BIBLID 1026-7220（2007）97:1 p.35-38

國家圖書館網站典藏先導系統

賴重仁　國家圖書館採訪組書記

一、發展緣起

網路資源已成爲目前重要的資訊管道，其內容豐富多元，可提供一般民眾資訊檢索及學術界研究參考所需。本館爲一國家級圖書館，除致力於國內實體書刊文獻典藏，近十年來對館藏期刊、善本古籍及臺灣研究等文獻進行數位化及資訊服務，獲致相當成就，深受國內外學術界及民眾一致肯定。但隨著資訊科技的不斷發展，圖書館由傳統以紙本爲館藏重心的圖書館進入以紙本與電子媒體並行的圖書館，更朝著虛擬圖書館邁進，電腦網路的日益普及，使得圖書館可瀏覽、檢索、取得位於館外遠端的資訊，所謂館藏不再侷限於館內所有的資料，加上資訊的類型也由紙本、微捲、線上書目資料庫、光碟資料庫擴展至各式網路資源，網路科技使得圖書館對館藏及館藏發展的定義不同於以往。圖書館需要蒐藏的文獻資源必須滿足讀者需求除既有紙質媒體外，尚且包括各類網路資源及電子媒體。

然而網站成長的快，消失的也快，今日所看到的網站內容很可能隨時因爲伺服器關閉、網站移除、網站名稱變更…等因素無法再次被利用。因此站在資源保存與協助學術研究的立場上，實有必要針對網站內容進行保存。而現今世界先進國家之網站典藏 (Web archive) 計畫執行，均由國家圖書館負責。如世界聞名已具 10 年發展歷史的澳洲國家圖書館 “PANDORA” 計畫及美國國會圖書館等均積極投入網站典藏作業中。因此網站典藏已成爲兼負文獻典藏任務的國家圖書館，在進行國家文獻典藏時，不能忽略的重要職責。基於上述理由，國家圖書館於 96 年度著手建置國家圖書館網站典藏先導系統（以下簡稱本系統）。本系統將對網站進行定期擷取內容，並根據擷取時間建立個別的典藏版本，存放於伺服器中，除建構完備之模擬、封裝、更新、轉置等永久典藏機制外，可進一步提供讀者進行資料查詢與其他加值應用。

二、系統簡介

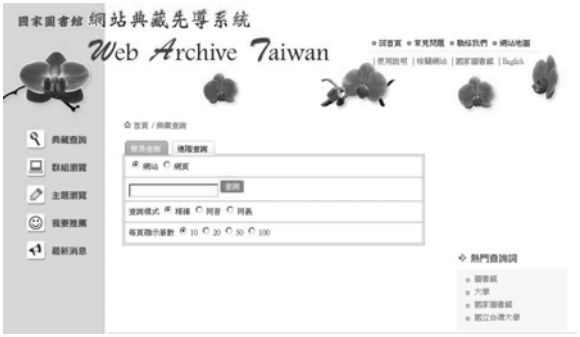
本系統建置之重點爲對擇定之網站內容進行定期擷取，根據擷取時間建立個別的典藏版本，並存置於本館伺服器典藏。但典藏的目的在於提供讀者應用，爲提高讀者的使用意願，是否提供一個友善的使用者使用環境就十分重要。本系統的前端網站服務可概分爲查詢、典藏網站內容瀏覽及讀者服務等三類，茲分別介紹如下：

1. 查詢

讀者可使用簡易查詢、進階查詢、群組瀏覽或主題瀏覽等功能來找尋所需網站資源。

簡易查詢

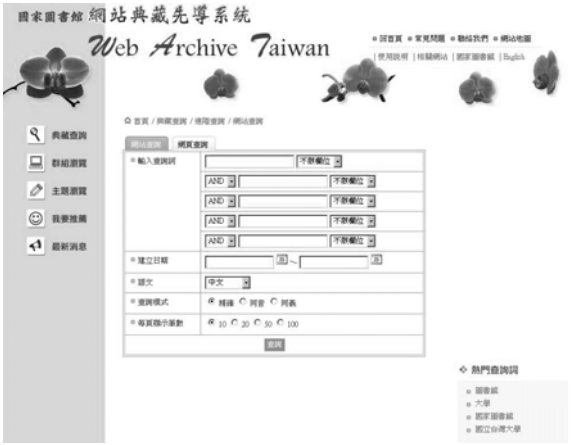
簡易查詢提供使用者以全文檢索方式進行網站資料（網站 Metadata）或網頁資料 (網頁內容全文) 查詢。並提供精確、同音、同義三種查詢模式進行資料檢索。



圖一、簡易查詢

進階查詢

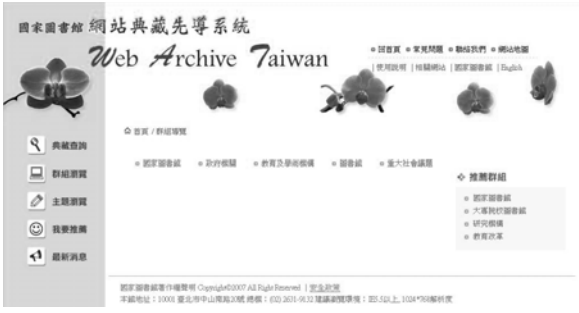
進階查詢之網站查詢提供使用者以欄位式方式進行典藏網站及網頁資料查詢。使用者可選擇所欲查詢欄位爲網站題名、主題名稱、群組名稱、網站簡述、或是網站的資料類別，並可選擇各欄位之 AND/OR/NOT 布林組合運算。且亦有建立日期及語文等限制條件做爲資料篩選檢索。此查詢功能亦提供精確、同音、同義三種查詢模式進行資料檢索。



圖二、進階查詢

群組瀏覽

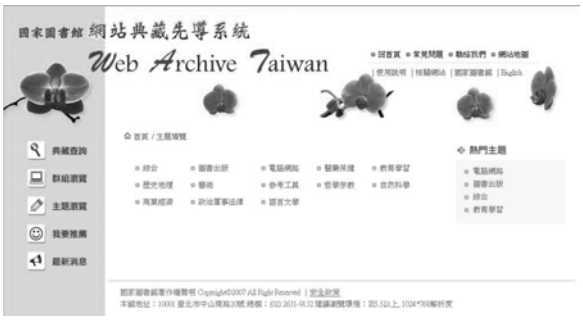
本系統所有典藏網站資料均依群組分類，目前群組做爲二層群組架構。因此讀者可依群組進行典藏網站資料進行瀏覽。



圖三、群組瀏覽

主題瀏覽

本系統亦可依網站主題進行瀏覽，目前主題爲二層群組架構。讀者可依網站主題進行典藏網站資料查詢。

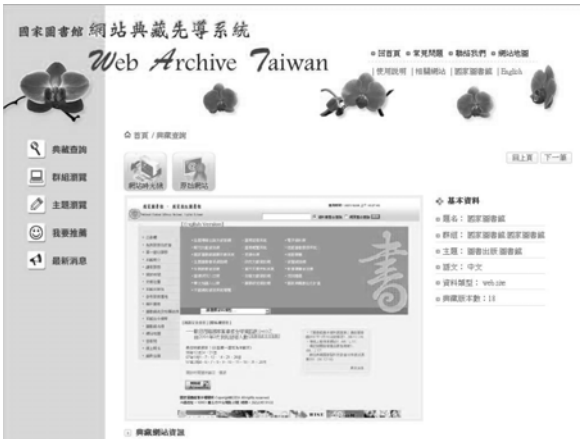


圖四、主題瀏覽

2. 網站內容瀏覽

網站基本資料瀏覽

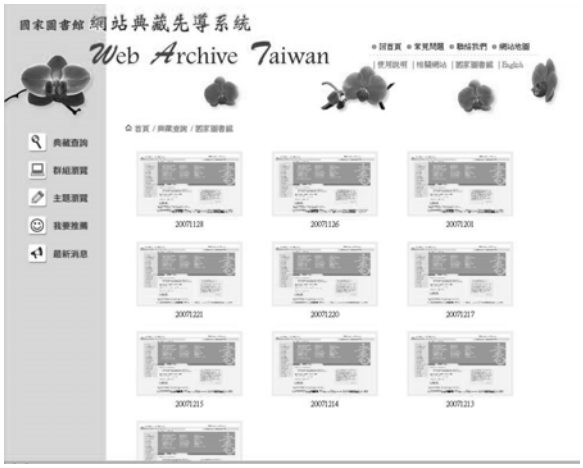
讀者依不同查詢方式尋得並點選進入所需網站後，首先可瀏覽該網站的基本資料及其他相關連結，包括有網站 Metadata(如題名、語文、群組、主題、簡述等)、典藏版本數、原網站連結、網站時光機等資訊，使用者可根據自身需求進行操作。



圖五、網站基本資料瀏覽

網站時光機

本系統會自動對網站進行定期擷取內容，並根據擷取時間建立個別的典藏版本，而讀者可於網站時光機中對個別網站的所有版本進行瀏覽，如圖六所示即爲於不同時間所建立之國家圖書館網站典藏版本。

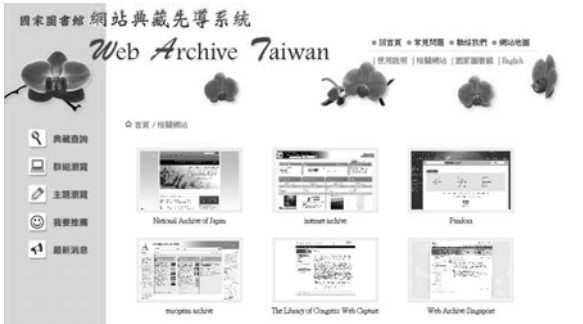


圖六、網站時光機

1. 讀者服務

相關網站

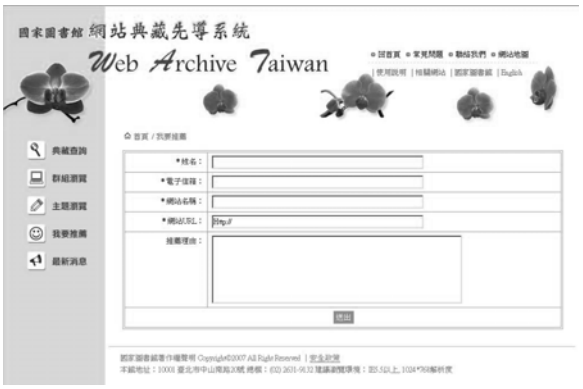
本系統收集國內外網站典藏系統之連結 (如 PANDORA、Internet Archive、Library of Congress Web Capture 等)，匯集於相關網站中。讀者可於此瀏覽國內外之網站典藏內容。



圖七、相關網站

我要推薦

讀者可使用推薦網站功能介紹網站供系統典藏。系統會對您推薦的網站進行審核以決定是否納入典藏之中。審核之結果會以 E-mail 方式通知推薦人。



圖八、我要推薦

一、系統後續發展

1. 著作權問題

依據圖書館法第十五條的規定，國家圖書館爲全國出版品的法定送存機關，但網站資源是否屬於出版品送存之範疇則仍有疑慮，且網站資源之著作權歸屬極爲複雜，雖有國外相關計畫經驗可供資鑑，但考慮國情之不同，故本計畫所儲存之數位資源在存取方面仍採較謹慎的態度，目前擬取得著作權人完全授權方對公眾公開。

2. 網站擷取問題

網站擷取技術雖已發展一段時間，但不論是商用、Open Source 或是國圖自行開發的軟體均存有部分未能克服的技術困難，舉其大者約有以下數端。一、網站 Flash 及以 Javascript 撰寫的網站擷取問題，二、網站的版本控制問題。以國圖目前的典藏分析，網站的擷取不全常導因於網頁 Flash 的擷取失敗或是被 Javascript 導回原網站而無法順利擷取，而經查國內外網站典藏系統針對同一網

站的擷取也有相似的現象，此爲目前網站擷取普遍存在的技術困難；此外爲系統長遠發展計，網站的版本比對是相當重要的，這樣可以節省相當大量的儲存空間，但如何達到目標則尚無定論，版本比對首先遇到的問題將是如要長時間對所有典藏網站進行監控及內容比對，這對系統及網路頻寬來說是相當大的負荷，此外，多大的變動需要被視爲應建立一個新的版本，這也是一個值得思考的問題。

隨著網際網路的快速發展，網路資源已成爲一般大眾或學術研究者獲得資訊的重要管道。而蒐集、整理、保存網路資源本爲圖書館的主要任務，對於典藏國家重要網路資源國圖更應扮演領航的角色。國家圖書館於 96 年度開始投入本系統的開發，現已獲得初步成果，但網站資源是一個成長的有機體，而資訊科技亦日新月異，本系統在未來仍應持續投入資源，力求突破相關擷取技術瓶頸，廣納國內具參考價值的重要網站，以善盡保存國家文獻的重要任務。

（此先導系統於規劃階段獲館外學者專家王梅玲教授、歐陽崇榮教授、陳光華教授、陳亞寧先生、陳雪華教授、陳昭珍教授、宋雪芳教授等諮詢委員們熱心提供建議及指導，以及本館宋建成副館長、王佩瑛主任、吳英美主任、林巧敏主任、莊建國組長、林淑芬副主任、阮靜玲助理編輯、俞小明主任、蔡佩玲編輯等工作小組成員的費心指正，在此謹致謝忱。）

四、參考資料

1. 國家圖書館採訪組，《人文社會科學相關中文網站網頁典藏成果報告》，96 年 12 月。
2. 楊志津，美國與澳洲國家圖書館數位保存計畫之比較研究，國立政治大學圖書資訊與檔案學研究所碩士論文，96 年 6 月。
3. 臺大圖書館網頁典藏庫，http://webarchive.lib.ntu.edu.tw(僅限臺灣大學網域內使用)。
4. Archipol: Archive of web sites of political parties in the Netherlands, http://www.archipol.nl/.
5. Internet Archive, http://www.archive.org/web/web.php.
6. Preserving and Accessing Networked Documentary Resources of Australia (PANDORA), http://pandora.nla.gov.au/index.html.
7. Web Archive Singapore, http://was.nlb.gov.sg/.