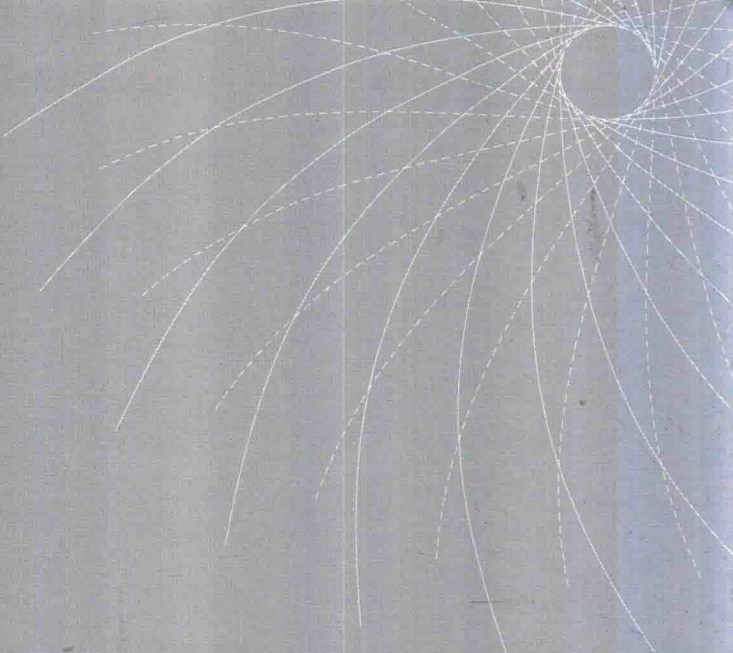




數位人文研究叢書 5
Series on Digital Humanities

數位人文研究 與技藝

項潔 編

Digital Humanities and Craft: Technological Change

數位人文研究叢書 5
Series on Digital Humanities

數位人文研究 與技藝

項潔 編

**Digital Humanities
and Craft:**
Technological
Change

目 錄

Contents

序

Preface

導論 Introduction

009 數位人文——學科對話與融合的新領域

An Interdisciplinary Dialogue through Digital Humanities

◆ 項潔、陳麗華

025 數位人文學之可能性及限制——一個歷史學者的觀察

The Possibilities and Limitations of Digital Humanities: A Historian's Perspective

◆ 王汎森

Part I 知識組織

Knowledge Organization

039 數位典藏的知識組織系統之模式建構與應用研究

——以故宮中國節慶詞彙為例

A Study of Knowledge Organization System for Digital Archives in the National Palace Museum Collections of "Chinese Festivals"

◆ 陳淑君、郭巧蓁

067 資料擷取與描述比對——清代臺灣方志物產篇分析系統之建置與解釋

Data Extraction and Mapping: The Taxonomy of Fauna and Flora of Taiwan Local Gazetteers in the Qing Dynasty

◆ 張素玢、李鈺淳

Part II 文本探勘

Text Mining

- 093 利用文本採礦探討《紅樓夢》的後40回作者爭議
A Text-Mining Approach to the Authorship Attribution Problem of *Dream of the Red Chamber*
◆ 杜協昌
-
- 121 漢文文獻之外來語音譯詞擷取方法
Transliteration Extraction Methods for the 19th Century Chinese Literature
◆ 王昱鈞、呂翊瑄、蔡宗翰、劉青峰、金觀濤、劉昭麟
-
- 139 自動擷取中文典籍中人名之嘗試
——以PMI (Pointwise Mutual Information) 斷詞於《資治通鑑》的應用為例
Automated Name-Extraction in Chinese Classics: Applying PMI (Pointwise Mutual Information) Segmentation to *Zizhi Tongjian*
◆ 彭維謙、劉士綱、杜協昌、翁稷安、項潔
-

Part III 語料分析

Corpus Linguistics

- 167 藏漢佛教語料品目之自動對列
Automatic Chapter-Level Alignment for Tibetan and Chinese Buddhist Texts
◆ 陳光華、闕慧貞、李家名、唐國銘、黃乾綱
-
- 193 句本位語法圖解析句系統的設計與實現
The Design and Implement of Diagrammatical Sentence-Based Grammar Parsing System
◆ 彭煒明、何靜、宋繼華
-
- 211 漢字構形系統的網絡分析——以《說文解字》小篆構形系統為例
Network Analysis of Chinese Ideography: The Formation of Seal Script in *Shuowen Jiezi*
◆ 胡佳佳
-

數位人文研究叢書 5
Series on Digital Humanities

數位人文研究 與技藝

項潔 編

Digital Humanities and Craft: Technological Change



序

從全球各國數位典藏發展的歷史來看，臺灣開始的不算太早，但也不算晚。暫且不論個別學校、單位或個人的研究和努力，從國科會在 1998 年以「數位博物館計畫」開始投入國家資源，有系統地發展數位典藏來算，到現在也已將近 15 年了。在這段不算短的時間裡，我國投入了大量經費，也將大量的文化資產數位化，其中最重要，從 2002 年開始執行的「數位典藏國家型計畫」，更產生了指標性的作用，讓「數位典藏」在臺灣成為一個大眾語言內的詞彙。

數位典藏國家型計畫的成功，至少有一部分歸功於許多資深且傑出的研究人員，尤其是人文學者，不計報酬地全心投入。但是從大約 2005 年開始，我們漸漸發覺這些優秀的學者在他們本身的研究中，似乎並沒有充分利用他們花了大量精力數位化的檔案；這令我感到困惑，因為這樣不是事倍功半嗎？為什麼不能將數位化工作和本身的研究結合在一起呢？觀察到這個現象後，我開始去探討它的環節。我發現其中的一個重要原因是數位典藏系統的設計往往沒有考慮到使用者——尤其是研究者——的需求，以致於一直到現在絕大多數的研究者還是認為檢索系統只能幫他們找資料罷了，而不能幫忙整理或分析資料。這是很可惜的，因為許多資訊技術已經十分成熟，如果能夠與數位檔案結合並有效地運用在這些系統中，應對人文研究產生非常大的助益。在更進一步的探討後，我發現這個問題並不是臺灣獨有的，在國外亦有學者思考這個問題，而且已有一個研究社群，那便是「數位人文」。

用最簡單的話來講，「數位人文」就是結合大量數位材料，運用資訊科技，來從事人文研究。顧名思義，這是一項跨領域整合的工作。要達到這個目標，除了要有大量高品質的數位資源可供使用外，更需要人文學者與資訊學者密切的互動與合作。有鑑於此，國立臺灣大學數位人文研究中心從 2009 年開始舉辦的「數位典藏與數位人文國際研討會」，希望藉由這個會議，不但能夠介紹世界最新的研究成果，並能提供國內外人文與資訊學者交流的機會，讓臺灣的數位人文研究能夠遍地開花，並與國際接軌。

《數位人文研究叢書》即是會議產出一項成品。叢書中的每一篇文章均在研討會中發表過，再經修改及至少兩位審查人的審查通過。在此特別感謝臺大數位人文研究中心的蔡炯民博士、陳怡君小姐與全體同仁對本叢書投注的心力。我們希望透過這個系列的叢書，提升國內學界對數位人文的認知，並激發進一步的研究。

項潔

2011 年 9 月於臺大

Preface

Taiwan started its cultural digitization effort in earnest in 1998, when the National Science Council initiated the Digital Museum Program. Since then the government has invested a large amount of money and has digitized a significant portion of Taiwan's cultural heritage. Indeed, the effort has been so successful that 數位典藏, the Chinese translation of "digital archives", has become a household term in Taiwan.

The success of the digitization enterprise, in particular the National Digital Archives Program, is at least partly due to the altruistic devotion of many senior scholars. However, around 2005 I started to notice that many of these scholars do not seem to utilize the outcome of their digitization work in their own research. This phenomenon baffled me. Since an important purpose of digitization is to make materials more accessible, why couldn't they use the fruits of their own hard work to expedite their own research? After some studies, I observed that at least one reason is that the digital archive systems are not designed with the need of the users, especially researchers, in mind. Thus most people still think that a retrieval system can do just that, retrieve, but not to help them organize, observe, and explore what has been retrieved. This is a pity, because combining information technology with massive high quality digital objects should provide tremendous opportunities for humanities research. I then learned that this phenomenon is not unique to Taiwan. Indeed, many scholars in the world have been thinking about this challenge, and the community is called Digital Humanities.

To put it in the simplest term, digital humanities is humanities research with the help of digital materials and information technology, with an emphasis on the type of research that cannot be accomplished otherwise. It is interdisciplinary by nature. In 2009, the Research Center for Digital Humanities of the National Taiwan University started an annual International Conference on Digital Archives and Digital Humanities to serve as a forum for researchers from different disciplines and different parts of the world to showcase their work and exchange ideas. This series, the Series on Digital Humanities, is a product of this effort. All papers in the volumes have been presented in one of the conferences. The final version is then submitted and went through rigorous reviews. We hope that the Series will promote digital humanities and stimulate further research.

Jieh Hsiang

September, 2011

National Taiwan University

目 錄

Contents

序

Preface

導論 Introduction

009 數位人文——學科對話與融合的新領域

An Interdisciplinary Dialogue through Digital Humanities

◆ 項潔、陳麗華

025 數位人文學之可能性及限制——一個歷史學者的觀察

The Possibilities and Limitations of Digital Humanities: A Historian's Perspective

◆ 王汎森

Part I 知識組織

Knowledge Organization

039 數位典藏的知識組織系統之模式建構與應用研究

——以故宮中國節慶詞彙為例

A Study of Knowledge Organization System for Digital Archives in the National Palace Museum Collections of "Chinese Festivals"

◆ 陳淑君、郭巧蓁

067 資料擷取與描述比對——清代臺灣方志物產篇分析系統之建置與解釋

Data Extraction and Mapping: The Taxonomy of Fauna and Flora of Taiwan Local Gazetteers in the Qing Dynasty

◆ 張素玢、李鈺淳

Part II 文本探勘

Text Mining

- 093 利用文本採礦探討《紅樓夢》的後40回作者爭議
A Text-Mining Approach to the Authorship Attribution Problem of *Dream of the Red Chamber*
◆ 杜協昌
-
- 121 漢文文獻之外來語音譯詞擷取方法
Transliteration Extraction Methods for the 19th Century Chinese Literature
◆ 王昱鈞、呂翊瑄、蔡宗翰、劉青峰、金觀濤、劉昭麟
-
- 139 自動擷取中文典籍中人名之嘗試
——以PMI (Pointwise Mutual Information) 斷詞於《資治通鑑》的應用為例
Automated Name-Extraction in Chinese Classics: Applying PMI (Pointwise Mutual Information) Segmentation to *Zizhi Tongjian*
◆ 彭維謙、劉士綱、杜協昌、翁稷安、項潔
-

Part III 語料分析

Corpus Linguistics

- 167 藏漢佛教語料品目之自動對列
Automatic Chapter-Level Alignment for Tibetan and Chinese Buddhist Texts
◆ 陳光華、闕慧貞、李家名、唐國銘、黃乾綱
-
- 193 句本位語法圖解析句系統的設計與實現
The Design and Implement of Diagrammatical Sentence-Based Grammar Parsing System
◆ 彭煒明、何靜、宋繼華
-
- 211 漢字構形系統的網絡分析——以《說文解字》小篆構形系統為例
Network Analysis of Chinese Ideography: The Formation of Seal Script in *Shuowen Jiezi*
◆ 胡佳佳
-

導論

Introduction

■ 數位人文——學科對話與融合的新領域

An Interdisciplinary Dialogue through Digital Humanities

■ 數位人文學之可能性及限制——一個歷史學者的觀察

The Possibilities and Limitations of Digital Humanities:
A Historian's Perspective



數位人文——學科對話與融合的新領域

項潔*、陳麗華**

摘要

自 1990 年代以來，數位人文在國際及在臺灣均成為引人注目的新興研究領域。2011 年之後，伴隨學術社群的變化及學科規範的建立，西方圍繞數位人文的認識論問題引發了一場爭論：一種說法認為應將數位人文視為人文研究的新理論方法；另一種說法則認為應將焦點放在製造數位工具或方法上，而不強調解釋的效用。本文的目的是透過對於本書相關研究論文的評述，結合筆者之一十幾年來從事數位人文研究的經驗，提出對於這場論爭的理解。在我們看來，二分法並不能概括這個流動領域中思想交換和轉變的過程，但理論性實作和實證性人文的結合，將有助於數位人文成為新研究範式衍生之地。

* 國立臺灣大學資訊工程學系特聘教授。

** 國立臺灣大學數位人文研究中心博士後研究員。

An Interdisciplinary Dialogue through Digital Humanities

Jieh Hsiang*, Li-hua Chen**

Abstract

Digital Humanities is rapidly emerging as a most exciting new research field in the humanities. With the emergence of academic attentions and new research communities, a debate about the epistemology of DH was formed. While some consider DH as a new approach to humanities research, others emphasize on prototyping and tool building, with less regard to the narrative nature of humanities.

In this introductory article, we provide, through the papers presented in this book, our view on this controversy. In our opinion, this dichotomy does not reflect the process of the exchange and transformation of thoughts. The combination of implementation (prototyping) of a theory and its validation through empirical studies in the humanities might hold the key to the paradigm shift.

* Distinguished Professor, Department of Computer Science and Information Engineering, National Taiwan University.

** Postdoctoral Researcher, Research Center for Digital Humanities, National Taiwan University.

一、前言：「數位人文」的發展及論爭

作為臺灣大學數位人文研究中心的成員，筆者之一曾經無數次面對這樣的問題，「到底什麼是數位人文？」對這本書感興趣的讀者，這個問題也一定在頭腦中縈繞。

這不是一個容易回答的問題。事實上至今為止，學術界對於到底什麼是「數位人文」，並沒有一個統一的答案。（項潔、涂豐恩，2011：11-26）在學術社群逐漸形成的共識當中，數位人文是結合數位資料，運用資訊科技來從事人文研究的新興領域。但問題是「數位」科技日新月異，層次多元，「人文」研究本身亦涵蓋極廣，難以一言以蔽之。不同學科訓練的學者，對它有著相當歧義的想像：人文學者可能認為這是人文學科利用新的資訊技術、網路和多媒體等工具，更快速而有效地處理人文問題的領域；資訊學家則可能認為是其成員在向人文學科領域拓展的過程中，由輔助性工作轉化為智識性活動的結果；近年來則有愈來愈多的學者，將數位媒介本身當作了其人文研究的對象。（Berry, 2012: 3-4）

回顧一下數位人文的歷史，或許對理解它有幫助。數位人文萌發的土壤，可以追溯到 20 世紀中葉以來，伴隨計算機的發明和應用逐漸興起的「人文計算」學。早期參與者多半擁有人文學科的學位，從事當時尚小眾的研究，比較文學、計量史學、古典研究、語言學等都曾經和它走得很近。（Ramsay & Rockwell, 2012: 75）然而 1990 年代以來，網際網路技術的快速發展，亦引致資訊技術和人文學科的關係發生了難以想像的變化，「數位人文」多用來指稱這一波之後的變化。這個新名詞的衍生被追溯至 2004 年，當時美國出版的一本回顧「人文計算」幾十年來發展歷程的權威性著作，為了吸引更多讀者的注意而在書名便採用了這個詞，很快受到熱烈歡迎，甚至推動了一個領域的建構，它涵化了之前已經存在的人文計算組織，亦催生出許多新的計畫和組織。（Schreibman, Siemens & Unsworth 2004; Fitzpatrick, 2012: 12）2005 至 2007 年，兩個數位人文的跨國性組織亦相繼成立，即「數位人文研究聯盟」（The Alliance of Digital Humanities Organizations）與「數位人文中心網絡」（centerNet），後者至 2012 年已有來自 20 個國家的 140 多個數位人文中心的會員。（Fraistat, 2012: 287）¹

臺灣近年也逐漸開始採用「數位人文」一稱，整合十幾年來參與相關研究的學術社群。2002 年起臺灣在國家力量的支持下，展開了持續 11 年的數位典藏國家型計畫，在國際上可謂前瞻作為。密切參與這一計畫的研究機構和中心，在結束前後亦逐漸轉型。2008 年臺灣大學設立的「數位人文研究中心」，便是其中最早的專門

1 可參見其網頁：<http://digitalhumanities.org/centernet/about/>

機構²。2013年成立的「中研院數位文化中心」，是國家型數位典藏計畫退場之後的延續性機構。³此外，政治大學頂大計畫「數位人文中心」、中興大學「中臺灣數位中心」計畫、法鼓山佛教學院亦有相關計畫。這些新的發展，與2012年起國科會開始推動「數位人文」主題研究計畫亦有很大關係，但亦有部分在此之前已經多年參與數位典藏的相關工作。

不難看出，近十餘年「數位人文」一詞已被迅速地擴展和挪用，變成一個包容性極強的涵蓋性詞彙。新生力量的湧入改變了學術社群原本的生態，其領域邊界和成員資格如何界定問題更顯紛繁。2011年之後，許多西方學者捲入了一場「數位人文學者到底意味著什麼」的爭論，讓潛藏的矛盾公開化。當時爭論由內布拉斯加大學人文數位研究中心的Stephan Ramsay發起，他反對沒有門檻，任何人均可宣稱自己是數位人文學者。在他看來，一定要「建造」東西（如檔案、程式、系統等），才有資格成為其中一員。（Gold, 2012: x-xi）這一觀點引起很多爭議和反彈，今日討論亦未止息。

這一爭議不僅反映了研究取向的差異，更反映了深層認識論上的衝突。在部分學者看來，整個20世紀大部分時間，是知識在各種「主義」框架下被重新架構的時代。這一取向將19世紀末20世紀初，學者們在組織知識和建構學科的過程中曾極為重視的語言文字學、詞彙學、書目文獻學等，逐漸排擠出了嚴肅的學術殿堂之外。而今數位技術的興起，則是向這一被湮滅的方法論傳統的回歸，亦是學術潮流擺脫意識形態的宰制，學者重新組織知識及自身工作的契機。（Scheinfeldt, 2012: 124-125）然而大部分人文學者無法容忍詮釋活動地位的淪喪，他們的焦慮反應在如下的一系列問題上：數位人文中的技術方法，如何以及在怎樣的基礎上對人文發展有幫助？除了更快、更容易地處理大量的資料外，數位方法是否只是印證我們已有的想法？數位人文是否要將人文轉化成「能實證」和「能證偽」的實證科學？資訊工具能經由詮釋上的操作，深刻地改變、或甚至改造人文學科嗎？

臺灣社會對於這些爭論的感覺，並沒有西方那麼強烈。這一方面與臺灣數位人文尚未學科化、體制化有關，另一方面也與尚未完全脫離前一階段的典藏思維有關。本文的目的當然不是要建構某種類型的數位人文或劃定其邊界，而是試圖透過對於本書相關研究論文的評述，結合我們十餘年來參與數位人文研究的經驗，與大家分享對於數位人文研究的思考。在我們看來，二分法並不能概括這個流動領域中思想

2 臺灣大學數位人文研究中心的前身為「臺灣大學數位典藏研究發展中心」，2012年正式改為今名。但中心的英文名稱自始即為Research Center for Digital Humanities。可參見其網頁：<http://www.digital.ntu.edu.tw/>

3 <http://www.asdc.sinica.edu.tw/about.htm>

交換和轉變的過程，但理論性實作和實證性人文的結合，無疑有助於新研究範式的產生。

二、數位人文作為跨學科的「貿易區」

在西方有關數位人文的爭議中，有學者主張它應該成為不同學科的「貿易區」和「聚會所」。「貿易區」一詞暗示了市場和交換，這個詞彙曾被哈佛大學科學史系的 Peter L. Galison 所使用，借以分析物理學中屬於不同認識論傳統的範式所進行的合作研究。學者遂借來延伸到數位人文領域，強調其是跨學科工作進行的場所，一方面可以延續其不同的傳統，另一方面則鼓勵不同形式的深層合作。（Svensson, 2012: 45-46）

就臺灣而言，從早期人文計算的歷史開始討論的話，跨學科的交流亦行之有年。其中與計算機資訊領域交流最為密切的人文科系，大概首推圖書館學及語言學。1960 至 1970 年代，臺灣有關計算機的研究機構和系所開始設立，1964 年圖書館便開始用電腦處理西文資料的編目，而後擴展至中文領域，1980 至 1990 年代，隨著電腦科技和網路的發展，圖書館與資訊科學的交流合作亦日趨頻繁。（何揚，1981；李德竹，1993）在 1970 年代末至 1980 年代，資訊專家開始處理如何讓電腦讀中文字的問題，並開始和語言學家密切互動，由此在臺灣開拓了計算語言學領域，曾任職中研院的謝清俊、黃居仁等便是代表人物。（蘇克毅，1993）1980 年代中葉，謝清俊亦開始於中研院人文研究領域推動古籍的數位化。

而筆者之一參與數位人文研究的經歷，也是推動資訊工程學與人文學科合作的一個過程。1995 年時，因為對數位圖書館的概念充滿興趣，筆者之一便和臺大資訊工程學系同事與圖書館學系（幾年後更名為圖書資訊學系）教授定期聚會討論，當時國際上這一領域才剛剛開始起步，而校內氛圍亦鼓勵跨院的合作，因此次年便開始推動「臺灣大學數位圖書館計畫」，當時參與的人包括了圖書館館長吳明德、歷史學系吳密察、人類學系胡家瑜、圖書館學系陳雪華，以及資訊工程學系項潔和歐陽彥正等。這個跨學科的團隊包含了熟稔資料內容與分析的人文學者，了解圖書組織與分類法則的圖書學者，還有提供技術工具與建立系統的資訊學者，是極為特別的組合。這一計畫和後來 1998 年國科會開始推動的「數位博物館專案計畫」，乃至後來「數位典藏與數位學習國家型科技計畫」之間也有延續性關係。（王汎森，2013: xi-xii）現在較為人知的「臺灣歷史數位圖書館」（THDL）⁴，它在 2006 年完成，2007 年開放線上使用，其中總共包含了近九萬件和臺灣有關的官方文書和民間契約等，是目前臺灣史研究資源最為豐富的資料庫，亦是這一團隊多年合作的成果之一。

4 <http://thdl.ntu.edu.tw/>