

# 數位圖書館中權威控制系統的設計 The Design for Authority-Control Systems in Digital Libraries

陳 光 華

Kuang-hua Chen

國立臺灣大學圖書資訊學系副教授

Associate Professor

Department of Library and Information Science

National Taiwan University

khchen@ccms.ntu.edu.tw

## 【摘要 Abstract】

數位圖書館已成為資訊社會重要的資訊集散中心，也是資訊時代國家總體力量的重要指標，其研究的急迫性不言而喻。數位圖書館之研究本質是科際性的，如何由不同的觀點檢視原屬其他領域的研究是一項值得深入探討的課題。本文藉由資訊組織與整理的觀點，說明現有數位圖書館資訊檢索子系統不足之處，提出應使用權威控制子系統，以服務使用者、資訊檢索子系統以及詮釋資料著錄子系統，同時討論該權威控制子系統的功能與規格，並以淡新檔案索引典為案例，實作詞彙瀏覽、詞彙檢索、以及詞彙對映等三種功能。

The digital library has become the important center of information collection and dissemination in the information society. It is also the important indicator of country power in the information era. Since the nature of DL researches is interdisciplinary, how to investigate the various researches of DL from different viewpoints is a critical issue. This paper argues the inadequacy of current information retrieval researches from the viewpoint of information organization, proposes an authority-control subsystem (dubbed Thesalogy) to serve users, IR subsystems and metadata subsystems in DL, and discusses the functionality and specification of Thesalogy. Finally, this paper reports the implementation of three functions of Thesalogy, i.e., browsing vocabularies, retrieving vocabularies, and mapping vocabularies in terms of Dan-shin Archive thesaurus.

## 關鍵詞 Keyword

權威控制 數位圖書館 資訊檢索

Authority control; Digital library; Information retrieval



## 壹、緒論

隨著網際網路的快速發展，數位圖書館／博物館的建置與研究越來越受到世界各國的重視。數位圖書館／博物館的研究非屬單一學門，其科際性（interdisciplinary）的本質促成各學門的相互交流<sup>①</sup>，簡而言之，吾人可以說至少需要三種不同的學者專家投入其中，一是電腦科技學者（computer scientists）、二是圖書資訊學者（library scientists）、三是專門學科學者（content experts），唯有此三類的學者專家各司其職，構成穩固的三角結構，再加上有效地交流合作，才能夠建構完整的數位圖書館／博物館系統。綜觀目前數位圖書館／博物館的研究核心，大致環繞在資源數位化、資源組織與整理、通訊協定、資源儲存系統、資訊檢索、資源展現、使用者行為、系統評估等課題。至於筆者近幾年參與數位圖書館／博物館研究計畫，主要的研究課題是資源組織與整理以及資訊檢索，屬於電腦科技學與圖書資訊學的研究範疇。資訊檢索的研究長期以來並未受到國內學術單位的重視，相關的課程僅有圖書館學系所開設的「資訊儲存與檢索」，然此課程的目標主要為教導學生學習檢索策略，並多數以光碟資料庫為學習應用的對象，而未深入整合探討其他如資源建立、儲存、傳播等相關議題。近來，資訊檢索研究的蓬勃發展，主要拜網際網路之賜，不僅是學者專家，一般的使用者也感受到資訊檢索系統的重要性。然而，目前網際網路上的檢索系統多數是由電腦科學專長的人士建構而成，主要提供全文檢索的功能，但並未以資源本身及使用者的角度切入，以致無法有效滿足檢索需求。本文將從資源組織與整理的觀點，討論數位圖書館應提供何種資訊檢索的服務，並擴展服務的概念，提出權威控制系統，協助使用者有效使用數位圖書館提供的各項資源，同時可以協助館員著錄館藏資料。

早期的圖書館以卡片記載圖書資料的各項書目資料，因應不同的檢索目的，而有多套不同的排卡系統（如著者卡、題名卡、標題卡、分類卡等）。當使用者對某一位著者的著作感興趣，會查詢著者卡以瞭解該著者一系列的著作；當使用者知道題名而不知其他書目資料，可查閱題名卡；當使用者對某一項主題感興趣時，可以查閱標題卡或分類卡，得知相關的著作。根據美國圖書館學會的調查，公共圖書館的使用者多數使用標題檢索，學術圖書館與專門圖書館的學者則多數使用題名或著者檢索；但耶魯大學的研究則顯示，56%是題名或著者檢索，而33%是主題檢索（包括標題與分類號）。<sup>②</sup>然而，隨著圖書館作業逐漸自動化，線上查詢成為檢索作業的主流，使用者的檢索方式也有所改變。根據台灣大學黃慕萱教授1996年所做的調查，有82.5%的檢索是屬於主題檢索<sup>③</sup>，顯示主題檢索已漸漸成為資訊檢索的主要方式。數位圖書館系統透過網際網路提供各類型的資訊服務，就資訊檢索的功能而言，主題檢索的重要性將益形顯著，因為使用者更加地多元化，更無法預期使用者具有何種背景知識，因此，數位圖書館的研究者與建置者必須更強化主題檢索的功能，提供額外的服務子系統，方能有效協助使用者主題檢索的需求。

由於電腦科技的日新月異，再加上商業體系的投入，整個世界就像是縮小了一般，訊息的傳遞很快，人類享有前所未見的資訊服務。全球各地的使用者透過網際網路此一新興媒體，檢索各式各樣的資料、文獻、或資訊，電子資訊的累積也變得極為快速，網際網路上的資訊量非常龐大，以往資訊多數是由有組織的機構提供，如圖書館、博物館、行政單位、企業組織等等。現在，任何人都可以作為資訊的提供者，藉由網際網路的連線，將個人的觀點、評述、作品傳播給讀者。由於資訊的來源益形多元，資訊的類型益形多變，某一類的資訊可能只有該領域的人了解，能夠使用比較高階的檢索方式



取得資訊。但是，大多數的時候，我們都是一般的資訊檢索者，即便是生化學家，在檢索有關經濟方面的資料時，也只是扮演一般使用者的角色。以主題作為主要的檢索點在資訊檢索行為中成為一重要的趨勢，這種檢索方式的改變對於檢索系統的服務方式具有重大的意義，為了讓使用者有效地透過主題檢索的方式，檢索出所需的文獻資料，又不必牢記檢索系統在建立內部索引時使用的控制詞彙，檢索系統必須提供檢索詞彙與索引詞彙對映的友善機制。由此觀之，隨著時間演進，詞彙對映的重要性將越來越顯著。

本文第二節將說明資訊檢索在數位圖書館的角色，以及應提供的檢索功能。第三節將討論數位圖書館服務的型式。第四節將以台大數位圖書館／博物館為例，提出權威控制系統，以服務使用者及其他子系統。第五節是簡短的結論。

## 貳、資訊檢索與資訊的組織

在目前網際網路應用益形重要的背景之下，資訊檢索愈來愈受到資訊科學研究者的重視。一般而言，資訊檢索大致可分為下列三種型式：

1. 全文檢索 (full-text retrieval)：根據使用者輸入的檢索詞彙，嘗試找出包含該詞彙的所有文件。基本上，全文檢索是針對文字型資料而設計的檢索方式，目前的技術多數仍侷限於 TF×IDF(詞彙頻率乘反文件頻率)的基本計算模式。<sup>④</sup>
2. 內容檢索 (content-based retrieval)：亦即提供使用者以內容查詢的方式 (query by content)，檢索所需的資源或資訊。例如，使用者可以使用特定的影像，檢索資料庫中類似的影像。在許多情況下，內容檢索是唯一可行的檢索方式，特別是多媒體資料，如果吾人無法賦予適切的詮釋資料，則僅能由內容檢索入手。

3. 詮釋資料檢索 (metadata retrieval)：類似圖書館 OPAC 系統提供的檢索功能，早期稱之為書目資料的檢索，但由於數位化資訊的型態不限於傳統所指稱的「書」，因此不以書目資料稱之，而名為詮釋資料。詮釋資料檢索的最大特色是可以達到一致性的檢索，無論是文字、影像、音訊、視訊等各種類型的資料，吾人皆可以檢索其詮釋資料。因此，檢索所得的資料，就可能包含多種資料類型。

對於數位圖書館系統而言，雖然典藏的資源型態眾多，但仍以文字型資訊為主，因此文字型資訊的檢索是非常重要的。紐西蘭學者 Alastair Smith 曾經仔細比較 11 個數位圖書館系統提供的檢索功能，多數系統提供布林運算功能，僅有五個系統提供控制詞彙的功能，有二個系統則提供相關詞彙的功能。<sup>⑤</sup>吾人可以發現，就使用者的角度而言，現有的數位圖書館系統在檢索功能上顯然並不完善，還必須提供更進一步的服務以滿足使用者的需求。本文主要討論的是詮釋資料的檢索，也就是在詮釋資料檢索系統的架構下，如何提供適當的服務，協助使用者有效地檢索資訊，並協助館員有效地著錄文獻之詮釋資料。

無論是實體圖書館或是數位圖書館，文獻資料在典藏之前，必須經過分類編目的過程。編目人員通常進行兩種不同方式的分析過程：一為實體分析 (physical analysis)，紀錄文獻的作者、題名、版本、出版項等資訊；另一為內容分析 (content analysis)，進行文獻的分類，並為該文獻選定若干可展現其內容的標題 (有時還必須加上階層式複分)。實體分析通常是沒有任何爭議的，然而內容分析必須經由編目人員對文獻內涵的了解與思考，才能妥善處理。為了使分類編目的結果較一致，圖書館採權威控制的方式進行，也就是說分類有既定的分類法，如國會圖書分類法 (LCC)、杜



威十進位分類法 (DCC)、中文圖書分類法；標題則有既定的標題表，如國會標題表 (LCSH)、醫學標題表 (MeSH)、中文圖書標題表。圖書館編目館員以前述權威控制的方式進行文獻資料的分類編目，而每一筆文獻的編目資料（即是一種 Metadata，也就是詮釋資料）就成為該文獻資料

可能的檢索點 (access point)，至於所使用的詞彙則被稱為索引詞彙。

一般而言，使用者可藉由類似 OPAC 或 WEBPAC 的線上公用目錄系統，使用自行決定的檢索詞彙，透過檢索系統取得所需的資料，如圖一所示。



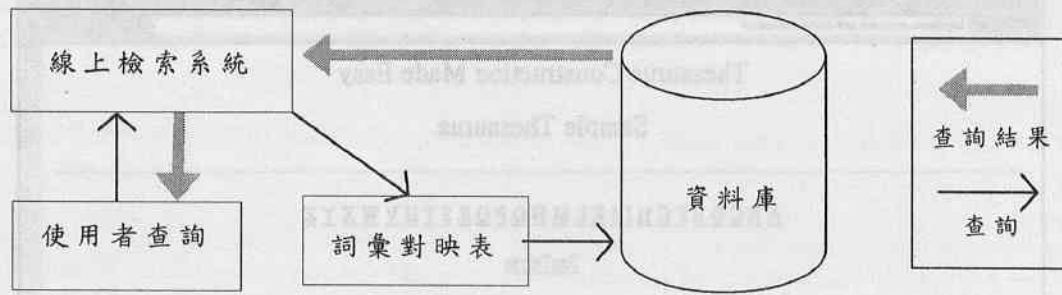
圖一、使用者使用檢索系統查詢資料

然而，使用者若是使用標題欄作檢索點查詢文獻資料，卻常常無法得到滿意的查詢結果，這是因為標題欄使用既定的控制詞彙（可能來自標題表或是索引典），而一般的使用者根本不知道系統使用的控制詞彙為何，使得檢索結果往往無法滿足使用者的需求，造成使用者抱怨連連。對於前述的困境，我們可以使用以下的解決方案：

1. 館員服務協助 (staff services)
2. 線上輔助系統 (online help system)
3. 線上詞彙展示 (online vocabulary display)
4. 自動詞彙對映 (automatic vocabulary mapping)

前二種解決方案使用者必須花費一番心力才可以得到解答；第三種方案雖然使用者必須花費時間，但是卻能透過瀏覽詞彙的過程，瞭解系統的知識架構，對於以後的檢索作業有很大的幫助；第四種方案 -- 自動詞彙對映，也就是在檢索系統將使用者的查詢送到資料庫之前，先進行詞彙的轉換，然後將轉換後的詞彙作為檢索的依據，對線上檢索系統的使用者而言，就如通透的介面一樣，使用者幾乎感受不到它的存在，就能夠解決前述的問題，請參見圖二的說明。





圖二、檢索詞彙／索引詞彙的對映

由前述的討論，若吾人使用標題表或索引典提供的控制詞彙描述數位圖書館的藏品，並在資訊檢索系統提供適當的服務功能，便可以有效協助使用者快速查檢所需的資訊。然而建構一部索引典或是標題表，必須投入相當多的人力、物力，一旦建構完成上線提供大眾使用後，若是使用者不知道這些詞彙，因而造成系統無法提供有效服務的現象時，則投入的人力、物力等於白白浪費。因此，若能夠將使用者的檢索詞彙轉換為索引詞彙，或是提供線上索引典或標題表讓使用者瀏覽，便能夠強化資訊檢索系統的功能，提昇檢索品質。以國家科學委員會科學技術資料中心的科技索引典為例，共收錄 49,270 個關鍵詞（包括 37,701 個描述語及 11,569 個非描述語），並依照科技分類典將關鍵詞分成 21 個大類、182 個中類，詞彙之間有下位語（NT）、上位語（BT）、關聯語（RT）等關係。如 IA 類為資訊工程類，複分為 00 - 資訊工程總論、01 - 電腦理論、02 - 電腦系統與硬體、03 - 電腦軟體、04 - 電腦應用技術。若專業的科技數位圖書館採用該索引典描述其藏品，而檢索系統又可以提供其與使用者的檢索詞彙之間的自動轉換，或是讓使用者可在線上瀏覽索引典，選擇適當的檢索詞彙，並廣泛

使用詞彙之間的等同關係（equivalent relation）、層級關係（hierarchical relation）及關聯關係（associative relation）豐富其檢索條件，則檢索結果必能更加滿足使用者的資訊需求。

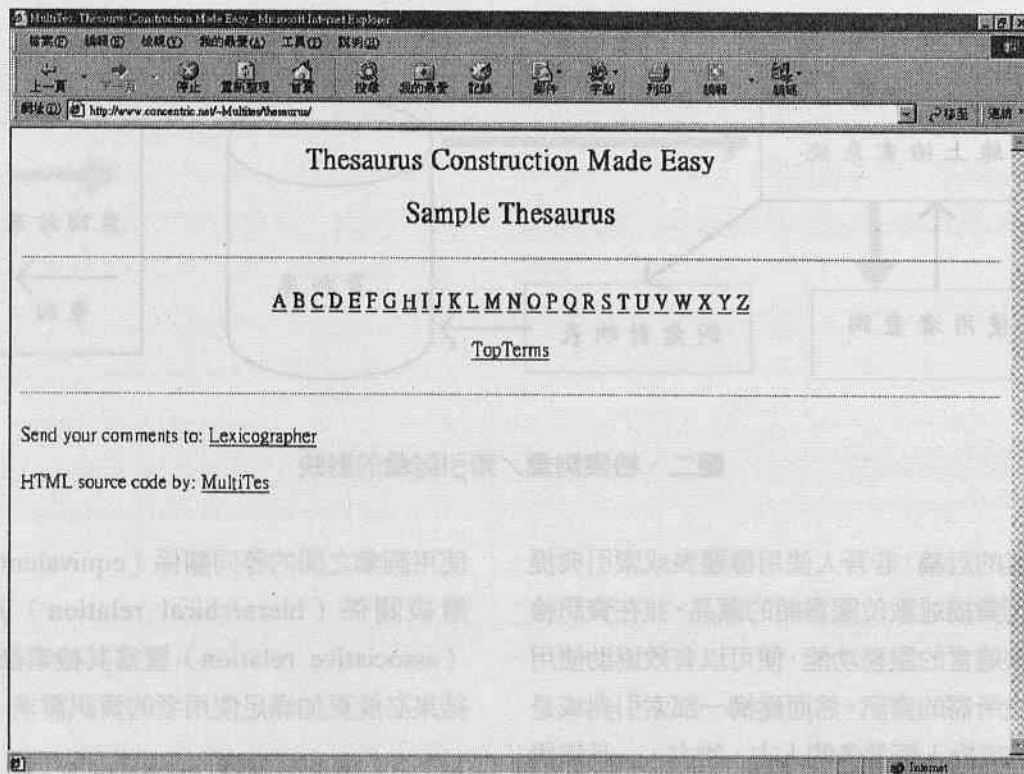
### 參、資訊檢索的服務型式

第二節曾指出四種線上檢索作業時可能的服務方案：館員服務協助、線上輔助系統、線上詞彙展示、以及自動詞彙對映，其中館員服務協助與資訊檢索系統無關，而線上輔助系統的概念早已為多數資訊系統採用，本文將不予討論，以下分二部分探討線上詞彙展示與自動詞彙對映之機制。

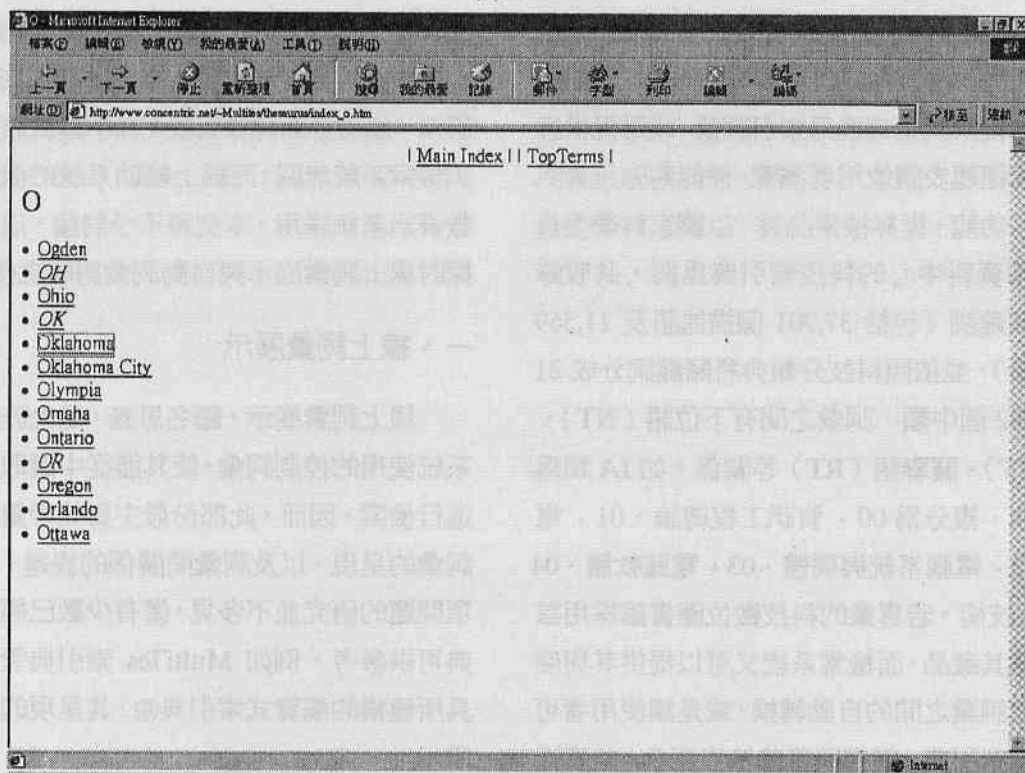
#### 一、線上詞彙展示

線上詞彙展示，顧名思義，是允許使用者瀏覽系統使用的控制詞彙，使其能從中選取適當的詞彙進行檢索。因而，此部份最主要的考量因素為控制詞彙的呈現，以及詞彙間關係的表達。目前探討這項問題的研究並不多見，僅有少數已經上線之索引典可供參考，例如 MultiTēs 索引典管理與建構工具所建構的瀏覽式索引典<sup>⑥</sup>，其呈現的方式如圖三所示：



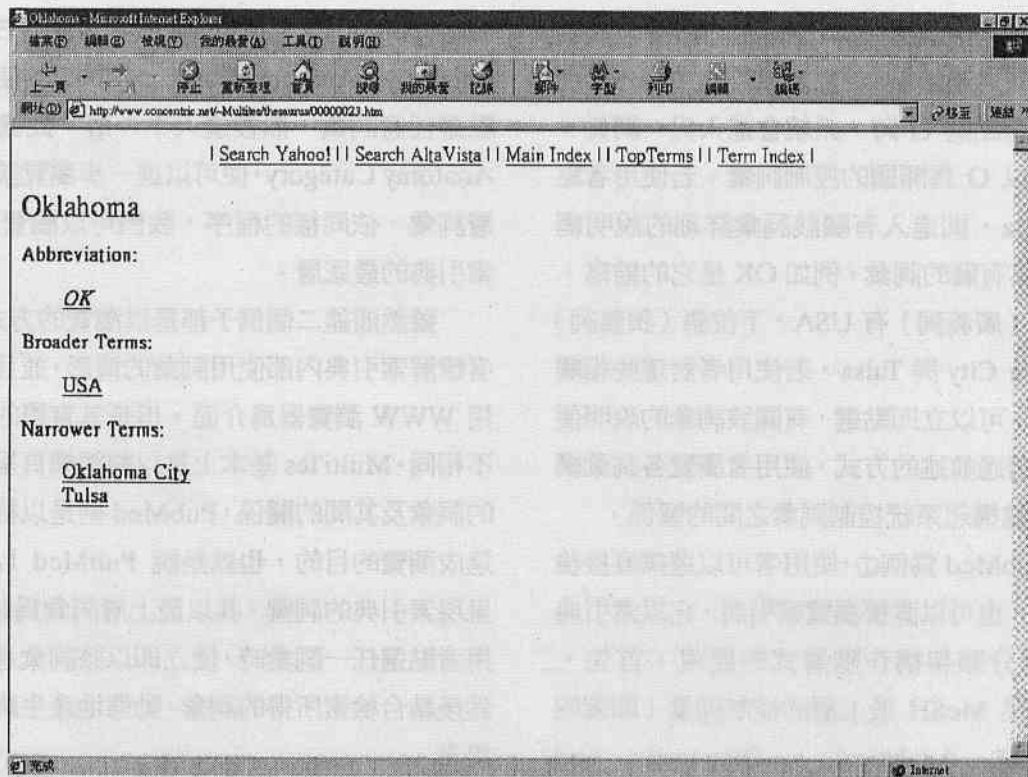


(a)

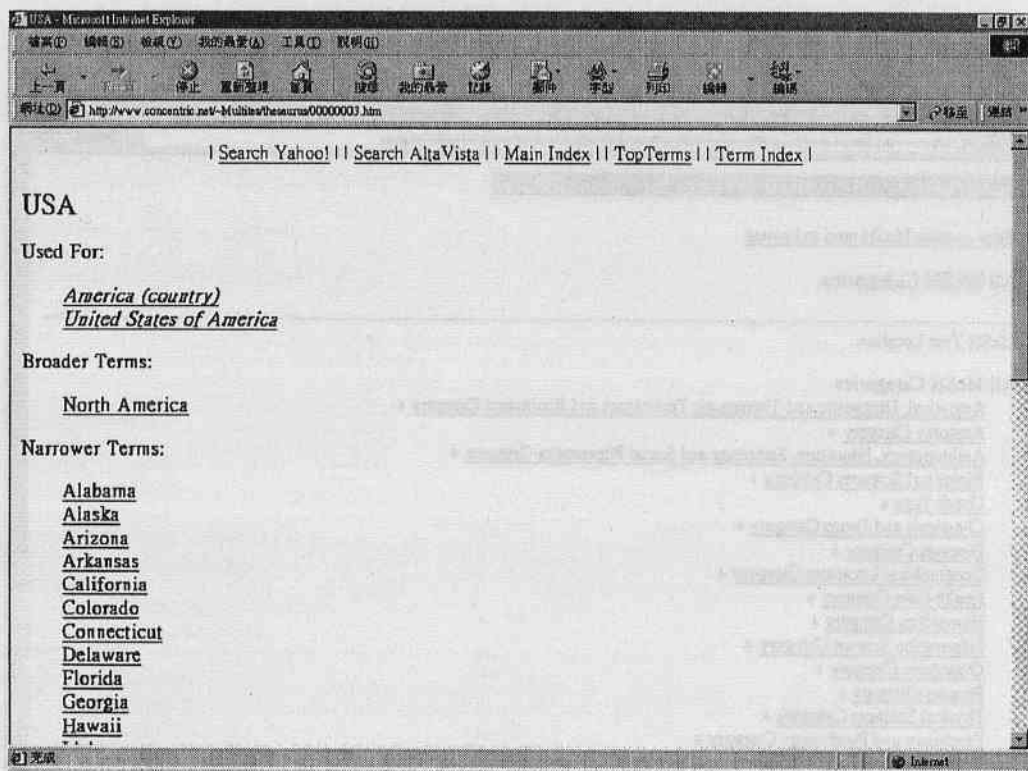


(b)





(c)



(d)

圖三、MultiTes 之 WWW 瀏覽式索引典

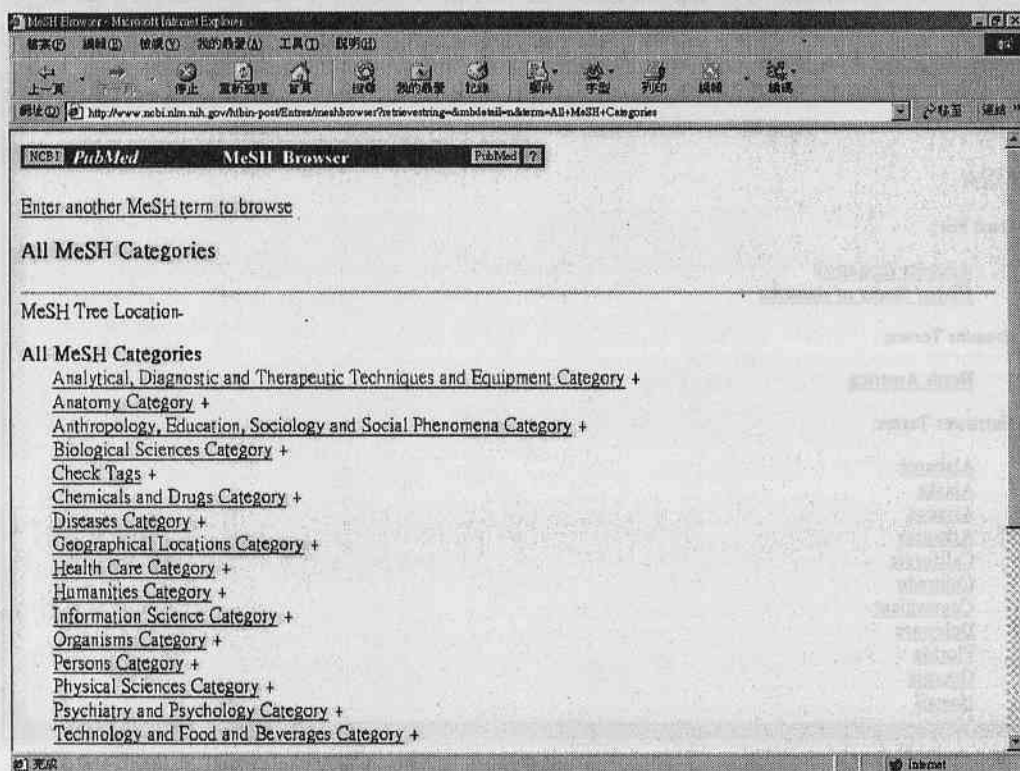


MultiTes 利用以字母順序為主的階層式展現架構作為瀏覽之起始點。首頁是以英文字母分門別類，若使用者點選 O 時，系統會進入另一網頁，並列出所有以 O 為開頭的控制詞彙。若使用者點選 Oklahoma，則進入有關該詞彙詳細的說明網頁，列出與其有關的詞彙，例如 OK 是它的縮寫，它的上位語（廣義詞）有 USA，下位語（狹義詞）有 Oklahoma City 與 Tulsa。若使用者對這些相關詞彙感興趣，可以立即點選，有關該詞彙的說明便立即呈現。透過前述的方式，使用者瀏覽各詞彙網頁，便可以建構起系統控制詞彙之間的關係。

再以 PubMed 為例<sup>⑦</sup>，使用者可以選擇直接檢索控制詞彙，也可以直接瀏覽索引典，它以索引典本身的主題分類架構作階層式的展現。首先，PubMed 提供 MeSH 最上層的控制詞彙（即圖四所呈現之 Analytical、Diagnostic and

Therapeutic、Techniques and Equipment Category、Anatomy Category 等等）。使用者可以點選任意詞彙，然後進入下一層，此處我們點選 Anatomy Category，便可以進一步瀏覽該詞彙之下層詞彙，依同樣的程序，我們可以瀏覽至 MeSH 索引典的最底層。

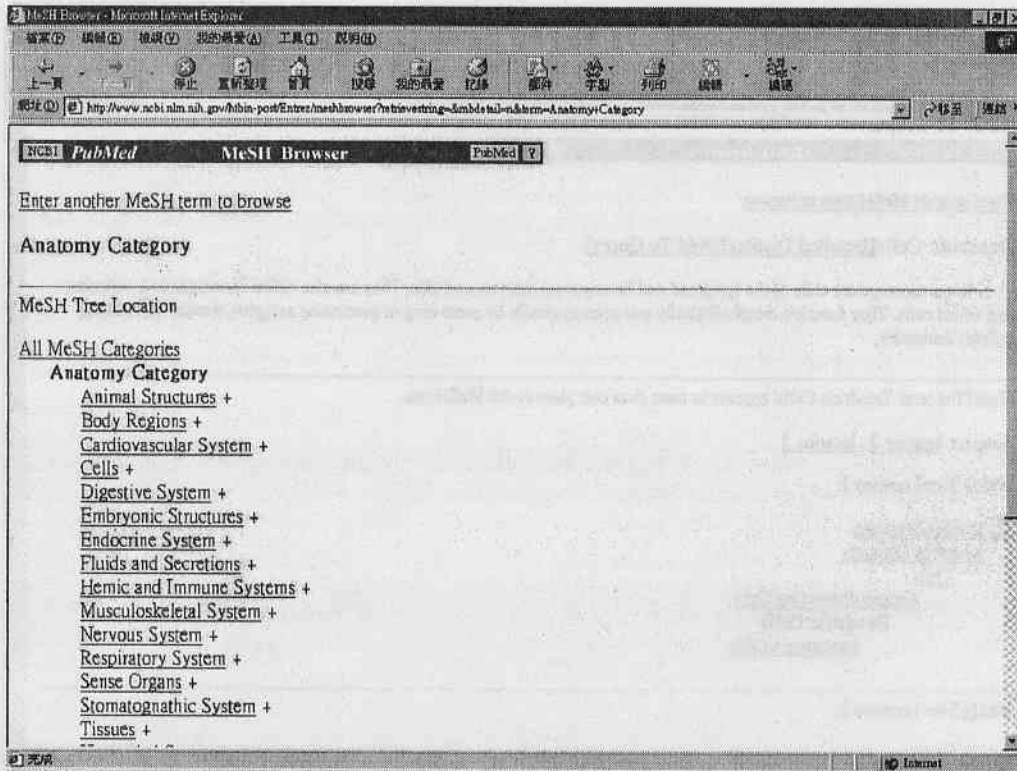
雖然前述二個例子都是以瀏覽的方式，讓使用者瞭解索引典內部使用詞彙的情形，並且也同樣使用 WWW 瀏覽器為介面，但是其實際的作法卻並不相同。MultiTes 基本上是以靜態網頁呈現索引典的詞彙及其間的關係，PubMed 則是以檢索的方式達成瀏覽的目的，也就是說 PubMed 以動態網頁呈現索引典的詞彙，其以最上層詞彙為起點，當使用者點選任一詞彙時，便立即以該詞彙進行檢索，然後結合檢索所得的詞彙，動態地產生網頁告知使用者。



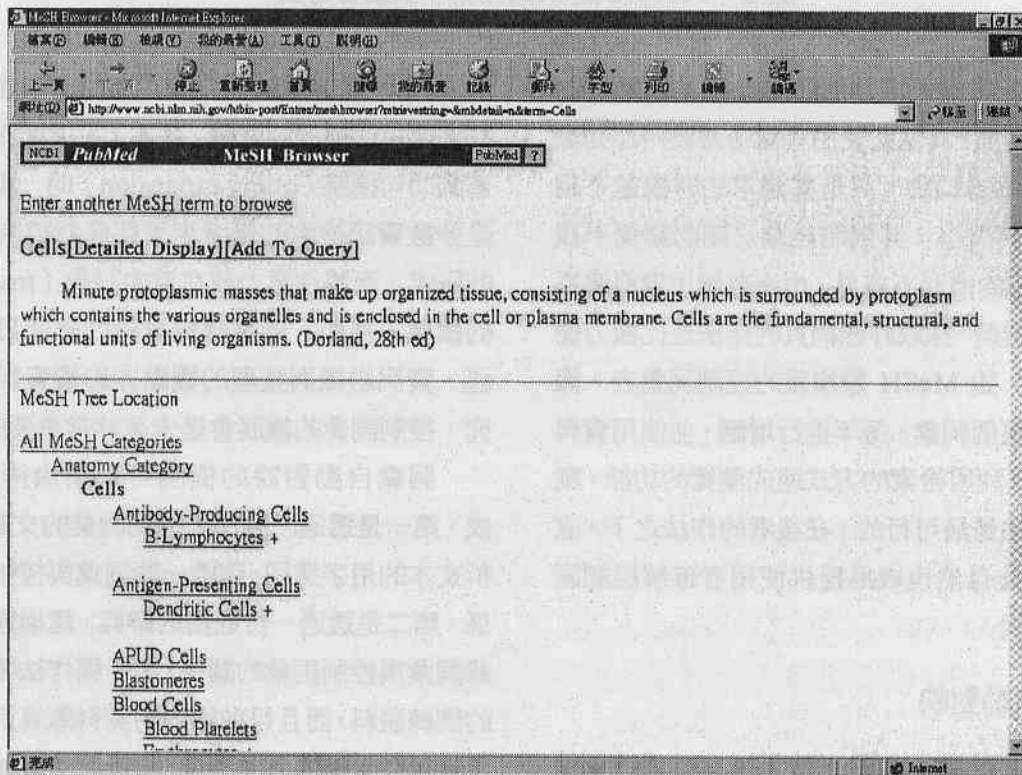
(a)





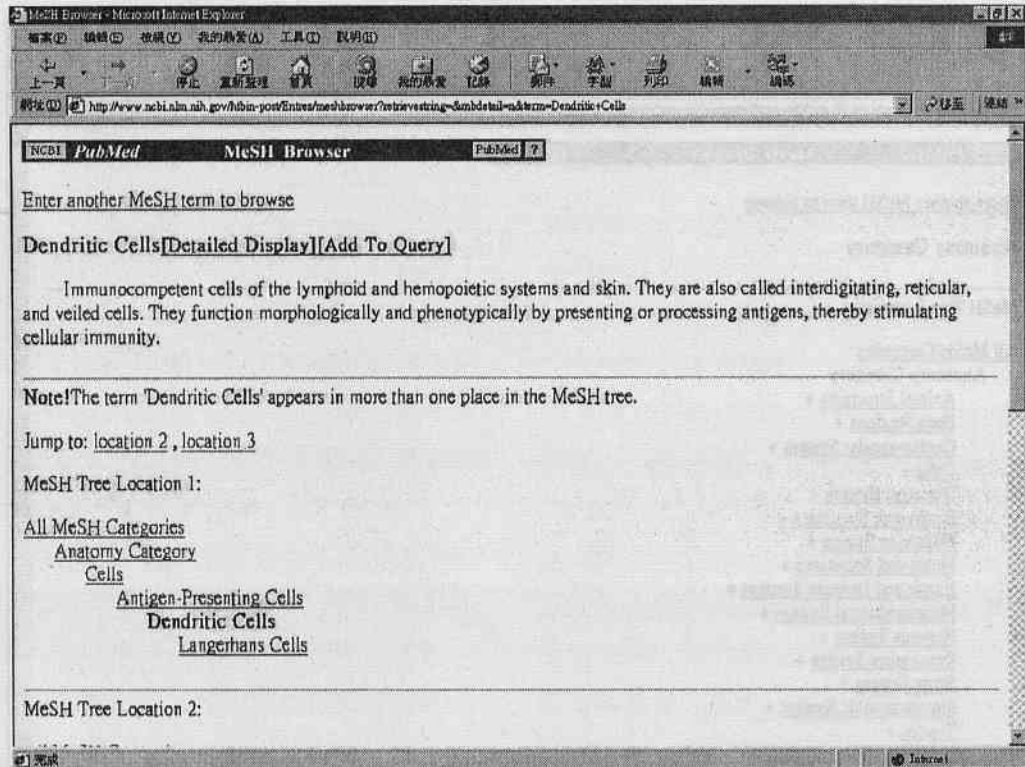


(b)



(c)





(d)

圖四、PubMed 之 MeSH 索引典瀏覽方式

無論系統內部使用靜態網頁或是動態網頁的方式呈現索引典，其協助使用者獲得系統內部知識結構的目的並無二致，只是其應用的時機並不相同。當索引典很小，其控制詞彙之間的關係不複雜，詞彙增刪的情形不多見，也沒有使用資料庫系統管理索引典時，採取靜態網頁的作法是比較方便的。相對地，如 MeSH 般複雜的控制詞彙表，擁有高達十數萬個詞彙，每年進行增刪，並使用資料庫系統管理，利用檢索的方式達成瀏覽的功能，動態網頁的作法是最可行的。在後者的作法之下，直接檢索引典自然也成為提供使用者瞭解控制詞彙的可能方式。

## 二、詞彙自動對映

資訊檢索的另一種服務型式便是所謂的檢索詞彙／控制詞彙的自動對映。對使用者而言，這是

最沒有負擔的作法，因為使用者完全不需顧慮系統到底使用何種控制詞彙。基本上，這種作法可視為查詢問句擴展（query expansion）的一種型式，但是多數資訊檢索的學者專家都忽略控制詞彙擴展的研究，而將注意力放在自由詞彙（free terms）的擴展。然而，隨著詮釋資料的研究日漸受到重視，資訊組織與整理的觀點亦影響資訊檢索的研究，控制詞彙的擴展會是未來非常重要的課題。

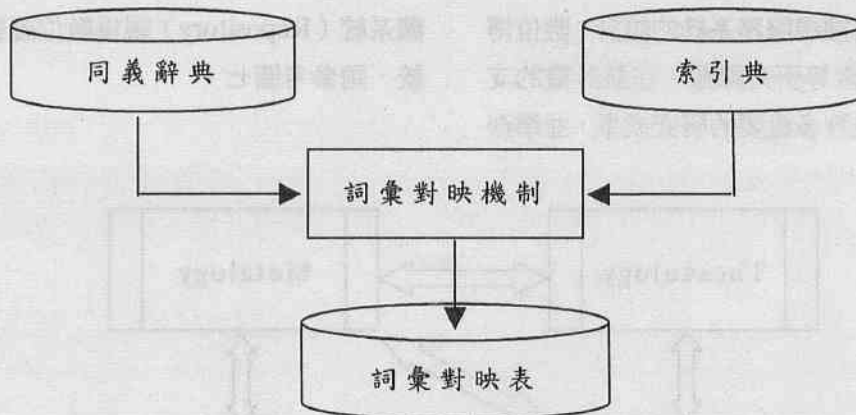
詞彙自動對映的機制可以藉由兩種方式達成，第一是透過已經加註控制詞彙的文獻資料，分析文本的用字遣詞，建構一般詞彙與控制詞彙的關係。第二是透過一份通用的辭典，建構通用辭典收錄詞彙與控制詞彙的關係。第一種作法需要大量的訓練語料，而且目前國內的資料庫真正採用索引典進行文獻描述者並不多，所以現階段並不可行；第二種作法則需要一部同義辭典，藉由這部辭典可



以連結索引典，然後利用鬆弛法 (relaxation) ⑧ 參見圖五與圖六。  
或其它方法強化各連結的權重 (weight)，請分別



圖五、文獻資料的詞彙對映



圖六、辭典與索引典的對映

如果沒有現成的同義辭典可供使用，則可以透過文件語料庫依下列程序先建立辭典：

1. 使用斷詞程式處理文件，排除停用詞。
2. 計算詞彙之頻率。
3. 設定一大小為  $n$  的文字框 (text window)，文字框內共現之詞彙，視為成對詞彙。
4. 計算成對詞彙之頻率。
5. 計算兩兩詞彙之相似度 (利用獨立詞彙頻率與成對詞彙頻率)。
6. 使用層級分類 (hierarchical clustering) 方法，建構同義辭典。

#### 肆、台大數位圖書館／博物館之權威控制系統

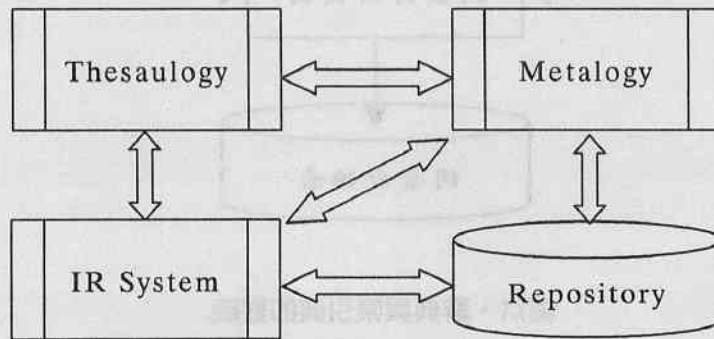
台大數位圖書館／博物館 <<http://ntudlm.ntu.edu.tw/>>計畫肇始於 1997 年，由國立臺灣大學提供數百萬的經費，並結合圖書館、人類學系、歷史學系、圖書資訊學系、資訊工程學系等十數位教授進行相關研究，首先完成的是「平埔族探源」，而後是「岸裡大社」、「伊能嘉矩」等主題。1998 年國家科學委員會有鑑於數位圖書館／博物館的研究越來越重要，因此推動「數位博



物館先導計畫」。爰此，台灣大學接續執行「淡水河溯源」與「資源組織檢索規範」二個先導計畫，其中「淡水河溯源」屬於主題型計畫，而「資源組織檢索規範」屬於支援型計畫。綜觀台灣大學近年的數位圖書館研究，主要著重於特定典藏詮釋資料的制訂<sup>⑨</sup>、使用者的研究<sup>⑩</sup>、系統的評估<sup>⑪</sup>、檢索服務系統的設計<sup>⑫</sup>、整合檢索的研究<sup>⑬</sup>等，目前均有許多重要的研究成果。「資源組織檢索規範」計畫（<<http://ross.lis.ntu.edu.tw/>>，ROSS）則進一步進行整合性研究，包括電子資源之組織與中文資源著錄格式的制定、使用者資訊尋求行為分析與使用環境之建立、索引典架構與權威資料庫建置之研究、資訊檢索及資料搜尋服務系統的設計、數位博物館/圖書館整合檢索等研究課題。在該計畫的支持之下，ROSS 發表許多重要的研究成果，並舉辦

多次學術討論會<sup>⑭</sup>，以及兩次的成果發表會。<sup>⑮</sup>本人於該計畫負責資訊檢索服務系統的設計，主要的研究工作著重於設計資訊檢索與詮釋資料著錄的服務系統。對於將數位圖書館視為是實體圖書館延伸的研究者而言，將權威檔或是標題表、索引典的觀念使用於數位圖書館系統是顯而易見的研究取向，因而本人將權威控制系統視為前述的服務子系統，提供權威控制詞彙給資訊檢索子系統與詮釋資料著錄子系統。

基本上本人將權威控制系統（稱之為 Thesaulogy）、詮釋資料著錄系統（稱之為 Metalogy）、資訊檢索系統（IR System）、資訊典藏系統（Repository）視為數位圖書館的核心子系統，請參考圖七。



圖七、數位圖書館核心子系統

Thesaulogy 扮演權威控制的角色，必須提供控制詞彙給 Metalogy 系統使用，以確定詮釋資料著錄之一致性。同時 Thesaulogy 必須將所使用的控制詞彙提供給各 IR System，以增進資料檢索的有效性。最後 Thesaulogy 也必須提供領域知識（domain knowledge），以加強使用者對數位圖書館典藏品的瞭解。在前述的考量之下，並配合當前電腦的使用環境以及網際網路已成形的標準，Thesaulogy 有如下的基本功能與規格。

1. 執行於 Windows 作業系統，透過網路與其他異質性系統互動。
2. 檔案匯格式入：可以讀取 XML 檔案（.xml）、平素檔案（Plain Text，.txt）、MS Access 檔案（.mdb）、Dbase 檔案（.dbf）。
3. 檔案匯出格式：可以轉出 HTML 檔案（.htm）、XML 檔案（.xml）、平素檔案（Plain Text，.txt）、MS Access 檔案（.mdb）、Dbase 檔案（.dbf）。



- 4.檔案儲存格式:以 MS Access 檔案格式為內部檔案的儲存格式。
- 5.檔案瀏覽型式:以首頁展現層級架構的第一層詞彙,每一詞彙皆可點選,以展現其下之第二層詞彙,同時展現詞彙的完整路徑,讓使用者藉由瀏覽方式瞭解系統的知識架構。
- 6.檔案檢索型式:使用 MS Active Server Page (ASP),讓使用者或是詮釋資料著錄者檢索權威控制詞彙。
- 7.紙本檔案呈現型式:Alphabetical Display、Rotated Display、Systematic Display。
- 8.詞彙關係的定義:除了層級關係、等同關係、關聯關係之外,權威檔案建置者可自訂詞彙關係。
- 9.詞彙的輸入與編輯:輸入詞彙,並依據定義的詞彙關係,建構輸入詞彙關係。
- 10.詞彙關係的驗證:檢查詞彙之間的關係是否相互衝突。
- 11.多語詞彙的處理:因應網際網路的應用需求

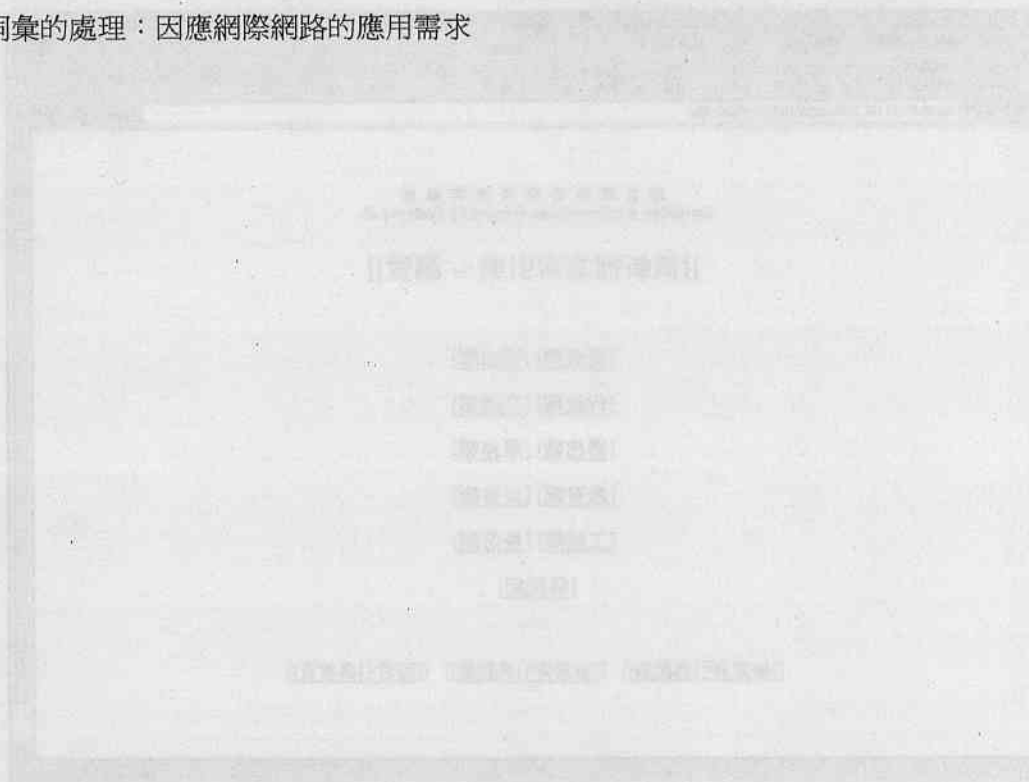
及其跨越國界的特性,必須處理多種語言的詞彙。

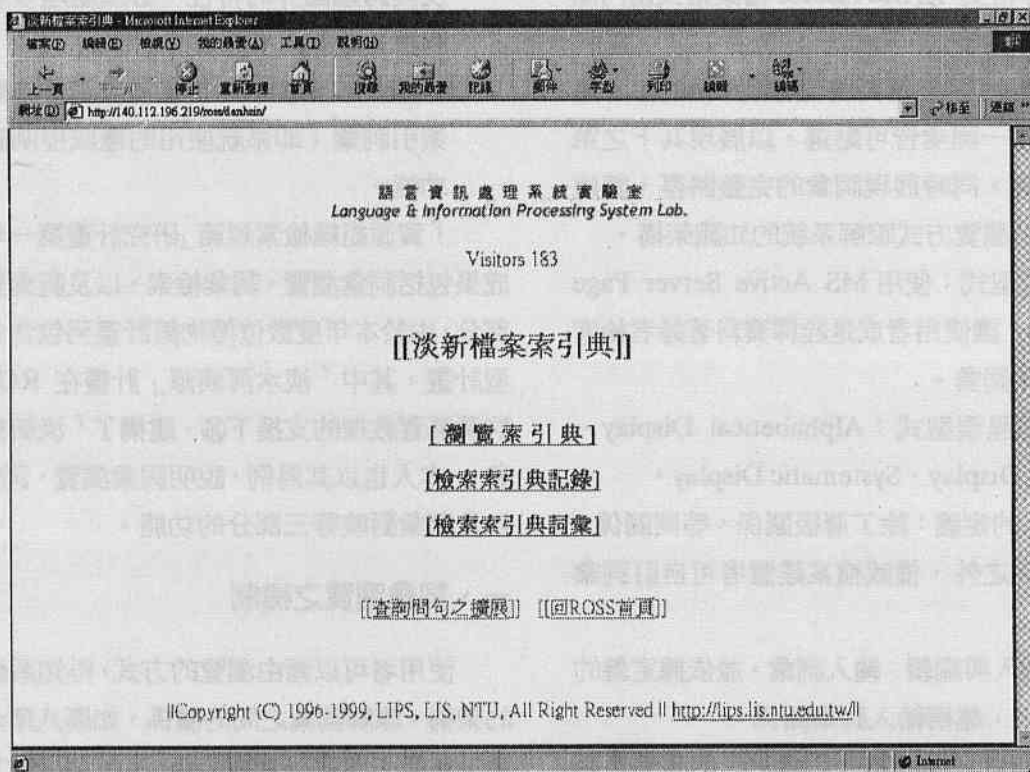
- 12.詞彙的自動對映:提供檢索詞彙自動轉換為索引詞彙(即系統使用的權威控制詞彙)的功能。

「資源組織檢索規範」研究計畫第一年的研究成果包括詞彙瀏覽、詞彙檢索、以及詞彙對映等三部分,由於本年度數位博物館計畫另包含多種主題型計畫,其中「淡水河溯源」計畫在 ROSS 研究群黃慕萱教授的支援下<sup>16</sup>,建構了「淡新檔案索引典」,本人也以其為例,說明詞彙瀏覽、詞彙檢索、以及詞彙對映等三部分的功能。

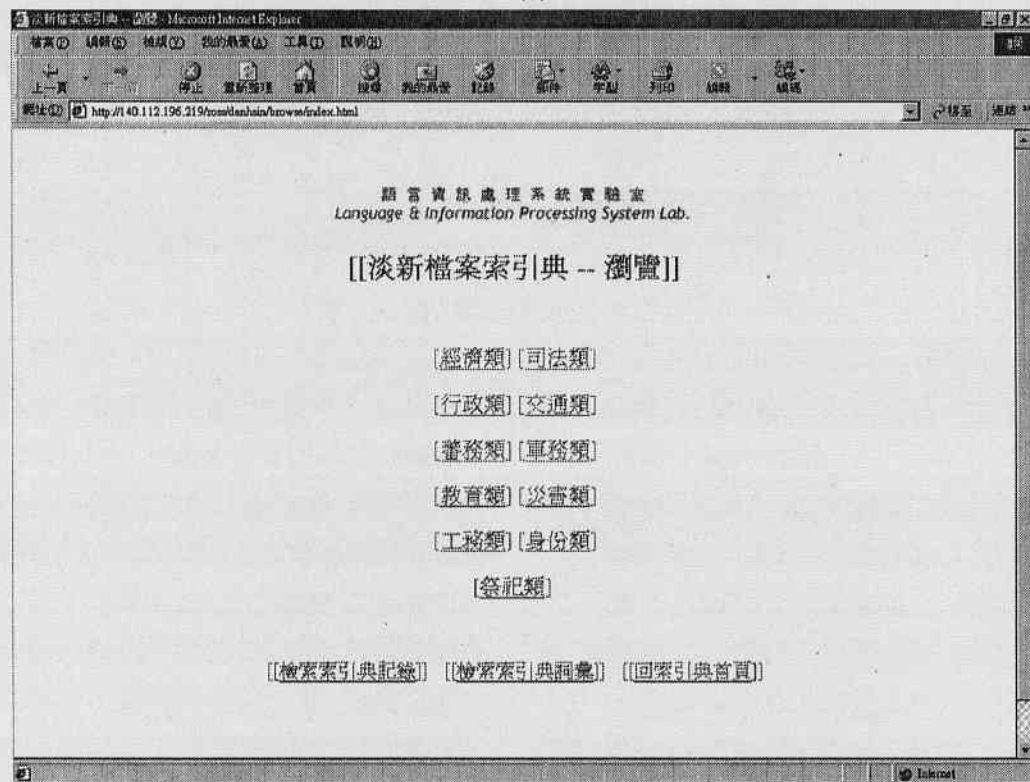
### 一、詞彙瀏覽之機制

使用者可以藉由瀏覽的方式,得知系統索引典的架構,瞭解詞彙之間的關係,如圖八所示。首頁顯示該索引典第一層的詞彙,藉由點選的方式可以瀏覽第二層的詞彙。



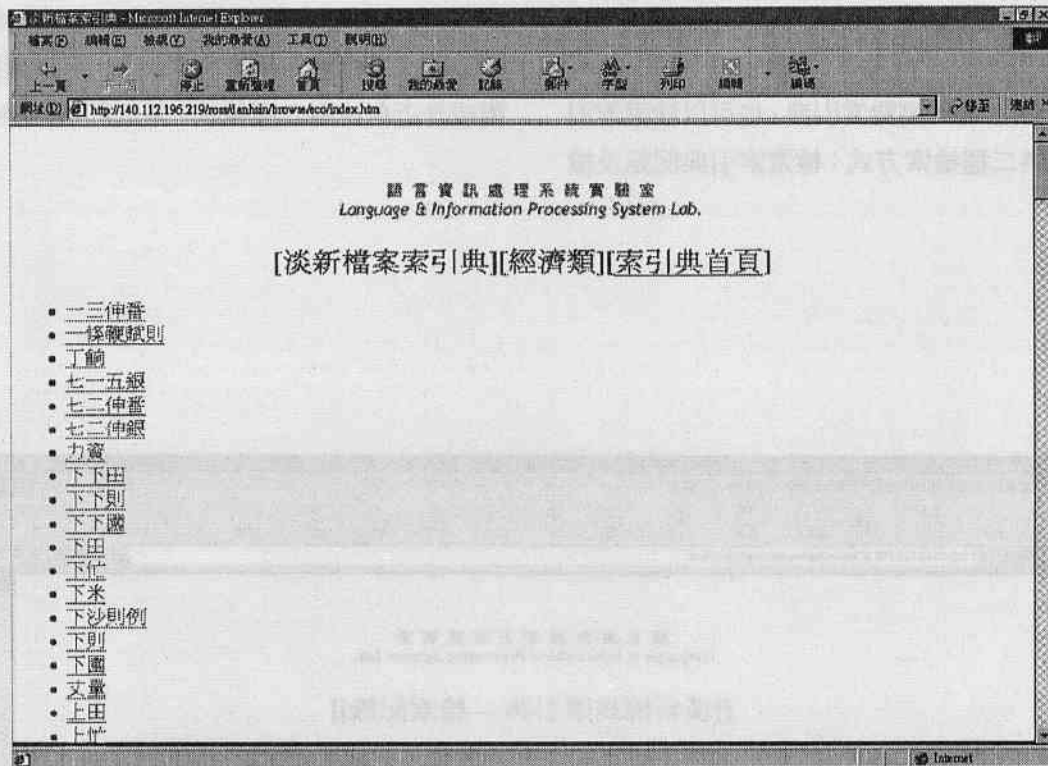


(a)

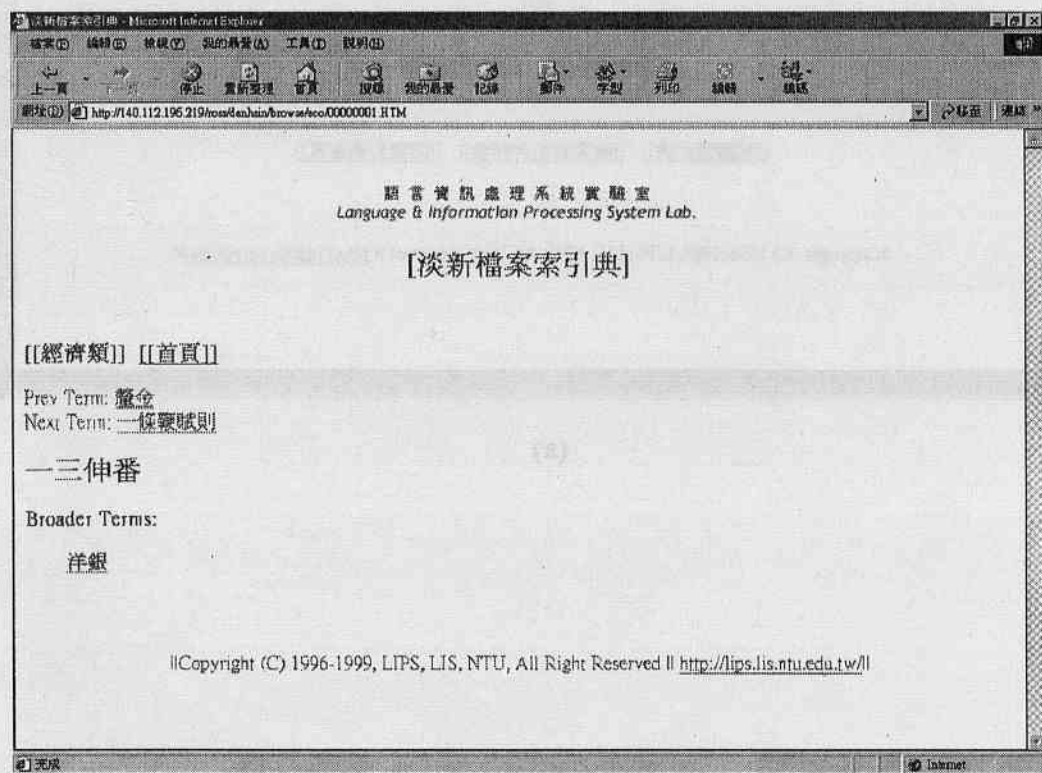


(b)





(c)



(d)

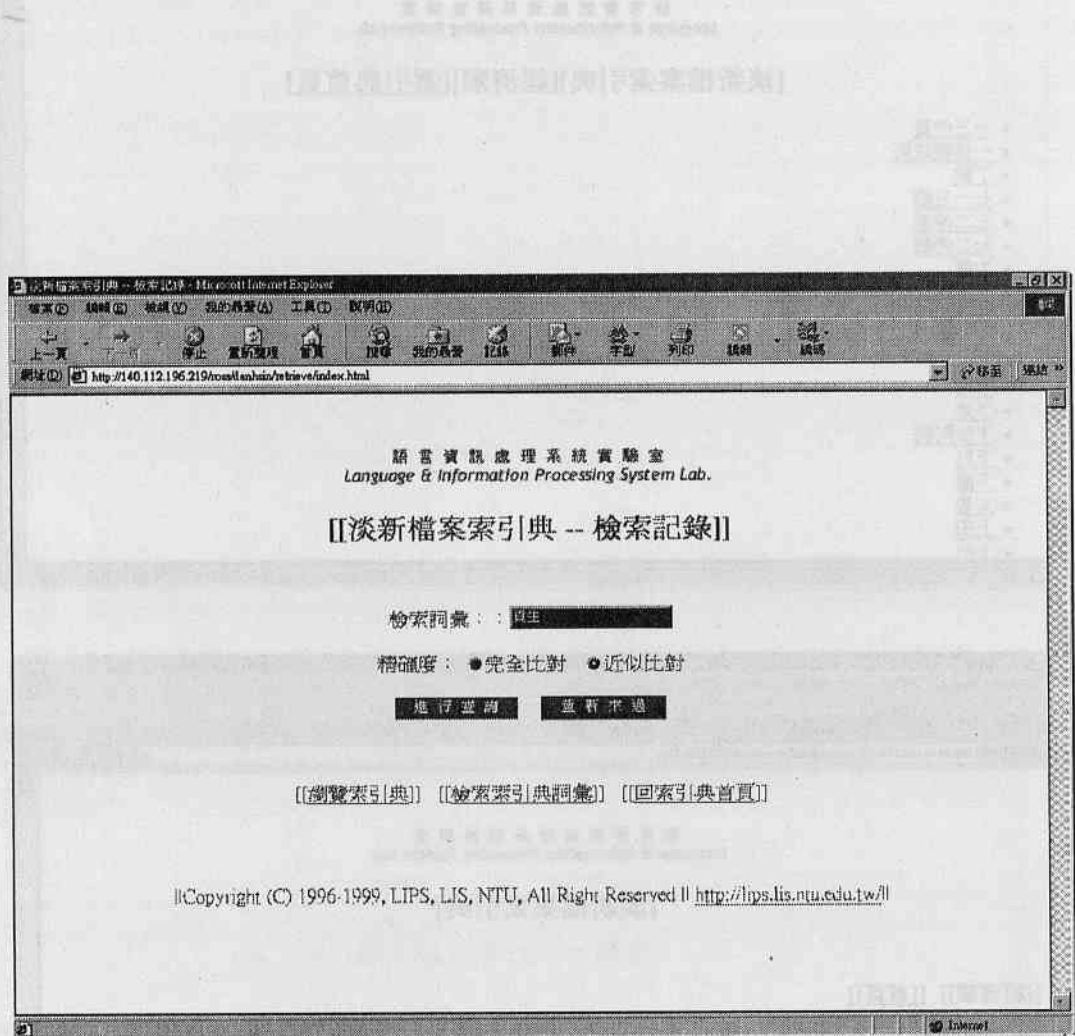
圖八、詞彙瀏覽之機制



## 二、詞彙檢索之機制

使用者除了可以瀏覽索引典，也可以檢索引典，系統提供二種檢索方式：檢索引典記錄及檢

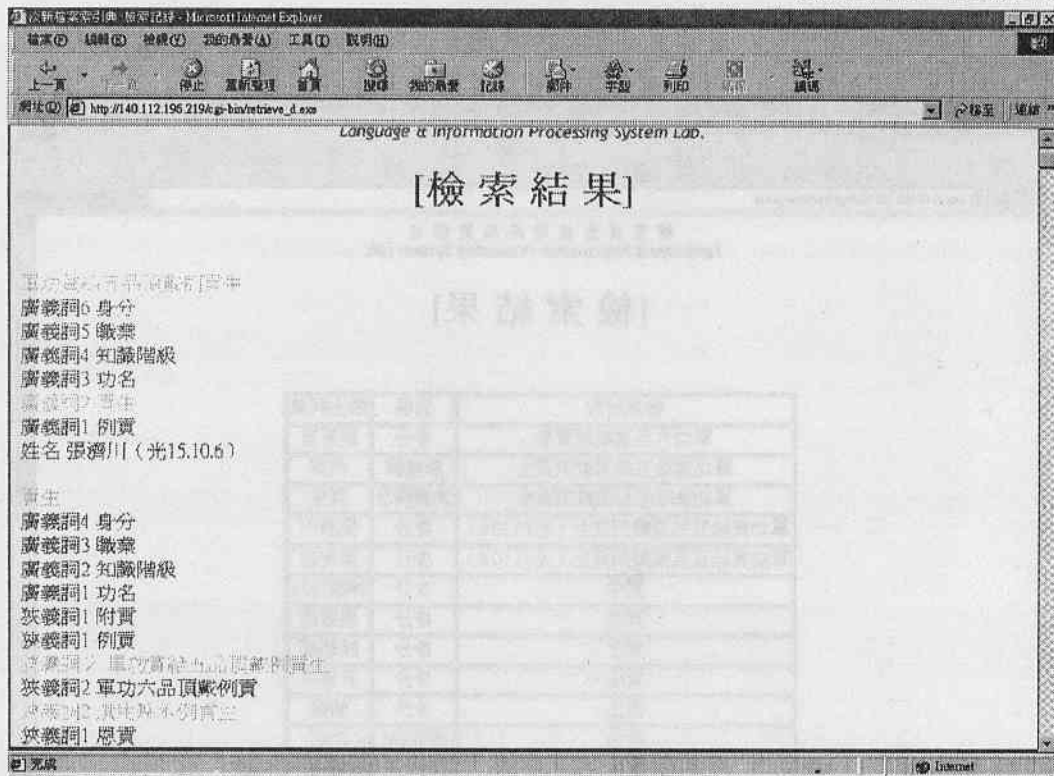
索索引典詞彙，如圖九所示。使用者於「檢索詞彙」之查詢文字框輸入欲查詢之詞彙，可以選擇完全比對或是近似比對二種模式，系統即送回檢索結果。



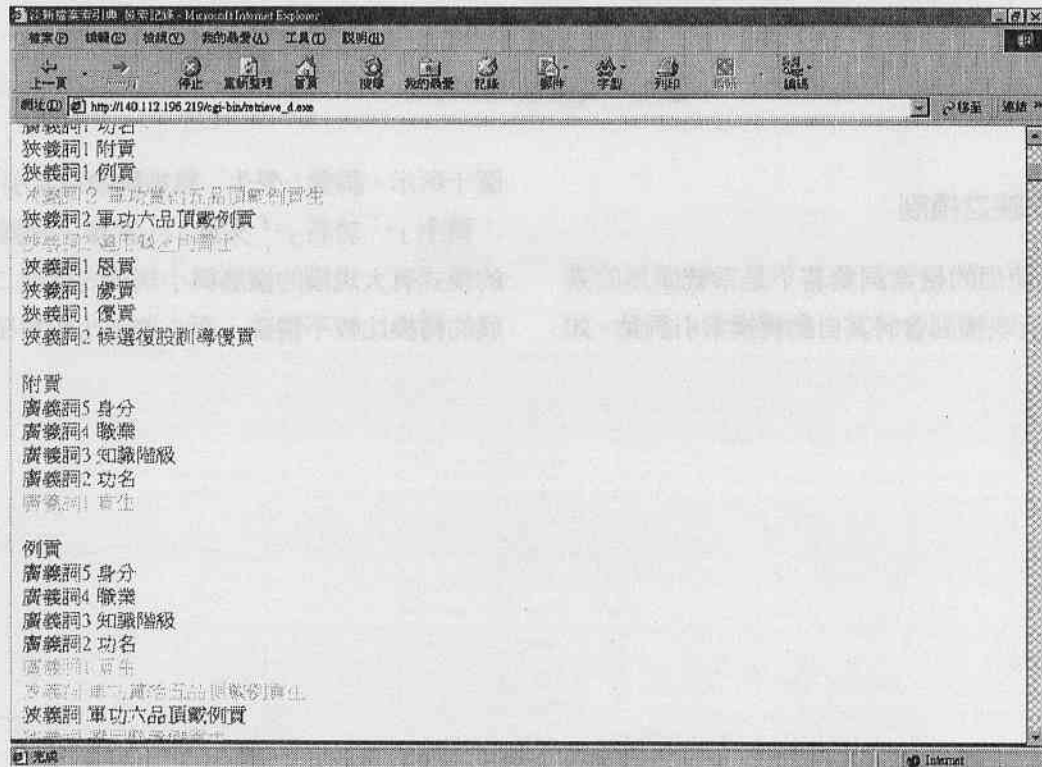
(a)





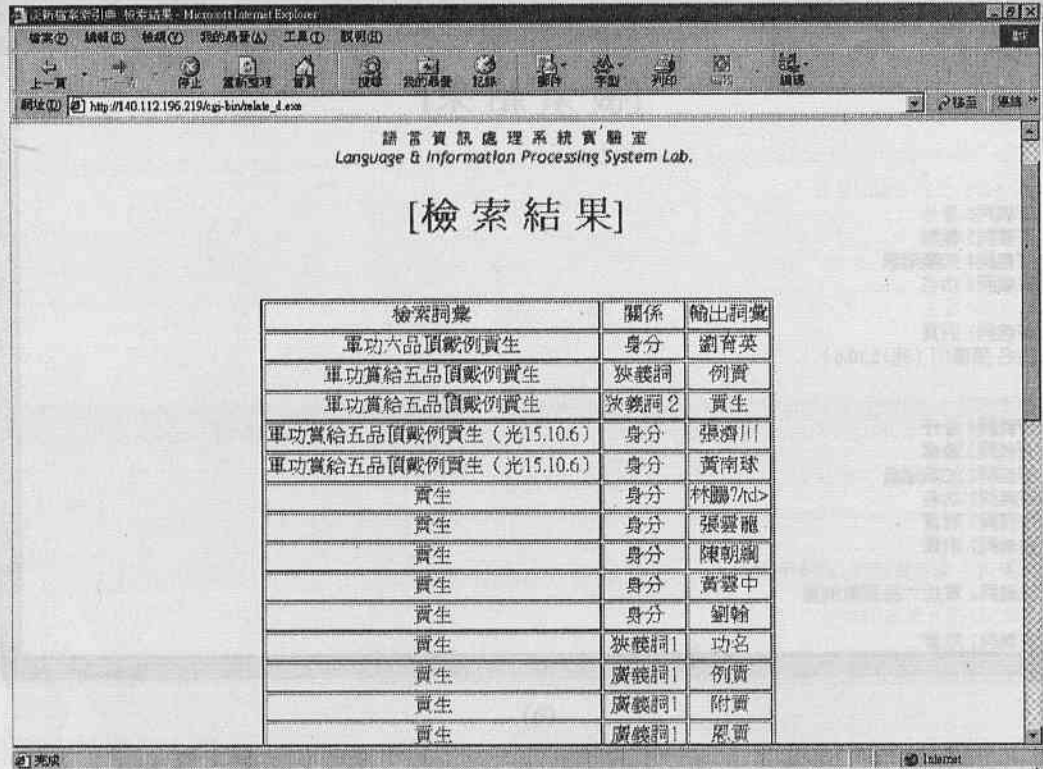


(b)



(c)





(d)

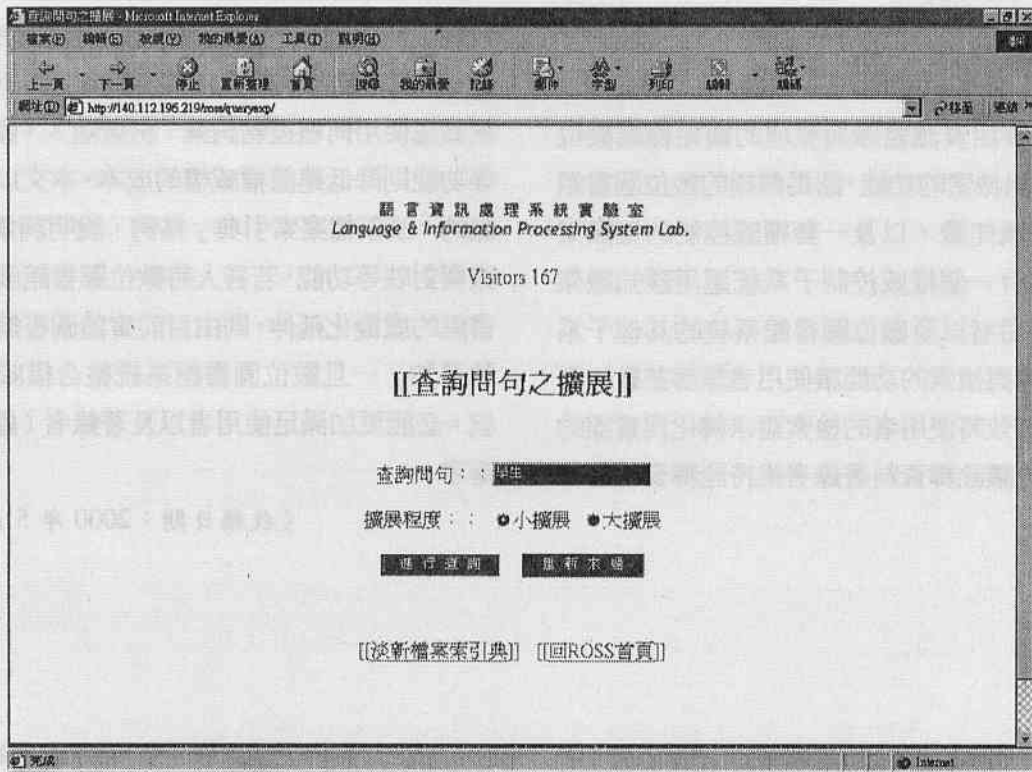
圖九、詞彙檢索之機制

### 三、詞彙對映之機制

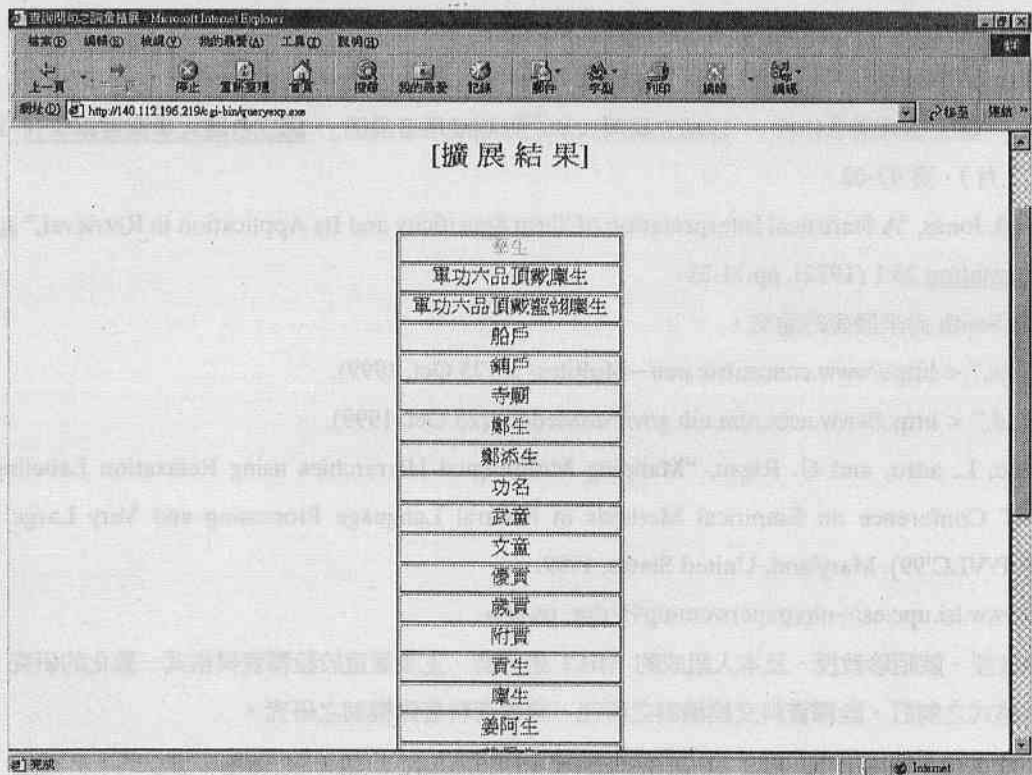
使用者使用的檢索詞彙若不是系統使用的索引詞彙，本對映機制會將其自動轉換索引詞彙。如

圖十所示，詞彙「學生」將被轉換為索引典使用的「貢生」、「功名」、「文章」、「武童」等詞彙，轉換的模式有大規模的擴展與小規模的擴展二種，大擴展的轉換比較不精確，而小擴展比較精確。





(a)



(b)

圖十、詞彙對映之機制



## 伍、結 論

本文主要由資源組織與整理的觀點檢視數位圖書館中資訊檢索的功能，認為嚴謹的數位圖書館應具有其領域知識，以及一套權威控制的知識架構，並且應有一個權威控制子系統運用該知識架構，服務使用者以及數位圖書館系統的其他子系統。詞彙瀏覽與檢索的功能讓使用者瞭解該數位圖書館系統，有效將使用者的檢索需求轉化為實際的檢索問題；也讓詮釋資料著錄者維持詮釋資料的一

致性；詞彙對映的功能讓使用者自由使用檢索詞彙，系統自動將其轉換為控制詞彙，而不必記憶系統到底使用何種控制詞彙；詞彙輸入、匯入、驗證等功能則降低建置權威檔的成本。本文以目前已上線的「淡新檔案索引典」為例，說明詞彙瀏覽、檢索與對映等功能，若吾人將數位圖書館視為實體圖書館的虛擬化延伸，則由目前實體圖書館的運轉經驗得知，一旦數位圖書館系統整合權威控制子系統，必能更加滿足使用者以及著錄者（編目員）的需求。

（收稿日期：2000年5月29日）

## 註 釋

註①：以國立臺灣大學數位圖書館研究計畫的推動而言，來自不同領域的學者專家共聚一堂，從彼此不瞭解各自的研究取向，不瞭解各自使用的語彙，到目前可以充分的意見交流，明瞭各自在數位圖書館研究的定位，其間所耗費的心力，相信各位學者專家都是心有戚戚焉。

註②：B.-A. Liptez, "Catalog Use in a Large Research Library," *Library Quarterly* 42 (1972), pp.129-139.

註③：黃慕萱，「線上檢索指令分析 -- 以國立臺灣大學之終端使用者為例」，*國立台灣大學圖書館學刊* 11（民國85年12月），頁47-62。

註④：K. Sparck Jones, "A Statistical Interpretation of Term Specificity and Its Application in Retrieval," *Journal of Documentation* 28:1 (1972), pp.11-21.

註⑤：Alastair Smith 尚未發表的論文。

註⑥："MultiTes," < <http://www.concentric.net/~Multites/> > (25 Oct. 1999).

註⑦："PubMed," < <http://www.ncbi.nlm.nih.gov/PubMed/> > (25 Oct. 1999).

註⑧：J. Daude, L. adro, and G. Rigau, "Mapping Multilingual Hierarchies using Relaxation Labeling," Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC'99). Maryland, United States, 1999.

<<http://ww.lsi.upc.es/~nlp/papers/emnlp99-dpr.ps.gz>>.

註⑨：陳雪華教授、陳昭珍教授、及本人組成的 MICI 研究群，主要著重於詮釋資料格式一致化的研究、中文詮釋資料格式之制訂、詮釋資料交換機制之研究、詮釋資料著錄機制之研究。

註⑩：林珊如教授指導的使用者行為、資訊尋求行為等相關研究。

註⑪：黃慕萱教授、吳美美教授、林珊如教授指導的系統評估相關研究。

註⑫：本人的 LIPS 實驗室主要從事資訊檢索、自然語言處理、數位圖書館相關研究，包括控制詞彙之自動索引，



權威控制系統之研究，資訊檢索系統測試集之研究 <http://lips.lis.ntu.edu.tw/>。

註⑬：陳昭珍教授指導的分散式數位圖書館的整合檢索相關研究。

註⑭：如民國 87 年 10 月 8 日舉辦的「建構使用者導向的數位圖書館」專題演講，民國 87 年 11 月 21 日於國立臺灣大學舉辦的「資源組織與檢索之規範」討論會，民國 88 年 6 月 22 日於國立臺灣大學舉辦的「數位博物館資料庫架構與著錄實務專業課程」。

註⑮：民國 88 年 6 月 22 日在國立臺灣大學，民國 88 年 10 月 23 日在中央研究院。

註⑯：黃慕萱教授在「資源組織與檢索之規範」計畫中負責索引典架構與權威資料庫建置之研究，為配合主題性研究計畫「淡水河溯源」，建置了「淡新檔案索引典」。相關資料請參考：

黃邦欣，「淡新檔案行政篇索引典建構之研究」(碩士論文，國立臺灣大學圖書資訊學研究所，民國 88 年 7 月)。

【結 語】

Abstract text, likely a summary or conclusion of the research, discussing the design of authority control systems in digital libraries.

Abstract text, likely a summary or conclusion of the research, discussing the design of authority control systems in digital libraries.

Keywords

