

資訊搜尋與異質資料庫整合查詢系統 建置之初探

羅思嘉

中興大學圖書資訊學研究所副教授

壹、前言

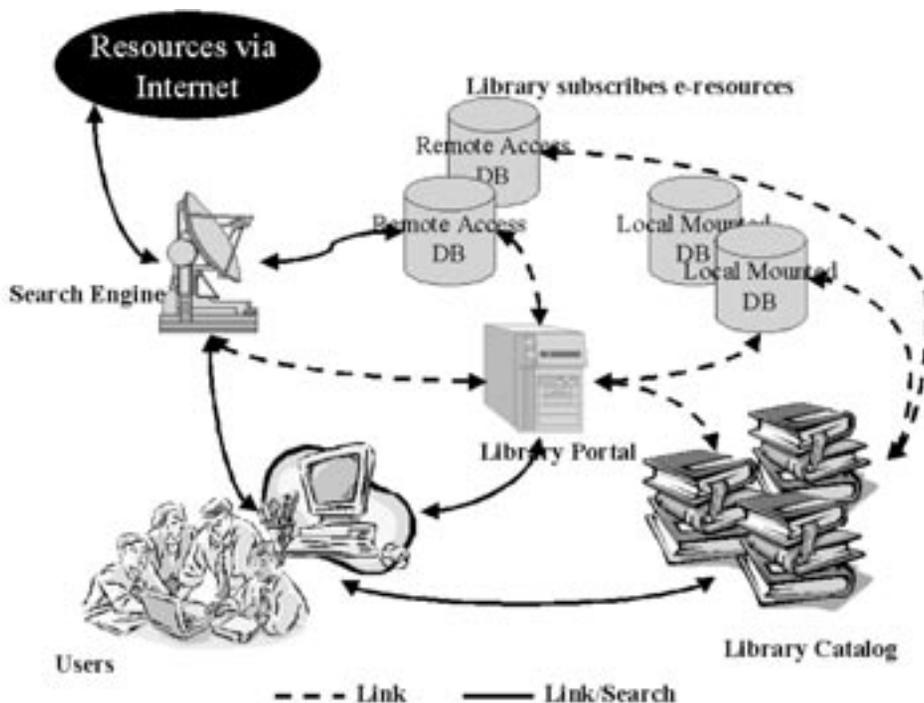
資訊科技隨著網路技術與資訊數位化的發展，讓圖書館得以擴展與過去不同深度與廣度的資訊服務。不單是資訊搜尋與相關服務可以透過網路完成，資訊的數位化使得使用者在資訊取得與利用上更為便利。目前圖書館所提供的資訊搜尋介面不再侷限於圖書館館藏資源搜尋的網路化，不同類型的資料庫及資訊系統在現今圖書館的資訊服務上均扮演重要角色，讓資訊搜尋的範圍從圖書館館藏資源書目資訊的搜尋，擴及到內容以及館藏以外的資源。數位全文，包括電子書、電子期刊、各種形式的參考工具，讓資訊的取得與利用更加便利。同時，網際網路介面提供個人與

單位機構一個便利的出版平台，讓資訊的產出不再受限於過去既定的出版機制，依照個人或單位機構的需求，都可以透過網際網路介面發表各型各樣的文獻，透過此一出版型態產出的資訊亦逐漸成為資訊使用者另一豐富的資訊來源。不單是資訊產出形式有所改變，網路資源搜尋引擎的發展與便利使用，更是在有形、無形間取代圖書館所提供的資訊搜尋介面，成為多數資訊使用者主要的搜尋窗口，其中還包括相當比例的使用者充分依賴網路資源搜尋引擎所提供的資訊搜尋服務，一旦脫離搜尋引擎的服務範圍，資訊的檢索成為一個無法掌握的黑洞。由於網路資源搜尋引擎使用的便利，讓資訊的發掘對多數的使用

者似乎成為一件輕而易舉的事情。但是，是否真是如此？而圖書館在資訊服務究竟應扮演什麼樣的角色，更是提供資訊服務者所需要思考的問題。

目前圖書館所提供的一個典型資訊搜尋介面，其中提供的資源包括圖書館各種資訊服務的訊息、圖書館館藏目錄、連結至各項圖書館所提供之電子資源入口介面、與圖書館服務有關之網路資源以及網路資源搜尋引擎（圖一）。從資訊使用者的角度來看，其主要依賴的檢索核心可能是搜尋引擎，而非圖書館所提供的資訊入口；而從圖書館的角度來看，圖書館則

是希望館藏目錄以及所提供的入口平台能成為讀者搜尋資訊的核心。但是在便利使用及必要性上，便利使用往往成為主要的決定因素。圖書館所提供的檢索介面，使用者若需要充分利用介面所包含的各項資源，其需要克服的問題與掌握的訊息包括介面中所涵蓋的各種資源及其屬性、各種資源連線的方式以及基本登入之必要訊息，通過資源使用之身分認證機制後，使用者仍需進一步瞭解各資源所提供的資訊檢索介面與檢索方法，再透過不同介面輸入搜尋資訊指令、自行整合所獲得的檢索資訊。



圖一：圖書館資訊入口架構示意圖



面對技術與使用行為的改變，圖書館一方面要思考與不同模式資訊服務的競爭過程中，所提供的資訊服務如何具有其一定價值，並取得使用者對這樣價值的認同，同時圖書館也必須思考如何解決現有的資訊利用問題，讓資訊使用者可以容易地掌握其所需要的資訊。事實上類似的問題在提供資訊服務的過程不斷出現，對於這樣便利方法的追求一直是圖書館努力的目標。1945年 Vannevar Bush 在其 *As We May Think* 一文中所提出 Memex 的概念，其所刻劃的即是透過一個便利使用的單一介面--Memex，資訊使用者可以查詢與獲得所需要的資訊（Bush, 1945）。館藏目錄、Gopher 的選項畫面、圖書館網頁，都是在這樣概念下所形成的產品，而隨著資訊處理技術與服務模式的變化，既有的 Memex 形式顯然已不符合使用者的期待與提供服務者的要求。下一代的 Memex 又是何種樣貌？目前被熱烈討論的異質資料庫整合檢索系統是否就是下一代的 Memex？本文將從異質資料庫整合查詢系統的本質討論整合查詢系統應有的功能，並進一步說明異質資料庫整合檢索系統評選與建置的步驟與其中需要注意之事項。

貳、異質資料庫整合檢索系統之基本概念

異質資料庫整合搜尋的概念並非是近兩年的產物，對於這樣的資訊需求也不是近幾年才出現，如同前言所述，在 1940 年代即有對檢索效能的要求，只是受限於技術的發展，能夠落實的程度有限。隨著資訊技術的發展，使用者及圖書館有越來越多的檢索工具可以選擇，並提供整合程度高的檢索服務。與整合檢索相關的概念包括在 1970 年代開始發展的 Z39.50 及 1994 年與網路資源搜尋有關的 Metacrawler，均是架構於透過一個共通的搜尋介面，檢索異質資訊來源概念所形成的資訊檢索通訊協定或程式工具。這類型的通訊協定與程式工具的設計目的，都是希望能夠透過一個共通的檢索介面，降低資訊使用者檢索多個資訊來源、重覆輸入多次相同檢索策略的繁瑣。Z39.50 通訊協定雖然在 1970 年代已經發展出來，經過二次改版，直到 1990 年代才運用於異質圖書館館藏目錄之間的檢索，不過成效有限。除了運用 Z39.50 進行系統之間資訊的傳輸與交換，亦有不同的 API 程式從類似概念出發，企圖透過檢索介面的分析、指令的包裝與資料的重整達到館藏目錄檢索的整合。Metacrawler 則是網路資源搜尋引擎出現後，為了滿足同時查詢多個搜尋引擎需求而出現的工具，設計目



的同樣是希望經由單一介面，一次輸入檢索指令同時檢索多個搜尋引擎所提供的網路資源。不論是 Z39.50、API 程式或者是 Metacrawler，檢索範圍多限於單一型態資源，整合程度有限。近幾年發展的 OpenURL 以及相關的程式工具提高進一步改變整合檢索成效的可能性。

近幾年出現的幾個相關名詞以不同的訴求與說法反應異質資料庫整合檢索的概念，包括 Metasearch、Parallel Search、Federated Search、Broadcast Search、Cross-Database Search、Common User-Interface Search、Search Portal 等 (Fryer, 2004)。上述名詞的運用分別從不同角度強調解決查詢異質資料來源的概念。Metasearch 沿用 Metacrawler 的概念，主要是指涉及搜尋多種資源的功能；Parallel Search 強調同時檢索的處理程度；Federated Search 強調檢索不同資訊來源的整合程度；Broadcast Search 從將檢索指令傳輸至不同資訊來源解釋整合檢索的概念；Cross-Database Search 強調跨越不同的檢索平台；Common User-Interface Search 則以共同的檢索介面為主要訴求；Search Portal 則是從其想要達到的目標強調這樣的資訊檢索是取得資訊的搜尋入口，透過該入口使用者可以獲得需要的訊息。上述名詞雖從不同的角度強調

經由單一介面，以單一檢索指令同時檢索不同資源平台所能夠提供的資源，但是各種名詞強調的面向所顯示的功能都是目前異質資料庫整合檢索所需具備的功能。除了上述所提及檢索各種資訊資源的功能，亦有從廣義角度來看整合檢索，認為異質資料庫整合檢索機制應包括使用者個人的相關資訊，以大學圖書館的服務範圍來說，異質資料庫整合檢索系統除了檢索圖書館所提供的各種資訊以及網路相關資源，系統的檢索範圍亦包括使用者在學校的各項紀錄 (Carden, 2004)。以目前的發展趨勢來看，還是以檢索資訊資源的功能為發展主要訴求。

參、異質資料庫整合檢索系統之功能需求

異質資料庫整合檢索系統要能夠達到從多種資源檢索資訊的功能，必須處理認證、資訊對應與檢索指令轉譯等問題，才能透過整合檢索介面的系統機制處理不同資訊資源的使用與檢索要求，使其檢索與使用能夠簡化，降低使用的困難程度。但過程簡化並不同功能的簡單化，在使用上，系統要盡可能操作簡易，但是通訊與檢索方面，則必須具有處理與不同資訊系統連線檢索的功能。另外，透過介面處理的問題包括處理不同系統的認證機制、提供單一介面處理並轉換為不同系統可辨識



的檢索指令、處理轉換不同系統所提供的檢索結果並重新顯現於單一介面、去除檢索結果的重複資料、記錄使用者的檢索歷程，提供使用者能夠於檢索時設定處理檢索結果的機制或修改檢索策略之功能、具有考慮不同使用者之個人偏好與使用需求的功能。

一、單一使用者介面

使用者透過單一使用者介面 (Common User Interface) 查詢各種不同資訊資源，包括身分認證、資訊的選擇、檢索策略的輸入以及檢索結果的瀏覽。使用者透過此一單一介面從不同資訊資源提供搜尋所需之資訊，降低使用者學習各種不同資訊資源提供之檢索介面與語法可能產生的困難。

二、認證機制

目前電子資源的使用認證機制 (Authentication) 可分為網址認證、帳號認證與混合認證等三種，認證後，系統提供資訊的使用權。網址認證是透過連線網址的限制，確認連線者是否具有使用權；帳號認證則是透過連線者輸入帳號與密碼，確認使用者是否可合法使用所提供的電子資源；混合方式認證則是網址、帳號兩者並用，登入帳號又可區分為機構帳號及個人帳號。

整合檢索機制除了提供單一使用介

面，必須能夠依據所連線系統的要求設定認證方式，甚至進一步設定系統可使用人數。對使用者來說，所需要做的是單一登入，也就是說一旦經由系統認證，後續相關之資源使用無須再處理登入認證的程序，都由系統依據該次登入提供之系統認證資訊處理相關的認證程序。

三、支援多種通訊協定與資料交換格式

不同系統支援整合檢索系統的方式不同，整合檢索系統必須依照需要連線搜尋的系統來支援不同的通訊方式與資料交換格式 (Communication Protocol)。常見的通訊協定包括 Z39.50、OpenURL、HTTP …等，資料交換格式則是包括 MARC、XML、Dublin Core …等。部分資源甚至需要與資訊或系統提供者相互配合撰寫不同的 API 程式，以利系統間的資訊交換。

四、檢索功能

整合檢索系統提供單一介面進行搜尋，檢索功能 (Search Functionality) 上首需解決的問題為不同系統所提供之索引之對照，介面中之各項檢索所對應的欄位範圍必須能夠依照各系統之特色進行適當的設定。整合檢索系統策略上雖講求簡易操作與使用，但是在簡易操作的同時，也必須考慮提供讓檢索結果更加精確或回收率高的需求。一般檢索系統常見的布林



邏輯運算、詞彙對照、剔除字、萬用字元等，在整合檢索過程中仍舊不應被忽略。

五、策略廣播機制

配合單一介面的搜尋，整合檢索系統應配合使用者所選擇的資訊資源，將檢索指令依照符合的語法傳送至被選定的目標，進行檢索後再回傳檢索結果，過程中使用者無須重複輸入其檢索策略。期間所涉及的問題在於系統是否可以正確地依照系統需求轉譯檢索策略，充分掌握系統所提供的資訊。

六、記錄檢索歷程與設定

記錄檢索歷程與設定 (User Interaction) 分兩層次來看，一是機構層次，另一則是檢索者個人的檢索記錄。從機構層次來看，系統要能夠記錄不同 (群組) 使用者，使用各項資源以及使用各項資源的歷程，以利管理之需求；從個人層次來看，檢索過程系統必須記錄檢索者進行之檢索，檢索者並可以依照檢索需求處理檢索結果。

七、客制化

客制化 (Customization) 可由兩部分討論，一是從機構層次考慮，另一則是從使用個人層面考量。機構部分是指系統必須提供一個簡易操作介面，使機構使用者能夠依照所提供的資源、簽訂的使用條件、使用者型態設定使用規範。從個人使

用者的角度來說，系統要提供使用者自行設計其長期需要使用的資訊資源、符合個人使用習慣的介面、儲存個人之檢索歷史以供再利用等。

從資訊搜尋角度，整合檢索系統必須要能夠克服通過不同的資源使用認證機制、提供整合的檢索介面，透過單一介面搜尋各種不同型態的資訊資源、有效的對應不同資訊資源的欄位並製作搜尋索引、檢索結果的擷取與處理，包括結果的回傳，排序與去除重複資料。

肆、整合系統之評選與建置程序

如同圖書館作業自動化系統的發展，整合檢索系統可以採取自行開發、廠商配合開發以及引進套裝系統等模式處理。但是亦如同圖書館自動化系統的發展，前兩項選項所面臨的問題一是人力成本過高，二是檢索效益不確定，三是後續發展較受到限制。考慮發展成本，套裝軟體自是不如自行開發來的客制化，但在經費掌握以及功能確定方面佔有極大優勢。但即使如此，面對所提供的不同產品，圖書館如何規劃、選擇、建置，過程中還是有相當多的細節需要注意與思考。為了提供圖書館與系統發展廠商一個可供參考遵循標準，NISO 於 2003 年成立了 MetaSearch Initiative (<http://www.niso>.)



org/committees/MS_initiative.html)，其主要任務是針對異質資料庫整合檢索，從使用管理、館藏敘述以及資訊檢索與搜尋等不同面向提出基本的功能需求與標準設立。紐西蘭國家圖書館也曾委託 Victoria University of Wellington 的資訊管理學院（School of Information Management）就市場所提供的套裝系統進行評估，瞭解不同相關產品所能滿足的檢索功能（Dorner & Curtis, 2003）。相關的研究與調查結果可作為圖書館發展異質資料庫整合檢索系統的基本參考資料。

成大圖書館自 2003 年開始，選擇以引進套裝軟體的方式發展異質資料庫整合檢索系統，期間歷經需求書撰寫、徵求需求答覆書、系統評選、系統測試等步驟。雖尚未找到合適的產品，但其間的歷程可作為圖書館之參考。成大圖書館 2003 年所進行的規劃因與其他兩所圖書館合作，因此除圖書館本身需求的考量外，並需就當時的硬體架構、其中涉及的系統認證問題以及三所圖書館對於系統功能需求之差異進行考量與討論。以下所提出的步驟程序是回歸到單一圖書館進行類似計畫時所應注意的歷程細節。

一、基本需求分析

系統需求分析分為兩部分，第一個部

分是基本需求分析。基本需求分析所需要考慮的是檢索系統需要涵蓋的主要功能範圍除了資訊搜尋，是否需要包括電子資源管理的功能，若需要包括後者功能，那麼管理功能需要管理哪些資源，同時需要達到哪些要求？基本需求第二項需要考慮的是，建置的整合檢索系統與現行系統之間的關係為何？是取代現有系統還是與現有部分系統並存。若是取代現有系統，那麼現有系統所具備的功能自是轉為系統需求；若是並存，不同系統之間是否需要具備某種程度的連結功能則是另一層不同的需求考量。無論系統是以資源檢索為主要訴求或是管理功能為取向，系統需要處理哪些類型資源的功能是基本需求分析第三項需要考慮的。雖說所需要管理的都是電子形式的資源，但是不同類型的電子資源在搜尋及管理上仍有不同的需求。

二、功能需求分析

確立基本需求後，第二部分則是需就細部的系統功能需求進行分析，從資源檢索的角度來看，功能細項包括不同資訊型態所需要考量的檢索功能、不同資訊資源所支援的通訊協定系統與認證機制處理模式、各類型資源的索引建置與對照、檢索功能的要求，包含策略組合與詞彙處理、搜尋結果的處理與呈現、使用者個人化設



定；從管理角度來看，功能細項包括系統建置與管理資訊資源的方法、使用者類型與授權程度的要求、使用狀況的記錄管理與追蹤、統計報表的產生等。每項都需要充分的思考與討論。

三、撰寫系統需求書

完成架構前述兩項工作結果後，可以開始撰寫系統需求書。除了在功能文字描述上需要詳加推敲，在撰寫過程中，可以重新審視需求的內涵以及需求程度，並可尋求已建構類似系統圖書館的支援，從其經驗及提供可能的實際操作蒐集使用資訊，以利需求書的撰寫與後續評選過程。需求書內容包括系統整體要求、各項功能及其詳細描述，還需包括功能評比權重。

四、尋求答覆書與產品評估

系統需求書完成後，聯繫廠商提供相關的系統介紹與實務操作可作為評估的基礎。除了廠商所提供的書面答覆資料審查之外，對圖書館最有利的評估是能建立測試系統，讓館員能夠有機會實際就圖書館的作業環境與所提供的資源進一步進行測驗評估。此階段的工作有其執行上需要克服的問題，除了時間方面需要圖書館員相當程度的投入，有意提供服務的廠商亦需要有配合的意願，同時採購程序以及評

估結果對系統評選必須具有其實質意義。因配合作業程序，成大圖書館此次的進行方式分為兩階段，評選前的評估是以書面審查為主，而深度的實際操作評估則是在評選後針對入選系統進行測試。不論是決選前後，作者均建議圖書館應投入一定的人力與時間成本進行深度評估，以充分瞭解系統功能。以成大圖書館此次的經驗來說，書面審查所獲得的資訊與實地深度評估的結果有相當的落差。

書面審查部分包含就需求答覆書以及評比權重評分，深度操作評估則包括整體功能項目檢查、認證測試以及檢索問題測試。整體功能項目檢查主要是確立系統主要功能選項以執行成果均符合需求；認證測試是確定使用者登入、連線與使用權限設立與操作的正確性；檢索問題測試則是設計檢索測試集，透過選定的資源、設計的問題，以實際的檢索結果比較確定整合檢索功能的運作正常。設計的程序包括選定各類型資源具代表性的測試典範、設計檢索問題、進行整合檢索與直接連線檢索、比較檢索結果，結果包括檢索結果的筆數以及回傳的資訊內容。過程中需保留完整的測試記錄。由於成大圖書館系統能夠提供詳細的使用記錄，因此使用記錄與統計是另一項評估重點，檢索測試過程所保留的記錄，正可作為系統提供使用記錄



與統計功能的驗證依據。

五、系統測試

深度評估測試通過後，圖書館就可考慮將所提供的各種資源逐一鍵入系統，進行完整測試。資訊建構方面多可要求提供服務單位的協助，一方面是服務單位有相關類似的經驗，操作上會較圖書館容易掌握，另一方面圖書館可以藉此機會觀察學習，累積經驗並將主力集中於驗證與規劃各種資訊資源在系統上呈現的樣貌。

前一階段的評估，已就系統功能進行深度的測試，上線前可針對深度評估階段較具不確定性的部分進行測試，同時亦可擴大範圍進行檢索測試，例如主題領域、同一資料類型、資訊連結等進行操作測試。另一方面圖書館可以考慮加入讀者測試，包括檢索資訊經驗豐富者的測試以及新手上路的測試，從其中擬定推廣的策略以及觀察未來上線後可能產生的問題。

六、上線推廣

異質資料庫整合檢索系統雖以單一操作、高度親和力為主要訴求，但實際操作上仍具有一定的學習與磨合期。如何與既有的系統相互配合、推廣活動的安排、主

要檢索訴求以及可能的 Q&A，是系統上線前與上線初期圖書館必須面臨的階段，需要做好投入人力與時間成本的準備。

伍、討論與思考

提供一個單一整合介面讓資訊使用者能夠容易、精準地掌握與取得所需要的資訊是從事圖書館專業人員一直以來希望能夠達到的目標。異質資料庫整合檢索系統讓圖書館朝這目標向前邁進了一步。但是就現階段的發展來看，我們也需要有這樣的體認，如果運用得當整合檢索系統相對來說確實提供了一個操作容易、節省時間的搜尋資訊的起點；但是整合檢索系統在現階段並非是一個檢索功能強大的搜尋工具，在滿足複雜檢索需求上有其限制，同時絕不應是檢索的終點，資訊使用者仍需小心謹慎的審視檢索結果。

對於資訊檢索效能的追求是無止盡的，整合檢索系統是另一個 Google 還是應具有其不同的定位？要營造什麼樣的資訊搜尋環境恐怕也是一個圖書館必須持續自我挑戰的問題。但相信在謹慎的規劃、投入的時間以及確實的執行，在資訊傳播者角色的扮演上，圖書館應能再往前邁進一步。



參考書目

- Bush, Vannevar (1945) . As we may think. The Atlantic Monthly, 176 (1) : 101-108.
Retrieved April 7, 2006, from <http://www.theatlantic.com/doc/194507/bush>.
- Carden, Mark (2004) . Library portals and enterprise portals: why libraries need to be at the centre of enterprise portal projects. Information Services & Use, 24 (4) : 171-177.
- Dorner, Daniel G. & Curtis, AnneMarie (2003) . A comparative review of common user interface software products for libraries. Retrieved April 7, 2006, from http://www.iub.edu/~librfef/1526/CUI_Report_Final.pdf.
- Fryer, Donna (2004) . Federated search engines. Online, 28 (2) : 16-19.

本篇文章轉載自主講者於 CONCERT 2006 年南區座談會之演講，經主講者修改同意刊登。羅思嘉副教授原任職於成功大學圖書館資訊服務組，自 95 年 8 月轉任現職。