東亞觀念史集刊　第十期
2016 年 6 月　　頁 389-399

「東亞聚焦：2015第六屆數位典藏與數位人文國際研討會」會議報導

The observation Report of 6th International Conference of Digital Archives and Digital Humanities

彭　醴　璃*

Lily Perng

「東亞聚焦：2015 第六屆數位典藏與數位人文國際研討會」於 2015 年 11 月 30 日至 12 月 2 日在臺灣大學舉行，臺、中、日、韓與歐美地區學者齊聚一堂，三日之內，計有 3 場專題演講、35 篇論文發表討論，所論以數位典藏、數位人文領域相關問題爲主軸，綜之約有三端：一、數位平台建置與技術開發；二、人文領域研究方法之開拓與革新；三、數位人文學發展之挑戰與反思。茲分述如下。[1]

---

* 作者現爲政治大學中國文學系碩士生。

1 本文所論研究成果，詳見項潔主編：《東亞聚焦：2015 第六屆數位典藏與數位人文國際研討會論文集》，國立臺灣大學數位人文研究中心主辦，會議時間：2015 年 11 月 30 日 -12 月 2 日。

一、數位平臺建置與技術開發

數位人文乃一橫跨人文、資訊、社會科學等學門之領域，借重資訊科學與社會科學的技術，人文學科得以突破傳統治學、論述之法，在舊材料中發現新觀點，而數位資料庫、平臺之建置與相關輔助技術之開發，便是啟動數位人文研究的起始點。

設置可供研究者檢索、分析資料的數位平臺，乃本次會議中諸多學者關懷所在。日本關西大學外國語學科教授內田慶市〈東亞文獻資料的電子化的現狀和未來〉，便著力強調設置文獻電子化、數位化的資料庫之重要，資料庫不僅便於文獻保存，亦便於研究者取用、整合資料，世界諸多大學、機構皆以著手進行文獻電子化的工作，可見資料庫建置乃世界共同的大趨勢。這些資料庫若能建立國際聯絡網，數位資料的運用空間將會更大、更彈性。然內田氏並不認爲電子化資料可完全取代紙本，「進入書籍海」仍是身爲研究者不可缺少的功夫。

至於資料庫實際建置的成果，首先是中央研究院史語所研究員暨傅斯年圖書館主任劉錚雲〈我們如何建全文資料庫：中央研究院史語所漢籍電子文獻資料庫的回顧與展望〉，詳細介紹中央研究院漢籍電子文獻資料庫之涵蓋內容與建置流程，該資料庫內容涵蓋經、史、子、集四部，收錄歷代典籍925 種，計5 億1 百萬餘字，且每年以新增2 千萬字的規模成長。「漢籍」資料庫建置方法、態度與坊間商業公司大不相同，在將典籍數位化的過程中，透過人工嚴謹考察，將原始典籍影印本描字錯誤、漏頁、錯印、錯置、史實缺誤等問題一一改正，力求輸入數位文本之正確無誤。除此之外，該資料庫尚提供以文找文、內文比對、標誌（tag）管理等服務，以期打造適合研究者查找、分析資料的研究平臺。

而美國哈佛大學費正清中國研究中心博士後研究員德龍〈自動化與合作：數位媒介的開發〉，則介紹其經營多年之 Ctext.org 平臺，即「中國哲學書電子化計畫」（Chinese Text Project），該平臺目前收錄

600 萬種傳世文獻、1300 萬頁影印資料、30 億字維基資料，且提供各種輔助功能，如：應用程式介面（API）插件，使用者可依其所需進行功能擴充；整合檢索系統；自動參照通知（pingbake）；標記功能（Xml）等，以便使用者快速查找、統合平臺中的資料。此外，該平臺利用群眾外包（crowd-sourcing）的方式進行數位文本校對，鼓勵群眾參與，目前已有6 千位自願貢獻者投入其中，不僅展現數位人文協作之精神，亦利用數位優勢，讓數位平臺提供給使用者更豐富、全面的服務。

南通大學人文學院教授、副校長、中國屈原協會副會長周建忠〈東亞楚辭文獻數據庫建設及語義化應用研究進展綜述〉，則展示南通大學建置之「東亞楚辭文獻數據庫」，該數據庫目前收錄楚辭研究文獻1 萬5 千餘條，然其並不以收羅資料爲足，而是將焦點放在建構資料庫知識本體與開發文本語義標註系統，力求將改善關鍵字檢索時所得資料的孤立、零散的局面，連結不同文獻中相互關聯的文本，並以視覺化形式呈現，並藉由切分詞語、半自動標引技術，提供使用者進行語義化研究。未來將建成多語種楚辭語義辭典，並改善語義標註、檢索系統，以便使用者更好地對楚辭文獻進行跨學科、地域之研究。此外，中國社會科學院民族文學研究所資料中心助理研究員郭翠瀟、王憲昭等，亦投入資料庫之建置，〈中國神話母題編目數據庫的設計、實施及應用〉一文，便以其團隊開發之「中國神話母題 W 編目數據庫」爲焦點。該數據庫乃據王憲昭《中國神話母題 W 編目》一書建置，該書收錄1 萬2 千6 百篇中國各民族神話，下分3 萬3 千餘神話母題，是中國第一部全面提取中國各民族神話母題名稱、系統擬定母題代碼的神話著作。以該書爲紐帶，「中國神話母題 W 編目數據庫」得以將大量神話文本有系統的組織起來，並進行多維度的檢索，且可於互聯網上開放使用，信能大大嘉惠海內外神話研究學者。

除了單一資料庫的建置外，獨立資料庫之整合亦是使用者能否順利取用資料的關鍵。中央研究院近史所張哲嘉副研究員〈MHDB 的

數位歷史網路建構與發展〉，便論及中央研究院「近代史數位資料庫群組」（Modern History Databases）之建置。MHDB 的是一個鏈接中研院各大主題資料庫的入口網站，使用者可藉之連結至「胡適檔案檢索系統」、「報刊資料檢索系統」、「英華字典資料庫」、「清代奏摺檔案資料庫」等多種中央研究院開發之主題資料庫，這個入口網站對全世界免費開放，任何人皆能自由取用其中的數位史料，並使用各別資料庫提供之詞目比對、時間軸比對、詞頻綴詞分析、圖表統計等數位分析工具，藉此開展數位人文的研究路徑。而日本立館大學專門研究員 Biligasikhan Batjargal 等〈日本人文資料庫的跨資料庫雙語近用：以英文與日文關鍵字進行文本檢索〉，亦論及跨資料庫整合之重要性。該團隊爲日本已上線之人文資料庫以及立命館大學藝術研究中心轄下 40 個不同資料庫，分別建置聯合查詢雛型系統，可利用雙語（英、日）關鍵字進行跨資料庫檢索，成功建置一可取用多樣化資訊的跨庫檢索系統。此亦是資料庫整合的成功典範。

建置資料庫，除全文搜尋外，提供使用者分析、整理資料的功能亦同等重要。臺灣大學資訊工程學系特聘教授暨數位人文研究中心主任項潔〈電子文獻的再脈絡化〉與法鼓文理學院佛教學系副教授暨圖書資訊館館長洪振洲〈建構以輔助學術研究爲導向之漢籍佛典數位平臺〉二文便是對資料庫深度資料搜尋、整合等問題的回應。前者認爲資料庫全文檢索忽略原始資料的脈絡與結構，故試圖解構原典脈絡，重新整合知識本體，如其建置之「『藝文類聚・太平御覽』資料庫」，便將《太平御覽》全文依其書之門類、部目層層劃分，使用者可全文展讀，亦可從部目、小目去觀察資料的分層結構，甚至可與結構相近之《藝文類聚》相較，進行不同文本間的參照分析。此外，其更以 TEI（Text Encoding Initiative）文獻編碼標註文本，將文獻中詞與分類（如：人名、地名、動物名等）標註，方便使用者以類別查詢、整理資料。而後者則爲突破目前佛學研究瓶頸，建立一整合良好、內容豐富的佛典數位資料平臺，即「佛學數位整合平臺」。該平臺不僅

協助整合 CBETA（中華電子佛典）、佛教規範資料庫、佛學術語字辭典等資料庫，同時亦提供文字量化分析工具（Ngram 量化統計）、資料編修與紀錄工具、多種查詢方式（依據部類、冊別、經目等）等服務，期能達到輔助漢籍佛典文獻學術研究的目的。藉由二者的努力，信可更加完善數位平臺資料搜索、整理、分析等不同層次的功能。

時序推移，身處資訊化時代，人文學者處理資料的方式亦須因勢革新，借重資訊領域相關技術，人文研究遂可有不同以往的新視野、新取徑，中央研究院歷史語言研究所特聘研究員邢義田演講〈居延漢簡資料庫的建置與未來〉，就多次談及數位技術（如資料庫、紅外線掃瞄、地理資訊系統等）對資料保存、現行研究方法的重大影響。金門大學土木工程管理學系教授吳宗江、林宜君博士〈整合空間資訊與 SfM 理論於較大區域古蹟數位保存之研究：以金門明遺老街爲例〉，便欲藉由2D 影像塑形3D 結構場景的 SfM（Structure from Motion）理論與技術，試圖保存金門明遺老街的影像，建立其3D 數位模型，雖然其實驗結果因模型組構誤差累積導致影像末端變形，然透過技術改良（整合高精度空間控制座標點位），仍不失爲大區域古蹟保存的可行方案。

臺灣大學資工所博士候選人杜協昌〈半自動詞彙擷取：簡化的詞夾子方法以及其 JavaScript 元件的開發與應用〉，則藉由 JavaScript 元件開發一套從文本擷取詞彙的系統——簡化的詞夾子。這套詞夾子程式可幫助使用者從大量數位文本中，擷取出有意義的詞彙，杜氏於會議中便以《紅樓夢》爲底本，成功夾取其中的人物名稱。藉由詞夾子優異的擷詞功能，使用者可更快速、全面的對大量文本進行分類研究。而日本學者與蒙古學者 Biligsaikhan Batijargal、Garmaabazar Khaltarkhuu 等〈利用機器學習擷取蒙古歷史檔案之人名〉，同樣提供一套資料擷取的方式，該團隊利用支持向量機（Support Vector Machine，SVM）監督式學習方法擷取蒙古文獻中的人名與地名，訓練標註資料庫學習人名、地名擷取的規則，便可對文獻進行標註，大

幅節省歷史文本整理所需的人力。

至於臺灣大學工程暨海洋工程研究所副教授黃乾綱等〈利用文脈強化文字辨識正確率的中文古籍數位工具〉則提供一套提高光學文字辨識軟體正確率的半自動數位化資訊系統，以便加速古籍數位化工作。由於古籍文字與一般現代印刷書籍在板式（古籍一頁分兩欄、三欄）、樣態（古籍有雙行小註、夾註）、字體（古籍手刻、手抄本同一字字體仍略有差異）多有差異，以致一般光學文字辨識軟體運用於古籍上效果不佳，故該團隊改以機器學習法提升文字辨識率，包括加強切字正確率、文字影像分群等技術來協助辨識結果校對等，其實驗結果準確率可達68% 以上，可大幅降低典籍數位化所需人工文字輸入量。待其程式更加完善後，對研究資料建置實大有裨益。

最後是臺灣大學生物產業傳播暨發展學系副教授闕河嘉、臺灣大學圖書資訊學系教授陳光華〈中文獨立語料庫分析工具之開發與應用〉。數位人文的研究工作需要資訊與人文學者攜手共進，然而，當人文學者發想出新的研究議題，身邊卻沒有可配合的技術人員時，便會陷入研究停擺的窘境，而該團隊提出之「庫博中文語料庫分析工具」便是對此一問題之回應。該語料庫是以語料庫語言學（corpus linguistics）爲基礎的電腦輔助文本分析軟體工具，提供七大功能：1.KWIC（關鍵詞檢索上下文工具）2. 搭配詞 3. 詞彙分布 4. 語料庫比較與正負關鍵詞分析 5. 自建辭典 6. 子語料庫 (sub-corpus) 建立技術 7. 詞彙權威控制功能，專以輔助人文學者進行數位人文研究。幫助人文學者跳脫傳統研究方法的限制，可更容易進行數位人文方法的研究。闕氏於會議中展示庫博的資料庫介面與處理數位資料的方式，操作主題爲「有機農業」概念在《一步一腳印，發現新臺灣》節目中的意義，以見電腦軟體工具對人文研究之重大革新。

二、人文研究方法之開拓與革新

傳統人文研究重視學者對資料精讀（Close Reading）後所作的質性分析，長於細膩深刻的探討發掘，然短處是不夠具象、客觀，也無法應對過於龐大的文獻資料，而數位技術的建置與發展，正可輔助傳統人文研究之不足，開創以電腦輔助取徑（Computer-aided Approach）的人文研究範式。

暨南大學東南亞學系教授兼系主任李美賢、闕河嘉〈臺灣「東南亞新二代」的形象建構〉透過語料庫語言學工具之輔助，分析2005至2015年10月之間《自由時報》、《中國時報》與《聯合報》對東南亞新二代住民的形象建構，以及背後反映之臺灣社會價值。北京大學高等人文研究所博士後研究員郜謐俠〈以數量方式判定《老子河上公章句》成書年代〉則透過CHANT database（漢達文庫）比對五十種先秦、西漢典籍與《老子河上公章句》中100個專有術語的疊合程度，藉由數量分析考證《老子河上公章句》應是介於1世紀中、後期，以客觀證據證實饒宗頤等人對該書年代之推定。德國馬克斯・普郎克科學史研究所研究員馬君蘭、陳詩沛〈運用中國方志重建地方上的物質文化與認同〉則採用愛如生中國方志庫與數位分析工具、標記（tagging）技術，以中國8000餘種方志中的物產爲焦點，探討各地方紀錄物質、建構物質認同背後的思維模式。至於德國馬克斯・普郎克科學史研究所博士後徐源、荷蘭萊登大學區域研究所研究員何浩洋〈以電腦地圖科技跨越六朝草藥於地理及知識領域的分界〉同樣透過資料庫、半自動文本標記系統MARKUS以及馬克斯普朗克科學史研究所建置的數位工具平臺，分析道藏、大藏經中的草藥，使醫學史研究更加事半功倍。

以上四份研究成果，皆非單一領域之傳統研究方法所能企及，誠如陳詩沛於會中所言，數位人文研究方法的建立乃是對漢學家與資訊學家的挑戰，前者需要擁有開放的心態（mindset），願意接受、發展

新的研究方式，並嘗試學習使用數位工具，後者則須學習與人文學者溝通，創立符合其研究需求的工具。二者相輔相成，不僅過往難以處理之議題有突圍之機，亦可打開傳統研究領域的新局面，如結合數位工具的概念史研究。

韓國翰林大學翰林科學院 HK 研究教授宋寅在〈韓國近代期刊資料庫的建設以及殖民時期韓國的關鍵字〉中，結合概念史研究與數位人文方法（資料庫建構），對韓國近代思想世界進行宏觀考察，其運用韓國翰林大學所建置之資料庫（HCKCH），結合頻率分析、共現詞頻分析、主題模型（topic model），分析韓國在 1945-1985 年間重要的關鍵詞與觀念群（概念地圖）。宋氏認爲資料庫建置開創觀念史研究的新局面，使學者得以突破人力的侷限，宏觀看待思想世界。而政治大學中國文學系教授鄭文惠等所發表之〈概念關係的數位人文研究：以《新青年》中的「世界」觀念爲考察核心〉與〈情感現象學與色彩政治學：中唐詩歌白色抒情系譜的數位人文研究〉，前者結合統計、資科之多種關係性統計法，考察《新青年》中「世界」概念與其他重要概念之間共現離合的變化；後者則藉由數位技術勾勒、錨定中唐詩中前後詞綴爲「白」的構詞，觀察白色構詞與對仗詞、搭配詞的聯組關係，勾勒「句鍊」的結構模型，使文本概念的分析從以前的「斷詞」前進到「構詞」思維，並藉此看出中唐文人的心理情感與思想觀念及社會文化變革。兩者皆是透過數位人文方法研究文本概念，實是傳統概念研究上一大突破。

數位工具發展帶動人文研究之革新，然而，並非每一個人文研究者都能找到資科、社科背景的研究夥伴，能夠針對自己的研究打造相應的工具，此時，學習使用現成的數位工具，亦是涉入數位人文研究的取徑。中央研究院人社中心地理資訊科學研究中心研究助技師廖泫銘〈從歷史 GIS 朝向空間人文學發展〉便論及現有的 GIS 軟體爲史學研究、教學開拓之新路徑。而美國學者 Peter Broadwell 等〈從韓國演歌到文化科技：韓國流行音樂生產網絡的歷史發展〉、捷克馬薩里

克大學漢學系副教授路丹妮、臺灣師範大學英語學系助理教授陳正賢〈用數位人文研究方法重建臺灣戰後初期的文學地景〉、清華大學博士生趙薇〈「社會網絡分析」（SNA）在現代漢語歷史小說研究中的應用：以李劼人的《大波》三部曲爲例〉與金門大學閩南文化研究所教授李宗翰等〈范成大《吳郡志》中的社會關係網絡：以 CBDB 與 Pajek 作爲分析工具〉等四篇研究，則分別採用社會網絡分析軟體如 Pajek、Gephi 等，將文獻中散亂錯綜的訊息透過節點、線段連接成具象化的圖示，以說明物件間的互動、關係、影響，爲人文研究提供量化、視覺化的綜合分析方法。

當然，數位人文的方法除對學者撰寫文章有所裨益外，對圖書館領域的工作亦有所助，英國大英圖書館研究員 Nora McGregor 等〈利用群眾外包增加大英圖書館中的中文以及印尼文的資料公共取用〉就利用群眾外包（crowd sourcing）的方式將大量的中文、印尼文館藏目錄數位化，從此讀者不須親臨大英圖書館的閱覽室，在線上就能找到自己所需的資料。此一創舉既省時力，亦充分發揮數位人文的協作精神。

三、數位人文學之挑戰與反思

人文、資訊、社科等領域交叉合作，激盪出數位人文的研究取徑，新方法造就新議題、新視野，能量無限、蓄勢待發，但走上數位人文研究之路，仍有其隱憂。中山大學中文系教授劉文強於會中報告〈三皇五帝、五帝三王：數位人文下的新方法與新議題〉一文間，便不斷談及人文學者面對數位平臺挾帶的大量數據時的徬徨與迷惘，大數據有助學者對手邊議題更宏觀、全面的理解，然海量資料的閱讀、處理對習慣傳統研究方法的學者而言實是一大負擔。數位資料數量龐大且便於取得，如何找到正確方式安置、消化、編排這些大量資料，便成爲研究者不可迴避的第一項挑戰。

除面對大量資料外，學者還需時時警惕數位資料的可信程度。日本京都大學東亞訊息研究中心教授 Christian Wittern〈值得信賴的數位文本〉便再三提醒文本可信度與轉錄數位文本所應有的編輯規則與程序。京都大學人文研究中心便致力於開發一款能更貼近使用者習慣，並且值得信賴、可持續使用的數位平臺。惜該平臺現在仍在測試階段，尚未公開，然其平臺開發的精神，確實值得所有研究者深思。

彰化師範大學歷史學研究所副教授暨所長李宗信〈學科邊界的消融或強化？淺談當今歷史 GIS 的侷限和挑戰〉則解析當今歷史研究引入地理資訊系統之挑戰，一是基本史料數位化仍有不足；二是各家史料地圖數位化的標準不一致；三是大型資料庫未整合；四是現有的 GIS 工具，如臺灣歷史文化地圖系統（THCTS）Web-GIS 系統仍有種種使用上的不便。而臺灣師範大學臺史所教授兼所長張素玢〈歷史 GIS 的新境與困境〉則言 GIS 系統加入史學領域，在教學上缺乏專業人才，且學生多缺乏相關操作經驗，學習亦有無力感；而在研究上，常有研究者不問 GIS 軟體中圖層的來源，隨意使用，用圖卻未好好判圖，無法透過地圖呈現欲表達的結構。且古地圖與今地圖時常有定位誤差，史料地點難與今日地點之對應。由二人的報告可知，數位工具縱然方便使用、易於取得，然若躁進粗心，未能深入驗證軟體提供資料的虛實，研究的新工具反成新阻力。

清華大學中文系副教授祝平次〈臺灣數位人文發展的問題〉則直指臺灣目前數位人文學發展的問題，包括國家機構重理工而輕人文，理工學者與人文學者合作時，兩邊皆自認爲是研究中心，容易產生摩擦。又人文機構對數位技術的支援不足，傳統人文學者除非自我加強數位知識與能力，否則難以涉入數位人文的研究，畢竟數位人文學不僅止步於全文檢索。此外，數位人文若要推廣，教學也扮演重要的腳色，然相關人才稀缺，需透過線上學習與國際合作等方式改善。其於會中又發表另一篇論文〈文字資料 vs. 資料庫：朱子語類研究〉，運用 NodeXL、MSexcel、正規表達式、QGis、MARKUS 等軟體將《朱

子語類》文本整理成類似資料庫般容易取用、分析資料的狀態，正是人文學者自行充實數位能力後，完成數位人文研究的最好示範。

最後，在數位時代，當所有資料都資訊化以後，圖書館還有存在的必要嗎？德國柏林國立圖書館東亞部主任 Matthias Kaun〈圖書館在當今和未來是否還有存在的必要？從一所德國圖書館談資訊服務之於東亞研究的作用〉便論及未來圖書館存續問題，其主張圖書館應致力於開發整合性數位平臺，以便數位人文研究人員更方便取用、分析資料。此外，建立跨國分散式全文平臺以支援數位資料存取遞送，亦勢在必行，圖書館須因應數位時代的挑戰努力轉型，方能永續經營，與人文學者的研究一樣，在數位浪潮中，開闢嶄新領域。