念，已計算該叢集加權。給予一「分數」 α 至各已匹配單元 s_i ，根據其相關頻率；尤其是， $\sigma(s_i) = 1 - (R1i - R2i)$ ，在此 $R1i$ 與 $R2i$ 定義為上述範例 3 中。該叢集加權給定為：

$$W(u_1, u_2) = \sum_{s_i} \sigma(s_i) * 1 / (1 - C * \rho(s_i)), \quad (5)$$

常數 C 的值可根據經驗分析最佳化；在一具體實施例中， $C=0.5$ 。

回到第 5 圖，於步驟 510，做成一決策是否包含該候

選單元於具有該基本單元之一叢集中。舉例來說，若其叢集加權太低，一單元可自一叢集排除。

在某些具體實施例中，叢集可停止於單元對。在其他具體實施例中，較大的叢集藉由選擇一不同候選單元與重複步驟 506、508、集 510 加以形成。在又其他具體實施例中，二或多個單元的叢集可用於該基本單元處，以產生較大叢集。一叢集係使用為一基本單元處，其鄰域可以各種方式定義，如做為該成員單元之鄰域之聯集或交集，如做為該成員單元之鄰域之聯集或交集之單元組、如做為該成員單元之至少某最小分數(如 25%、50%、80%)之一相鄰者等等。不論大小，該叢集可使用為超級單元種子。

在某些具體實施例中，叢集可進一步定義至具有較強或較近關係於該成員單元間之「支系」中。在一具體實施例中，一「支系」係一組單元，其中每個成員單元呈現於每個其他成員單元形成之叢集中。支系可使用於各種目的，如區別拼字錯誤其替代字形式，或區別環繞其一叢集形成之基本單元之不同字義。舉例來說，其基本單元為「New York」之一叢集可包含其他城市的名稱(如「Boston」、「Seattle」等等)之單元，且亦可包含相同程式之替代名稱(如「NY」、「NYC」)之單元。從這些單元，可形成包含不同城市(「New York」、「Boston」、「Seattle」)之一支系與包含「New York City」之替代名稱(「New York」、「NY」、「NYC」)之一不同支系。

作為另一範例，一叢集具有基本單元「Yahoo」之一

支系可包含其他電子郵件提供者之名稱(如「AOL」、「Hotmail」)以及其他搜尋引擎之名稱(如「Google」)。一叢集具有基本單元「Google」之一支系可包含「Yahoo」但無「AOL」或「Hotmail」。因此,「Yahoo」與「Google」可能是一支系之成員,而「Yahoo」、「AOL」與「Hotmail」可能為另一支系之成員。

第6圖係根據本發明一實施例可用於自一叢集群形成具有成員單元 q_k 之一支系 Q 之一過程600之一流程圖。在這些範例中,已形成一數量 N 之叢集,各具有一不同基本單元 b_i ($1 \leq i \leq N$)。該叢集在此表示為 $C(b_i)$,而為支系 $C(b_i)$ 之一成員之一單元 u 具有由 $W(u, b_i)$ 表示之一叢集加權,其可被計算,如使用任何上述給定之公式。(對於並非於叢集 $C(b_i)$ 中之一單元 u 來說,加權 $W(u, b_i)$ 可被指派一為零的值。)應了解一給訂單元若為任何數量叢集 $C(b_i)$ 之一成員,且在某些範例中,一叢集 $C(b_i)$ 可僅由該基本單元 b_i 組成。在過程600中,形成具有成員 q_k 之一支系 $Q(b_i)$,藉由基本單元 b_i 作為支系 $Q(b_i)$ 之第一成員開始,且為下列 $C(b_i)$ 之其他單元 u_j :(1)該支系 Q 之所有成員 q_k 為支系 $C(u_j)$ 之元件;及(2)單元 u_j 對該支系 Q 之各成員 q_k 來說係該叢集 $C(q_k)$ 之一元件。

尤其是,於步驟602,該支系 Q 建立有一成員 b_i 。於步驟604,獲得該叢集 $C(b_i)$ 之下一成員單元 u_j 。於步驟606,獲得該叢集 $C(u_j)$ 。於步驟608,決定是否支系 Q 之所有成員亦為支系 $C(u_j)$ 之成員。若否,則單元 u_j 不加入

支系 Q 中，且過程 600 跳至步驟 616。否則，於步驟 610，對支系 Q 之各成員 q_k 來說，獲得該叢集 $C(q_k)$ 。於步驟 612，決定是否單元 u_j 於步驟 610 獲得之各叢集 $C(q_k)$ 中。步驟 610 與 612 可藉由循環支系 Q 之成員 q_k 加以執行，或叢集為多個成員 q_k 可平行測試。若單元 u_j 並非於步驟 610 獲得之每個叢集 $C(q_k)$ 中，則 u_j 不加入支系 Q ，且過程 600 會跳至步驟 616。若單元 u_j 係於每個叢集 $C(q_k)$ 中，則單元 u_j 會於步驟 614 加入支系 Q 。

於步驟 616，不論是否單元 u_j 加入支系 Q ，決定是否有更多單元 u_j 仍然於叢集 $C(b_i)$ 中。若是，則過程 600 會回到步驟 604，以處理下一個成員單元 u_j 。

在所有單元 u_j 皆已經處理後，於步驟 618，為支系 Q 之各成員 q_k 決定一成員分數。在一具體實施例中，該分數根據支系 Q 之各其他成員單元藉由加入該叢集中單元 q_k 之叢集加權加以計算，即，

$$\text{Score}(q_k) = \sum_{i \neq k} W(q_k, q_i), \quad (6)$$

在此 $W(q_k, q_i)$ 表示該叢集加權為單元 q_k 作為支系 $C(q_i)$ 之一成員。亦可使用其他公式以指派一支系成員分數。在某些具體實施例中，支系成員可以降冪或升冪分數之順序排列。

應了解在此描述之過程為舉例說明，且可能有變化與修改。所述為連續的步驟可平行執行，步驟之順序可能相異，且步驟可被修改或結合。舉例來說，用以加入一單元

至一支系之條件可放鬆為要求，如(於步驟 608)該支系 Q 之成員 q_k 之至少一分數 $f1$ 為叢集 $C(u_j)$ 之元件，或(於步驟 612) u_j 係該叢集 $C(q_k)$ 之至少一分數 $f2$ 之一元件。該分數 $f1$ 與 $f2$ 可依需求選擇，且可為如 50%、70%、90% 等等；該二分數於各種具體實施例中可能或可能不相等。過程 600 可以不同基本單元 b_i 重複，以產生任何數量的支系。在支系產生處，支系可使用為超級單元種子而非叢集，或支系與叢集之一組合可使用為超級單元種子。

2. 根據外部來源之種子

在另一具體實施例中，超級單元種子模組 412 藉由參照一或多個外部來源產生種子(一般顯示為第 4 圖中之方塊 416)。外部來源之範例包含一編輯或編輯團隊所建立之相關項目之一清單(如受歡迎歌手之一清單或該團隊已知之汽車製造商之一清單)；一權威網站(如維護一疾病字典或其他清單之一醫學參考網站)等等。在此具體實施例中，超級單元種子模組 412 可執行極少或甚至沒有處理於該外部來源資料上。舉例來說，若一字清單由一編輯團隊提供，超級單元種子模組 412 可僅傳送該清單至超級單元建立器 414。超級單元種子模組 412 亦可刪減該清單，以移除並非概念網路 408 中單元之任何項目。應注意此一超級單元種子不需要為一窮盡清單，且可包含一小數量(如二、五、或十)之單元。

3. 根據使用者行為之種子

在一第三具體實施例中，超級單元種子模組 412 藉由

分析使用者行為產生種子。舉例來說，一搜尋伺服器(如第 2 圖之伺服器 160)可藉由提供一搜尋結果頁面至客戶端 120 回應至一查詢。該搜尋結果頁面包含一「點擊」清單(連結至包含關於該查詢之內容之網頁或網站)。該點擊清單可包含如頁面標題、顯示該相關內容之摘錄、及/或其他資訊。該使用者檢視該清單並選擇一點擊，如藉由按下該已顯示連結。(此動作稱為「點選」，雖然應了解不需要連結與敲擊。)查詢登錄 402 可為某些或所有查詢提供點選資料，指示一使用者自該搜尋結果頁面跟隨何連結。超級單元種子模組 412 可能於輸入不同查詢之使用者點選相同頁面處接收此資料與識別範例。此使用者行為建議該查詢間之一通用性，且種子模組 412 可將具有相似(或相同)點選行為之查詢(或其已選擇單元)分類至一種子中。種子模組 412 最好用以配置僅於偵測到該頁面之行為建議相關性之一型態時(如當點選一特殊頁面剛好具有一特定最小頻率時)分類該查詢(或單元)。

4.根據文件分析之種子

在一第四具體實施例中，種子模組 412 藉由該搜尋語言資料庫中一或多個「來源」文件之分析產生種子(如於一網路搜尋具體實施例情況中之網頁)。在此具體實施例中，種子模組 412 根據相同文件中其出現推斷單元間之通用性。舉例來說，種子模組 412 可將一文件分析為成分單元，如藉由匹配文字字串至單元字典 406 中之項目，或至概念網路 408 中之單元(節點)。在一具體實施例中，於該文件

中找到之所有單元收集至單一種子清單中。在另一具體實施例中，過濾該單元，如藉由要求出現之一最小頻率、藉由僅若其互相接近出現時包含單元對(或較大群組)等等。該單元之結果清單可使用為一超級單元種子。可使用任何數量之來源文件執行文件分析，且可使用各種規則以自動或手動選擇文件加以分析。

應了解種子模組 412 之前述具體實施例為舉例說明且並非限制。種子可藉由上述或其他單元之任何一或多種，或藉由一技巧組合加以產生。在又其他具體實施例中，各單元(或該單元之某子組之各者，如最常見)可使用為一分離種子。

C. 超級單元建立器

不論種子如何產生，種子模組 412 提供該種子至超級單元建立器 414，其使用該種子及 CN 建立器 410 所產生之概念網路 408 以建構超級單元。在某些具體實施例中，超級單元由識別一簽章(即該種子傾向共同具有之單元之一或多個關係)、搜尋匹配該簽章之概念網路中之額外單元、加入這些單元至該超級單元、及修改該簽章以反映該超級單元之目前內容之一循環過程加以建構。

尤其是，第 7 圖係根據本發明一具體實施例用以產生可實施於超級單元建立器 414 中之超級單元之一過程 700 之一流程圖。於步驟 702，超級單元建立器 414 自種子模組 412 接收一種子。該種子視為一初始超級單元。

於步驟 704，為該(初始)超級單元決定一簽章。該簽

章最好根據關於該超級單元之一或多個成員單元之一組單元加以定義，在此並沒有簽章單元為該超級單元之一成員。舉例來說，超級單元建立器 414 可放置該超級單元之成員單元於概念網路 408 中，且比較各成員單元之相鄰單元，以決定何相鄰單元對該成員單元為共同者（且本身並非成員單元）。於一具體實施例中，簽章單元最好根據二規則加以選擇：(1)該超級單元之一成員係該簽章單元之一鄰居之可能性；及(2)該簽章單元之一鄰居係超級單元之一成員之可能性。這些規則識別傾向為該超級單元之成員與非成員間有效鑑別器之簽章單元。

做為該第一規則之一範例，一簽章單元可能要求具有與至少 5% (或 10% 或 50%) 成員單元之一特殊關係；或一簽章單元至成員單元之某分數之關係可能要求具有一最小邊界加權；或一簽章單元與該成員單元間之邊界加權總合可能要求超過某閾值。在某些具體實施例中，該簽章單元關於可能為該超級單元之成員單元與各簽章單元間之關係反映一平均邊界加權 (或邊界加權之分散) 之加權界線。

做為該第二規則之範例，該簽章單元之相鄰單元之一最小分數可能要求為該超級單元之成員；或該簽章單元與該成員單元間之關係之邊界加權對上該簽章單元與非成員單元間之關係之邊界加權可能要求滿足一特定關係。下方說明簽章定義之另外範例。

於步驟 706，評估候選單元——即不在該超級單元或該簽章中之單元——以決定是否其匹配該簽章。當其對該簽章

單元之關係符合預先建立之規則時，一候選單元匹配該簽章。例如，該候選單元可給定一成員分數，反映其關係如何接近符合該簽章。該分數可以各種方式計算，且一最小分數可作為一「匹配」規則。在一具體實施例中，該成員分數根據關於該候選單元之簽章單元之分數，具有 50% (或 40% 或 90% 等等) 之一最小分數。在其他具體實施例中，簽章單元關於加權界線處，該候選者可根據簽章單元之分數加以評估，其中該候選者之關係之邊界加權於該加權界線中。在又其他具體實施例中，上述任何演算法或用以決定一叢集過程 (第 5 圖) 期間二單元之相似性之其他適用演算法可適用於為一候選單元決定一成員分數，使用該超級單元作為另一候選單元，及該簽章單元做為該超級單元之相鄰單元。

欲評估之候選單元之選擇可被簡化，如藉由僅考慮直接關於一或多個簽章單元之單元。如上所注意者，已經為該超級單元或該簽章之成員之單元可能被排除於該候選清單之外。

於步驟 708，匹配該簽章之任何候選單元 (如具有超過某閾值之一成員分數者) 加入該超級單元。於步驟 710，為該更新超級單元產生一新簽章。步驟 710 最好使用與步驟 704 相同之簽章產生技術，使得該新簽章與先前簽章間之任何差異係由於超級單元中變更之成員。

於步驟 712，藉由移除不匹配該新簽章之任何成員單元淨化該超級單元。步驟 712 最好使用與步驟 706 相同之

匹配規則。在某些具體實施例中，該種子單元於步驟 712 測試與淨化為任何其他成員單元；在其他具體實施例中，不淨化種子單元。在又其他具體實施例中，省略步驟 712，使得超級單元成員可成長但不縮水。

在步驟 714，決定是否該超級單元已收斂；若否，則該過程回到步驟 706，以循環根據該簽章更新該超級單元接著更新該簽章之步驟。收斂發生於當該超級單元或其簽章(或兩者皆是)之成員不於一循環期間變更時。某些具體實施例可為收斂實施一放鬆條件，允許於該超級單元或簽章中之一足夠小變化發生時找到收斂。

於步驟 716，一旦該超級單元收斂，該新超級單元加入單元字典 406。舉例來說，該超級單元可呈現為成員單元之一向量及成員加權之一向量(在此各成員之成員加權係自該最終簽章決定之其最終成員分數)。或者，該超級單元可呈現為為單元字典 406 之所有單元使用成員加權之一向量。在此情況中，並非該超級單元之成員之單元之加權可設定為零，或一成員加權可根據該最終簽章為非成員單元加以計算。該超級單元之最終簽章亦最好儲存於單元字典 406 中。

第 8A-B 圖舉例說明根據過程 700 超級單元建構期間之一概念網路 800 之一部分。第 8A-B 圖之概念規則大致類似於第 3 圖之概念規則，除了網路 800 僅顯示結合關係，且不標示關係邊界。(應了解在其他情況中，亦可考量除了結合以外的關係)。第 8A 圖顯示步驟 704 後概念網路 800

之狀態。節點(單元「艾薇兒」802與「席琳狄翁」804為超級單元「X」(虛線方塊806)之一種子之成員。節點「mp3」808、「歌詞」810、「照片」812、「相本」814、及「導覽」816已識別(於步驟704)為超級單元X之簽章「Y」(虛線方塊818)之成員。節點「芭芭拉史翠珊」820、「電影」822、而「阿諾史瓦辛格」824並非超級單元X或簽章Y之成員。

在此範例中，由於各為簽章Y之至少一單元之一鄰居，超級單元產生可能藉由識別該單元「芭芭拉史翠珊」820與「阿諾史瓦辛格」824為候選單元加以處理。各候選單元接著會根據某規則為至該簽章之一匹配加以評估。舉例來說，該候選者可能要求關於該簽章單元至少75%。該「芭芭拉史翠珊」節點820結合簽章Y中五個單元中的四個，且會於步驟708加入超級單元X。該「阿諾史瓦辛格」節點824僅結合簽章Y中單元中的一個，且不會於步驟708加入超級單元X。第8B圖顯示步驟706與708後概念網路800之狀態，其中該「芭芭拉史翠珊」節點820加入超級單元X'(虛線方塊806)。

接下來，簽章Y為超級單元X'更新(步驟710)。舉例來說，簽章Y可定義為僅包含結合超級單元X之成員至少50%之單元。該「芭芭拉史翠珊」單元820結合該「電影」單元822，但非其他成員；因此，「電影」並不加入簽章Y。該「照片」單元812不結合該「芭芭拉史翠珊」成員單元820，且結合三個單元中之其他兩個；因此，「照片」維持於該簽章中。

在此範例中，由於可能候選單元之成員分數不變更，簽章 Y 不於循環期間變更，且會找到收斂。應了解此範例被高度簡化；概念網路可被視為比第 8A-B 圖中所示之部分更大且更複雜，且許多循環可為一超級單元要求以收斂。

現在將為關於藥品之一超級單元描述根據過程 700 之超級單元產生之另一範例。在此範例中，該概念網路自一大量查詢產生(如由一主要網際網路搜尋提供者接收之一週內有效查詢，例如雅虎)。從該概念網路，使用一特殊藥品之廠牌名稱(如「Vicodin」)形成一支系作為一基本單元。根據上述過程 600 形成之支系包含為特殊藥品名稱(如「Oxycontin」、「Propecia」等等)之一小量(在此情況中為九)其他單元。

此支系使用為一超級單元種子(步驟 702)，用以產生超級單元組 X。該超級單元種子之各成員單元 x_i 指派初始化為一常數值(如對所有 x_i 來說 $W(x_i) = 1$)之一成員加權 $W(x_i)$ ，在其他具體實施例中，該叢集加權(如使用上述任何叢集演算法)或該支系成員分數(如從上述公式(6))可使用為該初始成員加權。

接著為該超級單元種子建立一簽章(步驟 704)。用於該「藥品」超級單元之種類之一簽章產生過程之一範例顯示於第 9 圖中為過程 900。於步驟 902，形成一初步簽章組 P，在此組 P 係超級單元組 X 之各成員單元 x_i 之鄰居之組 $V(x_i)$ 之聯集。在某些具體實施例中，該組 $V(x_i)$ 可包含少於該成員單元 x_i 之所有鄰居；舉例來說，可要求一最小邊

界加權或一特殊類型之關係，或該組可被摘取以移除重複單元(如僅可保留「map of Spain」或「Spain map」其中之一)。

於步驟 904，為初步簽章組 P 中各單元 p_j 計算一第一分數。該第一分數為一單元 p_j 最好反映超級單元組 X 之一成員 x_i 將為單元 p_j 之一鄰居之可能性。在該「藥品」超級單元範例中，為單元 p_j 之第一分數係「相關比例」(RP)分數，根據為該單元 p_j 之鄰居之單元 x_i 之成員加權 $W(x_i)$ 。舉例來說，若 $L(x_i, p_j)$ 定義為等於 1，若單元 x_i 係單元 p_j 之一鄰居且等於 0，否則該 RP 分數可計算為：

$$RP(P_j) = \frac{1}{N[X]} \sum_{i=1}^{N[X]} L(x_i, p_j) * W(x_i) \quad (7)$$

其中 $N[X]$ 表示超級單元組 X 中成員單元之總數。

於步驟 906，為初步簽章組 P 中各單元 p_j 計算一第二分數。該第二分數為一單元 p_j 最好反映潛在簽章單元 p_j 之一相鄰單元(即相鄰組 $V(p_j)$ 之一成員)之可能性亦為超級單元組 X 之一成員。在該「藥品」超級單元範例中，該第二分數係依相關頻率比(RFR)，給定為：

$$RFR(p_j) = 100 * \rho [V(p_j), X] / \rho [V(p_j)] \quad (8)$$

在此 $\rho [V(p_j), X]$ 表示相鄰組 $V(p_j)$ 之成員與組 X 之成員單元間關係之頻率(或邊界加權)之總合，而 $\rho [V(p_j)]$ 表示相鄰組 $V(p_j)$ 之所有成員之累積頻率。

於步驟 908，藉由結合該第一與第二分數為初步組 P 中各單元 p_j 計算一最終分數 $S_j(p_j)$ 。在一「藥品」超級單元之範例中，分別使用上述等式(7)與(8)中定義之 RP 與

RFR 分數，該最終分數給定為：

$$S_j(p_j) = RP(p_j) * \log RFR(p_j) \quad (9)$$

在其他具體實施例中，該最終分數 $S_j(p_j)$ 可為該 *RP* 與 *RFR* 分數之一不同組合。

於步驟 910，一閾值應用至該最終分數 $S_j(p_j)$ ，且具有高於該閾值之分數之單元 p_j 變成超級單元 *X* 之簽章組 *Y* 之簽章單元 y_j 。在該「藥品」超級單元範例中，該閾值由為各單元 p_j 將該最大值 $S_j(p_j)$ 除以一常數值加以決定；在此情況中，使用一常數值 6，但亦可選擇其他值。對於包含於簽章組 *Y* 中之各單元 y_j 來說，該最終分數 $S_j(y_j)$ 儲存為一成員加權 $W(y_j)$ 。

在產生該簽章後，為可能加入超級單元組 *X* 測試候選單元 c_k (過程 700 之步驟 708)，藉此建立一已修改超級單元 *X*。這些候選單元 c_k 從為至少一簽章單元 y_j (在此 y_j 係組 *Y* 之一成員) 之鄰居且還不是組 *X* 或組 *Y* 之成員之單元選取。對於各候選單元 c_k 來說，計算一成員分數，部分根據簽章組 *Y* 中之候選單元 c_k 與簽章單元 y_j 之相鄰單元 $V(c_k)$ 。為超級單元成員之成員分數之計算一般類似於上述為簽章說明之過程 900，且其本身之分數也以相似的方式定義。

尤其是，一分數係與上述等式 (7) 類似定義之一相關比例分數。也就是說，若 $L(y_j, c_k)$ 定義為等於 1，若單元 y_j 係單元 c_k 之一鄰居且等於 0，否則該 *RP* 分數為候選單元

c_k 定義為：

$$RP(c_k) = (1/N[Y]) * \text{Sum} [L(y_j, c_k), *W(y_j)] \quad (10)$$

在此 $N[Y]$ 係簽章組 Y 中單元 y_j 之總數，而 $W(y_j)$ 係成員分數（如上所注意之等式 (9) 之結果）。該第二分數係類似上述等式 (8) 定義之一相關頻率比分數。也就是說，若 $V(c_k)$ 表示候選單元 c_k 之所有相鄰單元， $\rho [V(c_k), Y]$ 表示相鄰組 $V(c_k)$ 之成員與簽章組 Y 中簽章單元間之關係之頻率或邊界加權之總合，而 $\rho [V(c_k)]$ 表示相鄰組 $V(c_k)$ 之所有成員之累積頻率，接著：

$$RFR(c_k) = 100 * \rho [V(c_k), X] / \rho [V(c_k)] \quad (11)$$

該最終分數 $S_j(c_k)$ 藉由結合該 RP 與 RFR 分數加以決定；即：

$$S_j(c_k) = RP(c_k) * \log RFR(c_k) \quad (12)$$

與上述等式 (9) 類似。一閾值應用至該最終分數 $S_j(c_k)$ ，以決定是否候選者 c_k 應加入超級單元組 X' 。此閾值藉由將所有候選單元 c_k 之最大值 $S_j(c_k)$ 除以一常數值；在此情況中，使用一常數值 6，但亦可選擇其他值。對於加入為一單元 x_i 之各候選者 c_k 來說，其成員加權 $W(x_i)$ 設定等於其最終分數。此成員加權使用於過程 700 之簽章更新步驟 710 之下一循環中。

在處理所有候選者之後，該超級單元產生過程繼續至

步驟 710，在此簽章組 Y 根據已更新超級單元組 X' 之成員更新至一新組 Y'。此藉由使用超級單元組 X' 之目前成員重新執行過程 900 加以完成。接著，於步驟 712，評估超級單元組 X' 之成員單元，以決定其是否應被移除；此過程使用與步驟 708 相同之分數計算與成員規則。

於步驟 714，分別藉由比較組 X' 與 Y' 和組 X 與 Y 決定收斂或非收斂。各組對間之沒有變化或一足夠小變化導致收斂。

第 10 圖為該「藥品」超級單元顯示結果。如上所註，該種子係根據單一廠牌名稱 (VICODIN) 之一支系；簽章加權藉由上述等式 (7)、(8)、及 (9) 加以決定；而超級單元成員加權藉由上述等式 (10)、(11)、及 (12) 加以決定。第 10A 圖顯示該簽章單元與其各自成員加權，於八個循環後，而第 10B 圖顯示某些超級單元成員與其各自加權，亦於八個循環後。這些結果從一大量實際使用者查詢產生，且該完整超級單元包含超過一百個成員，代表者如圖所示。

對於此範例來說，該簽章組由第 10A 圖中所列之六個單元組成。應注意有些單元可能吾人會預期一人會在搜尋關於一藥品之資訊時包含在內，但在不關於一藥品之搜尋中不包含在內。該超級單元成員，某些顯示於第 10B 圖中，包含大量各種藥品之廠牌名稱。(從「Vicodin」旁，其為該超級單元種子環繞其形成之基本單元，這些廠牌名稱列示於第 10B 圖中為 <brand A> 等等，由於該特定廠牌與其順序對本發明並不相關。) 亦包含藥品 (如「ibuprofen」、

「drug」、「caffeine」)、非法藥品(如「heroin」)、食物附加(如「aspartame」, 以及數種不同維他命(未列出))、及其他藥品相關詞(如「chemotherapy」)之一般名稱。

應了解此範例為舉例說明, 變化與修改是可能的, 且超級單元成員、簽章單元、及/或分數一般會與此範例中所提及者相異, 如若一不同概念網路使用做為該輸入。此外, 為簽章與超級單元成員分數描述之公式為舉例說明, 可依需求相異。

例如, 在某些具體實施例中, 潛在簽章單元之分數可不須參照該超級單元成員之成員加權 $W(x_i)$ 計算。在此一具體實施例中, 在此 $N[X \cap V(p_j)]$ 表示超級單元組 X 之成員且亦為一單元 p_j 之相鄰組 $V(p_j)$ 之成員之數目, 而 $N[X]$ 表示超級單元組 X 之成員總數, 一第一分數 S_1 為單元 p_j 反映超級單元組 X 之一成員為單元 p_j 之一鄰居之可能性, 可計算為:

$$S_1(p_j) = N[X \cap V(p_j)] / N[X] \quad (13)$$

同樣的, 一第二分數為單元 p_j 反映該單元 p_j 之一相鄰單元為超級單元組 X 之一成員之可能性, 可計算為:

$$S_2(p_j) = \rho [V(p_j), X] / \rho [V(p_j)] \quad (14)$$

在此 $\rho [V(p_j), X]$ 與 $\rho [V(p_j)]$ 定義於上。做為該第二分數為單元 p_j 之另一範例可計算為:

$$S_2'(p_j) = N[V(p_j) \cap X] / N[V(p_j)] \quad (15)$$

在此 $V(p_j)$ 表示一單元 p_j 之相鄰單元組， $N[V(p_j) \cap X]$ 表示單元 p_j 之相鄰組 $V(p_j)$ 中單元亦為 X 之成員之數目，而 $N[V(p_j)]$ 表示相鄰組 $V(p_j)$ 中相鄰單元之總數。

第一與第二分數可以想要之任何方式結合，以為應用包含於簽章組 Y 中之一閾值之目的決定一最終分數。或者，一分別截止可分別應用至各分數；如若對某些閾值 t_1, t_2 來說 $S_1(p_j) > t_1$ 且 $S_2(p_j) > t_2$ ，則一單元 p_j 係簽章組 Y 之一成員單元 y_j 。若使用分別截止於兩個分數上，兩個分數可儲存為成員加權。

應了解視為包含於超級單元 X 中之候選單元 c_k 之類似分數可以類似方式計算。舉例來說，若 $N(V(c_k) \cap Y)$ 表示為組 Y 中簽章單元之候選者 c_k 之相鄰單元數，則 $N[V(c_k)]$ 表示候選單元 c_k 之相鄰單元之總數，且 $N[Y]$ 表示簽章單元 Y 之總數，則兩個成員分數 S_1 與 S_2 可定義為：

$$S_1(c_k) = N[V(c_k) \cap Y] / N[V(c_k)] \quad (16)$$

且

$$S_2(c_k) = N[V(c_k) \cap Y] / N[Y] \quad (17)$$

與上述等式(13)與(15)類似。代表頻率之一定義亦可使用於一或兩個分數。無論是否加入一候選者單元 c_k 可根據該個別分數之一或兩者或其組合加以決定。

如上所述，不需要為該超級單元或簽章考慮一候選單元之所有鄰居。該候選單元可被限制，如根據一特殊關係

(如僅擴充)、一最小邊界加權、或其他規則。在一具體實施例中，使用的相鄰單元為該候選單元之「建議」，在此建議使用上方參照之申請案第 10/713,576 號中之細節說明之技術加以辨識。

應了解在此說明之超級單元建構過程為舉例說明，且可能有變化與修改。描述為連續的步驟可平行執行，步驟的順序可相異，且步驟可被修改或合併。多個超級單元可平行(或依序)建構，從任何數量之種子開始。除此之外，相異超級單元組可藉由使用該超級單元及/或簽章中之成員之不同規則自相同之概念網路(且選擇項相同種子)建構，藉此產生具有不同內容之超級單元。此外，上述範例參照結合關係，可考慮超級單元成員與簽章單元間之其他關係類型。同樣的，上述的範例僅考量該成員單元正鄰居之簽章單元；其他具體實施例可根據間接關係、查詢中超過兩單元之共同出現等等選擇簽章單元。

在某些方面，該超級單元建構過程係上述叢集產生過程 500(第 5 圖)之延伸。如在此所使用者，一「叢集」指的是根據其鄰域相似性相關之一單元群(即相關單元)；如此一來，一叢集可被視為超級單元之一類型，該簽章根據該共同鄰域定義。應了解亦可建立其他類型的超級單元，以抓取其他類型的關係，包含一超級單元的成員間之直接關係。舉例來說，該單元「britney spears」與「brittany spears」(一共同錯誤拼法)可能具有一共同鄰域，且與例如「芭芭拉史翠珊」與「席琳狄翁」之單元包含於一叢集類

型超級單元中，其清楚的指其他歌手。為了抓取該正確拼法「britney spears」與各種不正確拼法間之特殊關係，可建立替代之一超級單元。此類型之超級單元之簽章可包含存在一「替代」關係，具有某數量之其他成員(或具有單一「較佳」成員)以及(或取代)該共同鄰域。

再次參照第 4 圖，查詢引擎 404 最好用以配置執行其查詢處理操作於一循環基礎上(如每週、每日、每小時、當查詢接收及時等等)。在某些具體實施例中，根據一組新的查詢登錄檔中接收之查詢更新一既有單元字典；在其他具體實施例中，可從自該新查詢登錄檔組刮下產生一新單元字典。在任一情況中，應了解可自然包含該概念網路與該超級單元，以回應變更使用者行為。例如，若一新歌手變得受歡迎，他或她可能變成第 8 圖中之超級單元 X 之部分，由於使用者可能為該新歌手的名稱結合超級單元 X 之簽章單元開始搜尋。

在較佳具體實施例中，該超級單元傾向反映真實世界的概念關係(如屬於一類別之單元，例如歌手或城市)，即使查詢處理引擎 404 不需要提供有關於單元或查詢之真實世界知識或語義資訊。舉例來說，一超級單元可包含「New York City」、「San Francisco」、及「Chicago」，而該簽章為該超級單元可包含「hotel」、「restaurant」、及「night club」。此一超級單元可反映紐約市、舊金山、及芝加哥皆為旅遊目的地(或城市)，但查詢引擎 404(第 4 圖)不要求處理任何先前概念「目的地」(或「城市」)的知識。此概念知識可

自動從分析查詢樣式成長。應了解本發明詳述標示超級單元具有執行對人類有語義意義的詞，很方便促進了解本發明。事實上，查詢引擎 404 或單元字典 406 使用的任何超級單元標示架構不需要具有此屬性；例如，一超級單元標示可僅為一數字、為該超級單元至一加權向量或簽章之一參照等等。

在某些具體實施例中，超級單元可進一步藉由指派有語義的有意義標籤至某些或所有超級單元加以增強。舉例來說，與一標籤(如「artist」)結合一或多個關鍵字(如「lyrics」、「mp3」等等)之一分類關鍵字資料庫可提供以供指派標籤使用。一超級單元之簽章單元可比較該關鍵字，以決定是否應用該標籤。人賴索引編輯可參與此過程，如藉由建立該關鍵字資料庫及/或驗證指派標籤至超級單元。

一單元可能屬於多個超級單元；舉例來說，一模糊的單元，例如「Java」，可能結束於一「computer programming」超級單元、一「food and drink」超級單元、及一「travel」或「places」超級單元。在某些具體實施例中，亦可能有不屬於任何超級單元之單元。欲建立之超級單元數量可提前建立(不論為一特殊數目或一數目範圍)，且可為如 100、500、1500、或 5000。在其他具體實施例中，不預先決定該超級單元之數目。

應了解在此描述之系統與過程為舉例說明，且可能有變化與修改。描述為連續的過程步驟可平行執行，步驟可

合併，且步驟的順序可修改。舉例來說，視為超級單元中成員之候選者之單元組可以各種方式限制(如藉由限制候選者為出現相當頻繁的單元)，且該簽章單元組亦可被限制。在一具體實施例中，簽章單元可包含或限制於結合該超級單元至少一成員之「建議」。在此內容中，建議為已識別為輸入一特定查詢(或單元)之一使用者可能有興趣的可能事項之單元，且根據單元與頻率資訊之分析。用以產生建議之技術詳細描述於上方參照之美國申請案第 10/713,576 號中。

III. 查詢回應中超級單元之應用

超級單元資訊可以各種方式使用，以增強對一查詢之一回應。第 11 圖顯示可由第 2 圖之系統 110 使用以回應至一查詢之一方法。客戶端 120 傳送一查詢至搜尋伺服器系統 160。搜尋伺服器系統 160 傳送該查詢與/或其成分單元至存取單元字典 406 之一概念伺服器 180。概念伺服器 180 傳回關於該查詢之概念資料，例如自該查詢識別之一或多個單元以及為該各種單元之統計與超級單元資訊。此資訊可藉由散列該查詢取出，以識別其中包含之單元，並存取單元字典 406 以為各已識別單元取出項目。在此具體實施例中，單元字典 406 包含關於查詢處理期間有效之單元之任何資訊，且可包含完整或部分之一概念網路之一呈現方式。在一具體實施例中，該已傳回資訊包含關於與該查詢或其個別單元相關之超級單元之資訊。

搜尋伺服器系統 160 最好使用自概念伺服器 180 接收

之概念資料以回應該查詢。搜尋伺服器系統 160 傳回之結果最好包含回應該使用者查詢之結果，以及其他相關資訊，例如根據抓取於單元中與其關係包括超級單元了解使用者需求，關於該使用者接下來可能想要揭示的暗示與提示。現在將說明超級單元資訊可使用以回應一查詢之方法之數個範例；應了解這些範例為舉例說明而非限制。

A. 解決模糊性

在某些具體實施例中，搜尋伺服器系統 160 可使用一多單元查詢之成分單元以解決該成分單元其中之一中之模糊性。舉例來說，假設一查詢包含可能使用於超過一內容中之一模糊詞，例如「Java」。此一詞可能屬於多個超級單元，如一「food and drink」超級單元、一「computer」超級單元、及一「location」超級單元。在將該查詢分析為單元，並偵測該單元「Java」中之模糊性後，搜尋伺服器 160 可比較查詢之其他成分單元與各超級單元之簽章。因此，若該查詢亦包含一詞，例如「shop」或「coffee」，搜尋伺服器系統 160 可推斷該使用者最可能有興趣的是「food and drink」超級單元，而例如「program」或「script」的詞可能指示該「computer」超級單元等等。結果(如連結至回應查詢之頁面)可呈現於群組中，對應至不同超級單元，最有可能的超級單元最先出現。在另一具體實施例中，來自不同超級單元(或內容)之結果可排列於該結果頁之不同「標籤」，允許使用者藉由按下該想要之標籤選擇一內容。在預設的情況下可顯示最有可能的內容。

超級單元亦可使用以其他方法解決模糊性，如藉由檢查該使用者可能在相同區段中所作之其他查詢。舉例來說，該單元「jaguar」可能指一動物或指一車。若該使用者於「jaguar」前之一查詢關於汽車但非動物(如「kelly blue book」或「porsche」)，可推斷該使用者比較可能有興趣汽車而非動物。此一推斷可藉由檢查相同使用者輸入之不同查詢中之單元之超級單元成員自動化；具有兩個單元為成員之一超級單元可識別為比另一個不具有兩個單元為成員之一超級單元要有可能。可考量任何數量的該使用者先前查詢，如給予最大加權之最近查詢。

搜尋伺服器系統 160 可使用各種技巧以決定如何分類結果。舉例來說，產生該頁面索引(如第 2 圖之頁面索引 170)之搜尋相關演算法可用以使用來自單元字典 406 之目前超級單元資料，以指派該索引中各頁面或網站(或其他內容單元)至一或多個超級單元；該超級單元指派可儲存於該索引(如作為一內容識別碼 172)中，以及關於特殊詞或單元出現之其他資料。

B. 建議相關搜尋

在某些具體實施例中，搜尋伺服器系統 160 可根據超級單元資訊建議相關搜尋。例如，假設一查詢包含「New York City」且此單元已知屬於一「destination」超級單元。搜尋伺服器系統 160 可能使用關於該超級單元之簽章以建議額外搜尋，例如搜尋「restaurant」或「hotel」結合「New York City」。此建議可根據該超級單元之簽章單元。

C. 建議「側向」搜尋

在某些具體實施例中，搜尋伺服器系統 160 亦可使用超級單元資訊以建議類似或相關網站之「側向」搜尋。舉例來說，假設一使用者有興趣於 W 日自 A 點飛至 B 點。該使用者可能直接存取一航空公司網站，如一美國航空網站，且執行該網站中之一搜尋，或該使用者可能要求「airlines」或「air travel」或「American airlines」等等之一搜尋，自顯示之搜尋結果中從連結中存取一特定網站（如該美國航空網站），且於該存取網站中要求關於 W 日自 A 點至 B 點之航班之資訊。該使用者現在從美國航空的網站檢視關於該要求資訊之資訊，也許包含關於各種有效航班之價格資訊。一「側向」搜尋讓該使用者利用相同資訊搜尋另一網站，如 A 與 B 點及 W 日，以獲得類似結果，而不需要手動存取新網站並重新輸入想要的資訊。

在本發明一具體實施例中，搜尋伺服器系統 160 可能提示該使用者於建議「相關」網站上執行側向搜尋，使用超級單元資訊以識別相關網站。例如，該單元「American airlines」可能屬於一「airline」或「transportation」超級單元；搜尋伺服器系統 160 可識別該超級單元中其他單元（如「United Airlines」）並建議執行關於該單元之一網站上之搜尋。若該使用者選擇該側向搜尋，該系統處理該已識別網站之介面，以提供想要的搜尋結果，舉例來說，列出於 W 日自 A 點至 B 點之航班之價格資訊之已識別網站之一頁面。在該使用者已經直接存取一網站且輸入搜尋資訊

至關於該已存取網站之一表單中之情況中，該搜尋模組 126 儲存此輸入資訊，並使用此資訊於當要求一側向搜尋需要用以填入相關網站中之表單處。當然，該使用者可能必須輸入額外資訊於一新網站中，根據該已選擇網站之需求而定。依此方式，該使用者提供有為類似資訊跨不同網站流線型類似搜尋之能力。

D. 解決拼字錯誤

在某些具體實施例中，超級單元與簽章可使用以提供查詢處理期間之增強拼字檢查。舉例來說，若一使用者輸入一查詢包含「basset」，傳統搜尋伺服器系統可能辨認「bassett」或「basket」為可能的替代品，且可能建議一或兩者給使用者。具有對超級單元資料存取之搜尋伺服器 160 可產生概念網路，以決定何替代拼字是該使用者最有可能想要的。

舉例來說，假設包含「basset」之先前查詢具有一簽章靠近「bassett」而非「basket」(如由於「basset」與「hound」一起出現比與「weaving」一起出現要頻繁得多)。在此情況中，該搜尋伺服器可能建議「basset」作為最佳替代形式。在另一實施方式中，完整查詢可能針對關於含有「basket」之一超級單元之個別簽章比較含有「bassett」之另一超級單元，根據何簽章較靠近匹配該查詢做成一建議。因此，搜尋伺服器 160 可以一建議搜尋「bassett hound」回應至該查詢「basset hound」，且以一建議搜尋「basket weaving」回應至查詢「basset weaving」。

E. 支援目錄型搜尋

在另外具體實施例中，超級單元資訊可使用以建構一階層性分類單元。在一具體實施例中，執行多階段的超級單元建構。在第一階段，可使用相當嚴格的成員規則，藉此建立呈現低層次階層之超級單元。例如，一「cities」超級單元、一「states」超級單元、及一「nations」超級單元可於此階段建構。在一稍後階段中，可再次建構超級單元(選擇性具有較不嚴格規則)，從初始超級單元組開始，藉此建立一較高階超級單元(例如包含城市、洲、及國家之一「places」超級單元)。或者，一循環超級單元建構過程(如第 7 圖之過程 700)中之不同階段可使用以辨識階層之不同階層。

根據超級單元之一階層性分類可使用以提供目錄型搜尋功能，類似於目前由雅虎及其他搜尋服務提供者所提供者。傳統目錄型搜尋系統單獨依賴人類編輯團隊建構該目錄；自超級單元建構一目錄使該過程自動化，且可導致較快速調適變更使用者興趣與行為之一目錄。

F. 其他應用

超級單元亦可使用於其他方式。例如，在某些具體實施例中，一網站操作人員或其他實體可「贊助」一超級單元，使得顯著的顯示由該贊助者提供之一廣告(或只是至該贊助者網站之一連結)，無論一查詢何時包含關於該贊助者超級單元之一單元。在其他具體實施例中，一查詢中之詞可比較超級單元名稱，且可建議超級單元之其他成員之相關

搜尋。在又其他具體實施例中，若一查詢詞符合一超級單元名稱，關於其他查詢詞之頁面可能根據是否該內容對應至該超級單元而加以排名。

IV. 另外具體實施例

本發明已針對特定具體實施例加以描述，熟知該項技術人士應了解可能有許多修改。例如，超級單元之數目與特性可能不同，且一單元可能屬於超過一超級單元。根據實施方式，可能或可能不需要每個單元屬於至少一超級單元。超級單元與簽章可動態定義，且概念發掘與/或概念分析可時時執行(如每日或每週)以更新單元、超級單元、及/或簽章資料，以回應變更使用者之行為。如上所提及的，可使用各種用以識別與結合單元以建立超級單元之技術。超級單元可傾向反映概念之真實世界關係，沒有需要所有的超級單元(或任何超級單元)做到任何特定程度。除此之外，該超級單元不需要反映從真實世界知識建立之一階層性目錄結構或其他分類，例如雅虎目錄。在此描述之自動化系統與方法可修改或增加所有或部分結果單元字典、超級單元、簽章、特殊索引頁面或網站之超級單元指派等等之人類檢閱。

在此描述之具體實施例可參照網站、連結、及其他特定範例之術語，在此全球資訊網(或其子組)做為該搜尋語言資料庫。應了解在此說明之系統與過程可適用於使用一不同搜尋語言資料庫(例如一電子資料庫或文件儲存)，且結果可包含內容以及可找到內容之位置之連結或參照。

因此，雖然本發明已針對特定具體實施例加以描述，應了解本發明意欲涵蓋下列申請專利範圍之範疇中之所有修改與對等變化。

【圖式簡單說明】

第 1 圖係根據本發明一具體實施例之一資訊取出與通訊系統之一簡化高階方塊圖。

第 2 圖係根據本發明一具體實施例用以通訊媒體內容之一資訊取出與通訊網路之一簡化方塊圖。

第 3 圖係根據本發明一具體實施例之一概念網路之一圖形呈現方式。

第 4 圖係根據本發明一具體實施例之一查詢處理引擎之一簡化方塊圖。

第 5 圖係根據本發明一具體實施例用以產生可使用為超級單元種子之叢集之一過程之一流程圖。

第 6 圖係根據本發明一具體實施例用以產生可使用為超級單元種子之支系之一過程之一流程圖。

第 7 圖係根據本發明一具體實施例用以自種子建構超級單元之一過程之一流程圖。

第 8A-B 圖係第 7 圖中舉例說明之超級單元產生過程中不同階段之一概念網路之圖形呈現方式。

第 9 圖係根據本發明一具體實施例用以為一超級單元建構一簽章組之一過程之一流程圖。

第 10 圖顯示第 7 圖之超級單元產生過程之一範例結果，其中第 10A 圖顯示簽章單元而第 10B 圖顯示代表性超

級單元成員。

第 11 圖係根據本發明一具體實施例包含一單元字典與相關處理智慧之一系統之一簡化方塊圖，包含某些方面之一查詢處理引擎。

【主要元件符號說明】

10 資訊取出與通訊網路	180 概念伺服器
20 客戶端	220 客戶端 2
22 使用者介面裝置	302 節點
40 網際網路	304 邊界
50 ₁ 伺服器	400 系統
50 _N 伺服器	402 查詢登錄
110 資訊取出與通訊網路	404 查詢處理引擎
120 客戶端	408 概念網路
125 模組	410 概念網路建立器
126 特定搜尋模組	412 超級單元種子模組
127 使用者介面模組	414 超級單元建立器
128 應用程式介面模組	416 外部種子來源
140 網際網路	500 過程
150 ₁ 內容伺服器	502 提供概念網路至超級單元種子 模組
150 _N 內容伺服器	
160 搜尋伺服器	503 為叢集選擇基本單元
162 項目	504 選擇欲加至叢集之候選單元
170 頁面索引	506 辨識以基本&候選單元出現之某 相鄰單元之範例
172 單元字典	

- 508 根據相鄰/已匹配單元計算叢集
加權
- 510 決定是否包含候選單元於具有
基本單元之叢集中
- 600 過程
- 602 將單元 b_i 放入支系 Q 中
- 604 獲得叢集 $C(b_i)$ 之下一個成員單
元 u_j
- 606 獲得叢集 $C(u_j)$
- 608 Q 的所有成員都在 $C(u_j)$ 中嗎?
- 610 對於支系 Q 的各成員單元 q_k 來
說，獲得叢集 $C(q_k)$
- 612 u_j 在所有 $C(q_k)$ 中嗎?
- 614 加入單元 u_j 至支系 Q 中
- 616 更多 u_j 在 $C(b_i)$ 中嗎?
- 618 為各成員 q 計算成員分數
- 700 過程
- 702 接收超級單元種子
- 704 為種子決定簽章
- 706 為匹配至簽章評估候選單元
- 708 加入匹配的候選單元至超級單
元
- 710 根據已擴充超級單元成員修改
簽章
- 712 移除不匹配新簽章之成員單元
- 714 收斂?
- 800 概念網路
- 802 艾薇兒
- 804 席琳狄翁
- 806 簽名 X
- 806' 簽章 X'
- 806 mp3
- 810 歌詞
- 812 照片
- 814 相簿
- 816 導覽
- 818 簽名 Y
- 820 芭芭拉史翠珊
- 822 電影
- 824 阿諾史瓦辛格
- 900 過程
- 902 建立初步組 P (為超級單元 X 之所有
成員之相鄰組聯集)
- 904 為組 P 中之各單元計算相關比例
(RP) 分數
- 906 為組 P 中之各單元計算相關頻率
比(RFR) 分數
- 908 從 RP 與 RFR 分數計算最終分數

910 應用閾值截止至最終分數，以決定組 P 之何成員包含於簽章組 Y 中

伍、中文發明摘要：

本發明在一搜尋處理系統中，藉由分析該查詢為單元，且部分根據一起出現於查詢中之單元模式，以定義該單元間各種關係，以自一組查詢產生一概念網路。具有某些相似特性之概念網路中之單元被分組為超級單元。對於各超級單元來說，有一對應簽章定義該群組之類似特性。一查詢之處理係藉由識別成分單元、決定某些或所有成分單元之超級單元成員、及使用該資訊以訂定對該查詢之一回應。

陸、英文發明摘要：

In a search processing system, a concept network is generated from a set of queries by parsing the queries into units and defining various relationships between the units based in part on patterns of units that appear together in queries. Units in the concept network that have some similar characteristic(s) are grouped into superunits. For each superunit, there is a corresponding signature that defines the similar characteristic of the group. A query is processed by identifying constituent units, determining the superunit membership of some or all of the constituent units, and using that information to formulate a response to the query.

拾、申請專利範圍：

1. 一種用以從一概念網路產生超級單元之電腦實施方法，該概念網路包含複數單元與定義於該複數單元對間之複數關係，其中各關係具有一相關邊界加權，該方法包含之動作有：

識別包含至少一成員單元之一超級單元種子，其中各成員單元係該概念網路之複數單元其中之一；

為該超級單元種子定義一簽章，該簽章包含一或多個簽章單元，其中各簽章單元與至少一最小數量之成員單元於該概念網路中具有一關係；

藉由加入來自該概念網路之一或多新成員單元擴充該超級單元種子，其中各新成員單元根據該簽章滿足一匹配規則；

根據該已擴充超級單元種子修改該簽章；

重複擴充與修改之動作，直到滿足一收斂規則為止，其中一旦滿足該收斂規則，形成一最終超級單元與一最終簽章；及

為該最終超級單元之各成員單元儲存超級單元成員資訊。

2. 如申請專利範圍第 1 項所述之方法，其中上述概念網路係自一組先前之搜尋查詢產生。
3. 如申請專利範圍第 1 項所述之方法，其中上述儲存該超級單元成員資訊之動作包含之動作有：

為該最終超級單元之各成員單元計算一成員加權，其中上述成員加權係基於該成員單元與該最終簽章之簽章單元間之概念網路中之關係，

其中上述已儲存超級單元成員資訊包含該成員加權。

4. 如申請專利範圍第 1 項所述之方法，其另包含自該先前查詢產生該概念網路之動作。
5. 如申請專利範圍第 1 項所述之方法，其另包含之動作有：

於修改該簽章之動作後與該重複動作前，藉由移除根據該已修改簽章不符合該匹配規則之一成員單元淨化該超級單元，

其中上述淨化動作於重複該修改步驟之後重複。

6. 如申請專利範圍第 1 項所述之方法，其中由於該重複動作，若該超級單元種子之成員數目之變更不超過該等單元數目的最大值，則滿足上述收斂規則。
7. 如申請專利範圍第 1 項所述之方法，其中由於該重複動作，若該簽章之成員數目之變更不超過該等單元數目的最大值，則滿足上述收斂規則。
8. 如申請專利範圍第 1 項所述之方法，其中上述識別該超級單元種子之動作包含形成二或多單元之一叢集做為該超級單元種子之動作，其中在上述叢集中之各單元具有與該叢集之一基本單元相同之至少一相鄰單元。

9. 如申請專利範圍第 8 項所述之方法，其中上述形成該叢集之動作包含的動作有：

自該概念網路選擇一基本單元與一候選單元；

識別該基本單元之複數相鄰單元，其中各相鄰單元於該概念網路中具有對該基本單元之一關係；

識別該相鄰單元至少其中之一為一已匹配單元，其中各已匹配單元於該概念網路中具有對該候選單元之一關係；

根據包含該至少一已匹配單元之複數相鄰單元，為該候選單元計算一叢集加權；及

根據該叢集加權，決定是否包含該候選單元於具有該基本單元之一叢集中。

10. 如申請專利範圍第 1 項所述之方法，其中上述識別該超級單元種子之動作包含形成二或多密切相關單元之一支系。

11. 如申請專利範圍第 1 項所述之方法，其中上述識別該超級單元種子之動作包含自一外部來源接收一單元清單之動作，該單元清單可用作一超級單元種子。

12. 如申請專利範圍第 11 項所述之方法，其中上述外部來源包含一網頁。

13. 如申請專利範圍第 11 項所述之方法，其中上述識別該超級單元種子之動作另包含刪減該單元清單之動作，以移除不在該概念網路中之一單元。

14. 如申請專利範圍第 1 項所述之方法，其中上述識別該超級單元種子之動作包含之動作有：

接收關於該先前查詢之使用者行為資料；及

偵測關於含有不同單元之先前查詢之使用者行為資料中之相似性。

15. 如申請專利範圍第 14 項所述之方法，其中上述使用者行為資料包含藉由該先前查詢之資訊點選。

16. 如申請專利範圍第 1 項所述之方法，其中上述辨識該超級單元種子之動作包含之動作有：

偵測一來源文件中該概念網路之單元之出現；及

根據該已偵測出現產生一超級單元種子。

17. 如申請專利範圍第 1 項所述之方法，其中上述概念網路之單元間之關係包含一結合關係、一擴充關係、及一替代關係之一或多者。

18. 如申請專利範圍第 1 項所述之方法，其中上述定義該簽章之動作包含之動作有：

識別該概念網路中複數單元具有與該相關超級單元種子之至少一最小數量成員單元之一特定關係者為該簽章單元；及

建立一閾值數量，

其中上述擴充該超級單元種子之步驟包含：

自該概念網路選擇一候選單元；及

於該候選單元具有與至少該閾值數量之簽章單元

之特定關係之事件中，加入該候選單元至該超級單元種子。

19. 如申請專利範圍第 18 項所述之方法，其中藉由參照一預定分數，相對於一總數簽章單元，建立上述閾值數量。

20. 如申請專利範圍第 1 項所述之方法，其中上述定義該簽章包含之動作有：

將在該概念網路中具有與該相關超級單元種子之至少一最小數量成員單元之一特定關係之複數單元識別為該簽章單元；

為各簽章單元建立一邊界加權範圍；及

建立一閾值數量，

其中上述擴充該超級單元種子之步驟包含：

自該概念網路選取一候選單元；

決定等於一數量簽章單元之一第一數量，其中該候選單元與其具有該特定關係，且具有該簽章單元於該邊界加權範圍中之一邊界加權；及

於該第一數量等於或大於該閾值數量之事件中，加入該候選單元至該超級單元種子。

21. 如申請專利範圍第 1 項所述之方法，於該儲存動作之後另包含之動作有：

接收一目前查詢；

將該目前查詢分析為一或多成分單元；

為該一或多成分單元取出該已儲存超級單元成員

資訊；及

至少部分根據該已取出超級單元成員資訊，訂定對該目前查詢之一回應。

22. 如申請專利範圍第 21 項所述之方法，其中上述訂定該回應之動作包含使用該超級單元成員資訊，建議一相關搜尋查詢之動作。
23. 如申請專利範圍第 22 項所述之方法，其中上述相關搜尋查詢包含一第一單元，其係該超級單元之成員單元其中之一，其中上述第一單元並非該目前查詢之一成分單元。
24. 如申請專利範圍第 22 項所述之方法，其中上述相關搜尋查詢包含一第一單元，其係該超級單元之簽章單元其中之一，其中上述第一單元並非該目前查詢之一成分單元。
25. 如申請專利範圍第 21 項所述之方法，其中上述訂定該回應之動作包含使用該超級單元成員資訊，為一旁支搜尋建議一網站之動作。
26. 如申請專利範圍第 21 項所述之方法，其中上述成分單元其中之一係多於一超級單元之一成員，且其中上述訂定該回應之動作包含：根據該成分單元其中之一附屬之超級單元，使用該超級單元成員資訊，群組化回應資料之動作。
27. 如申請專利範圍第 21 項所述之方法，其中上述訂定該

回應之動作包含：根據對該組成單元之另一者與一或多超級單元之簽章單元之比較，使用該超級單元資訊解決該成分單元之一第一成分單元之一模糊性之動作，其中該第一成分單元係一或多超級單元之一成員。

28. 如申請專利範圍第 21 項所述之方法，其中上述訂定該回應之動作包含：使用該超級單元資訊，以選擇欲顯示之贊助者內容。

29. 一種用以自使用者搜尋查詢產生超級單元之系統，該系統包含：

一概念網路建立器模組，其經調整以自複數先前查詢產生一概念網路，該概念網路包含複數單元與複數定義於該複數單元對間之複數關係，其中各關係具有一相關邊界加權；

一超級單元種子模組，其經調整以識別包含至少一成員單元之一超級單元種子，其中各成員單元係該概念網路之複數單元其中之一；

一超級單元建立器模組，其經調整以建構超級單元及以該超級單元種子開始之簽章，其中各超級單元包含複數成員單元，且其中各簽章關聯於該超級單元其中之一，其中各簽章包含一或多簽章單元，其中各簽章單元在該概念網路中具有與該相關超級單元之至少一最小數量成員單元之一關係；及

一儲存模組，其經調整以儲存該成員單元之超級單

元成員資訊，其中上述超級單元成員資訊係由該超級單元建立器模組提供。

30. 如申請專利範圍第 29 項所述之系統，其中上述超級單元建立器模組另經調整以為各超級單元種子定義一簽章，以藉由加入來自該概念網路之一或多新成員單元擴充該超級單元種子，其中各新成員單元根據該簽章滿足一匹配規則，以根據該已擴充超級單元種子修改該簽章，且重複該擴充與修改步驟直到滿足一收斂規則為止，其中一旦滿足該收斂規則，形成一最終超級單元與一最終簽章。
31. 如申請專利範圍第 30 項所述之系統，其中上述超級單元建立器模組另經調整以為該最終超級單元之各成員單元計算一成員加權，其中上述成員加權係基於該概念網路中該成員單元與該最終簽章之簽章單元間之關係，且儲存該成員加權於該儲存模組中。
32. 如申請專利範圍第 30 項所述之系統，其中於重複期間，若該超級單元種子之成員變更少於該等單元數目的最大值，則滿足上述收斂規則。
33. 如申請專利範圍第 30 項所述之系統，其中於重複期間，若該簽章之成員變更少於該等單元數目的最大值，則滿足上述收斂規則。
34. 如申請專利範圍第 29 項所述之系統，其中上述超級單元種子模組另經調整以將二或多單元之一叢集識別為

該超級單元種子，其中該叢集中之各單元具有相同之至少一相鄰單元。

35. 如申請專利範圍第 34 項所述之系統，其中上述超級單元種子模組另經調整以自該概念網路選擇至少二候選單元，以識別該候選單元之複數相鄰單元，其中各相鄰單元於該概念網路中具有對一或多候選單元之一關係，以根據該複數相鄰單元為該候選單元計算一叢集加權，且根據該叢集加權決定是否自該候選單元形成一叢集。

36. 如申請專利範圍第 34 項所述之系統，其中上述超級單元種子模組另經調整以自一外部來源接收一單元清單，該單元清單係可用作一超級單元種子。

37. 如申請專利範圍第 34 項所述之系統，其中上述超級單元種子模組另經調整以接收關於該先前查詢之使用者行為資料，且偵測關於含有不同單元之先前查詢之使用者行為資料之相似性。

38. 如申請專利範圍第 34 項所述之系統，其中上述超級單元種子模組另經調整以偵測一來源文件中之概念網路之單元之出現，且根據該已偵測出現產生一超級單元種子。

39. 如申請專利範圍第 29 項所述之系統，其另包含：

一查詢回應模組，其耦合至該儲存模組，且經調整以接收一目前查詢，以將該目前查詢分析為一或多成分

單元，以自該儲存模組取出一或多該成分單元之該超級單元成員資訊，且以至少部分根據該已取出超級單元成員資訊訂定對該目前查詢之一回應。

40. 一種包含以程式碼編碼之一電腦可讀取媒體之電腦程式產品，該程式碼包含：

用以識別一超級單元種子之程式碼，該超級單元種子包含至少一成員單元，其中各成員單元係一概念網路之複數單元其中之一，該概念網路包含複數單元與該複數單元對間定義之複數關係，其中各關係具有一相關邊界加權；

用以為該超級單元種子定義一簽章之程式碼，該簽章包含一或多簽章單元，其中各簽章單元於該概念網路中具有與至少一最小數量成員單元之一關係；

用以擴充該超級單元種子之程式碼，其係藉由加入來自該概念網路之一或多新成員單元，其中各新成員單元根據該簽章滿足一匹配規則；

用以修改該簽章之程式碼，其係根據該已擴充超級單元種子；

用以重複該擴充與修改步驟之程式碼，其係進行到滿足一收斂規則為止，其中一旦滿足該收斂規則，則形成一最終超級單元與一最終簽章；及

用以儲存超級單元成員資訊之程式碼，其係供該最終超級單元之各成員單元之用。

41. 如申請專利範圍第 40 項所述之電腦程式產品，其中上述程式碼另包含：

用以接收一目前查詢之程式碼；

用以將該目前查詢分析為一或多成分單元中之程式碼；

用以為一或多成分單元取出該已儲存超級單元成員資訊之程式碼；及

用以至少部分根據該已取出超級單元成員資訊，訂定對該目前查詢之一回應之程式碼。

42. 一種用以形成來自一概念網路之一叢集的電腦可實施方法，該概念網路包含複數單元與該單元間定義之複數關係，其中各關係具有一相關邊界加權，該方法包含之動作有：

自該概念網路選擇一基本單元與一候選單元；

識別該基本單元之複數相鄰單元，其中各相鄰單元於該概念網路中具有對該基本單元之一關係；

將至少一相鄰單元識別為一已匹配單元，其中上述已匹配單元於該概念網路中具有對該候選單元之一關係；

根據包含該至少一已匹配單元之複數相鄰單元為該候選單元計算一叢集加權；及

根據該叢集加權，決定是否包含該候選單元於具有該基本單元之一叢集中。

43. 如申請專利範圍第 42 項所述之方法，其另包含之動作有：

選擇一第二候選單元；及

使用該第二候選單元重複將至少一相鄰單元識別為一已匹配單元之動作，計算一叢集加權，及決定是否包含該第二候選單元於該叢集中。

44. 一種用以自一概念網路形成一支系之電腦實施方法，該概念網路包含複數單元與定義於該單元間之複數關係，其中各關係具有一相關邊界加權，該方法包含之動作有：

形成複數叢集，其中各叢集包含至少一基本單元；

選擇該複數叢集其中之一作為一開始叢集；

初始化一支系以僅包含該開始叢集之基本單元；且

為該開始叢集之各成員單元 u ，於以下事件中加入該成員單元 u 至該支系中：

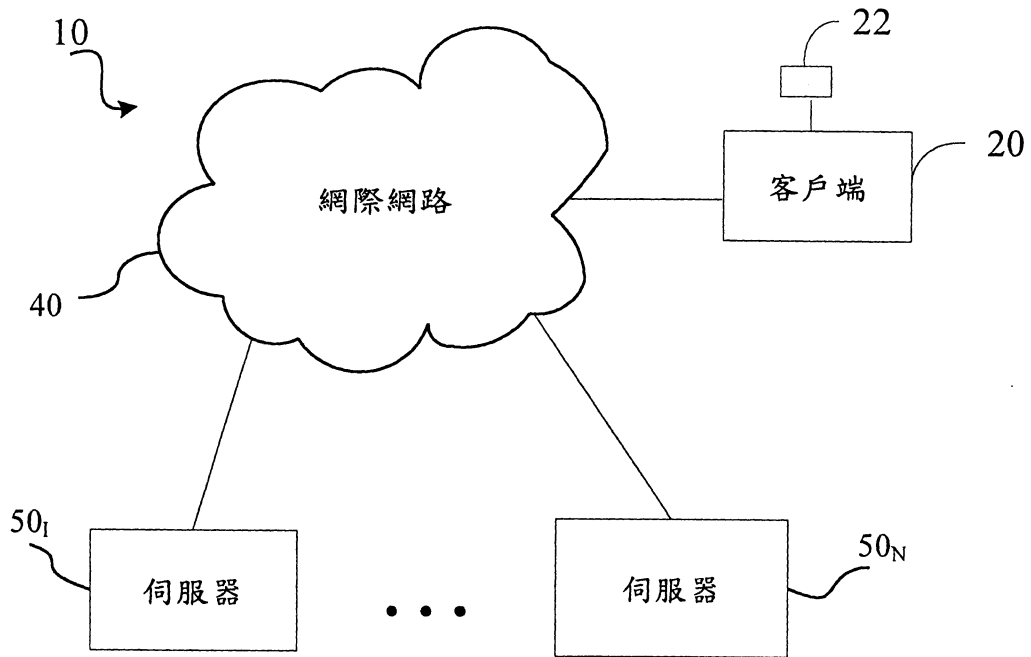
(a) 亦為具有成員單元 u 做為該基本單元之其中一叢集之成員之支系之目前成員之分數等於或大於一第一閾值；及

(b) 具有目前支系成員作為亦包含成員單元 u 之基本單元之叢集分數等於或大於一第二閾值。

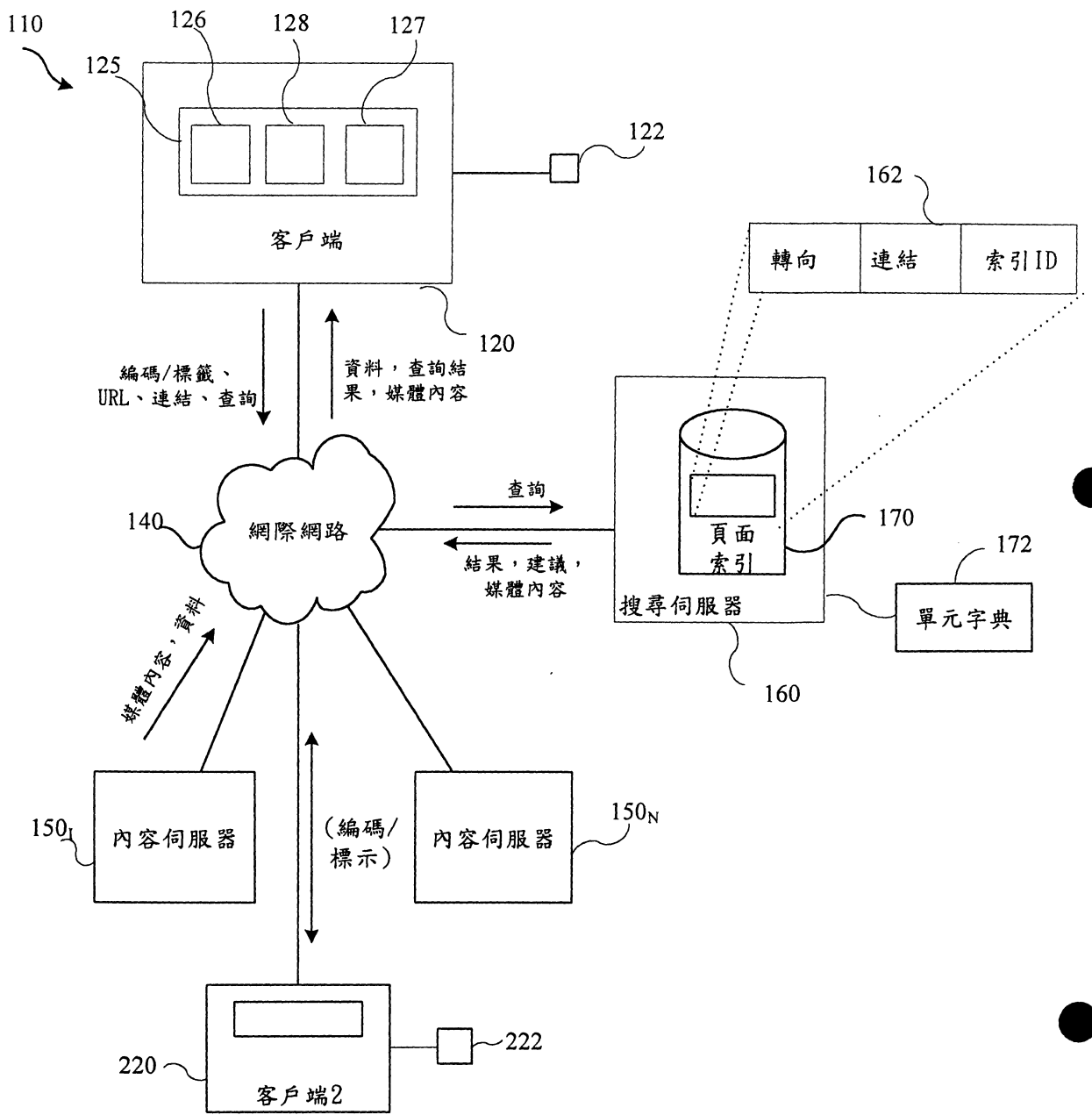
45. 如申請專利範圍第 44 項所述之方法，其中上述第一閾值與上述第二閾值各等於 100%。

46. 如申請專利範圍第 44 項所述之方法，其中上述第一閾

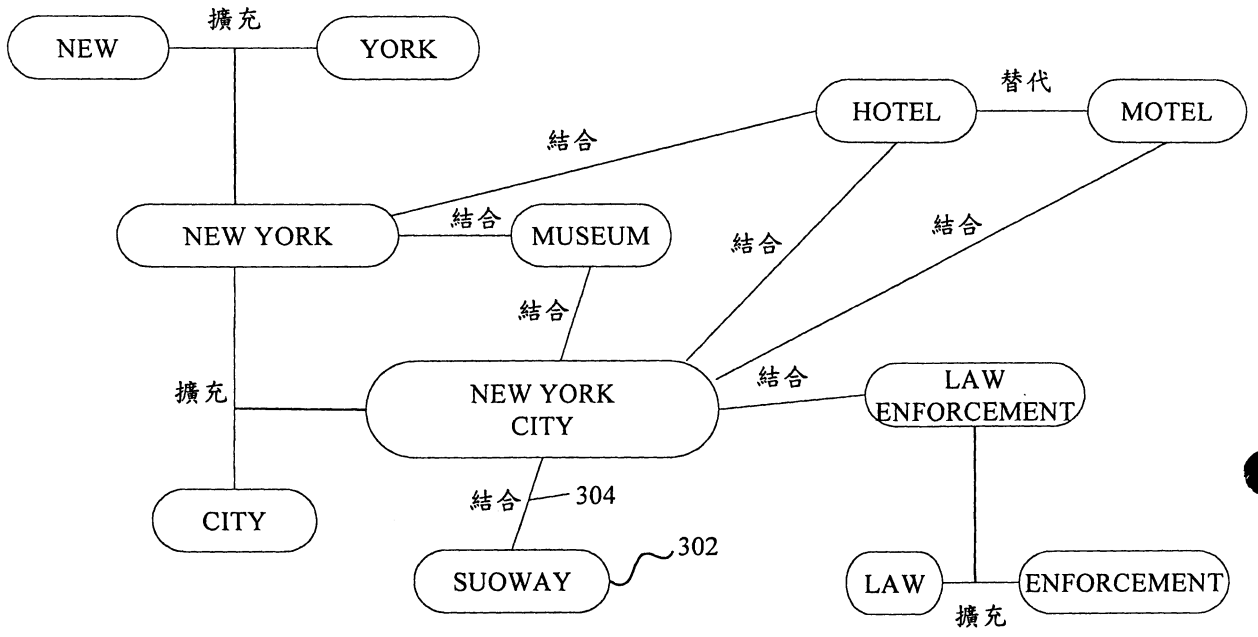
值與上述第二閾值各等於約 70%。



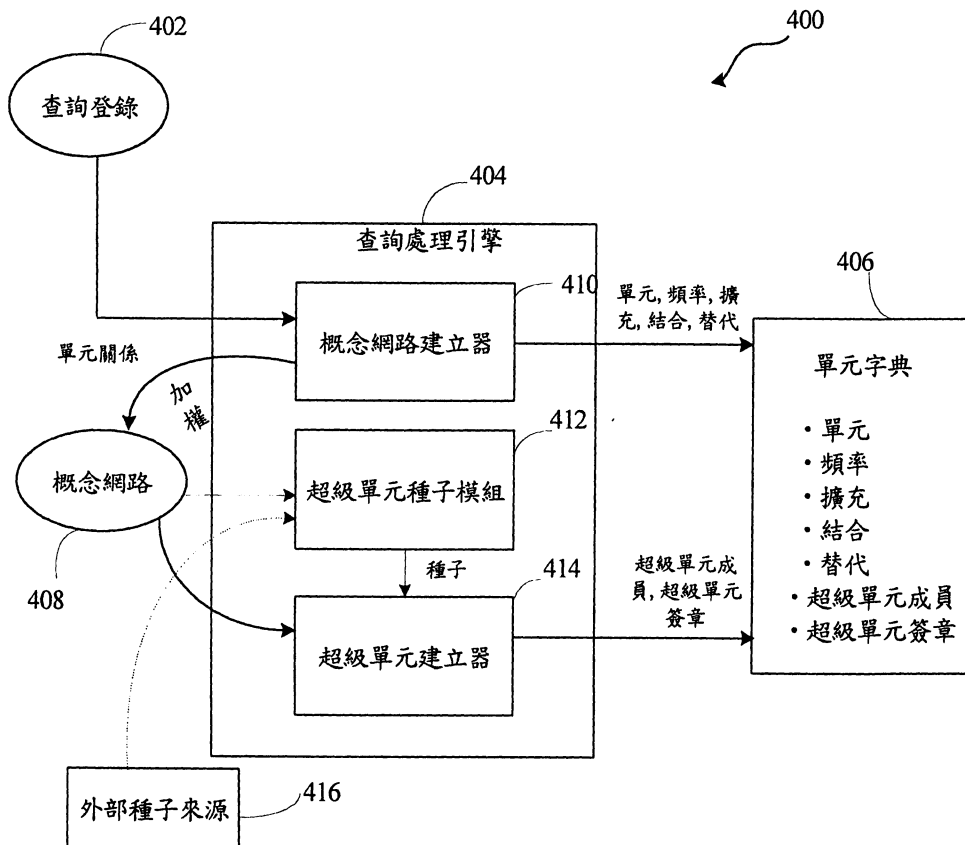
第 1 圖



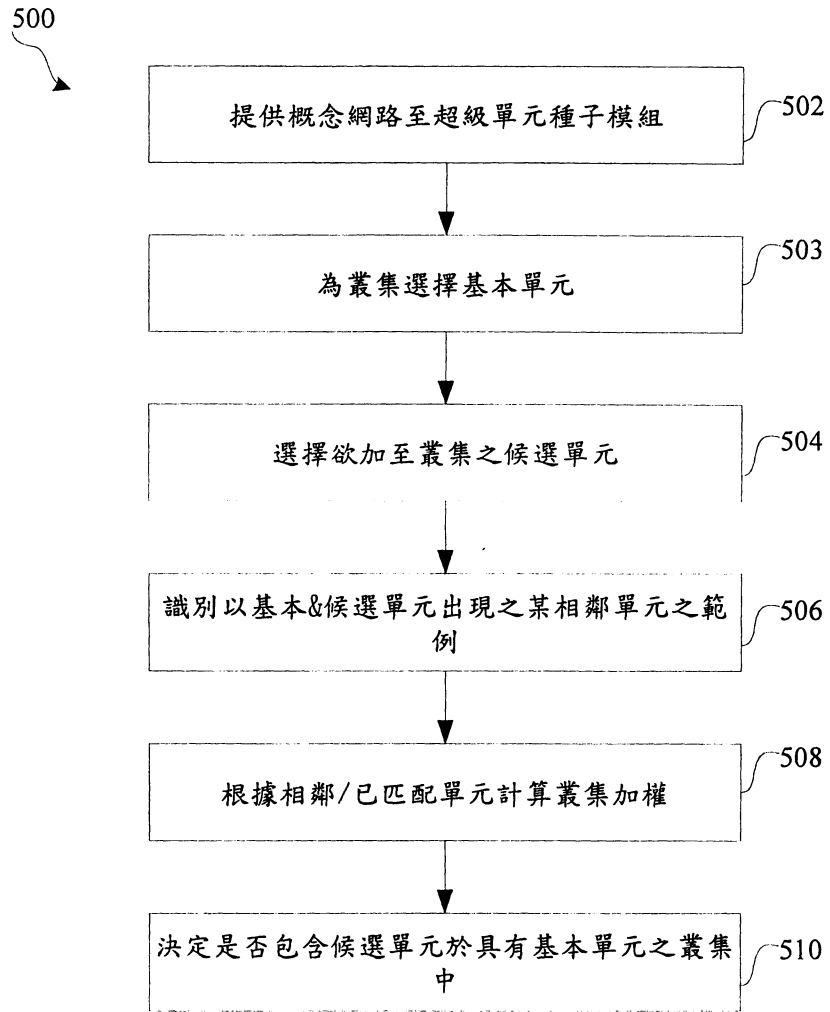
第 2 圖



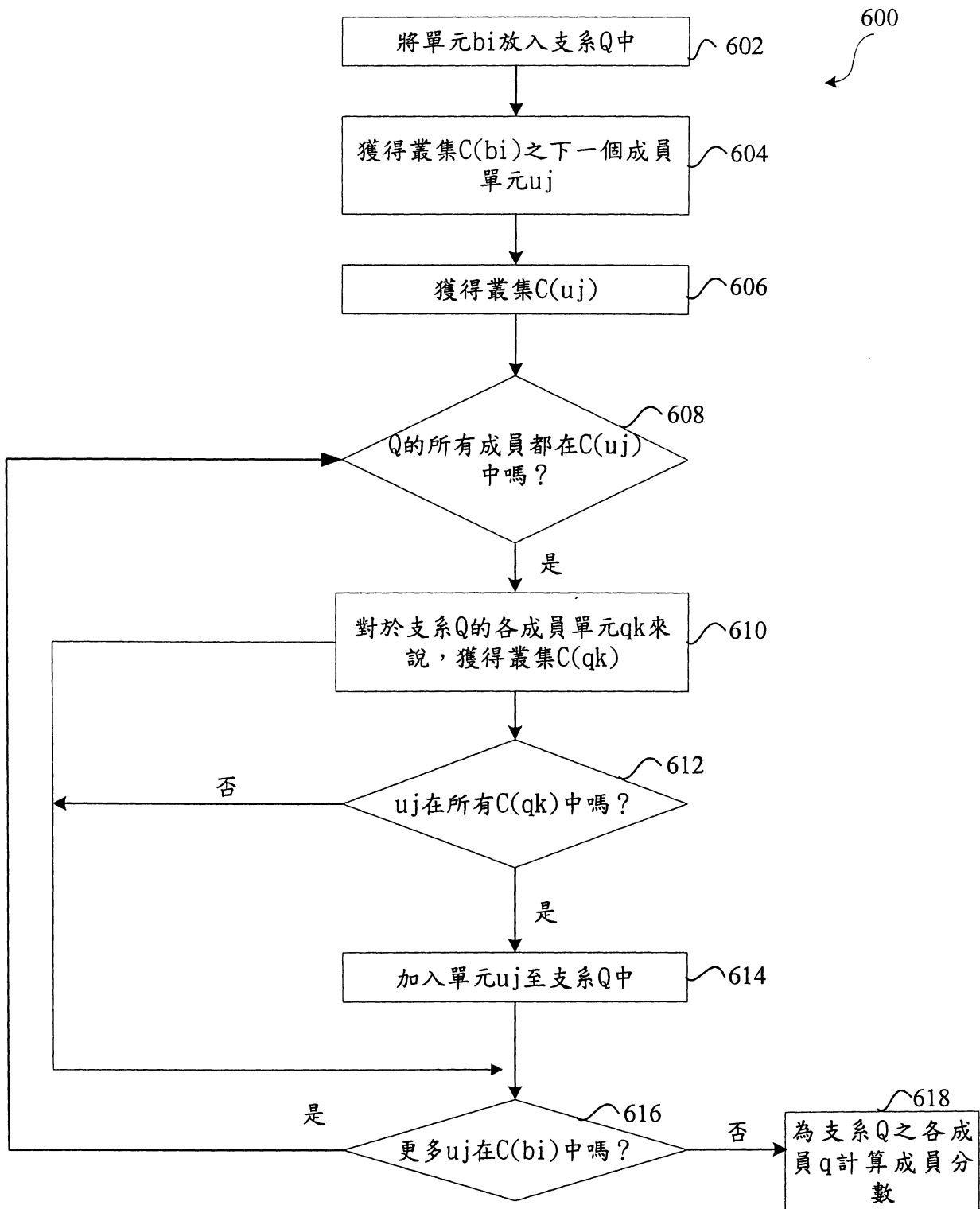
第 3 圖



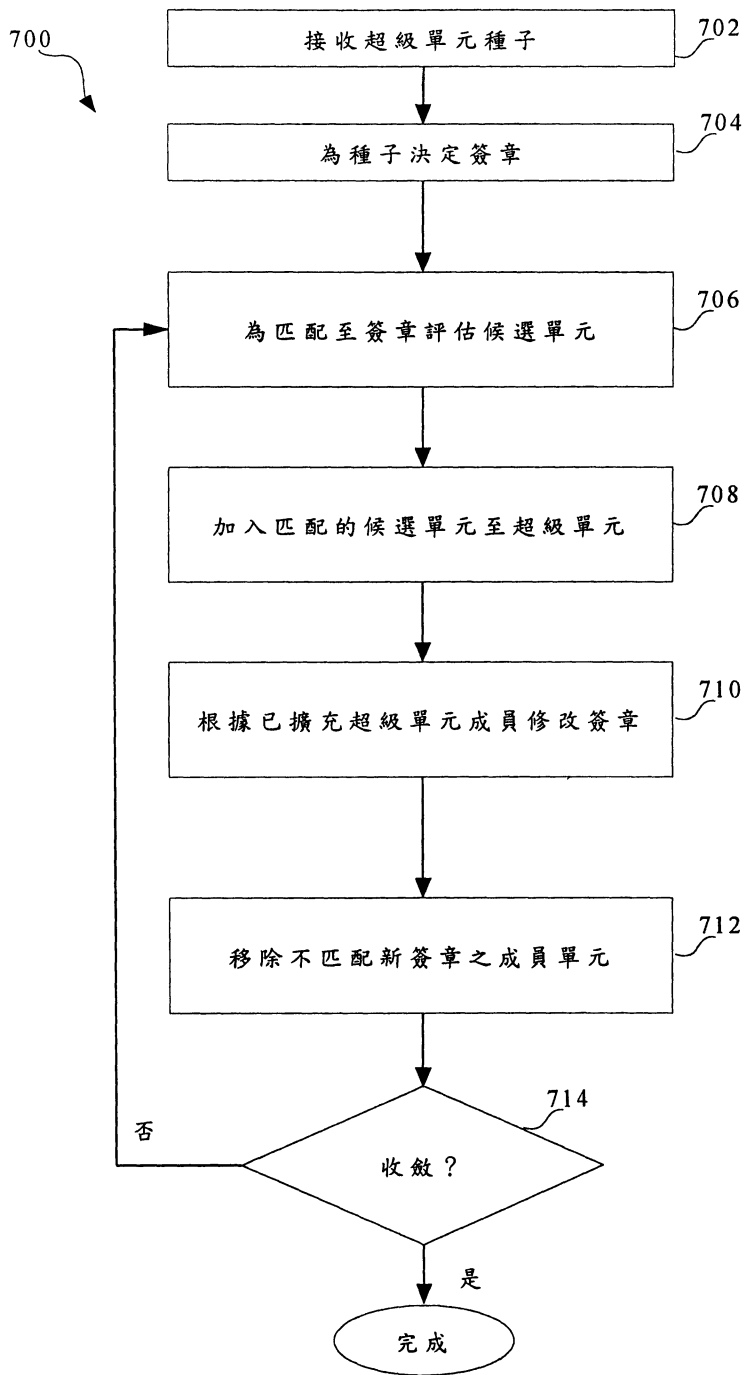
第 4 圖



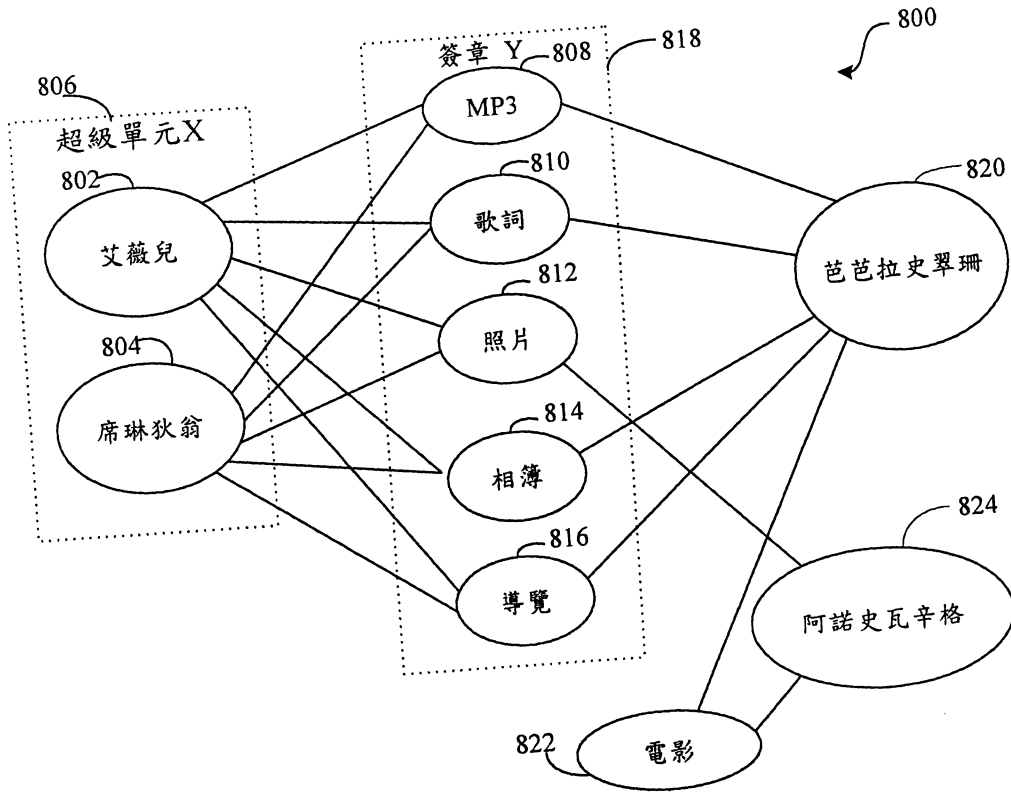
第 5 圖



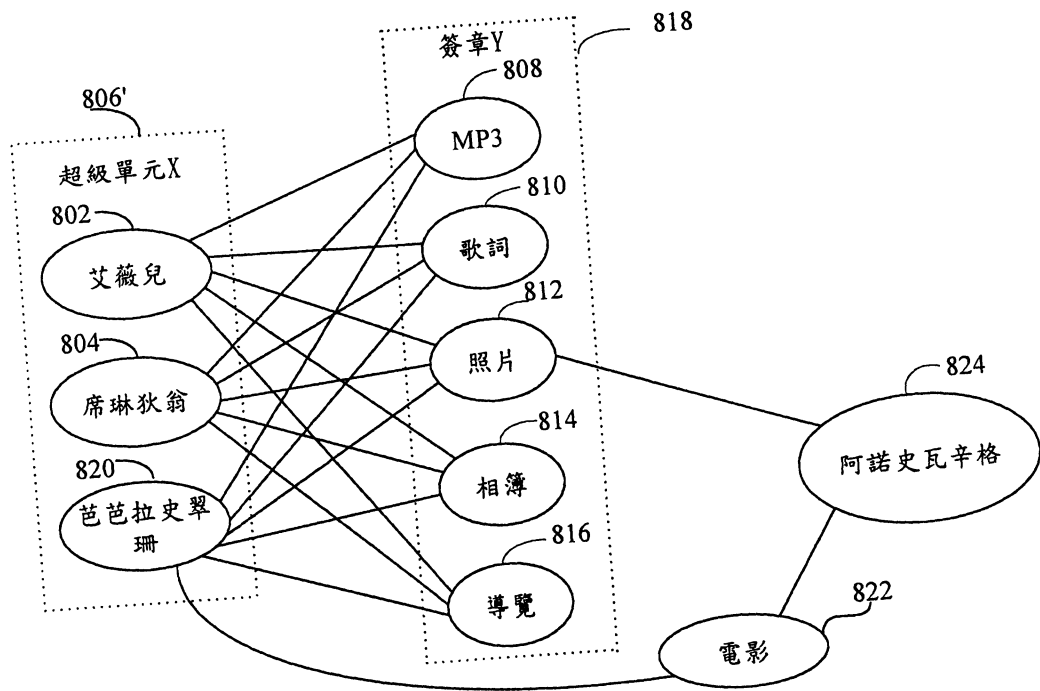
第 6 圖



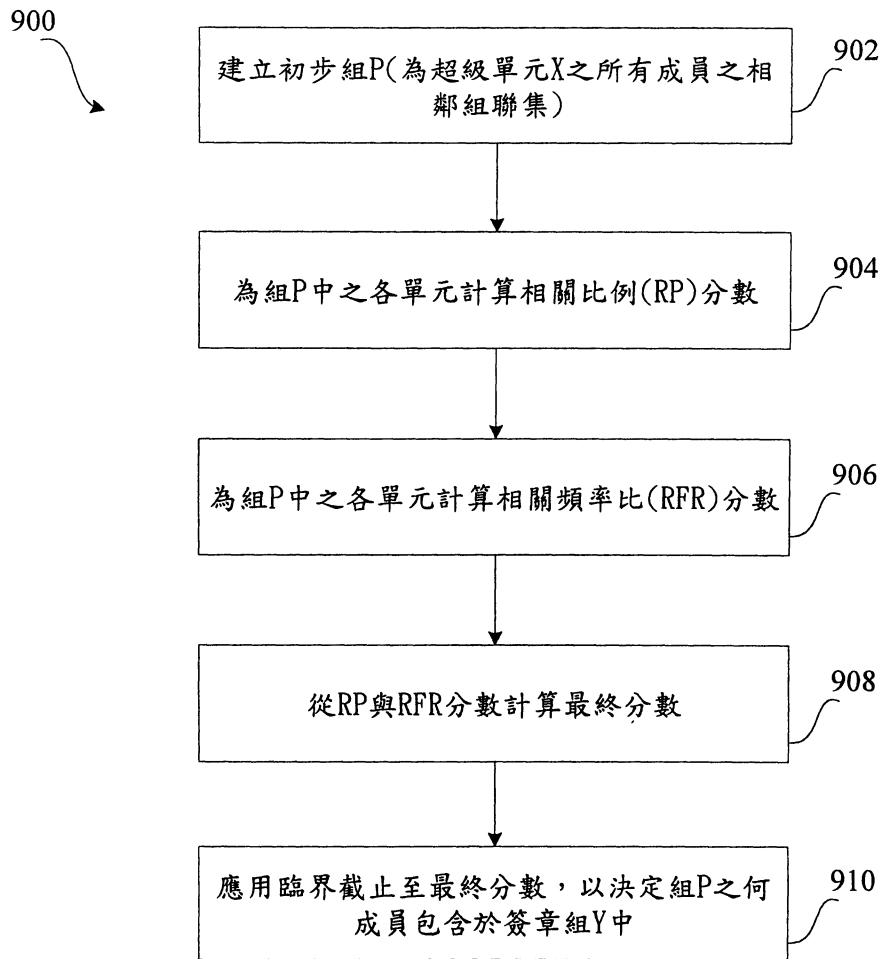
第 7 圖



第 8A 圖



第 8B 圖



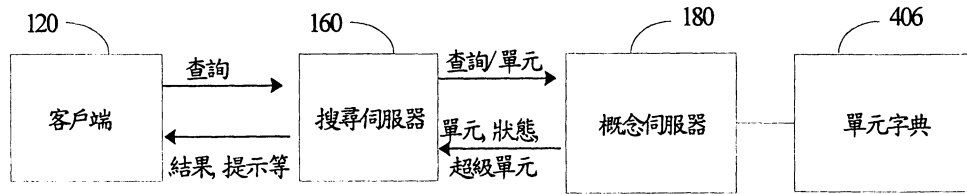
第 9 圖

排名	單元	加權 W(xi)
1	"side effect"	2535.1081555555
2	"dosage"	1715.60970691241
3	"overdose"	1012.67890927915
4	"generic"	962.26525161571
5	"withdrawal"	903.478068429062
6	"addiction"	374.038397276877

第 10A 圖

排名	單元	加權 W(xi)
1	<brand A>	5058.45052935936
2	<brand B>	4986.57847034358
3	<brand C>	4686.94999118965
4	<brand D>	4443.85674161793
5	<brand E>	4395.6107815323
6	<brand F>	4362.31743340316
7	<brand G>	4359.4591947467
8	<brand H>	4141.491219357
9	<brand I>	4008.98929826121
10	<brand J>	3834.00725848058
11	<brand K>	3831.09009064923
12	<brand L>	3769.71768473659
13	<brand M>	3681.19081682748
⋮		
19	"ibuprofen"	3238.3484470847
⋮		
61	"drug"	1818.41003172791
62	<brand N>	1783.6272487977
63	<brand O>	1771.02988169388
64	<brand P>	1768.27751636375
65	"caffeine"	1763.63682236685
⋮		
87	"chemotherapy"	1407.43386959684
⋮		
102	"heroin"	1161.6056979906
103	"aspartame"	1151.13482596151
⋮		
117	<brand R>	1009.53779331573

第 10B 圖



第 11 圖

柒、指定代表圖：

(一)、本案指定代表圖為：第 4 圖。

(二)、本代表圖之元件代表符號簡單說明：

400 系統

410 概念網路建立器

402 查詢登錄

412 超級單元種子模組

404 查詢處理引擎

414 超級單元建立器

408 概念網路

416 外部種子來源

捌、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

無