

「北京清華大學同方知網技術有限公司」考察報告

讀者服務組 李美幸

今年5月底隨「兩岸中文資源交流友好訪問團」到北京參訪7天。以參觀北京清華大學同方知網技術有限公司（簡稱同方知網）為主，了解目前大陸中文數位化發展的狀況及出版的情形，作為日後徵集大陸資源的參考。

隨者兩岸的交流日益密切，蒐集大陸方面的資訊早已不是禁忌了，然而我們要掌握大陸方面的出版狀況或購買出版品，相較於對西文圖書的了解，仍有隔層紗看世界的感覺。

同方知網是北京清華大學下的一支分機構，受到大陸教育部的支持，將大陸出版的圖書、期刊、會議論文、學位論文、報紙、工具書有計劃的進行數位化，這是一個繁重的計劃，也是一個高度作業科學化、勞力密集的計劃，參與的工作人員上千人。

同方知網下有中國學術期刊電子雜誌社及同方光盤股份有限公司。分掌資源組織、編輯出版、信息加工及儲存產品研製等工作（見圖一至圖三）。中國學術期刊電子雜誌社是目前中國規模最大最具權威的電子出版機構之一；而同方光盤股份有限公司是中國目前規模最大的全文資料庫與網路出版文獻標準化及生產加工機構。全文文獻數位化的年產量達400萬篇，年信息處理能力達5億條，數位化平均速度7天。

2000年起中國開始積極將紙本出版品數位化，對於核心期刊皆從創刊號開始數位化，以達到保存文獻的目的。這次參訪清華同方知網所發行的「中國學術期刊網絡出版總庫」（見圖四），是在統一的標準規範下加工、分類、整合，所形成的一個超大規模的網絡資料庫，收錄



圖一 各地寄來的文件，等待做掃描處理



圖二 加工廠現場，以系統「揀沙子」的方式控制品質

內容包含自建8,200種期刊、1,500種中國年鑑、工具書、博碩士學位論文、會議論文集、報紙、多媒體資料庫，總量達5,000萬篇，並且與相關單位合作建置專利、專準、科技成果、NSTL等國內外1,200個資料庫，提供二次文獻免費和全文收費服務。



圖三 加工廠的報紙剪報



圖四 CNKI首頁

3. 中國博碩士學位論文全文數據庫：收錄中國著名大學2000年起的學位論文。
4. 中國重要會議論文全文數據庫：收錄2000年以來中國國家二級以上重要會議論文。
5. 中國重要報紙全文數據庫：以重要報紙刊載的學術性、資料性文獻為收錄對象。
6. 中國年鑑網絡出版總庫：至2008年收錄種數將達2,387種。
7. 中國工具書網絡出版總庫：收錄中國200多家出版社正式出版的工具書1,500部。

以上自建資料庫加上合作資料庫（包含專利、科技研究成果報告、標準、NSTL）都可以在CNKI整合平台上檢索。該中心網站一天下載超過100萬篇文獻，每天超過40萬人次造訪，在線人數高峰3萬人。CNKI平台以檢索與挖掘技術，可進行研究主題的時間、領域分析以及熱門主題發現、專題技術追蹤、讀者社群挖掘、學術社群挖掘以及中英引文連結的功能。

由於網絡出版總庫內的各類型資料庫，均建置在同一CNKI的平台上，因此不只使用者可以查詢資料，同時，平台管理端也可提供資源的徵集、加工與編輯、製作出版、以及用戶服務。這樣的平台提供下面三項功能：

重要自建資料庫簡述如下：

1. 世紀期刊：遴選3,964種核心期刊回溯至創刊到1993年。
2. 中國期刊網：此期刊庫從1994年起至今8,460種期。
1. 跨資料庫同一檢索介面
2. 跨資料庫統一知識分類結構
3. 跨資料庫知網節：每篇文獻提供引證文獻、參考文獻、相關文獻連結

Librarian 2.0

簡而言之，在這次參訪中，對大陸在資源的數位化上的努力以及強烈的企圖心印象深刻。文化產業的內容經數位化後，無論在保存上、組織上、傳播上、利用上均發展出與紙本資料迥然不同的風貌，數位化的資料不會因使用的次數增加而耗損，且在經過無遠弗屆的傳播方式後，反而愈利用，資料的價值愈高。我想大陸已深知知識經數位化的影響力深遠，所以除了投注很多心力在建立各類型的中文資料庫外，也盡力解決了版權上的問題，據告知這個網路出版總庫總共花了上億人民幣（可能大於此數）在處理版權問題上，因此使用者在使用此資料庫時無版權上的困擾。

現在我們正處於知識經濟的時代，因著數位化的資料易於做加值服務，該資料庫所發展的知網節的功能，好似擁有虛擬研究團隊一樣，二十四小時為你服務。同時這資料庫也盡力解決了各種類型資料的整合，想想看若所需資料可能遍及期刊、會議集、學位論文甚至報紙及工具書，將需花費多少時間蒐集資料，而現在只要輸入一個檢索詞，彈指一按就可達到「上窮碧落，下黃泉」完整蒐集資料的效果。因此，這資料庫雖然花費大量的人力、資金及時間來建置，但是所達到附加的效果難以想像。在大陸幅員廣闊，各校資源狀況差異性極大的情況下，數位資料可以解決空間上的限制以及資料獲得便利性的問題；在文化傳播上，則可讓國外人士瞭解大陸文化及各項科技發展的成果，也藉著綿密的網際網路傳播到世界各個角落需要的地方，具有宣傳「國力」的效果。

幾星期前，歷史所雷祥麟老師詢問台灣是否有類似整合查詢的資料庫，目前勉強可算是的只有「遠距圖書服務系統」，該資料庫提供期刊目次、政府公報、政府出版品相關的整合查詢，但若要符合老師及一般讀者常用的期刊、報紙、會議論文、學位論文的整合性資料庫，即使只是索

摘型的資料庫，目前也尚未出現。老師還詢問可向何單位或會議場合反應，以推動國內類似資料庫之建置。依個人的淺見，要做到像大陸資料庫的完整性，除了需要有強而有力的主持者來整合人力及技術上的問題，還需提出有效合法獲得授權的途徑，而這已超出台灣民間企業所能做的事了。但這樣的需求其實是非常迫切的，以一個圖書館館員來說，最能體會到資料的分散所帶來的不便，需要花數倍時間幫讀者找資料；對使用者而言，則會因為收集的資料不完整，而影響到研究發展的判斷或造成研究結果錯誤。

總之，隨著大陸資訊愈來愈容易獲得，要研究大陸各領域知識的發展，日益便利。我覺得中文資料庫的建置具有多方面的意義，有經濟上的、政治上的，以及文化上的意義。此外，大陸藉著文化的輸出，有助於加速簡體字的使用遍及，反觀繁體字目前只有香港與台灣在使用，若我們沒有積極的面對這樣的狀況，繁體字將會逐漸被邊緣化，愈來愈少人使用了。