



# 從人文到數位人文： 知識微縮革命與人文研究範式的轉向

鄭文惠\*

數位時代是一個具革命性與顛覆性的知識媒體新時代，知識與研究，乃至人類行為與國家治理、區域發展等，因著數位媒介的轉變，產生了劇烈的變動。就研究而言，隨著文獻的數位化工程，啟動了知識微縮革命，改變也主導了知識的產製與文明的型態，當然，無國界、無地域，不分你我而同潤共享的龐大的數位資源，也對人文研究造成巨大的衝擊與影響，全球學術研究更產生前所未有的蛻變與進展。

2008年前後，隨著大規模語言文字及多媒體文化文本資料庫建置技術的積累與精進、數位文本標引的標準化與格式化、新傳播媒體和呈現技術與元資料（metadata）觀念及 GIS 系統方法迅速發展，資訊累積呈指數成長，人類世界進入了一個嶄新的大數據（Big Data）時代。無以計數的數位化文獻如何為學術研究所用？或是學術研究如何帶動文獻數位化工程更加智慧可用？無疑是一個全新的挑戰。

數位資源確為學術研究帶來巨大的轉變，以往學者以少數物件或有限文獻進行研究的模式，因著巨量的數位資源產生根本性的改變，學術研究越來

\* 國立政治大學中國文學系教授，為「中國近現代思想及文學史專業數據庫（1830-1930）」主編及科技部「觀念・事件・行動：中國近現代觀念形成與演變的數位人文研究」整合型計畫總主持人。「中國近現代思想及文學史專業數據庫（1830-1930）」目前只開放八千萬字於國立政治大學圖書館及校內使用，預計 2015 年校對完成。在數據庫所收錄文獻智財權尚未處理完畢，未能正式公開給全球學界使用期間，國內外任何研究者需要相關資料，數據庫計畫辦公室均可提供服務，請逕至「中國近現代思想及文學史專業數據庫（1830-1930）」辦公室查詢或 e-mail 至 dsmtcl@gmail.com 信箱，計畫助理會協助查詢並傳送相關資料以供研究。2010 年至今，全球學界運用本數據庫資料，已完成百篇以上論文。本文部分論述，參見鄭文惠：〈近代中國知識轉型與概念變遷 / 觀念形塑——觀念史 / 概念史方法與視域〉，《東亞觀念史集刊》第 4 期（2013 年 6 月）一文，以及鄭文惠：〈觀念史方法與中國研究〉，收入《東亞的知識交流——關鍵概念詞的再檢討》（京都：日本國際日本文化研究センター，2013 年 12 月）一書。

越傾向跨時段、跨地域的長期觀察與研究，以及跨領域、跨國跨區的交流、對話與結盟。然而，在大數據時代，伴隨著知識的數量化，又面對巨量的數位文獻，如何從龐雜的資料中萃取重要資訊，在在挑戰著人文學者，因而人文思維結合數位技術，勢必是一條實然且應然必走的路。

臺灣學界面對這一全新的時代挑戰，除了從九〇年代開始啟動長達二十年的數位典藏計畫外，2013年科技部也開啟了數位人文主題研究計畫，共分(1)善用新工具：數位科技支援人社研究下的新方向；(2)理解新現象：由人文及社會科學反思融入生活脈絡裡的數位科技；(3)改變生活世界：設計科學(design science)觀點下數位與文化創意的結合。早在科技部啟動數位人文主題研究計畫之前，臺灣大學項潔教授主持「數位典藏與自動推論實驗室」，除建置相關臺灣文史資料庫外，並以資訊技術開發切合人文研究的數位工具，希望提供人文研究一個全新的視野。又持續數年召開「數位典藏與數位人文國際研討會」，出版數位人文叢書，大大推進數位人文的研究。

金觀濤和劉青峰教授1997-2007年在香港中文大學因著近現代中國政治思想轉型研究上的需求，建立了1830-1930年間包含中國近現代報刊、雜誌、傳教士和西方著作中譯本及各種文集，共1.2億字的「中國近現代思想史專業數據庫(1830-1930)」。<sup>1</sup>在研究型數據庫的基礎上，兩位教授利用數據庫方法進行觀念史研究，十年間完成《觀念史研究：中國現代重要政治術語的形成》<sup>1</sup>一書，探究「社會」、「共和」、「權利」、「個人」、「經濟」、「科學」、「革命」等重大現代觀念形成及相應近百個重要術語的演變。所謂「數據庫方法」是指研究者在進行研究時處理電腦數據的新方法，它有助於研究者思考如何從數據庫中提取數據，並通過分析發現研究對象中易被忽略的關係與向度。運用在觀念史研究上，主要是指從專業數據庫中提取某一觀念相對應的關鍵詞例句，進而分析該關鍵詞不同的意義類型及演變過程與年代分布頻度，以發現並勾勒觀念的形成與演變軌跡及與社會事件互動的進程。<sup>2</sup>將數據挖掘方

<sup>1</sup> 金觀濤、劉青峰：《觀念史研究：中國現代重要政治術語的形成》(香港：香港中文大學當代中國文化研究中心，2008年)。

<sup>2</sup> 金觀濤和劉青峰分別發表過〈數位人文研究的理論基礎〉和〈觀念史研究與數據庫的建立和應用〉二文，擊劃出數據庫方法與數位人文學作為觀念史研究之理論模型與方法建構，同時也揭示了數位人文學作為一門新興學科的可能性與可行性。參見項潔編：《數位人文研究的新視野：基礎與想像》(臺北：國立臺大出版中心，2011年)。



法系統運用到觀念史的研究中，實質為觀念史研究帶來新的視野與觀點，也帶來新的方法和技術，大大改變了傳統研究的方式與方法。然而，這僅止於指涉於觀念的一個個關鍵詞的研究，且以詞頻判斷關鍵詞的重要性，仍存在著指標性不足的疑慮。

迨至 2010 年筆者有幸與金教授在原有基礎上共同主持「中國近現代思想及文學史專業數據庫(1830-1930)」，結合數位技術人才持續開發數據庫功能，並由金教授帶領「中國現代認同的形成」研究群，進一步落實人文思維結合數位技術的研究探索，在原有關鍵詞數據挖掘方法中，逐而開展出一個以人文研究為主體，結合數位技術與統計方法的新興的數位人文學領域。

從 2010 至 2012 年，本研究團隊持續探索有關觀念史研究的數位人文方法：如以指涉於觀念的關鍵詞為研究核心，開發出前 / 後詞綴數位方法：以前 / 後詞綴為主，透過數位技術快速勾勒出相同詞綴的不同詮釋向度。如以「後詞綴」數位方法，觀察「主義」觀念的形成、發展與變化的過程，除可從「中國近現代思想及文學史專業數據庫(1830-1930)」共五萬一千多筆「主義」資料中篩除不相關資料 92% 外，又可快速確定近代中國百年間共出現 1,680 種不同名稱的主義，並給出每年「主義」的種類數目、各種主義在每年的使用頻度、任一「主義」的發展趨勢、特殊高峰期相關「主義」的關係和演變脈絡、「主義」的脈絡化轉向與時代語境，宏觀的掌握「主義」這一關鍵詞所指涉的近代中國的重要概念 / 觀念的形成與演變軌跡，以及其在近代中國產生的影響。如 1903 年出現次數最多的「民族主義」、「帝國主義」及「社會主義」，從中即可進一步分析三者之間的指涉關係，以及其作為取代傳統儒家意識型態，成為新的意識型態之意義。<sup>3</sup>

或以關鍵詞叢擷詞技術，勾勒出指涉於重要概念 / 觀念關鍵詞出現前的醞釀語境，進行「前概念 / 觀念」的數位人文研究，主要在觀察概念 / 觀念如何從時代思潮中逐漸凝縮成「短語」，最終形成「關鍵詞」的詞化過程。<sup>4</sup>或以事件為核心，擷取共現的觀念詞叢，著重詞彙的語境與共現的詞叢，進行

<sup>3</sup> 參見詹筌亦、王乃昕：〈「主義」的數位人文研究〉，收入項潔編：《數位人文在歷史學研究的應用》（臺北：國立臺大出版中心，2011 年），頁 219-245。

<sup>4</sup> 參見邱偉雲：〈關鍵字叢與文本意義挖掘的嘗試：以《清季外交史料》為例〉，收入項潔編：《數位人文在歷史學研究的應用》（臺北：國立臺大出版中心，2011 年），頁 159-188。

「共現」詞頻之分析，進一步發展出事件與概念／觀念互動的數位人文方法，<sup>5</sup>藉以觀察在事件發展過程中，不同概念／觀念在事件發展史上，如何受事件的影響而產生變化的過程。

另，又利用 PAT-Tree 數位擷詞技術，以及根據 Zipf-Mandelbrot 模型計算出的理論曲線，統計出「準詞彙」的偏離值，從而得知最大偏離段及確定關鍵詞，藉此一則解決主觀判斷的爭議及減少人工篩選的負擔；二則也能更加準確的篩選出關鍵詞，以細緻的觀察思想變化及反映在文本上的語言現象。<sup>6</sup>此外，或論析自然語言處理技術於中文史學文獻分析之應用；<sup>7</sup>或探索漢文文獻之外來語音譯詞擷取方法等。<sup>8</sup>

2013 年 8 月通過科技部「觀念・事件・行動：中國近現代觀念形成與演變的數位人文研究」整合型計畫，並創立「歷史與思想數位實驗室」研究團隊。總計畫（主持人：國立政治大學中文系鄭文惠教授）之下共有五個子計畫：「觀念・事件・行動：中國近現代『國家／世界』觀念形成與演變的數位人文研

<sup>5</sup> 參見金觀濤、邱偉雲、劉昭麟：〈「共現」詞頻分析及其運用——以「華人」觀念起源為例〉，收入項潔等主編：《數位人文要義：尋找類型與軌跡》（臺北：國立臺大出版中心，2012 年），頁 141-170。

<sup>6</sup> 參見金觀濤、梁穎誼、姚育松、劉昭麟：〈統計偏離值分析於人文研究上的應用——以《新青年》為例〉，《東亞觀念史集刊》第 6 期（2014 年 6 月）一文。

<sup>7</sup> 參見劉昭麟、金觀濤、劉青峰、邱偉雲、姚育松：〈自然語言處理技術於中文史學文獻分析之初步應用〉，國立臺灣大學數位典藏研究發展中心、國立臺灣大學圖書館主辦：「第三屆數位典藏與數位人文國際研討會」（2011 年 12 月 1-2 日）會議論文集。

<sup>8</sup> 參見王昱鈞、呂翊瑄、蔡宗翰、劉青峰、金觀濤、劉昭麟：〈漢文文獻之外來語音譯詞擷取方法〉，第四屆數位典藏與數位人文國際研討會，國立臺灣大學數位典藏研究發展中心，2012 年 11 月 29-30 日，頁 247-263。另，臺灣學界也有進行社會網絡模型研究，或運用歷史空間化數位方法，重新解讀歷史，或進行譯者、作者風格量化分析研究。分別參見劉吉軒、柯雲娥、張惠真、譚修雯、黃瑞期、甯格致：〈以文本分析呈現臺灣海外史料政治思想輪廓〉，收入項潔等主編：《數位人文要義：尋找類型與軌跡》（臺北：國立臺大出版中心，2012 年），頁 83-116。張嘉文、歐仲翔、翁稷安、項潔：〈開放式歷史地理資訊系統——日治時期臺灣主題式分布圖繪製系統之開發與應用〉，國立臺灣大學數位典藏研究發展中心、國立臺灣大學圖書館主辦：第三屆數位典藏與數位人文國際研討會（2011 年 12 月 1-2 日）。余清祥：〈統計在《紅樓夢》的應用〉，《政大專報》第 76 期（1998 年 6 月），頁 303-327。謝承恩、洪振洲、馬德偉：〈結合漢典古籍虛詞常見字與統計量化分析進行漢譯佛典譯者風格辨別〉，收入項潔等主編：《數位人文要義：尋找類型與軌跡》（臺北：國立臺大出版中心，2012 年），頁 117-140。杜協昌：〈利用文本採礦探討《紅樓夢》的後四十回作者爭議〉，國立臺灣大學數位典藏研究發展中心、國立臺灣大學圖書館主辦：第四屆數位典藏與數位人文國際研討會（2012 年 11 月 29-30 日）。凡此種種新開發的數位工具與方法及其研究與成果，均是臺灣人文學界面對大數據時代知識微縮革命衝擊下的回應與探索。



究」(主持人：國立政治大學中文系鄭文惠教授)、「具有『中國特色』的主義：近代中國的『家國』想像與數位人文研究」(主持人：國立政治大學歷史系楊瑞松副教授)、「文學關鍵詞的指涉與詮釋研究——『中國文學』在戰後臺灣的意義(1945-1965)」(主持人：國立臺灣大學臺文所所長黃美娥教授)、「爭辯中國：中國研究在戰後臺灣的觀念變遷」(主持人：國立中山大學中國與亞太區域研究所副教授)，主要聚焦於中國近現代/臺灣冷戰時期的國族認同議題，另「以人工智慧與自然語言處理技術為基礎之數位人文資訊服務：資訊檢索、時空分析、音譯詞彙、實體名詞辨識與社群計算」計畫(主持人：國立政治大學資科系劉昭麟特聘教授)，則致力於開發數位技術，協助人文學者進行跨領域研究。本整合型計畫主要結合文學、歷史、思想、國際外交、資訊科學、統計等領域跨校系及國外學者十餘人，從數位人文學科建置的角度出發，以中國/臺灣近現代觀念形成與演變為觀察對象，透過數位技術與人文研究跨領域的合作與應用，冀望能拓進數位人文技術，提出數位人文學方法論，並期望能在學術研究上有新的革命性的蛻變與突破，俾使數位人文學作為一門新興學科的建置更為落實，從中更冀望能創構出一種新的人文學研究範式。

至今年年底短短不到一年多時間，團隊成員在科技部與臺灣各大學及美國萊斯大學(Rice University)、韓國翰林大學(Hallym University)、日本關西大學、捷克科學院(Academy of Sciences of the Czech Republic)等演講相關議題近20場，在美國、瑞士、葡萄牙、日本、中研院近史所等發表超過30篇論文，每週更帶領「DHRG 政治大學數位人文學研究群」(國立政治大學中文系鄭文惠教授、國立政治大學中文系博士後邱偉雲、國立政治大學統計系博士候選人梁穎誼共同主持)舉辦共計28場次研討活動，積極培育數位人文研究人才，以厚植數位人文學科發展潛力。今年11/21-11/22本計畫與「中國近現代思想及文學史專業數據庫(1830-1930)」、「中國形塑與現代國家的形成」(主持人：國立政治大學中文系鄭文惠教授；共同主持人：國立政治大學金觀濤講座教授)在日本將與關西大學東西學術研究所亞洲文化研究中心、韓國翰林大學共同主辦「大型數據庫與數位人文研究：東亞近現代觀念的形成與演變」國際工作坊，12/5-12/6與國立政治大學中文系共同承辦科技部人文司「從人文到數位人文」學術研習營。本計畫希望無論在研究成果或技術創新、學科建置、人才培育等方面，能盡棉薄之力。

當然，作為一門新興的學科，數位人文學只是剛起步，還有很多課題等待拓展、很多問題有待克服，也亟需建立理論模型與方法架構。本研究團隊目前聚焦在時空分析、社群計算、引文分析及關鍵詞篩選技術上，嘗試將物種聚落統計<sup>9</sup>或單字組合頻率統計計算方式等概念，匯入關鍵詞篩選技術上，希望能為人文研究者先行篩選出大部分非詞彙部分，以降低研究所耗費的時間；另，為了使詞頻判斷關鍵詞更具指標性，也導入統計學敘述統計（Descriptive statistics）視野及立基於人文思維，提升 Network 技術中的關係計算效能，著重於概念／觀念共現的多節點網絡關係性的挖掘，<sup>10</sup>以掌握巨量語料中更具指標性意義的關鍵詞，以及其指涉於概念／觀念的發展軌跡與論述結構。未來希望在「技術」、「應用」、「方法」層面取得更進一步的拓進、精深，除著重於以數位方法尋找核心關鍵詞，為中文歷史文獻分析建立基礎外，並將時空分析、音譯詞彙、實體名詞辨識與社群計算重作檢索環境；此外，亦擬將分析工具如高頻詞彙擷取與分析、關鍵詞彙的時序分析、關鍵詞彙與共現詞彙（collocations）的時序分析的繪圖等整合於新的檢索系統，過去程式分段運作，也希望在未來朝向數位人文技術自動化發展。除了希望厚植數位人文學這一新興學科的建置基礎外，也希望進一步促使概念史／觀念史有新的方法上的更新與革命性的研究突破。在以人文研究為主體並結合數位運算技術的數位人文方法，實質為人文研究提供了一種新方法與新視野。對本計畫主要研究的觀念史／概念史而言，在研究過程中，一則須不斷精進數位人文技術與方法，建構數位人文學方法論，驗證數位人文方法的有效性與可行性，以奠定數位人文學新興學科的建置基礎；二則在勾勒中國近現代及臺灣戰後國族認同的形塑與轉變過程中，用數位技術與方法突出觀念・事件・行動緊密關聯、彼此作用的複雜關係與脈絡化轉向，實質提供了一種「觀念・事件・行動」關係研究的詮釋模型，從中可勾勒出觀念轉化為社會行動、社會行動又改變觀念的互動鏈的大歷史結構，相信透過不斷精進的觀念史／概念史的數位人文研究，當能有效促使歐美觀念史／概念史研究從語言論轉向於數位人文的新研究範式。

<sup>9</sup> 主要由團隊成員國立政治大學統計系余清祥教授提出。

<sup>10</sup> 主要由團隊成員國立政治大學中國文學系博士後邱偉雲與國立政治大學統計系博士候選人梁穎誼提出。



因應著數位時代知識微縮革命及大數據時代的快速推進，當代研究者勢必將面對各式各類的資料庫與巨量的數位文獻。在巨量多元的數位文本作為研究的文獻類型語境下，研究者必須開始嘗試轉向「人文研究思維」與「數位推論技術」結合下所形成的「數位思維邏輯」，方能適切而有效的運用巨量的數位文獻。相信會通人文與科技的數位人文學，除了可以「繼往」——驗證傳統人力可及的研究議題與成果外；更可以「開來」——開展出傳統人力所不及之研究議題，而以一种超越傳統研究視域與方法，獲致更豐碩、更深入、更具開展性、前瞻性的議題與成果，甚至可能建立新的人文學研究典範，促成「從人文到數位人文」之人文研究範式的移位與轉向。期盼在不久的將來，臺灣成為數位人文學的發展重鎮，也希望有更多研究人才共同投入這個新興的研究領域。