

圖書館古籍整理與資訊科技的應用

專文

圖書館古籍整理與資訊科技的應用

兼述國家圖書館數位化古籍整理的發展

■ 顧力仁 蔡慧瑛

一、前言

由於資訊數位化及網路技術的演進，促使數位圖書館的興起，也由於社會大眾對文化及歷史資產的重視，圖書館珍藏的歷史文獻遂成為數位化的重要對象。目前有許多保護文化資產的計畫以及圖書館紛紛以珍藏文獻為數位化的主題，例如：「聯合國世界文化遺產數位化計畫（UNESCO Memory of the World）」、大英圖書館的「十一世紀盎格魯——撒克遜史詩手稿數位化計畫（The Electronic Beowulf）」、IBM公司的「梵蒂岡圖書館計畫（The Vatican Library Project）」等等。從資訊科技對古籍整理的影響及其發展現況來看，可以預知資訊及網路科技將提供圖書館古籍整理的新方向，也提供了對古籍研究的新境界。

本文先列舉若干海內外知名的古籍數位化相關計畫及電子資料庫，分別介紹其概要及特色，藉以瞭解資訊科技對圖書館古籍整理的影響及應用現況；其次再以國家圖書館（以下簡稱「國圖」）在數位化古籍整理的成果為例，說明該館利用現代資訊科技在古籍整理的發展現況，最後建議未來圖書館要借諸現代化的手段來進行古籍的整理、組織與利用，建立具有中國特色的數位化圖書館，以謀民族文化的延續。

二、資訊科技在古籍整理的應用現況

就中央研究院「漢籍電子文獻資料庫」、上海圖書館「中國古籍善本查閱系統」、臺灣地區「數位典藏國家型科技計畫」中有關「古籍數位典藏計畫」以及香港迪志文化出版有限公司「文淵閣四庫全書電子版」，分別介紹其概要及特色，藉以瞭解資訊科技對圖書館古籍整理的影響及其發展現況。

（一）中央研究院「漢籍電子文獻資料庫」

中央研究院所推動的漢學研究電子文獻包含全文資料庫、研究參考工具資料庫、主題研究多媒體資料庫、書目及關連式資料庫以及利用影像處理技術進行的檔案與古文書光碟影像資料庫等5種。其中以「全文資料庫（現稱漢籍電子文獻資料庫）」的規模最大，應用最廣。民國73年，中央研究院歷史語言研究所與計算中心，合作開發廿五史（食貨志）全文資料庫，第二年起逐漸擴大到廿五史全部，接者相繼輸入十三經、佛藏、醫書、政書等資料，截至88年9月底，已完成的漢籍電子文獻資料庫累計達1億2千萬字，正在進行校對、標誌的資料約有2億字，並以每年超過1千萬字的速度擴增，堪稱為全球最大的中文文獻資料庫¹。中央研究院於84年成立「漢籍電子文獻協調委員會」，未來計畫結合相關學術研究機構，共同推動「新四庫全書資料庫」計畫，以擴增全文資料庫的內容。中央研究院「漢籍電子文獻資料庫」的網址為<http://www.sinica.edu.tw/ftms-bin/ftmsw3>（圖14）。

中央研究院「漢籍電子文獻資料庫」的特色包括²：



圖14 中央研究院「漢籍電子文獻資料庫」選例



- 1.品質最佳：文獻失誤率低於千分之一。
- 2.規模最大：開發及完成的合計超過3億字，每年平均以3千萬字建檔，校對完成者約1千萬字。
- 3.資料經過一定規模的組織。
- 4.軟體設計最佳。

(二) 上海圖書館「中國古籍善本查閱系統」

上海圖書館收藏古籍文獻170萬冊，包括善本書25,000餘種，17萬冊。該館於1966年開發「中國古籍善本查閱系統」，將古籍掃描錄入光碟，並提供古籍標引、檢索及查閱等功能；此外，並將數位化館藏古籍善本陸續分期放上網路，目前已上網善本20種19,794影幅，包括宋刻本17種、元刻本1種、稿本2種，內容詳見<http://www.libnet.sh.cn/digilib/gj/index.htm>。

上海圖書館「中國古籍善本查閱系統」(圖15)的特色包括³：

- 1.分為兩個子系統：系統分為「製作標引」及「存儲查詢」等兩個子系統，前者提供掃描、壓縮、編碼及燒錄等功能；後者提供檢索及調閱影像等功能。
- 2.影像畫面：即古籍書頁原貌，每兩個半葉組成一個畫面，黑白、彩色或兩者相間，凡彩色印本、彩色批校、題跋或重要藏印皆為彩色畫面。
- 3.閱讀功能：可選擇全書，也可任選封面、扉頁、序、跋、凡例、目錄、圖、附錄、添加頁或正文，其中正文又可任選不同卷次。以上皆可隨意向後、向前逐頁翻閱，也可指定其中一頁跳躍翻閱。可顯示全書或某卷、某部份的總頁數，及正在閱讀的頁數。閱讀時可隨意調整為全畫面或局部畫面，也可同時顯示兩者。圖像可以任意作黑

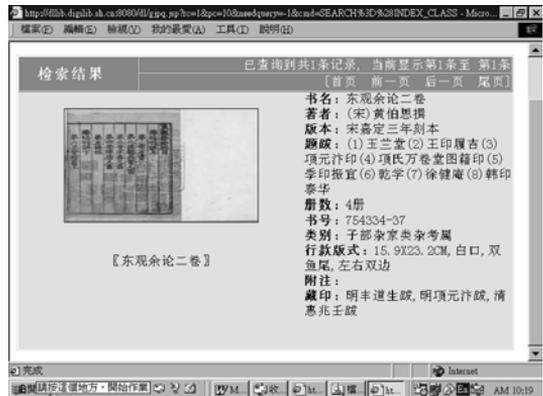


圖15 上海圖書館館藏古籍影像選例

白反轉、正反反轉及旋轉位置。貼有粘簽的古籍書頁，除可閱讀其有粘簽的書頁外。也可使粘簽「飛去」，顯示其粘簽下遮蓋的文字原貌。

- 4.檢索功能：檢索途徑包括分類、書名、著者、索書號等項目。
- 5.光盤及網路併行。

(三) 臺灣地區「數位典藏計畫」

臺灣地區為建立重要文化資產的影像化，針對重要圖書館及博物館所典藏的珍貴文物規劃數位典藏計畫，自91年起由行政院國科會推動「數位典藏國家型科技計畫(National Digital Archives Program, 簡稱NDAP)」其中已進行與古籍有關的包括「國圖」所進行的「古籍文獻典藏數位化：善本古籍典藏數位化子計畫」以及中央研究院「歷史語言研究所珍藏文物數位典藏計畫：史語所傅斯年圖書館藏善本圖籍」，前者的發展詳如後文，現就後者簡介說明如下。(圖16)

中研院史語所傅斯年圖書館的珍貴典藏甚多，該館於77年開始進行「善本書光碟影像計畫」，目前收錄善本圖籍及俗文學資料影像近125萬頁，並

提供「傅斯年圖書館珍藏圖籍全文影像資料庫」及「傅斯年圖書館珍藏善本圖籍書目資料庫」⁴，各界人士可以透過網際網路（Internet）檢索已數位化的善本圖籍書目性資料，也可以透過網路授權使用或在該圖書館閱覽室內閱覽善本圖籍影像檔，並可列印所需資料。該館目前係與「國圖」、故宮博物院圖書館合作，共同進行「明版明人詩文集」以及「古籍附圖數位典藏」等數位計畫，日後將針對該館所藏醫學、考古及金石等主題的古籍予以數位化。該館數位工作包括：

1. 影像製作：掃描製作300-600dpi/tiff全彩影像。
2. Metadata書目資料庫建置：訂定館藏善本書（及明人文集）詮釋資料及著錄範例⁵，並建置Metadata書目資料庫。
3. 整合已數位化的善本圖籍影像與Metadata書目資料庫，提供「傅斯年圖書館珍藏圖籍全文影像資料庫」，透過授權，可在線上閱覽原文影像，全文影像資料庫目前內容收錄22,574筆資料。

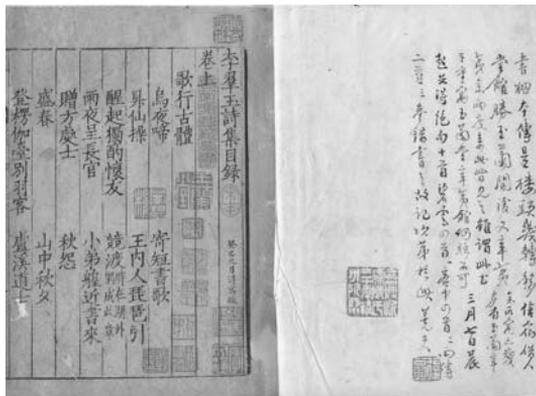


圖16 傅斯年圖書館珍藏善本圖籍影像選例
（集部，李群玉詩集）

（四）香港迪志文化公司「文淵閣四庫全書電子版」 （圖17）

香港迪志文化出版有限公司於1998年推出「文淵閣四庫全書電子版」。四庫全書原於清乾隆年間纂修，共收圖書3,469多種，約7億字，36,000餘冊，該公司以掃描方式將原書圖像數位化，再用光學文字辨識技術，將圖像轉為近7億的電腦編碼文字，並研究出多種檢索功能，「文淵閣四庫全書電子版」的簡介網址為<http://www.unihan.com.cn/skqs/skqshome.htm>。

「文淵閣四庫全書電子版」的製作包括了輸入、辨識、校對及檢索等四項程序，其特色則包括⁶：

1. 圖文並列：「電子版」分為「原文及標題檢索版」及「原文及全文檢索版」，前者包含原書頁的圖像及標題檢索，後者另增加全文檢索功能。
2. 使用Unicode字集：於Unicode的基礎上，建立近3萬中文字集。
3. 跨平臺：適用於臺灣（Big5）及大陸（GBK）中



圖17 文淵閣四庫全書電子版選例（史部，別史類，東都事略，卷一）



文視窗，以及其他如英文及日文等語文平臺。

繼「文淵閣四庫全書電子版」，陸續又推出了「康熙字典電子版」及「四部叢刊電子版」，使用Unicode文字平臺及XML（可擴展標識語言）標識，同樣提供原書影像及全文檢索等功能。

三、國家圖書館數位化古籍整理的發展

臺灣地區的中國古籍文獻是以「國圖」為主要典藏機構，不僅數量居冠，種類亦多；關於善本舊籍共12,532部、126,359冊，其中宋金元版有497部、7,613冊，明本逾6,000部、鈔本近3,000部、稿本、批校本各500部左右，敦煌卷子143部、151冊；另有普通線裝舊籍9,475部、119,141冊，總數量為12,532部、245,500冊。其他又包含：金石拓片6,477種、11,997幅，居延漢簡30枚，民間版畫1,156種、1,852幅，當代名人手稿245家，契券23種，縮影微捲（片）41,814捲（片）；以及版片印模63件、書畫、明信片、老照片、古文書等收藏品，不勝枚舉⁷。

「國圖」秉持著維護、發揚固有文化遺產為職志，對於古籍文獻的整理不遺餘力，除修裱珍貴舊籍、編印資料目錄等外，並於78年辦理「古籍整編計畫」，組成「古籍整編小組」，從事古籍整理、研究，歷時多年，出版《善本叢刊》7種，編印《標點善本題跋集錄》、《國立中央圖書館善本序跋集錄》、《四庫經籍提要索引》，以及《國家圖書館善本書志初稿》等⁸。近年來，科技資訊日新月異，為配合時代需求，各領域、各學術均紛紛走上數位化一途，「國圖」不落人後，在資訊科技的推進上全力以赴，古籍文獻的數位化亦是重要項目之一，可由書目查詢、影像建置、全文檢索等成就上見諸端倪，茲分述於后。

（一）書目查詢

古籍文獻浩如煙海，若無完善的書目提供查詢，難以獲致事半功倍之效；「國圖」有鑑於此，始終致力編製圖書目錄，以方便讀者利用查詢。惟古籍文獻不同於一般圖書，除了具有撰、著者及出版者、時間、地點外，另有若干特性，如：批校者、題跋者、刻書者、序跋者以及藏印者等，這些均是古籍的重要檢索項目；此外，為了因應實際需求，在目錄的提供上採取不同方式；「國圖」特藏組目前所製作的相關目錄共4種，分別說明如下：

1. 善本古籍機讀編目格式目錄

「國圖」於72年開始建置善本古籍書目資料庫，其主要是依據「中國善本圖書機讀編目格式」建檔。建檔內容包含書名、卷冊、著者、版本、版式、行款、刊記、扉頁題署、序跋、避諱、刻工、藏印、手書題跋、版本類型、裝訂形式、卷端行題，並間附版本考訂，著錄項目十分完整。

「中國善本圖書機讀編目格式」是以69年所訂定《中國機讀編目格式（Chinese MARC Format）》為基礎，再針對中國善本古籍的特殊性設計而成，其過程曾參考美國獨立研究圖書館協會所擬訂的「善本書及特藏資料機讀編目格式標準建議」中的黎波里各項構想，並實際參訪國內相關古籍收藏單位及專家學者，於《中國機讀編目格式》的原架構中加添欄位，包含：有關善本圖書的批校題跋、殘缺卷葉、印記、版式行款、序跋、紙墨裝訂及考證校勘等；此外，另增設善本輔助檢索項（欄號608），可依版本、刻書地、刻書者、刻工、刻書年、裝訂形式、藏印者等作為檢索相關資料⁹。

關於「國圖」善本書機讀書目所使用的系統，最初係用王安電腦公司所開發的「善本書編目系

統」，81年改採URICA系統編目模組，繼續建檔，其成效良好；截至91年11月為止，已建檔的舊籍達79,096筆，包括叢書所涵蓋的各子目在內。

2.善本古籍Metadata書目

Metadata是為資料描述格式、詮釋資料，其目的在促進資料系統中對資料的檢索、管理與分析。Metadata不僅具有傳統的著錄功能，更是資料的管理維護者與使用者之間的橋樑，透過Metadata更容易了解，且增強資料的辨識，進而予以利用與管理資料，其功效遠較機讀編目理想¹⁰。

有鑑於Metadata日漸重要，「國圖」於89年邀請國內相關的專家學者及館內同人成立「Metadata研究小組」，並收集國內所設計的Metadata格式，編印為「中文詮釋資料（Metadata）格式彙編」；此外，為了配合古籍的特殊性質，另設計「古籍善本詮釋資料及著錄範例」¹¹。據此又開發建置「善本古籍Metadata書目資料庫」，逐步將「國圖」所珍藏的善本古籍的相關書目資料轉為Metadata格式書目（圖18）。

3.臺灣地區善本古籍聯合目錄

過去，臺灣地區曾出版善本公藏目錄，係屬書本式目錄，其資料並未隨著各館資料的異動而更新。87年4月「全國圖書資訊網路系統」（簡稱NBINet，後改稱「全國圖書書目資訊網」）新系統啓用後，為推廣「國家書目」資料庫的發展，並藉由現代科技，以擴大資料收集範圍，建立較完備的古籍善本資料庫，因而有「臺灣地區善本古籍聯合目錄建檔計畫」並建立臺灣地區古籍聯合目錄資料庫。此資料庫共收錄「國圖」、故宮博物院圖書館、臺灣大學圖書館、東海大學圖書館、國立中央圖書館臺灣分館、中研院史語所圖書館、中研院文

哲所圖書館等7個單位，（截至91年11月）已建檔者有116,034筆古籍書目資料（<http://rbook.hyweb.com.tw/frameset4.htm>）¹²。

由於古籍聯合目錄資料庫合作單位的擴增以及數位影像的快速發展，「國圖」特別選定若干核心欄位，建置「古籍聯合目錄Metadata著錄測試系統」（<http://nclcc.ncl.edu.tw/ttsweb/metadata.htm>）（圖19），以Metadata格式表達呈現。此測試系統所有書目資料，除臺灣地區各重要古籍收藏機構的11萬餘筆外，另有大陸、港、澳等地區11個機構所送交測試完成的600餘筆。日後將更進一步提供Metadata書目資料的匯出功能，以利國際間古籍書目資料的交換。

書號	11857	機讀格式	篇目資料
題名	李氏居室記		
卷數	五卷		
創作者	（明）李濂（撰）		
篇目影像	正文前	卷之一	卷之二 卷之三
	卷之四	卷之五	
標題	集部-別集類-明之屬		
序跋者	（明）李莘（跋）		
序跋	序：「李氏居室記序 嘉靖癸卯嵩渚山人李 濂川父志」，跋：「書居室記後 嘉靖甲辰男莘 叟頓首識」		
收藏印記	「吳興劉氏嘉 / 業堂藏書記」朱文長方印、「國立中 / 央圖書 / 館考藏」朱文方印		
版本	明嘉靖甲辰（23年：1544）大梁李氏家刊本		
裝訂	線裝		
版式行款	9行，行18字。單欄。版心白口，單白魚尾，上方記書名。		
數量	2冊		
高廣	（匡17.9x13.4公分）		
索書號	402.6 11857		

圖18 國家圖書館善本書機讀書目轉為古籍Metadata選例



近年來，為擴大共建共享功能，「國圖」有意將此聯合目錄資料庫發展成全球性的構想。曾兩度邀約海、內外著名古籍收藏單位共同開會討論，並請其將已建檔或紙本書目資料提供測試，目前完成「全球中文古籍聯合目錄測試系統」，期盼「全球中文古籍聯合目錄」能順利完成建置，早日公諸於世(圖19)。

4.臺灣地區家譜聯合目錄

家譜是記載民族血緣的歷史圖籍，在歷史發展中具有特殊價值地位，尤有助於人文社會科學的研究。目前，臺灣地區收藏中國家譜(含紙本及微縮資料)約有6,000餘種，在中國家譜總數中佔有相當的比重；「國圖」十分重視這份彌足珍貴的文化遺產，於是展開調查公藏家譜工作，詢求合作編製一份聯合目錄。

91年3月間邀約有關收藏單位，召開「臺灣地區家譜聯合目錄合作編製相關事宜座談會」，會中決議由各收藏單位提供書目資料，並由「國圖」訂定欄位、進行測試，匯集各單位的書目資料，建置家譜聯合目錄。經過數月的努力，已於11月初順利完成「臺灣地區家譜聯合目錄測試系統」(<http://nclcc.ncl.edu.tw/ttsweb/nclfamily.htm>)，參與單位共有11個，除「國圖」外，另包含：中研院民族所、中研院傅斯年圖書館、臺北市文獻會、臺灣分館、臺灣省各姓淵源研究學會、宜蘭縣史館、故宮博物院、國史館、國史館臺灣文獻館及萬萬齋等，共收錄6,052筆書目。

家譜聯合目錄的資料庫除可供相關單位匯入CMARC、CNMARC以及USMARC等MARC格式書目資料，以利查詢、顯示外，並兼具提供著錄、維護等功能；未來「國圖」將進一步擴充該系統，提

供符合國際標準的Metadata書目資料匯出功能，以利資料的交換¹³。

(二) 影像建置

「國圖」的古籍文獻資料在影像建置方面，早期係拍攝縮微影像，即目前所使用的微捲、微片；近期則是建立數位檔案，已建置者有明人詩文集影像系統及古籍影像資料庫，茲說明於下：

1.明人詩文集影像系統

「國圖」於87年擇取館藏明代前期作家17種，建構製作成「善本叢刊影像先導系統」，共有7,000影幅，其係以明人詩文集為主(<http://www.ncl.edu.tw/flyweb/nclbook/index.htm>)。

此先導系統結合電腦科技，利用拍攝及掃描方式，將古書的各葉影像數位化；一方面得以顯示館藏特色，另一方面則可展現難得一見的版本。本系統的影像全部為彩色，為顧及存真，以呈現原書面目，製作時大抵不作去底色、去污跡的處理，故逼真肖原書。檢索點各為詩文集的書名、撰者、版本、篇名、序跋者及題跋者等，可以不同的尺寸瀏覽並



圖19 全球中文古籍聯合目錄測試系統選例

列印。此系統設置於「國圖」的區域網路上，提供讀者使用；另擷取各書部分內容掛於網際網路上，讀者亦可上網利用。此外，所編製的各書目次，偶爾會出現缺字情況，另提供了造字檔，讀者可下載使用。建置於「國圖」館區的區域網路系統十分完整，具檢索功能，檢索欄位包括：書名、撰者、版本、書號、卷次、篇名、序跋者、題識者等。

此系統係採逐次製作專題的方式，仿叢刊型式，故稱為「叢刊」者，即具有相當彈性。因屬起步階段，試驗性質居多，故暫以「先導系統」為名，期望未來的建置更為精良，並且持續的編製下去。

2. 古籍影像資料庫

「國圖」為能迎合科技潮流，進一步建置古籍影像資料庫。90年參與行政院國科會計畫，「國家圖書館善本古籍典藏數位化」為其中子計畫之一，預計3至5年內要將館藏重要善本古籍縮微資料轉製為數位化影像；截至91年12月底止，已轉製完成善本逾1,500部，近100萬影幅，已於網路上開放供大眾使用（圖20）。古籍影像資料庫為方便查詢利用，建置有檢索系統，其系統已開發建置為「善本古籍Metadata書目資料庫」，使用者可從書名、著者、版本，乃至於序跋者、刻工、卷次、篇目等各角度去檢索。

（三）全文檢索

「國圖」致力於古籍的保存與數位化，特別與中央研究院資訊研究所的文件分析與辨識實驗室共同研究，開發出一套「古籍無定型工整手寫文字辨識、檢索與管理系統」，專門針對古書中的各種不同版本及字體，如刻本、抄本、明體、楷體等為對象，進行實驗研究，期望能建立一套最適合製作古

籍影像及全文資料庫的理想方案。

89年，「國圖」開始委外建置該系統，為了提升「文字辨識率」，因而選擇文字清晰、工整，且沒有眉批及校點的古籍作為初步實驗樣本（圖21）。現已建立完成「嘉興楞嚴寺方冊藏經」中的「金剛般若波羅蜜經」、「摩訶般若波羅蜜經」兩部大藏經，其具有3項特色：（1）文字原貌重現：將原文件的影像以完全不失真的方式呈現在電腦螢幕上；（2）全文容錯性檢索：此系統不但製造影像檔，同時也製造其他各種文字資訊，以方便使用者檢索、搜尋所欲擷取的資料，由於所檢索的資訊是電腦辨識的文字，故具有容錯性功能；（3）製作及查詢簡易：利用掃描機將文件輸入電腦後，此系統即自動從事影像處理、文件分析、文字辨識及資料庫建構等工作¹⁴。

四、結語

謝清俊認為電子化的古籍是使古籍活出最佳現代面貌、也是唯一的選擇，因為¹⁵：

1. 古籍的電子版本可無限地複製，是取之不盡、用之不竭的資源，可供全民共享。
2. 透過網路，電子古籍可以瞬息千里，沒有運輸和



圖20 國家圖書館善本古籍典藏數位化計畫影像選例



分配的問題。

3. 電子版本的古籍容易匯集，鉤稽參照後，能產生新的訊息。
4. 電子古籍好儲存，體積小，便於檢索、應用及處理。

圖書館古籍整理隨著資訊科技的影響，產生相當大的變化，武亞民論及「數字圖書館與古籍整理的關係」中認為¹⁶：

1. 數字圖書館（即「數位圖書館」，另有譯成「數據圖書館」）將各種不同類型的文獻信息有機地結合在一起，因此古籍成為數字圖書館中不可缺少的信息源之一。
2. 中國的數字圖書館在建設的過程中必須考慮和首先解決古籍的中國傳統文化特徵，這是具有中國特色的數字圖書館的關鍵。

由此可知，未來圖書館要借諸現代化的手段來進行古籍的整理、組織與利用，建立具有中國特色的數位化圖書館，以謀民族文化的延續。

【註釋】

1. 黃寬重、劉增貴，〈中央研究院人文計算的回顧與前瞻〉，《漢學研究通訊》17卷：2期（民國87年5月）：頁146。
2. 同註1。
3. 陳秉仁，〈古籍善本數字化的嘗試—中國古籍善本查閱系統述略〉，《現代圖書》1998年1期（民國87年1月）：頁22-25, 45。
4. 〈傅斯年圖書館珍藏檢索〉，<http://ultra.ihp.sinica.edu.tw/~fsnlib/c/database/raredb.htm>（accessed November 26, 2002）。
5. 陳昭珍主編，《中文詮釋資料（Metadata）格式彙編》（臺北市：國家圖書館，民國89年）。
6. 張軸材，〈《四庫全書》電子版工程與中文信息技術〉，<http://www.unihan.com.cn/~skqs/skqshome.htm>（accessed November 26, 2002）。

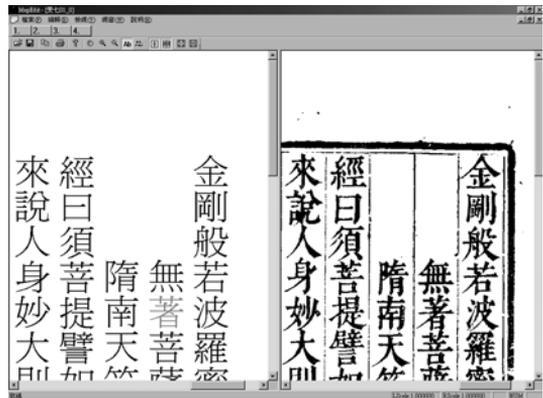


圖21 古籍文字辨識、檢索與管理系統選例（系統辨識結果的對照：左半部是辨識完文字，右半部是原文重現部分）

7. 根據國家圖書館特藏組於民國91年10月25日書庫清點清單。
8. 盧錦堂，〈國立中央圖書館古籍蒐藏與整理〉，《國立中央圖書館館刊》新26卷第1期（民國82年4月）：頁161。
9. 中國機讀編目格式工作小組，〈中國善本圖書機讀編目格式——實驗報告〉，《國立中央圖書館館刊》新15卷1/2期（民國71年12月）：頁28。
10. 陳雪華，〈史料數位化與Metadata〉，《臺灣古文書數位化研討會暨成果發表會之會議資料》（臺北市：國立臺灣大學，民國87年），頁55-56。
11. 同註5。
12. 參閱「古籍聯合目錄資料庫合作建置研討會」會議資料。
13. 參閱「臺灣地區家譜聯合目錄合作編製相關事宜座談會」會議資料。
14. 張復計畫主持，《古籍無定型工整手寫文字辨識、檢索與管理系統計畫執行構想報告書》（臺北市：中央研究院資訊研究所文件分析與辨識實驗室，民國89年，影印本），頁63-64。
15. 謝清俊、林晰，《中央研究院古籍全文資料庫的發展概要》（1997年），頁2，http://www.sinica.edu.tw/~cdp/paper/1997/19970301_1.htm。
16. 武亞民，〈數字圖書館與古籍整理〉，《圖書館學刊》1998年2期（民國87年3月）：頁25。