(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 201211798 A1
(43)公開日： 中華民國 101 (2012) 年 03 月 16 日

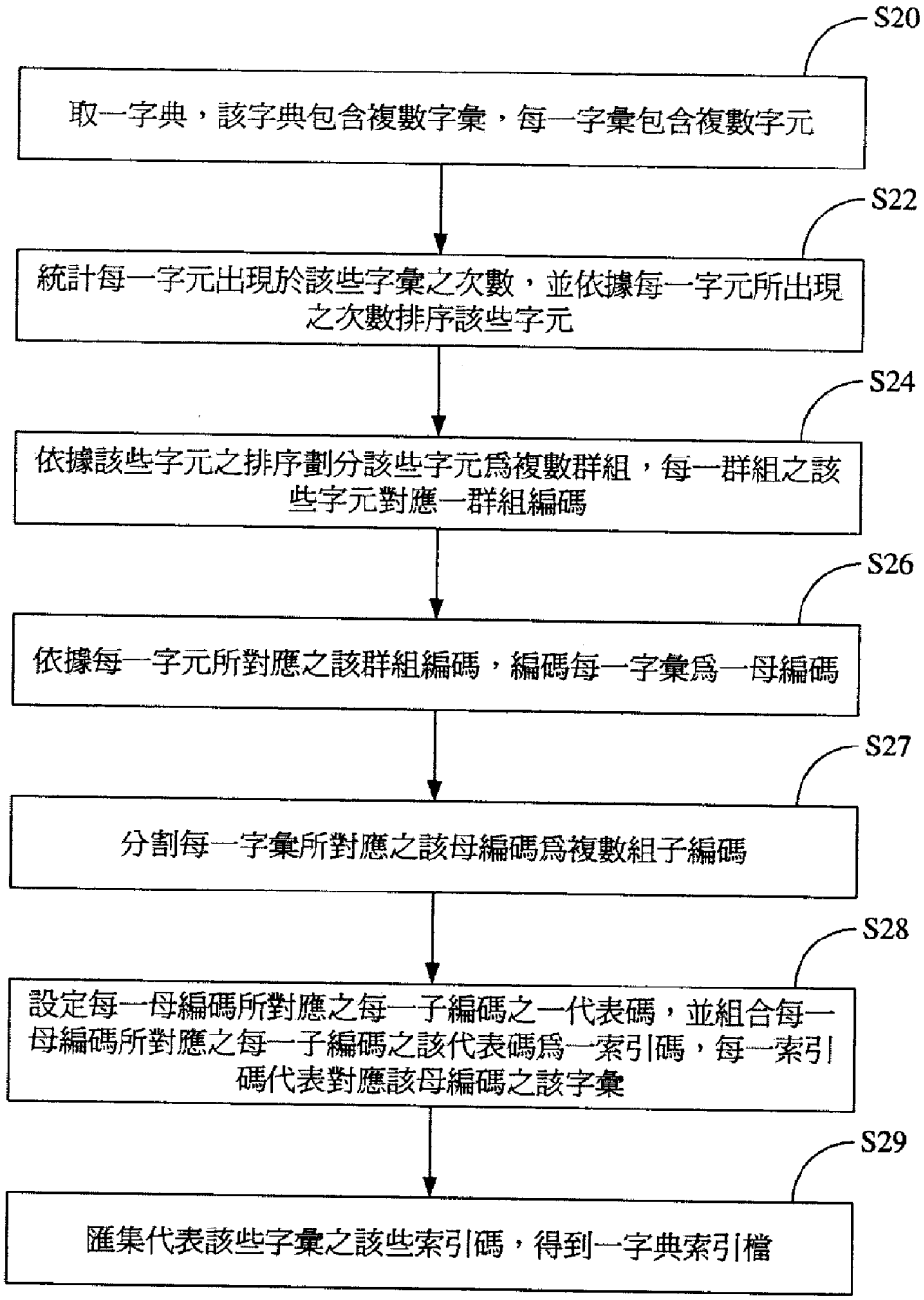
(21)申請案號：099130818 (22)申請日： 中華民國 99 (2010) 年 09 月 13 日
(51)Int. Cl.： *G06F17/30 (2006.01)*
(71)申請人：李俊達 (中華民國) (TW)
臺南市永康區中正路 529 號
(72)發明人：李俊達 (TW) ；沈肇基 (TW)
(74)代理人：蔡秀玫
申請實體審查：有 申請專利範圍項數：13 項 圖式數：7 共 26 頁

(54)名稱

字典索引之建立方法及使用方法

(57)摘要

本發明有關於供一種字典索引之建立方法及使用方法，該字典索引之建立方法主要將每一字典之每一字彙轉換成一索引碼，匯集該些索引碼得到一字典索引檔。如此可將每一種字典依據本發明所提供之字典索引之建立方法建立一字典索引檔，使用者可依據自己的需求選擇適合的字典，並下載該字典對應之該字典索引檔。然後本發明所建立之該字典索引檔之使用方法為該字典索引檔包含該些索引碼，每一索引碼對應複數字彙，當確認欲搜尋之字彙對應之該索引碼存在於該字典索引檔內時，可依據該索引碼對應之該些字彙搜尋欲搜尋之字彙，如此有效減少搜尋字彙的數量，更縮短搜尋字彙時間。


S20
取一字典，該字典包含複數字彙，每一字彙包含複數字元
S22
統計每一字元出現於該些字彙之次數，並依據每一字元所出現之次數排序該些字元
S24
依據該些字元之排序劃分該些字元爲複數群組，每一群組之該些字元對應一群組編碼
S26
依據每一字元所對應之該群組編碼，編碼每一字彙爲一母編碼
S27
分割每一字彙所對應之該母編碼爲複數組子編碼
S28
設定每一母編碼所對應之每一子編碼之一代表碼，並組合每一母編碼所對應之每一子編碼之該代表碼爲一索引碼，每一索引碼代表對應該母編碼之該字彙
S29
匯集代表該些字彙之該些索引碼，得到一字典索引檔

六、發明說明：

【發明所屬之技術領域】

[0001] 本發明係有關於一種字典索引之建立方法及使用方法，尤指一種可控制字典索引佔記憶體容量之大小及有效減少搜尋字彙之數量及時間之字典索引之建立方法及使用方法。

【先前技術】

[0002] 按，對於電腦上的電子字典來說，一般在使用者輸入單詞的過程中，會利用輸入單詞的部分內容(例如前幾個字母)來查詢單詞的相關內容，例如釋義，同時還必須快速地查詢出所輸入單詞部分內容的後續單詞，換言之，每當使用者輸入待查詢單詞的任何一個字母時，電子字典就必須執行一遍查詢動作，因此查詢動作的快慢勢必影響到使用者對於電子字典整體性能的觀感。

以英文字典為例，為了能快速檢索到使用者欲查詢之英文字彙，有許多字典編排方法被提出以建置索引，例如：可依據每個英文字彙的第一個英文字母從A~Z的順序，進行26種分類以建置索引，或可將英文字典裡的每個字彙依其字彙長度大小，進行49種分類(從字彙長度2至50)以建置索引，或同時結合英文字彙第一個字母順序及字彙長度大小進行更詳細的分類(總共會有26*49=1274種分類)，其目的是以減少每項分類所內含的英文字彙數，達到快速檢索之需求。從上述所提及之方法可看出，進行更詳細的分類雖可達到快速檢索之目的，但為了建置此分類繁雜的索引方式會耗費較多的記憶體，有時可能光建置索引所佔的記憶體就比整本字典

所佔的記憶體更多。

有鑑於上述問題，本發明提供一種字典索引之建立方法及使用方法，本發明之字典索引之建立方法及使用方法可控制子編碼之組數，決定該字典索引檔佔記憶體容量之大小。

【發明內容】

[0003] 本發明之目的，在於提供一種字典索引之建立方法及使用方法，本發明之使用方法為二階段搜尋方式，先於本發明所建立之一字典索引檔搜尋是否存在欲搜尋之一字彙對應之一索引碼，當確認欲搜尋之字彙對應之該索引碼存在於該字典索引檔內時，可依據該索引碼對應之一字典字彙檔之複數字彙搜尋欲搜尋之該字彙，如此有效減少搜尋的字彙數量，更縮短搜尋時間。

本發明之目的，在於提供一種字典索引之建立方法及使用方法，於字典索引之建立過程中可控制子編碼之組數，決定該字典索引檔佔記憶體容量之大小。

本發明之目的，在於提供一種字典索引之建立方法及使用方法，可將每一種字典依據本發明所提供之字典索引之建立方法建立一字典索引檔，使用者可依據自己的需求選擇適合的字典，並下載該字典對應之該字典索引檔。

為達到上述之目的，本發明提供一種字典索引之建立方法，係包含：取一字典之一字典字彙檔，該字典字彙檔包含複數字彙，每一字彙包含複數字元；統計每一字元出現於該些字彙之次數，並依據每一字元所出現之次數排序該些字元；依據該些字元之排序劃分該些字元

為複數群組，每一群組之該些字元對應一群組編碼；依據每一字元所對應之該群組編碼，編碼每一字彙為一母編碼；　分割每一字彙所對應之該母編碼為複數組子編碼；設定每一母編碼所對應之每一子編碼之一代表碼，並組合每一母編碼所對應之每一子編碼之該代表碼為一索引碼，每一索引碼代表對應該母編碼之該字彙；以及匯集代表該些字彙之該些索引碼，得到一字典索引檔。

本發明提供一種字典索引之使用方法，係包含：下載一字典之一字典索引檔，該字典索引檔包含複數索引碼，每一索引碼對應一字典字彙檔之複數字彙；輸入欲搜尋之一字彙；分析欲搜尋之該字彙，得到對應欲搜尋之該字彙之一索引碼；判斷該索引碼存在於該字典索引檔之該些索引碼；以及依據該索引碼對應之該字典字彙檔之該些字彙搜尋欲搜尋之該字彙。

【實施方式】

[0004]　茲為使　貴審查委員對本發明之結構特徵及所達成之功效有更進一步之瞭解與認識，謹佐以較佳之實施例及配合詳細之說明，說明如後：

本發明提供一種字典索引之建立方法及使用法，利用本發明所建立之索引取代習知字典索引，大幅改善習知字典索引之檔案佔大量記憶體之問題，並縮短搜尋字彙之數量及時間。使用本發明之建立方法對每一種字典建置其索引，得到該字典之一字典索引檔。本實施例所提供之字典索引之建立方法及使用方法，其皆透過一電腦系統內之至少一種程式進行統計、分析及演算

請參閱第一圖，係本發明之一較佳實施例之使用字

典索引檔之流程圖。如圖所示，本發明之字典索引之使用方法係先執行步驟S10，先下載一字典之一字典索引檔，該字典索引檔包含複數索引碼，每一索引碼對應一字典字彙檔之複數字彙。接著執行步驟S12，輸入欲搜尋之一字彙。然後執行步驟S14，分析欲搜尋之該字彙，得到對應欲搜尋之該字彙之一索引碼，而分析欲搜尋之該字彙依據本發明之建立方法進行分析，以得知該字彙之該索引碼。接著執行步驟S16，判斷該索引碼是否存在於該字典索引檔之該些索引碼，若該索引碼不存在於該字典索引檔時，即表示該字彙不在該字典內，不無須進行下一步驟；反之，若該索引碼存在於該字典索引檔時，即表示該字彙在該字典內，進入步驟S18，依據該索引碼對應之該字典字彙檔之該些字彙中可快速搜尋出欲搜尋之該字彙。由上述可知，本發明使用二階層檢索方法，先利用欲搜尋之該字彙對應之該索引碼判斷是否存在於該字典內，若存在於該字典內，只要再對該索引碼對應之該些字彙進行搜尋即可，有逐步縮小搜尋範圍，有效減少搜尋字彙的數量及搜尋時間。

請一併參閱第二圖，係本發明之一較佳實施例之建立字典索引檔之流程圖。如圖所示，本實施例提供一種字典索引之建立方法，其用以建立每一字典之該字典索引檔及分析欲搜尋之該字彙。該建立方法係先執行步驟S20，取一字典之該字典字彙檔，該字典字彙檔包含複數字彙，每一字彙包含複數字元，該字典可以為一語言字典，例如：英語字典、法語字典、德語字典或其他語言之字典。該字典亦可為專業字典，例如：醫學字典、化

學字典、機械字典或其他專業領域之字典，專業字典內之字彙為英語字彙、法語字彙或其他語言之字彙。該字典內之每一字彙包含該些字元，該些字元包含字母、標點符號或空格，例如：該字彙A為「abatement」，所以該字彙A所包含之該些字元皆為字母，有「a」、「b」、「t」、「e」、「m」、「n」；該字彙B為「abatement, air pollution」，所以該字彙B所包含之該些字元包含字母、標點符號及空格，該些字元有「a」、「b」、「t」、「e」、「m」、「n」、「,」、「空格」、「i」、「r」、「p」、「o」、「l」、「u」。

復參閱第二圖，接著執行步驟S22，統計每一字元出現於該些字彙之次數，並依據每一字元所出現之次數排序該些字元。請一併參閱第三圖，係本發明之另一較佳實施例之字元出現次數統計圖。如圖所示，本實施例使用醫學字典，並統計該醫學字典內之複數字彙所包含之複數字元出現於該些字彙之次數，然後依據每一字元之出現次數排序該些字元，例如：字元「e」之出現次數為202760，其為所有字元中出現次數最多之字元，字元「_」之出現次數為3，其為所有字元中出現次數最少之字元。排序該些字元之方式為從出現次數最多排至出現次數最少。

排序該些字元後，執行步驟S24，依據該些字元之排序劃分該些字元為複數群組，每一群組之該些字元對應一群組編碼。請一併參閱第四圖，係本發明之另一較佳實施例之群組編碼示意圖。如圖所示，依據第三圖所顯示之該些字元之排序，本實施例主要將該些字元劃分成

四個群組，分別為一第一群組、一第二群組、一第三群組及一第四群組，依據該些字元之排序依據將該些字元分配至該第一群組、該第二群組、該第三群組及該第四群組，例如：排序在前四個之該些字元分別為「e」、「a」、「i」、「o」，然後依序將該些字元分配至該第一群組、該第二群組、該第三群組及該第四群組，該字元「e」分配至該第一群組，該字元「a」分配至該第二群組，該字元「i」分配至該第三群組，該字元「o」分配至該第四群組，排序上述字元之後四個之該些字元分別為「r」、「s」、「t」、「n」，字元「r」分配至該第四群組，該字元「s」分配至該第三群組，該字元「t」分配至該第二群組，該字元「n」分配至該第一群組。剩下之該些字元依據上述方式分配至該第一群組、該第二群組、該第三群組及該第四群組。

由圖中可知每一群組所包含之該些字元，該第一群組內每一字元之出現次數總和為568604，該第二群組內每一字元之出現次數總和為544478，該第三群組內每一字元之出現次數總和為539607，該第四群組內每一字元之出現次數總和為504268，由此可知，上述分配方式使每一群組之該些字元之出現次數總合相當平均，相差不會超過7萬次。劃分該些字元至該第一群組、該第二群組、該第三群組及該第四群組後，給予每一群組一群組編號，每一群組編號係由複數二進位編碼構成，例如：該第一群組之該群組編號為00，該第二群組之該群組編號為01，該第三群組之該群組編號為10，該第四群組之該群組編號為11。

復參閱第二圖，接著執行步驟S26，依據每一字元所對應之該群組編碼，編碼每一字彙為一母編碼。請一併參閱第五圖，係本發明之另一較佳實施例之字彙與母編碼之對應圖。如圖所示，像圖中之字彙「a1-acid glycoprotein」對應之母編碼為「01010101101000010100001011111111101001000」，字彙「a1-lipoprotein」對應之母編碼為「0101010010111111111101001000」，每一字彙對應之該母編碼之長度主要依據該字彙之長度，該字彙的長度越長，對應之該母編碼之長度越長；反之，該字彙的長度越短，對應之該母編碼之長度越短。

然後執行步驟S27，分割每一字彙所對應之該母編碼為複數組子編碼，分割每一字彙所對應之該母編碼前，先設定分割後之該子編碼之組數，並判斷每一組子編碼所包含之複數二進位編碼之個數是否相同，若每一組子編碼所包含之該些二進位編碼之個數不同，必須於該母編碼後面補上至少d個二進位編碼，其計算式為d=(C-(TX mod C)) mod C，在此C為子編碼之組數、TX為該母編碼之長度。舉例來說，該字彙所對應之該母編碼為「00100110011010101」，該子編碼之組數為C=4組，該母編碼之長度為TX=18，即表示由十八個二進位編碼構成，若該母編碼欲分割成四組子編碼，又要使每一組子編碼所包含之該二進位編碼之個數相同，所以必須於該母編碼後面補上兩個二進位編碼。經過上述計算式d=(4-(18 mod 4)) mod 4=2，可得出d=2，本實施例係於該母編碼後面補上兩個0，所以該母編碼變成「

00100110011011010100」，然後將該母編碼分割成四組子編碼，第一組子編碼為「00100」，第二組子編碼為「11001」，第三組子編碼為「10110」，第四組子編碼為「10100」。

接著執行步驟S28，設定每一母編碼所對應之每一子編碼之一代表碼，並組合每一母編碼所對應之每一子編碼之該代表碼為一索引碼，每一索引碼代表對應該母編碼之該字彙。延續上述例子，第一組子編碼為「00100」之該代表碼為「0」，第二組子編碼為「11001」之該代表號為「1」，第三組子編碼為「10110」之該代表號為「1」，第四組子編碼為「10100」之該代表號為「0」，從上述可知，每一組子編碼對應之該代表碼為一二進位編碼，該代表碼之設定規則判斷每一組子編碼中的「1」及「0」何者出現個數最多，當該子編碼之「1」之出現個數最多時，該子編碼對應之該代表碼為「1」；反之，當該子編碼之「0」之出現個數最多時，該子編碼對應之該代表碼為「0」。另外一種特殊狀況為當該子編碼之「1」與「0」之出現個數一樣多時，該子編碼對應之該代表碼設定為「1」。組合上述每一組子編碼對應之該代表碼所得到之該索引碼為「0110」，所以該索引碼亦由複數二進位編碼構成，而該索引碼「0110」即代表對應該母編碼之該字彙。下述為演算每一字彙對應之該索引碼之程式：

Input(輸入):TX(母編碼),C(子編碼之組數)

Output(輸出):The signature K(索引碼) of TX

Step1：Calculate the length of binary code

of TX as |TX|(計算母編碼之長度)

Step2：d=(C-(|TX| mod C)) mod C

if (d<>0)

TX=TX+ the string of d zeroes //" +" represents the concatenation of string //

(上述計算是否須於母編碼後面補上0，及須補上幾個0)

Step3：Divide TX evenly into TX1, TX2,…, and TXC

(上述分割母編碼為複數組子編碼)

Step4：P=|TX|/C

FOR i=1 TO C

FOR t=0 TO 1

FOR j=1 TO P

rj=(t) XOR (j-th bit of TXi)

Next j

Compose the total bit string R1 as r1r2…rp

Next t

Compare R0 and R1 to check the number of 1's in

These two bit strings. If R0 has least number of 1's then

set Ki=0 else set Ki=1

(上述計算出每一組子編碼所對應之該代表碼Ki)

Step5：Output the signature of TX as bit

Intellectual Property Office

string K as K1K2…KC

(上述輸出每一字彙對應之該索引碼K)

復參閱第二圖，最後執行步驟S29，匯集代表該些字彙之該些索引碼，得到一字典索引檔。請一併參閱第六圖，係本發明之另一較佳實施例之字典索引檔之示意圖。如圖所示，本實施例之該字典索引檔內包含複數索引碼，該些索引碼為八個二進位編碼構成，即表示該子編碼之組數C設定為8，如此將該母編碼分成八組子編碼，然後再設定每一組子編碼對應之該代表碼，最後由每一組子編碼對應之該代表碼構成該索引碼。該字典索引檔內之每一索引碼對應複數字彙，如此可知某一字彙對應之該索引碼可能與其他字彙對應之該索引碼相同，然後使用者日後欲查詢一字彙是否存在於此字典內，係先利用本發明之字典索引之建立方法分析該字彙，以得到該字彙對應之該索引碼，接著判斷該索引碼是否存在於該字典索引檔內所包含之該些索引碼，若該字彙對應之該索引碼存在於該字典索引檔內，即表示該字彙可能存在於該字典內，當知道該字彙對應之該索引碼存在於該字典索引檔內時，可從該索引碼對應該字典內之複數字彙進行搜索，如此可有效縮短搜尋字彙的時間。

請參閱第七圖，係本發明之另一較佳實施例之子編碼之組數與搜尋時間之關係圖。如圖所示，當子編碼之組數C設定越大時，索引碼之數量會增加，亦佔較多的記憶體，但每一索引碼對應之字彙數量越少，如此得知欲搜尋之字彙對應之該索引碼存在於該字典之該字典索引檔，並從對應該索引碼之複數字彙中尋找欲搜尋之字彙

，所以對應該所以碼之字彙數量越少，搜尋欲搜尋之字彙的時間越短。所以子編碼之組數將會影響日後搜尋字彙的時間，但子編碼之組數將依據電腦系統之記憶體容量設定較佳，防止該字典索引檔佔據較多記憶體容量。

由上述可知，本發明係為提供一種字典索引之建立方法，可將每一種字典依據本發明所提供之字典索引之建立方法建立一字典索引檔，使用者可依據自己的需求選擇適合的字典，並下載該字典對應之該字典索引檔。然後本發明所建立之該字典索引檔之使用方法係先利用本發明之建立方法分析欲搜尋之一字彙，得知該字彙對應之該索引碼，當確認欲搜尋之該字彙對應之該索引碼存在於該字典索引檔內時，可依據該索引碼對應之該些字彙搜尋欲搜尋之字彙，如此有效減少搜尋的字彙數量，更縮短搜尋時間。然後本發明之建立方法更可控制子編碼之組數，決定該字典索引檔佔記憶體容量之大小。

綜上所述，本發明具有新穎性、進步性及可供產業利用者，應符合我國專利法所規定之專利申請要件無疑，爰依法提出發明專利申請，祈　鈞局早日賜准專利，至感為禱。惟，以上所述者，僅為本發明之一較佳實施例而已，並非用來限定本發明實施之範圍，舉凡依本發明申請專利範圍所述之形狀、構造、特徵及精神所為之均等變化與修飾，均應包括於本發明之申請專利範圍內。

【圖式簡單說明】

[0005] 第一圖：本發明之一較佳實施例之使用字典索引檔之流程圖；

第二圖：本發明之一較佳實施例之建立字典索引檔之流程圖；

第三圖：本發明之另一較佳實施例之字元出現次數統計圖；

第四圖：本發明之另一較佳實施例之群組編碼示意圖；以及

第五圖：本發明之另一較佳實施例之字彙與母編碼之對應圖；

第六圖：本發明之另一較佳實施例之字典索引檔之示意圖；以及

第七圖：本發明之另一較佳實施例之子編碼之組數與搜尋時間之關係圖。

【主要元件符號說明】

[0006] 無。

201211798

智專收字第0992054063-0

DTD版本：1.0.1

專利案號：099130818

日期：99年09月13日

發明專利說明書

※申請案號：099130818　　※ＩＰＣ分類：

※申請日：99.9.13

G06F17/30 (2006.01)

一、發明名稱：

字典索引之建立方法及使用方法

二、中文發明摘要：

本發明有關於供一種字典索引之建立方法及使用方法，該字典索引之建立方法主要將每一字典之每一字彙轉換成一索引碼，匯集該些索引碼得到一字典索引檔。如此可將每一種字典依據本發明所提供之字典索引之建立方法建立一字典索引檔，使用者可依據自己的需求選擇適合的字典，並下載該字典對應之該字典索引檔。然後本發明所建立之該字典索引檔之使用方法為該字典索引檔包含該些索引碼，每一索引碼對應複數字彙，當確認欲搜尋之字彙對應之該索引碼存在於該字典索引檔內時，可依據該索引碼對應之該些字彙搜尋欲搜尋之字彙，如此有效減少搜尋字彙的數量，更縮短搜尋字彙時間。

三、英文發明摘要：

七、申請專利範圍：

1．一種字典索引之建立方法，係包含：

取一字典之一字典字彙檔，該字典字彙檔包含複數字彙，每一字彙包含複數字元；

統計每一字元出現於該些字彙之次數，並依據每一字元所出現之次數排序該些字元；

依據該些字元之排序劃分該些字元為複數群組，每一群組之該些字元對應一群組編碼；

依據每一字元所對應之該群組編碼，編碼每一字彙為一母編碼；

分割每一字彙所對應之該母編碼為複數組子編碼；

設定每一母編碼所對應之每一子編碼之一代表碼，並組合每一母編碼所對應之每一子編碼之該代表碼為一索引碼，每一索引碼代表對應該母編碼之該字彙；以及

匯集代表該些字彙之該些索引碼，得到一字典索引檔。

2．如申請專利範圍第1項所述之字典索引之建立方法，其中該字典可為一語言字典。

3．如申請專利範圍第1項所述之字典索引之建立方法，其中該字典可為一專業字典。

4．如申請專利範圍第1項所述之字典索引之建立方法，其中該些字元包含字母、標點符號或空格。

5．如申請專利範圍第1項所述之字典索引之建立方法，其中該群組編碼為包含複數二進位編碼。

6．如申請專利範圍第1項所述之字典索引之建立方法，其中

Intellectual Property Office

該些群組分別包含對應之該些字元，每一群組對應之該些字元出現次數總合相差不超過7萬次。

7．如申請專利範圍第1項所述之字典索引之建立方法，其中該母編碼包含複數二進位編碼。

8．如申請專利範圍第1項所述之字典索引之建立方法，其中該些組子編碼分別包含複數二進位編碼，每一子編碼所包含之該些二進位編碼之個數相同。

9．如申請專利範圍第8項所述之字典索引之建立方法，其中分割每一字彙所對應之該母編碼為該些組子編碼前，更判斷每一組子編碼所包含之該些二進位編碼之個數是否相同。

10．如申請專利範圍第9項所述之字典索引之建立方法，其中判斷每一組子編碼所包含之該些二進位編碼之個數不相同，則於該母編碼後面補上至少一二進位編碼。

11．如申請專利範圍第8項所述之字典索引之建立方法，其中該代表碼係依據每一子編碼所包含之該些二進位編碼選擇一二進位編碼。

12．如申請專利範圍第1項所述之字典索引之建立方法，其中該索引碼包含複數二進位編碼。

13．一種字典索引之使用方法，係包含：

下載一字典之一字典索引檔，該字典索引檔包含複數索引碼，每一索引碼對應一字典字彙檔之複數字彙；

輸入欲搜尋之一字彙；

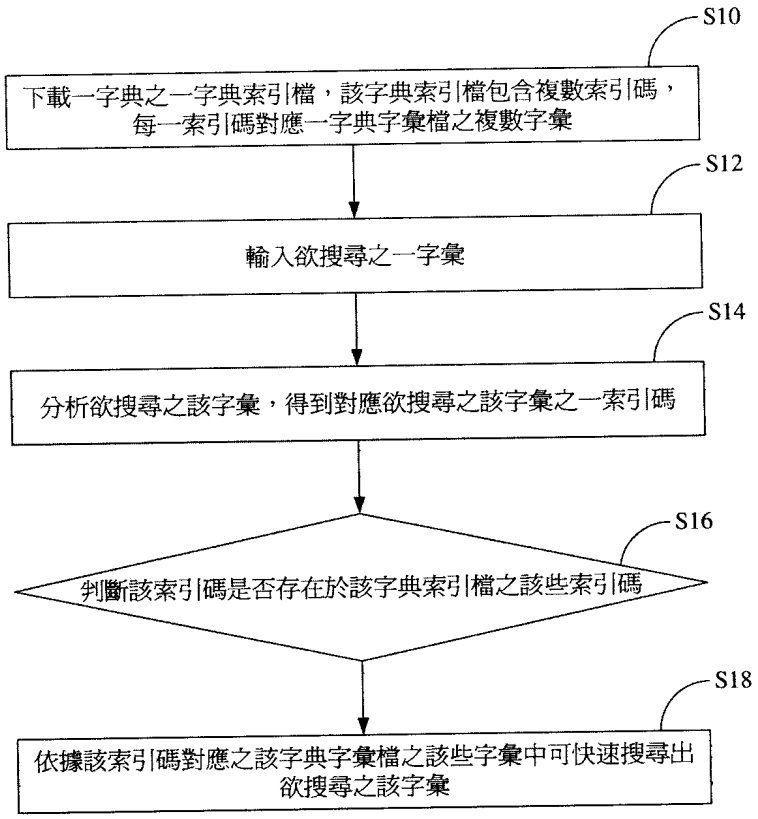
分析欲搜尋之該字彙，得到對應欲搜尋之該字彙之一索引碼；

判斷該索引碼存在於該字典索引檔之該些索引碼；以及

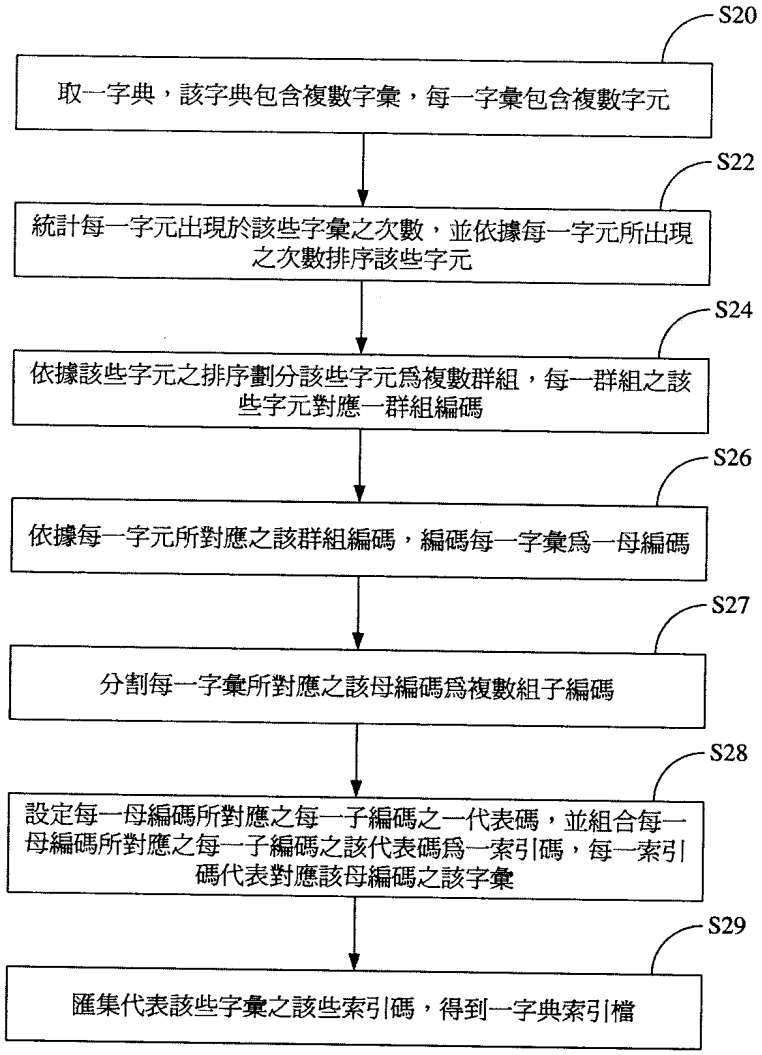
依據該索引碼對應之該字典字彙檔之該些字彙搜尋欲搜尋之該字彙。

Intellectual Property Office

八、圖式：


S10
下載一字典之一字典索引檔，該字典索引檔包含複數索引碼，每一索引碼對應一字典字彙檔之複數字彙
S12
輸入欲搜尋之一字彙
S14
分析欲搜尋之該字彙，得到對應欲搜尋之該字彙之一索引碼
S16
判斷該索引碼是否存在於該字典索引檔之該些索引碼
S18
依據該索引碼對應之該字典字彙檔之該些字彙中可快速搜尋出欲搜尋之該字彙

第一圖


S20
取一字典，該字典包含複數字彙，每一字彙包含複數字元
S22
統計每一字元出現於該些字彙之次數，並依據每一字元所出現之次數排序該些字元
S24
依據該些字元之排序劃分該些字元爲複數群組，每一群組之該些字元對應一群組編碼
S26
依據每一字元所對應之該群組編碼，編碼每一字彙爲一母編碼
S27
分割每一字彙所對應之該母編碼爲複數組子編碼
S28
設定每一母編碼所對應之每一子編碼之一代表碼，並組合每一母編碼所對應之每一子編碼之該代表碼爲一索引碼，每一索引碼代表對應該母編碼之該字彙
S29
匯集代表該些字彙之該些索引碼，得到一字典索引檔

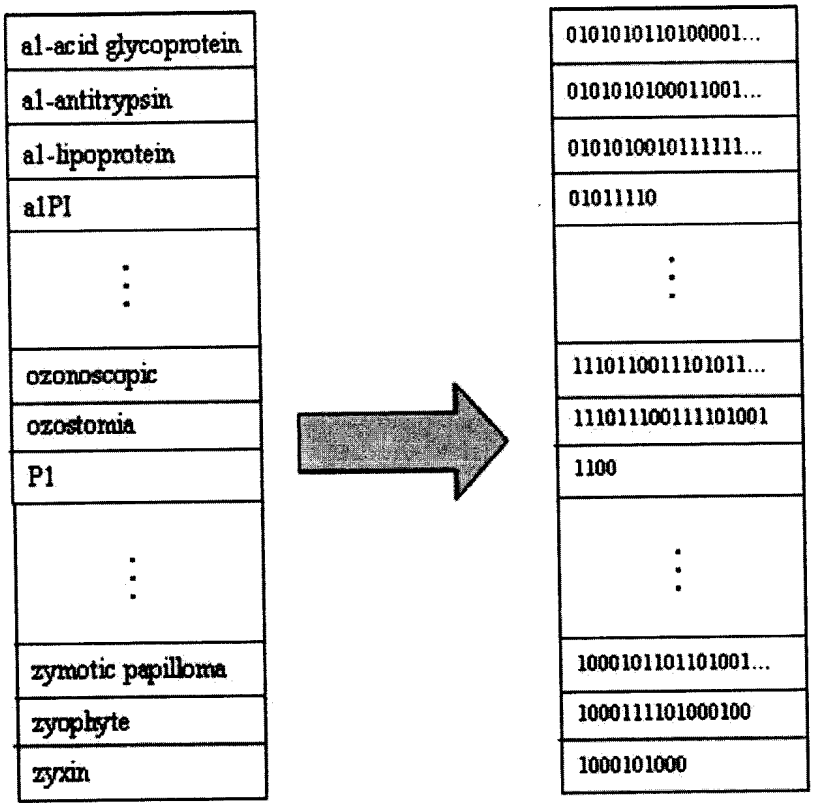
第二圖

序號	字元	出現次數	序號	字元	出現次數	序號	字元	出現次數
①	e	202760	⑰	y	45348	㉝	3	386
②	a	191554	⑱	g	39335	㉞	4	329
③	i	182916	⑲	b	28847	㉟	5	289
④	o	159617	⑳	f	25856	㊱	(	252
⑤	r	149568	㉑	v	19116	㊲	)	252
⑥	s	138910	㉒	x	9334	㊳	6	224
⑦	t	136899	㉓	-	9232	㊴	7	141
⑧	n	130193	㉔	k	8251	㊵	0	109
⑨	l	117922	㉕	,	6594	㊶	9	95
⑩	Space	103379	㉖	w	6268	㊷	8	90
⑪	c	102408	㉗	z	4267	㊸	/	61
⑫	u	71942	㉘	'	4228	㊹	:	51
⑬	p	70857	㉙	j	2498	㊺	.	31
⑭	m	70257	㉚	q	2070	㊻	+	24
⑮	h	56494	㉛	1	749	㊼	&	8
⑯	d	56454	㉜	2	489	㊽	_	3

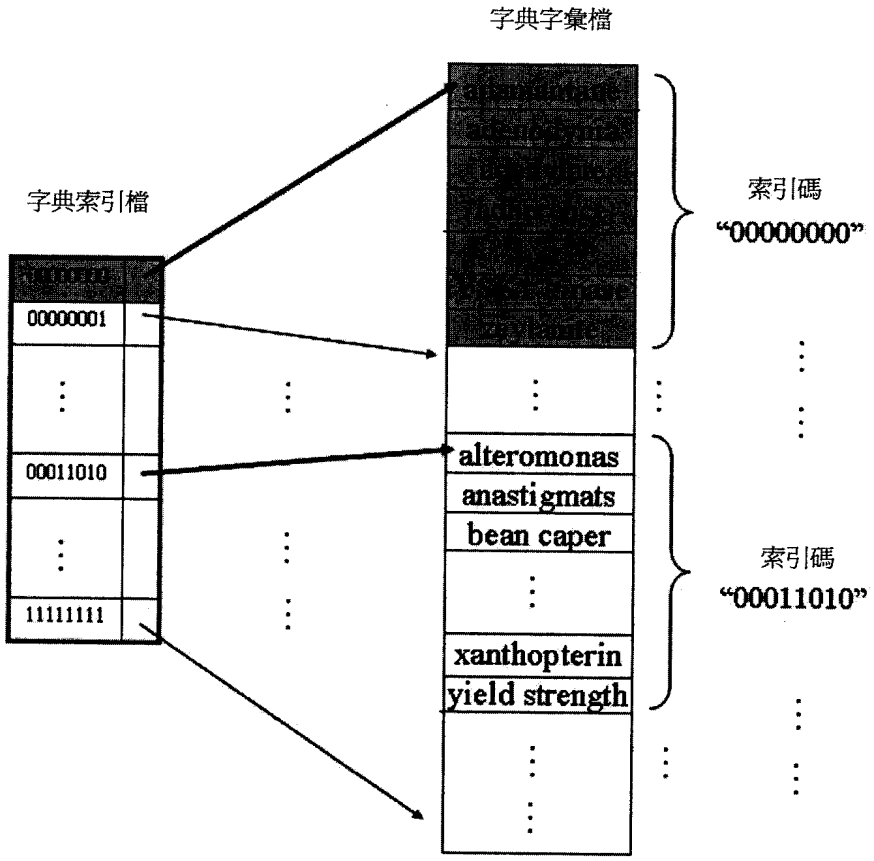
第三圖

群組編碼	包含字元	包含序號	總出現次數
00	“e”, “n”, “l”, “d”, “y”, “k”, “ ”, “2”, “3”, “0”, “9”, “ ”, “_”	① ⑧ ⑨ ⑯ ⑰ ㉔ ㉕ ㉜ ㉝ ㊵ ㊶ ㊽	568,604
01	“a”, “t”, “Space”, “h”, “g”, “-”, “w”, “1”, “4”, “7”, “8”, “&”	② ⑦ ⑩ ⑮ ⑱ ㉓ ㉖ ㉛ ㉞ ㊴ ㊷ ㊼	544,478
10	“i”, “s”, “c”, “m”, “b”, “x”, “z”, “g”, “5”, “6”, “/”, “+”	③ ⑥ ⑪ ⑭ ⑲ ㉒ ㉗ ㉚ ㉟ ㊳ ㊸ ㊻	539,607
11	“o”, “r”, “u”, “p”, “f”, “v”, “”, “j”, “(”, “)”, “:”, “.”	④ ⑤ ⑫ ⑬ ⑳ ㉑ ㉘ ㉙ ㊱ ㊲ ㊹ ㊺	504,268

第四圖


a1-acid glycoprotein
a1-antitrypsin
a1-lipoprotein
a1PI
ozonoscopic
ozostomia
P1
zymotic papilloma
zyophyte
zyxin
0101010110100001...
0101010100011001...
0101010010111111...
01011110
1110110011101011...
111011100111101001
1100
1000101101101001...
1000111101000100
1000101000

第五圖


字典字彙檔
字典索引檔
00000001
00011010
11111111
alteromonas
anastigmats
bean caper
xanthopterin
yield strength
索引碼
"00000000"
索引碼
"00011010"

第六圖

子編碼之組數	索引碼之總數	佔記憶體容量	搜尋字彙之平均數量	平均搜尋時間
8	256	2048	1212	4.63
9	512	4608	728	3.72
10	1024	10240	610	2.31
11	1928	21208	466	1.48
12	3881	46572	304	1.42
13	5769	74997	259	1.25
14	11543	158830	124	0.78
15	15364	230460	116	0.76
16	25103	401648	107	0.67

第七圖

201211798

0992054063-0

表單編號 A0101

099130818

四、指定代表圖：

(一)本案指定代表圖為：第二圖

(二)本代表圖之元件符號簡單說明：

無。

五、本案若有化學式時，請揭示最能顯示發明特徵的化學式：