

第一章 緒論

本章旨在闡述本論文之研究背景、研究動機、研究目的與問題、研究方法與步驟、研究範圍與限制及名詞解釋，因此以下共分六節詳述之。

第一節 研究背景

由於網際網路的普及，產生了大量的電子資源，這些電子資源被快速的生產、傳播、交換和使用，造成了今天以網路傳輸資訊的生活型態。電子資源越豐富，越容易被取用，將可擴大個人學習的視野與速度，但資訊越分散、越動態，使用者的資訊檢索負擔也就越沈重。如何將這些龐大的資訊量，以整合性、系統化的方式提供使用者檢索，也成為圖書館資訊服務的重要課題。

就目前的網路環境而言，使用者在檢索資訊時，最大的問題即是資料分散、許多資料庫的檢索介面不一，造成使用者檢索困難，陳昭珍(2000)即指出「康乃爾大學圖書館在一項使用者調查當中發現，由於資料庫的種類愈來愈多，使用者往往須浪費時間於資料庫的選擇與專指性的資訊檢索介面之學習上，因此共通的檢索介面成為使用者共同的心聲。」

此外，黃華明(2000)在其碩士論文中也提到，使用者在面對眾多的資料庫時，會希望可以：

1. 學習單一的操作介面；
2. 能同時檢索多個資料庫，而不必重複輸入檢索步驟；
3. 能夠針對同一個檢索操作做深入的檢索，如重複使用該結果；
4. 可以同時對不同資料庫的檢索結果進行排序。

Walker(2001)指出，早在1945年Vanner Bush就提出的「關聯性索引」(Associate Indexing)概念，希望藉由與主題之間相關的屬性連結，提供使用者資訊，而非經由一一的檢索才可獲得所需的資訊。

然而以目前資料庫的設計而言，不同資料有不同的資料屬性，不同的資料組織方式，因此所建立的資料庫系統，也會產生不同的資料結構、控制詞彙，而且不同的資料格式會導致不同的資訊存取方式及索引方法；當然，使用者介面的檢索功能就更不可能一樣(Cleveland, 1998；陳昭珍，2000)。具體而言異質性(Heterogeneous)與分散性(Distributed)的資訊環境，會有下列各種狀況存在(施保旭，1997；陳昭珍，2000；Paepcke, Chang, García-Molina & Winograd, 1998)：

1. 資料屬性的異質性：資料屬性的異質有文字、聲音、影像、數值、程式碼，所以資料庫管理系統的設計會針對各個資料模式做不同的存取處理方式，而此方面的異質將會產生資料管理模式的異質性。

因為所要連結的資料是被各個不同的資料庫系統，針對其不同的屬性內容以不同的資料管理模式儲存在各個不同的資料庫系統裡，因此所面對資料管理模式當然也就不相同；

2. 資料庫管理系統廠牌的異質性：即使是同一種資料屬性，因為使用的是不同資料庫管理系統產品，則其提供功能的使用亦有出入；
3. 資料定義與詮釋的異質性：因為不同的資料屬性有其不同的詮釋方法，所以資料庫管理系統的設計者會依此來設計其資料庫的綱要(Schema)乃至於表格(Table)的切割亦會有所不同；
4. 網路協定與系統之異質性：各個資料庫系統廠商可能因設備特性或採購或管理體制等因素，因此採用的網路協定與網路系統當然也不一樣如：TCP/IP，Novell 等；
5. 檢索語言、字元集編碼方式的異質性；
6. 語意異質(Semantic Heterogeneity)：包括因為詮釋資料、屬性命名、資料儲存格式的不同，和其在進行整合檢索時所遇到的語意流失及衝突值等問題。

除了上述的異質性外，還要面對的是各種不同的硬體與作業平臺之技術；而整合檢索的重要目標之一，就在整合各資料庫系統所採取的資訊組織規範，以提供使用者具有一致性的服務，滿足使用者更多的資訊需求(徐代昕，2003)。

而國家圖書館(以下簡稱國圖)居於領導全國圖書館之角色，對於資訊系統及資料數位化之工作向來也走在時代的尖端。目前國圖之資訊系統及資料庫主要可分為三大部份：

1. 書目資料庫系統；
2. 國圖自建的數位化資訊系統；
3. 國圖引進之國內外光碟或遠端存取之資料庫。

國圖之書目系統包括全國圖書資訊網、全國新書資訊網、館藏目錄等；第二部份是近年來國家圖書館積極從事的數位化或自建資料庫工作，且已頗有成效；然而這些系統歷年來因為開發背景與經費的關係，多以各單位個自開發為主，系統與系統間無法整合，導至使用者要檢索很多系統才能查到他所需要的資料。第三部份是國圖引進的光碟或遠端取用之資料庫，這個部份主要包含書目性資料庫、電子期刊、電子工具書、報紙全文、文史資料庫、自然科學資料庫等。

除了國圖內部讀者使用外，九十二年度政府推動的公共建設之公共圖書館強

化計畫中，國圖也要為公共圖書館購買共用資料庫，該計畫並要求國圖應結合學者專家規畫國內資料庫整合檢索協定，並建置檢索系統，要求各資料庫廠商配合，以提升資料檢索的親和及便利性。

因此，國圖將再建立『華文知識入口網站』，最主要的目的即在整合檢索這些國內的中文資料庫。

第二節 研究動機

過去圖書館界也曾解決過各資料庫系統整合檢索的問題，並發展 Z39.50 資訊檢索協訂，不過正如 Caplan (2003)所指出的，圖書館的自動化系統乃以 MARC 為資訊組織的傳統圖書館，因為其資料格式為 MARC 21、編目時採取通用的規則且儲存記錄的檔案系統(File System)是同質性(Homogeneous)的，因此當利用 Z39.50 與各圖書館 OPAC 間產生的虛擬聯合目錄(Virtual Union Catalog)，不但可以讓使用者透過此一虛擬聯合目錄，同時檢索多個書目資料庫，而且不需要另外改變圖書館自動化系統間的組織與產生的程式(Coyle, 2000；黃華明，2000)。

但是數位圖書館時代所面臨的問題，已不再只是以 MARC 建檔的資料庫系統，而是以各種不同的詮釋資料格式建檔的資料庫系統，圖書館所建立的、或所購買的資料庫，往往多達數十種，而這些系統皆各有其資訊組織與檢索方式。

在此複雜多元的資訊服務環境中，圖書館需籌措經費購買各種資料庫以充實館藏，提供快速的服務，然而一方面也面臨這些花了龐大經費買來的資料庫使用量非常少的窘境。資料庫使用量不高，其中原因可能有多種如：學術研究風氣不佳(否則國內應產生像 Van de Sompel 這樣的人，感覺到使用上的不便)、外文閱讀能力不佳，但也可能是因為資料庫太多，不知用那一個，以及不知如何使用。

如果再加上有些資料要在圖書館的書架上取得、有些要透過館際互借、有些要透過文獻傳遞等這麼多元的方式，則一個可以整合各種資源的機制實在非常迫切；目前各種國內數位典藏計畫及商業性資料庫也越來越多，如何建立比較好用的服務環境，實在值得探討。

要整合這麼多種不同資料結構的資料庫系統，在國內仍是個創舉，而『華文知識入口網站』擬以開放檔案詮釋資料擷取協定(Open Archives Initiative Protocol for Metadata Harvesting, 以下簡稱 OAI-PMH)來達到整合檢索的目的，其成效為何值得探討；尤其各資料庫系統的資訊組織模式不同，是否會影響整合檢索的執行與成效，都是本論文極欲探討的問題。

因此本論文希望藉由以上問題的探討，使得以確實力行數位圖書館的一個重要目標：整合散佈且獨立運作的資料庫系統與資源，讓使用者僅需透過單一系統或介面即可取得所需的資訊，以得到透通性(Transparency)的資訊服務(Cleveland, 1998)，使其資訊需求獲得滿足。

第三節 研究目的與問題

本論文之研究目的在於瞭解各資料庫系統間整合檢索會面臨到的問題，如需採用的技術，以及各系統不同之資訊屬性與資訊組織模式對此整合檢索成效的影響為何；具體而言，本論文之研究目的如下：

1. 瞭解面對如此複雜的各資料庫系統應以何種技術、標準達到整合檢索的目的；
2. 探討各資料庫系統因資訊組織、屬性命名和資料儲存格式的不同，其資訊組織模式的互通模式應為何，方能使整合檢索的效能表現最佳化。

因此根據研究目的，本論文的問題為：

1. 各資料庫系統利用 OAI-PMH 建置的聯合目錄之內涵為何？
2. 各資料庫系統在資訊組織模式方面的互通共識為何？
3. 為使整合檢索的效能最佳化，各資料庫系統所會面對的問題為何？

第四節 研究方法與步驟

本論文所探討的是各資料庫系統在透過整合檢索的機制之下，其不同資訊組織模式在技術、內容層面所遇到的異質性問題為何，及其因應之道，又為何方能使整合檢索的效能良善；因此採用的研究方法說明如下：

1. 文獻分析法：先蒐集、整理相關文獻瞭解數位圖書館整合檢索協定的內涵如：OAI-PMH、Z39.50、OpenURL 等，及國內外相關計畫實作的情形。

而後將蒐集的文獻作為文獻分析的工具，進行分析各資料庫系統間在運用 OAI-PMH 協定建立一集中式聯合目錄，需在技術、內容層面取得的互通共識為何，方得以便利使用者進行整合檢索；

2. 訪談法：根據文獻分析所得，擬定『華文知識入口網站』之資料庫系統提供廠商的訪談大綱。

訪談對象為『華文知識入口網站』的資料庫系統提供廠商；訪談目的在

瞭解其如何執行詮釋資料的對映工作，方便整合檢索之效能有良善的表現，從而探討各資料庫系統間的資訊組織需取得的互通共識為何，方能得到令使用者滿意的整合檢索效能；

3. 內容分析法：針對『華文知識入口網站』，進行內容分析，選擇分析單位和建立分析的類目，將記錄的資料進行內容編碼，以瞭解『華文知識入口網站』在整合檢索效能方面的表現情形。

接著根據記錄資料的內容進行分析與比較，從而評估『華文知識入口網站』中各資料庫系統的資訊組織對整合檢索的效能所產生的影響。

第五節 研究範圍與限制

本論文因時間、人力因素，有以下之限制：

1. 本論文所指的整合檢索協定：以 OAI-PMH 為主要的研究範圍；而所參考依據的協定內容以 2003 年 11 月底以前公佈在官方網站的最新版本為主，以後再經修改的內容，本論文將不再納入探討；
2. 本論文所指的資訊組織模式：僅以『華文知識入口網站』在 2003 年 11 月底以前所收錄的資料庫類型為研究範圍；
3. 本論文之文獻：以中、英文文獻為範圍，不涉及其他語文的探討。

第六節 名詞解釋

一、『華文知識入口網站』

『華文知識入口網站』是國圖在圖書館事業發展三年計畫裡其中的一項目；國圖希望透過建置『華文知識入口網站』，提供使用者藉由 Single Sign On 及單一的檢索與瀏覽介面找到所需資源，而這些資訊資源在本論文中是指收錄於國圖網站中的公共圖書館共用資料庫(http://dblink.ncl.edu.tw/common_db.asp)。

二、資訊組織模式

在資訊量龐大的網路環境裡，須做好資訊組織的工作，才得以快速且正確地找尋資料，而這也正是詮釋資料概念出現的原因；因為詮釋資料是指在經資訊組織後，所定出對資訊屬性、內容、特色的描述，可用來協助資訊的辨識、描述與指示其位置的任何資料(柯皓仁、黃夙賢與楊維邦，2001；王麗蕉 1999)。

而由於使用者需求與資訊特質的不同，因此在做資訊組織時所採用描述資料

的方法也不盡相同，使之所產生的資訊組織模式也就不同，亦即所產生的詮釋資料也就不同，如描述政府資訊時是採用 GILS(Government Information Locator Service，以下簡稱 GILS)、描述地理資料時即採用 FGDC(Federal Geographic Data Committee，以下簡稱 FGDC)而描述檔案資料時採用 EAD(Encoded Archival Description，以下簡稱 EAD)(陳昭珍，1998)。

是以本論文所指的資訊組織模式等同於詮釋資料一詞的概念。

三、 整合檢索

整合檢索就是利用共通檢索介面把分散各處的資料庫系統加以整合，令使用者不需要逐一連結資料庫系統與學習使用其各別不同的檢索功能。

亦即其主要概念為利用共通檢索介面與操作指令，讓使用者可以透過此單一檢索介面(而此單一介面可以利用建置聯合目錄的方式來達成)就可取得需要的所有資訊，以滿足現代人對資訊的大量需求(陳昭珍，2000；柯皓仁等，2001；林大彰，1998；徐代昕，2003)。

四、 數位圖書館

文獻中關於數位圖書館一詞的解釋繁多，而根據研究動機發現與Cleveland (1998)在「Digital Libraries: Definition, Issue and Challenges」一文中指出：「數位圖書館是各資料庫系統透過網路整合在一介面下，而非單一完整的數位系統。」的概念不謀而合。

是以在本論文中的數位圖書館一詞乃是指：「由各資料庫系統組成，用以提供並延伸傳統圖書館各種服務的分散式技術環境。」