

(由本局填寫)

承辦人代碼：
大 類：
I P C 分類：

A6
B6

本案已向：

國(地區) 申請專利，申請日期： 案號： ，☐有 ☐無主張優先權
美 國 1998年4月10日 09/058,635 ☐有 ☐無主張優先權

有關微生物已寄存於： ，寄存日期： ，寄存號碼：

(請先閱讀背面之注意事項再填寫本頁各欄)

裝

訂

線

經濟部智慧財產局員工消費合作社印製

五、發明說明(1)

發明背景

1、發明領域

概略而言，本發明係關於資訊檢索，特別係關於有效率及有效檢索超文字鏈結文件例如全球資訊網之方法及裝置。

2、相關技術之說明

廣域電腦網路稱作網際網路，特別部分網際網路稱作全球資訊網(WWW)，提供使用者存取大量資訊。並不意外地，若干搜尋引擎可供使用者輸入查詢，搜尋引擎使用多種方案架構響應查詢而送返網站明細，俾輔助由網路挖掘資訊。此等網站通常表示電腦儲存文件，使用者可存取該文件而獲得有該關特定網站主題之相關資訊。

典型地，類似大半電腦搜尋方法，網路搜尋引擎使用某種形式之關鍵字搜尋策略，其中使用者輸入的查詢項以某種方式匹配網路文件的各項，而送返相關網站明細至查詢的使用者。但大半查詢的長度僅有一至三字，因此查詢範圍極寬廣。如此表示相當大量網址含有一或多個查詢字，若搜尋引擎送返全部可能的候選文件項，則使用者可能需要篩選數百個且甚至數千個文件。

此外，響應一查詢，該查詢最相關之網址可能絲毫也未送返。特別，一查詢可能使用未出現於最貼切該查詢之網站之文件項。例如”瀏覽器”一項絲毫也未出現於目前最普及的二瀏覽器網址。反而該等網址使用”瀏覽器”以外的字來表示網址的主題。結果當使用者輸入”瀏覽器”至搜尋

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明(2)

引擎，而該引擎係使用簡單關鍵字搜尋策略時，此等網址可能絲毫也不送返。

但如本發明瞭解，網際網路使用者不知不覺地協力合作於搜尋讀取、綜覽及判定網路文件的品質。此種協力合作大半反映於網頁的編譯上，許多(即使非大半)網頁典型地描述且指向而鏈結其他高品質網頁。

更特別，一網頁係以超鏈結形式指向其他網頁，其主要係於第一文件(亦即第一網頁)參考項至其他文件(亦即其他網頁)。超鏈結對使用者提供利用電腦滑鼠或其他指標裝置及點選裝置，藉由「點選」超鏈結而可即刻選擇存取另一網頁。如此可知此種參考項網頁可能為該被參考項網頁普遍關聯之項之豐富文件項來源，即使被參考項網頁本身未使用該等文件項亦如此。結果，此等文件項可用以改進網路搜尋查詢結果。本發明進一步瞭解有效擴散傳播物件(以項形式)經由參考項至文件(例如超鏈結)之原理不僅可應用至網路，同時也可應用至任何被鏈結的文件實體，例如專利案、學術報告、文章、書籍、電子郵件等。

如此本發明之一目的係提供一種擴散可傳播物件經由超鏈結用之方法及系統。本發明之另一目的係提供響應一查詢，於一文件集合排序文件之方法及系統。本發明之又另一目的係提供一種於一文件集合找出關鍵字之方法及系統。本發明之又另一目的係提供一種於電腦儲存文件，介於文件項與查詢主題(由一或多查詢項代表的查詢主題)間找出關聯之方法及系統。本發明之另一目的係提供一種使

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明(3)

用容易且具成本效益之網路搜尋方法及系統。

發明概述

本發明為一種根據此處所述發明步驟之通用用途電腦，經程式規劃成根據本發明步驟而響應一查詢，排序一文件集合的文件。本發明也具體表現為一製造物件，亦即一機器組件，該製造物件係由數位處理裝置使用而具體表現由該數位處理裝置執行之指令程式，於電腦儲存文件中找出介於文件項與查詢主題間之關聯。本發明可實現一臨界機器組件，該組件可讓一數位處理裝置執行此處所述本發明之方法步驟。

根據本發明，電腦包括可讀程式碼裝置用於識別於第一文件至第二文件的參考項。可讀程式碼裝置接收一辭彙距離，一辭彙距離定義文件項的數目。又電腦包括可讀程式碼裝置用於接收一查詢，該查詢包括一或多個查詢項，以及可讀程式碼裝置用於決定至少一查詢項出現在第一文件的次數，而該查詢項係在參考項至第二文件之辭彙距離以內，基於如此而排序各文件。

一具體實施例中，文件可透過廣域電腦網路存取，而參考項包括一統一資源明細(URL)。若有所需，可基於查詢而建立辭彙距離。

較佳地，一電腦可包括一可讀程式碼裝置，此可讀程式碼裝置可利用文件包括被查詢項的數目來排序文件，而這些查詢項必須在參考項之辭彙距離內。此外，電腦包括可讀程式碼裝置用於接收一文件集合「U」。可讀程式碼裝

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明(4)

置係用於對集合「U」的至少一試驗文件「u」，定義集合「U」文件包括至少一試驗文件「u」參考項的文件作為鄰近文件「N(u)」。此外，可讀程式碼裝置對至少一鄰近文件「N(u)」中之至少一文件項，決定該至少一文件項是否位在鄰近文件「N(u)」至試驗文件「u」參考項的預定距離以內(亦即在預定項數以內)。遵照本發明，可讀程式碼裝置隨後響應該裝置，輸出一信號，用於決定是否至少有一文件項係在參考項的預定距離以內。當該至少一文件項係在參考項至試驗文件「u」的預定距離以內時，輸出裝置遞增關聯該至少一文件項之計數器。

除了前文摘述的邏輯外，電腦也包括可讀程式碼裝置，用於響應包括查詢接收一文件集合「U」，各文件含有一或多個文件項，而該查詢含有一或多個查詢項。當第一文件項及第一文件的參考項都在查詢的預定距離以內時，可讀程式碼裝置係供定義介於至少一第一文件與至少一第一文件項間之交互關係。若有所需，該交互關係關聯一權值，權值係基於該第一文件項及第一文件的參考項係在文件集合「U」的查詢項的預定距離以內的次數。

另一方面，一電腦程式規劃裝置包括一電腦程式儲存裝置，其可由數位處理裝置讀取，及一於程式儲存裝置上的程式規劃裝置，且包括可由數位處理裝置執行的指令，該指令係用於執行於一文件集合中找出關鍵字之方法步驟。方法步驟包括：接收文件集合，然後決定該文件集合中之參考文件及被參考文件，參考文件是在文件集合中含

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明(5)

有被參考文件的參考項之文件。對一參考文件之各文件項，決定該文件項出現於被參考文件的參考項的預定距離以內的次數。文件中至少有部分文件項係基於個別次數排序。本發明也揭示一種電腦，用於執行前文摘述之程式規劃裝置。

又另一方面，本發明揭示一種電腦執行方法，用於於電腦儲存文件中，找出介於文件項與(由一或多查詢項代表之)查詢主題間之關聯。根據本發明，文件各有文件名稱。該方法包括響應查詢項，接收至少一文件明細，然後，當文件項及文件名稱都出現在查詢項的預定距離以內的文件時，輸出一信號，表示該文件項與查詢主題間有關聯。本發明也揭示一種執行前述方法之電腦。

另一方面，一電腦包括一資料儲存裝置，而資料儲存裝置又包括一電腦可使用媒體，其具有電腦可使用程式碼裝置，該裝置用以響應一查詢，排序一文件集合的文件。電腦可使用程式碼裝置具有可讀程式碼裝置用於接收一文件集合「U」，及可讀程式碼裝置用於對集合「U」中之至少一試驗文件「u」，定義包括試驗文件「u」的至少一參考項的集合「U」的文件，作為鄰近文件「N(u)」。此外，可讀程式碼裝置對於至少一鄰近文件「N(u)」中之至少一文件項，決定該文件項是否在鄰近文件「N(u)」至試驗文件「u」的參考項的預定距離以內。然後可讀程式碼裝置響應該決定裝置輸出一信號。

有關本發明之結構及操作細節。經由參考項附圖將最

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明(6)

為明瞭，附圖中類似的參考編號表示類似部件，附圖中：

圖式之簡單說明

圖 1 為本電腦系統用於經由超鏈結擴散可傳播文件特徵之示意圖；

圖 2 為一電腦程式產品之示意圖；

圖 3 為生長一網址明細之邏輯之流程圖，該網址明細已經響應一查詢而被提供；

圖 4 為由響應一查詢而產生的網頁明細中，送返「高品質」頁之邏輯之流程圖；

圖 5 為流程圖顯示經由超鏈結找出描述項(此處也稱作可傳播特徵)之邏輯；以及

圖 6 為流程圖顯示於電腦儲存文件中，找出介於文件項與(由一或多查詢項代表之)查詢主題間之關聯之邏輯。

較佳具體實施例之詳細說明

初步參考項圖 1，顯示經由超鏈結找到描述項之系統，系統概略標示為 10。所示特定架構中，系統 10 包括一數位處理裝置，如電腦 12。一具體實施例中，電腦 12 可為紐約安莫克國際商務機器公司(IBM)製造的個人電腦(如圖所示)，或電腦 12 可為任一種電腦，包括以商品名如 AS400 伴隨 IBM 網站出售的電腦。或者電腦 12 可為 Unix 電腦，或 OS/2 伺服器，或視窗 NT 伺服器，或 IBM RS/6000 250 工作站附有 128 MB 主記憶體執行 AIX 3.2.5.，或 IBM 膝上型電腦。

電腦 12 存取網際網路搜尋引擎 14。一具體實施例中，

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明(7)

搜尋引擎 14 為 Alta Vista 製造，但需瞭解也可使用其他搜尋引擎。搜尋引擎 14 接收來自電腦 12 之查詢項，及響應查詢送返一串電腦儲存文件明細給電腦 12，更特別是送返一串網站 16 明細，電腦 12 可透過稱作全球資訊網 18 的網際網路與該等網站通訊。

此外，電腦 12 包括一特徵擴散器模組 19，其可由處理器於電腦 12 內部執行作為一系列電腦可執行指令。此等指令例如駐在電腦 12 之 RAM。

流程圖舉例說明由本發明模組 19 從事的程式規劃指令結構，如於電腦程式規劃軟體具體表現。業界人士瞭解根據本發明發揮功能之邏輯元件結構，例如電腦程式碼元件或電子邏輯電路之流程圖。本發明之具體實施例係藉一機器組件實施，該機器組件可使邏輯元件指令一數位處理裝置(亦即電腦)執行對應於所示之功能步驟的步驟。

換言之，模組 19 可一電腦程式，該程式由電腦 12 處理器執行為一系列電腦可執行指令。

另外，指令可使用資料儲存裝置含於電腦可讀媒體，例如圖 2 所示之電腦磁碟 20。磁碟 20 包括一電腦可使用媒體 22，其係以電子方式儲存電腦可讀取程式碼元件 A-D。另外，指令可儲存於 DASD 陣列、磁帶、習知硬碟機、電子唯讀記憶體、光學儲存裝置、或其他適當資料儲存裝置。本發明之一具體實施例中，電腦可執行指令可為多行經過編譯的 C⁺⁺ 相容碼或超文字標記語言(HTML)相容碼。

圖 1 也顯示系統 10 包括業界已知之周邊電腦設備，包

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明 (8)

括輸入裝置如電腦鍵盤 24 及/或電腦滑鼠 25。也可使用如圖所示以外之輸入裝置，例如軌跡球、鍵台、觸控螢幕、及語音辨識裝置。也設置輸出裝置如視訊監視器 26。也可使用其他輸出裝置如印表機、其他電腦等。

現在參考項圖 3 有關模組 19 執行之第一程序(此處稱作"程序 A")邏輯。始於方塊 28，接收到一使用者查詢，該查詢可係使用鍵盤 24 輸入。使用者查詢係由一或多查詢項組成，例如「高山」。

前進至方塊 30，建立辭彙距離「1」。較佳具體實施例中，辭彙距離「1」係以整數項數界定一視窗(容後詳述)。辭彙距離「1」有固定值，或另外，辭彙距離「1」值可基於查詢項之項數建立。例如辭彙距離「1」值係與查詢項數成反比。

於方塊 32，查詢項送至搜尋引擎 14。根據搜尋引擎原理，搜尋引擎 14 送返符合查詢項之網站 16 明細。該明細係呈結果集合「R」送返，結果集合「R」係於方塊 34 接收。典型地，結果集合係以網站名稱明細被送返，網站名稱明細稱作統一資源定位器或稱 URL。

前進至方塊 36，邏輯擴張結果集合「R」如後。首先結果集合「R」中存在有超鏈結至一或多個元件「r」之全部「s」URL 添加至結果集合「R」。如此於方塊 36 於第一文件識別第二文件的參考項。

其次，於方塊 38，全部「t」URL 添加至結果集合「R」，「t」URL 具有特徵為一種 URL 對其存在有來自

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明 (9)

結果集合「R」之任何元件「r」的超鏈結，被擴大的集合標示為「S」。如此，於方塊 36 及 38，結果集合「R」藉由增添 URL 至結果集合「R」而生長成為擴大集合「S」，該增添的 URL 係於「R」由 URL 之超鏈結指向者，或利用超鏈結指向 R 之一 URL 者。

由方塊 38，邏輯前進至方塊 40 而進入對擴大集合「S」之各文件之「DO」迴路。於決策菱形 42，決定任何查詢項是否出現在文件的 URL「u」之辭彙距離以內，亦即是否任何查詢項出現於超鏈結至擴大集合「S」之第 u 個文件之辭彙距離以內的接受測試文件。若是，則於方塊 44 關聯擴大集合之第 u 個文件之計數器_u被遞增 1，然後於方塊 46 檢索次一文件。如此邏輯決定至少一查詢存在於第二文件的參考項之辭彙距離以內的第一文件之次數，用於基於此結果排序文件，容後詳述。

若於決策菱形 42 之測試為否定，則邏輯直接前進至方塊 46。由方塊 46，邏輯前進至決策菱形 48 而決定「DO」迴路是否完成；若否，則邏輯迴路返回決策菱形 42。它方面，當「DO」迴路完成時，處理前進至方塊 50 而以計數器值遞減順序送返已經排序後的 URL 集合。

現在參考項圖 4，瞭解程序「B」，程序「B」尋求基於某些項之意義重新排序由程序「A」送返的最高「N」URL。始於方塊 52，接收一文件集合。此集合例如為於方塊 50 輸出的最高「N」(例如 20)個 URL。該集合進入「DO」迴路，及於方塊 54 設定索引變數「v」等於接受測

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明 (10)

試之 URL。

前進至方塊 56，決定全部(或一子集的)URL「U」其引述接受測試的 URL「v」(例如經由含有超鏈結至接受測試的 URL「v」)。其次，前進至方塊 58，檢索於 URL 中有關超鏈結至 URL「v」之全部錨文字。

“錨文字”一詞表示直接關聯一文件之超鏈結或其他參考項或引述的文件內部。例如於”最早高能核子加速器之一係建造於<AHREF=「http://www.CERN.ch」>CERN，歐洲粒子物理學實驗室 ”，一句中，超鏈結為「http://www.CERN.ch」片語，及錨文字為”<A>...”對中包括的內容。採用本例，對例如於錨文字之五項辭彙距離以內之項為「核能加速器係建造於」；而非屬於錨文字之辭彙距離以內者為「最早高能之一」。

然後於方塊 60 對查詢項進入巢套「DO」迴路。前進至決策菱形 62，決定接受測試之文件中接受測試之查詢項頻率是否大於錨文字之某個參考項集合中之參考項頻率，可藉多種習知統計技術之一決定。

當接受測試文件中之接受測試查詢項頻率係高於參考項頻率時，處理前進至方塊 64，而旗標標示接受測試文件為有意義。否則，接受測試文件未被旗標為有意義。任一例中，各文件可關聯一計數器或其他如上測試表示其意義之值。前述「DO」迴路結束時，頂端「N」個 URL 利用其意義排序。

現在參考項圖 5，顯示經由超鏈結尋找描述項之程序

五、發明說明 (11)

「C」邏輯。始於方塊 68，接收 URL 「U」之集合「U」，對集合「U」之個別 URL 「U」進入「DO」迴路。於方塊 70，決定接受測試之 URL 「U」之鄰近 N(u) 集合。「鄰近」一詞表示於 URL 集合「U」中含有超鏈結至接受測試文件「U」之文件。換言之，鄰近集合 N(u) 可視為參考文件至被參考文件「U」之集合。

於方塊 72 對鄰近集合 N(u) 之各個元件 (亦即文件項) v 進入巢套「DO」迴路。前進至方塊 74，一計數器 v 關聯鄰近集合 N(u) 之各項 v。然後進入雙重巢套「DO」迴路。前進至決策菱形 76，決定接受測試之項 v 是否在 (例如超鏈結) 至接受測試文件「U」的參考項的預定距離以內。此種預定距離可為前述辭彙距離。若接受測試項 v 係在接受測試文件「U」的參考項的預定距離以內，則該項之計數器 v 於方塊 78 遞增 1。否則計數器未遞增。當對文件集合「U」之全部文件「U」之鄰近集合 N(u) 中之全部鄰近項皆已經如前述測試時，邏輯前進至方塊 80，而藉由其個別計數器值篩選各項，及送返篩選後的明細。

如本發明，方塊 80 之輸出為文件集合「U」各項之排序明細。此而排序明細可使用於提示額外查詢項給使用者。又可使用於關聯遨遊中資料庫。此外，方塊 80 之輸出可使用於註解被超鏈結的文件叢集及項叢集作為許多搜尋引擎之後處理步驟。

圖 6 顯示程序「D」之邏輯，用於於電腦儲存文件中找出介於文件項與由一或多個查詢項代表之查詢主題間之

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明 (12)

關聯。始於方塊 82，接收一查詢「Q」。查詢項「Q」係由一或多查詢項「q」組成。

查詢項於方塊 84 發送至搜尋引擎，響應此查詢項，接收由搜尋引擎送返之一文件明細。前進至方塊 86，構成二分圖 $G=((T,U)E)$ ，其頂點項(T)及文件(U)係於方塊 84 送返，其中 T 及 U 分別表示二分圖之文件項分支及 URL 分支，及其中 E 表示分支間之邊。

前進至方塊 88，對各文件進入「DO」迴路。前進至方塊 90，對文件掃描 URL「U」及查詢項「q」。其次前進至方塊 92，對各文件項「t」及 URL「U」找出屬於查詢項「q」的預定距離以內者，進入「DO」迴路，其中於方塊 94，邊 $(t,u) \in E$ 之權值遞增 1。使用此種邏輯，當一文件項及文件名稱或引述(呈超鏈結形式)皆出現於查詢項的預定距離以內之一文件時，輸出一信號代表該文件項與查詢主題間之關聯。

若有所需，「DO」迴路包括前進至方塊 96，其中對矩陣 A 決定單一值分解(SVD)，矩陣 A 係由邊 $E: a_{ij}$ 定義，其中 a_{ij} 為邊由第 i 項至第 j URL 之權值。如業界眾所周知，SVD 於方塊 96 有效決定因子 $A=UT\Sigma V$ ，此處 Σ 為含有 A 單數值之對角矩陣，U 及 V 為執行正交轉換的正交矩陣。業界稱作延遲語意學索引(LSI)之技術例如美國專利 4,839,853(併述於此以供參考)揭示之技術可使用於前處理主體，特別用於將文件項矩陣 A 處理成 $U\Sigma VT$ ，此處 U 表示由項空間至所謂之 LSI 或構想空間的線性投影。數百 LSI 大

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

五、發明說明 (13)

小「k」即足。

但 LSI 搜尋未處理 U 矩陣，本發明確實使用 U 矩陣如後。各項映射至 LSI 空間，各文件以 k 維向量序列表示。查詢項本身轉變成此種向量之短序列。然後文件被掃描，及邏輯嘗試匹配查詢項向量與文件之向量小窗。若存在有低成本(亦即良好)匹配，則大量投向附近引述，亦即超鏈結。成本可使用最低成本匹配策略評估，此處匹配對應項 t_1 及 t_2 之向量之邊緣成本單純為其於 U 之投影間距。舉例言之，查詢項「汽車製造商」可以低成本匹配至文字「汽車製造公司」序列，支持接近此等類似片語的引述。

與 LSI 相反，本發明對各文件保持一序列 LSI 向量。換言之，本發明不似 LSI，本發明考慮匹配 LSI 向量序列及使用分數來支持鄰近引述。

若有所需，處理於方塊 98 送返提示搜尋項給使用者。為了決定此等被提示項，邏輯篩選於方塊 96 決定投影於 SVD 左向量(亦即「U」第一欄)之項俾便減低值。被篩選明細之最上方「k」項隨後於方塊 98 送返，其中「k」為預定整數，例如 5。

雖然此處顯示且詳細說明之特定經由超鏈結之特徵傳播全然可達成前述本發明目的，但需瞭解此乃目前較佳本發明之具體實施例，如此表示可由本發明廣義涵蓋之主題，本發明範圍全然涵蓋業界人士顯然易知之其他具體實施例，如此本發明範圍除了隨附之申請專利範圍之外未受任何限制。

四、中文發明摘要(發明之名稱：經由超鏈結之特徵傳播)

一種響應一查詢，藉普遍性而排序廣域電腦網路(例如網路)網頁之系統及方法。進一步，使用一查詢及其來自搜尋引擎的回應，該系統及方法找出額外關鍵字其可能為良好延伸的搜尋項，特別於查詢時間於掃過時產生一局部資料庫。

英文發明摘要(發明之名稱：FEATURE DIFFUSION ACROSS HYPERLINKS)

A system and method for ranking wide area computer network (e.g., Web) pages by popularity in response to a query. Further, using a query and the response thereto from a search engine, the system and method finds additional key words that might be good extended search terms, essentially generating a local thesaurus on the fly at query time.

六、申請專利範圍

1. 一種電腦，其包括一資料儲存裝置其包括一電腦可使用媒體具有電腦可使用程式碼裝置，用於響應一查詢排序一文件集合的文件，該電腦可使用程式碼裝置具有：

識別用之電腦可讀程式碼裝置，用於識別於第一文件第二文件的參考項；

接收用之電腦可讀程式碼裝置，其係用於接收一辭彙距離，該辭彙距離定義多項文件項；

接收用之電腦可讀程式碼裝置，用於接收包括一或多查詢項之一查詢；及

決定用之電腦可讀程式碼裝置，其係用於決定至少一查詢出現於第一文件內於第二文件的參考項之辭彙距離以內的次數，用於基於此而排序各文件。

2. 如申請專利範圍第1項之電腦，其中該等文件可透過廣域電腦網路存取，及該參考項包括一統一資源明細(URL)。
3. 如申請專利範圍第2項之電腦，其中該辭彙距離係基於查詢項建立。
4. 如申請專利範圍第2項之電腦，其進一步包含排序用之電腦可讀程式碼裝置，用於基於查詢項存在於文件的參考項的辭彙距離以內之次數而排序多個文件。
5. 如申請專利範圍第2項之電腦，其進一步包含：

接收用之電腦可讀程式碼裝置，用於接收一文件集合「U」；

定義用之電腦可讀程式碼裝置，其係用於對集合「U」中之至少一試驗文件「u」定義作為集合「U」中的鄰近

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

六、申請專利範圍

文件「N(u)」，該鄰近文件「N(u)」包括試驗文件「u」的至少一參考項；

決定用之電腦可讀程式碼裝置，其係用於對至少一鄰近文件「N(u)」之至少一文件項，決定是否該至少一文件項係在鄰近文件「N(u)」中試驗文件「u」的參考項的預定距離以內；及

輸出用之電腦可讀程式碼裝置，其係用於響應裝置決定該至少一文件項是否在一參考項的預定距離以內而輸出一信號。

6. 如申請專利範圍第5項之電腦，其中該輸出裝置於該至少一文件項係在測試文件「U」的參考項的預定距離以內時，遞增關聯該至少一文件項之計數器。

7. 如申請專利範圍第2項之電腦，其進一步包含：

接收用之電腦可讀程式碼裝置，其係用於響應包括一或多查詢項之一查詢接收一文件集合「U」，各文件含有一或多文件項；及

定義用之電腦可讀程式碼裝置，其係用於當該文件項及第一文件的參考項係在該等文件中之至少一者的一查詢的預定距離以內時，定義至少一第一文件與至少一文件項間之交互關係。

8. 如申請專利範圍第7項之電腦，其中該交互關係係關聯一權值，而該權值係基於該文件項及第一文件的參考項屬於文件集合「U」中之一查詢預定距離以內之次數。

9. 一種電腦程式規劃裝置，及包含：

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

六、申請專利範圍

一 電腦程式規劃儲存裝置，其可由一數位處理裝置讀取；及

一 電腦規劃裝置於程式儲存裝置及包括可由數位處理裝置執行指令，該指令係用於執行於一文件集合中找出關鍵字之方法步驟，該等方法步驟包含：

接收該等文件集合；

測定於該文件集合中之參考文件及被參考文件，參考文件為於含有被參考文件集合的參考項的文件；

對一參考文件之各文件項，決定該文件項出現於被參考文件的參考項的預定距離以內之次數；及

基於個別次數排序文件中之至少部分文件項。

10. 如申請專利範圍第9項之電腦程式規劃裝置，其中該等方法步驟進一步包含每次一文件項出現於被參考文件的參考項的預定距離以內時，遞增個別計數器。

11. 如申請專利範圍第9項之電腦程式規劃裝置，其中該等方法步驟進一步包含：

響應一查詢接收一文件集合「U」，該查詢包括一或多查詢項；及

當該文件項與第一文件的參考項係屬一查詢之預定範圍以內時，定義至少第一文件與該至少一文件項間之交互關係。

12. 如申請專利範圍第11項之電腦程式規劃裝置，其中該交互關係關聯一權值，及該權值係基於該文件項及第一文件的參考項係屬文件集合「U」中之一查詢的預定距離

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

東

六、申請專利範圍

以內的次數。

13. 如申請專利範圍第9項之電腦程式規劃裝置，其中該等方法步驟進一步包含：

接收一辭彙距離，該辭彙距離定義多項文件項；

接收一查詢包括一或多查詢項；及

決定其中至少一查詢出現於第二文件的參考項之辭彙距離以內之第一文件之次數，用於基於此次數排序該等文件。

14. 如申請專利範圍第13項之電腦程式規劃裝置，其中該等文件係可透過廣域電腦網路存取，且該參考項包括一統一資源明細(URL)。

15. 如申請專利範圍第14項之電腦程式規劃裝置，其中該辭彙距離係基於查詢項建立。

16. 如申請專利範圍第14項之電腦程式規劃裝置，其中該等方法步驟進一步包含基於查詢項出現於文件中參考項的辭彙距離以內之個別次數排序多文件。

17. 一種電腦執行方法，其係用於找出於電腦儲存文件中介於文件項與一或多查詢項表示之查詢主題間之關聯，該等文件具有個別文件名稱，該方法包含下列步驟：

響應查詢項接收至少一文件明細；及

當一文件項及一文件名稱皆出現於一文件內位在一查詢的預定距離以內時，輸出一信號代表該文件項與查詢主題間之關聯。

18. 如申請專利範圍第17項之電腦執行方法，其中該輸

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

象

六、申請專利範圍

出步驟包含：

構成一張二分圖具有多個頂點，頂點表示於各步驟接收之文件；及

對至少若干文件(u)，及對各文件(u)中之至少若干文件項(t)，當文件項(t)及文件(u)之文件名稱皆出現於一查詢的預定距離以內時，遞增邊(t,u)於二分圖之權值，邊(t,u)表示文件(u)及文件項(t)。

19. 如申請專利範圍第17項之電腦~~程式~~執行方法，其進一步包含下列步驟：

決定於文件明細及參考文件及~~被~~參考文件，參考文件乃明細中含有被參考文件的參考項之文件；

對一參考文件中之各文件項，決定該文件項出現於被參考文件的參考項的預定距離以內之次數；及

基於個別次數排序文件中之至少若干文件項。

20. 如申請專利範圍第17項之電腦~~程式~~執行方法，其進一步包含下列步驟：

接收一辭彙距離，該辭彙距離定義多個文件項；

決定其中至少一查詢出現於第一文件內位在第二文件的參考項之辭彙距離以內之次數用於基於該次數而排序各文件。

21. 如申請專利範圍第20項之電腦~~程式~~執行方法，其中該等文件係可透過廣域電腦網路存取，及該參考項包括一統一資源明細(URL)。

22. 如申請專利範圍第20項之電腦~~程式~~執行方法，其中該辭

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

六、申請專利範圍

彙距離係基於查詢項建立。

23. 如申請專利範圍第20項之電腦，其進一步包含基於查詢項存在於文件中參考項的辭彙距離以內之個別次數排序多個文件之步驟。

24. 一種電腦，其包括一資料儲存裝置包括一電腦可使用媒體，其具有電腦可使用程式碼裝置用於於電腦儲存文件中找出介於文件項與由一或多查詢項表示之查詢主題間之關聯，該等文件具有個別文件名稱，該電腦可使用程式碼裝置具有；

接收用之電腦可讀程式碼裝置，用於響應該查詢接收至少一文件明細；及

輸出用之電腦可讀程式碼裝置，用於一文件項及一文件名稱皆出現於一查詢的預定距離以內之文件時，輸出代表該查詢與該查詢主題間之關聯之信號。

25. 如申請專利範圍第24項之電腦，其中該輸出裝置包含：

構成用之電腦可讀程式碼裝置，其係用於構成一二分圖具有多個頂點，頂點表示於各步驟接收之文件；及

增量用之電腦可讀程式碼裝置，其係用於對至少若干文件(u)，及對各文件(u)中之至少若干文件項(t)，當文件項(t)及文件(u)之文件名稱皆出現於一查詢的預定距離以內時，遞增邊(t,u)於二分圖之權值，邊(t,u)表示文件(u)及文件項(t)。

26. 如申請專利範圍第25項之電腦，其進一步包含：

決定用之電腦可讀程式碼裝置，其係用於文件明細及

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

泉

六、申請專利範圍

參考文件及被參考文件，參考文件乃明細中含有被參考文件的參考項之文件；

決定用之電腦可讀程式碼裝置，其係用於對一參考文件中之各文件項，決定該文件項出現於被參考文件的參考項的預定距離以內之次數；及

排序用之電腦可讀程式碼裝置，其係用於基於個別次數排序文件中之至少若干文件項。

27. 如申請專利範圍第26項之電腦，其進一步包含：

接收用之電腦可讀程式碼裝置，其係用於接收一辭彙距離，該辭彙距離定義多個文件項；

決定用之電腦可讀程式碼裝置，其係用於決定其中至少一查詢出現於第一文件內位在第二文件的參考項之辭彙距離以內之次數用於基於該次數而排序各文件。

28. 如申請專利範圍第27項之電腦，其中該等文件係可透過廣域電腦網路存取，及該參考項包括一統一資源明細(URL)。

29. 一種電腦，其包括一資料儲存裝置包括一電腦可使用媒體，其具有電腦可使用程式碼裝置用於找出一文件集合之關鍵字，該電腦可使用程式碼裝置具有：

接收用之電腦可讀程式碼裝置，其係用於接收該等文件集合；

決定用之電腦可讀程式碼裝置，其係用於測定於該文件集合中之參考文件及被參考文件，參考文件為於含有被參考文件集合的參考項的文件；

(請先閱讀背面之注意事項再填寫本頁)

裝

訂

線

六、申請專利範圍

決定用之電腦可讀程式碼裝置，其係用於對一參考文件之各文件項，決定該文件項出現於被參考文件的參考項的預定距離以內之次數；及

排序用之電腦可讀程式碼裝置，其係用於基於個別次數排序文件中之至少部分文件項。

30. 如申請專利範圍第29項之電腦，其進一步包含增量用之電腦可讀程式碼裝置用於每次當文件項出現於被參考文件的參考項的預定距離以內時遞增個別計數器。

31. 如申請專利範圍第30項之電腦，其進一步包含：

接收用之電腦可讀程式碼裝置，其係用於響應一查詢接收一文件集合「U」，該查詢包括一或多查詢項；及

定義用之電腦可讀程式碼裝置，其係用於當該文件項與第一文件的參考項係屬一查詢之預定範圍以內時，定義至少第一文件與至少一文件項間之交互關係。

32. 如申請專利範圍第31項之電腦，其中該交互關係係關聯一權值，而該權值係基於該文件項及第一文件的參考項屬於文件集合「U」中之一查詢預定距離以內之次數。

33. 如申請專利範圍第32項之電腦，其進一步包含：

接收用之電腦可讀程式碼裝置，其係用於接收一辭彙距離，該辭彙距離定義多項文件項；

接收用之電腦可讀程式碼裝置，用於接收包括一或多查詢項之一查詢；及

決定用之電腦可讀程式碼裝置，其係用於決定至少一查詢出現於第一文件內於第二文件的參考項之辭彙距離

(請先閱讀背面之注意事項再填寫本頁)

訂

泉

六、申請專利範圍

以內的次數，用於基於此而排序各文件。

34. 如申請專利範圍第33項之電腦，其中該等文件可透過廣域電腦網路存取，及該參考項包括一統一資源明細(URL)。

35. 一種電腦，包括一資料儲存裝置包含一電腦可使用媒體其具有電腦可使用程式碼裝置用於響應一查詢排序一文件集合的文件，該電腦可使用程式碼裝置具有：

接收用之電腦可讀程式碼裝置，用於接收一文件集合「U」；

定義用之電腦可讀程式碼裝置，其係用於對集合「U」中之至少一試驗文件「u」定義作為集合「U」中包括至少一次試驗文件「u」的參考項之鄰近文件「N(u)」文件；

決定用之電腦可讀程式碼裝置，其係用於對至少一鄰近文件「N(u)」之至少一文件項，決定是否該至少一文件項係在鄰近文件「N(u)」中試驗文件「u」的參考項的預定距離以內；及

輸出用之電腦可讀程式碼裝置，其係用於響應裝置決定該至少一文件項是否在一參考項的預定距離以內而輸出一信號。

36. 如申請專利範圍第35項之電腦，其中該輸出裝置於該至少一文件項係在測試文件「U」的參考項的預定距離以內時，遞增關聯該至少一文件項之計數器。

(請先閱讀背面之注意事項再填寫本頁)

訂

線

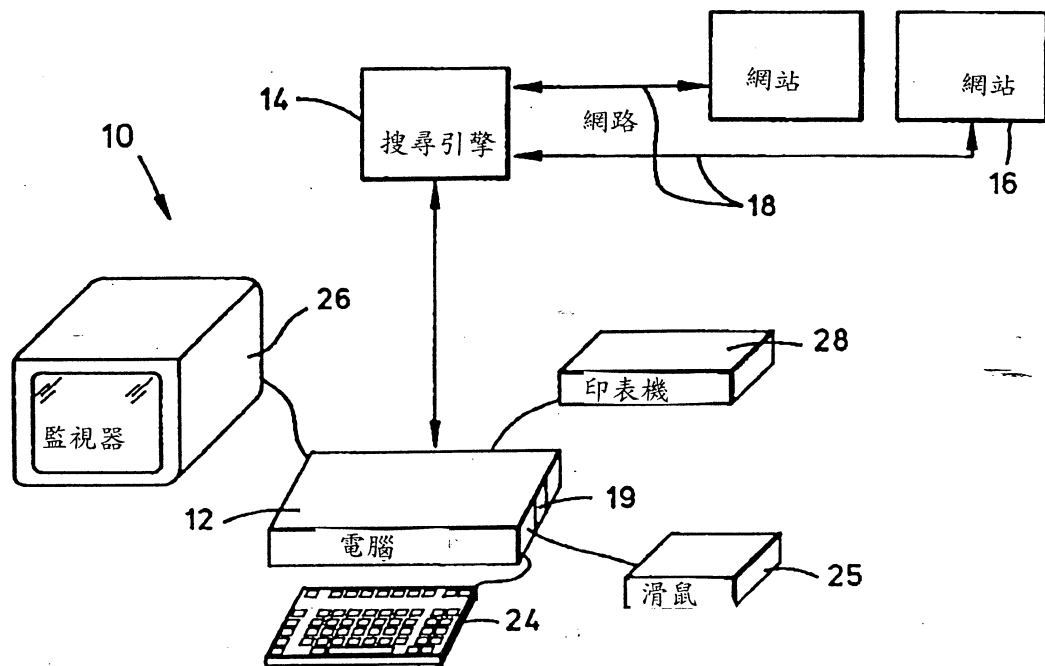


圖 1

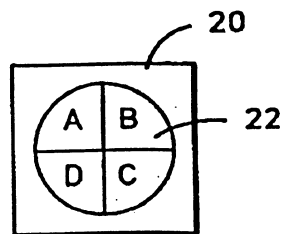
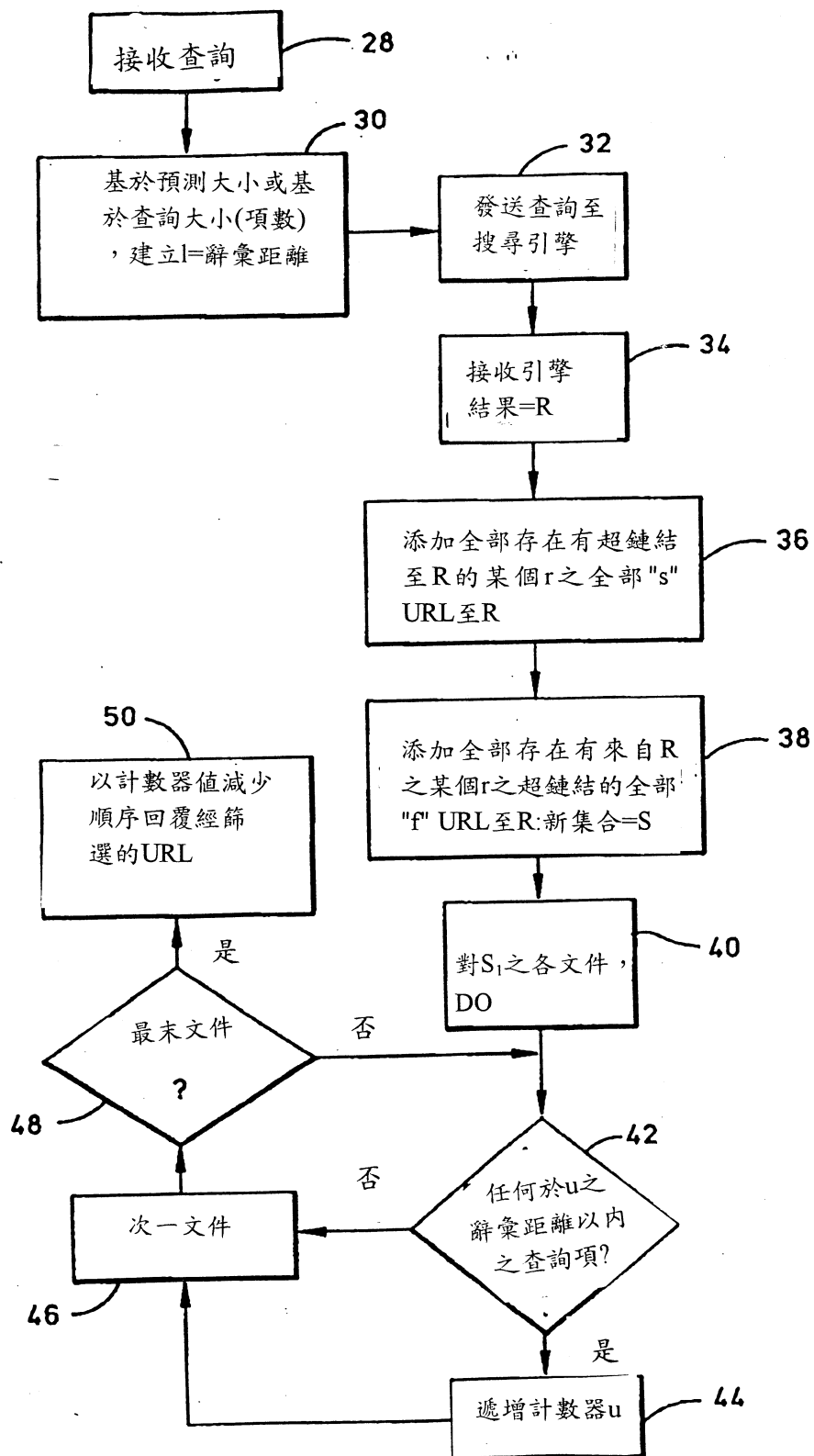


圖 2



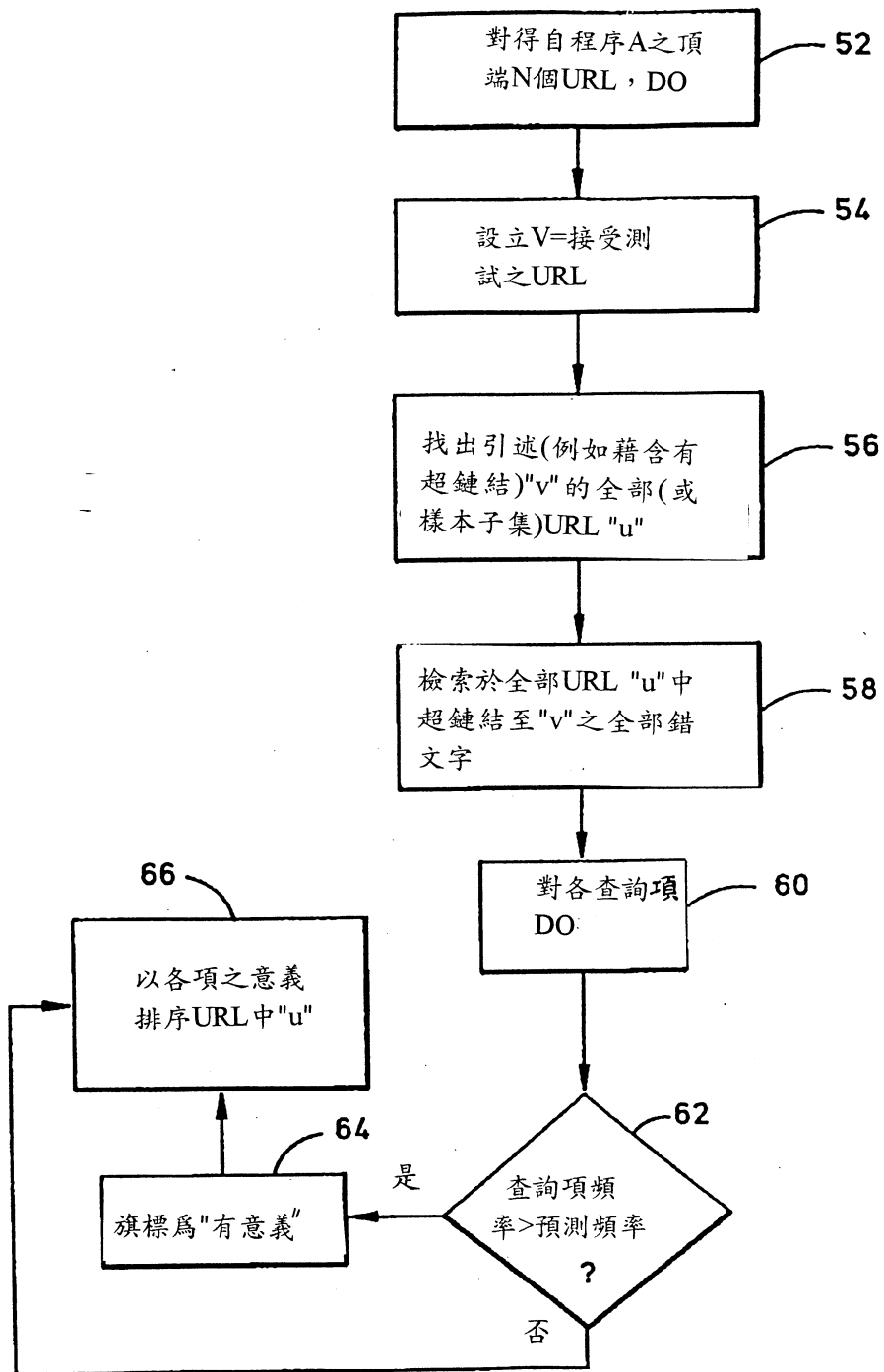


圖 4
程序B

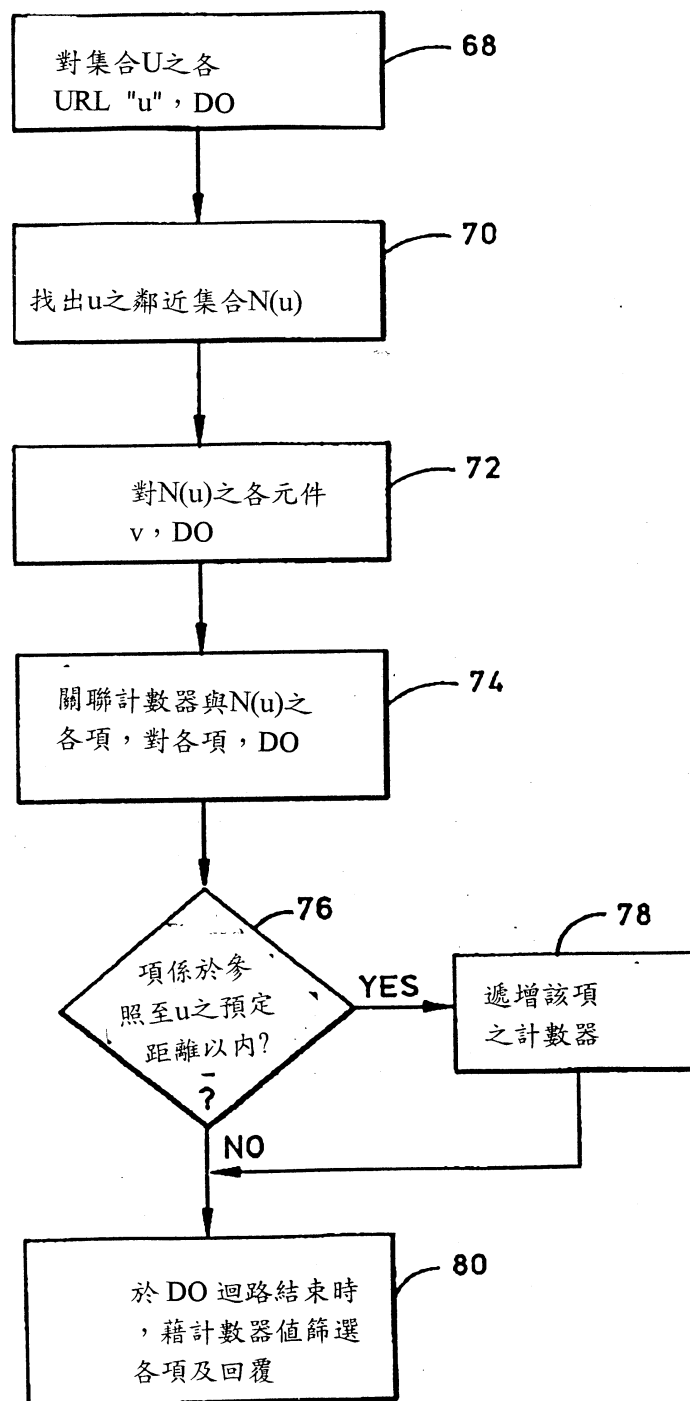


圖 5
程序C

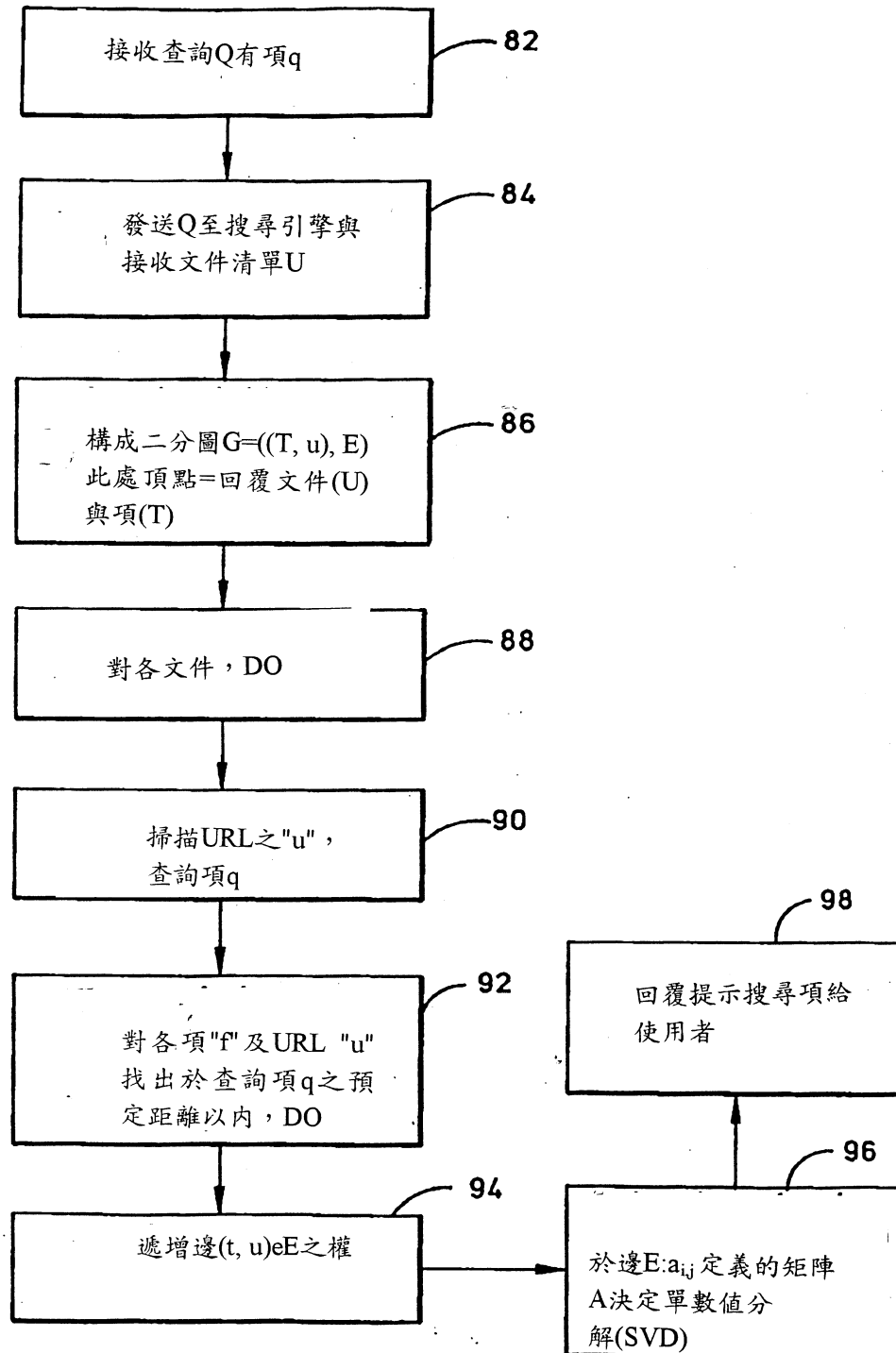


圖 6
程序D

公告本

89年9月18日修正/更正/補充

申請日期	88.02.09
案 號	88101870
類 別	G06F 17/50

A4
C4

526432

(以上各欄由本局填註)

說明書修正本(89年9月)

發明專利說明書

一、發明 新型名稱	中 文	經由超鏈結之特徵傳播
	英 文	FEATURE DIFFUSION ACROSS HYPERLINKS
二、發明 創作人	姓 名	1.索門 卡拉芭堤 2.拜朗 艾德華 鬥
	國 籍	1.印度 2.美國
	住、居所	1.美國加州喬荷西市卡圖路6174號 2.美國加州洛杉磯市希巴斯坦路111號
三、申請人	姓 名 (名稱)	美商萬國商業機器公司
	國 籍	美國
	住、居所 (事務所)	美國紐約州阿蒙市新果園路
	代 表 人 姓 名	費羅普